This paper was a finalist for the Robert F Wagner F. All Conference Best Student Paper Award

MPWC-Net++: Evolution of Optical Flow Pyramidal Convolutional Neural Network for Ultrasound Elastography

Ali K. Z. Tehrani[a,b] and Hassan Rivaz[a,b]

[a]Electrical and Computer Eng., Concordia University, Montreal, Canada
[b]PERFORM Centre, Concordia University, Montreal, Canada

ABSTRACT

Recently, Convolutional Neural Networks (CNNs) have been very successful in optical flow estimation in computer vision. UltraSound Elastography (USE) displacement estimation step can be performed by optical flow CNNs. However, there is a large domain gap between ultrasound Radio-Frequency (RF) data and computer vision images which reduces the overall accuracy of displacement estimation. Some modifications of the network architecture are required to be able to extract reliable information from RF data. Modified Pyramidal Network (MPWC-Net) which is based on the well-known PWC-Net was among the first attempts that adopts the optical flow CNNs to USE displacement estimation. However, MPWC-Net suffers from several shortcomings that limit its application especially for unsupervised training. In this paper, we propose additional modifications to substantially improve MPWC-Net. We also publicly released the network's trained weights.

Keywords: Ultrasound Elastography, Optical Flow, Pyramidal Convolutional Neural Network.

1. INTRODUCTION

UltraSound Elastography (USE) is an imaging technique which provides relative stiffness properties of the tissue. It has been successfully used for image guided surgery.[1] Free-hand palpation is one of the simplest and most popular method in USE. In free-hand palpation, the operator compresses the tissue by the probe. The images of the tissue before and after compression are compared to obtain displacement map which indicates movement of each sample in axial and lateral direction.[2, 3]

In order to obtain the displacement map, optical flow methods used in computer vision applications could be utilized. CNNs have been successful in many computer vision applications including optical flow. Recently, many CNN architectures such as FlowNet,[4] FlowNet2,[5] PWC-Net[6] and LiteFlowNet[7] have been introduced.

A growing number of researchers have adopted optical flow CNNs for USE.[8, 9] These architectures are optimized for computer vision images and there is a large domain gap between computer vision images and USE.[10, 11] To reduce the domain gap, many researchers tried to fine-tune the network by ultrasound images.[9, 10, 12] However, their improvements were not significant. We showed that the architecture has larger impact compared to simple fine-tuning.[10] We proposed that by modifying the structure of the network, substantial improvements can be achieved. We modified PWC-Net by removing the strides to allow the network to use high frequency Radio-Frequency (RF) data. We named our architecture as MPWC-Net.

MPWC-Net modification is straightforward and does not require re-training the network. However, this simplicity is not without any cost. In this paper, we introduce these problems and try to mitigate them by some additional modifications. We name the new modified network, MPWC-Net++, which is based on MPWC-Net with additional changes to facilitate use of optical flow CNNs in USE. We publicly made available the trained network weights at code.sonography.ai.

2. METHOD

Before introducing shortcomings of MPWC-Net, we need to briefly review the PWC-Net structure.

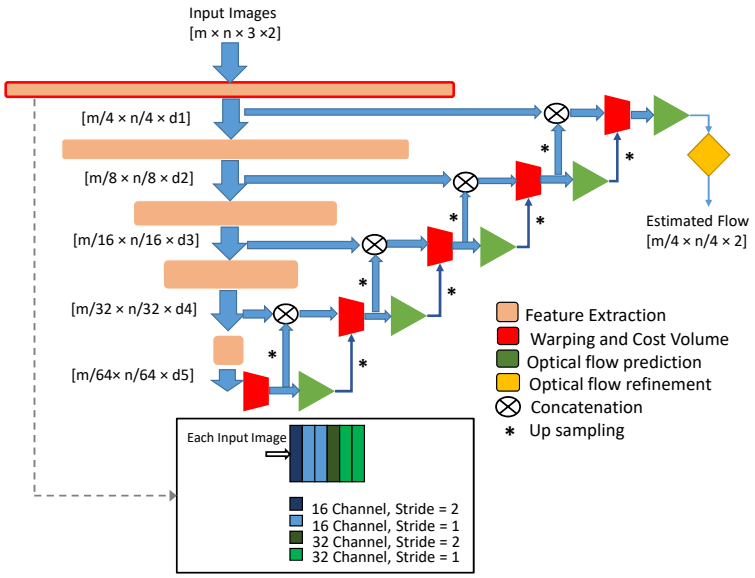

Figure 1. PWC-Net architecture. Reproduced from[10] with permission. In iterative residual refinement variant,[13] cost volume and optical flow prediction layers have shared weights.

2.1 PWC-Net

This network is based on pyramidal estimation of optical flow. In each level, the displacement is estimated and features of the second image are warped with this displacement toward the first image to reduce the displacement range to be estimated in higher levels. In each pyramid level, feature extraction, cost volume, warping and optical flow estimation are used. Cost volume performs correlation between two feature maps, which has been shown to be more informative compared to feature maps alone.[5] Each cost volume of each level gives correlation of limited search range (in PWC-Net it is 4). The basic architecture of PWC-Net is shown in Fig. 1. As we can see some of the blocks such as cost volume and optical flow prediction are repeated for each pyramid level. Recently, a variant of PWC-Net called iterative residual refinement PWC-Net (PWC-Net-irr) was proposed which used shared cost volume and optical flow prediction blocks.[13] By this modification, the required number of parameters were reduced and accuracy was also improved due to having less over-fitting.

2.2 MPWC-Net

RF data, which has high frequency contents, is required for accurate displacement estimation in USE. However, the input images are downsampled by a factor of 4 at the first feature extraction blocks in PWC-Net. This leads to a loss of information and degrades the strain image quality. Recently, we proposed to remove the first two downsampling by replacing stride=2 with stride=1 such that reliable feature can be extracted by the feature extractor blocks.[10] Furthermore, we proposed to use B-mode and envelope alongside RF data to avoid unreliable displacement estimation in coarse levels where RF data does not provide reliable information.

2.2.1 Limitations of MPWC-Net

In PWC-Net, each pyramid level is responsible to estimate the displacement in its range and then remove this displacement by warping the feature map of the second image toward the first one. This facilitates the task of the following pyramid level since it requires only to estimate the residual displacement. This pyramidal structure also allows to limit the search range of the cost volumes to further reduce the computation complexity. Let c, M and S be the cost volume search range, the number of pyramid levels and rate of downsampling for the first pyramid level, respectively. The higher bound of the maximum displacement (D_m) that PWC-Net detects is:

$$D_m = \sum_{i=0}^{M-1} 2^i \times S \times c \tag{1}$$

In the original PWC-Net, $c = 4$, $M = 5$ and $S = 4$ which results in the maximum displacement of 496. In MPWC-Net, $S = 1$ hence the higher bound of the maximum displacement would be 124 pixels which limits the application of MPWC-Net to small displacements. This is a limiting factor especially or large images and unsupervised training where the ground truth and maximum displacement is unknown.[11, 14]

The optical flow networks are mostly trained on Flyingchair dataset and fine-tuned on other available datasets such as Sintel.[5, 6] Flyingchair is a synthetic dataset produced for training the networks which its displacement range and distribution mimic Sintel dataset. The maximum displacement of Sintel is around 450[4] and the histogram is shown in Fig. 2. We can see that only a few training data have displacements more than 310 and most of maximum displacements are below 200. The network only sees very limited training data that have displacements close to the higher bounds of PWC-Net (496 pixels). This means that in large displacements, the network cannot perform as well as smaller displacements. This is another factor that further limits the displacement range. The practical displacement range of PWC-Net and MPWC-Net are lower than 496 and 126, respectively.

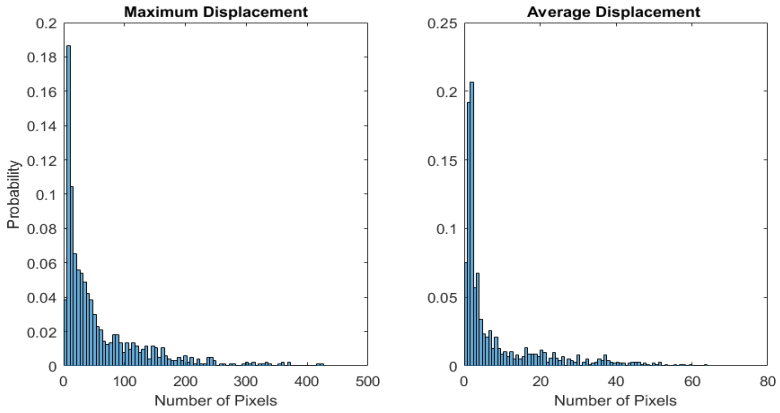

Figure 2. Histogram of maximum displacement and average displacement of Sintel training dataset.

Another limiting factor is the input image size. In original PWC-Net, the image is downsampled by a factor of 4 in the first two feature extraction convolution blocks. Whereas, in MPWC-Net there is no downsamplping; therefore, there are 16 times more pixels in train and inference. The much higher image and feature size results in significant reduction of speed in both training and test. The inference times are reported in section 3.3.

2.3 MPWC-Net++

In order to solve the aforementioned problems, we proposed to modify MPWC-Net structure. We used PWC-Net-irr instead of original PWC-Net structure to reduce the number of learning weights.

We also increased the maximum displacement range by changing values of c and S. Different values of these parameters are given in Table 1. The new structure must have a search range close to the Sintel and Flyingchair since it would be trained by these datasets and if it cannot estimate large displacement values, the incorrect loss would be propagated during the training. As shown in Table 1, case 3 has a good balance between maximum displacement and the downsampling factor and can track a large portion of large displacements.

According to the selected value of S, we needed to decrease the S from 4 to 2. One of the first two feature extraction blocks needed to have a stride of 2 and the other had the stride of 1 ($S : 2 \times 1 = 2$). We selected the first feature extraction layer and the second one to have stride of 2 and 1, receptively. By this setting, smaller image size is fed to the second feature extraction layer and the computation time is improved. While, this leads to loss of information especially in axial direction which has a high frequency content. To avoid this loss of information, we also changed the kernel size of first feature extraction layer from 3×3 to 5×3 to capture more information in axial direction.

Another modification is that we replaced the input B-mode image with the imaginary part of the RF data which has been shown to improve the performance.[15]

Table 1. Maximum search range (D_m) for different c and S.

Model	c	S	D_m
PWC-Net	4	4	496
MPWC-Net	4	1	126
case 1	4	2	252
case 2	5	1	155
case 3 (MPWC-Net++)	5	2	310
case 4	6	2	372

Table 2. The validation EPE of the networks trained on FlyingChairOcc.; FlowNet2 and PWC-Net are reported from.[13]

Network	EPE of FlyingChairOcc
FlowNet2	2.39
PWC-Net	1.89
FlowNet2 (irr)	2.22
PWC-Net (irr)	1.83
MPWC-Net++ (irr)	1.51

3. RESULTS

3.1 Training and Results on Validation Set

Due to the change of parameters, we needed to train MPWC-Net++ from scratch. We followed the same training schedule of PWC-Net-irr.[13] The decrease of stride (S) and having higher search range (c) had substantial impact on training time; therefore, we decreased the number of epochs from 220 to 180. The network was trained using FlyingChairsOcc dataset.[13] The dataset is composed of 22870 image pairs with forward and backward ground truth flows. The displacement distribution is the same as original FlyingChair dataset. The network was trained for 10 days (due to larger search range and reduced stride it took longer time than PWC-Net-irr).

We also investigated the error on the validation set of the mentioned dataset. We used End Point Error (EPE) which is defined as:

$$EPE = \sqrt{(\widetilde{U_x} - U_x)^2 + (\widetilde{V_x} - V_x)^2} \tag{2}$$

where $\widetilde{U_x}$, U_x, $\widetilde{V_x}$, V_x denote estimated and ground truth displacement in axial and lateral, receptively. The EPE of validation set of FlyingChairsOcc for different architecture are given in Table 2. Using the blocks iteratively (irr) results in a modest improvement in both FlowNet2 (2.22 *vs* 2.39) and PWC-Net (1.83 *vs* 1.89). MPWC-Net++ results in substantial improvements compared to PWC-Net with only a few small modifications. After investigating the results, we found the main reason of the improvement is that the stride (S) is reduced from 4 to 2. This leads to larger input images in each pyramid level and more accurate features can be extracted from the images even for computer vision images.

3.1.1 USE Results

The problem with MPWC-Net is that it has high inference time due to large size of image and cannot handle large displacements. We present simulation and experimental phantom results for large and small displacements. The quantitative metrics used for the experimental phantom evaluation are CNR and SR. These metrics require a background and target window to be computed and they are defined as follows:

$$SR = \frac{\bar{\varepsilon}_t}{\bar{\varepsilon}_b}, \qquad CNR = \sqrt{\frac{2(\bar{\varepsilon}_b - \bar{\varepsilon}_t)^2}{{\sigma_b}^2 + {\sigma_t}^2}}, \tag{3}$$

where $\bar{\varepsilon}_t$ and $\bar{s}_b$ are mean values of strain in the target and background windows. The σ_t and σ_b are variance values of the strain images in the target and background windows, respectively. It should be mentioned that to

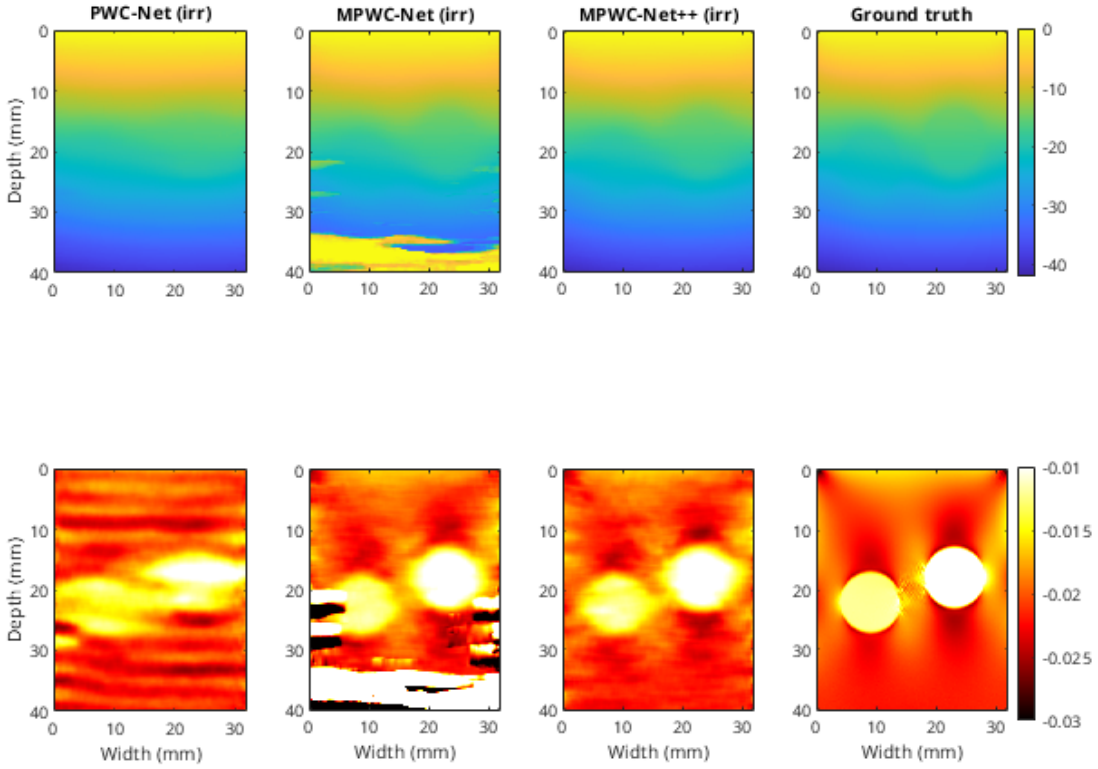

Figure 3. Axial displacement (top) and strain images (bottom) of MPWC-Net, PWC-Net, MPWC-Net++ and ground truth.

compute the SR and CNR, several small windows inside the selected region are used to compute them to obtain more reliable estimates. The mean and standard deviation of the results are reported. It is important to note that CNR is sensitive to both bias and variance error. Whereas, SR is only sensitive to bias error hence it is a suitable metric for bias error evaluation.[10, 11]

3.1.2 Simulation Results

A simulation data with displacement range of 0 to -42 samples is chosen to compare the methods. The results are shown in Fig 3. We can see that MPWC-Net has outlier regions where the displacement is high. It should be noted that this network fails in displacement values much lower than the computed higher bound (it fails around 35 which is much lower than 126). PWC-Net does not fail but the strain image quality is low due to a limited use of the high frequency information. MPWC-Net++ results in the strain image as good as MPWC-Net without any outlier region.

3.1.3 Experimental Phantom Data

Phantom data has been collected at Concordia University's PERFORM Centre by an E-Cube R12 research ultrasound machine (Alpinion, Bothell, WA, USA). We used the prob with 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) was used. The phantom has tissue elasticity of $15-25kPa$ and contains hard inclusions which have elasticity at least more than twice the elasticity of the tissue.

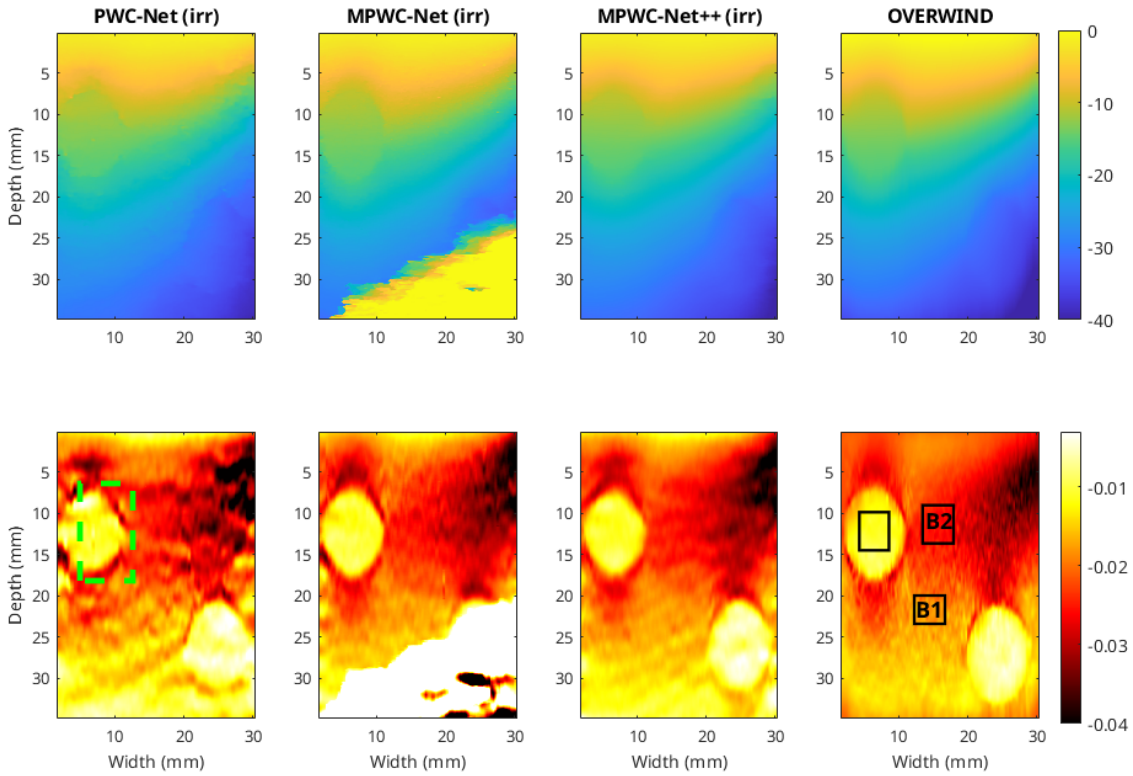

Figure 4. Axial displacement (top) and strain images (bottom) of PWC-Net, MPWC-Net, MPWC-Net++ and OVERWIND. Background 1 (B1), background 2 (B2) and the target regions are specified by black windows. The 3D view of the region highlighted by the green window is illustrated in Fig. 5.

PWC-Net, MPWC-Net and MPWC-Net++ are compared with OVERWIND which is considered as the state-of-the-art method in USE. OVERWIND is an optimization-based method that has been proposed by Mirzaei *et al.*[16] The results are shown in Fig. 4. As can be seen from the figure, PWC-Net results in low-quality strain. MPWC-Net produces higher quality strain image compared to PWC-Net but it fails when the displacement is high (bottom of the phantom). MPWC-Net++ produces high quality strain without failure in regions where the displacement is high. The strain quality of MPWC-Net++ is also comparable with OVERWIND.

The quantitative results are reported in Table 3. The regions for computation of CNR and SR are highlighted in Fig. 4 (black windows). We used smaller windows inside the specified regions to compute CNR and SR. The mean and standard deviation are reported. The CNR and SR are reported for one target region against two background regions. OVERWIND outperforms other methods in terms of CNR for background1 (comparing the target region with the background 1). MPWC-Net++ results in a CNR close to OVERWIND for background 1. Whereas, MPWC-Net++ has the highest CNR for background 2 (comparing the target region with the background 2). It is higher than OVERWIND by 1.66 dB and it is higher than other networks by a large margin. It is evident from the results that MPWC-Net++ not only fixes the low displacement range of MPWC-Net, but also improves the CNR and the overall quality of the strain image considerably.

Regarding SR, the lower value indicates lower bias.[10] MPWC-Net has the lowest SR value among the compared methods which represents that it has the lowest bias error. MPWC-Net++ has SR values slightly higher than MPWC-Net but the SR values are very close to MPWC-Net.

The 3D view of the region specified in Fig. 4 by the green window is shown in Fig. 5. It can be seen from the figure that OVERWIND and MPWC-Net++ have low heterogeneity in the specified region.

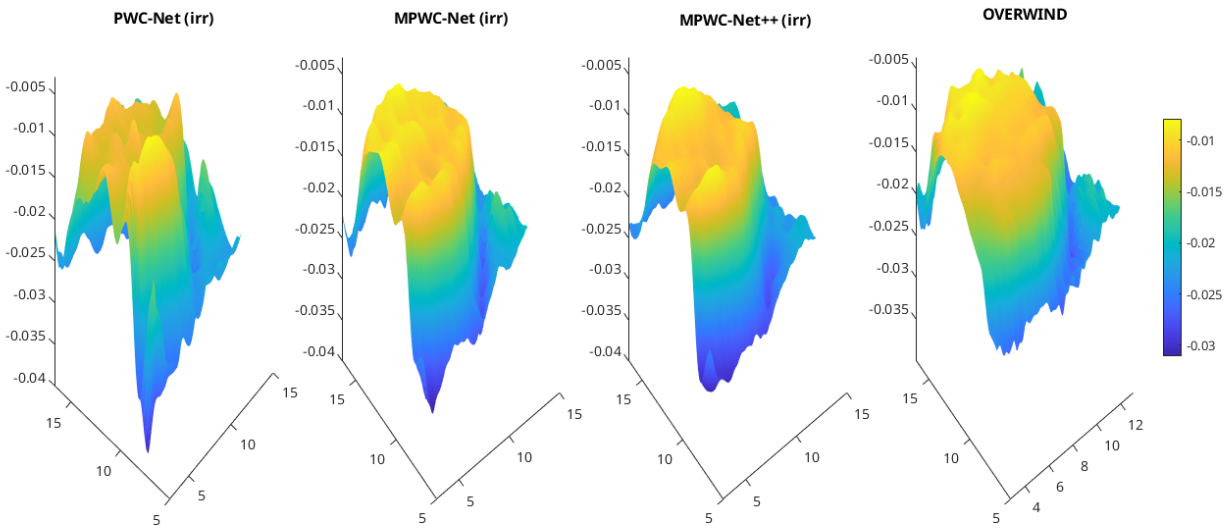

Figure 5. 3D side view of the strain of the region specified in the Fig. 4.

Table 3. Quantitative results of the experimental phantom comparing target with background 1 and background 2: mean±standard deviation.

Method	Background 1		Background 2	
	CNR	SR	CNR	SR
OVERWIND	**25.21**±4.82	0.413±0.019	12.86±1.85	0.639±0.022
PWC-Net	12.67±3.14	0.418±0.021	7.26±2.60	0.624±0.074
MPWC-Net	20.59±3.46	**0.392**±0.023	8.63±4.17	**0.526**±0.209
MPWC-Net++	24.99±2.67	0.425±0.014	**14.52**±2.72	0.632± 0.029

3.2 *In Vivo* Results

In vivo data was obtained at Johns Hopkins Hospital from the liver of a patient. A research Antares Siemens system with a VF 10-5 linear array having sample frequency of 40 MHz and center frequency of 6.67 MHz was used. The study was approved by the institutional review board with consent of all patients. For more information about the experimental details of the obtained data see.[17]

The results of the compared methods are given in Fig. 6. OVERWIND obtains the most visually appealing strain image compared to other methods. However, the parameters of this method need to be adjusted carefully. In fact, we spent some time to find the best parameters. OVERWIND optimizes data in a window. Due to the fact that the data is noisy and low-quality, we set a large window (33×7) to increase the robustness.

Considering the deep learning methods, they do not need any adjustment and they are plug-and-play which makes them suitable for real-time applications such as image-guided surgery. We can see that MPWC-Net obtains a high quality strain image but it fails in some noisy regions. MPWC-Net++ does not fail but the sharpness especially in the tumor boundaries is slightly degraded.

The quantitative results are presented in Table 4. The quantitative results confirm the visual assessments. MPWC-Net++ and OVERWIND both have high CNR. Similar to the experimental phantom results, the best

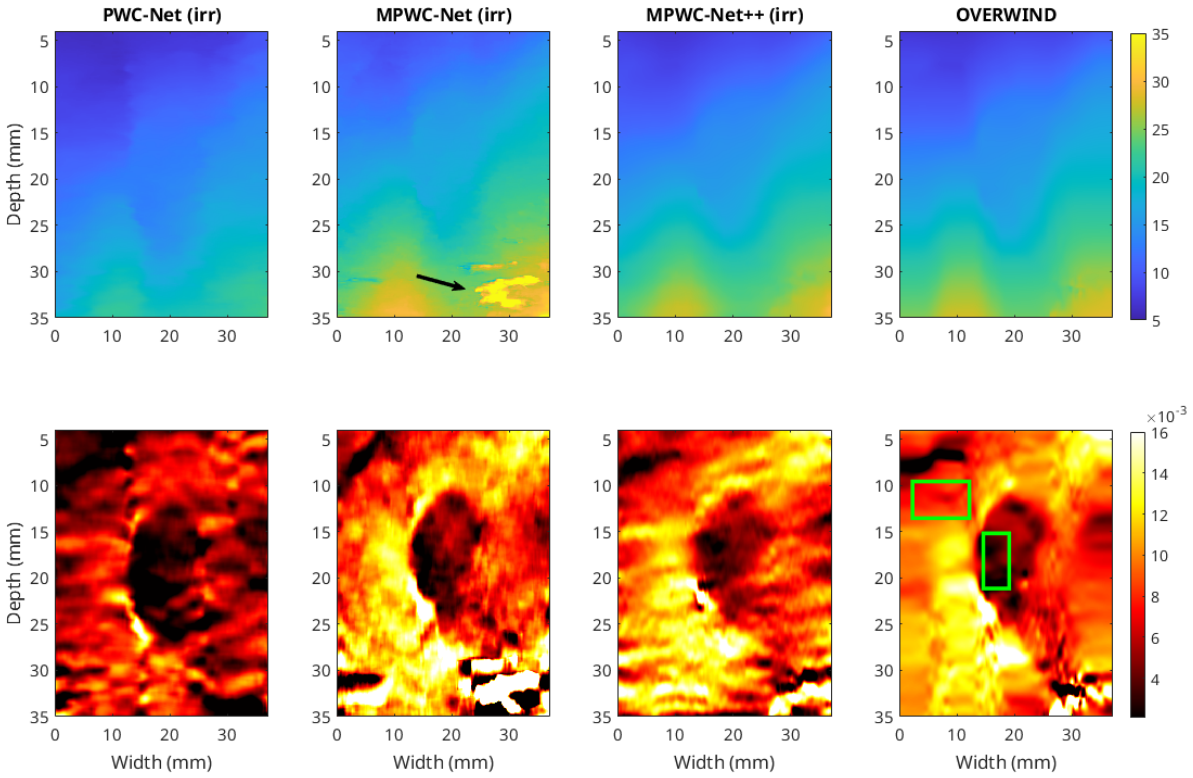

Figure 6. Axial displacement (top) and strain images (bottom) of PWC-Net, MPWC-Net, MPWC-Net++ and OVERWIND. Background and the target regions are specified by green windows. The failed region in MPWC-Net is marked by an arrow.

Table 4. Quantitative results of the experimental phantom comparing target with background: mean±standard deviation.

Method	CNR	SR
PWC-Net	3.94±2.63	0.630±0.255
MPWC-Net	6.31±1.47	**0.361**±0.071
MPWC-Net++	**7.85**±2.38	0.435±0.068
OVERWIND	7.40±1.46	0.394±0.098

SR is obtained from MPWC-Net which shows that this method has lowest bias error. The SR value of MPWC-Net++ is higher than SR value of MPWC-Net which means that it has higher bias.

3.3 Inference Time

Inference time is another important factor that needs to be evaluated. The inference time is crucial for real-time applications. We excluded OVERWIND due to the fact that it runs on CPU, whereas the other methods run on GPU. The inference times of single image pair of size 2048×256 are reported in Table 5. MPWC-Net++ is three times faster than MPWC-Net due to the fact that $S = 2$ instead of $S = 1$.

Table 5. Inference time of different architectures in ms.

Method	Time in ms
PWC-Net	42.28
MPWC-Net	1409.40
MPWC-Net++	495.20

4. CONCLUSIONS

In this paper, we presented a modification to MPWC-Net based on physics of ultrasound data and distribution of the computer vision training data. These modifications mitigated the limitations of MPWC-Net in estimating large displacements in USE and improved the inference time which is crucial for real-time applications such as image-guided surgery.

5. ACKNOWLEDGEMENT

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) RGPIN-2020-04612. The authors would like to thank Drs. E. Boctor, M. Choti and G. Hager for allowing us to use the *in vivo* data.

REFERENCES

[1] Billings, S., Deshmukh, N., Kang, H. J., Taylor, R., and Boctor, E. M., "System for robot-assisted real-time laparoscopic ultrasound elastography," in [*Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*], **8316**, 83161W, International Society for Optics and Photonics (2012).

[2] Ophir, J., Alam, S. K., Garra, B., Kallel, F., Konofagou, E., Krouskop, T., and Varghese, T., "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* **213**(3), 203–233 (1999).

[3] Sigrist, R. M., Liau, J., El Kaffas, A., Chammas, M. C., and Willmann, J. K., "Ultrasound elastography: review of techniques and clinical applications," *Theranostics* **7**(5), 1303 (2017).

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

[5] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and 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*], 2462–2470 (2017).

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

[7] Hui, T.-W., Tang, X., and Change Loy, C., "Liteflownet: A lightweight convolutional neural network for optical flow estimation," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 8981–8989 (2018).

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

[9] Peng, B., Xian, Y., Zhang, Q., and Jiang, J., "Neural network-based motion tracking for breast ultrasound strain elastography: An initial assessment of performance and feasibility," *Ultrasonic Imaging* , 0161734620902527 (2020).

[10] Tehrani, A. K. and Rivaz, H., "Displacement estimation in ultrasound elastography using pyramidal convolutional neural network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* (2020).

[11] Tehrani, A. K., Mirzaei, M., and Rivaz, H., "Semi-supervised training of optical flow convolutional neural networks in ultrasound elastography," *International conference on Medical Image Computing & Computer Assisted Intervention (MICCAI)* (2020).

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

[13] Hur, J. and Roth, S., "Iterative residual refinement for joint optical flow and occlusion estimation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 5754–5763 (2019).

[14] Delaunay, R., Hu, Y., and Vercauteren, T., "An unsupervised approach to ultrasound elastography with end-to-end strain regularisation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 573–582, Springer (2020).

[15] Tehrani, A. K., Amiri, M., and Rivaz, H., "Real-time and high quality ultrasound elastography using convolutional neural network by incorporating analytic signal," in [*42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society (EMBC)*], (2020).

[16] Mirzaei, M., Asif, A., and 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).

[17] Rivaz, H., Boctor, E. M., Choti, M. A., and Hager, G. D., "Real-time regularized ultrasound elastography," *IEEE transactions on medical imaging* **30**(4), 928–945 (2010).