Deep Learning Based Motion Tracking of Ultrasound Image Sequences

Skanda Bharadwaj, Mohamed Almekkawy
School of Electrical Engineering and Computer Science
The Pennsylvania State University, University Park
Pennsylvania, USA
{ssb248, mka9}@psu.edu

***Abstract*—Conventional motion estimation techniques in ultrasound images such as exhaustive search-based block matching (ES-BM) have been studied exhaustively and are known to be computationally expensive and slow. Consequently, they are not feasible for real-time processing. On the other hand, several deep learning-based techniques are being developed for real-time motion estimation of day-to-day objects. In this paper, we attempt to bridge the gap between tracking techniques being used for ultrasound images and recent deep learning-based techniques used for non-medical real-world objects. We propose to adopt the deep neural network-based Fully-Convolutional Siamese tracker (SiamFC) to track regions of interest (ROI) in ultrasound images. We prove that siamese architecture-based tracker is feasible for motion tracking in ultrasound images and performs better than conventional ES-BM technique. We applied SiamFC and ES-BM on 10 different image sequences to track the motion of the transverse section of the carotid artery. Our experiments showed that SiamFC was almost six times faster with slightly better performance compared to ES-BM in most of the cases.**

***Index Terms*—Speckle Tracking, Siamese Tracker, Convolutional Neural Network, Block Matching**

I. Introduction

Motion tracking has a large number of applications in diagnostic ultrasound. It is used in techniques such as elasticity imaging [1], elastography [2], and echocaridography [3]. In ultrasound images, speckle tracking is one of the well-known methods of motion tracking to assess the elastic properties and stiffness of soft tissue such as carotid artery. Several techniques including block matching [4], 2D - tracking using parabolic polynomial expansion with Riesz transform [5] and deep neural networks [6] have been proposed for speckle tracking. Among the many available techniques, ES-BM is one of the most commonly used speckle tracking techniques. Owing to its importance, a number of applications have been developed to improve the accuracy of motion estimation such as carotid artery wall motion tracking [7], subsample displacement estimation using kriging interpolation [8] and shear strain within the arterial wall [9].

Block matching techniques involve defining an ROI (or a reference block) in the first frame and tracking it through the entire image sequence. Two factors that play a vital role for motion tracking are search strategy and cost function for similarity matching. Typical block matching techniques use exhaustive search strategy to locate the reference block in the subsequent frames. Mean absolute difference (MAD), mean squared error (MSE) or normalized cross-correlation (NCC) are the commonly used cost functions to find the best candidate block (any block subject to search against reference block). Exhaustive search strategy, however, is bottleneck to achieve real-time processing speeds. In exhaustive search strategy, cost function is calculated for every possible candidate block, which makes the process computationally very expensive. In order to overcome the limitation of exhaustive search in the context of block matching, faster search algorithm has also been proposed [10].

Convolutional neural networks are slowly being adopted for motion tracking in medical imaging. Recently, Peng et al. [6] applied Flownet2.0 to ultrasound elastography. Typically, CNNs make use of feature maps to characterize the image. For tracking, CNNs can be used to obtain feature maps of both the object under consideration and the candidate blocks. Such netwroks are referred to as twin networks or Siamese networks. In this paper, we propose to adopt one such Siamese network-based architecture to track the motion of carotid artery. We adopt the Fully-Convolutional Siamese Networks (SiamFC) developed by Bertinetto et al [11]. Siamese neural networks are most popular among tasks that involve similarity matching. SiamFC uses a Siamese network to locate the reference block within a larger search region. SiamFC uses similarity learning approach to create a correlation map of the reference block and all possible candidate blocks within a predefined search region in a single evaluation. This helps improve the performance of the tracker. In addition, SiamFC assumes small and gradual changes of the ROI under consideration. Since motion of carotid artery is also gradual (also due to high frame rate), SiamFC can be used to track the motion of the carotid artery. Towards this end, our objective in this paper is to adopt SiamFC to prove that it has higher tracking accuracy and less computational time than the conventional ES-BM tracking technique.

II. Materials and Methods

A. Siamese Tracker

Siamese networks use same transformation function on two different inputs to compare the similarity between them. Siamese architectures formulate motion estimation as convolutional feature cross-correlation between a reference block and a search region. Fig. 1 represents the architecture of

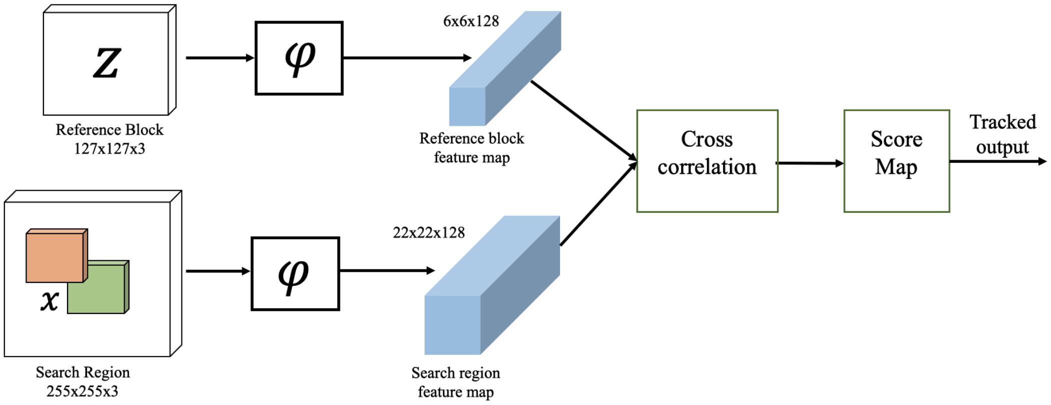


Fig. 1. Architecture of Fully-Convolutional Siamese Network – SiamFC

SiamFC. In the original paper, the authors address the problem of tracking an arbitrary object in a given video sequence. In other words, they aim to build a class-agnostic tracker. The authors propose to learn a function $f(z, x)$, that compares the reference block z to a candidate block x. If z and x depict the same object, the function outputs a high score and a low score otherwise. The network is trained in an initial offline phase and is then just evaluated during tracking. In SiamFC, the convolutional stage of the architecture resembles that of AlexNet [12], which is used as the embedding function (φ) for both the inputs. Fully-convolutional (does not contain dense layers) property of the architecture allows us to input images of different sizes as shown in Fig. 1. SiamFC attempts to find the reference block within a larger search region considering all translated sub-windows. Hence, the output of the network is a score map. The position of the maximum score relative to the center of the score map helps to find the target from frame to frame. Similarity learning approach is achieved by applying identical transformation φ to both the inputs and then combine their representations using a different function $f(z, x) = g(\varphi(z), \varphi(x))$.

The convolutional stage of SiamFC, is trained with non-medical image classes from ILSVRC dataset [13]. Since the architecture assumes small and gradual changes to the object scale, we propose to adopt this architecture directly to track the motion of the carotid artery and prove that it performs better than ES-BM. In our experiments, the ground truth ROI available from the first frame was considered as z and a predefined search region in subsequents frame around the ROI of the current frame was considered as the search region. The network then compares the reference block to candidate blocks x of the same size to create a correlation score map. The candidate block with highest score is considered as tracked output and as reference block for the next frame.

B. Dataset and Experiments

In this paper, B-mode image sequences of the perpendicular cuts of the carotid artery (cross-section view) obtained from [14]–[16] were used to validate the efficacy of SiamFC against ES-BM technique. 10 different image sequences from this dataset (henceforth called the CA dataset) were used to compare the two tracking techniques. Each image sequence consisted of $20-25$ frames. Ground truth of the arterial movement was known a priori. Fig. 2 represents one of the frames of the CA dataset along with the ROI. The red circle in the image represents the location of the blood vessel. Such ROIs were tracked in all 10 image sequences using both SiamFC and ES-BM. Ground truth ROI available from the first frame was given as input to both the trackers and were tracked through the rest of the sequence. For ES-BM tracker, as the name suggests, best matching candidate block was located using exhaustive search and normalized cross-correlation (NCC) was used as the similarity matching cost function.

Three different metrics were used to compare the two trackers. First, the Root Mean Squared Error (RMSE) of the centroid locations of the tracked ROIs obtained using SiamFC and ES-BM were compared with the centroid locations obtained from the ground truth. Second, Computational Time per Frame (CTF), defined as the total time taken by the tracker to process an entire frame, was compared between the two trackers. Lastly, intersection over Union (IoU) or bounding box overlapping intersection (an evaluation metric used to measure the accuracy of detection of a particular technique with respect to the known ground truth) of the two trackers were calculated with respect to the ground truth.

III. Results

As mentioned in the previous section, the CA dataset consisted of 10 different image sequences. For tracking, ground truth ROI from the first frame was given as input to both SiamFC and ES-BM. Fig. 3 represents the detections of SiamFC and ES-BM against the ground truth for a sample frame from the CA dataset. The two trackers were compared against each other based on the three aforementioned metrics.

A. RMSE of Centroids

Fig. 4a and 4b represent the centroid locations obtained from ground truth, ES-BM and SiamFC in both axial and

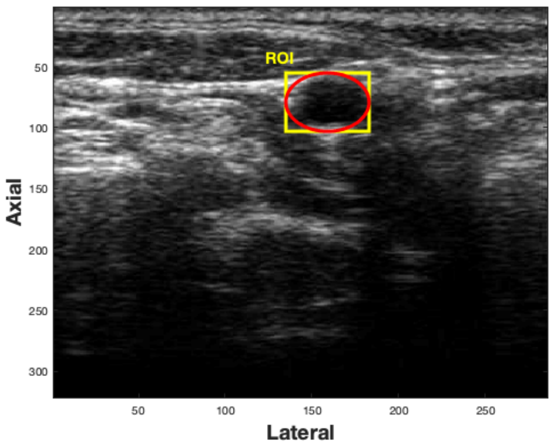


Fig. 2. Sample frame from the carotid artery (CA) dataset. The red circle represents the blood vessel and the yellow box represents the ROI.

TABLE I
RMSE VALUES IN PIXELS FOR SIAMFC AND ES-BM WITH RESPECT TO THE GROUND TRUTH AVERAGED OVER ALL IMAGE SEQUENCES OF CA DATASET.

RMSE	SiamFC	ES-BM
Axial	3.55	4.11
Lateral	3.22	4.23

lateral directions. Fig. 5a and 5b represent the performance of SiamFC against ES-BM for individual image sequences. Table I summarizes the RMSE values of SiamFC and ES-BM with respect to the ground truth averaged over all 10 image sequences. It can be seen that RMSE for SiamFC is reasonably lower than ES-BM in both axial and lateral directions.

B. Computational Time per Frame (CTF)

CTF was computed for both SiamFC and ES-BM for all 10 image sequences. Fig. 6 represents the error bar plot for the CTFs for both the trackers. Averaged over all frames for each sequence, CTF for SiamFC was found to be $0.29s$ and $1.69s$ for ES-BM. In terms of CTF, SiamFC was roughly 6 times faster than ES-BM.

C. Intersection Over Union (IoU)

Fig. 7 represents the IoU calculated for both SiamFC and ES-BM with respect to the ground truth over all 10 image sequences. The IoU averaged over all image sequences was found to be 82.38% for ES-BM and 83.08% for SiamFC.

IV. Discussion

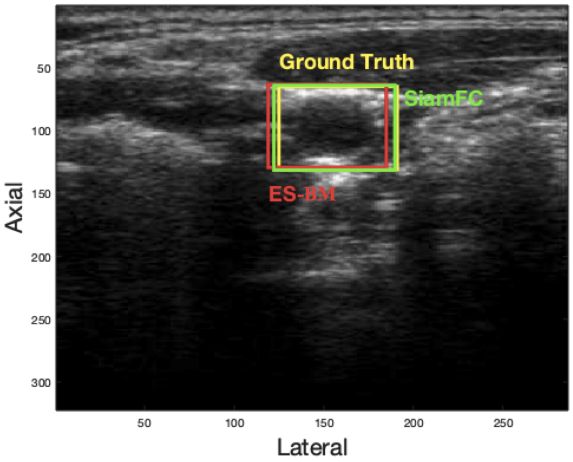
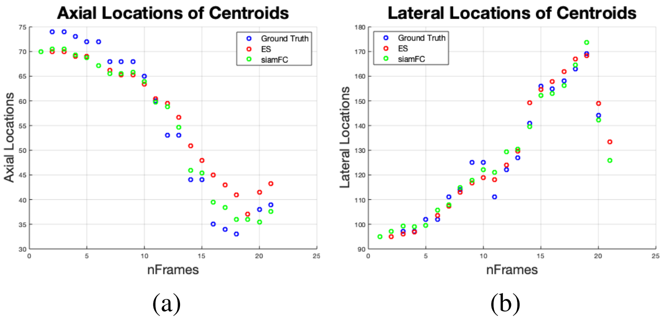


Fig. 3. Ground truth and detections: yellow represents the ground truth ROI, red represents the detection from ES-BM and green represents the detection from SiamFC.



(a)

(b)

Fig. 4. Centroid locations of ground truth, ES-BM and siamFC through all the frames of a sample image sequence from the CA dataset.

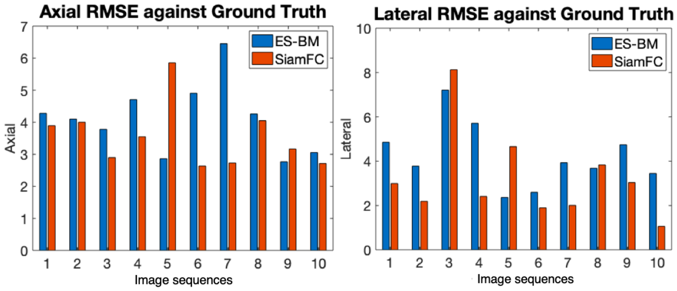
In this paper, we proved that applying SiamFC in place of conventional ES-BM for motion tracking in ultrasound images was feasible. Results from our initial experiments were found to be promising. It was evident that SiamFC performed better than ES-BM both in terms of accuracy and processing time. To the best of our knowledge, this is for the first time that a Siamese architecture has been adopted to estimate motion in ultrasound image sequences. We also found that SiamFC had much higher processing speed compared to ES-BM.

Since SiamFC is a class-agnostic tracker, any ROI can be tracked without having to explicitly re-train the network. This comes as the greatest advantage of using a Siamese architecture-based tracker. The deep convolutional neural network used in SiamFC is trained on the ILSVRC dataset [13]. We would like to highlight that this dataset consists of non-medical classes such as cats, dogs, cars, fish, etc. Yet, the performance of SiamFC was better than ES-BM in terms of every metric that we have used in this paper.

It should be noted that SiamFC is a detection-based tracker. That is to say that the reference ROI provided in the first frame is tracked through subsequent frames based solely on detections obtained using similarity matching. The tracker does not consider any motion model. This allows us to adopt the required motion model based on the target tissue or organ making the tracker more versatile.

V. Conclusion

A siamese architecture-based tracker was used to estimate the motion of the carotid artery. It was compared to the conventional exhaustive search-based block matching tracking. Our results proved that Siamese architecture-based tracker is



(a) (b)

Fig. 5. Represents RMSE in (a) axial direction and (b) in lateral direction, for all image sequences.

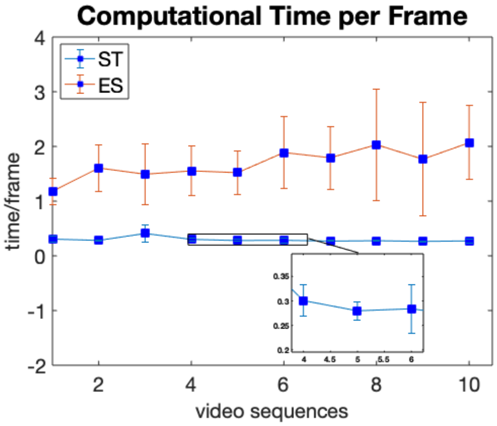


Fig. 6. Error bar plot of Computational Time per Frame (CTF) for SiamFC and ES.

feasible for motion estimation in ultrasound images. We also proved that it is better than conventional methods both in terms of accuracy and processing time. Our future work will focus on an improved version of the Siamese tracker with an additional motion model for carotid artery to further enhance tracking accuracy.

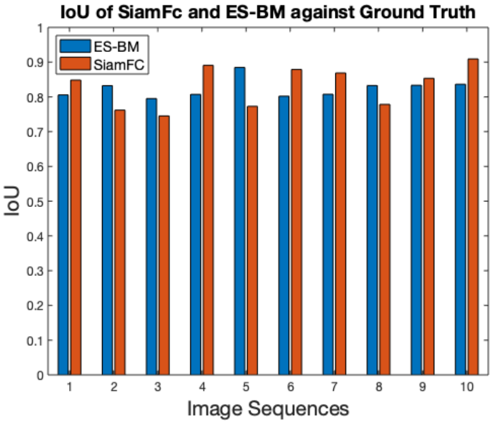


Fig. 7. IoU of SiamFC and ES-BM against the ground truth for all image sequences.

References

[1] Jingfeng Jiang and Timothy Hall, "A robust real-time speckle tracking algorithm for ultrasonic elasticity imaging," in *2009 IEEE International Ultrasonics Symposium*. IEEE, 2009, pp. 451–454.

[2] Jonathan Ophir, Ignacio Cespedes, Hm Ponnekanti, Youseph Yazdi, and Xin Li, "Elastography: a quantitative method for imaging the elasticity of biological tissues," *Ultrasonic imaging*, vol. 13, no. 2, pp. 111–134, 1991.

[3] Yong Yue, John W Clark Jr, and Dirar S Khoury, "Speckle tracking in intracardiac echocardiography for the assessment of myocardial deformation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 416–425, 2008.

[4] Andrea Giachetti, "Matching techniques to compute image motion," *Image and Vision Computing*, vol. 18, no. 3, pp. 247–260, 2000.

[5] Mohamed Almekkawy and Emad Ebbini, "Two-dimensional speckle tracking using parabolic polynomial expansion with riesz transform," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 201–205.

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

[7] Mohamed Khaled Almekkawy, Yasaman Adibi, Fei Zheng, Emad Ebbini, and Mohan Chirala, "Two-dimensional speckle tracking using zero phase crossing with riesz transform," in *Proceedings of Meetings on Acoustics 168ASA*. Acoustical Society of America, 2014, vol. 22, p. 020004.

[8] Brandon Rebholz and Mohamed Almekkawy, "Efficacy of kriging interpolation in ultrasound imaging; subsample displacement estimation," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2137–2141.

[9] Magnus Cinthio, Asa Ryden Ahlgren, Jonas Bergkvist, Tomas Jansson, Hans W Persson, and Kjell Lindstrom, "Longitudinal movements and resulting shear strain of the arterial wall," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 291, no. 1, pp. H394–H402, 2006.

[10] Skanda Bharadwaj and Mohamed Almekkawy, "Faster search algorithm for speckle tracking in ultrasound images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2142–2146.

[11] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr, "Fully-convolutional siamese networks for object tracking," in *European conference on computer vision*. Springer, 2016, pp. 850–865.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[14] Radek Benes, Jan Karasek, Radim Burget, and Kamil Riha, "Automatically designed machine vision system for the localization of cca transverse section in ultrasound images," *Comput. Methods Prog. Biomed.*, vol. 109, no. 1, pp. 92–103, Jan. 2013.

[15] Kamil Řiha and Igor Potůček, "The sequential detection of artery sectional area using optical flow technique," in *Proceedings of the 8th WSEAS International Conference on Circuits, systems, electronics, control & signal processing*. World Scientific and Engineering Academy and Society (WSEAS), 2009, pp. 222–226.

[16] Kamil Říha, Jan Mašek, Radim Burget, Radek Beneš, and Eva Závodná, "Novel method for localization of common carotid artery transverse section in ultrasound images using modified viola-jones detector," *Ultrasound in medicine & biology*, vol. 39, no. 10, pp. 1887–1902, 2013.