3D Ultrasound Anatomical Landmark Tracking via Fast Normalized Cross Correlation Network and Robust Outliers Rejection

Anonymous ECCV submission

Paper ID 11

Abstract. In image-guided abdominal radiotherapy, accurate localization of targets can minimize damage to crucial structure. Due to abdominal movements caused by heartbeat and breathing, however, margins would be added around target by surgeons to ensure target can be covered and treated, which would cause additional trauma. To alleviate motion uncertainties and minimize trauma, we propose an accurate algorithm based on a novel deep tracker and outliers rejection method for anatomical landmark tracking in 3D liver ultrasound sequences. Firstly, we couple normalized cross correlation filter (NCC) with fully convolutional network (FCN) and reformulate NCC as a differentiable layer to generate a novel and effective deep tracker. Meanwhile, we introduce the channel attention mechanism to generate the effective features. Finally, we derive a fast implementation form of NCC, which enables the algorithm to track in real time. The organizers of the Challenge of Liver Ultrasound Tracking (CLUST) evaluate the proposed algorithm, which yields mean and 95%ile tracking error of 1.70 0.98 mm and 3.05 mm, on 22 landmarks across 10 3DUS sequences. Comparison between our and published algorithms shows our algorithm achieves state-of-the-art performance. Moreover, it is proved by ablation study that the leading tracking results significantly benefit from fast NCC and channel attention mechanism.

Keywords: Fast normalized cross correlation filter, Channel attention module, Outliers rejection, Abdominal intervention therapy.

1 Introduction

In abdominal interventions, due to abdominal movement caused by heartbeat and breathing, surgeons would spend more time locating targets and set margins around target to ensure it could be covered and treated, which cause additional trauma. Therefore, effective motion management in intervention is crucial for minimizing the trauma. The approaches [9, 16] of motion mitigation are used to reduce the speed of the target to reduce the difficulty of target localization. However, these approaches may cause additional damage to health tissue and increase the surgery time. Therefore, more and more surgeons pay attention to the approaches of direct target localization.

In order to reduce the risk of surgery, surgeons want to be able to observe the movement of the target directly. Therefore, medical image-guided intervention therapy has become the most common way of abdominal intervention therapy. Compared with other types of medical images, such as computed tomography (CT) and magnetic resonance imaging (MRI), ultrasound (US) is attractive as its non-radiation and rich temporal resolution. Moreover, 3D US can show the abundant structure information and real motion pattern in liver. Therefore, 3D US is an ideal choice to guide surgeons in surgery. To locate the target tissue, the 3D US-based approaches [1, 4, 10, 13, 14, 17] that track the landmark in target tissue have been proposed. Most of these methods mainly focus on establishing the physical or physiological model of the target tissue, and then tracking the target. Inevitably, this will increase the complexity of these approaches. Inevitably, this will increase the complexity of these approaches and limit their clinical application. Recently, the approach [6] based on deep network has been proposed. However, how to increase the its robustness needs further discussion.

In this paper, a tracking algorithm based on deep tracker and outliers rejection method is proposed. Our main contributions are: first, we couple normalized cross correlation filter (NCC) with fully convolutional network (FCN) to generate a novel and effective deep tracker. Second, we introduce channel attention mechanism and train network end-to-end to generate more appropriate convolutional features for NCC. Third, we reformulate the tradition NCC as fast NCC, which enables the algorithm to track landmark in real time. Our algorithm achieves the state-of-the-art performance on CLUST dataset [9] for landmark tracking in 3D US. Based on the above contributions and the performance of the proposed network, proposed algorithm is potential to guide surgeons to accurate track target and minimize trauma in 3D ultrasound-guided intervention therapy.

2 Related work

In order to alleviate the uncertainties from breathing, heartbeat, and drift of patients, motion management approaches, such as respiratory gating technology [10] and anesthesia [17] are utilized to decrease the movement speed of liver. However, these approaches may cause additional damage to health tissue and increase the surgery time.

3D ultrasound-guided tracking of anatomical landmarks draws more and more attention [10], as it has large potential for reducing complication rates and minimizing surgical trauma [14]. Over the last decade, several approaches that utilize 3DUS for the anatomical landmark tracking have been proposed. A preoperative modeling approach [4] first employs 3D US to model the respiratory movement of the abdomen. However, this approach does not achieve intraoperative liver motion tracking. Further an intraoperative localization approach [17], which achieves a high landmark location accuracy has been proposed. Nevertheless, due to its high complexity, its clinical application is limited. In order to reduce the complexity, a morphological and geometrical constraints based outliers rejection method [1] has been proposed. Additionally, a dense motion

estimation and mechanical model simulation approach [13] achieves impressive tracking performance. While it can effectively simulate the real liver motion pattern, parameter tuning limits its application. Moreover, this approach depends on initial anatomical contours, which is potentially increase extra burden on surgeon when employing them into clinical application.

Recently, deep learning based approaches has shown potential for fast and robust US landmarks tracking. SSPMNet [6] couples the Siamese network with fully connection network to track 3D US landmarks, while it adds the spatial pyramid pooling (SPP) [7] after Siamese network. This approach uses block-matching based tracking strategy to locate the position of landmarks in 3D US, which makes its tracking speed unsatisfactory.

3 Method

3.1 Network Architecture

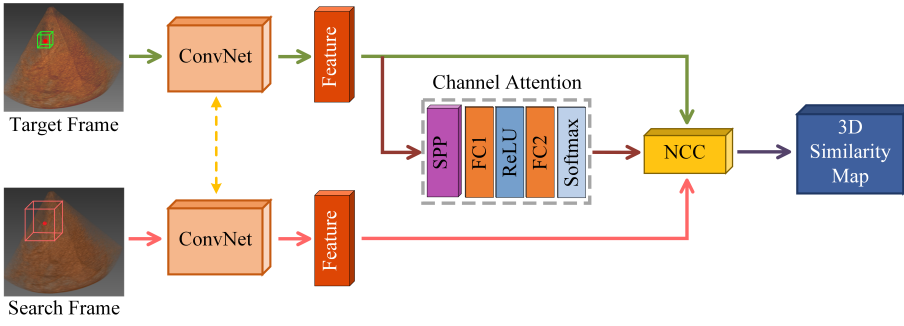

Fig. 1. The normalized cross correlation network in training process. The cube in target and the cube in search frame denote target and search block respectively. The red dots are manual annotation landmarks. The ConvNet means convolutional network.

As depicted in Fig.1, we couple the NCC with FCN and train it end-to-end to generate an effective deep tracker. Firstly, the FCN network is utilized to learn a feature representation of 3D US. Meanwhile, we introduced the channel attention module to generate the more effective convolutional features. Secondly, to alleviate the influence of ultrasonic gain variation, we introduce the NCC to calculate the 3D similarity map. Finally, as the loss function is optimized, the 3D similarity map is closer to the label. The FCN consists of three convolution layers. After each convolutional layer, batch normalization (BN) and rectified Linear Units (ReLU) are used.

3.2 Normalized Cross Correlation Layer

The Foundation of NCC. NCC is a classical method for stereo vision matching. The operation process of NCC is similar to the cross correlation. Compared with cross correlation, NCC is robust to ultrasonic noise and gain variation. Let T^l, F^l represent the l-th channel of transformation of target block $t \in \mathbb{R}^{w\times h\times d}$ and search block $f \in \mathbb{R}^{W\times H\times D}$. Let $F^l_{xyz} \in \mathbb{R}^{w\times h\times d}$ denotes the block, whose size is same as the size of target block, and center coordinate is (x, y, z) in search block. Therefore, single channel NCC can be represented by Eq. (1).

$$\begin{aligned} R^l_{xyz} &= \frac{\sum_{ijv}\left[F^l_{xyz}(i,j,v) - \overline{F^l_{xyz}}\right]\left[T^l(i,j,v) - \bar{T}^l\right]}{\sqrt{\sum_{ijv}\left[F^l_{xyz}(i,j,v) - \overline{F^l_{xyz}}\right]^2 \sum_{ijv}\left[T^l(i,j,v) - \overline{T^l}\right]^2}} \\ &= \frac{1}{N}\sum_{i,j,v}\frac{\left(F^l_{xyz}(i,j,v) - \mu_F\right)}{\sqrt{\sigma_F}}\frac{\left(T^l(i,j,v) - \mu_T\right)}{\sqrt{\sigma_T}} \end{aligned} \tag{1}$$

where, R^l_{xyz} refers to the similarity score or response value on position (x,y,z) of l-th similarity map. $\bar{}$ denotes the average value, and $N = w\times h\times d$. Multi-channel NCC can be calculated by Eq. (2), where β^l is calculated by the CAM:

$$R = \sum \beta^l R^l \tag{2}$$

Loss Function and Back-propagation. To better train the proposed network, we minimize the hingeloss:

$$l(\omega) = \sum \max(0, 1 - R\cdot y) + \rho\|\omega\|^2 \tag{3}$$

where, y is the label, whose calculation method is similar to that of SiameseFC [3], but the value of negative example is set to -1. In order to make convolutional features more appropriate for NCC, we reformulate the traditional NCC as a differentiable layer and then train the network end-to-end. In fact, both the calculation process of NCC and that of correlation can be regarded as that of block matching [5] with different similarity calculation methods. Therefore, when the target blocks slide to different positions on search block, the calculation process of similarity is independent and same. Therefore, the back-propagation of NCC is independent after each sliding. The back-propagation form of upper branch in Fig. 1 can be written as Eq.4 [15]:

$$\begin{aligned} \frac{\partial NCC\left(T^l, F^l_{xyz}\right)}{\partial T^l(i,j,v)} = &\frac{1}{N\sigma_T}\left(\frac{F^l_{xyz}(i,j,v) - \mu_F}{\sigma_F}\right. \\ &\left. - \frac{NCC\left(T^l, F^l_{xyz}\right)\left(T^l(i,j,v) - \mu_T\right)}{\sigma_T}\right) \end{aligned} \tag{4}$$

Additionally, the back-propagation of lower branch in Fig. 1 is:

$$\frac{\partial \mathrm{NCC}\left(T^l, F^l_{xyz}\right)}{\partial F^l_{uv}(i,j,v)} = \frac{1}{N\sigma_F}\left(\frac{\left(T^l(i,j,v)-\mu_T\right)}{\sigma_T} - \frac{\mathrm{NCC}\left(T^l, F^l_{xyz}\right) F^l_{uv}(i,j,v)-\mu_F}{\sigma_F}\right) \tag{5}$$

With Eq. (4) and Eq. (5), the back-propagation can be performed from NCC to FCN and then network can be trained end-to-end.

Fast Implementation Form of NCC. According to [5], the μ_T and σ_T of the target block only need to be calculated once. However, the μ_E and σ_F of search block at each position must be calculated, which are very time-consuming operation. In fact, they involve a lot of repeated calculations. By simplifying NCC, therefore, the calculation of NCC can be accelerated. Let $T' = T^l - \bar{T}^l$. For the numerator of Eq. 1, we can get the following formula:

$$NCC_{num} = \sum F^l_{xyz}(i,j,v)T'(i,j,v) - \overline{F^l_{xyz}}\sum T'(i,j,v) \tag{6}$$

Note that T' is zero-mean, hence, the second part of Eq. (6) can be removed. For denominator of Eq. (1), the problematic quantities are those in the expression $\sum_{ijv}\left[F^l_{xyz}(s)(i,j,v)-\overline{F^l_{xyz}}\right]^2$, which can be expanded easily by binomial theorem. Furthermore, note that:

$$\sum_{ijv}(\overline{F^l_{xyz}})^2 = N\left(\frac{1}{N}\sum_{ijv}F^l_{xyz}(i,j,v)\right)^2 \tag{7}$$

By the expansion of $\sum_{ijv}\left[F^l_{xyz}(s)(i,j,v)-\overline{F^l_{xyz}}\right]^2$ and Eq. (7), the denominator of Eq. (1) can be simplified as :

$$NCC_{den} = \left[\left(\sum_{ijv}\left(F^l_{xyz}(i,j,v)\right)^2 - \frac{1}{N}\left(\sum_{ijv}F^l_{xyz}(i,j,v)\right)^2\right)(T')^2\right]^{0.5} \tag{8}$$

Therefore, we can implement the fast NCC by utilizing four correlation:

$$FNCC\left(T^l, F^l\right) = \frac{T' \star F^l}{\sqrt{\left(F^l\right)^2 \star U - \left(F^l \star U\right)^2/N}\sqrt{(T')^2 \star U}} \tag{9}$$

Where, $\star$ is the correlation operator. U is the 3D matrix whose size is same as that of target block and whose elements are all 1. In fact, the above correlation can be converted into Hadamard product in frequency domain, and the calculation speed would be faster. However, due to the boundary effect, we just implement FNCC on the original domain.

3.3 Channel Attention Module

Different convolutional feature channels often correspond to different types of visual pattern [16]. In other words, the contributions of different convolution feature channels to the final result are different. Therefore, we introduce the channel attention module (CAM) to select the more significant convolutional feature. To keep the adaptation ability of the network to the different input size and target appearance variation, we utilize SPP to squeeze the convolutional features and employ a learn-based module to weight the features of different channels. [6] proves that the hierarchical features extracted by SPP at different scales are helpful to improve the accuracy of target localization in US. Additionally, note that Softmax is used as the final activation function, whose benefit is that the sum of weights of different channel features is 1. And then the output of NCC layer can be fixed between -1 and + 1.

3.4 Online Tracking and Outliers Rejection

To improve the robustness of tracking algorithm, an outliers rejection approach that can capture the motion pattern of local tissue to correct the displacement of the target. First, we sample several blocks around the target, which are called context blocks and have the same size as target block. In the search frame, our sampling strategy is the same as that of the training process. Second, by feeding the blocks to the proposed network, the most similar central coordinates of target block and context blocks in search frame can be obtained. In this way, we can use the movement information of the surrounding tissue to compensate the displacement of the target. However, ultrasound suffers from low signal-to-noise ratio and low spatial resolution [14], which may lead to inaccurate tracking results.

To reject outliers and improve the tracking performance, therefore, we introduce a robust outliers rejection method. Though, due to respiration and heartbeat, the movement of the liver is very intense, the displacement of different parts of the same tissue in the liver is very close. Therefore, the distance between outliers and inliers varies greatly at different time or frame, while the inliers preserve their geometric distances. Additionally, note that the higher the tracking score, the more reliable the result is. Let P^t be the set of central coordinates of sampling blocks from target frame. Correspondingly, let P^s denote the set of tracking central coordinates in search frame. $Ms \in \mathbb{R}^m$ is the set of the corresponding tracking scores. With the above prior knowledge, we construct geometric model and score model [1]:

$$G_{uo} = \exp\left(-\frac{d_{uo}^2}{2\vartheta_G^2}\right), \quad d_{uo}^2 = \frac{|\|P_u^t - P_o^t\| - \|P_u^s - P_o^s\||}{\|P_u^t - P_o^t\| + \|P_u^s - P_o^s\|} \tag{10}$$

$$M = \exp\left(-\frac{|1 - Ms|^2}{2\vartheta_M^2}\right) \tag{11}$$

where $G \in \mathbb{R}^{m \times m}$ is the adjacency matrix, whose elements represent the distance variation corresponding pair of points at different time or frame. M is the score vector. In order to combine the above models for robust outliers rejection, the following objective function is proposed:

$$GSCM = X^T K X, \quad K = G + \frac{\gamma}{2}\left(EM^T + ME^T\right) \tag{12}$$

where, X is the confidence, which states whether the tracking result is outlier or not. $X \in \Delta$, which is unit simplex, and defined as: $\Delta = \left\{X \in \mathbb{R}_+^m : E^T X = 1\right\}$ where, $E = [1, 1 \cdots 1]^T$ Replicator equation is utilized to optimize the Eq. (12) [11]. Replicator equation is:

$$X_{tj}(ti+1) = X_{tj}(ti)\frac{(KX(ti))_{tj}}{X(ti)^T K X(ti)} \tag{13}$$

where, X_h is the tj^{th} term of X. The detailed iterative update process is same as [1]. In addition, in order to utilize the temporal information, we set up the following time consistent model.

$$D'_{te} = \begin{cases} D_{te}, & \overline{Ms} \geq \text{ threshold} \\ D_{te-1}, & \text{other} \end{cases} \tag{14}$$

where, D'_{te}/D_{te} is the displacement before/after correction of landmark in current frame, and D_{te-1} is the displacement after correction of the landmark in the previous frame.

4 Experiments

4.1 CLUST Dataset

In this work, the CLUST 2015 challenge dataset [9] is used to train and evaluate the proposed algorithm. The ground truth of landmark is established using manual annotations by the radiologist. The summary of this data is shown in Table 1. An example image is shown in Fig 2.

Table 1. Summary of CLUST dataset. Train and Test denote the training set and test set of the CLUST dataset.

Source	Objects(Train / Test)	Volume Size	Resolution (mm)	Frame Rate (Hz)
EMC	6/8	$192 \times 246 \times 117$	$1.14 \times 0.59 \times 1.19$	6
ICR	1/1	$480 \times 120 \times 120$	$0.31 \times 0.51 \times 0.67$	24
SMT	9/13	$227 \times 227 \times 229$	$0.70 \times 0.70 \times 0.70$	8

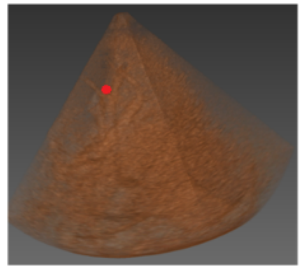

Fig. 2. An example image in CLUST dataset. The red dos is the tracking anatomical landmark.

4.2 Implementation Details

In training process, the size of target block and search block are $\{17\times17\times17, 45\times45\times45\}$. With the prior knowledge of physiological movement [1], we augment the data around three axes and rotating to -2,-1.5,-1,-0.5,0,0.5,1.25,1.5,2 degrees. We apply Adam [8] with a learning rate of 0.001 and a batch size of 8 to train the network. For SPP layer, the level of pyramid pooling is set to $\{1\times1\times1, 2\times2\times2\}$. In tracking process, the size of search block is $91 \times 91 \times 91$. Because the time resolution and spatial resolution of different types of sequences are different, we set different parameters for different type data. We set thresholds in time consistent model of (EMC, ICR, SMT) as (0.5, 0.2, 0.6). Further, for ultrasound sequences with spatial resolution greater than 1mm, the target block size is set to $17 \times 17 \times 17$, and for ultrasound sequences with spatial resolution less than 1mm, the target block size is set to $19 \times 19 \times 19$. For outliers rejection method, we set $\{\gamma, \vartheta_G, \vartheta_M\}$ as $\{0.1, 0.1, 0.1\}$.

Besides, all experiments are conducted on a workstation with Intel Core i9-9900X at 3.5GHz and NVIDIA RTX 2080 Ti 11GB GPU.

4.3 Results on CLUST Test Dataset

Euclidean distance is applied to evaluate tracking performance between the tracked points and manual annotations. Error statistics are summarized by mean, standard deviation (SD), and 95%ile error. Table. 2 shows that the tracking results of the proposed algorithm on CLUST test dataset. Further, the comparison between our and published algorithms is shown in Table 3. It shows that our algorithm achieves an improvement in mean of 0.04 mm (approx-imately 2.30%) and 95%ile error 0.36 mm (approximately 10.56%). Additionally, the tracking speed of the proposed algorithm is also improved. The average tracking speed of the proposed algorithm is 5 frame per seconds (fps), which is close to the imaging speed of 3D US. The processing speed of fast NCC is about 15 times that of traditional NCC. These support that the proposed algorithm achieves state-of-the-art performance and real-time tracking.

Table 2. Tracking results of the proposed algorithm on CLUST test dataset in millimeter

Landmark	Objects	Mean	SD	95ile
EMC	8	1.97	0.90	3.60
ICR	1	1.85	0.49	2.53
SMT	13	1.67	0.99	2.96
Total	22	1.70	0.98	3.05

Table 3. Comparison between our and published algorithms in millimeter. # means that no access to 20% of all data before computation of the tracking results. More details please see https://clust.ethz.ch/results.html (the first anonymous on results of 3D point-landmark tracking)

Approach	Tracked objects	Mean	SD	95ile
Royer et al. [13]	80%	1.74	**0.92**	3.65
Banerjee et al. [2]	80%	1.80	1.64	3.41
Our	100%	**1.70**	0.98	**3.05**

4.4 Ablation Study

In order to understand the benefits from different part of the proposed network, in this part, we perform ablation study on CLUST training set (20% for training, 80% for testing). Table 4 shows the tracking performance with different similarity calculation methods. Table 3 supports that the proposed algorithm significantly benefits from fast NCC. Besides, when Euclidean distance and correlation are used to calculate the similarity between target block and search block, there may be cases of tracking failure, which is similar to the conclusion of [14]. Specifically, we note that when L2 and Xcorr are used as similarity calculation methods, the tracking moving curves in some cases may be the straight line, while the target are moving.

Table 4. Tracking performance with different similarity calculation methods in millimeter. – denotes that there are cases of tracking failure. SCM denotes the different similarity calculation method.

SCM	Mean	SD	95ile
L2	-	-	-
Xcorr	-	-	-
Cosine	2.98	1.87	5.88
NCC	1.84	0.95	3.23

Table 5. Tracking performance with different Siamese network architecture in millimeter.

Network	Mean	SD	95ile
0 Convlayer	3.26	2.62	9.20
1 Convlayer	1.86	0.97	3.25
2 Convlayer	1.86	0.96	3.24
3 Convlayer	**1.84**	**0.95**	**3.23**
4 Convlayer	1.87	0.97	3.26
5 Convlayer	1.95	1.04	3.49
without CAM	1.94	1.03	3.51

Table 4 shows that proposed algorithm can benefit from CNN features and CAM. Note that the result, which does not use CNN feature is the tracking result of [2] on CLUST training set. Compared with the other two jobs, we found that we have made a greater improvement in the training set. This comparison is shown in Table 6. In fact, this is because that tracking algorithm faces different types of difficulty in training set and test set. In training set, some tissues undergo large deformation, and in test set, some targets may go out of the field of view (FOV). This shows that the algorithm can greatly reduce the influence of target tissue deformation.

Table 6. Tracking performance with different similarity calculation methods on training set in millimeter.

Approach	Mean	SD	95ile
Royer et al [13]	2.16	0.93	3.28
Banerjee et al [2]	3.26	2.62	9.20
Our	1.84	0.95	3.23

5 Conclusion

In this paper, we propose an accurate and real-time tracking algorithm for anatomical landmark tracking in 3D ultrasound-guided intervention. By coupling the FCN and NCC and training it end-to-end, an effective tracker is generated after coding 3D ultrasound images. Meanwhile, channel attention module is introduced to generate the more effective convolutional features. Then, fast NCC and outliers rejection method are introduced into this work to improve the computing speed and the robustness of the proposed algorithm. Finally, compared to other published work, the algorithm achieves the state-of-the-art

performance on CLUST dataset. Extensive ablation study proves that proposed algorithm significantly benefits from fast NCC and convolutional features. In conclusion, the proposed algorithm is potential to guide surgeons to accurately locate target with less time and minimize trauma in 3D ultrasound-guided interventional therapy. Future, different feature fusion approaches and scale learning are the potential choices to further improve the tracking performance.

6 References

1. Banerjee, J., Klink, C., Peters, E.D., Niessen, W.J., Moelker, A., van Walsum, T.: Fast and robust 3D ultrasound registration–block and game theoretic matching. Medical Image Analysis 20(1), 173-183 (2015).
2. Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., Torr, P. H.: Fully-convolutional siamese networks for object tracking. In: European conference on computer vision (ECCV), pp. 850-865. Springer, Heidelberg (2016)
3. Banerjee, J., Klink, C., Vast, E., Niessen, W.J., Moelker, A., van Walsum, T.: A combined tracking and registration approach for tracking anatomical landmarks in 4D ultrasound of the liver. In: MICCAI workshop: Challenge on Liver Ultrasound Tracking, pp 36–43, 2015
4. Dürichen, R., Davenport, L., Bruder, R., Wissel, T., Schweikard, A., Ernst, F.: Evaluation of the potential of multi-modal sensors for respiratory motion prediction and correlation. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5678-5681. IEEE, (2013)
5. Hallack A., et al.: Robust Liver Ultrasound Tracking using Dense Distinctive Image Features. In: MICCAI 2015 Challenge on Liver Ultrasound Tracking. Munich, Germany (2015)
6. He J., Shen C., Huang Y,. Wu J.: Siamese Spatial Pyramid Matching Network with Location Prior for Anatomical Landmark Tracking in 3-Dimension Ultrasound Sequence. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) pp. 341-353. Springer, Heidelberg (2019).
7. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence 37(9), 1904-1916 (2015).
8. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. In: arXiv:1412.6980. (2014)
9. Luca V., et al.: The 2014 liver ultrasound tracking benchmark. Physics in Medicine and Biology, 60(14), 5571-5599 (2015)
10. Mageras, G.S., Yorke, E.: Deep inspiration breath hold and respiratory gating strategies for reducing organ motion in radiation treatment. Seminars in Radiation Oncology 14(1), 65-75 (2004).
11. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE transactions on pattern analysis and machine intelligence, 29(1), 167-172 (2007)
12. Riley C., et al.: Dosimetric evaluation of the interplay effect in respiratory-gated RapidArc radiation therapy. Medical Physics. 41(1), 011715 (2014)

13. Royer, L., Krupa, A., Dardenne, G., Le Bras, A., Marchand, E., Marchal, M.: Real-time target tracking of soft tissues in 3D ultrasound images based on robust visual information and mechanical simulation. Medical Image Analysis 35, 582-598 (2017).
14. Shen C., He J., Huang Y., Wu J.: Discriminative Correlation Filter Network for Robust Landmark Tracking in Ultrasound Guided Intervention. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp 646-654, Springer, Heidelberg (2019).
15. Subramaniam, A., Chatterjee, M., Mittal, A.: Deep neural networks with inexact matching for person re-identification. In Advances in Neural Information Processing Systems (NIPS) pp. 2667-2675 (2016)
16. Wang Q., Teng Z., Xing J., Gao J., Hu W., Maybank S.. Learning attentions: residual attentional siamese network for high performance online visual tracking. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4854-4863).
17. Vijayan, S., Klein, S., Hofstad, E.F., Lindseth, F., Ystgaard, B., Langø, T.J.M.p.: Motion tracking in the liver: Validation of a method based on 4D ultrasound using a nonrigid registration technique. Medical physics, 41(8Patr1), (2014).