Transfer Learning for Motion Estimation in Ultrasound Imaging

Ewan Evain\textsuperscript{1,2} \hspace{1cm} EWAN.EVAIN@PHILIPS.COM
\textsuperscript{1} Medisys Research Lab, Philips France
\textsuperscript{2} CREATIS, INSA-Lyon, University of Lyon 1, Villeurbanne

Khuram Faraz \textsuperscript{2} \hspace{1cm} FARAZ@CREATIS.INSA-LYON.FR
Damien Garcia \textsuperscript{2} \hspace{1cm} GARCIA.DAMIEN@GMAIL.COM
Thomas Grenier \textsuperscript{2} \hspace{1cm} THOMAS.GRENIER@CREATIS.INSA-LYON.FR
Mathieu De Craene \textsuperscript{1} \hspace{1cm} MATHIEU.DE.CRAENE@PHILIPS.COM
Olivier Bernard \textsuperscript{2} \hspace{1cm} OLIvier.BERNARD@INSa-lyon.FR

Editors: Under Review for MIDL 2019

Abstract

In recent years, deep learning (DL) has been successfully applied to ultrasound (US) image processing for image segmentation and view recognition. Whether DL can also outperform standard tracking algorithms in US (e.g. speckle tracking) has yet to be evaluated. Assessing the performance of DL for motion estimation is also clinically relevant, especially in the context of cardiac image analysis. Obtaining training data for motion algorithms is a challenging task as manually annotating data is not an option, thus calling for building simulated US databases on which different network architectures can be trained and compared to state-of-the-art algorithms. Such synthetic databases are therefore keys to determine if DL is suited for tissue and flow motion estimation, and if the performance gap is similar or not to other application fields. To this aim, we chose to focus on simplified geometry and motion fields and benchmarked the performance of convolution neural networks for the case of a rotating disk, both for synthetic and in-vitro images.

Keywords: Motion estimation, Optical Flow, Deep Learning, Ultrasound, Neural Networks

1. Introduction

Translating DL advances to the field of medical imaging often requires dedicated approaches for each imaging modality. Ultrasound is a widespread modality in many clinical applications, including abdominal, cardiac and fetal imaging. A lot of researchers are therefore evaluating DL solutions in ultrasound for guiding acquisition (Prevost et al., 2018), recognizing standard imaging views (Raynaud et al., 2017) and segmenting images (Leclerc et al., 2019). Besides these applications, motion quantification from deep learning algorithms has yet to be benchmarked against state of the art algorithms.

Obtaining reference values from medical images is often a challenging task. For instance, validating image segmentation requires contouring by several experts, which is a manual
tedious task. For tracking algorithms, reference values are even more difficult to obtain as it is almost impossible to manually follow all trajectories in image sequences with enough accuracy. A practical way to address this issue is to use synthetic datasets. Novel approaches mixing real images with biomechanical and ultrasound models (Alessandrini et al., 2018) have recently been introduced for benchmarking cardiac motion quantification algorithms. It thus makes sense to use a similar approach to train and evaluate CNN architectures aiming at reconstructing motion fields.

A prominent method for CNN-based motion quantification is FlowNet2 (FN2) (Ilg et al., 2017). Indeed, this architecture currently ranks among the best tracking methods in computer vision. We therefore chose to apply standard transfer learning concepts to evaluate FN2 performance on synthetic and real US images. We limited ourselves to the case of a rotating disk that can easily been reproduced in vitro (Porée et al., 2016). We performed transfer learning as described in the next sections for adapting the FN2 architecture to handle US images.

2. Methodology

2.1. Database

We designed a simulation pipeline for generating training images and reference displacement fields according to the principles described in (Schmerr Jr, 2014). Foreground scatterers were placed on a rotating disk. The resulting database was composed of 2850 image pairs with reference frame-to-frame displacement fields. We simulated 3 different angular speeds: 1, 2 and 3 radians per second. For each value of the angular speed, 10 sequences, each containing 95 pairs of frames were simulated: 5 sequences with a homogenous disk and 5 with four anechoic cysts positioned symmetrically with respect to the centre of the disk. The position of scatterers were randomly distributed for each simulated sequence, inducing small texture changes in the produced images. The settings of the simulated ultrasound equipment are the same as the ones used to acquire the in vitro database described below.

The second database corresponds to real acquisitions of a rotating disk performed with a Verasonics research scanner (V-1-128, Verasonics Inc., Redmond,WA) and a 2.5 MHz phased-array transducer (ATL P4–2, 64 el-ements). Each sequence is composed by 300 frames with a constant value for the rotational speed of 1, 2 or 3 rad/s.

2.2. Training

Regarding the training, we followed a transer-learning strategy, starting from the pre-trained weights of FN2 architecture (weights learned from synthetic sequences of flying chairs (Dosovitskiy et al., 2015)). In particular, FN2 is composed of five U-Net (Ronneberger et al., 2015) divided into two branches, one with three stacked networks and the other with one U-Net. The last U-Net takes the results of the two branches as input and outputs the final displacement field result. Only weights of the last U-Net were optimized during the transfer learning phase.

As training loss, we use the $L_{p,q}$ norm and the endpoint error (EPE, expressed in pixels) as metric which is the Euclidean distance between the predicted flow vector and the ground truth, as this is a standard error measure for optical flow estimation (Baker et al., 2011).
Transfer Learning for Motion Estimation in Ultrasound Imaging

<table>
<thead>
<tr>
<th></th>
<th>PIV</th>
<th>FlowNet2</th>
<th>Transfer Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulation</td>
<td>Real</td>
<td>Simulation</td>
</tr>
<tr>
<td>EPE Speed</td>
<td>EPE Speed</td>
<td>EPE Speed</td>
<td>EPE Speed</td>
</tr>
<tr>
<td>1 rad/s</td>
<td>0.1</td>
<td>1.0</td>
<td>0.4</td>
</tr>
<tr>
<td>2 rad/s</td>
<td>0.1</td>
<td>2.0</td>
<td>0.5</td>
</tr>
<tr>
<td>3 rad/s</td>
<td>0.3</td>
<td>3.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 1: Average EPE in the disk and estimated speed on synthetic and real databases for three algorithms: PIV, FN2 and FN2 after transfer learning.

The loss and the solver are the same as in FN2 with the same parameters. After careful tuning, the learning rate was set to $\lambda = 1e^{-4}$ with a batch size of 4. Our network was implemented in Keras with the Tensorflow backend.

We chose to train the network on the synthetic databases with each rotational speed (1, 2 and 3 rad/s). This database is split into three parts: 60% for the training set, 20% for the validation set and 20% for the test set. Datasets in the validation and test sets were not chosen randomly. Indeed, for each speed value, as the database contains 5 sequences with anechoic cysts and 5 without, we ensured that the validation and test sets did both contain one sequence of each category.

3. Results

The final results are obtained on both databases: real and simulated images. On the test fold of both databases, we assessed accuracy using EPE with the corresponding global estimated rotational speed. From Table 1, we can observe that our transfer learning method outperforms PIV and FN2 in terms of EPE on both databases. This validates the fact that we have optimized our network during the learning process based on this metric. With regard to the estimation of the rotational speed, PIV made the best estimate on the synthetic data for 2 rad/s and 3 rad/s but it is interesting to see that FN2 after transfer learning was as accurate for 1 rad/s. PIV was also less accurate than our method on real disk sequences and tended to underestimate the rotational speed with respect to our method.

4. Conclusion

In this paper, we evaluated the accuracy of a CNN architecture after transfer learning to estimate motion in ultrasound. Starting from the FN2 architecture and weights, the network was trained on a database of simulated ultrasound images of a rotating disk with various speeds and speckle patterns. A significant improvement in EPE accuracy was obtained after transfer learning and tuning of the hyper-parameters, thus suggesting the applicability of FN2 based networks to ultrasound. This method is also effective in estimating a rotation speed on both real and simulated sequences where it gave consistent results. Although we focused here on a simplified geometry and motion, future work will extend this preliminary study by considering more complex geometries, and more elaborated motion fields.
References


