Convolutional neural network (CNN) applied to respiratory motion detection in fluoroscopic frames

Authors: Christoph Baldauf[1], Alex Bäuerle[2], Timo Ropinski[2], Volker Rasche[1], Ina Vernikouskaya[1]

[1]*Internal Medicine II, Ulm University Medical Center, Ulm, Germany*

[2]*Institute of Media Informatics, Ulm University, Ulm, Germany*

Purpose: X-ray (XR) fluoroscopy provides real-time images and allows physicians to see the internal structure and function of a patient thus yielding an essential technique for guiding cardiovascular interventions. The weaknesses of XR in imaging soft tissues can be addressed by augmenting real-time fluoroscopy with organ shape models derived from *e.g.* pre-interventionally acquired CT angiography (CTA).

Following initial registration, respiratory motion is a major cause of introducing mismatch to the superposition. If this motion can be extracted from the fluoroscopy, the model overlay position can be adjusted accordingly, and the mismatch can be reduced.

Convolutional neural networks (CNN) have been shown to be a powerful technique in image related tasks. This project aims to evaluate CNNs as a novel approach to extract respiratory motion from fluoroscopic runs.

Methods: Respiratory motion correlates well with the diaphragm position, which is often the most prominent anatomic structure in the fluoroscopic images. Thus the CNN is trained on fluoroscopic frame pairs as input with the target value being the displacement of the diaphragm between these frames. The displacement is restricted in head-feet direction only and is determined by defining a region of interest (ROI) on the diaphragm edge and tracking its motion with cross-correlation using Matlab. In order to gather target values of the training data from a similar region, selection of a ROI was restricted to the most right 17% of the frame. Figure 1 shows two fluoroscopic frames with displacement of the diaphragm indicated between these frames and ROI defined in the left image.

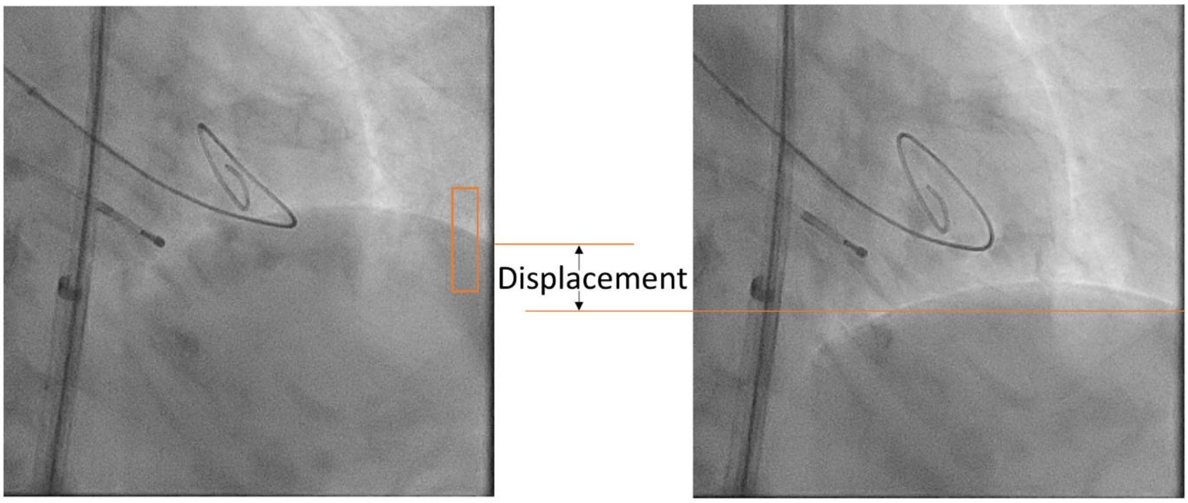

Figure 1 -- Two fluoroscopic frames showing the displacement of the diaphragm. The orange rectangle (left image) indicates the defined ROI.

Training data were derived from transcatheter aortic valve implantation (TAVI) procedures performed at the Ulm University Medical Center. 100 fluoroscopy runs (length 47 – 299 frames, 512 x 512 resolution) containing the diaphragm were preprocessed and labeled. All possible permutations of frame pairs were created for each run, which results in a total of 550.000 samples with target values in the range ±119 pixels. Samples were split into training-, validation- and test-sets with a ratio 86%/9%/5%.

The problem is addressed with a rather generic network, allowing the network to decide itself how to process the image pairs and extract the displacement. The network consists of multiple convolutional- and pooling operations before getting the output value from the final fully-connected layer. Implementation was done in Keras[1] and trained on a GeForce GTX 1080 graphics card.

Results: Left image in Figure 2 visualizes the superposition of target values and prediction for a fluoroscopy run of the validation set. Despite a mean absolute error of 2.35 pixels, the overlay provides reasonable accuracy.

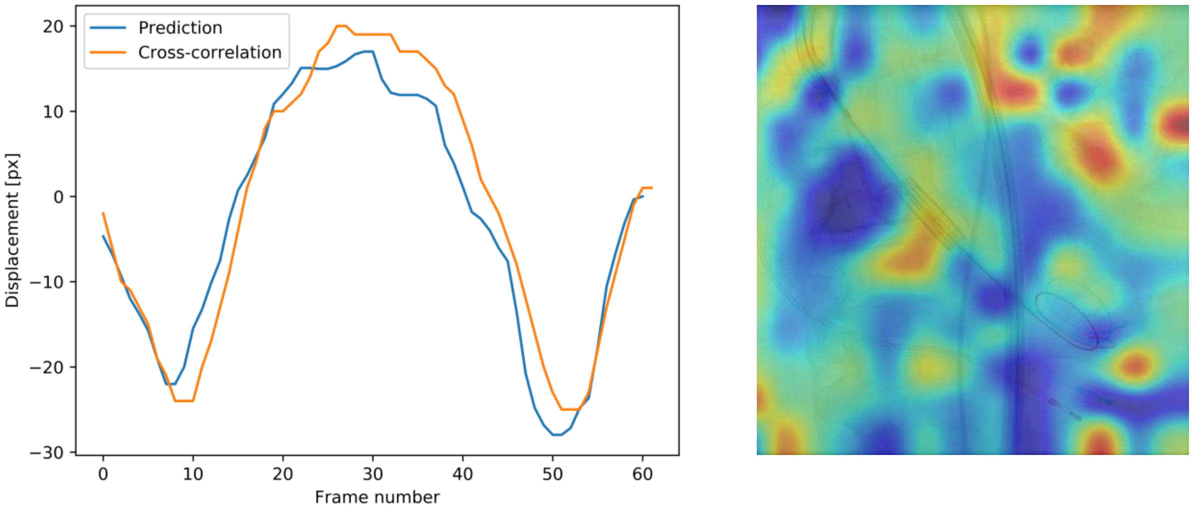

*Figure 2 -- **Left**: Overlay of network prediction (blue) and cross-correlation results (orange). The curve represents the displacement of the diaphragm within one fluoroscopy run from the validation dataset. **Right:** Attention heatmap plot. Red parts show regions of the fluoroscopic frame that cause the prediction strongly to decrease, i.e. predict a negative displacement, while blue regions do not influence the prediction significantly.*

The mean absolute error on the whole validation set after 10 epochs of training is 4.49 pixels. The deviation to the cross-correlation occurs mainly in the respiratory transition phase, *i.e.* while the displacement values are rather high. While these often are underestimated by the network, the respiratory waveform can be extracted from all runs and can be used for respiratory rate and phase detection.

Attention heatmaps were used to visualize the importance of different regions of the input images for the prediction. Although training data were generated based on the diaphragm's displacement, the network's attention rarely has a strong focus on the diaphragm, but rather composes its decision from multiple features in the images (Figure 2, right image). Furthermore, these features vary greatly across different fluoroscopy runs.

[1] https://keras.io/

Conclusion: Convolutional neural network has been shown to be capable to extract respiratory motion from fluoroscopic frames. Accuracy is sufficient to reliably detect respiratory phase and rate. Deviations from the cross-correlation results were observed especially for large respiratory amplitudes. One particular issue may result from the huge individual variability of respiratory motion and respective diaphragm displacement. Further, calculating the displacement with the help of cross-correlation is depending on the selected ROI. This is a weakness of the dataset, since the ROI cannot always be set on the same location because some trial and error is often necessary before finding a ROI that leads to a plausible result.

Further work should address how to overcome the drawbacks of the current dataset. Training on a feature that is displaced uniformly at every location could improve the performance. In contrast to the diaphragm, these features are much more difficult to be automatically labeled.