Deep-Learning-Based CT Motion Artifact Recognition in Coronary Arteries

T. Elss[#,⊗], H. Nickisch[#], T. Wissel[#], H. Schmitt[#], M. Vembar[†], M. Morlock[⊗], M. Grass[#]

[#]Philips Research, Hamburg, Germany
[⊗]Hamburg University of Technology, Germany
[†]Philips Healthcare, Cleveland, OH, USA

ABSTRACT

The detection and subsequent correction of motion artifacts is essential for the high diagnostic value of non-invasive coronary angiography using cardiac CT. However, motion correction algorithms have a substantial computational footprint and possible failure modes which warrants a motion artifact detection step to decide whether motion correction is required in the first place. We investigate how accurately motion artifacts in the coronary arteries can be predicted by deep learning approaches. A forward model simulating cardiac motion by creating and integrating artificial motion vector fields in the filtered back projection (FBP) algorithm allows us to generate training data from nine prospectively ECG-triggered high quality clinical cases. We train a Convolutional Neural Network (CNN) classifying 2D motion-free and motion-perturbed coronary cross-section images and achieve a classification accuracy of $94.4\% \pm 2.9\%$ by four-fold cross-validation.

Keywords: cardiac CT, motion artifact measure, coronary angiography

1. INTRODUCTION

Non-invasive coronary angiography using cardiac CT imaging has become a preferred technique for the detection and diagnosis of coronary artery disease (CAD). However, motion artifacts due to cardiac motion may limit or even preclude the evaluation of portions of coronary arteries or cause misinterpretations. Several motion compensation approaches have been developed which deal with motion estimation via 3-D/3-D registration of multiple heart phases [1, 2, 3]. These approaches are computationally demanding and have possible failure modes which offers potential applications for reliable motion artifact measures. The classification of each location in the coronary artery tree in *motion-free* and *motion-corrupted* could decide whether and where motion correction is required in the first place. Additionally, the success of motion compensation methods could be verified by an appropriate measure for motion artifacts. Rohkohl et al. [4] also introduced a motion compensation method

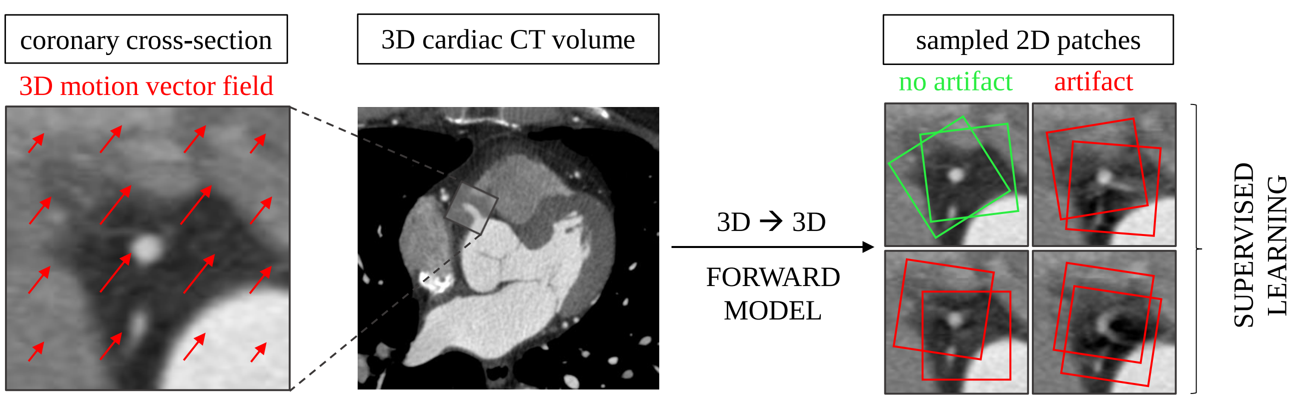

Figure 1: Multiple CT image volumes with varying artifact levels are created by the forward model which introduces motion locally around the coronary arteries by means of an artificial motion vector field. Randomly rotated, cropped and flipped coronary cross-sections are used as input features for supervised learning. The corresponding target label (*no artifact* or *artifact*) depends on the utilized motion vector field.

which is based on motion vector field estimation by minimizing handcrafted motion artifact measures. We found current handcrafted measures [4, 5] like normalized entropy and positivity not to be sufficiently robust with regard to the high variability of background intensities and the noise level of clinical cases (see Figure 7). Motion artifacts in the coronary arteries manifest in typical patterns containing arc-shaped blurring and intensity undershoots due to the CT reconstruction geometry (see Figure 1, 5 or 6). Over the past few years CNNs have been driving advances in many pattern recognition tasks [6] and by means of machine learning, the challenging task of designing a handcrafted measure with the required robustness, might be circumvented. Therefore, the goal is to investigate the feasibility of deep-learning-based motion artifact recognition in coronary arteries. In contrast to typical supervised learning, the required feature and label data is generated by a forward model which introduces artificial motion to cases with high quality.

2. MATERIAL AND METHODS

Nine prospectively ECG-triggered clinical datasets acquired with a 256-slice CT scanner (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA) are included. The mean heart rates range between 45.8 bpm and 66.0 bpm. The target phase was chosen at mid-diastole between 70% and 80% R-R interval, respectively. It is verified by visual inspection that the nine reconstructed *step-and-shoot* cases exhibit no motion artifacts. Additionally, a well-segmented coronary tree including the centerline and the lumen contour is created for each case using the Comprehensive Cardiac Analysis Tool (IntelliSpace Portal 9.0, Philips Healthcare, Cleveland, OH, USA). The required data for supervised learning is generated by applying the following forward model on these nine cases.

2.1 FORWARD MODEL

The forward model simulates patient motion by applying the motion compensated filtered back projection (MC-FBP) algorithm [1, 7]. This reconstruction algorithm includes angular weighting for gated reconstruction and aperture weighting for avoidance of cone-beam artifacts. Usually, motion vector fields are estimated to shift image voxels according to the point in time of each specific projection. Thus, the subsequent back projection leads to a correction of the detected motion. In contrast, the forward model as shown in Figure 1 inverts this process. It takes the CT image with corresponding projection data, the segmented coronary tree and the target motion strength $s \in \mathbb{R}^+$ as input and delivers a motion degraded CT image as output. The heart motion, or strictly speaking the reversing motion is simulated by an artificial continuous motion vector field $\vec{m}_t \colon \Omega \to \mathbb{R}^3$ which describes the displacement of each voxel coordinate $\vec{x} \in \Omega \subset \mathbb{R}^3$ in the CT volume at each time point t in millimeters. The time $t \in [0\%, 100\%]$ is measured in percent cardiac cycle and the continuous motion vector field is separable in a time and a location component:

$$\vec{m}_t(\vec{x}) = w(\vec{x}) \cdot \vec{\delta}(t) \tag{1}$$

The location-dependent weight mask $w \colon \Omega \to [0,1]$ limits the motion to the area of the coronary arteries and is generated by dilation of the binary input lumen mask and a subsequent uniform filtering. The smoothing is necessary to avoid reconstruction artifacts, because the elastic tissue structure forbids abrupt changes of motion in a local neighborhood. The time component $\vec{\delta}(t) \in \mathbb{R}^3$ is obtained by piecewise linear interpolation between five sample vectors $\vec{\delta}_i \in \mathbb{R}^3, i \in \{1, \ldots, 5\}$ (see Figure 2 and Figure 4). The corresponding time points are given by $t_i \in \{r - 10\%, r - 5\%, r, r + 5\%, r + 10\%\}$, where r denotes the reference heart phase of the input CT volume. The first and the last sample vectors $\vec{\delta}_1$ and $\vec{\delta}_5$ are constant extrapolated for time values outside the covered range of 20% R-R interval. In practice, the angular weighting window, which depends on the heart rate and the gantry rotation speed, is usually narrower. Therefore, the influence of the sample vectors may differ. They are computed with the following formula:

$$\vec{\delta}_i = \frac{s}{\max_{j,k} \|\vec{\rho}_j - \vec{\rho}_k\|_2} \cdot \vec{\rho}_i \tag{2}$$

The motion direction is given by random uniform vectors $\vec{\rho}_i \sim \mathcal{U}[-1,1]^3$ for $i \in \{1, 2, 4, 5\}$ and no motion occurs at the reference heart phase, i.e. $\vec{\rho}_3 = \vec{0}$. The first term in Equation (2) scales the motion vector fields so that the target motion strength s corresponds to the maximal displacement during 20% R-R interval in millimeters (see Figure 3). Of course, alternative normalization factors instead of the maximum pairwise Euclidean distance are possible. Empirical results by visual inspection show that the current choice of maximum normalization

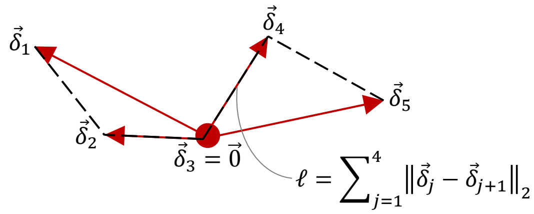

Figure 2: Schematic drawing of the motion trajectory (dashed line) which is determined by the sample vectors $\vec{\delta}_i$ (red arrows). The total displacement length of the trajectory ℓ corresponds to the sum of the Euclidean distances.

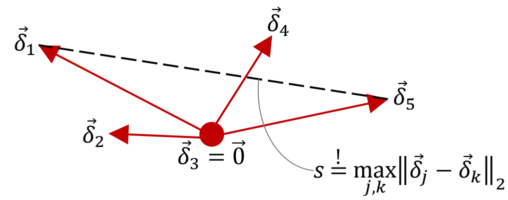

Figure 3: The sample vectors $\vec{\delta}_i$ (red arrows) are scaled so that the target motion strength s corresponds to the maximum pairwise Euclidean distance (length of the dashed line).

delivers a better correlation to the artifact level than the total displacement length (visualized in Figure 2). The resultant role of s as motion level regulator is illustrated in Figure 5. It also has to be noted that the current choice of random uniform vectors for $\vec{\rho}_i$ may yield no realistic physical simulation of the cardiac motion. Restrictions should be specified in further investigations to create an appropriate subset of motion manifestations. Nevertheless, CT volumes created by the proposed forward model show the typical motion artifact pattern of arc-shaped blurring and intensity undershoots at the coronary arteries (see Figure 5 and Figure 6).

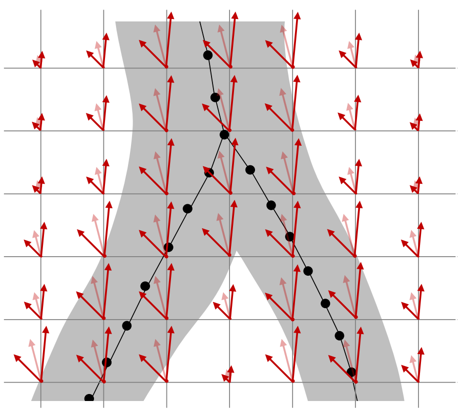

Figure 4: Schematic drawing of $\vec{m}_t(\vec{x})$. The displacement vectors (bright red) are linearly interpolated from the sample vectors (dark red) in time domain. The motion directions are spatially constant, while the displacement length decreases with increasing distance to the coronary centerline shown in black.

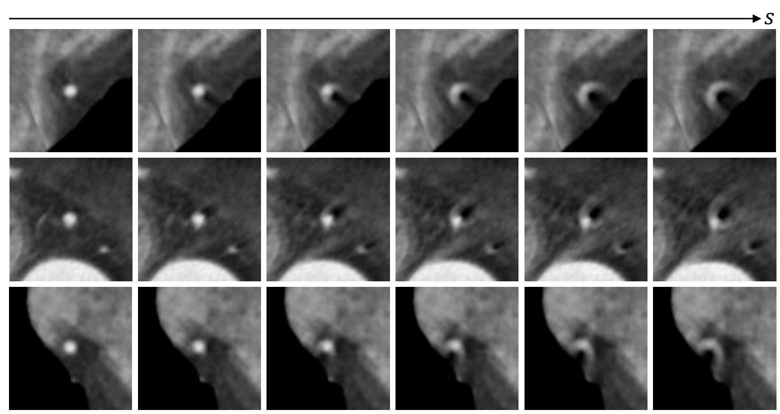

Figure 5: Coronary cross-section images sampled from 3D CT image volumes which were perturbed by the forward model. The same coronary cross-section images are given row-wise in different motion states. The motion strength s increases from left to right, while the motion directions, i.e. $\vec{\rho}_i, i \in \{1, \ldots, 5\}$ were fixed. The lengths of the displacement vectors in the motion trajectory and consequently the artifact level are regulated by the target motion strength s.

2.2 SUPERVISED LEARNING

The forward model enables the creation of multiple motion-degraded 3D CT image volumes with controlled motion levels at the coronary arteries. It is important to consider that the motion level is not equivalent to the artifact level. Phantom studies show that the artifact level additionally depends on the relation of the motion direction, the tube positions during acquisition and the orientation of the coronary arteries. Also other factors like background intensities and the angular weighting window have an impact. Therefore, the motion strength s is merely an approximate measure for the artifact level in the coronary arteries.

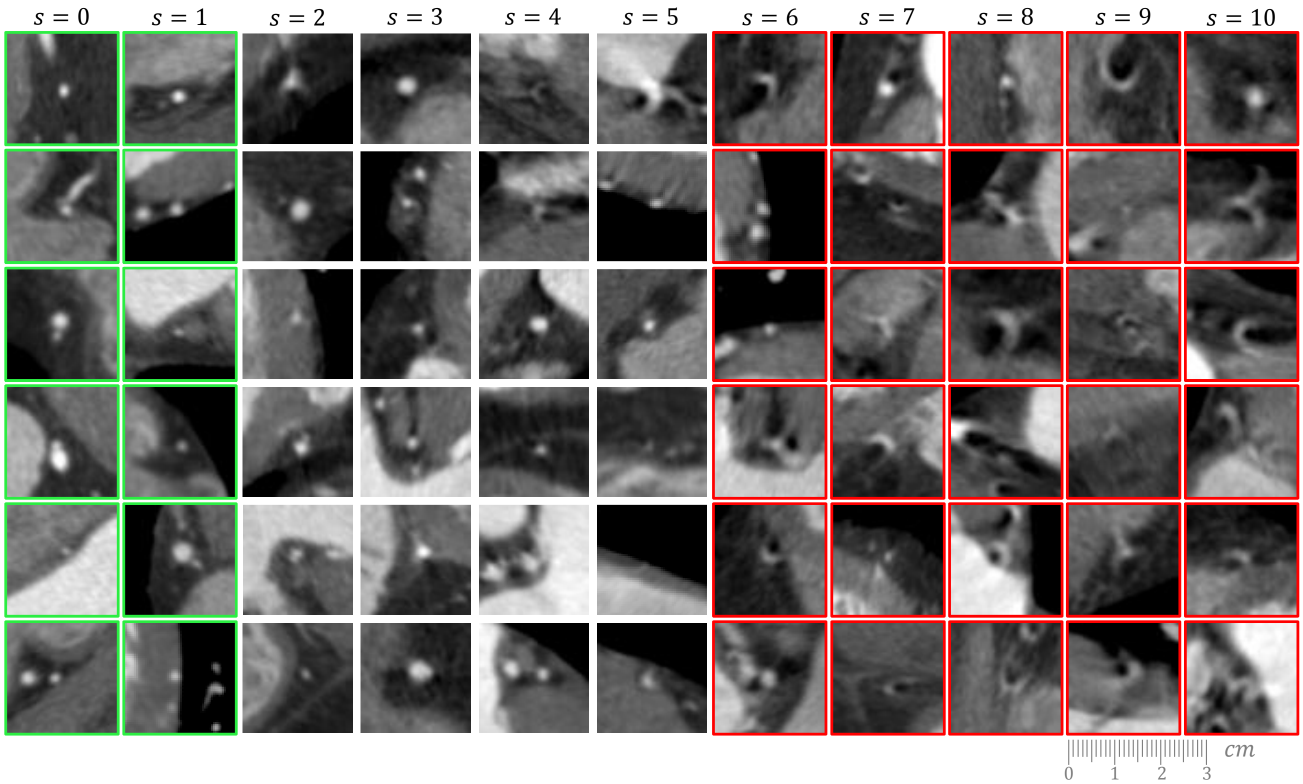

Figure 6: Coronary cross-section images are sampled from 3D CT volumes which are generated by the forward model. The visualized example patches $I^{60\times60}(\vec{c})$ are randomly rotated, flipped and cropped to a size of 60×60 pixels. The corresponding motion strength s is given above. The artifact level in the coronary arteries is not monotonically increasing with the underlying motion strength s, but on average growing intensity undershoots and blurring can be observed. Patches highlighted in green or red are assigned to the classes *no artifact* or *artifact* respectively and defined as input feature data of the neural network. Due to the ambiguous class label, the non-highlighted patches with a motion level between 2 and 5 are excluded from the learning process.

Figure 6 shows randomly sampled coronary cross-section images from 3D CT image volumes perturbed with a varying motion strength $s \in \{0, 1, \ldots, 10\}$. On the basis of velocity measurements at the coronary arteries by Vembar et al. [8], the data generation process for the supervised learning task was limited to the maximal displacement of 10 millimeters during 20% cardiac circle. Figure 6 reveals the difficulties of the given learning problem. The neural network has to be robust regarding variations in noise level, background intensities, vessel radius and effective intensity of the contrast agent. Another requirement on a reliable motion artifact measure is the differentiation between blurring artifacts and branching coronary arteries.

In this paper, the supervised learning problem is defined as a classification task to separate 2D *artifact* and *no artifact* coronary cross-section images. The database is generated by applying the proposed forward model seven times per clinical case and the target label l (0: *no artifact*, 1: *artifact*) is given by the utilized input motion strength:

$$l = \begin{cases} 0, & \text{if } s \in \{0, 1\} \\ 1, & \text{if } s \in \{6, 7, 8, 9, 10\} \end{cases} \tag{3}$$

The gap in s is chosen to assure a better class division. As input for the supervised learning approach, coronary cross-sections of the size 96×96 pixels are sampled perpendicular to the centerline with a resolution of 0.4 millimeter per pixel along the whole coronary tree. For this purpose, the original no-motion coronary centerlines are used. The samples are clipped to the relevant intensity range with a window/level setting of 750/100 HU. To guarantee balanced classes, a subset of two fifths of the samples from class *artifact* is randomly selected. The dataset includes a total of 18k samples. These are case-wise separated for training, validation and testing with a ratio of 6 : 2 : 1. During training, online data augmentation is performed by randomly rotating (by 0 to 360 degrees), horizontal mirroring and cropping of the image patches to the final CNN input size of 60×60 pixels (see Figure 1). The forward model may cause a shifting of the original coronary centerline point $\vec{c} \in \Omega$ (see Figure 7, bottom). Therefore, the image translation by cropping is necessary to avoid a bias from the in-plane coronary position. The neural network NN: $I^{60\times60}(\vec{c}) \mapsto p(\vec{c})$ takes a 2D coronary cross-section image $I^{60\times60}(\vec{c})$ as input and delivers a predicted artifact probability $p(\vec{c})$ as output.

3. EXPERIMENTS AND RESULTS

We use the Microsoft Cognitive Toolkit (CNTK) as the framework for deep learning. The neural network is a feed-forward 20-layer ResNet [9], where the number of filters is doubled to $\{32, 64, 128\}$. The Adam optimizer [10] with an initial learning rate of 0.01 (decreasing with a factor of 2 after every 10^{th} epoch), a minibatch size of 100 and a momentum of 0.8 is defined as the learning setup. A classification accuracy of $94.4\% \pm 2.9\%$ by stratified four-fold cross-validation is achieved by the proposed network architecture and hyperparameter selection. The result is divided into a ratio of 47.1% : 47.3% : 2.8% : 2.9% for the rates TN : TP : FN : FP, where *positive* refers to the class *artifact*.

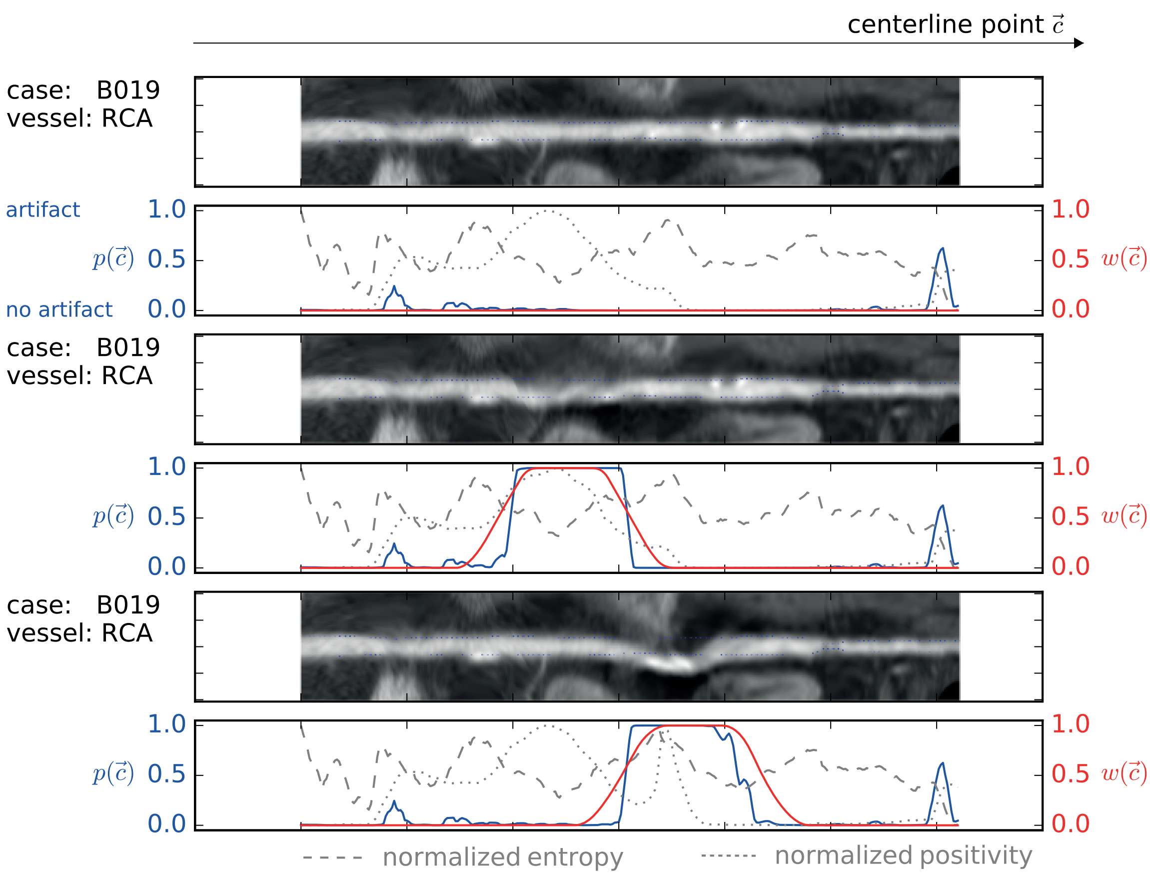

Figure 7: The multiplanar reformation (MPR) is visualized in the upper row. The predicted artifact probability $p(\vec{c})$, the normalized entropy, the normalized positivity and the corresponding weight value $w(\vec{c})$ of each centerline point $\vec{c}$ are plotted in the lower row. The reference state of no motion is given in the first subplot. The vessels in the second and the third subplot are locally perturbed with a motion strength of $s = 8$. A vessel shift compared to the original coronary centerline position can be observed in subplot three. In contrast to the hand-designed measures, the areas of high activations in class *artifact* conform with the regions of motion influence.

3.1 RECOGNITION CHALLENGE

In order to visually evaluate the quality of the learned motion artifact measure, the following challenge is developed. Starting from the original CT image volume of the test case, local motion is introduced at an arbitrary point in the coronary tree. Therefore, a little adjustment in the forward model is performed. The weight mask $w(\vec{x})$, which limits the motion to a desired area, is generated now by dilation and subsequent uniform filtering of a single centerline point in the coronary tree. For each centerline point $\vec{c}$, a cross-section of the local motion perturbed image volume is sampled and classified via CNN. The experiment investigates, whether the trained artifact measure is able to detect the region of motion, given the approximate location of the coronary artery.

Figure 7 shows an exemplary result of this recognition challenge. The red marked weight $w(\vec{c})$ corresponds to the relative displacement width. The blue marked predicted artifact probability $p(\vec{c})$ is smoothed by a five point running average to increase the robustness. For comparison, the gray marked handcrafted artifact measures from [4] are provided. In the paper of Rohkohl et al., the normalized entropy and the normalized positivity are merely used as relative artifact measures to compare the same image volume in different motion states. In contrast to our deep-learning-based measure, the handcrafted ones are obviously not suitable for a section-wise classification, due to a missing robustness regarding the variations mentioned in section 2.2. Apart from a few exceptions, high motion artifact measures $p(\vec{c})$ predicted by our neural network are correctly located at the created motion window. The experiment shows that the network is able to learn motion artifact pattern.

3.2 EXTENSION TO 3D CROSS-SECTIONAL VOLUMES

The sampling and learning processes are also adapted to 3D input data to investigate how far these additional information provide a benefit. For example the differentiation between bifurcations and blurring artifacts might be simplified. The input feature size of the network is extended to $60 \times 60 \times 11$ voxels, where the third dimension is sampled with a lower resolution of 0.8 mm per voxel orthogonal to the cross-section plane. This delivers a depth range of ± 4 mm around each centerline point. The transformations of the online data augmentation are limited to the first two dimensions. Intensity clipping, network architecture (except for the first layer) and hyperparameter selection remain the same. By this configuration, a classification accuracy of $95.6\% \pm 2.7\%$ with a ratio of 47.9% : 47.7% : 2.1% : 2.3% for the rates TN : TP : FN : FP is archived by stratified four-fold cross-validation. Even without extra hyperparameter tuning on the changed input dimension, the measure has slightly improved. Nevertheless, higher memory requirements and execution time has to set against it.

4. CONCLUSIONS

We developed a machine-learning-based measure for motion artifacts in coronary arteries. A forward model is presented, which uses artificial motion vector fields and the MC-FBP algorithm for ground truth data generation. Supervised learning of a CNN is performed and high predictive accuracy is achieved on the artificially motion perturbed data. Areas of artificial motion are correctly identified, given the approximate location and orientation of the coronary arteries. In future work, we are planning to adapt the forward model, to produce data with more accurate ground truth artifact level. Therefore, the orientation of the gantry and the coronary arteries will be considered in the motion model.

The proposed measure is based on nine clinical datasets. In practice, CT images for non-invasive coronary angiography are acquired with a wide variety of scanner types and imaging protocols. For instance, cardiac motion leads to different artifact shapes in helical and step-and-shoot scans. In this paper, we demonstrate the feasibility of accurate motion artifact recognition in the coronary arteries using deep learning in the first place. These promising results warrant the collection of hand-labeled data and further studies to assess the transferability of these initial results to motion artifact prediction in clinical practice. So, for the next stage, it has to be investigated, whether the achieved performance also holds on a large number of real datasets and to what extent a network fine-tuning is required.

REFERENCES

[1] van Stevendaal, U., von Berg, J., Lorenz, C., and Grass, M., "A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography," *Medical Physics* **35**(7), 3239–3251 (2008).

[2] Isola, A. A., Grass, M., and Niessen, W. J., "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," *Medical Physics* **37**(3), 1093–1109 (2010).

[3] Bhagalia, R., Pack, J. D., Miller, J. V., and Iatrou, M., "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Medical Physics* **39**(7), 4245–4254 (2012).

[4] Rohkohl, C., Bruder, H., Stierstorfer, K., and Flohr, T., "Improving best-phase image quality in cardiac CT by motion correction with MAM optimization," *Medical Physics* **40**(3) (2013).

[5] McGee, K. P., Manduca, A., Felmlee, J. P., Riederer, S. J., and Ehman, R. L., "Image metric-based correction (autocorrection) of motion effects: analysis of image metrics," *Journal of Magnetic Resonance Imaging* **11**(2), 174–181 (2000).

[6] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[7] Schäfer, D., Borgert, J., Rasche, V., and Grass, M., "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography," *IEEE Transactions on Medical Imaging* **25**(7), 898–906 (2006).

[8] Vembar, M., Garcia, M., Heuscher, D., Haberl, R., Matthews, D., Böhme, G., and Greenberg, N., "A dynamic approach to identifying desired physiological phases for cardiac imaging using multislice spiral CT," *Medical Physics* **30**(7), 1683–1693 (2003).

[9] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[10] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).