CT-Value Conservation based Spatial Transformer Network for Cardiac Motion Correction

Xuan Xu, Peng Wang, Liyi Zhao, Guotao Quan*

ABSTRACT

Artifact correction is a great challenge in cardiac imaging. During the correction of coronary tissue with motion-induced artifacts, the spatial distribution of CT value not only shifts according to the motion vector field (MVF), but also shifts according to the volume change rate of the local voxels. However, the traditional interpolation method does not conserve the CT value during motion compensation. A new sample interpolation algorithm is developed based on the constraint of conservation of CT value before and after image deformation. This algorithm is modified on the existing interpolation algorithms and can be embedded into neural networks with deterministic back propagation. Comparative experimental results illustrate that the method can not only correct motion-induced artifacts, but also ensure the conservation of CT value in the region of interest(ROI) area, so as to obtain corrected images with clinically recognized CT value. Both effectiveness and efficiency are proved in forward motion correction process and backward training steps in deep learning. Simultaneously, the visualized motion vector field transparentizes the correction process, making this method more interpretable than the existing image-based end-to-end deep learning method.

Keywords: Interpolation, Cardiac Motion Correction, Convolutional neural networks

1. INTRODUCTION

Recent study shows that cardiovascular disease is still the largest worldwide. Coronary Computed tomography angiography(CCTA) is a crucial technology to diagnose coronary heart disease as a simple, fast, noninvasive and safe imaging method. However, the beating characteristics, especially patients with high heart rates, introduce motion artifacts to the reconstruction, which significantly decreases the quality and confidence of the image and potentially limits the evaluation of coronary arteries or even makes misinterpretion. Existing technologies try to suppress artifacts from both hardware and software. Limited by the physical and mechanical properties of CT equipment, even the small incremental gain of the frame rotation time needs to make great efforts in engineering design.

Some methods try to improve image quality during reconstruction[1].[2]Rohkohl et al[3]initially proposed a Metric-based correction method and later improved and extended. Some registration-based also have shown good performance in compensating for strong motion artifacts.A classical non-rigid registration algorithm[4] uses the motion vector field estimated by bidirectional labeled point matching (BLPM) algorithm to perform 3D warping on a series of partial reconstructions. This algorithm uses thin plate spline interpolation algorithm (TPS) for interpolation. TPS, as a very robust spatial data interpolation method, was introduced by Duchon et al.[5] into geometric design, which is commonly used for non-rigid registration. Since the structure of TPS is differentiable, Spatial Transformer Networks (STN)[6] applies it in the network to achieve spatial alignment of feature maps.

Deep learning based cardiac motion correction method, as a particular case of image deblurring, usually follows two common ways: using deep network to estimate the motion vector field, and then combined with the traditional warp algorithm to deform. Another way is to learn from image to image, that is, the trained neural network can output the corrected image directly. Methods proposed by S. Jun[7] and N. Fu[8] have successfully proved that CNN has the ability to generate and learn coronary motion patterns.[9][10] has successfully applied

Corresponding author: Guotao Quan* is with the Shanghai United Imaging Healthcare Co., Ltd email: guotao.quan@united-imaging.com
Xuan Xu is with the Shanghaitech University
Peng Wang and Liyi Zhao is with the Shanghai United Imaging Healthcare Co., Ltd,2258 Chengbei Rd, Jiading District, Shanghai China

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 123041X
 · 0277-786X · doi: 10.1117/12.2646430

STN for end-to-end training. The obtained images can be well registered in shape, but due to the limitations of traditional interpolation in value conservation, the accuracy of the CT value needs to be investigated. Based on the principle of CT imaging, the overall integral value of the reconstruction image is not related to the states of motion of object in the fixed Field of View. However, various system biases may be introduced in the reconstruction process, resulting in differences between motion and static reconstruction results.

Separating raw data to generate multiple partial angle reconstructions and applying different MVFs with affine transformation is one way to eliminate the interpolation issue, but it requires more detailed and exact motion patterns for each subset. In order to solve the interpolation issue and meanwhile avoid increasing the complexity of correction process, the conservation integration constraint interpolation method is designed. This paper takes into account the proportional coefficient between the integral value and the area of deformation grid. This new deformation interpolation method is based on the existing interpolation method and can be embedded into the classical spatial transform network for back propagation.

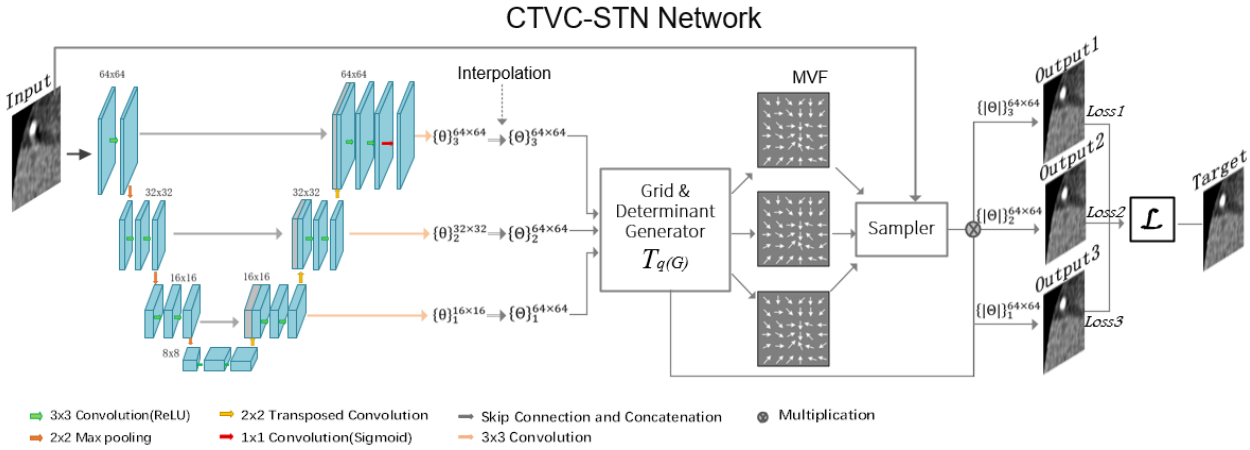

Figure 1: Patch-based coronary correction pipeline of the proposed interpolation method. The front network receives an input which can be central axial, sagittal, or coronal slices of a motion-corrupted volumetric ROI and generate deformed parameters $\theta_1, \theta_2, \cdots, \theta_k$. CTVC-STN network feed with $\theta_1, \theta_2, \cdots, \theta_k$ to interpolate to get $\Theta_1, \Theta_2, \cdots, \Theta_n$ $(k \leq n)$ and sampling to output an corrected image.

2. MATERIALS AND METHODS

The outline of the proposed coronary correction pipeline is shown in Fig.1. Firstly, the front Network is designed to output three sets of deformation parameters θs.Then the deformation parameters of each pixel in the whole image are obtained by interpolation of θs, and are used to warp the coronary ROI. Simultaneously, calculate the deformation coefficient of the area of the deformed grid. Then the deformation coefficient is multiplied by each corresponding pixel to obtain the final image which is the closest to the ground truth, not only in shape but also in CT value. The whole process can be back-propagated in the convolutional Network.

2.1 Deformation parameters estimation

Inspired by literature,[11] a changed 2D-UNet network with deep supervision[12] is selected as the front network to generate deformation parameters $\theta_1, \theta_2, \cdots, \theta_k (k \geq 1)$ in decoder path, and it can be substituted by other suitable network structures. Since convolutional layers of different depths have different receptive fields, different from Unet,[13] this network simultaneously outputs the learned deformation parameters from features extracted at different scales. For the specific coronary artery correction task in this paper, the features of three different scales are selected to estimating deformation parameters at the same time and calculate loss respectively, and finally the total loss is calculated by the combination of the three losses. This configuration can not only increase the stability of the network during training, but also support pruning the network during testing, which can increase

the testing speed while ensuring the correction accuracy, thereby reducing the amount of network parameters within a controllable range.

2.2 CT Value Conservation Network based on Spatial Transformer (CTVC-STN)

In order to realize CT value conservation while ensuring the deformation, CTVC-STN network is proposed by improving on the basis of Spatial transformer Network as STN has shown some deficiencies in the end-to-end training of motion correction. First, STN introduces full connection layers to output an affine matrix θ with the size of 2×3 which increases the difficulty of training and limits that STN can only be used for small-size features.Secondly, STN uses conventional affine transformation and interpolation function to warp, in which the interpolation function can be bilinear interpolation, bicubic interpolation and thin plate spline interpolation. However, for coronary images with artifacts caused by different motion patterns, the simulated data as Fig.2 show that, ideally, the sum of CT value of the stationary and motion reconstructions are conserved. Therefore, this paper designs a deformation interpolation network CTVC-STN to keep the CT value conserved.

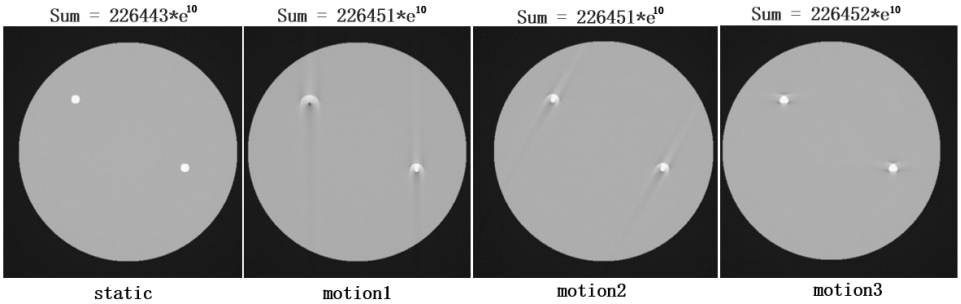

Figure 2: One simulated case of the stationary reconstruction and motion reconstruction.The first image is the result of stationary reconstruction, and the rest three motioned images are reconstructed from different gantry starting positions and reconstruction angles. The CT value sums of the four images are approximatly equal. The CT value difference among them comes from system resolution error.

As illustrated in Fig.1, The Grid & Determinant Generator feeds with deformation parameters $\theta \mathrm{s} = \theta_1, \theta_2, \cdots, \theta_k$, which are generated by front network. According to $\theta_1, \theta_2, \cdots, \theta_k$, input image is averagely discretized into k pixels. For the presupposed conditions, $\theta_1, \theta_2, \cdots, \theta_k$ are the accurate deformation parameters corresponding to k pixels, so the coordinates of these points can be directly obtained. Furtherly, combined with sampling and bilinear interpolation, the coordinates and pixel values of all points in the initial deformed image are also obtained. Backward mapping is used for sampling, that is, the pixel value of each point of the deformed image is traversed to find the corresponding coordinates on the original image, and then the surrounding pixels are used for simple interpolation. This method avoids holes generated during forward mapping.

To realize CT value conservation of an image before and after deformation, ideally, when one pixel p of deformed image is contributed by q pixel grids on the original image where $q > 1$,as shown in Fig.3 e), $q > 2$, for these q pixels, calculating the ratio of covered area of each pixel grid of initial image. Then, the sum of the pixel values of these q grids multiplied by the corresponding ratio should be the exact pixel value of that pixel. An alternative solution is used here to alleviate the situation that computational complexity increase as initial image size becomes larger. To our knowledge, in the two-dimensional space, the geometric mathematical significance of the determinant of the matrix represents the directed area surrounded by two vectors.[14] Based on this mathematical theory, the area of each pixel in the affine transformation grid can be obtained.

A sample interpolation example is shown in Fig.3 a)-c). Geometrically, consider the pixels of the image as squares rather than points and the pixel value as the center points of the input's corner pixels. Fig.3 a) is the initial image with a size of 4x4, Fig.3 b) is the result of bilinear interpolation and sampling with scaling factor of 1/2. As it illustrated in Fig.3 b), the summation of all pixel value is non-conservation. According to the affine matrix of each pixel of this transformation and its determinant are as follows:

$$\theta = \begin{pmatrix} 2 & 0 & \delta x \\ 0 & 2 & \delta y \end{pmatrix}, \|\theta\| = 4 \tag{1}$$

Therefore, the value summation of the image obtained by multiplying Fig.3 b) by $\|\theta\|$ is equal to the initial image, as shown in Fig.3 c).

The proposed method maintains CT value conservation meets the following constraints: First, if the deformed image exceeds the size of the deformed grid, the boundary pixels will be lost in the sampling process,which are illustrated in Fig.3 e),f). Second, at the ideal limit resolution, even very exaggerated deformations will become very smooth. Therefore, the method can realize the conservation of CT value under ideal conditions. However, due to the difficulty of implementation and computational complexity, very precise grids are not used during implementation, which will cause errors. However, the following experiments show that the method can control the error within the clinically acceptable range, as shown in Fig.5.

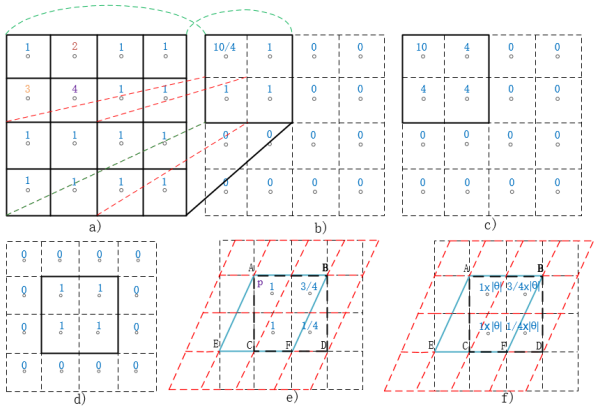

Figure 3: A sample interpolation example. a) initial image of 4x4. b) result of bilinear interpolation, 2x2. c) result of proposed interpolation method with CT value-sum conservation. d)initial image with 2x2 focused ROI. e) the result of bilinear with deformation parameters θ. f)the result of proposed method. b),c) is still in the focused ROI, while e),f) not, so CT value conservation can be obtained in b),c), but invalid in e),f).

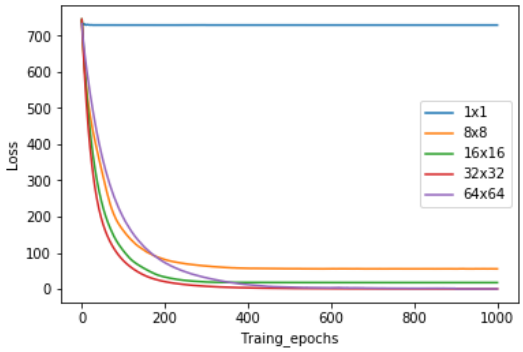

Figure 4: The trend of different numbers of deformation parameters on the loss convergence as the training epoch increase. The loss is Mean Square Error.

3. RESULTS

3.1 Image acquisition

The combination network above is performed using supervised end-to-end training strategy. To train this model, the ground truth used in the training process are anonymous motion-artifacts-free cases with United Imaging Healthcare(UIH) uCT ATLAS devices. Referring to the forward model for simulating cardiac motion method proposed in related literature[15,16] and our knowledge of cardiac beating patterns, artificial motion vector fields is generated to simulate all kinds artifacts. The artificial blurred data are input data for training. 9600 samples of 2D coronary patches with the size of 64x64 based on the above artifact simulation methods were generated. The samples were divided into 80% training data and 20% validation data. Test data set involves real motion blurred cases to examine the effectiveness of network and simulation method.

3.2 Neural Network Training

The training was performed on an NVIDIA TITAN RTX for 1000 epochs using an Adam optimizer with 0.1 decay, The batch size is 16 and the loss function is Mean Square Error(MSE).

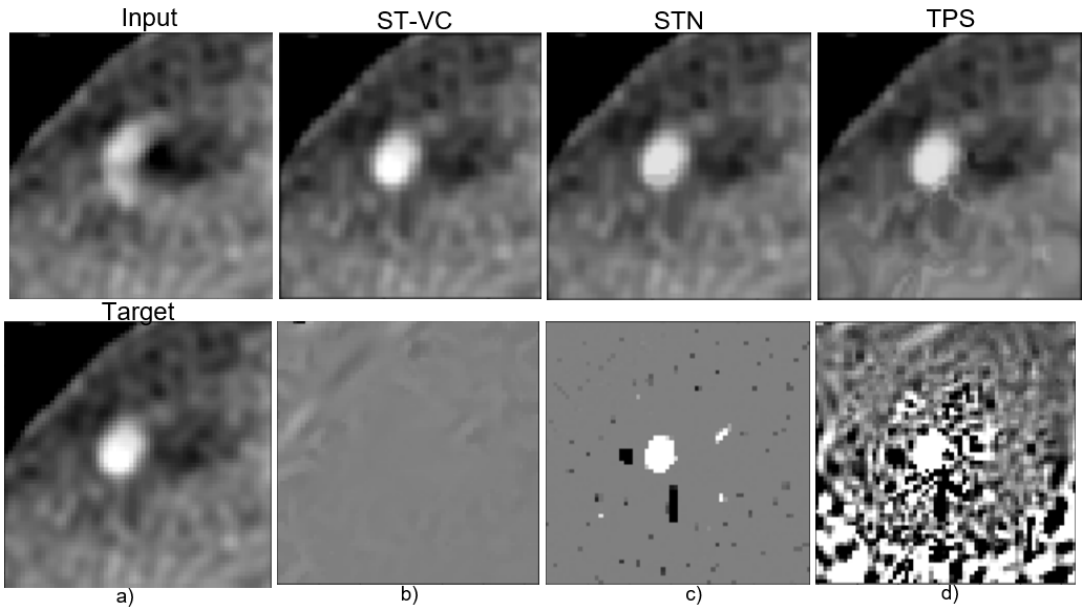

Figure 5: The x-y plane of one image patch is visualized before and after different methods. Among these, CTVC-STN shows best correction both in shape and CT value. The CT value sum of Input is 3914156, CTVC-STN: 3914182, STN-biliear: 3909930, STN-TPS: 3908637, and Target: 3914188. CTVC-STN: can obtain an aproximate value coservation with a value error ratio of 1e-6. Figure b),c),d) show the differences between ground truth and the Results of STN-bilinear c), STN-TPS d) , our method b).

3.3 Evaluation

Real data from several clinical patients with severe artifacts were used to test the trained network model.

The number of deformation parameters directly affects the accuracy of artifact correction. Fig.4 shows the trend of different deformation parameters on the loss convergence as the training epoch increases. It can be seen that since coronary motion is a relatively complex non-rigid deformation, it is impossible to correct the deformation of the entire image with a single parameter, so the loss is maintained at a relatively high level. With the increase of deformation parameters, the network can gradually learn complex motion deformation, and its number is positively correlated with the correction result. When the number of θ generated by the network is the same as the number of pixels in the input image, it is equivalent that each pixel has its own specific displacement vector, and a more accurate shape correction can be achieved under this configuration.

When Mean Square Error (MSE) and structural similarity index measure (SSIM) were used as the loss function, the above corrected CT value will be slightly deviated. The network needs to add the directed area of the deformation vector for further numerical correction. As shown in Fig.5, TPS and STN failed to maintain the image CT value conservation before and after correction, while CTVC-STN can achieve approximate conservation of CT value within the range of loss not exceeding 1e-6. Real cases were also tested to demonstrate the effectiveness of this method as shown in Fig.6, which shows that the designed network has a good performance in correcting drastic artifacts of coronary images in three planes.

4. CONCLUSION

The key contribution of this work is the solution that provides individual deformation of each pixel of the image, and can maintain the approximate conservation of the CT value in deformation. This method is a supplement to the existing interpolation algorithm and can be used in the network to support back propagation. A novel framework for motion correction of CCTA were experimented to verify the validity of this method. Compared with the existing interpolation method such as bilinear interpolation and TPS, it can get images with more

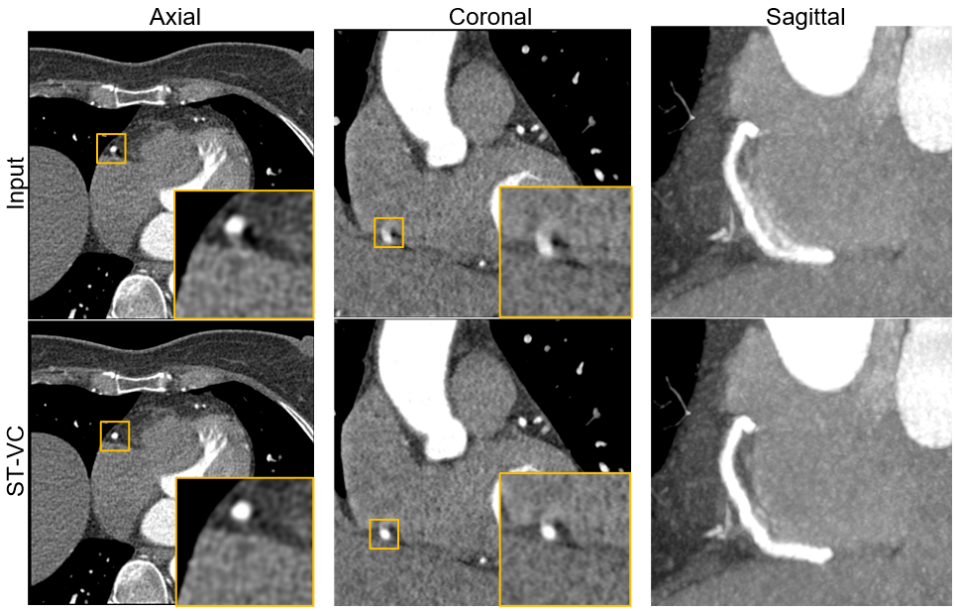

Figure 6: One clinical case to show that the proposed pipeline is robust in related artifacts in three planes images of patient's coronary .

accurate CT value and lower MSE. For the further works, we would focus on improving the network capacity that allows a full field of view image as input and realize self-attention to the regions of coronary or anywhere artifacts appear. The second improvement would take place to extend this method into a 3D network to achieve direct 3D coronary volume correction.

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**(7Part1), 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,"

[3] 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).

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

[5] Duchon, J., "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in [*Constructive Theory of Functions of Several Variables*], *Lecture Notes in Mathematics*, 85–100, Springer Berlin Heidelberg, Berlin, Heidelberg (2006).

[6] Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K., "Spatial transformer networks," *CoRR* **abs/1506.02025** (2015).

[7] Jung, S., Lee, S., Jeon, B., Jang, Y., and Chang, H. J., "Deep learning cross-phase style transfer for motion artifact correction in coronary computed tomography angiography," *IEEE Access* **PP**(99), 1–1 (2020).

[8] Fuin, N., Bustin, A., Küstner, T., Oksuz, I., and Prieto, C., "A multi-scale variational neural network for accelerating motion-compensated whole-heart 3d coronary mr angiography," *Magnetic Resonance Imaging* **70** (2020).

[9] Fish, N., Zhang, R., Perry, L., Cohen-Or, D., and Barnes, C., "Image morphing with perceptual constraints and stn alignment," (2020).

[10] Yoo, I., Hildebrand, D., Tobin, W. F., Lee, W., and Jeong, W. K., "ssemnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features," (2017).

[11] Zhou, Z., Siddiquee, M., Tajbakhsh, N., and Liang, J., "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging* **39**(6), 1856–1867 (2020).
[12] ŏ01FSaining Xie ŏ01FPatrick W. Gallagher, C.-Y. L., "deep1y supervised nets,"
[13] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], (2015).
[14] Kolman, B. and Shapiro, A., "Matrices and determinants," *Algebra for College Students (Revised and Expanded Edition)* , 419–440 (1982).
[15] "Deep-learning-based ct motion artifact recognition in coronary arteries," in [*SPIE Medical Imaging Conference*],
[16] Maier, J., Lebedev, S., Erath, J., Eulig, E., Sawall, S., Fournié, E., Stierstorfer, K., Lell, M., and Kachelrie, M., "Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac ct," *Medical Physics* .