

Received April 14, 2020, accepted April 27, 2020, date of publication April 30, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991445

Deep Learning Cross-Phase Style Transfer for Motion Artifact Correction in Coronary Computed Tomography Angiography

SUNGHEE JUNG^{1,4}, (Member, IEEE), SOOCHAHN LEE², (Member, IEEE),
BYUNGHWAN JEON³, YEONGGUL JANG¹, AND HYUK-JAE CHANG^{4,5}

¹Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul 03722, South Korea

²School of Electrical Engineering, Kookmin University, Seoul 02707, South Korea

³School of Computer Science, Kyungil University, Gyeongsan 38428, South Korea

⁴CONNECT-AI Research Center, Yonsei University College of Medicine, Seoul 03721, South Korea

⁵Division of Cardiology, Severance Cardiovascular Hospital, Yonsei University College of Medicine, Yonsei University Health System, Seoul 03721, South Korea

Corresponding author: Soochahn Lee (scllee@kookmin.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) (Intelligent SW Technology Development for Medical Data Analysis and Autonomous Digital Companion Framework and Application) under Grant 2018-0-00861 and Grant 2017-0-00255, in part by the National Research Foundation of Korea (NRF) under Grant NRF-2019R1F1A1063656, and in part by the Korean Government (Ministry of Science and ICT (MSIT) and Ministry of Education (MOE)).

ABSTRACT Motion artifacts may occur in coronary computed tomography angiography (CCTA) due to the heartbeat and impede the clinician's diagnosis of coronary arterial diseases. Thus, motion artifact correction of the coronary artery is required to quantify the risk of disease more accurately. We present a novel method based on deep learning for motion artifact correction in CCTA. Because the image of the coronary artery without motion (the ground-truth data required in supervised deep learning) is medically unattainable, we apply a style transfer method to 2D image patches cropped from full-phase 4D computed tomography (CT) to synthesize these images. We then train a convolutional neural network (CNN) for motion artifact correction using this synthetic ground-truth (SynGT). During testing, the output motion-corrected 2D image patches of the trained network are reinserted into the 3D CT volume with volumetric interpolation. The proposed method is evaluated using both phantom and clinical data. A phantom study demonstrates comparable results to other methods in quantitative performance and outperforms those methods in computation time. For clinical data, a quantitative analysis based on metric measurements is presented that confirms the correction of motion artifacts. Moreover, an observer study finds that by applying the proposed method, motion artifacts are markedly reduced, and boundaries of the coronary artery are much sharper, with a strong inter-observer agreement ($\kappa = 0.78$). Finally, evaluations using commercial software on the original and resulting CT volumes of the proposed method reveal a considerable increase in tracked coronary artery length.

INDEX TERMS Computed tomography, deep learning, image restoration, motion correction and analysis, coronary angiography.

I. INTRODUCTION

Coronary artery disease (CAD), also known as ischemic heart disease, is the leading cause of death globally [1]. Recently, non-invasive coronary computed tomography angiography (CCTA) has been widely adopted for the diagnosis of CAD. The diagnostic accuracy of CCTA is comparable to that of conventional invasive coronary angiography, with a significantly lower risk of complications [2]. However, motion

artifacts may occur in the acquisition of CCTA, which can cause errors in tracking or segmentation of the coronary artery.

Prospective electrocardiography (ECG)-gating can be used to address this problem by timing the CCTA acquisition to the most quiescent phase of the heartbeat, although motion artifacts can occur if the heart rate is very high or irregular. Drugs such as beta-blockers may generally be administered to slow down the patient heart rate when it is higher than 65 beats per minute, but often with limited efficacy [3] or more frequent angina and ischemia [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu¹.

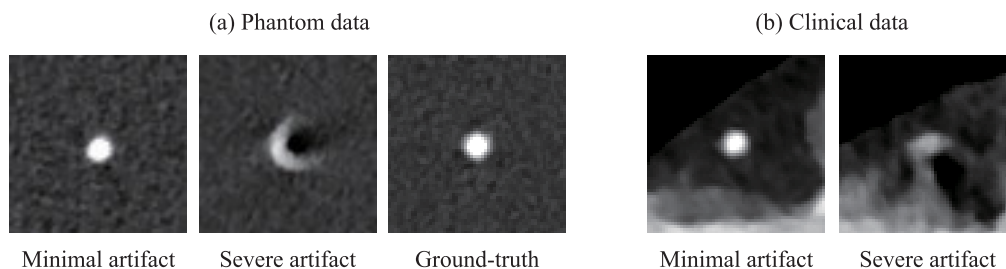


FIGURE 1. (a) CT images of the phantom data that imitates the coronary artery. The degree of motion artifacts varies from minimal, when the heart motion is slow, to severe, when fast. Ground-truth data, with no absolute motion, can be obtained when the phantom is completely stopped. (b) CT images of the coronary artery of clinical data. Ground-truth data cannot be obtained because the heart cannot be stopped.

To the best of our knowledge, previous image processing methods for CCTA motion compensation are based on motion estimation through image registration or minimizing a motion artifact metric. Methods based on 3D-3D non-rigid image registration [5]–[7] have demonstrated excellent motion compensation results. However, they require multiple phases of CT images, and it is possible that motion estimation is erroneous in the presence of motion artifacts, which in turn would lead to the degradation of motion compensation. Rohkohl *et al.* [8] presented the method of improving best-phase image quality using a single 3D reconstructed image; its motion estimation is based on minimizing motion artifact metrics. This method has produced superior results in more quiescent cardiac phases; in contrast, in more rapid cardiac phases, 3D-3D nonrigid image registration-based methods produce superior image quality because they use information from neighbor phases. Kim *et al.* [9] used a partial angle reconstructed (PAR) image, which is reconstructed using a smaller angular range than a short scan, to improve temporal resolution. They estimate motion using PAR images based on 3D-3D nonrigid registration, so they require multiple phases of CT images. Hahn *et al.* [10] presented a coronary motion-compensation method based on PAR images from short-scan data, that has benefits concerning dose efficiency. That method optimizes an image artifact measuring cost function for motion estimation, as proposed by Rohkohl *et al.* [8]. Using the motion vector field estimated by various approaches, such as registration and minimizing an artifact metric, the aforementioned methods require motion-compensated reconstruction from raw or projection data.

In this study, we present an image-based coronary motion correction method that does not require projection or back-projection steps and can be applied when raw data are not available. Our approach assumes that a physician can estimate motion artifacts and envision a motion-corrected image more accurately when he or she has more experience. Our hypothesis is that motion correction in CCTA can be learned with a large set of training images. Coincidentally, methodologies based on deep learning have demonstrated revolutionary performance in various

domains [11], [12]. Deep learning-based methods effectively address relevant problems, including image super-resolution [13], image denoising [14], and image deblurring [15]. These methods adopt a supervised learning approach where it is assumed that accurate ground-truth corresponding to the input data is available.

Critically, it is medically impossible to obtain precise corresponding coronary CT images devoid of motion artifacts, which are the ground-truth images required in supervised learning. This problem and comparison using phantom data are visualized in Fig. 1. This problem is among the many various situations when ground-truth is unattainable, which often occurs when learning from images, including urinary bladder segmentation [16], 3D pose estimation [17], the social behavior of honeybees [18], learning eye gaze direction estimation [19], and object detection in indoor scenes [20]. These methods offer effective ways to generate the ground-truth images or realistic synthetic images appropriate for their particular problems by utilizing Position Emission Tomography (PET) acquisitions for automatic urinary bladder segmentation in CT images [16], rendering various configurations of a single 3D model [17]–[19] or rendering a composite 2D image of various partial 2D images [20]. However, these processes are most likely insufficient for the motion correction of coronary arteries in 3D CCTA because a single 3D model cannot represent variations in the 3D structure, nor can a composite image of various CCTA slices be used to generate a 2D slice consistent with a realistic 3D volume.

We propose a novel method using a style transfer method, such as those proposed in [21], [22], to generate a synthetic ground-truth (SynGT). We apply style transfer to image patches from 4D CT volumes containing phases with large and small amounts of motion artifacts, as depicted in Fig. 2. Compared to using the patches directly, the local deformations that occur from the heartbeat motion can be considered using style transfer. Our aim is to suppress the effect of genuine appearance change and isolate only the effect of motion artifacts. Using the SynGT, we can subsequently learn to generate images with reduced motion artifacts from the corresponding training input images.

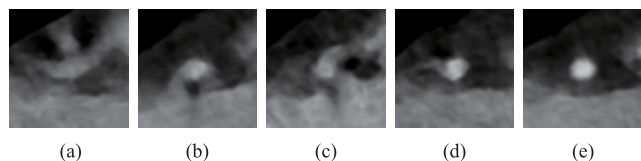


FIGURE 2. Appearance of the motion artifacts of a coronary artery in different phases within a full heartbeat cycle, sampled from a 4D CT. Based on a 5-point Likert scale, as described in Sec. III-B4, the image patches are (a) completely unreadable or have (b) significant motion artifacts, (c) apparent motion artifacts, (d) minor motion artifacts, or (e) no motion artifacts.

The overall process of the proposed method is as follows. First, we compile a dataset of spatially corresponding 2D image patches, extracted from 3D CT volumes at different phases within a 4D CT. Next, we apply style transfer with the patch from the phase with significant motion artifacts as the source, and the patch from the phase with minimal motion as the target, to generate the SynGT. We then train the motion artifact compensation network (MAC-net) using the patches with significant motion and the SynGT. At the test stage, we assume a 3D CT volume with motion artifacts is given and that the centerline of the coronary artery has been annotated. We generate the 2D cross-sectional image patches along the centerline and feed them into the trained motion artifact correction network. The output motion-corrected 2D image patches are reinserted into the original 3D CT volume with volumetric interpolation to obtain the final motion-corrected 3D CT volume.

The three main contributions of our work are as follows: First, we propose a method for motion correction using deep learning, in which the ground-truth is synthesized using style transfer between corresponding 2D image patches of the coronary artery extracted at different phases within a 4D CT. Second, we propose a method to perform motion correction on the coronary arteries of the 3D CT volume by reinserting motion-corrected output 2D patches with volumetric interpolation. Third, we provide extensive quantitative and qualitative evaluations to demonstrate the degree of motion correction after applying the proposed method.

This paper is an extension of our previous work in [23], with extended quantitative and qualitative evaluations. We include a comparative analysis with other methods using the phantom dataset, which is publicly available, and the results are presented in Sec. III-A. In Sec. III-B, we evaluate the proposed method quantitatively by measuring motion artifact metrics [24] and image quality metrics for the clinical dataset. An observer study, a straightened curved planar reformation (CPR) example, and the tracking results of the proposed method are presented in Sec. III-B for qualitative analysis.

II. METHODS

Our approach is to solve the problem of motion correction using a deep neural network, such as a convolutional neural network (CNN). Thus, given a new 3D CT volume at the

test stage, we want to generate a motion-corrected version of that volume as the output of our trained CNN. Accordingly, the proposed method is designed based on the following decisions:

- 1) The network input and output are defined in terms of corresponding image patches of the coronary artery with and without motion artifacts.
- 2) 4D CTs are used to achieve corresponding pairs of patches extracted from the same patient at relatively slower and faster heart motions within the heartbeat cycle. It is clinically impossible to achieve images with no motion artifacts because the heart motion cannot be arbitrarily stopped.
- 3) Style transfer is used to synthesize a patch with the corresponding phase, in which the appearance is preserved, but motion artifacts are reduced. The corresponding patch at the slower motion acts as a reference to guide artifact correction while preserving local appearance at that motion phase. This synthesized patch, rather than the patch at the slower motion, is defined as the ground-truth to learn motion correction.
- 4) A very deep CNN [13] is used as the deep learning network for motion correction. The deep structure helps to generalize the complex changes between the original and synthesized patches.
- 5) Motion-corrected patches are reinserted and interpolated into the original 3D CT volume to compensate for the motion artifacts of the entire coronary artery.

Based on these decisions, the overall framework comprises the following subprocesses: 1) extraction of corresponding coronary artery patches from 4D CT, 2) synthesis of morphed source patches using style transfer, 3) training and applying the motion artifact correction network, which we term MAC-net for the patch-wise motion artifact correction, and 4) reinsertion of motion-corrected patches to the 3D CT volume. It is important to note that the style transfer network for generating synthetic motion-corrected patches and MAC-net have different roles. Style transfer is performed between pairs of patches on corresponding points on the coronary arteries in different phases in a 4D CT and alters only the style (local texture), not the content (structure). Our assumption is that motion artifacts are closer to style and thus can be reduced using this process. In contrast, patches with motion artifacts are paired with the generated SynGT from style transfer for supervised learning of the MAC-net for motion compensation. While corresponding patch pairs from 4D CT are required for removing motion artifacts by style transfer, once the MAC-net is trained, any 3D CT patch with motion artifacts can be given as an input to generate a motion-compensated patch at test time. Fig. 3 provides a visual summary of the second and third subprocesses, which are technically the most critical. The subsequent subsections describe these processes and the original 4D CT data in detail.

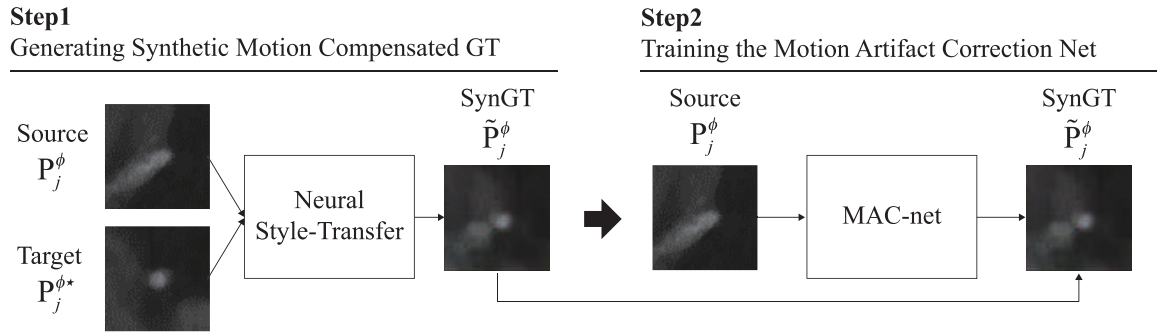


FIGURE 3. Workflow of the proposed method. In step 1, generate synthetic motion-corrected patch (SynGT) using style transfer method. In step 2, train the motion artifact correction network (MAC-net) using SynGT. Detailed descriptions of steps 1 and 2 are in Sections II-B and II-C, respectively.

A. EXTRACTING CORRESPONDING CORONARY PATCHES FROM 4D CT

We used 4D CT images acquired by retrospective gating using a dual-source CT scanner (SOMATOM Definition Flash, Siemens). All raw data 0% – 90% were reconstructed in 10% increments of the heartbeat cycle (R-R interval). The images reconstructed at 40% and 70% usually have the highest image quality in terms of motion artifacts because they are around the end-systole and end-diastole, respectively. The images reconstructed at other phases mostly contain considerable motion artifacts of the coronary arteries.

Without loss of generality, we will hereby specifically focus only on the middle of the right coronary artery (mid-RCA) as the region of interest, which generally has the most motion. Given the temporally sampled 3D CT volumes, the mid-RCA is manually annotated by an experienced reader in each volume using commercial coronary analysis software (QAngioCT, Medis Medical Imaging Systems, Leiden, the Netherlands). The first right ventricle (RV) branch and acute marginal branch are defined as the respective start and end points of the mid-RCA.

The mid-RCA centerline C^ϕ of a 3D volume at phase ϕ is represented as a discretized set of ordered 3D coordinates $C^\phi = \{c_i^\phi | 0 \leq i \leq N_c^\phi - 1\}$, where c_i^ϕ denotes the i th 3D point coordinate of C^ϕ . N_c^ϕ denotes the total of number of points within C^ϕ . The exact centerline is approximated as a piecewise linear function between the points in C^ϕ . Thus, the entire length of the mid-RCA centerline is defined as the sum of all distances between subsequent point pairs, and denoted, as $l^\phi = \sum_{i=0}^{i < N-1} ||c_{i+1}^\phi - c_i^\phi||_2$.

To extract the corresponding 2D image patches on C^ϕ , the corresponding points must first be determined. We assume that the start and end points for all ϕ will correspond because they correspond to the same anatomical landmarks: first RV branch and acute marginal branch. A fixed number of M equidistant points $\mathcal{V}^\phi = \{q_j^\phi | 0 \leq j \leq M - 1\}$ each spaced $\frac{l^\phi}{M}$ are sampled between the start and end points of C^ϕ . Because the mid-RCA centerline is being approximated as a piecewise linear function, we applied interpolation

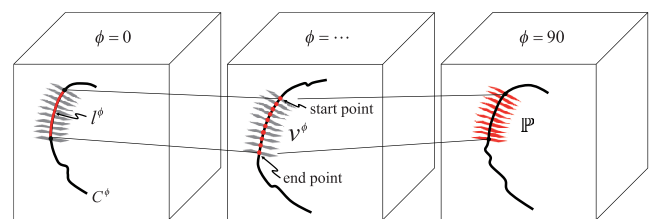


FIGURE 4. Determining positions and normals for corresponding patches of mid-RCA in 3D CT volumes at different temporal phases, included within the full-phase 4D CT volumes. The centerlines of the mid-RCA are assumed to have been manually annotated, and the start and endpoints are manually annotated by anatomical landmarks in the volumes.

to compute the exact equidistant point coordinate. Finally, we define the normal directions \vec{n}_j^ϕ for the planar patches centered at each q_j^ϕ as the tangential direction of C^ϕ at q_j^ϕ . Fig. 4 visualizes this process of determining the corresponding points and the 3D CT volumes at different temporal phases.

The corresponding patches $\mathbb{P} = \{P_j^\phi | 0 \leq j \leq M - 1\}$ are extracted by sampling the voxel intensities on an $R \times R$ discrete grid centered at q_j^ϕ with normal \vec{n}_j^ϕ within the corresponding 3D CT volume. To align the spatial distribution of the grid points physically, we constructed a 2D grid (on the xy-plane as a reference) with 3D coordinates considering the physical dimensions of the CT, and applied translation based on the center point and rotation based on the normal direction, to obtain the projected grid coordinates. Because these coordinates are not integers, bilinear interpolation is applied when assigning intensity values to each pixel in the extracted patch.

B. GENERATING SYNTHETIC MOTION-CORRECTED PATCHES USING CROSS-PHASE STYLE TRANSFER

The corresponding patches extracted from coronary arteries within different phases of a 4D CT are not just different in terms of the severity of the motion artifacts. The motion during the heartbeat also causes differences in its local appearance. We want to obtain the corresponding patch with identical local appearance but without motion artifacts,

because we would like to train a CNN to remove only the artifacts. Because this is clinically unattainable, we aim to synthesize this same-phase-no-artifact patch, \tilde{P}_j^ϕ , using style transfer to the source patch P_j^ϕ with a different-phase-no-artifact patch as the target patch $P_j^{\phi\star}$. We refer to this process as cross-phase style transfer, where $\phi\star$ denotes the phase within the heartbeat when the motion is the slowest, resulting in the minimum number of motion artifacts.

Style transfer is the process of converting only the style, not the contents, of a source image to the style of the target image. The aspects of the style include texture, color, and contrast, both local and global. Content, in contrast, generally refers to the outlines, textures, and colors required to recognize the scene, including the specific objects or persons. In our framework, we assume motion artifacts as part of the style but local appearance as content.

In the proposed framework, we applied a recent method for style transfer using deep neural networks [21], often referred to as the neural style transfer method. The central concept is as follows. First, a CNN—the Visual Geometry Group (VGG) network [25] pre-trained on the ImageNet database [26]—is used to compute local image features subsequently defined as the numerical representation of the content. If we denote the tensor of the CNN features at layer l as F_x^l and F_c^l for the synthesized image \vec{I}_x and content reference image \vec{I}_c , respectively, the loss function for the content is defined as

$$\mathcal{L}_{content}(\vec{I}_x, \vec{I}_c) = \frac{1}{2} \|\vec{I}_x - \vec{I}_c\|_2^2. \quad (1)$$

Next, the numerical representation of the style is defined using the Gram matrix G^l , where each element is the inner product between different CNN features at layer l , as

$$G_{ij}^l = F_i^l \cdot F_j^l, \quad (2)$$

where G_{ij}^l denotes the element at row i , column j of G^l , F_i^l and F_j^l denote the i_{th} and j_{th} features, respectively, corresponding to the i_{th} and j_{th} convolutional kernels, respectively, at layer l . The loss function for style is subsequently defined as

$$\mathcal{L}_{style}(\vec{I}_x, \vec{I}_s) = \frac{1}{2N_x^{l2} \times 2N_s^{l2}} \|G_x^l - G_s^l\|_2^2, \quad (3)$$

where G_x^l and G_s^l are the Gram matrices, and N_x^l and N_s^l are the number of features at layer l , for \vec{I}_x and style-reference image \vec{I}_s , respectively.

Finally, \vec{I}_x is determined by using a gradient descent to minimize the balanced loss, defined as

$$\mathcal{L}_{total}(\vec{I}_x, \vec{I}_c, \vec{I}_s) = \alpha \mathcal{L}_{content}(\vec{I}_x, \vec{I}_c) + \beta \mathcal{L}_{style}(\vec{I}_x, \vec{I}_s), \quad (4)$$

where α and β are coefficients to balance the effect between the content and style loss terms. The process of optimizing the loss in Eq. (4) does not involve training the network, which is fixed. Instead, a modified version \vec{I}_x of the input images is generated. For further details, we refer the reader to [21].

From the review above, \tilde{P}_j^ϕ , P_j^ϕ , and $P_j^{\phi\star}$ correspond to \vec{I}_x , \vec{I}_c , and \vec{I}_s , respectively. Whereas the phase $\phi\star$ with the minimum amount of motion is determined manually, the patches from all other phases ϕ can be assigned as the source, i.e., the reference patch for content P_j^ϕ .

C. TRAINING AND APPLYING THE MOTION ARTIFACT CORRECTION NETWORK

We adopted the Very-Deep network for Super-Resolution (VDSR) [13] to our problem of motion artifact correction. The VDSR is a specific type of CNN, configured with deep layers and gradient clipping suitable for high learning rates in the gradient descent of the training process. The input and output dimensions are set to be identical, we apply supervised learning with our training dataset so that a motion-corrected version of the input image patch is given as the output. We chose VDSR because 1) our problem is primarily a noise reduction problem, and noise reduction is similar to achieving super-resolution, 2) the input data in VDSR is upsampled such that patch sizes of the input and output are assumed to be the same, which is the configuration of our case, and 3) it illustrates acceptable performance and fast convergence during training.

The acceptable performance is primarily due to the deep structure of the network, which combines the very deep CNN model of [25] together with the residual learning of [27]. Whereas skip connections were added at every other convolutional layer in [27], only a single skip-connection from the input to output is created in the VDSR network. This connection learns the difference between the input and output and prevents the vanishing gradient problem. Furthermore, there is a gradient clipping scheme [28] that is often used in training recurrent neural networks to expedite training convergence. We adopt an adjustable gradient clipping scheme [13], in which the gradients are clipped to $[-\frac{\theta}{\gamma}, \frac{\theta}{\gamma}]$ to boost the convergence, where γ denotes the current learning rate and θ denotes gradient clipping. VDSR exhibits the fast convergence of a CNN with a 20-layer network using adjustable gradient clipping. Also, to keep the sizes of all feature maps the same, we pad zeros before convolutions.

The structure of the MAC-net follows the VDSR network, which comprises 20 convolutional layers and 19 rectified linear unit (ReLU) non-linear activation functions, as depicted in Fig. 5. We used 64 3×3 kernels for each convolutional layer. The corresponding cross-phase style transferred patch \tilde{P}_j^ϕ is assigned as the GT output for the input patch P_j^ϕ . The loss function is defined as the mean squared error $\frac{1}{2} \|\tilde{P}_j^\phi - (P_j^\phi - f(P_j^\phi))\|^2$, where f denotes the network prediction of the residual between \tilde{P}_j^ϕ and P_j^ϕ . Subsequently, the final result of the network becomes $f(P_j^\phi) + P_j^\phi$.

We can denote the training dataset as

$$\{P_j^{\phi k,l}, \tilde{P}_j^{\phi k,l}\}_{j=0, k=0, l=0}^{j=M-1, k=K-1, l=L-1},$$

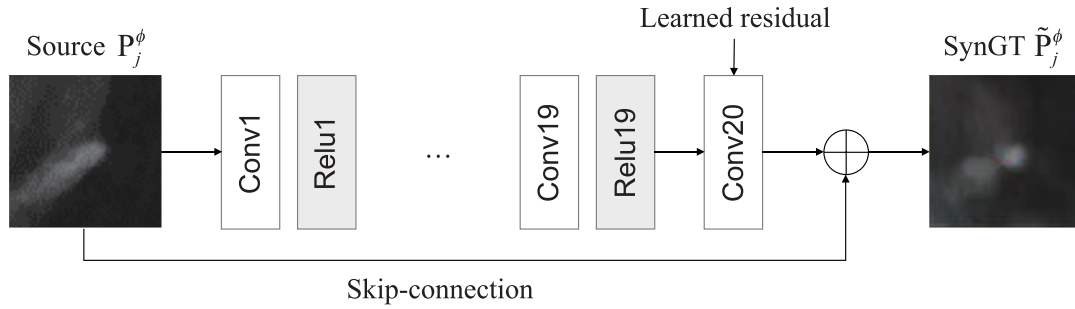


FIGURE 5. Architecture of the MAC-network based on the VDSR network [13]. A pair of convolutional layers and an activation function are cascaded repeatedly. The last convolutional layer denotes a learned residual image. A single skip-connection from the input to output is applied.

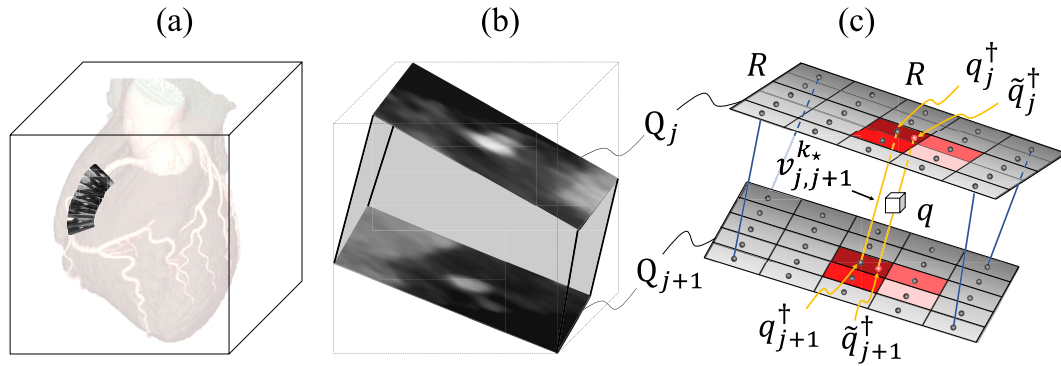


FIGURE 6. Volumetric interpolation of motion-corrected patches. (a) reinserted motion-corrected patches into a 3D volume (b) dotted line is a bounding box which covers two patches P_j^ϕ, P_{j+1}^ϕ and shaded region is the ROI of volumetric interpolation (c) bilinear interpolation is applied to all of voxels in the ROI.

where j represents the index of the patch within a single coronary artery centerline, k represents the index of phase within the 4D CT volume, excluding ϕ_* , which is assigned as the phase without motion correction, and l represents the index for the 4D CT volumes. Thus, the total number of patches in our training data should be the product $M \times K \times L$ of the M points in K phases in L 4D CT volumes.

We used the Caffe [29] framework for our implementation. The hyperparameters for the training are set as follows: a batch size of 64, a learning rate of 0.0001, an epoch of 100, and a weight decay of 0.0001, while using the optimizer “Adam” [30].

During testing, the MAC-net is applied after sampling M equidistant coronary patches based on a manually annotated artery centerline, as described in Sec. II-A. All patches are then fed independently into the trained MAC-net.

D. REINSERTION AND VOLUMETRIC INTERPOLATION OF MOTION-CORRECTED PATCHES INTO 3D CT VOLUME

The 2D patch outputs of the MAC-net are reinserted back into the CT volume, as illustrated in Fig.6, to apply motion correction to the entire 3D volume. Volumetric interpolation must be performed to propagate the motion correction to the 3D volume of the coronary artery and ensure a continuous appearance.

Because the center point \vec{q}_j and the normal \vec{n}_j of each patch are already known through the patch extraction process described in Sec. II-A, output patches P_j are first reinserted into the 3D volume by the inverse of the known transform. We can denote the planar grid of 3D coordinates corresponding to the pixel coordinates, obtained from the projection as Q_j .

Volumetric interpolation is performed for two adjacent patches P_j and P_{j+1} . Within the bounding box enclosing the two reinserted patch coordinate grids Q_j and Q_{j+1} , we define $R \times R$ vectors $v_{j,j+1}^k = q_j^k + t(q_{j+1}^k - q_j^k)$ defining 3D lines that pass through the corresponding voxel coordinates q_j^k and q_{j+1}^k , respectively, where $k, 1 \leq k \leq R^2$ denotes the index for the reinserted 3D coordinate grid and t is the parameter for the line equation. For each voxel with coordinate q within the bounding box, and in the volume between the two patches, we determine the vector $v_{j,j+1}^{k*}$ among $\{v_{j,j+1}^k, 1 \leq k \leq R^2\}$, that is closest with the voxel coordinate, i.e., that has minimum point-to-line distance $d_{p2l}(q, v_{j,j+1}^k) = \frac{|(q - q_j^k) \times (q - q_{j+1}^k)|}{|q_{j+1}^k - q_j^k|}$.

Then, we determine the two 3D coordinates q_j^\dagger and q_{j+1}^\dagger which are on the line defined by $q + t(q_{j+1}^{k*} - q_j^{k*})$ and the planes corresponding to the grids Q_j and Q_{j+1} . The final value for voxel q is the weighted average defined as

$I(q) = w_j^q P_j(\tilde{q}_j^\dagger) + w_{j+1}^q P_{j+1}(\tilde{q}_{j+1}^\dagger)$, where $P_j(q_j^\dagger)$ is the intensity obtained from bilinear interpolation on P_j at 2D non-integer coordinate \tilde{q}_j^\dagger , and $w_j^q = \frac{d_{p2p}(q, Q_{j+1})}{d_{p2p}(q, Q_j) + d_{p2p}(q, Q_{j+1})}$ is the weight defined by the point-to-plane distances d_{p2p} of q to the planes containing Q_j and Q_{j+1} . We compute d_{p2p} using the equation $d_{p2p}(q, Q_j) = \frac{|(q - \tilde{q}_j) \cdot \tilde{n}_j|}{|\tilde{n}_j|}$, where \tilde{q}_j and \tilde{n}_j are the grid center coordinate and normal vector of the plane containing the grid Q_j . The definitions are similar for \tilde{q}_{j+1}^\dagger and w_{j+1}^q .

III. RESULTS

A. PHANTOM STUDY

We evaluated our method on the C AVAREV platform [31], which is based on simulated dynamic projections based on the 4D XCAT phantom with contrasted coronary arteries derived from patient data. We used the dataset \mathcal{D}_c (cardiac motion only) for the proposed methods. Geometry calibration was obtained from a real-world clinical angiographic C-arm system.

To apply the proposed method, we reconstructed 10 C-arm CT volumes on a 256^3 grid with an isotropic voxel size of 0.5 mm with 10 different target reconstruction heart phases. Similar to the previous studies at the C AVAREV website [32], we used the reconstructed volume at heart phase 0.9, which shows quiescent motion, as the target phase for cross-phase style transfer, as described in Section II-B. Based on the description in Section II-C, the training data are comprised of patches from $M = 25$ points in $K = 10 - 1 - 2 = 7$ phases in $L = 1$ 4D CT volume. A total of 175 pairs of mid-RCA patches were extracted, which were augmented to 1,050 using vertical and horizontal flips and rotations. Because of the lack of data, we sampled the centerline points M as densely as possible and used seven volumes for training and two volumes for testing. The testing data are comprised similarly as the training data.

To evaluate our motion correction results, we used the 3D metric introduced as Q_{3D} as defined by C AVAREV [31]. We evaluated the similarity using the Dice similarity coefficients (DSCs) for the overlap of two binary images ranging from zero (no overlap) to one (perfect match). The motion corrected volume is binarized through thresholding and then evaluated with the ground-truth, which is the segmentation mask of the coronary artery within the volume reconstructed at the quiescent heart phase. The DSC of the proposed method is the mean of the DSCs of two test volumes. A comparison of the Dice score with other methods that are introduced on the C AVAREV website [32] is presented in Table 1. For use in urgent medical sites, the processing time is also an important consideration, so computation time is compared for the right coronary artery (RCA). The proposed method requires less than 1 minute of duration: 0.15 seconds for the test step of the MAC-net (Sec. II-C) and approximately 50 seconds for the reinsertion and volumetric interpolation (Sec. II-D). The computation time was obtained on Intel(R) Core(TM) i7-8700 CPUs (3.19GHz) with 16 GB of memory and

TABLE 1. Dice score comparison with the other methods using the C AVAREV dataset.

Method	Approach ^a	DSC	Computation Time
Dynamic Level Set [33]	CM	0.691	≤ 1 hr
Streak-Reduced ECG-Gated FDK [35]	REG3D	0.744	-
Residual Motion Compensation [36]	REG2D	0.776	> 1 hr (RCA)
Motion Compensation [37]	REG2D	0.823	> 1 hr (RCA)
Spatio-temporal TV [34]	CM	0.876	-
MAC-net (proposed)	DL	0.829	< 1 min (RCA) 0.15 s ^b + 49.83 s ± 1.82 ^c

^a CM: cost minimization, REG2D: 2D–2D registration, REG3D: 3D–3D registration, DL: deep learning

^b The test step of MAC-net, described in Sec. II-C, takes 0.15 seconds when considering only a right coronary artery (RCA).

^c The mean ± standard deviation computation time is 49.83 ± 1.82 seconds for the reinsertion and volumetric interpolation, described in Sec. II-D when considering only a right coronary artery (RCA), so the total computation time of the proposed method is within 1 minute.

a NVidia Titan XP GPU (12GB). The comparison methods [33], [34] are cost minimization-based approaches, and [35]–[37] are 2D–2D or 3D–3D registration-based approaches. In Schwemmer *et al.* [36], registration times were obtained on two Intel(R) Xeon(R) E5540 CPUs (2.53 GHz) with 16 GB of memory, and since [37] was their follow-up, the hardware specs would be the same or similar. When considering a trade-off between the DSC and computation time, the proposed method is comparable to other methods in terms of the DSC and outperforms in terms of computation time.

Fig. 7 illustrates examples of the test volumes with minor ($\phi = 80$) and severe ($\phi = 50$) motion artifacts. Given that the image quality is not high, motion artifacts are reduced in the results of the proposed method. The proposed method, as a deep learning-based approach, demonstrates promising results quantitatively and qualitatively using the C AVAREV dataset.

B. CLINICAL DATA

1) DATASETS

Based on the description in Section II-C, the training data comprised patches from $M = 10$ points in $K = 10 - 2 = 8$ phases in $L = 100$ 4D CT volumes. The number of phases $K = 8$ was determined by the number of temporal-phase quantizations, 10, minus the 40% and 70% phases designated as the targets for cross-phase style transfer. Several 4D CT volumes had K less than 8, in which phases with extreme motion artifacts were excluded because manual annotation of the coronary artery was impossible. The final dataset comprised a total of 5,868 pairs of mid-RCA patches, which were then augmented to 35,208 using vertical and horizontal flips and rotations. With $R = 60$, each patch sampled from the 3D volume was constructed to be of size 60×60 pixels. Validation and test sets, comprising 2,152 patches from $L = 30$ 4D CT volumes and 734 patches from $L = 10$ 4D CT volumes, were similarly constructed.

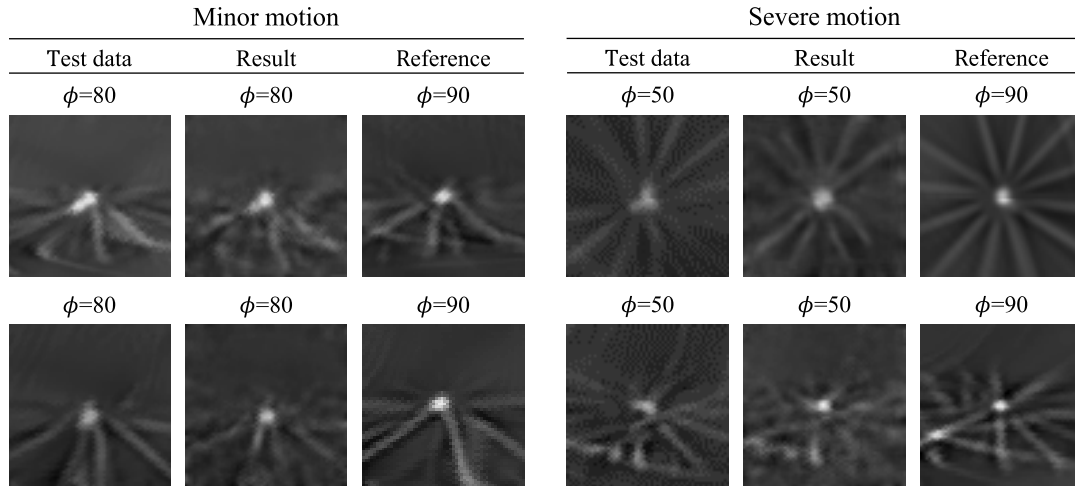
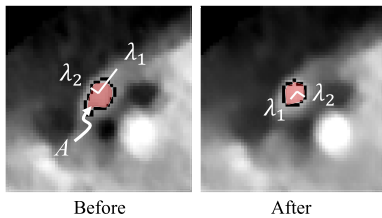


FIGURE 7. Qualitative examples with minor and severe motion artifacts among the test data in the C AVAREV dataset. Each pair consists of the test patch data containing the mid-RCA (left), results of the proposed MAC-net (middle), and the corresponding reference patch (right). Source information (phase ϕ in 4D CT) is presented above the patches. The radial artifacts are generated because of the number of projections ($n = 133$).

(a) Isotropy (higher is better)

$$L_{iso} = \frac{\lambda_1}{\lambda_2}, \quad \lambda_1 \leq \lambda_2$$

A : vessel region
 λ_1, λ_2 : principal component eigenvalues of pixel coordinates of A

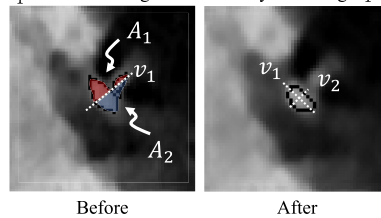


(b) Fold Overlap Ratio (higher is better)

$$L_{FOR} = \min(L_{FOR_{v_1}}, L_{FOR_{v_2}})$$

$$L_{FOR_{v_1}} = \frac{A_1^{v_1} \cap A_2}{A_1^{v_1} \cup A_2}$$

v_1, v_2 : principal component eigenvectors of pixel coordinates of A
 A_1, A_2 : area of vessel subregions divided by v_1
 $A_1^{v_1}$: area of subregion obtained by reflecting A_1 on v_1



(c) Low-Intensity Region Score (higher is better)

$$L_{LIR_{IS}} = \frac{\bar{I}_{LIR}}{\bar{I}_{my}}, \quad L_{LIR_{AS}} = 1 - \frac{A_{LIR}}{A_{ves}}$$

$$L_{LIRS} = \frac{L_{LIR_{IS}} + L_{LIR_{AS}}}{2}$$

A_{LIR}, \bar{I}_{LIR} : area, mean intensity of low intensity region
 \bar{I}_{my} : mean intensity of myocardium
 A_{ves} : area of vessel region

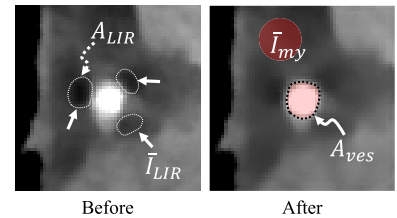


FIGURE 8. Visual description of motion artifact metrics (a) isotropy, (b) fold overlap ratio (FOR), and (c) low-intensity region score (LIRS).

2) QUANTITATIVE EVALUATION: MOTION ARTIFACT METRICS

In quantitative evaluation, because the nature of the metric makes it difficult to consider plaque, we cover only normal coronary arteries and exclude diseased coronary arteries. To quantitatively evaluate the proposed method, we first measure the isotropy of the vessel region to quantify the level of motion artifacts for the coronary artery. Because the cross-section shape of arteries is generally round, the segmented vessel area in the image patches should have an isotropic shape if it has not been corrupted by motion artifacts. We measured the isotropy using the ratio of the two eigenvalues (λ_1, λ_2) of the segment shape, as in [38].

We adopt the motion artifact metrics proposed in [24]. We present the results of three metrics having the highest agreement with three expert readers: fold overlap ratio (FOR), low-intensity region score (LIRS) and their product,

the motion artifact score (MAS). Like isotropy, the FOR also measures the shape of the vessel region, but measures the extent of mirror symmetry, rather than the overall isotropy. Given the vessel segmentation, an axis v that passes the region centroid is defined that subdivides the region in two. Then, one subregion is folded along the axis onto the other subregion, at which point the ratio of the intersection and union of those two regions are defined as L_{FOR_v} . In our implementation, we define the two orthogonal eigenvectors (v_1, v_2) of the region shape, corresponding to (λ_1, λ_2), as two different axes and determine the final FOR $L_{FOR} = \min(L_{FOR_{v_1}}, L_{FOR_{v_2}})$. The LIRS is defined as the mean of the low-intensity region intensity-score (LIR-IS) and low-intensity region area-score (LIR-AS). LIR-IS and LIR-AS respectively measure the intensity and the size of the low-intensity shading of the motion artifact that occurs on the myocardium. LIR-IS is defined as the ratio of the intensity values of the shaded

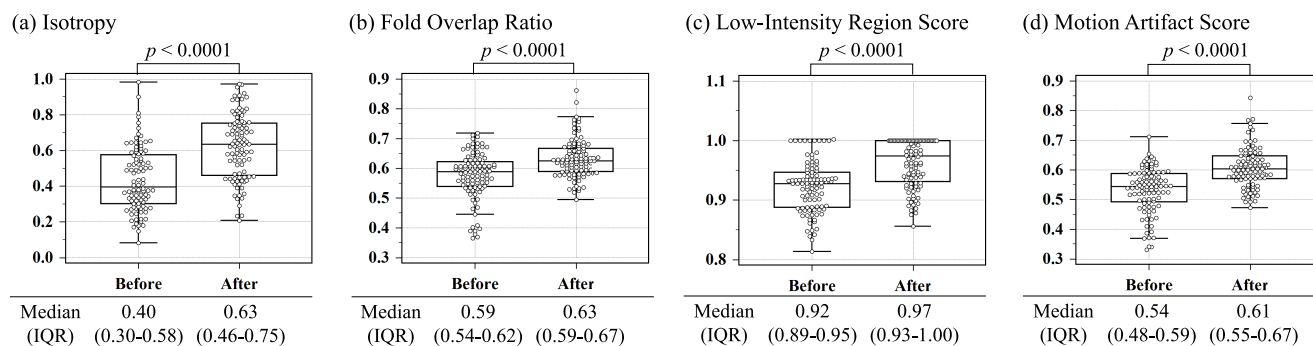


FIGURE 9. Results of motion artifact metrics (a) isotropy, (b) fold overlap ratio (FOR), (c) low-intensity region score (LIRS), and (d) motion artifact score (MAS). The results of randomly sampled 100 patches of test cases before and after applying the proposed motion correction method is depicted and their median and interquartile range (IQR). P-values were computed by a Wilcoxon ranksum test, and all indicate statistical significance.

area and the myocardium and LIR-AS are defined as the ratio of the area of the shading and the size of the artery. A visual description and the specific mathematical notations and equations are presented in Fig. 8.

To compute these metrics, the specific vessel region, regions where the low-intensity shading artifact occurs, and the myocardium region must be determined. To perform this process efficiently, we apply a simple seed-based region growing method to annotate the vessel and low-intensity artifact regions semi-automatically. Not only is the region growing more efficient, but it also aids in delineating the segment boundary and maintaining the consistency of the region intensity. Regions of the myocardium were manually annotated, in contrast, because it may have various appearances and shapes, it is thus unsuitable for region growing.

For comparative evaluation, the results of the four motion artifact metrics for the input and output of the MAC-net for the randomly sampled 100 patches of the test set are presented in Fig. 9. The isotropy, FOR, LIRS, and MAS, all of which increased, represent improved results. The median and interquartile range (IQR) of each metric are presented in Fig. 9 and increased 57.5% from 0.40 (0.30 – 0.58) to 0.63 (0.46 – 0.75), 6.8% from 0.59 (0.54 – 0.62) to 0.63 (0.59 – 0.67), 5.4% from 0.92 (0.89 – 0.95) to 0.97 (0.93 – 1.00), and 13.0% from 0.54 (0.48 – 0.59) to 0.61 (0.55 – 0.67), respectively. To test statistical significance, p-values were computed by a Wilcoxon ranksum test. The statistical significance of improvements is supported with p-values all smaller than 0.0001.

3) QUANTITATIVE EVALUATION: BACKGROUND PSNR AND SSIM

In the previous subsection, a quantitative evaluation for the coronary artery region is presented, and in this subsection, a quantitative evaluation for the background region is presented. To evaluate the degradation of the background of the 2D patches of the test set, we analyze two well-known image quality metrics: the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index, and the results

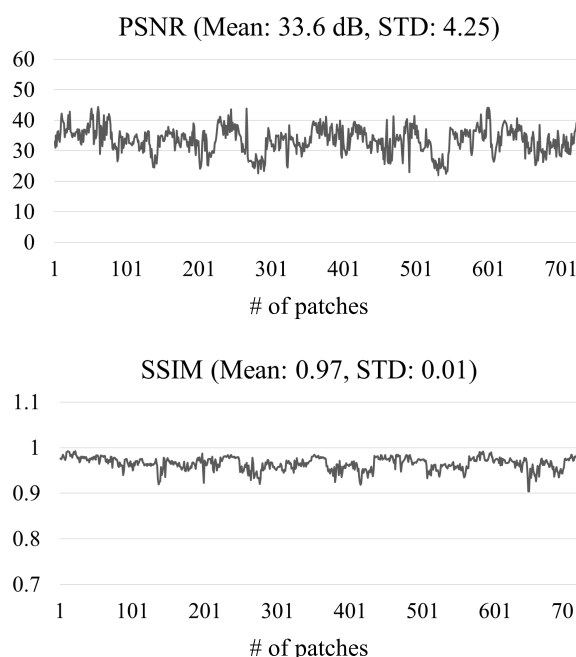


FIGURE 10. Background PSNR (top) and SSIM (bottom) of the patches in test cases. The Y-axis represents PSNR and SSIM, respectively, and the X-axis represents the number of test patches.

are presented in Fig. 10. The mean of PSNR is 33.6 dB (standard deviation of 4.25), and the mean of SSIM index is 0.97 (standard deviation of 0.01). Both metrics illustrate that there was minimal to no image quality degradation.

4) QUALITATIVE EVALUATION

Two experienced readers evaluated the degree of motion artifacts based on a 5-point Likert scale [39], where 1 = completely unreadable, 2 = significant motion, 3 = apparent motion, 4 = minor motion, and 5 = no motion. We performed a blind test on this inter-observer study: the observer did not know whether the patches were from before or after applying

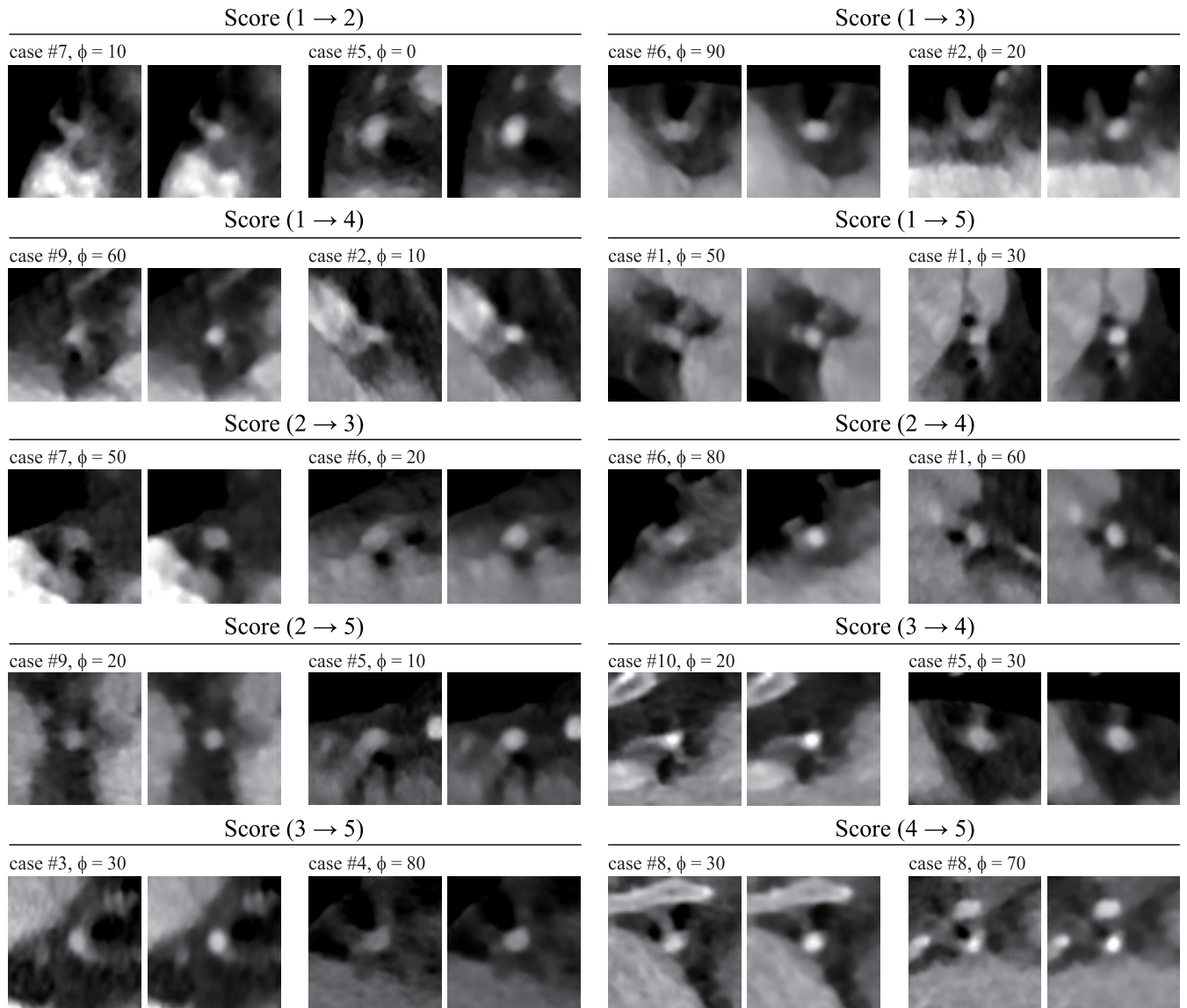


FIGURE 11. Qualitative examples of the original patch containing the mid-RCA (left) and results of the proposed MAC-net (right). Samples represent variational range of improvements. Expert evaluated scores based on 5-point Likert scale [39] (1 = completely unreadable, 2 = significant motion, 3 = apparent motion, 4 = minor motion, 5 = no motion) and source information (case number and phase ϕ in 4D CT) are presented above. For all patches in the test set, motion artifact metrics are measured quantitatively in Sec. III-B2.

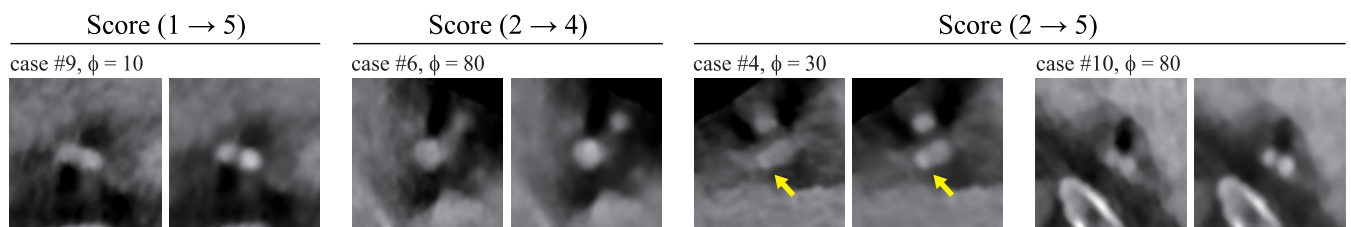


FIGURE 12. Qualitative examples of the input mid-RCA patch (left) and MAC-net results (right). Samples represent specific cases that distinguish the primary vessel from the branch. The third sample pair also illustrates correction for the artery plaque. Expert-evaluated scores and source information are presented above.

the proposed method. Both original and motion-corrected patches are depicted completely randomly to the observers without any delay.

Table 2 presents the ratio of frequencies of each category for test patches before and after motion correction using the MAC-net. The proportion of images presented with

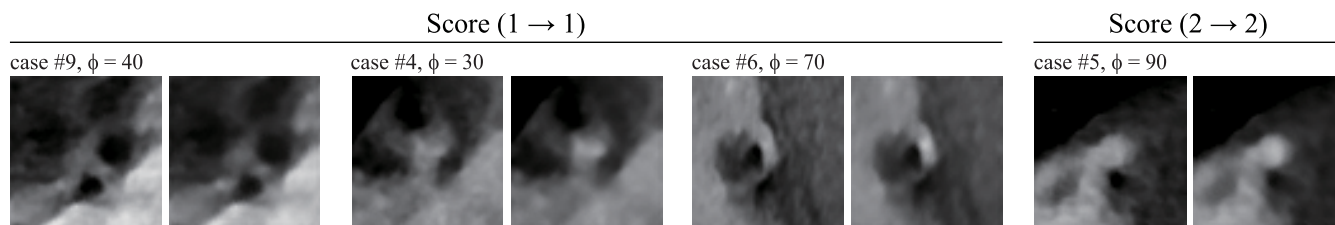


FIGURE 13. Qualitative examples of the input mid-RCA patch (left) and MAC-net results (right). Samples represent worst cases where there is no improvement. There were no cases where the score decreased. Expert-evaluated scores and source information are presented above.

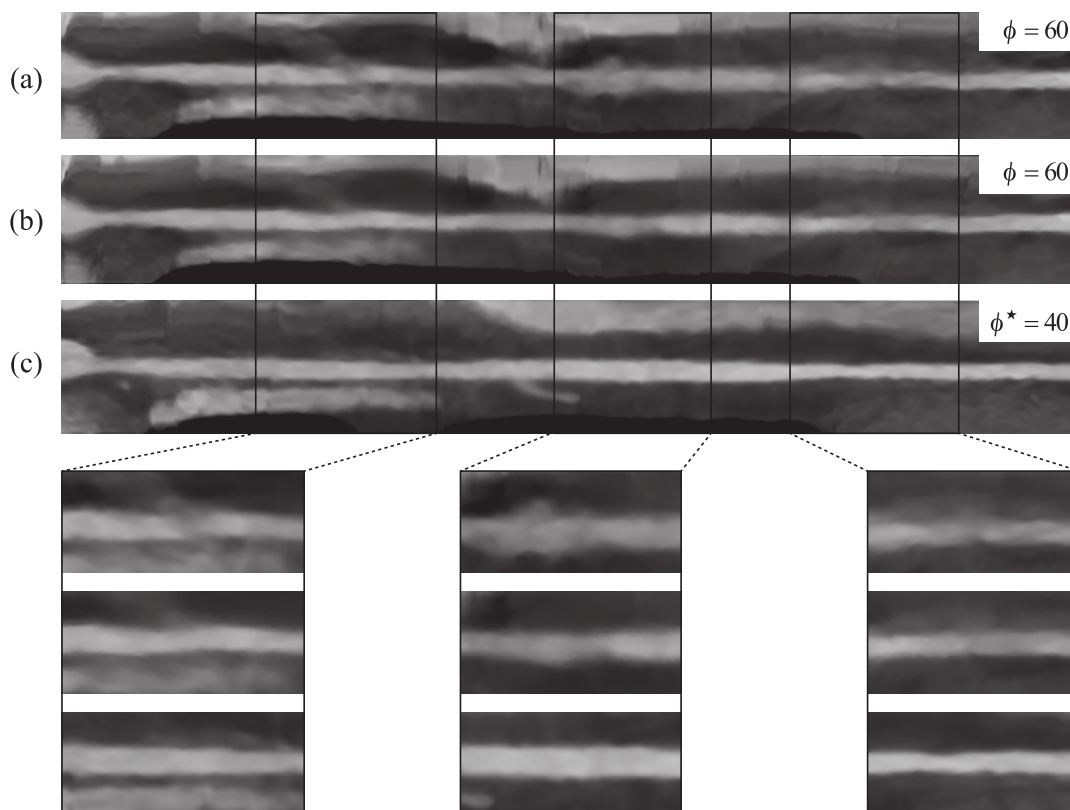


FIGURE 14. Straightened curved planar reformation (CPR) example of a right coronary artery (RCA) of case #7. (a) straightened CPR at phase ($\phi = 60$) with motion artifacts (b) straightened CPR of MAC-net result at phase ($\phi = 60$) (c) straightened CPR at the best-phase ($\phi^* = 40$). Below represents the magnified view of three parts blurred by motion artifacts.

completely unreadable, significant, and apparent motions (Likert scale 1, 2, and 3) were 98.5% previously but decreased to 35% for the results of the MAC-net ($p < 0.001$). The mean \pm standard deviation based on the Likert scale was significantly improved from 1.43 ± 0.66 to 3.80 ± 0.87 . ($p < 0.001$). The inter-observer agreement was calculated with the kappa (κ) statistics for the motion score and exhibited strong agreement as before $\kappa = 0.85$; 95% confidence interval (CI) 0.76–0.95 and after $\kappa = 0.70$; 95% CI 0.61–0.81.

The sample results of the motion-corrected patches are presented in Figs. 11, 12, and 13. Each pair of images represents those before and after applying the MAC-net. Changes in the Likert scores and the source information

of each patch are presented above the patches. Fig. 11 provides a comparison of various levels of improvement. Improvements range from extremely positive (from completely unreadable [score 1] improved to no motion [score 5]) to moderate (significant motion [score 2] improved to minor motion [score 4]) to incremental (apparent motion improved to minor motion). In all cases, after applying the MAC-net, the edge of the coronary artery more visible and motion artifacts reduced, while distortion of local appearance is limited. Fig. 12 demonstrates the robustness of the MAC-net for special cases where the coronary artery diverges or contains plaques. The MAC-net performs only a gentle modification on the artery appearance to improve visibility.

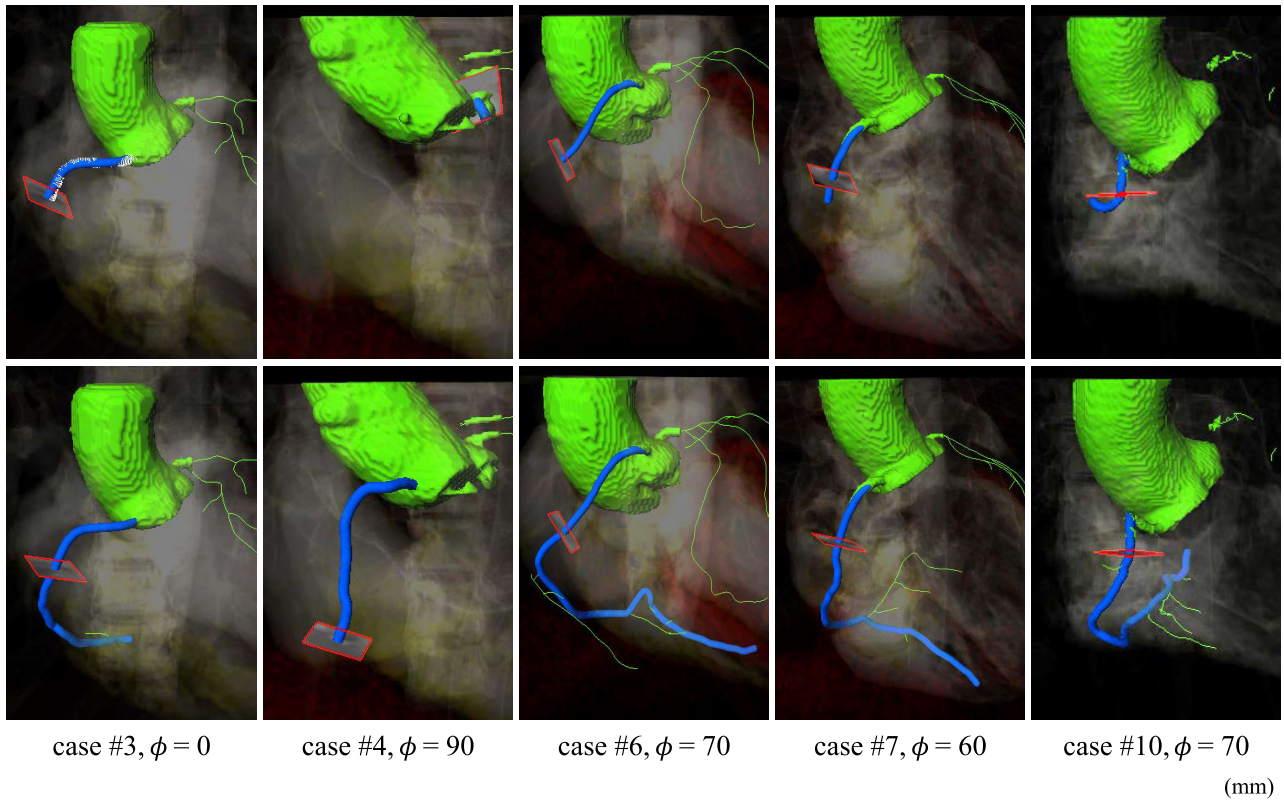


FIGURE 15. Results of RCA tracking using an automatic tracking method of commercial software (QAngioCT, Medis Medical Imaging Systems, Leiden, the Netherlands) on the original (top row) and the motion-corrected (second row) CT volumes. The table (bottom row) presents the total length(mm) of the tracked RCA in each case. ϕ^* (Ref) in the table represents the best-phase of each case. Source information is presented below the second row.

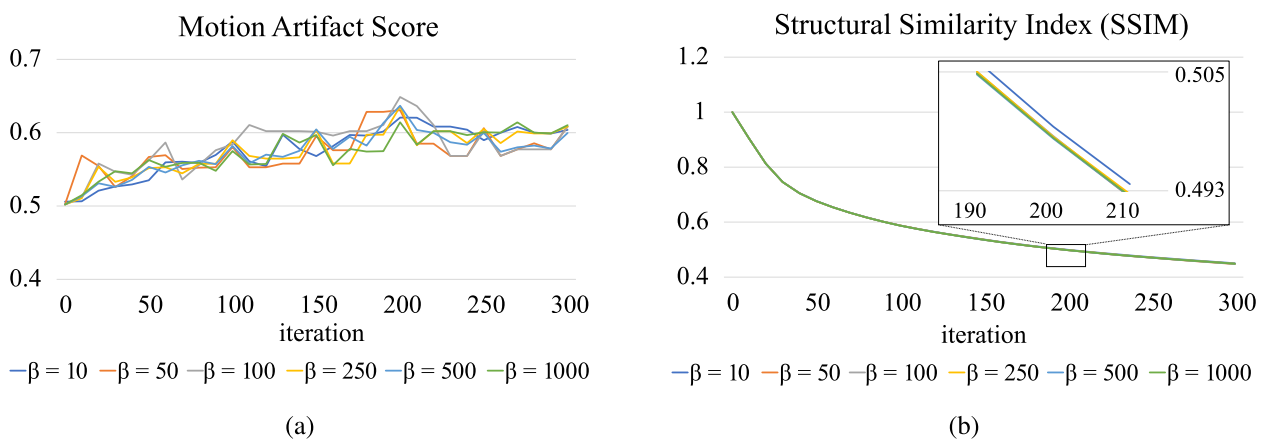


FIGURE 16. Ablation study for hyper-parameters for style transfer value of β in 4 (with α fixed to 1) and number of iterations on (a) motion artifact score circularity and (b) structural similarity. Results were measured on 10 random patches from 5 different cases.

Samples of worst cases are depicted in Fig. 13, in which there was no change in the score. These cases may occur when motion artifacts are too severe or the intensity of the coronary

artery is indistinguishable from the right atrium or the right ventricle. There were no cases where the score decreased. Overall, the MAC-net is highly likely to improve the image

TABLE 2. Motion artifact before and after applying proposed method.

	Before	After	p-value
Likert scale			<0.001
1 = Completely unreadable	66.5 %	2.5 %	
2 = Significant motion	26 %	2.5 %	
3 = Apparent motion	6 %	30 %	
4 = Minor motion	1.5 %	41.5 %	
5 = No motion	0 %	23.5 %	
Score			<0.001
Mean	1.43	3.80	
Standard deviation	± 0.66	± 0.87	
Interobserver agreement			
Kappa value	0.85	0.70	
Standard error	± 0.05	± 0.05	
95 % CI ^a	0.76-0.95	0.61-0.81	

^a CI = confidence interval

quality with a very small probability of causing harm. We also present straightened curved planar reformations (CPRs) of case #7 in Fig. 14. These include straightened CPRs of a right coronary artery (RCA) at phase ($\phi = 60$) with motion artifacts and the MAC-net results at the same phase and at the best-phase ($\phi = 40$). The boundary of the RCA is blurred by motion artifacts in Fig. 14(a), whereas in Fig 14(b), the border is clearer than before.

Furthermore, we present qualitative results of the volumetric motion correction, as described in Sec II-D, in Fig. 15. We collected five example cases in which the automatic tracking method [40] provided by the commercial software QAngioCT fails to track the RCA due to motion artifacts among ten test cases. After applying the proposed method, the RCAs were tracked 62% longer on average. Again, there were no cases where tracking was reduced, supporting our claim that even for the worst case, the proposed method is unlikely to cause harm.

C. ABLATION: HYPER-PARAMETERS FOR STYLE TRANSFER

For most of the hyper-parameters of components applied off-the-shelf in the proposed method, such as the pre-trained network used with style transfer or parameters for training the MAC-net, little or no tuning was required to obtain acceptable results. Thus, we present ablations regarding the hyper-parameters for iterative optimization within the style transfer process, including the β value and the number of iterations gradient descent for the style loss before termination, in Fig. 16.

We measure the amount of motion correction using the MAS and the amount of change in local appearance using the SSIM [41]. As described in Sec. III-B2, MAS is measured on the regions containing the vessel and motion artifacts, annotated manually, whereas SSIM is measured on the remaining background regions. Although the results are unexpectedly stable for different values of β , the MAS increases and SSIM

decreases during the iterations as expected. In all experiments, we used $\beta = 100$ and terminated the style transfer after 100 iterations to balance the motion correction and appearance change.

IV. CONCLUSION

We have proposed a novel framework for motion correction of CCTA. With practicality and efficiency in mind, the proposed framework has the following characteristics: 1) actual learning is performed on 2D patches of the coronary artery, 2) corresponding patches are extracted from 3D volumes from different phases with strong and weak motion in a 4D volume, 3) style transfer [21] is used to generate synthetic motion-corrected ground-truth images, 4) a separate deep CNN is used to learn the motion correction from the synthesized ground-truth data, 5) during testing, patches are extracted from manually annotated coronary artery in an input 3D volume, motion-corrected, and reinserted and interpolated back into the volume.

We have performed a quantitative and qualitative evaluation that confirms the effectiveness of the proposed method using phantom and clinical dataset. In the phantom study, the proposed method shows acceptable results when considering the DSC and computation time. In the experiments using clinical dataset, in terms of the MAS—the product of fold-overlap ratio and low-intensity region score, as proposed in [24]—the proposed method resulted in a 13.0% improvement in the median compared to the uncorrected input. Qualitatively, for the motion-corrected 2D patches, the mean \pm standard deviation values of the 5-point Likert scale graded by expert readers was significantly improved from 1.43 ± 0.66 (between completely unreadable and significant motion) to 3.80 ± 0.87 (between apparent motion and minor motion). For the motion-corrected 3D volume after reinsertion and interpolation, commercial software QAngioCT tracked the RCA 62% longer on average. Furthermore, we provide an ablation study concerning the setting of the style transfer method [21]. We found that the results were consistent with significantly varying values of β —the ratio of scale between style and content loss—and that 100 iterations were sufficient to obtain acceptable results.

For our future work, we plan to explore alternative methods to generate ground-truth motion correction images. The recent success of using generative adversarial networks (GAN) in synthesizing realistic training data [19] upholds its potential to address the motion correction problem.

REFERENCES

- [1] M. P. Heron, "Deaths: Leading causes for 2012," *Nat. Vital Statist. Rep.*, vol. 64, no. 10, pp. 1–93, 2015.
- [2] J. M. Miller, C. E. Rochitte, M. Dewey, A. Arbab-Zadeh, H. Niinuma, I. Gottlieb, N. Paul, M. E. Clouse, E. P. Shapiro, and J. Hoe, "Diagnostic performance of coronary angiography by 64-row CT," *New England J. Med.*, vol. 359, no. 22, pp. 2324–2336, 2008.

- [3] M. D. Shapiro, A. J. Pena, J. H. Nichols, S. Worrell, F. Bamberg, N. Dannemann, S. Abbara, R. C. Cury, T. J. Brady, and U. Hoffmann, "Efficacy of pre-scan beta-blockade and impact of heart rate on image quality in patients undergoing coronary multidetector computed tomography angiography," *Eur. J. Radiol.*, vol. 66, no. 1, pp. 37–41, Apr. 2008.
- [4] P. G. Steg, R. Ferrari, I. Ford, N. Greenlaw, J.-C. Tardif, M. Tenders, H. Abergel, and K. M. Fox, "Heart rate and use of beta-blockers in stable outpatients with coronary artery disease," *PLoS ONE*, vol. 7, no. 5, 2012, Art. no. e36284.
- [5] A. A. Isola, M. Grass, and W. J. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," *Med. Phys.*, vol. 37, no. 3, pp. 1093–1109, Feb. 2010.
- [6] Q. Tang, J. Cammin, S. Srivastava, and K. Taguchi, "A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT," *Med. Phys.*, vol. 39, no. 7, pp. 4291–4305, Jun. 2012.
- [7] R. Bhagalia, J. D. Pack, J. V. Miller, and M. Iatrou, "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Med. Phys.*, vol. 39, no. 7, pp. 4245–4254, Jun. 2012.
- [8] C. Rohkohl, H. Bruder, K. Stierstorfer, and T. Flohr, "Improving best-phase image quality in cardiac CT by motion correction with MAM optimization," *Med. Phys.*, vol. 40, no. 3, Feb. 2013, Art. no. 031901.
- [9] S. Kim, Y. Chang, and J. B. Ra, "Cardiac motion correction based on partial angle reconstructed images in X-ray CT," *Med. Phys.*, vol. 42, no. 5, pp. 2560–2571, Apr. 2015.
- [10] J. Hahn, H. Bruder, C. Rohkohl, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data," *Med. Phys.*, vol. 44, no. 11, pp. 5795–5813, Nov. 2017.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [12] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [13] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [14] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 241–246.
- [15] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 1790–1798.
- [16] C. Gsaxner, B. Pfarrkirchner, L. Lindner, A. Pepe, P. M. Roth, J. Egger, and J. Wallner, "PET-train: Automatic ground truth generation from PET acquisitions for urinary bladder segmentation in CT images using deep learning," in *Proc. 11th Biomed. Eng. Int. Conf. (BMEiCON)*, Nov. 2018, pp. 1–5.
- [17] W. Chen, H. Wang, Y. Li, H. Su, Z. Wang, C. Tu, D. Lischinski, D. Cohen-Or, and B. Chen, "Synthesizing training images for boosting human 3D pose estimation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 479–488.
- [18] L. Sixt, B. Wild, and T. Landgraf, "RenderGAN: Generating realistic labeled data," 2016, *arXiv:1611.01331*. [Online]. Available: <http://arxiv.org/abs/1611.01331>
- [19] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 3, no. 4, Jul. 2017, p. 6.
- [20] G. Georgakis, A. Mousavian, A. C. Berg, and J. Kosecka, "Synthesizing training data for object detection in indoor scenes," 2017, *arXiv:1702.07836*. [Online]. Available: <http://arxiv.org/abs/1702.07836>
- [21] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.
- [22] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Cham, Switzerland: Springer, 2016, pp. 694–711.
- [23] S. Jung, S. Lee, B. Jeon, Y. Jang, and H.-J. Chang, "Deep learning based coronary artery motion artifact compensation using style-transfer synthesis in CT images," in *Proc. Int. Workshop Simulation Synth. Med. Imag.*, Cham, Switzerland: Springer, 2018, pp. 100–110.
- [24] H. Ma, E. Gros, A. Szabo, S. G. Baginski, Z. R. Laste, N. M. Kulkarni, D. Okerlund, and T. G. Schmidt, "Evaluation of motion artifact metrics for coronary CT angiography," *Med. Phys.*, vol. 45, no. 2, pp. 687–702, Feb. 2018.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [28] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, pp. 1310–1318.
- [29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2014, pp. 675–678.
- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [31] C. Rohkohl, G. Lauritsch, A. Keil, and J. Hornegger, "CAVAREV—An open platform for evaluating 3d and 4d cardiac vasculature reconstruction," *Phys. Med. Biol.*, vol. 55, no. 10, p. 2905, 2010.
- [32] CAVAREV. Accessed: Mar. 10, 2020. [Online]. Available: <https://www5.cs.fau.de/research/software/cavarev/>
- [33] A. Keil, J. Vogel, G. Lauritsch, and N. Navab, "Dynamic cone-beam reconstruction using a variational level set formulation," in *Proc. Int. Meeting Fully Three-Dim. Image Reconstr. Radiol. Nucl. Med. (Fully3D)*, 2009, pp. 319–322.
- [34] O. Taubmann, M. Unberath, G. Lauritsch, S. Achenbach, and A. Maier, "Spatio-temporally regularized 4-D cardiovascular C-arm CT reconstruction using a proximal algorithm," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 52–55.
- [35] C. Rohkohl, G. Lauritsch, A. Nottling, M. Prummer, and J. Hornegger, "C-arm CT: Reconstruction of dynamic high contrast objects applied to the coronary sinus," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct. 2008, pp. 5113–5120.
- [36] C. Schwemmer, C. Rohkohl, G. Lauritsch, K. Müller, and J. Hornegger, "Residual motion compensation in ECG-gated interventional cardiac vasculature reconstruction," *Phys. Med. Biol.*, vol. 58, no. 11, pp. 3717–3737, Jun. 2013.
- [37] C. Schwemmer, C. Rohkohl, G. Lauritsch, K. Müller, J. Hornegger, and J. Qi, "Opening windows-increasing window size in motion-compensated ecg-gated cardiac vasculature reconstruction," in *Proc. Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, 2013, pp. 50–53.
- [38] B. Jeon, Y. Hong, D. Han, Y. Jang, S. Jung, Y. Hong, S. Ha, H. Shim, and H.-J. Chang, "Maximum a posteriori estimation method for aorta localization and coronary seed identification," *Pattern Recognit.*, vol. 68, pp. 222–232, Aug. 2017.
- [39] I. Cho, K. Elmore, B. ó Hartaigh, J. Schulman-Marcus, H. Granser, V. Valenti, G. Xiong, P. M. Carrascosa, and J. K. Min, "Heart-rate dependent improvement in image quality and diagnostic accuracy of coronary computed tomographic angiography by novel intracycle motion correction algorithm," *Clin. Imag.*, vol. 39, no. 3, pp. 421–426, May 2015.
- [40] G. Yang, P. Kitslaar, M. Frenay, A. Broersen, M. J. Boogers, J. J. Bax, J. H. C. Reiber, and J. Dijkstra, "Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography," *Int. J. Cardiovascular Imag.*, vol. 28, no. 4, pp. 921–933, Apr. 2012.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.



SUNGHEE JUNG (Member, IEEE) received the B.S. degree in computer science engineering from the Hankuk University of Foreign Studies, South Korea, in 2013. She is currently pursuing the Ph.D. degree in medical science with Yonsei University, Seoul, South Korea. Since 2012, she has been a Researcher with the Medical Image Computing and Analysis Laboratory, CONNECT-AI Research Center. Her research interests include the application of computer vision techniques from image processing and deep learning to noise reduction and motion correction in medical images.



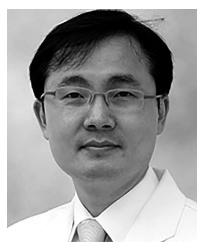
SOOCHAHN LEE (Member, IEEE) received the B.S. and Ph.D. degrees from Seoul National University, Seoul, South Korea, in 2004 and 2011, respectively. From 2011 to 2014, he was with the Digital Media and Communications Research and Development Center, Samsung Electronics. From 2014 to 2019, he was an Assistant Professor with Soonchunhyang University, Asan, South Korea. Since 2019, he has been an Assistant Professor with the School of Electrical Engineering, Kookmin University, Seoul. His research interests include computer vision and medical image analysis.



BYUNGHWAN JEON received the B.S. degree in computer science engineering from the Hankuk University of Foreign Studies, South Korea, in 2013, and the Ph.D. degree in medical science from Yonsei University, Seoul, South Korea, in 2019. From 2012 to 2019, he was a Researcher, and he has been a Postdoctoral Researcher with the Medical Image Computing and Analysis Laboratory, CONNECT-AI Research Center, Seoul, since 2019. Since 2020, he has been an Assistant Professor with the Department of Computer Science, Kyungil University, Gyeongsan, South Korea. His research interests include the application of computer vision techniques, from Bayesian estimation, geometric analysis, and deep neural networks to localization, tracking, and segmentation tasks in medical images.



YEONGGUL JANG received the B.S. degree in computer science engineering from the Hankuk University of Foreign Studies, South Korea, in 2014. He is currently pursuing the Ph.D. degree in medical science with Yonsei University, Seoul, South Korea. Since 2012, he has been a Researcher with the Medical Image Computing and Analysis Laboratory, CONNECT-AI Research Center. His research interests include the application of computer vision techniques, from geometric analysis, image processing, and deep learning to segmentation and localization in medical images.



HYUK-JAE CHANG received the M.D. and M.M. degrees from the College of Medicine, Yonsei University, Seoul, South Korea, in 1994 and 1999, respectively, and the Ph.D. degree in medicine from the Graduate School of Medicine, Ajou University, Suwon, South Korea, in 2003. From 2002 to 2004, he was an Assistant Professor with the Department of Internal Medicine, College of Medicine, Ajou University. From 2004 to 2009, he was an Assistant Professor with the Department of Internal Medicine, Seoul National University College of Medicine, Seoul. From 2007 to 2008, he was a Postdoctoral Fellow with the Department of Internal Medicine, Johns Hopkins University, Baltimore, MD, USA. Since 2009, he has been a Professor with the Department of Internal Medicine, Yonsei University College of Medicine. Since 2016, he has been the CIO with the Division of Medical Information and Technology, Severance Hospital. Since 2017, he has been the CEO with PHI Digital Healthcare, Inc., Seoul. His research interests include cardiovascular disease, pulmonary hypertension, and medical image processing.

...