Deep Learning Based Coronary Artery Motion Artifact Compensation Using Style-Transfer Synthesis in CT Images

Sunghee Jung[1], Soochahn Lee[2](✉), Byunghwan Jeon[1], Yeonggul Jang[1], and Hyuk-Jae Chang[3]

[1] Brain Korea 21 PLUS Project for Medical Science, Yonsei University,
Seoul 03722, South Korea
sh.jung@yonsei.ac.kr

[2] Department of Electronic Engineering, Soonchunhyang University,
Asan 31538, South Korea
sclsch@sch.ac.kr

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

Abstract. Motion artifact compensation of the coronary artery in computed tomography (CT) is required to quantify the risk of coronary artery disease more accurately. We present a novel method based on deep learning for motion artifact compensation in coronary CT angiography (CCTA). The ground-truth, i.e., coronary artery without motion, was synthesized using full-phase four-dimensional (4D) CT by applying style-transfer method because it is medically impossible to obtain in practice. The network for motion artifact compensation based on very deep convolutional neural network (CNN) is trained using the synthesized ground-truth. An observer study was performed for the evaluation of the proposed method. The motion artifacts were markedly reduced and boundaries of the coronary artery were much sharper than before applying the proposed method, with a strong inter-observer agreement (kappa = 0.78).

Keywords: Motion artifact compensation · Coronary artery
Deep learning · Computed tomography · Style-transfer

1 Introduction

Coronary artery disease (CAD), also known as ischemic heart disease, is the leading cause of death globally [7]. Recently, non-invasive coronary computed tomography angiography (CCTA) has been widely adopted. If CCTA is acquired when the heart is beating, motion artifacts can inevitably be caused. Therefore, motion artifact compensation is required to quantify the severity of CAD more accurately.

A. Gooya et al. (Eds.): SASHIMI 2018, LNCS 11037, pp. 100–110, 2018.
https://doi.org/10.1007/978-3-030-00536-8_11

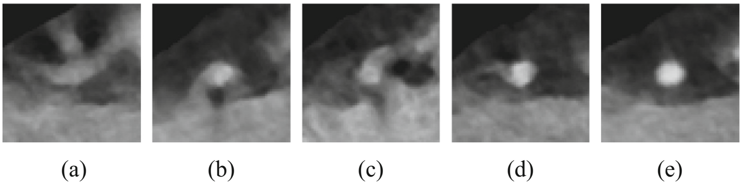

Fig. 1. Appearance of the motion artifacts of a coronary artery in different phases of same patient's 4D CT according to a 5-point Likert scale, described in Sect. 3.2 (a) completely unreadable, (b) significant motion artifacts, (c) apparent motion artifacts, (d) minor motion artifacts, (e) no motion artifacts

To solve this problem, prospective ECG-gating or drugs (e.g., beta-blockers) can be used. The former enables data acquisition when the heart is moving as quiescently as possible, and the latter enables the patient heart rate to be reduced. Nonetheless, motion artifacts can occur if the heart rate is irregular or due to the temporal resolution of CT.

Various approaches based on image processing have been proposed to solve this issue. Several methods were proposed that first perform coronary artery motion estimation, after which the cardiac CT images are obtained using motion compensated reconstruction [1,9,14,16]. However, the motion artifacts are likely to degrade the performance of motion estimation, ultimately leading to the degradation of motion compensation as well.

The advancement of deep learning has caused revolutionary improvements across many different disciplines [3,8,13]. It is reasonable to assume that a physician could estimate the image without the motion artifacts more accurately with more experience. Based on this assumption, together with the success of deep learning methods, we hypothesized that deep learning can be used for motion artifact compensation as well.

In this work, we approach the issue of reducing motion artifacts in CCTA by using deep learning, similar to denoising [17] or super-resolution methods [11] that have recently been shown successful. To apply deep learning to coronary artery motion compensation, ground-truth (GT) data are required. However, the image of a coronary artery without motion in a patient cannot be determined. That is, it is medically impossible to obtain exactly the corresponding coronary CT images with and without motion artifacts.

Our core idea is to use a style transfer method on the image patches from four-dimensional (4D) CT images, containing phases with large and small amounts of motion artifacts (Fig. 1), to generate a synthetic ground truth, which we term SynGT. We apply style transfer, instead of just using only the patches directly, owing to the local deformations that occur from the heartbeats. 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.

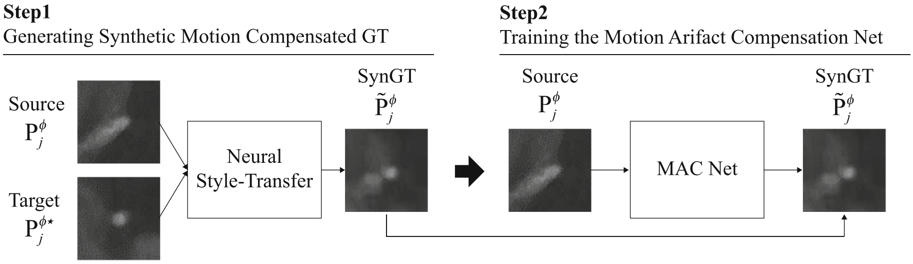

Fig. 2. Workflow of the proposed method. In step 1, generate synthetic motion compensated patch (SynGT) using style-transfer method. In step 2, training the motion artifact compensation network (MAC Net) using SynGT. The detailed descriptions of step 1 and step 2 are found in Sects. 2.2 and 2.3, respectively.

The primary contributions of our work are summarized as follows, (i) we applied the style transfer method in order to synthesize the motion compensated ground truth (SynGT) (Step1 in Fig. 2), (ii) trained the motion artifact compensation network, which we termed MAC Net, by utilizing the SynGT (Step2 in Fig. 2), and (iii) performed an observer study that scores the degree of motion artifacts before and after applying the proposed method.

Section 2 examines the proposed method and the details are presented in each subsection. Section 3 shows the dataset and experimental results of the proposed method. In Sect. 4, we present the conclusions and discussions.

2 Methods

2.1 Extraction of Corresponding Coronary Patches from 4D CT

We used 4D CT images which are acquired by retrospective gating using a dual source CT scanner (SOMATOM Definition Flash, Siemens). All raw data were reconstructed 0%–90% in 10% increments of the R-R interval.

Herein, we specifically focus only on the middle of the right coronary artery (mid-RCA), which generally has the most motion, as the region of interest when we trained the motion artifact compensation (MAC) network. Given the temporally sampled three-dimensional (3D) CT volumes, the mid-RCA was manually annotated by the experts in each volume using a commercial coronary analysis software (QAngioCT, Medis Medical Imaging Systems, Leiden, Netherlands). Here, the 1st right ventricle branch and acute marginal branch are defined as the start point and the end point, respectively.

The mid-RCA centerline $\mathcal{C}^{\phi}$ of a 3D volume at phase ϕ is represented as a discretized set of ordered 3D coordinates $\mathcal{C}^{\phi} = \left\{ c_i^{\phi} | 0 \leq i \leq N_c^{\phi} - 1 \right\}$, where c_i^{ϕ} denotes the i_{th} 3D point coordinate, among a total number of N_c^{ϕ}, of $\mathcal{C}^{\phi}$. The exact centerline is approximated as a piecewise linear function between the points in $\mathcal{C}^{\phi}$. Thus, the entire length of the mid-RCA centerline is defined

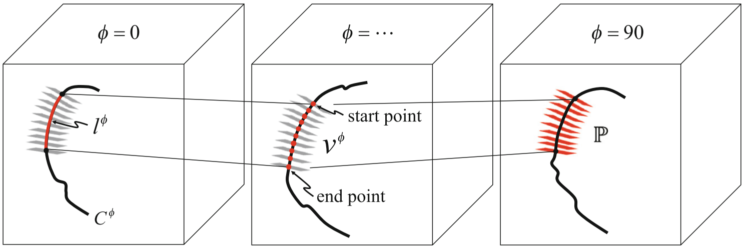

Fig. 3. Determining positions and normals for corresponding patches of mid-RCA in 3D CT volumes at different temporal phases, included within the full-phase 4D CT volumes. The centerlines of the mid-RCA, including the start and end points are manually annotated in full-phase volumes.

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 patches on $\mathcal{C}^\phi$, 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 landmark. A fixed number of M equidistant points $\mathcal{V}^\phi = \left\{q_j^\phi | 0 \leq j \leq M-1\right\}$ each spaced $\frac{l^\phi}{M}$ are sampled between the start and end points of $\mathcal{C}^\phi$. Because the mid-RCA centerline is approximated as a piecewise linear function, we applied interpolation to compute the exact equidistant point coordinate. Finally, we define the normal directions $\boldsymbol{n_j^\phi}$ for the planar patches centered at each q_j^ϕ as the tangential direction of $\mathcal{C}^\phi$ at q_j^ϕ. Figure 3 visualizes this process of determining the corresponding points along 3D CT volumes at different temporal phases.

The corresponding patches $\mathbb{P} = \left\{\mathbf{P}_j^\phi | 0 \leq j \leq M-1\right\}$ are extracted by sampling the voxel intensities on an $R \times R$ discrete grid centered at q_i^ϕ with normal $\boldsymbol{n_j^\phi}$ within the corresponding 3D CT volume. To align the spatial distribution of the grid points physically, we constructed a two-dimensional grid (on the xy-plane as 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, bicubic interpolation is applied when assigning intensity values to each pixel in the extracted patch.

2.2 Generating Synthetic Motion Compensated Patches Using Cross-Phase Style-Transfer

The motion of the heartbeat causes differences in its local appearance. However, we would like to obtain the corresponding patch with the identical local appearance but without motion artifacts because we would like to train a convolutional neural network (CNN) to remove only the artifacts. As this is clinically unattainable, we aim to synthesize this same-phase-no-artifact patch, $\tilde{\mathbf{P}}_j^\phi$, using style transfer to source patch $\mathbf{P}_j^{\phi}$ with a different-phase-no-artifact patch as the target $\mathbf{P}_j^{\phi\star}$, a process which we term *cross-phase style-transfer*. Here, $\phi\star$ denotes the phase within the heartbeat when the motion is the slowest, resulting in the least amount of motion artifacts.

In the proposed framework, we applied a recent method for style transfer using deep neural networks [5], often called the *neural style transfer* method, in our framework. The core of the method comprises the following components. First, a CNN, particularly the VGG network [15] pretrained on the ImageNet database [4], is used to compute local image features that are subsequently defined as the numerical representation of the content. If we denote the tensor of the CNN features at layer l as $\mathbf{F}_x^1$ and $\mathbf{F}_c^1$ for the synthesized image $\boldsymbol{I_x}$ and content reference image $\boldsymbol{I_c}$, respectively, the loss function for the content is defined as

$$\mathcal{L}_{content}\left(\boldsymbol{I_x}, \boldsymbol{I_c}\right) = \frac{1}{2}||\boldsymbol{I_x} - \boldsymbol{I_c}||_2^2. \tag{1}$$

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

$$G_{ij}^l = \mathbf{F}_i^l \cdot \mathbf{F}_j^l, \tag{2}$$

where G_{ij}^l denotes the element at row i, column j of $\mathbf{G}^l$. $\mathbf{F}_i^l$ and $\mathbf{F}_j^l$ 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

$$\mathcal{L}_{style}\left(\boldsymbol{I_x}, \boldsymbol{I_s}\right) = \frac{1}{2{N_x^l}^2 \times 2{N_s^l}^2}||\mathbf{G}_x^l - \mathbf{G}_s^l||_2^2, \tag{3}$$

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 $\boldsymbol{I_x}$ and style-reference image $\boldsymbol{I_s}$, respectively. Finally, $\boldsymbol{I_x}$ is determined by using a gradient descent to minimize the balanced loss, defined as

$$\mathcal{L}_{total}\left(\boldsymbol{I_x}, \boldsymbol{I_c}, \boldsymbol{I_s}\right) = \alpha\mathcal{L}_{content}\left(\boldsymbol{I_x}, \boldsymbol{I_c}\right) + \beta\mathcal{L}_{style}\left(\boldsymbol{I_x}, \boldsymbol{I_s}\right), \tag{4}$$

where α and β are coefficients to balance the effect between the content and style loss terms. We note that CNN is just applied as a tool to compute the features, and that optimization of Eq. 4 is for modifying the input image so that its style resembles that of the target image, not for learning parameters of the CNN.

From the review above, $\tilde{\mathbf{P}}_j^{\phi}$, $\mathbf{P}_j^{\phi}$, and $\mathbf{P}_j^{\phi\star}$ correspond to $\boldsymbol{I_x}$, $\boldsymbol{I_c}$, and $\boldsymbol{I_s}$, respectively. While the phase $\phi\star$ with the least 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 $\mathbf{P}_j^{\phi}$.

2.3 Training and Applying the Motion Artifact Compensation Network

We adopted the very deep CNN for super-resolution (VDSR) network [11], originally applied to super-resolution, in our problem of motion artifact compensation. We chose VDSR because (1) our problem is primarily a noise reduction

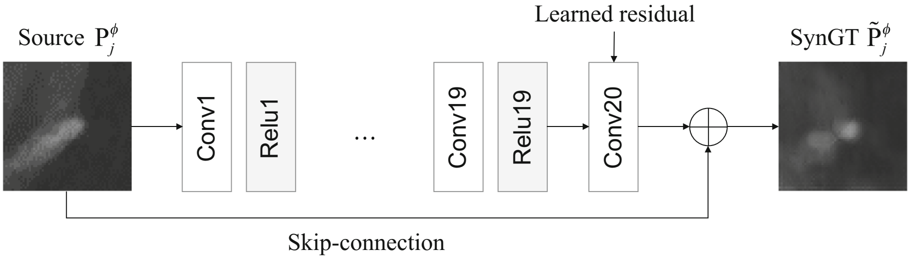

Fig. 4. Architecture of the MAC network, based on the VDSR network [11]. A pair of convolutional layers and an activation function are cascaded repeatedly. The last convolutional layer denotes a learned residual image. A single skip-connection from the input to output is applied.

problem, and noise reduction is similar to achieving super-resolution, (2) the input is upsampled such that patch sizes of the input and output are assumed to be the same for the VDSR as our configuration, and (3) it shows good performance and fast convergence during training.

The good performance is primarily due to the deep structure of the network, which combines the very deep CNN model of [15] together with the residual learning of [6]. Meanwhile skip connections were added at every other convolutional layer in [6], and 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. To expedite the training convergence, a high learning rate is used together with an adjustable gradient clipping scheme where the gradients are clipped to $\left[-\frac{\theta}{\gamma}, \frac{\theta}{\gamma}\right]$ to boost the convergence, where γ denotes the current learning rate and θ is the parameter for gradient clipping.

The structure of the MAC network follows the VDSR network, which comprises 20 convolutional layers and 19 ReLU nonlinear activation functions (Fig. 4). We used 64 filters of the size 3 × 3 for each convolutional layer. For the corresponding cross-phase style-transferred patch $\tilde{\mathbf{P}}_j^\phi$ is assigned as the GT output for the input patch $\mathbf{P}_j^\phi$, the loss function is defined as the mean squared error $\frac{1}{2}||(\tilde{\mathbf{P}}_j^\phi - \mathbf{P}_j^\phi) - f(\mathbf{P}_j^\phi)||^2$, where f denotes the network prediction of the residual between $\tilde{\mathbf{P}}_j^\phi$ and $\mathbf{P}_j^\phi$. Subsequently, the final result of the network becomes $f(\mathbf{P}_j^\phi) + \mathbf{P}_j^\phi$.

We used the Caffe [10] framework for our implementation. The hyperparameters for training are set as follows: batch size of 64, learning rate of 0.0001, and weight decay of 0.0001. The optimizer 'Adam' [12] is used.

The MAC network can be applied as follows. We assumed that a 3D CT volume of the coronary artery corrupted by motion artifacts is provided. From this volume, we extract the centerline of the coronary artery, sample M equidistant 3D point coordinates, and construct M patches, each centered at these points with normal direction as the centerline tangent direction, similarly as described in Sect. 2.1. All patches are fed into the trained MAC network, separately, where the output patches should have reduced motion artifacts.

3 Experimental Results

3.1 Datasets

In our experiments, we sampled a different number of phases from 100 4D CT volumes because some 3D volumes were excluded where the coronary artery could not be manually identified owing to extremely severe motion artifacts. A total of 5,868 mid-RCA patches were constructed. After a data augmentation process, including vertical and horizontal flips and rotation, the final training set contained a total of $35,208$ patch pairs. Each patch was constructed to be of size 60×60 when sampled from the 3D volume.

For validation, 2547 mid-RCA patches were extracted from 40 4D CT volumes and a total of 15,282 patches were constructed after a data augmentation. For testing, a total of 100 patches, extracted from 10 4D CT volumes, were used.

3.2 Qualitative Evaluation

The outputs of the trained MAC Net are presented in Fig. 5. After applying the proposed method, the edge of the coronary artery is visibly sharper than before. In addition, Fig. 6 shows that the proposed method can compensate the motion artifacts when the coronary artery diverges or contains plaques.

Two experienced readers evaluated the degree of motion artifacts based on a 5-point Likert scale as follows [2]: 1 = completely unreadable; 2 = significant motion; 3 = apparent motion; 4 = minor motion; 5 = no motion. The categorical variables are presented as the ratio of frequencies (see Table 1). The proportion of images presented with completely unreadable, significant, and apparent motions (Likert scale 1, 2, and 3) were 98.5% previously, and decreased to 35% for the MAC Net.

The mean score of the motion artifact is described as mean ± standard deviation. It 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 it shows a strong agreement: Before ($\kappa = 0.85$; 95% CI 0.76–0.95) and After ($\kappa = 0.70$; 95% CI 0.61–0.81).

4 Discussion

We proposed a motion compensation method of coronary artery in CT images based on deep learning. The key idea of the proposed method is to generate the synthetic motion compensated ground-truth by adopting the neural style-transfer method [5] using full-phase 4D CT images. It enables the patch with motion artifacts to mimic the style of the patches with small artifacts while

Table 1. 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	0.76–0.95	0.61–0.81	

retaining its content. The results of the proposed method improved qualitative readability scores from 1.43 to 3.80 on a 5-point Likert score.

While the proposed method showed improved results, in terms of the Likert score, for most cases, there were rare cases where there was no change in the score, as shown in Fig. 7. We assume that motion artifacts are too severe, or the coronary artery is too close to the right atrium or the right ventricle to distinguish its boundary from them. We note that there were no cases where the score decreased, so while there is a possibility that the proposed method does no good, there is very little possibility that it will do harm.

For future work, we intend to address the re-projection of the output patches into the original 3D CT volume and volumetric interpolation. This process is required to analyze the coronary artery in commercial software in practice. Further, we expect to quantify the motion artifacts based on the metric system and compare the performances before and after applying the proposed method. We also hope to use the quantitative measurements to perform comparative analysis with previous methods based on retrospective motion compensation based on motion estimation.

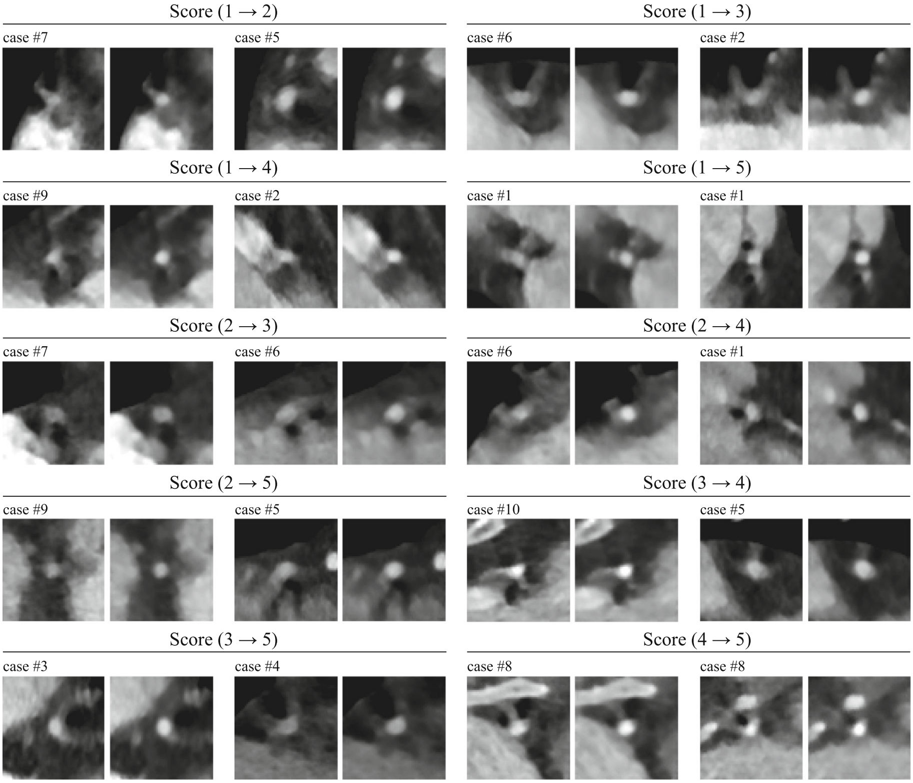

Fig. 5. Qualitative results of the test datasets. Left and right in each dataset mean before and after applying the proposed method, respectively. Expert evaluated scores based on 5-point Likert scale [2] and case number are presented above each dataset.

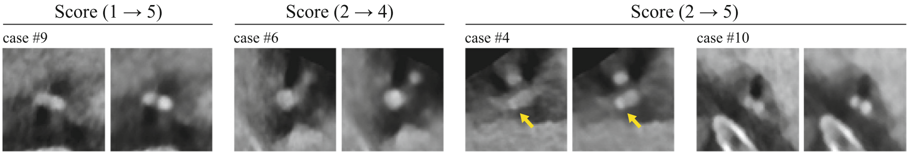

Fig. 6. Qualitative results of specific cases that distinguish well the primary vessel from the branch. The third sample pair also shows compensation for the artery plaque. Left and right in each dataset mean before and after applying the proposed method, respectively. Expert evaluated scores and case number are presented above.

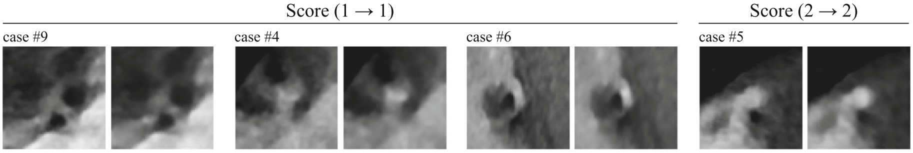

Fig. 7. Qualitative results of specific cases with little changes in motion artifacts. Left and right in each dataset mean before and after applying the proposed method, respectively. Expert evaluated scores and case number are presented above.

Acknowledgment. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (2012027176) and (NRF-2015R1C1A1A01054697).

References

1. Bhagalia, R., Pack, J.D., Miller, J.V., Iatrou, M.: Nonrigid registration-based coronary artery motion correction for cardiac computed tomography. Med. Phys. **39**(7), 4245–4254 (2012)
2. Cho, I., et al.: Heart-rate dependent improvement in image quality and diagnostic accuracy of coronary computed tomographic angiography by novel intracycle motion correction algorithm. Clin. Imaging **39**(3), 421–426 (2015)
3. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning, pp. 160–167. ACM (2008)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
5. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Heron, M.P.: Deaths: leading causes for 2012. National Vital Statistics Reports (2015)
8. Hinton, G.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig. Process. Mag. **29**(6), 82–97 (2012)
9. Isola, A.A., Grass, M., Niessen, W.J.: Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction. Med. Phys. **37**(3), 1093–1109 (2010)
10. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
11. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
12. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
14. Rohkohl, C., Bruder, H., Stierstorfer, K., Flohr, T.: Improving best-phase image quality in cardiac CT by motion correction with MAM optimization. Med. Phys. **40**(3), 031901 (2013)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
16. Tang, Q., Cammin, J., Srivastava, S., Taguchi, K.: A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT. Med. Phys. **39**(7), 4291–4305 (2012)
17. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**(Dec), 3371–3408 (2010)