Physica Medica 65 (2019) 67–75

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Using a deep neural network for four-dimensional CT artifact reduction in image-guided radiotherapy

Shinichiro Mori[a,*], Ryusuke Hirai[b], Yukinobu Sakata[b]

[a] Research Center for Charged Particle Therapy, National Institute of Radiological Sciences, Inage-ku, Chiba 263-8555, Japan
[b] Corporate Research and Development Center, Toshiba Corporation, Kanagawa 212-8582, Japan

ARTICLE INFO

Keywords:
Four-dimensional computed tomography
Artifact
Image quality
Markerless tumor tracking
Radiotherapy

ABSTRACT

Introduction: Breathing artifact may affect the quality of four-dimensional computed tomography (4DCT) images. We developed a deep neural network (DNN)-based artifact reduction method.

Methods: We used 857 thoracoabdominal 4DCT data sets scanned with 320-section CT with no 4DCT artifact within any volume (ground-truth image). The limitations of graphics processing unit (GPU) memory prevent importation of CT volume data into the DNN. To simulate 4DCT artifact, we interposed 4DCT images from other breathing phases at selected couch positions.

Two DNNs, DNN1 and DNN2, were trained to maintain the quality of the output image to that of the ground truth by importing a single and 10 CT images, respectively. A third DNN consisting of an artifact classifier and image generator networks was added. The classifier network was based on residual networks and trained to detect CT section interposition-caused artifacts (artifact map). The generator network reduced artifacts by importing the coronal image data and the artifact map.

Results: By repeating the 4DCT artifact reduction with coronal images, the geometrical accuracy in the sagittal sections could be improved, especially with DNN3. Diaphragm position was most accurate when DNN3 was applied. DNN2 corrected artifacts by using CT images from other phases, but DNN2 also modified artifact-free regions.

Conclusions: Additional information related to the 4DCT artifact, including information from other respiratory phases (DNN2) and/or artifact regions (DNN3), provided substantial improvement over DNN1. Interposition-related artifacts were reduced by use of an artifact positional map (DNN3).

1. Introduction

Image-guided radiotherapy (IGRT) and treatment planning go hand in hand with time information (4D) to increase treatment accuracy in the thoracoabdominal region [1,2]. Respiratory-induced organ motion compromises CT geometrical accuracy because the CT reconstruction algorithm was developed for static objects. The demand for time-resolved 3DCT imaging has increased and the 4DCT technique has been introduced [3,4]. 4DCT images reduce breathing artifact and provide 3D geometrical information with a time axis, which helps oncologists and medical physicists to contour targets/organs and calculate doses for respective respiratory phases. However, the different amplitudes apparent during the acquisition of different sections sometimes make it impossible to obtain a precise reproduction of the geometric shape (4DCT artifact) [5]. This can result in the degradation of image quality, which may in turn impair quantitative analysis and affect dose calculation [6]. One of the latest IGRT technology developments is real-time markerless tumor tracking using an X-ray fluoroscopic system. Our institute routinely uses this markerless tracking for carbon-ion pencil beam scanning (C-PBS) treatment [7]. Tumor tracking accuracy can be affected by image quality degradation due to 4DCT artifacts. Machine learning is key to performing markerless tumor tracking using 4DCT data sets [8–10].

Several publications have introduced 4DCT artifact reduction methods [11,12]; however, most of these have not yet been integrated into commercial CT scanners. Practically, when 4DCT artifact is observed in 4DCT images, the staff manually select tagging points on the respiratory data and try again to reconstruct the 4DCT images. Another approach is to use the area-detector CT, which has > 256 slices with a thickness of < 0.625 mm. It can acquire a 16-cm long segment in a single rotation. There will be no 4DCT artifact within the segment. Although this system is widely used for diagnostic CT imaging, few

* Corresponding author.
E-mail address: mori.shinichiro@qst.go.jp (S. Mori).

https://doi.org/10.1016/j.ejmp.2019.08.008
Received 22 December 2018; Received in revised form 5 August 2019; Accepted 8 August 2019
Available online 17 August 2019

centers use it for treatment planning CT.

Current neural networks are deepened by integrating several techniques, such as batch normalization and dropout [13,14]. Use of a graphics processing unit (GPU) helps to optimize huge network parameters. Current deep neural networks (DNNs) have achieved better performance than autoencoders by applying deep architecture [15]. In medical imaging, this technique has been used for image denoising [16,17], bone-density suppression [18], and image quality processing [19,20]. A similar approach to 4DCT artifact reduction is inpainting, which is an image restoration task. DNNs have been used in inpainting, mainly in industrial settings [21–23]. Since the features of cars, buildings, and other manufactured objects tend to be similar, inpainting produces good results despite low image spatial resolution. The human anatomical shape is so complex, however, particularly pulmonary vessels, that inpainting may not be applicable. It is difficult to provide rich information from low-spatial-resolution images due to the limitations of GPU memory. A deeper neural network might reduce the magnitude of 4DCT artifact by importing full 4DCT data sets and other information. However, such a system cannot be handled by GPU memory.

To solve these problems, we developed three types of 4DCT artifact reduction methods using DNN with/without additional information related to the cause of the 4DCT artifact, and compared them using thoracoabdominal 4DCT data.

2. Materials and methods

2.1. Image preparation

2.1.1. Image acquisition

A total of 236 patient thoracoabdominal examination data sets (107 thoracic regions (=103 lung, 1 breast and 3 esophagus), 96 abdominal regions (=31 liver and 65 pancreas), 33 pelvic regions (=29 bone-soft tissue sarcoma and 4 rectal cancer)) were randomly selected from our institution. 4DCT image sets had been acquired by a fast-rotating area detector CT (Aquilion One Vision®, Toshiba Medical Systems, Otawara, Japan) under free breathing conditions. This CT acquires an approximately 16-cm segment in a single rotation. Slice thickness and reconstructed field of view were 1.0 mm and 500 mm, respectively. To obtain a sufficient scan region for treatment planning required repeatable volumetric cine scans at serial couch positions. 4DCT artifact was not created within the respective couch positions. 4DCT images were organized into 10 CT data sets (T00: peak inhalation, T50: exhale) by sorting CT images at each couch position in the same respiratory phase. Since our treatment center uses orthogonal treatment beam ports, the treatment couch is rotated around its longitudinal axis and/or patient position is changed (supine or prone) to increase the available beam angles. If necessary, multiple 4DCTs are acquired in the same patient. A total of 857 CT data sets were used.

2.2. Simulation of 4DCT artifacts

To simulate 4DCT artifacts acquired by multi-slice CTs, we sorted 4DCT images at respective couch positions by selecting 4DCT images at respiratory phases ± 20% from other respiratory phases [6]. We had previously reported that the intrafractional exhalation position obtained with the external respiratory monitor varied by 7.5 ± 4.6% during treatment [24]. The monitor does not track actual tumor position, but rather abdominal surface motion, so we assumed that the external respiratory amplitude corresponded to the actual tumor position. When we considered two sigma statistical variation, exhalation positional variation could be 16.7%. Because 4DCT time coordinates are expressed by respiratory phase, the 4DCT phase at 16.7% position from peak exhalation was therefore close to T50 ± 20%. The couch moving steps were 8 mm, 16 mm, and 32 mm.

2.3. Image pre-processing

Due to the limitations of GPU memory, it is impossible to import 4DCT volume data directly into the DNN. One solution is 4DCT image size reduction. However, it is still difficult to import all 4DCT data even though it is a single phase, and small image size degrades minute structural information. We, therefore, generated the following image preprocessing to import 4DCT images into the DNN:

(i) Original CT data format is generally in 16-bit dynamic range, but it is sufficient to visualize patient geometrical structure on an 8-bit display.
(ii) The number of CT sections differed among patients. To handle data easily, we set 320 CT slices for all subjects.
(iii) 4DCT artifact was caused in respective section positions. It would therefore be better to import coronal/sagittal reconstructed images rather than axial images to detect and correct the 4DCT artifact. CT volume data was exported in the coronal and sagittal planes section by section. Air-only regions were removed and all images were resized 80% (320 × 256 pixels).

All image preprocessing was performed using a commercial programming environment (MATLAB R2016a®, Mathworks, Natick MA, USA).

2.4. Network architecture

Two of the three types of DNNs we examined were based on the image generator network, and the third one was constructed by an artifact classifier and image generator networks. In the following subsections, we describe the details of the DNNs:

2.4.1. Image generator network

Neural networks such as autoencoder (AE) are often used to restore image quality [21,25–27]. To obtain better performance, current neural networks apply deep architecture and other new techniques [13,28,29]. To extend these technologies to medical image processing, we used convolutional AE (CAE) with shortcut connections (Fig. 1a).

Our network structure for an image generator involved encoding and decoding procedures. For the encoding process, we added multiple sets of convolution (Conv), batch normalization (BN), and rectified linear units (ReLU), and then data dimensions were reduced to half size by applying max pooling (both kernel and stride size were 2 × 2 pixels). Convolutional kernel and stride size were 3 × 3 pixels and 1 × 1 pixels, respectively. The number of convolutional output channels was 64 in the first two convolutional layers; these were increased by a factor of two after respective max pooling layers. After the 5th max pooling layer, we added two sets of Conv + BN + ReLU layers with the convolutional kernel and filters of 7 × 7 pixels and 7000 × 7000 pixels.

For the decoding process, we added deconvolution (Deconv) + BN + ReLU layers with the deconvolutional kernel and filters of 7 × 7 pixels and 512 × 512 pixels, then added data from the encoding process in the same dimensions (shortcut connections) by applying an element-wise (Eltwise) layer. The shortcut connection is useful for preventing the vanishing gradient problem during training and keeping minute structures on the original image. Data dimensions were increased by a factor of two by applying a pixel shuffle layer (PS) [30], which is upscaled to the high resolution rather than the bicubic filter. Then two or three sets of Deconv + BN + ReLU layers and a single set of element-wise summation (Eltwise) + PS layers were repeatedly added to increase the data dimension. After adding the PS layer sets, the number of output channels of the deconvolution layer was decreased by a factor of two. Deconvolutional kernel and stride size were 3 × 3 pixels and 1 × 1 pixels, respectively. In the last convolutional layer, the number of output channels depended on the network architecture (=10 channels for DNN2, 1 channel for DNN1 and DNN3).

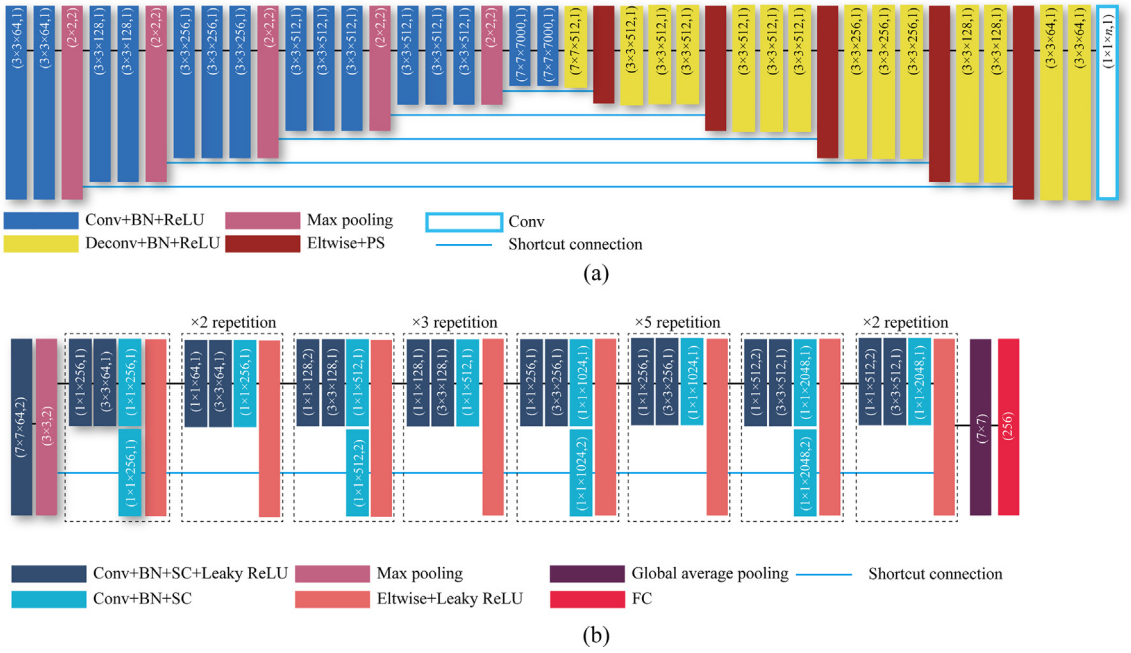

Fig. 1. Network structures for (a) the image generator based on CAE with shortcut connections, and (b) the artifact classifier based on ResNet50. Convolutional kernel size, the number of outputs channels and stride size are expressed by (kernel × kernel × outputs channels, stride), and pooling kernel size and stride size are expressed (kernel × kernel, stride) in the figures. Light blue line shows shortcut connection. Dotted square in Fig. 1b shows bottleneck building block.Abbreviations: CAE = convolutional autoencoder; Conv = convolution; BN = batch normalization; ReLU = rectified linear units; Deconv = deconvolution; Eltwise = element-wise summation; PS = pixel shuffle; SC = scale; FC = fully connection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Although we would have preferred to use more than three sets of Conv/Deconv + BN + ReLU, and to increase the number of output channels in our network, we reduced the number of Conv/Deconv + BN + ReLU in the first and last two sets and the number of output channels was not changed after the pooling/PS layer due to the constraints of the GPU memory.

2.4.2. Artifact classifier network

To detect CT section positions causing 4DCT artifact, we designed a ResNet50-based classifier network [15]. We modified the Resnet50 to replace ReLU with a leaky ReLU [31], which uses negative values in the activation function. By doing this, artifact classifier accuracy was increased by approximately 20%. The classifier network involves Conv + BN + scale layer (SC) + Leaky ReLU layers (Fig. 1b). The convolutional kernel and stride sizes were 7 × 7 pixels and 2 × 2 pixels, respectively. A max-pooling layer was then added to reduce the spatial size. We used bottleneck building block, which included three sets of Conv + BN + scale layer (Sc) + leaky ReLU layers (dotted square in Fig. 1b). The convolutional kernel and stride sizes of the three Conv layers were 1 × 1, 3 × 3, and 1 × 1 pixel, respectively. The number of output channels in the third convolutional layer was four times larger that of the first and second convolutional layers. To improve training accuracy, shortcut connections [32] were added between the first and second and after the third convolutional layers in the bottleneck building block. Four parallel bottleneck building blocks (each made up of Conv (kernel size of 1 × 1 pixel) + BN + SC layers) were included to adjust input and output dimensions (lower light blue box in Fig. 1b). A total of 16 bottleneck building blocks were placed to form a deeper structure. We then used two parallel fully-connected (FC) layers (256 output channels) after the last global average pooling layer with a kernel size of 7 × 7 pixels to output the artifact position in the CT section plane. The global average pooling layer reduced the number of parameters – as a result, GPU memory usage could be reduced and training time accelerated, even as classification accuracy was increased [33]. Details of the network parameters are shown in Fig. 1b.

2.4.3. DNN-based artifact correction methods

We evaluated the three types of artifact correction method using the following methods:

DNN1: The first artifact correction method consisted of the image generator network (Fig. 2a). It imported a single coronal image (= 320 × 256 × 1 dimensions) with artifact and predicted an artifact reduction coronal image (= 320 × 256 × 1 dimensions). To obtain 4DCT volumetric data sets, this process was repeated for all coronal images as a function of the respiratory phase.

DNN2: Tumor position is not always the same, even in the same respiratory phase [7]. This inconsistency between respiratory phase and tumor position could cause geometrical error in 4DCT images. However, corrected geometrical information might be included in 4DCT images in adjacent respiratory phases. Therefore, it could be useful to perform artifact correction by using 4DCT images in other phases. The second artifact correction method again used the image generator network, but coronal images of each respiratory phase (a total of 10 CT images) (= 320 × 256 × 10 dimensions) were imported into the DNN, and then 10 coronal images with artifact reduction (= 320 × 256 × 1 dimensions) were exported for each respiratory phase (Fig. 2b).

DNN3: Since both DNN1 and DNN2 can import single or multiple images only into the DNN due to the limitations of GPU memory, neither method considered any information about other CT images. As a result, the output image could cause geometrical inconsistency between adjacent coronal images. One solution is to correct artifact regions only for respective CT sections by using information for the 4DCT artifact region. We therefore constructed a third artifact correction method which includes the artifact classifier networks and the image generator networks (Fig. 2c).

4DCT artifact was caused section by section; therefore, the 4DCT artifact region was the same on respective coronal images at the same respiratory phase. To increase the accuracy of the artifact

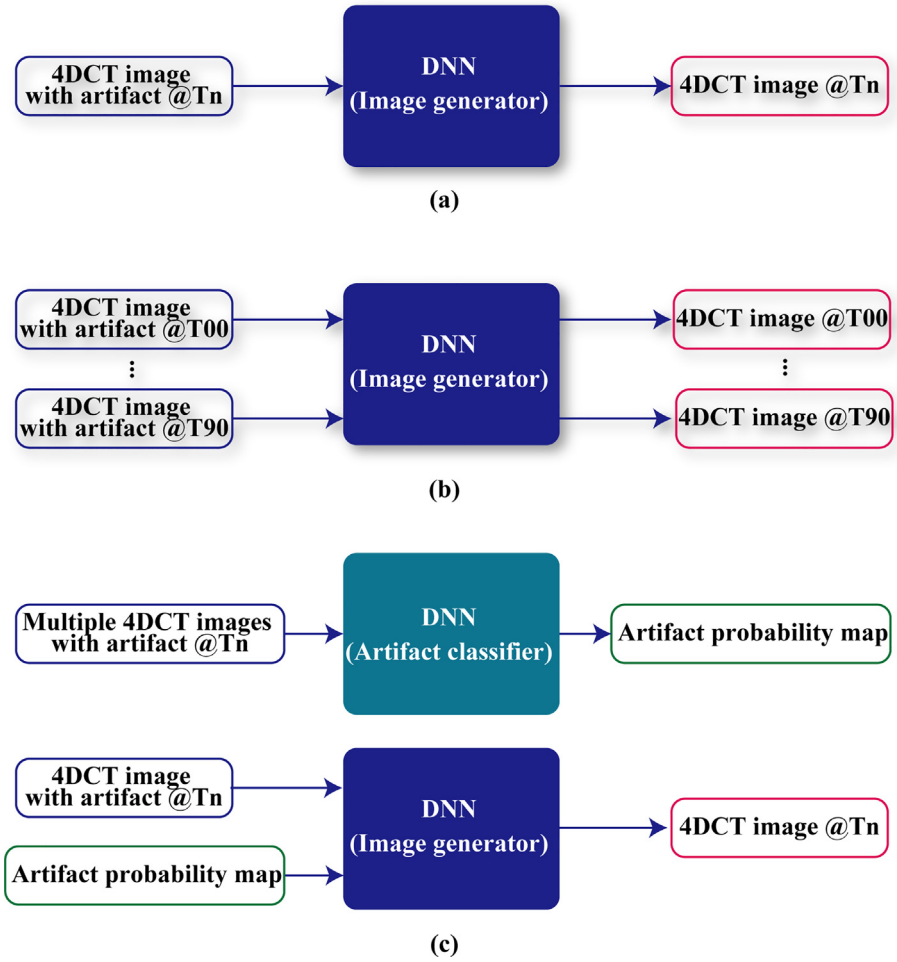

Fig. 2. Three types of DNNs for artifact reduction. (a) DNN1: a single image is input into the image generator deep neural network (DNN) and the DNN outputs the single CT image. (b) DNN2: single CT image for each of 10 respiratory phases positions are input into the image generator DNN, and the DNN outputs the 10 images for each phase. (c) DNN3: multiple CT images at the same respiratory phase are input into the artifact classifier DNN, which outputs the artifact probability map. Then, a single CT image and the artifact probability map are input into the image generator DNN, and the DNN outputs the single image.

discrimination, multiple coronal images with artifact were imported into the classifier network. In this study, five coronal images were imported into the classifier network (= 320 × 256 × 5 dimensions). The classifier network exported the artifact probability (0%: no artifact, 100%: artifact) at respective CT section positions (artifact probability map) (= 1 × 256 × 1 dimensions). The artifact probability map dimension was reshaped to 320 × 256 × 1 dimension by repeating the data. 4DCT artifact regions were finally defined by applying the threshold to the probability. Then, a single coronal image with artifact and the artifact probability (= 320 × 256 × 2 dimensions) were imported into the image generator network, and a single artifact-corrected coronal image (= 320 × 256 × 1 dimensions) was predicted. This process was repeated for respective CT images by using the same artifact probability, but it should be kept in mind that the artifact probability should be recalculated in respective respiratory phases.

2.4.4. Network training

All DNNs were trained to keep the quality of the output image close to that of the ground-truth image by importing artifact image. A total of 440,000 images in coronal and sagittal sections (179 patients) was used for training. To increase the number of training images, the CT data were rotated randomly along the longitudinal axis (CT slice direction), and 10–20 patterns of the simulated 4DCT images per image were prepared. An optimization procedure was performed for 1,870,000 iterations with a batch size of 3 using mini-batch stochastic gradient descent (SGD) to minimize the sum of squares of differences between output images through the network and the ground-truth images (Euclidean loss) [13]. The learning rate was 0.01 and SGD parameters of momentum, weight decay, and clip gradients were 0.9, 0.0001, and 50, respectively. We used the deep learning framework "Caffe" [34] in a 64-bit environment (Windows 7®, Microsoft Corp, Redmond WA, USA) and a single GPU on a board (NVIDIA Quadro P5000®, NVIDIA Corporation, Santa Clara, CA, USA), which is equipped with 2560 compute unified device architecture (CUDA) cores and has 16 GB of memory. Training time was 20 days.

2.5. Image quality evaluation

We evaluated the image processing performance in coronal sections predicted with the three DNNs. The magnitude of the incomplete artifact correction could be averaged out by using coronal rather than sagittal sections. We therefore imported coronal section image data into the network. We also evaluated the quality of sagittal images to analyze geometrical consistency in each coronal section.

The respective network models predicted the output image from the input image. These output images were compared with the ground-truth image using the mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) [35].

$$MSE = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}[IM_{ground-truth}(x,y) - IM_{output}(x,y)]^2, \quad (1)$$

where m and n are the number of image matrices, $IM_{ground\text{-}truth}$ and IM_{output} are ground-truth and output images, respectively.

$$PSNR = 10 \cdot \log_{10} \frac{MAX^2}{MSE} \quad (2)$$

where MAX is the maximum pixel values of the input image.

$$SSIM = \frac{(2\mu_{ground-truth}\mu_{output} + c_1)(2\sigma_{ground-truth,output} + c_2)}{(\mu^2_{ground-truth} + \mu^2_{output} + c_1)(\sigma^2_{ground-truth} + \sigma^2_{output} + c_2)} \quad (3)$$

where $\mu_{ground\text{-}truth}$, μ_{output}, $\sigma_{ground\text{-}truth}$, and σ_{output} are the average pixel values of the ground-truth image, the output image, the variance of the ground-truth image, and the output image, respectively. $\sigma_{ground\text{-}truth, output}$ is the covariance of the ground-truth and output images.

These metrics were evaluated in respective images within the subject region and then averaged.

A total of 1500 images and 100,000 images (57 patients) were used for evaluation of the classifier and generator networks, respectively.

3. Results

3.1. Artifact discrimination accuracy

One example of a coronal image with 4DCT artifact is shown in Fig. 3a. 4DCT artifacts are clearly detectable around the diaphragm; however, it is more difficult to recognize the artifact regions around the pulmonary and lower abdominal vessels. The artifacts can be visualized clearly by subtracting the artifact image from the ground-truth image (Fig. 3b). Artifact probability values (0–100%) derived from the classifier network in DNN3 were overlaid on the image difference. Artifact probability values were correctly derived on most regions; however, some regions showed wrong probability values explicitly (marked with red arrows in Fig. 3b).

We changed the artifact probability threshold to find the maximum classifier network accuracy (Fig. 3c). The maximum classifier network accuracy was 73.6% at an artifact probability threshold of 72%.

3.2. Image quality

3.2.1. Coronal section

The first case caused significant 4DCT artifact around the dome of

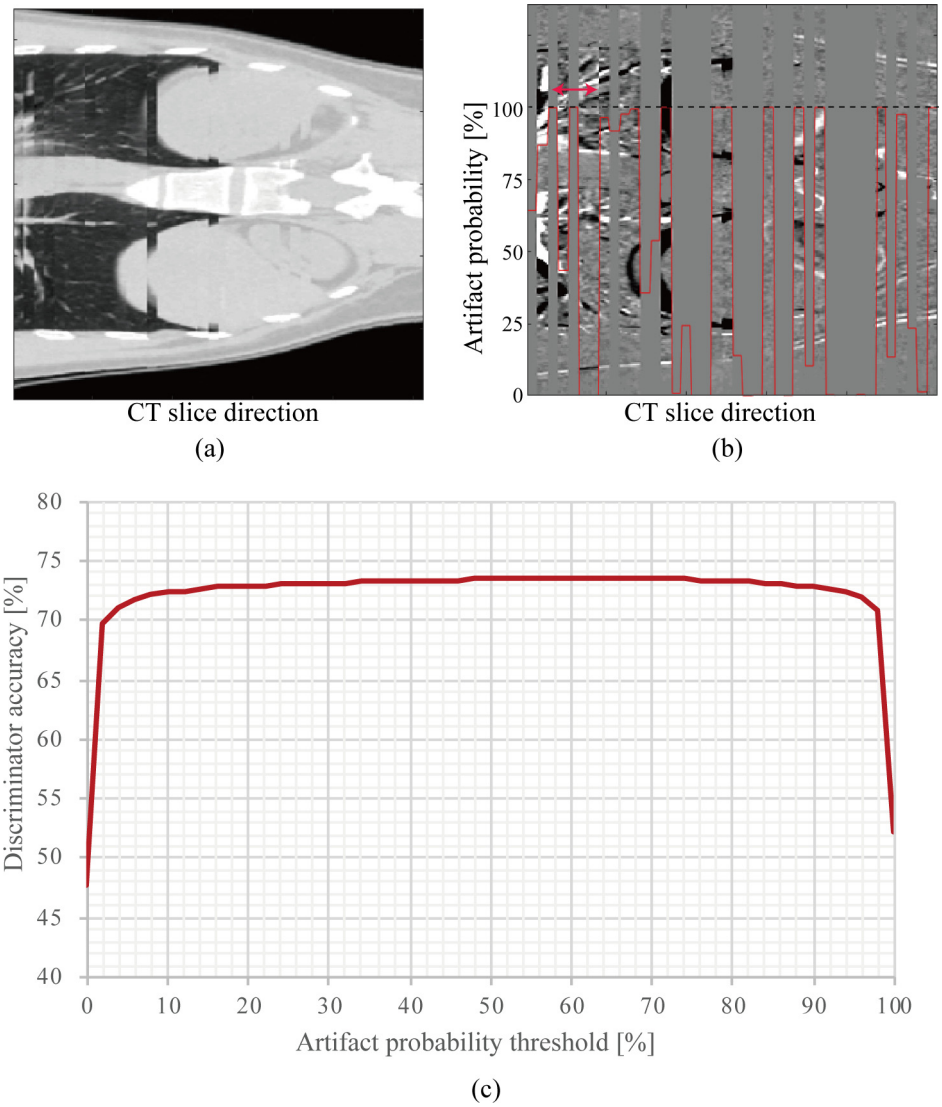

Fig. 3. (a) Input image with 4DCT artifacts (subject 35). (b) Artifact probabilities overlaid on image difference ([a] minus ground-truth images). (c) Artifact classifier network accuracy as a function of the artifact probability threshold over all patients.

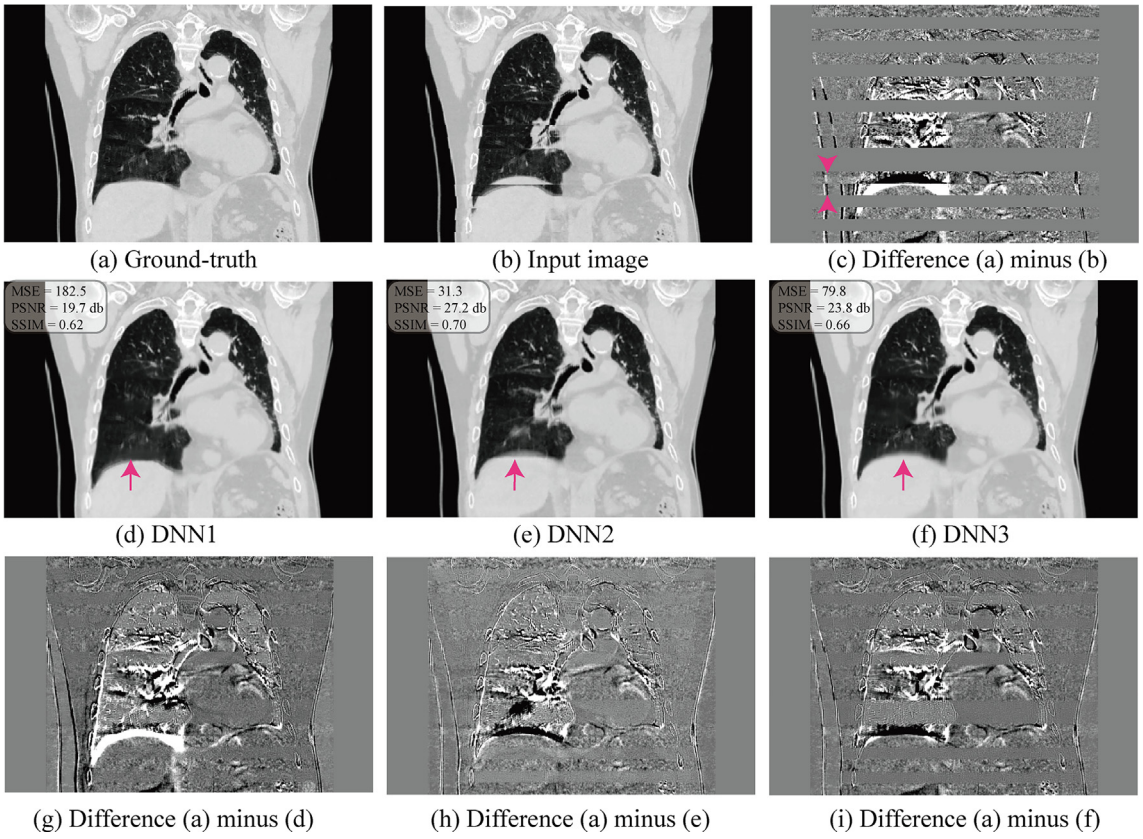

Fig. 4. 4DCT images at T60 in coronal section (patient 402). (a) Ground-truth image. (b) Input image (with 4DCT artifact). (c) Image differences ([a] minus [b]). (d) Output image with DNN1. (e) Output image with DNN2. (f) Output image with DNN3. (g) Image differences ([a] minus [d]). (h) Image differences ([a] minus [e]). (i) Image differences ([a] minus [f]). DNN = deep neural network; MSE = mean square error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measurement.

the diaphragm region (subject 402, T60). The 4DCT artifacts were observed in the input image, especially around the diaphragm (Fig. 4a–c). It is hard to locate the dome of the diaphragm due to the duplication artifact. The output image with DNN1 minimized the magnitude of the artifact, and the diaphragm shape seemed natural (MSE/PSNR/SSIM = 182.5/19.7 db/0.62). However, it was not correct when compared with the ground-truth image, and geometrical accuracy was still degraded (Fig. 4d and g).

While the output image with DNN2 improved geometrical accuracy better than DNN1 (MSE/PSNR/SSIM = 31.3/27.2 db/0.70) (Fig. 4e and h), DNN3 maintained the original image quality with no artifacts and most anatomical shapes close to the ground truth. In 4DCT artifact, particularly those involving the dome of the diaphragm, the predicted diaphragm shape seemed natural (Fig. 4f), but was moderately different from that in the ground-truth image (Fig. 4i).

While the same subject at a different respiratory phase (T70) also exhibited 4DCT artifacts (Fig. 5b), the artifact was not at the dome of the diaphragm (marked with red arrows in Fig. 5b and c). All output images by respective DNNs minimized the magnitude of the artifact. However, since DNN1 output images by using input images only, the DNN1 image was different from the ground-truth image (Fig. 5g). The diaphragm shape in the DNN1 image was significantly different from that in the ground-truth image (marked with a red arrow in Fig. 5g). The image quality in DNN1 (MSE/PSNR/SSIM = 308.4/17.4 db/0.57) was therefore lower than that of the others (Fig. 5e and f), while both DNN2 and DNN3 images improved the artifacts and the diaphragm positions were almost the same as those in the ground-truth image (Fig. 5h and i). No artifact around the diaphragm dome (marked with red arrows in Fig. 5c) was kept in DNN3 (Fig. 5i).

Image quality in DNN2 averaged over all subjects (MSE/PSNR/SSIM = 16.6 ± 7.9/31.1 ± 1.8 db/0.76 ± 0.06) ranked highest of all (Table 1). PSNR value in DNN3 (= 29.2 ± 2.0 db) was higher than that in DNN1 (= 28.1 ± 2.6 db); however, MSE and SSIM values in DNN3 (= 27.8 ± 15.9/0.71 ± 0.06) were smaller than those in DNN1 (= 41.8 ± 35.0/0.72 ± 0.06).

3.2.2. Sagittal sections

Sagittal 4DCT input image in the same case in Fig. 6 (patient 402, T70) showed 4DCT artifacts (Fig. 6b), not involving the dome of the diaphragm but rather the tumor (marked with arrows in Fig. 6c). DNN1 image showed incorrect tumor shape (marked with an arrow in Fig. 6d) because DNN predicted output images in each coronal section independently, causing geometrical inconsistency in the sagittal plane (Fig. 6g).

Both DNN2 and DNN3 successfully minimized the magnitude of the 4DCT artifacts and these anatomical shapes were close to the ground truth (Fig. 6e and f). the dome of the diaphragm and tumor shapes were correctly depicted (Fig. 6h and g).

Image quality metrics averaged over all subjects in DNN2 (MSE/PSNR/SSIM = 19.6 ± 8.3/30.5 ± 1.5 db/0.74 ± 0.06) were better than those in DNN3 (MSE/PSNR/SSIM = 30.5 ± 16.0/28.9 ± 1.7 db/0.70 ± 0.06). Image quality in DNN1 was lowest of all (MSE/PSNR/SSIM = 44.4 ± 34.9/28.0 ± 2.1 db/0.71 ± 0.06) (Table 1).

4. Discussion

We developed 4DCT artifact reduction DNNs compatible with limited GPU memory. By applying DNN to the whole image region (DNN1), the predicted CT image was improved but was not close to the ground truth. Considering the cause of 4DCT artifact, minimizing the

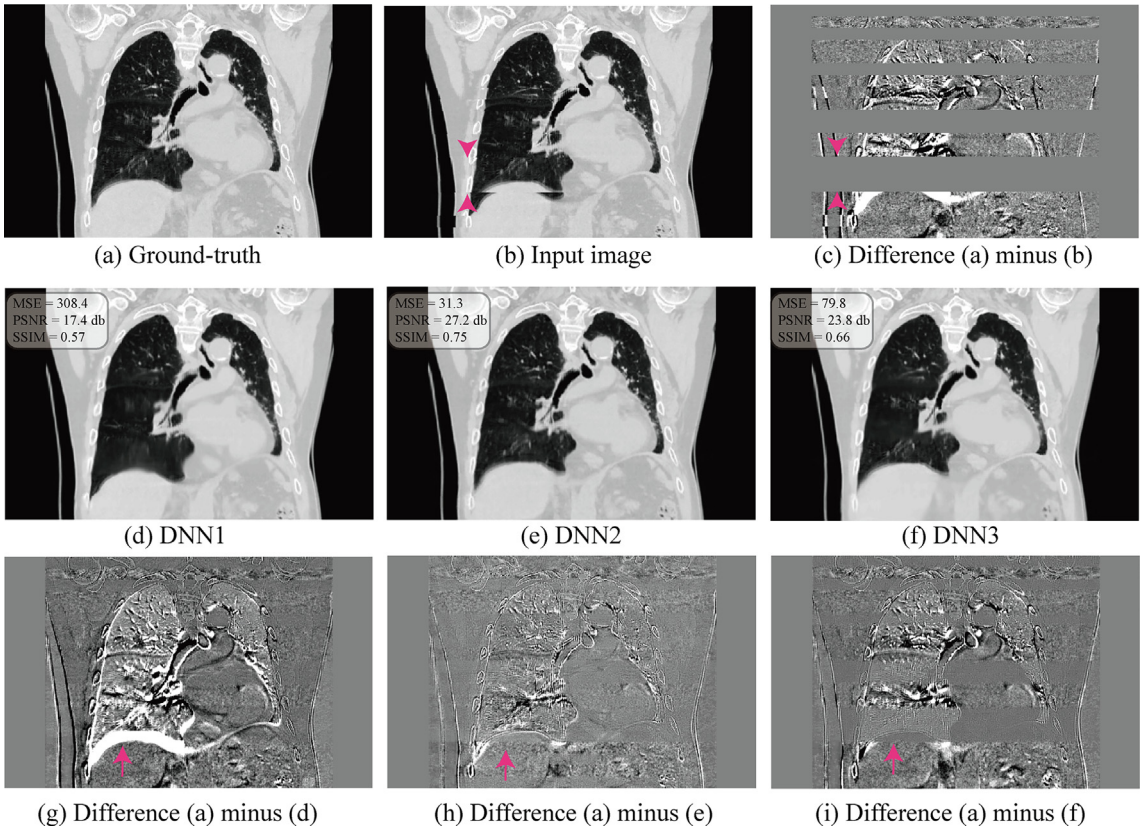

Fig. 5. 4DCT images at T70 in coronal section (subject 402). (a) Ground-truth image. (b) Input image (with 4DCT artifact). (c) Image differences ([a] minus [b]). (d) Output image with DNN1. (e) Output image with DNN2. (f) Output image with DNN3. (g) Image differences ([a] minus [d]). (h) Image differences ([a] minus [e]). (i) Image differences ([a] minus [f]).DNN = deep neural network; MSE = mean square error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measurement.

Table 1
Quality of 4DCT image assessment results with respective methods averaged over all patients.

Image type	DNN	Metrics	Mean ± SD	(Min–Max)
Coronal section	DNN1	MSE	41.8 ± 35.0	(6.0–471.8)
		PSNR	28.1 ± 2.6	(17.3–37.1)
		SSIM	0.72 ± 0.06	(0.48–0.91)
	DNN2	MSE	16.6 ± 7.9	(3.6–123.0)
		PSNR	31.1 ± 1.8	(23.7–37.8)
		SSIM	0.76 ± 0.06	(0.44–0.91)
	DNN3	MSE	27.8 ± 15.9	(4.9–100.0)
		PSNR	29.2 ± 2.0	(22.3–36.2)
		SSIM	0.71 ± 0.06	(0.60–0.94)
Sagittal section	DNN1	MSE	44.4 ± 34.9	(7.0–473.8)
		PSNR	28.0 ± 2.1	(18.8–35.6)
		SSIM	0.71 ± 0.06	(0.47–0.89)
	DNN2	MSE	19.6 ± 8.3	(4.6–125.7)
		PSNR	30.5 ± 1.5	(24.0–35.9)
		SSIM	0.74 ± 0.06	(0.42–0.90)
	DNN3	MSE	30.5 ± 16.0	(5.8–105.0)
		PSNR	28.9 ± 1.7	(22.3–34.9)
		SSIM	0.70 ± 0.06	(0.57–0.91)

SD = standard deviation; MSE = mean square error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measurement.

artifact by using CT image information from other respiratory phases at the same couch position (DNN2) is a useful approach, but DNN2 could also modify CT images without artifact. In DNN3, artifact reduction applied to CT images still caused artifacts by importing artifact positional information. DNN3 minimized the magnitude of residual artifact, and its quality of output image quality is better than that of DNN1. Since DNN3 imported single-phase CT image and artifact positions, it can be applied to gated CT for artifact reduction and provide better image quality than DNN1.

If possible, use of additional information such as CT images from other phases and artifact image regions (DNN2 and DNN3) is advisable. Integrating sufficient GPU memory to import CT volume data for deep learning might yet take a long time. However, it would not take so long to devise new DNNs to import CT images with artifact position information for 10 respiratory phases (DNN2 + DNN3). The accuracy of 4DCT artifact reduction could then be expected to be much improved over our current results. Therefore, when new GPUs with sufficient GPU memory become available, we will investigate this point.

Output images by respective DNNs show minimal blurring of lung vessels in some regions. This was reported in other studies [22,36]. This is a reason to require deeper neural networks and/or increase the number of convolutional channels to restore images closer to the ground truth. A second reason for blurring is the difficulty of predicting human anatomical shape due to its complexity.

We did not use other new DNNs for image generation such as a generative adversarial network (GAN) [29,37] and variational autoencoder (VAE) [38] because our purpose in this study was artifact reduction with additional information with limited GPU memory. GAN has demonstrated significant improvement in image restoration quality by judging generated image accuracy. The advantage of VAE is to generate a cine, making it suitable for 4DCT images. Though both might minimize the magnitude of 4DCT artifact, using additional information, such as 10 respiratory phase images and artifact positions, as presented in our study, would lead to further image quality improvement.

Recent DNN-based image segmentation used U-Net [28], which was extended to image restoration [39]. U-Net is based on CAE with skip connections similar to the ResNet, but with replacement of the element-wise summation with the concatenation operation. The increase in the

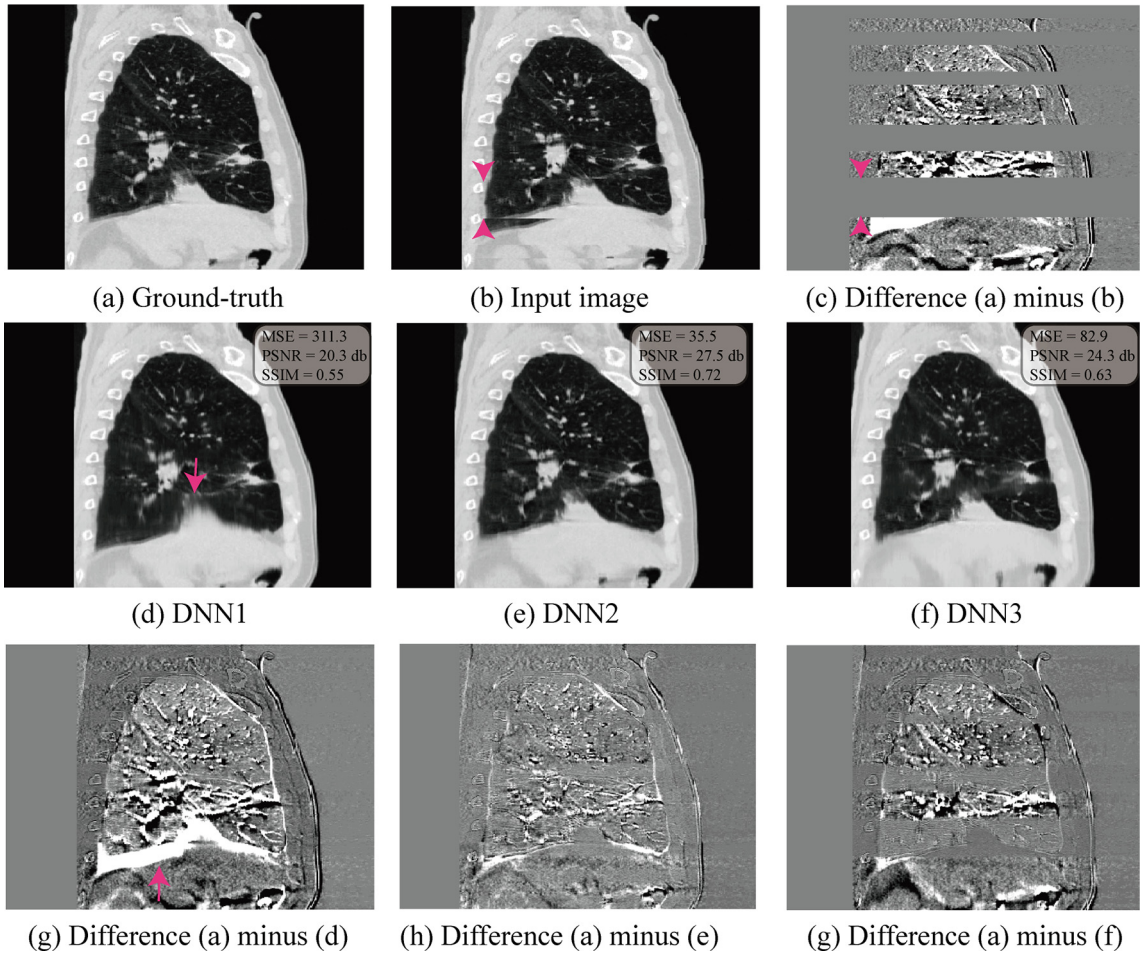

Fig. 6. 4DCT images at T70 in sagittal section (subject 402). (a) Ground-truth image. (b) Input image (with 4DCT artifact). (c) Image differences ([a] minus [b]). (d) Output image with DNN1. (e) Output image with DNN2. (f) Output image with DNN3. (g) Image differences ([a] minus [d]). (h) Image differences ([a] minus [e]). (i) Image differences ([a] minus [f]). DNN = deep neural network; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measurement.

number of channels required by the concatenation operation resulted in an increase in GPU memory usage. Because the network structure for the image generator in our study is based on U-net, the results with U-net might be almost the same as ours, therefore, we did not evaluate artifact reduction using U-Net.

Due to space limitations, we did not describe fluoroscopy-guided markerless tumor tracking with machine learning, which was trained by DRR images calculated from 4DCT data. Patient geometrical errors due to 4DCT artifact affected DRR images, even though DRR calculation averaged it out along each ray-tracing. Since tumor and diaphragm shapes are important information in machine learning training for lung and liver cases, respectively, the 4DCT artifact reduction by our DNNs could improve markerless tumor tracking accuracy. Our DNNs calculated CT volume data with artifact reduction by stacking the predicted coronal image. In our treatment room, dynamic flat panel detectors (DFPD) are installed on either side of the vertical irradiation port, and the respective X-ray tubes are installed under the floor [40]. Since the X-ray direction is close to the coronal plane, importing coronal images could minimize the impact of inconsistency of artifact correction in each image on DRR images.

The major problem in DNN is the limitations of GPU memory; therefore, most researchers decrease image spatial resolution and/or decrease the number of DNN parameters. These devices could reduce predictive accuracy and may not be suitable in practical situations. In 2018, the latest GPU for DNN has 32 GB memory (NVIDIA Tesla V100®/ Quadro GV100, NVIDIA Corporation, Santa Clara, CA, USA). It is, however, insufficient to import whole CT volume data. When new GPU with sufficient memory to import CT volume data is developed or new technology to combine GPU memories with multi-GPUs is developed, importation of whole CT data with DNN would improve image quality over that described in this study.

A few limitations of this study warrant mention. First, the DNN image generator did not completely remove artifacts; rather, it minimized the magnitude of the 4DCT artifact. Moreover, our results showed our predicted CT images were not completely the same as the ground truth. Even though medical staff can prevent misdiagnosis by paying close attention to distortions to infer that the tumor shape, etc., may be distorted due to the artifact, we do not recommend the use of our method for diagnostic purposes as a clinical application at the present time. However, since CT images predicted by DNN have a decreased magnitude of artifact, it is more difficult to recognize whether predicted patient anatomical information is true or not.

Second, CT data was resized to 80% and dynamic range was reduced to 8-bit to avoid the limitations of GPU memory error in this study. Clinical application will therefore require training the DNN with the original CT data using GPUs with sufficient memory.

Third, we did not analyze our DNNs using CT data with real respiratory artifact. This is because we only had 320-slice CT image sets, which do not have 4DCT artifact within the respective couch positions. Results for these might therefore slightly differ from those with the simulated 4DCT data sets. In particular, DNN2 corrected the artifact with other CT images obtained from different respiratory phases. This may substitute for correct CT images from other phases, albeit not completely. As a result, the corrected image was lightly blurred.

Moreover, it is difficult to say which DNN showed the correct geometry because we have no ground-truth image. In any case, we will evaluate our DNNs using real 4DCT artifact in a subsequent study.

5. Conclusion

In accounting for the causes of 4DCT artifact, prediction using input CT image information only (DNN1) to predict good image quality close to the ground truth is not optimum. The addition of information related to the 4DCT artifact, including that from other respiratory phases (DNN2) and/or artifact regions (DNN3), provides a substantial improvement over DNN1.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank Libby Cone, MD, MA, from DMC Corp. (www.dmed.co.jp) for editing drafts of this manuscript.

References

[1] Mori S, Zenklusen S, Knopf AC. Current status and future prospects of multi-dimensional image-guided particle therapy. Radiol Phys Technol 2013;6:249–72.
[2] Bert C, Herfarth K. Management of organ motion in scanned ion beam therapy. Radiat Oncol 2017;12:170.
[3] Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med Phys 2006;33:3874–900.
[4] Mageras GS, Pevsner A, Yorke ED, Rosenzweig KE, Ford EC, Hertanto A, et al. Measurement of lung tumor motion using respiration-correlated CT. Int J Radiat Oncol Biol Phys 2004;60:933–41.
[5] Yamamoto T, Langner U, Loo Jr. BW, Shen J, Keall PJ. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. Int J Radiat Oncol Biol Phys 2008;72:1250–8.
[6] Mori S, Kumagai M, Karube M, Yamamoto N. Dosimetric impact of 4DCT artifact in carbon-ion scanning beam treatment: worst case analysis in lung and liver treatments. Phys Med 2016;32:787–94.
[7] Mori S, Karube M, Shirai T, Tajiri M, Takekoshi T, Miki K, et al. Carbon-ion pencil beam scanning treatment with gated markerless tumor tracking: an analysis of positional accuracy. Int J Radiat Oncol Biol Phys 2016;95:258–66.
[8] Hirai R, Sakata Y, Taguchi Y, Mori S. Regression model of tumor and diaphragm position for marker-less tumor tracking in carbon ion scanning therapy for hepatocellular carcinoma. American Society for Radiation Oncology (ASTRO): Elsevier; 2016. p. E639–40.
[9] Richter A, Wilbert J, Baier K, Flentje M, Guckenberger M. Feasibility study for markerless tracking of lung tumors in stereotactic body radiotherapy. Int J Radiat Oncol Biol Phys 2010;78:618–27.
[10] Li R, Lewis JH, Cervino LI, Jiang SB. A feasibility study of markerless fluoroscopic gating for lung cancer radiotherapy using 4DCT templates. Phys Med Biol 2009;54:N489–500.
[11] Zhang Y, Yang J, Zhang L, Court LE, Balter PA, Dong L. Modeling respiratory motion for reducing motion artifacts in 4D CT images. Med Phys 2013;40:041716.
[12] Li T, Schreibmann E, Yang Y, Xing L. Motion correction for improved target localization with on-board cone-beam computed tomography. Phys Med Biol 2006;51:253–67.
[13] Sergey I, Christian S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: The 32nd International Conference on Machine Learning. 2015:pp. 448–456.
[14] Rumelhart ED, Hinton EG, Williams JR. Learning representations by back-propagating errors. Nature 1986;323:533–6.
[15] Simonya K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference for Learning Representations 2015.
[16] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010;11:3371–408.
[17] Jonathan M, Ueli M, Dan C, Jürgen S. Stacked convolutional auto-encoders for hierarchical feature extraction. International Conference on Artificial Neural Networks. Heidelberg: Springer; 2011. p. 52–9.
[18] Yang W, Chen Y, Liu Y, Zhong L, Qin G, Lu Z, et al. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. Med Image Anal 2017;35:421–33.
[19] Xiao H. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19.
[20] Mori S. Deep architecture neural network-based real-time image processing for image-guided radiotherapy. Phys Med 2017;40:79–87.
[21] Junyuan X, Linli X, Enhong C. Image denoising and inpainting with deep neural networks. Adv Neural Inform Process Syst 2012:341–9.
[22] Gao R, Grauman K. On-Demand Learning for Deep Image Restoration. arXiv. 2017:1612.01380v3.
[23] Mao X-J, Shen AC, Yang AY-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Proceedings of the 30th International Conference on Neural Information Processing Systems. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Spain: Curran Associates Inc.; 2016. p. 2810–8.
[24] Dobashi S, Mori S. Evaluation of respiratory pattern during respiratory-gated radiotherapy. Australas Phys Eng Sci Med 2014;37:731–42.
[25] Jain V, Seung S. Natural image denoising with convolutional networks. Adv Neural Inform Process Syst 2009:769–76.
[26] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans Image Process 2017;99:1.
[27] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning. 2008:pp. 1096–1103.
[28] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer; 2015. p. 234–41.
[29] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. International Conference on Learning Representations 2016.
[30] Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. pp. 1874–1883.
[31] He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE; 2015. pp. 1026–1034.
[32] Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press; 1995.
[33] Lin M, Chen Q, Yan S. Network In Network. International Conference on Learning Representations2013.
[34] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. Orlando, Florida, USA: ACM; 2014. p. 675-8.
[35] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13:600–12.
[36] Pathak D, Kra¨henbu¨hl P, Donahue J, Darrell T, E AA. Context Encoders: Feature Learning by Inpainting. arXiv. 2016:1604.07379.
[37] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. arXiv. 2014:1406.2661.
[38] Tomczak JM, Welling M. VAE with a VampPrior. arXiv. 2017:1705.07120v3.
[39] [Application of navigation system ExacTrac in radiation therapy of a female patient with disseminated pineoblastoma]. Zh Vopr Neirokhir Im N N Burdenko. 2010:pp. 43–46.
[40] Mori S, Shirai T, Takei Y, Furukawa T, Inaniwa T, Matsuzaki Y, et al. Patient handling system for carbon ion beam scanning therapy. J Appl Clin Med Phys 2012;13:3926.