Deep Neural Network-based Synthetic Image Digital Fluoroscopy Using Digitally Reconstructed Tomography

Shinichiro Mori (✉ shinichiromph@gmail.com)

  National Institutes for Quantum and Technology Quantum Medical Science Directorate Institute for Quantum Medical Science: Kokuritsu Kenkyu Kaihatsu Hojin Ryoshi Kagaku Gijutsu Kenkyu Kaihatsu Kiko Ryoshi Seimei Igaku Bumon Ryoshi Ikagaku Kenkyujo

Ryusuke Hirai

  Toshiba

Yukinobu Sakata

  Toshiba

Yasuhiko Tachibana

  National Institutes for Quantum and Technology Quantum Medical Science Directorate Institute for Quantum Medical Science: Kokuritsu Kenkyu Kaihatsu Hojin Ryoshi Kagaku Gijutsu Kenkyu Kaihatsu Kiko Ryoshi Seimei Igaku Bumon Ryoshi Ikagaku Kenkyujo

Masashi Koto

  National Institutes for Quantum and Technology Quantum Medical Science Directorate Institute for Quantum Medical Science: Kokuritsu Kenkyu Kaihatsu Hojin Ryoshi Kagaku Gijutsu Kenkyu Kaihatsu Kiko Ryoshi Seimei Igaku Bumon Ryoshi Ikagaku Kenkyujo

Hitoshi Ishikawa

  National Institutes for Quantum and Technology Quantum Medical Science Directorate Institute for Quantum Medical Science: Kokuritsu Kenkyu Kaihatsu Hojin Ryoshi Kagaku Gijutsu Kenkyu Kaihatsu Kiko Ryoshi Seimei Igaku Bumon Ryoshi Ikagaku Kenkyujo

Research Article

Abstract

We developed a deep neural network (DNN) to generate X-ray flat panel detector (FPD) images from digitally reconstructed radiographic (DRR) images.

FPD and treatment planning CT images were acquired from patients with prostate and head and neck (H&N) malignancies. The DNN parameters were optimized for FPD image) synthesis. The synthetic FPD images' features were evaluated to compare to the corresponding ground-truth FPD images using mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). The image quality of the synthetic FPD image was also compared with that of the DRR image to understand the performance of our DNN.

For the prostate cases, the MAE of the synthetic FPD image was improved (= 0.12 ± 0.02) from that of the input DRR image (= 0.35 ± 0.08). The synthetic FPD image showed higher PSNRs (= 16.81 ± 1.54 dB) than those of the DRR image (= 8.74 ± 1.56 dB), while SSIMs for both images (= 0.69) were almost the same. All metrics for the synthetic FPD images of the H&N cases were improved (MAE 0.08 ± 0.03, PSNR 19.40 ± 2.83 dB, and SSIM 0.80 ± 0.04) compared to those for the DRR image (MAE 0.48 ± 0.11, PSNR 5.74 ± 1.63 dB, and SSIM 0.52 ± 0.09).

Our DNN successfully generated FPD images from DRR images. This technique would be useful to increase throughput when images from two different modalities are compared by visual inspection.

Introduction

With recent advances in deep neural network (DNN) techniques, research and development in medical image processing has accelerated such that it now exceeds the performance of conventional image processing [1−3]. Their benefits have been applied to image-guided radiation therapy (IGRT) and treatment planning, auto-segmentation [4, 5], automated planning [6], planning CT image synthesis from cone-beam CT [7], deformable image registration [8], generating 4DCT images [9], and markerless tumor tracking [10, 11]. These techniques were developed to improve treatment accuracy and throughput in treatment planning. Patient positional verification, however, takes up the largest share of treatment room time [12].

Commercially available automatic image registration software speeds up patient positional verification [13−15]. The 2D-3D image registration software that registers X-ray flat panel detector (FPD) images to the reference digitally reconstructed radiography (DRR) images still takes time:10−20 sec to calculate, but 2−5 min to verify position. Different qualities of different image modalities may make visual comparison difficult.

One solution to this problem is to make both images in the same image modality, which is known as image synthesis. There are two types of image synthesis technique: intra-modality and inter-modality. The intra-modality image synthetic technique transforms an image acquired in one modality to another

form of the same modality, e.g. CT to low-dose CT [16, 17], MRI T1-weighted images to MRI T2-weighted images [18, 19], and PET [20, 21]. The inter-modality image synthetic technique transforms images from one modality to another: MR to CT [22–25], CT to MR [26, 27], and PET to CT [28–30]. Compared to image denoising and increasing image resolution [1, 2, 31], the major difficulties with image synthesis are large differences in pixel values, differences in visualized structures, and alignment error.

There have been no instances in which an image synthesis DNN has been used to generate an FPD image from a DRR image. With this purpose in mind, we developed an image synthesis DNN for FPD image data, and compared the quality of the synthetic FPD image data with those of the original FPD and DRR images of the pelvis and head and neck (H&N).

Materials and Methods

Patients and Image Acquisition

A total of 200 and 70 cases with tumors of the prostate or H&N undergoing carbon-ion beam scanning therapy (C-PBS) at our treatment center participated in this study, respectively. The study was conducted with the approval of the Institutional Review Board (N21-001) and performed in accordance with the Declaration of Helsinki. All the patients provided informed consent for use the data from their medical records. During image acquisition, all patients were positioned on the treatment table with immobilization devices (urethane resin cushion [Moldcare, Alcare, Tokyo, Japan]) and low-temperature thermoplastic shells (Shell Fitter, Kuraray Co., Ltd., Osaka, Japan).

Planning CT image and Projecting DRR image

Treatment planning CT image data were acquired under breath-hold in exhalation using a 320-detector CT (Aquilion One Vision, Canon Medical Systems, Otawara, Japan). Imaging conditions were based on our clinical protocols using automatic exposure control [32]. Reconstructed CT slice thicknesses were 2.0 mm for the prostate and 1.0 mm for H&N cases. Image field-of-view was 500 mm for both diseases.

A pair of DRR images was generated by projecting the CT data (converted to X-ray attenuation coefficients) along the X-ray imaging beam path using our in-house software [33]:

$$q\left(\mathrm{x}, \mathrm{y}\right) = \sum_{\mathrm{k}=1}^{\mathrm{n}} \Delta\mathrm{L} \bullet \mu_{\mathrm{k}}$$

1

where q(x, y) is the projection ray sum point on the DRR image position (x, y) and ΔL is the calculation grid size (= 1 mm in this study).

The CT image was shifted to position the tumor in the center of the DRR image. For some cases, the edge of the DRR image could not include the CT image completely due to the small number of CT slices, degrading DRR image quality. To solve this, we added additional CT slices before the first. This process was performed to the last CT slice (extended CT image region). The DRR image matrix size and pixel size were 768 × 768 pixels and 388 × 388 μm, respectively, the same dimensions as the FPD images. The DRR computation was programmed using commercial software (Compute Unified Device Architecture [CUDA] ver. 10.1, Microsoft Visual Studio 2013, Microsoft Corp, Redmond WA, USA) in a Windows 10 environment with a GPU (graphics processing unit) processor on an NVIDIA board (QuadroRTX 8000, NVIDIA Corporation, Santa Clara CA, USA), which is equipped with 4068 CUDA core units and 48 GB of memory, allowing a processing speed of more than 16.3 Tflops for a single precision calculation [34].

Fluoroscopic images

Digital fluoroscopic images from the prostate and H&N cases were acquired by imaging systems installed in the treatment room [12]. The X-ray imaging systems were set up according to the method of Mori [35]. The distance from the X-ray tube and the FPD was 239 cm, and from the X-ray tube to the room isocenter was 169 cm.

For the patient setup verification process, we performed 2D-3D image registration of the pair of FPD images and the planning CT data [35], coregistering patient anatomical structures on FPD images to those on the DRR images within a mean of 0.87 mm and 0.61 degrees expressed as the square root sum of squares for the respective three dimensions [36]. The prostate treatment protocol uses 90° and 270° beam angles. The H&N treatment protocol uses more than two beam angles by rotating the treatment couch around its long axis (ϕ: standard International Electrotechnical Commission [IEC] tabletop rolling angle) to extend the range of angles (− 10° to + 10°). The number of treatment fractions was 12 and 16 for prostate and H&N, respectively.

Network architecture

Our DNN was a modified 2D convolutional autoencoder with shortcut connections (U-net) [37] (Fig. 1). This DNN consisted of an "encoder block" and "decoder block." The encoder block extracts features representing the input data with reduced spatial dimensions via a combination of the following hidden layers: a convolutional layer (stride size of 1 × 1), a rectified linear unit (ReLU) layer, and an instance normalization layer [38]. Spatial dimensions were reduced by using a convolutional layer with a stride size of 2 × 2. The number of output channels after spatial dimensional reduction was doubled (64, 128, 256, 512, and 1024).

The decoder block reconstructed feature representations in the input data and included a combination of the upsampling layers, instance normalization layers, ReLU layers, dropout layers [39], and convolutional layers. The upsampling layers doubled the number of spatial dimensions. Subsequently, the number of

output channels was decreased by half. A dropout layer (rate = 0.2) was added to avoid overfitting. The convolutional layer with a of 1 × 1 and a single channel were added to export a single grayscale image for clinical use. The convolutional kernel size for all layers was 3 × 3 except for the last one (1 × 1). The shortcut connection is one solution to avoiding a vanishing gradient in the deep architecture [40–42]. It is applied before the dropout layer in the decoder block from the ReLU layer before the convolutional layer with the stride size of 2 × 2; however, it was not applied in the last convolutional layer with the stride size of 2 × 2 [43], because the input DRR image and the ground truth FPD image were not perfectly registered due to interfractional positional changes (misalignment).

Although original U-net uses pooling and batch normalization layers [40], we replaced U-net with a convolutional layer (stride size of 2 × 2) to instance the normalization layer to transfer the image style of the ground-truth image (FPD image) onto the output image (synthetic FPD image).

Network training

A total of 4000 and 2000 image pairs (DRR and FPD images) from prostate and H&N cases were randomly selected for the DNN training process, respectively. In this process, a pair of FPD and DRR images was subdivided into subimages (144 × 144 pixels) by changing position, rotation angle (± 3.0, 0.1–degree step, rotated by ± 90° and 180°), and flipping in the left-right or up-down directions. All FPD and DRR images were resized to 384 × 384 pixels with bicubic interpolation and normalized pixel values to the range of 0–1. A total of 50,000 and 10,000 subimage pairs were prepared for the prostate and H&N cases, respectively, with care to exclude subimages containing the irradiation port cover, air or bowel gas.

Treatment couch

The edge of the treatment couch included on the FPD image increased pixel values (marked as light green lines in Figs. 1a and 1c). The treatment couch positions on the FPD and DRR did not always match due to small differences in patient positioning. In the worst-case scenario, the treatment couch edge may have been absent from some images, resulting in large pixel value inconsistencies on the input and ground-truth images. To avoid this, we applied image processing to remove the treatment couch edge from the CT image, and then calculated the DRR image.

Irradiation port cover

The edge of the irradiation port cover was visualized on the FPD image (marked as arrows in Figs. 1a and 1c) but not on the DRR image.

Bowel gas

For the prostate cases, bowel gas positions differed on the FPD and DRR images, possibly due to interfractional changes or pixel value inconsistency. To avoid this, we outlined bowel gas regions of interest (ROIs) manually on the FPD images (marked with dotted yellow lines in Fig. 2a), and the DRR

image without the bowel gas was corrected by resetting the Hounsfield Units (HU) in the gas collections to 0. The upper thighs/male genitalia were variable in contour; as a result, the pixel values were sometimes inconsistent between FPD and DRR images (marked as light blue dashed lines in Fig. 2a).

Air

In the H&N cases, regions external to the patient included the treatment couch edge and/or air on the FPD images. Pixel values for air inconsistently measured showed zero on DRR images, which could affect DNN prediction accuracy. To solve this problem, a mask was applied to the DRR image using a pixel value threshold of 0 (marked as a light blue line in Fig. 2d). This patient mask was adapted to another FPD image of the pair of images at the same position (marked as a light blue dotted line in Fig. 2c). The proportion of air was kept at < 40% of the subimage.

Parameter optimization

The DNN parameters were optimized to predict an FPD image from a DRR image as follows: the optimization process was performed for 3000 epochs with a batch size of 70 using stochastic gradient descent (SGD) to minimize loss. We did not set early stopping criteria; however, the learning curves for the training and validation data were checked by plotting loss values for the training data and validation data. When the learning curves did not improve, or showed overfitting, we stopped the optimization process.

We calculated two types of loss: content loss and perceptual loss (Fig. 1c).

Content loss was calculated using the selected mean absolute error (MAE) because L1 loss function MAE can improve the robustness of outliers (image noise and image artifact) due to misalignment between DRR and FPD images

$$L_{content} = \frac{1}{n} \sum_{i=1}^{n} \left| I_i^{true} - I_i^{pred} \right| \quad (2),$$

where I_i^{true} and I_i^{pred}, are the ith pixel value in the ground-truth image and predicted image, respectively, and n is the total number of pixels in the image.

Perceptual loss was assessed using a pre-trained VGG19 model [44] as a feature extractor. It calculated the sum of six features of output from the respective convolutional layers (Fig. 1b). The error was computed as the mean square error of these features using the input DRR image and the predicted FPD image.

Perceptual loss ($L_{perceptual}$) was defined as:

$$L_{perceptual} = \sum_{k}^{P} \left[\sum_{i=1}^{n} \left(V_k \left(I_i^{true} \right) - V_k \left(I_i^{pred} \right) \right)^2 \right], (3)$$

$V_k(I^{true})$ and $V_k(I^{pred})$ were feature of kth layer in a VGG network when a ground-truth image and a predicted image, respectively. P was 1, 2, 5, 10, 15 and 20.

Finally, we calculated the total loss using the following equation:

$$L_{total} = 2.0 \bullet L_{content} + 10^{-5} \bullet L_{perceptual}, (4)$$

The learning rate, momentum, and decay were set to 10^{-5}, 0.9 and 10^{-5}, respectively. The learning rate was decreased 2×10^{-9} for every 4 epochs. The deep learning framework "TensorFlow 2.4" was used in Windows 10, 64-bit environment with a single GPU on the NVIDIA QuadroRTX 8000 board.

Post-processing

The ground-truth FPD image showed image noise from scatter. The predicted FPD image, however, did not fully reflect scattered radiation. Thus, we added image noise to the predicted FPD image.

The ground-truth FPD detected scattered radiation in the original image size (= 768 × 768); however, all training data were resized by half (= 384 × 384). The predicted FPD image was resized to 768 × 768, and Gaussian image noise was added (mean: 0, deviation: 0.001 and 0.0005 for prostate and H&N cases, respectively). Finally, the synthetic FPD image was decreased in size with image noise to 384 × 384 again.

Evaluations

We evaluated the quality of the synthetic FPD images using 2700 and 640 FPD (ground-truth) images for the prostate and H&N cases, respectively. These image data differed from the training data. The synthetic FPD images were compared with the ground-truth FPD image using the MAE, the peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) [45]. These metrics are widely used to quantify the similarity between two images. We also compared the image quality of the synthetic FPD image with that of the DRR image to understand the performance of our DNN.

The computation time for the prediction (not including the model file import) was evaluated.

Results

Pelvis

Large collections of bowel gas and the irradiation port cover edge were visualized on the ground-truth FPD image (marked as yellow and red arrows, respectively, in Fig. 3a). These regions were not visible in

the same positions on the DRR image (Fig. 3b). The same anatomical shape was extended in the CT slice direction (superior to inferior) because of the extended CT image region (marked as a blue dotted rectangle in Figs. 3b and 3d). The quality of the input DRR image compared to the ground-truth FPD image was expressed by the metrics MAE = 0.35, PSNR = 8.43 dB, and SSIM = 0.67. The synthetic FPD image was close to the ground-truth FPD image (Fig. 3c). The metrics improved to 0.13, 16.05 dB, and 0.64, respectively. Since the port edge was not contained in the input DRR image, it was not visualized on the synthetic FPD image. To clearly understand the image differences between the ground-truth FPD and the synthetic FPD images, we subtracted the synthetic FPD image from the ground-truth FPD image (Fig. 3d). Bowel gas positional variations caused large pixel value differences.

Results for image quality averaged over all prostate cases are summarized in Fig. 5 and Table 1. The box ranges in between the 25th and 75th percentiles for the synthetic FPD image were lower than those for the DRR image for all image quality metrics (Figs. 5a-5c). MAE for the synthetic FPD image was improved (= 0.12 ± 0.02) from that for the input DRR image (= 0.35 ± 0.08) (Fig. 5a). The synthetic FPD image showed a higher PSNR value (= 16.81 ± 1.54 dB) than the DRR image (= 8.74 ± 1.56 dB) (Fig. 5b). Although SSIM values for both images were almost the same (= 0.69), the number of outliers for the synthetic FPD image was smaller than that for the DRR image (Fig. 5c). Computation time was 89.6 ± 11.0 [msec].

		Pelvis			Head and neck		
		Mean	SD	95% percentile	Mean	SD	95% percentile
MAE	Synthetic FPD	0.12	0.02	0.17	0.08	0.03	0.13
	DRR	0.35	0.08	0.48	0.48	0.11	0.64
PSNR (dB)	Synthetic FPD	16.81	1.54	19.17	19.40	2.83	24.28
	DRR	8.74	1.56	11.41	5.74	1.63	8.72
SSIM	Synthetic FPD	0.69	0.03	0.74	0.80	0.04	0.85
	DRR	0.69	0.05	0.76	0.52	0.09	0.68
Time [msec]		89.6	11.0	108.9	90.3	10.6	109.3

Abbreviations: SD = standard deviation; DRR = digitally reconstructed radiography; FPD = flat panel detector; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.

Abbreviations: DRR = digitally reconstructed radiography; FPD = flat panel detector, H&N = head and neck.

Abbreviations: DRR = digitally reconstructed radiography; DNN = deep neural network; FPD = flat panel detector; ReLU = rectified linear units.

Abbreviations: DNN = deep neural network; DRR = digitally reconstructed radiography; FPD = flat panel detector; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.

Abbreviations: DNN = deep neural network; DRR = digitally reconstructed radiography; FPD = flat panel detector; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.

Abbreviations: DRR = digitally reconstructed radiography; FPD = flat panel detector; H&N = head and neck; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.

Head and neck region

For the H&N region, the ground-truth FPD and input DRR image are shown in Figs. 4a and 4b, respectively. The irradiation port cover edge and earlobe were observed in a horizontal direction (marked with red and blue arrows in Fig. 4a, respectively); however, these were not included in the input DRR image (Fig. 4b). The skull curvature emphasized the range of thickness depending on the angle of the incident beam to

the skull surface shape (marked as red arrows in Fig. 4b). Quality metrics of the input DRR image compared to the ground-truth FPD image were 0.60, 4.25 dB, and 0.38 for MAE, PSNR and SSIM, respectively. However, the quality of the synthetic FPD image was much closer to the ground-truth FPD image by visual inspection. Image quality metrics were also improved (MAE 0.04, PSNR 25.93 dB and SSIM 0.86) compared to those for the input DRR image (Fig. 4c). The irradiation port cover edge and earlobe were not visualized on the synthetic FPD image. The smoothness of the skull curvature was improved to the same degree as that of the ground-truth FPD image. Figure 4d shows image differences between the input DRR and synthetic FPD images. Relatively large pixel value differences were attributable to misalignment (red arrow in Fig. 4d).

Among all H&N cases, all metrics for the synthetic FPD image were improved (MAE 0.08 ± 0.03, PSNR 19.40 ± 2.83 dB and SSIM 0.80 ± 0.04) compared to those for the DRR image (MAE 0.48 ± 0.11, PSNR 5.74 ± 1.63 dB and SSIM, 0.52 ± 0.09) (Table 1). Computation time was 90.3 ± 10.6 msec. The box ranges between the 25th and 75th percentiles for the synthetic FPD image were smaller than those for the DRR image, except PSNR (Figs. 5d-5f).

DISCUSSION

We developed DNN for synthetic FPD images from DRR images and evaluated image quality between the synthetic FPD and original FPD using pelvic and H&N images. The quality of the synthetic FPD images was close to that of the original FPD images. Computation time for the prediction was approximately 90 msec per image on average.

Image quality

We evaluated quality of the synthetic FPD images using MAE, PSNR, and SSIM. Image quality metrics of the synthesized FPD image were improved compared to those of the DRR image for both pelvic and H&N regions, except the SSIM for the pelvic region. The SSIM values for the synthetic FPD image and the DRR image were roughly the same (= 0.69); bowel gas positional change between images might be a major cause of this. Objects that did not visualize on the DRR image were not visualized on the synthetic FPD image. For example, the earlobe was not visualized on the synthetic FPD image because it was not visualized on the input DRR image (Fig. 4). If an invisible object on the input DRR image, which was visualized on the ground-truth FPD image, was visualized on the synthetic FPD image, the DNN performed well on the training data but did not perform accurately in other images (overfitting). To avoid this problem, we checked the learning curve and predicted images during DNN during the network training process.

However, a DRR image should be created to visualize objects, which are visualized on the ground-truth FPD image. By doing this, the quality of the synthetic FPD image would be closer to that of the ground-truth FPD image.

Interfractional variation

We used DRR and ground-truth FPD images for the DNN training process because we wanted to generate realistic image quality using the DNN. Since planning CT and ground-truth FPD images were acquired on different days, organ and bowel gas positions were interfractionally different. We excluded image regions affecting interfractional changes in the training subimages; however, these might not exclude differences completely, especially with moving bowel gas. As long as original FPD images are used for the training data, this problem will not be completely resolved. One approach to solving this is to calculate a mimic FPD image by Monte Carlo simulation-based DRR calculation so that the quality of the mimic FPD image would be close to that of the original FPD image. This mimic FPD image would not include interfractional and intrafractional anatomical changes and misalignment in between the mimic FPD and the planning CT image. This approach could therefore be applied to the thoracoabdominal region. The quality of the mimic FPD image was not completely identical to that of the original FPD image; it is questionable whether it is possible to obtain sufficient image quality of a synthetic FPD image trained by using the mimic FPD images.

Loss function

The DRR image had lower spatial resolution and less image noise than the FPD image. It is more difficult to predict the synthetic FPD image from DRR image compared with the synthetic DRR image from the FPD image. Generally, L1 loss and L2 loss (such as MAE and mean square error [MSE]) were used as loss functions for image synthesis. However, use of these loss functions evaluated distortion only, and could therefore predict blurred images similar to the DRR image. To predict enough information and minute structures included on the original FPD image, it is necessary to evaluate image quality using both distortion and perceptual quality [46]. We used perceptual loss and content loss to evaluate quality and distortion, respectively. Weight factors for the respective loss functions were manually adjusted in this study (Eq. 3), although when more than two hyper-parameters were used, it was arduous to determine them manually. To solve this problem, it would be useful to use a single dynamic loss function combined with several loss functions, such as a generative adversarial network (GAN). GAN is also widely used for image synthesis and U-net [28, 47]. Although we did not use GAN in this study, GAN is able to generate a synthetic FPD image.

Applications to other IGRTs

Our group developed machine learning-based markerless tumor tracking software and integrated it into clinical applications [10, 11, 48]. The machine learning training process was performed using DRR images. The markerless tumor tracking detected tumor position on the FPD images in real time. Use of different image modalities in the training process and the prediction process makes it difficult to detect

tumor position. If the markerless tracking training process used the synthetic FPD image data instead of the DRR image data, tracking detection accuracy could be improved.

Conclusion

Our DNN successfully generated an FPD image from a DRR image. Its quality was close to that of the ground-truth FPD image. This technique would be useful to increase throughput when images from two different modalities are compared by visual inspection in IGRT.

Declarations

References

1. 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. Journal of Machine Learning Research. 2010;11:3371-408.

2. Jonathan M, Ueli M, Dan C, Jürgen S. Stacked convolutional auto-encoders for hierarchical feature extraction. International Conference on Artificial Neural Networks. Heidelberg: Springer Berlin; 2011. p. 52-9.

3. 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.

4. Dong X, Lei Y, Wang TH, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Medical Physics. 2019;46:2157-68.

5. Wang J, Lu J, Qin G, Shen L, Sun Y, Ying H, et al. Technical Note: A deep learning-based autosegmentation of rectal tumors in MR images. Med Phys. 2018;45:2560-4.

6. Shen CY, Nguyen D, Chen LY, Gonzalez Y, McBeth R, Qin N, et al. Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer

intensity-modulated radiation therapy treatment planning. Medical Physics. 2020;47:2329-36.

7. Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Phys Med Biol. 2019;64:125002.

8. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Isgum I. A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal. 2019;52:128-43.

9. Mori S, Hirai R, Sakata Y. Simulated four-dimensional CT for markerless tumor tracking using a deep learning network with multi-task learning. Phys Med. 2020;80:151-8.

10. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. Phys Medica. 2019;59:22-9.

11. Takahashi W, Oshikawa S, Mori S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. Br J Radiol. 2020;93:20190420.

12. 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.

13. Mori S, Kumagai M, Miki K, Fukuhara R, Haneishi H. Development of fast patient position verification software using 2D-3D image registration and its clinical experience. J Radiat Res. 2015;56:818-29.

14. Chang Z, Wang Z, Ma J, O'Daniel JC, Kirkpatrick J, Yin FF. 6D image guidance for spinal non-invasive stereotactic body radiation therapy: Comparison between ExacTrac X-ray 6D with kilo-voltage cone-beam CT. Radiother Oncol. 2010;95:116-21.

15. Penney GP, Weese J, Little JA, Desmedt P, Hill DL, Hawkes DJ. A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Trans Med Imaging. 1998;17:586-95.

16. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Medical Physics. 2017;44:e360-e75.

17. Kyong Hwan J, McCann MT, Froustey E, Unser M. Deep Convolutional Neural Network for Inverse Problems in Imaging. IEEE Trans Image Process. 2017;26:4509-22.

18. Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging. 2018;48:330-40.

19. Chartsias A, Joyce T, Giuffrida MV, Tsaftaris SA. Multimodal MR Synthesis via Modality-Invariant Latent Representation. IEEE Trans Med Imaging. 2018;37:803-14.

20. Dong X, Lei Y, Wang T, Higgins K, Liu T, Curran WJ, et al. Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. Phys Med Biol. 2020;65:055011.

21. Yang J, Park D, Gullberg GT, Seo Y. Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain F-18-FDG PET. Physics in Medicine and Biology. 2019;64.

22. Khoo VS, Joon DL. New developments in MRI for target volume delineation in radiotherapy. Br J Radiol. 2006;79 Spec No 1:S2-15.

23. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, et al. Medical Image Synthesis with Deep Convolutional Adversarial Networks. IEEE Trans Biomed Eng. 2018;65:2720-30.

24. Arabi H, Dowling JA, Burgos N, Han X, Greer PB, Koutsouvelis N, et al. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region. Med Phys. 2018;45:5218-33.

25. Shafai-Erfani G, Lei Y, Liu Y, Wang Y, Wang T, Zhong J, et al. MRI-Based Proton Treatment Planning for Base of Skull Tumors. Int J Part Ther. 2019;6:12-25.

26. McKenzie EM, Santhanam A, Ruan D, O'Connor D, Cao M, Sheng K. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. Med Phys. 2020;47:1094-104.

27. Lei Y, Dong X, Tian Z, Liu Y, Tian S, Wang T, et al. CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. Med Phys. 2020;47:530-40.

28. Dong X, Wang T, Lei Y, Higgins K, Liu T, Curran WJ, et al. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. Phys Med Biol. 2019;64:215016.

29. Landry G, Hansen D, Kamp F, Li M, Hoyle B, Weller J, et al. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. Phys Med Biol. 2019;64:035011.

30. Yuan N, Dyer B, Rao S, Chen Q, Benedict S, Shang L, et al. Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy. Phys Med Biol. 2020;65:035003.

31. Kim J, Lee JK, Lee KM. Accurateimagesuper-resolutionusing very deep convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition2016.

32. Papadakis AE, Perisinakis K, Oikonomou I, Damilakis J. Automatic exposure control in pediatric and adult computed tomography examinations: can we estimate organ and effective dose from mean MAS reduction? Invest Radiol. 2011;46:654-62.

33. Mori S, Inaniwa T, Kumagai M, Kuwae T, Matsuzaki Y, Furukawa T, et al. Development of digital reconstructed radiography software at new treatment facility for carbon-ion beam scanning of National Institute of Radiological Sciences. Australas Phys Eng Sci Med. 2012;35:221-9.

34. Vasiliadis G, Antonatos S, Polychronakis M, Markatos E, Ioannidis S. Gnort: High Performance Network Intrusion Detection Using Graphics Processors. Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection (RAID). 2008:116-34.

35. Mori S, Kumagai M, Miki K, Fukuhara R, Haneishi H. Development of fast patient position verification software using 2D-3D image registration and its clinical experience. J Radiat Res. 2015;56:818-29.

36. Mori S, Shibayama K, Tanimoto K, Kumagai M, Matsuzaki Y, Furukawa T, et al. First clinical experience in carbon ion scanning beam therapy: retrospective analysis of patient positional accuracy. J Radiat Res. 2012;53:760-8.

37. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39:2481-95.

38. Ulyanov D, Vedaldi A, Lempitsky VS. Instance Normalization: The Missing Ingredient for Fast Stylization. ArXiv. 2016;abs/1607.08022.

39. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov N. Dropout: A simple way to prevent neural networks from overfitting Journal of Machine Learning Research. 2014;15:1929-58.

40. Sergey I, Christian S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. The 32nd International Conference on Machine Learning. 2015:448-56.

41. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany: Springer; 2015. p. 234-41.

42. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv:150806576. 2015.

43. Hwang D, Kim KY, Kang SK, Seo S, Paeng JC, Lee DS, et al. Improving the Accuracy of Simultaneously Reconstructed Activity and Attenuation Maps Using Deep Learning. J Nucl Med. 2018;59:1624-9.

44. Simonya K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference for Learning Representations2015.

45. 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.

46. Yochai B, Tomer M. The Perception-Distortion Tradeoff. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018:6228-37.

47. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Med Phys. 2018.

48. Sakata Y, Hirai R, Kobuna K, Tanizawa A, Mori S. A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy. Phys Med Biol. 2020;65:085014.
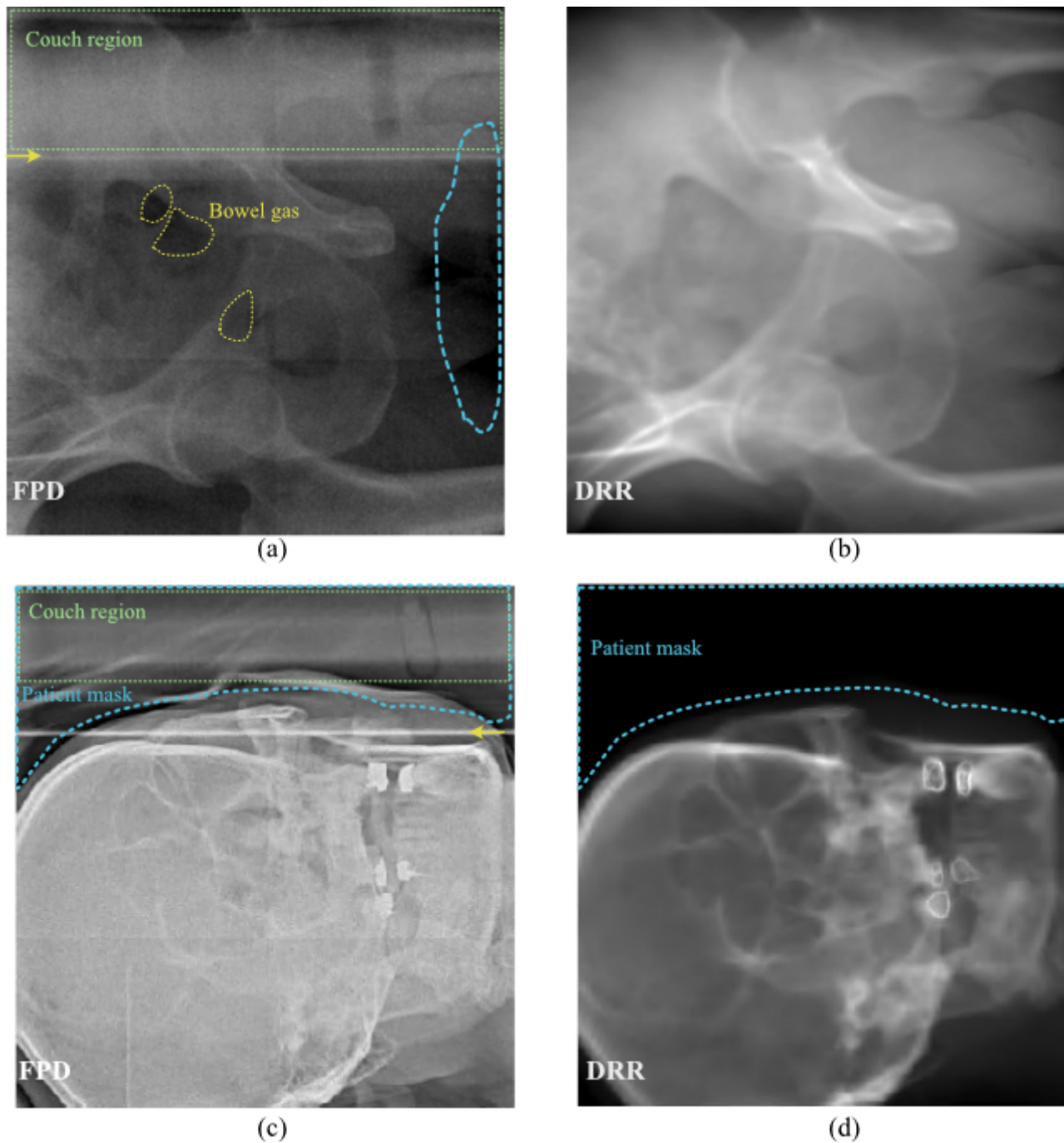
Figures

Figure 1

Excluded image regions on the training data for pelvic and H&N cases.

(a) FPD and (b) DRR images for the pelvic region. The yellow dotted and the light blue dashed lines show bowel gas, and the upper thighs/genitalia, respectively. (c) FPD and (d) DRR images for the H&N region. The light blue dotted line shows the patient mask region. Yellow arrows and light green dotted lines show the irradiation port cover edge and treatment couch edge, respectively.

Abbreviations: DRR = digitally reconstructed radiography; FPD = flat panel detector, H&N = head and neck.
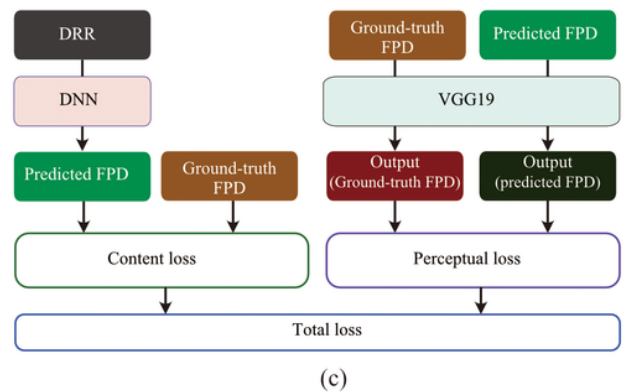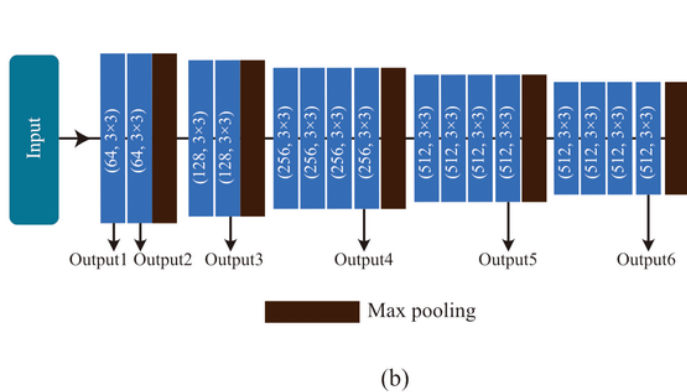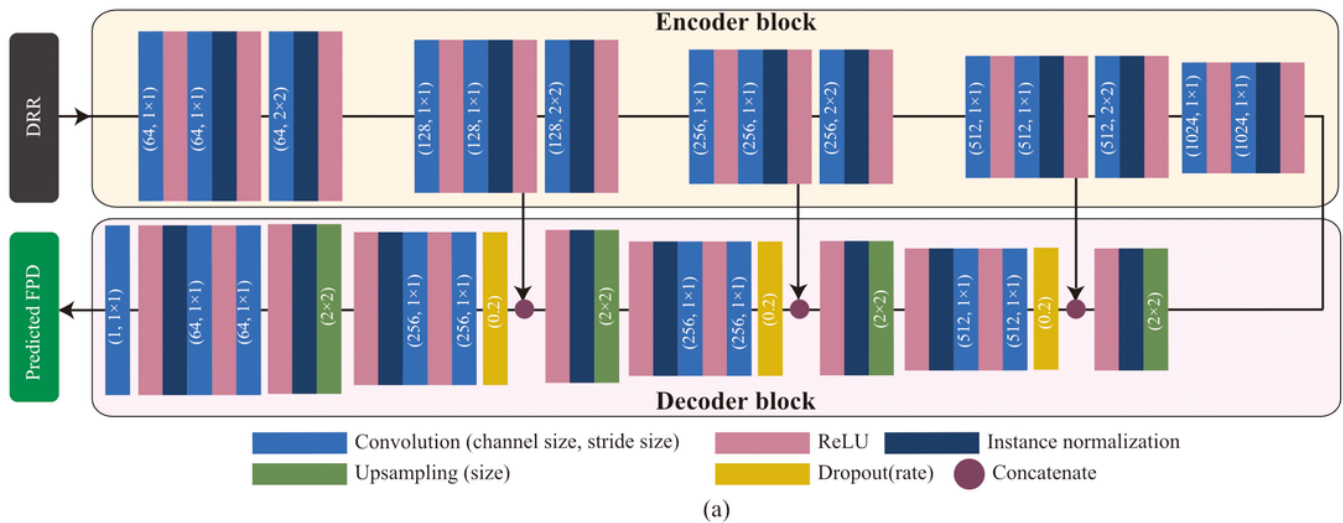
Figure 2

(a) Network structure for the DNN. (b) VGG19 network structure for the perceptual loss. (c) Loss calculation using content loss and perceptual loss.

Abbreviations: DRR = digitally reconstructed radiography; DNN = deep neural network; FPD = flat panel detector; ReLU = rectified linear units.
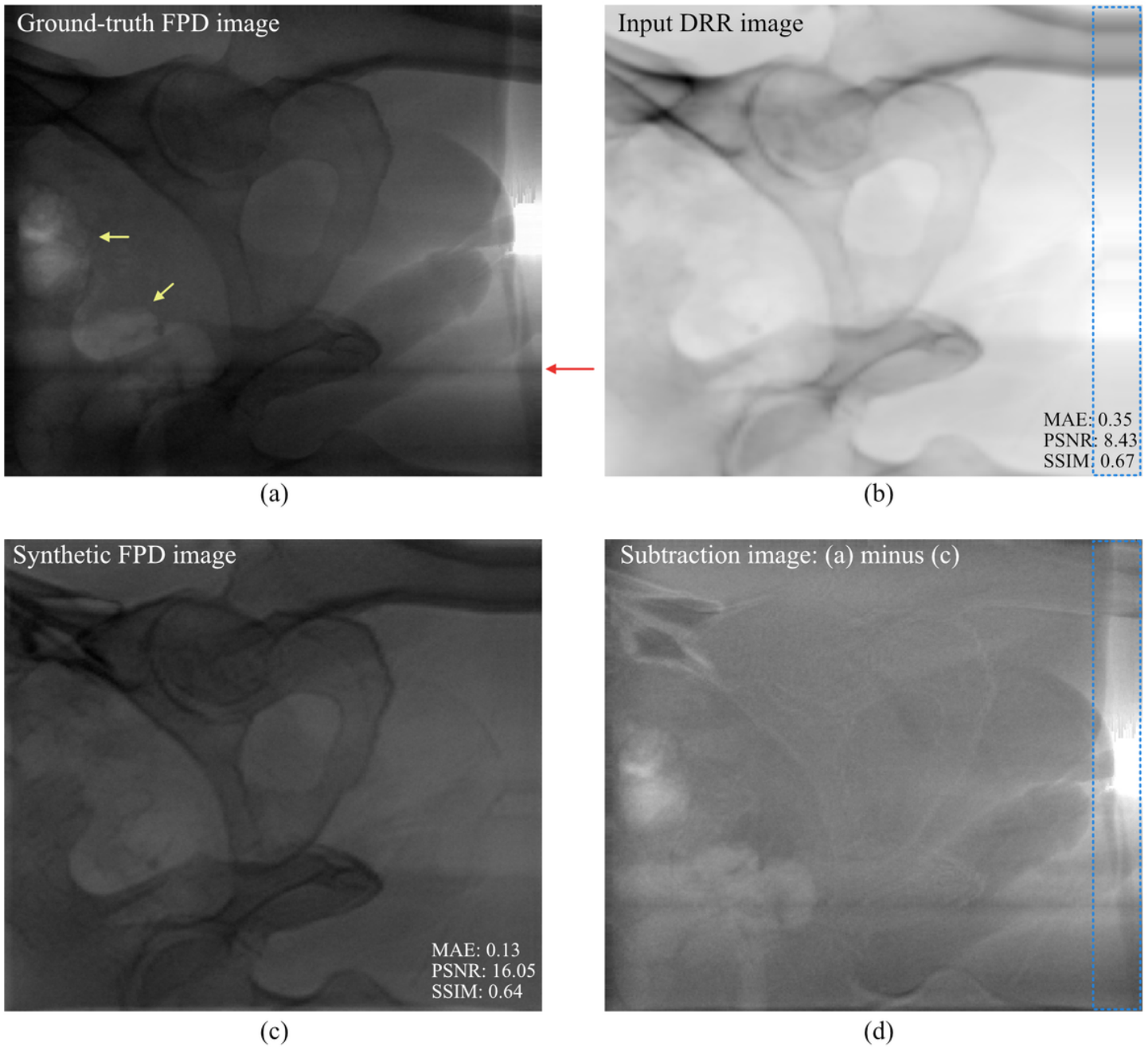
Figure 3

(a) Ground-truth FPD image. Yellow arrows show bowel gas collections. Red arrow shows the irradiation port cover edge. (b) Input DRR image. The blue dotted rectangle shows an extended image, generated by adding additional CT images after the last slice. (c) Synthetic FPD image with the DNN. (d) Image differences ([a] minus [c]).

Abbreviations: DNN = deep neural network; DRR = digitally reconstructed radiography; FPD = flat panel detector; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.
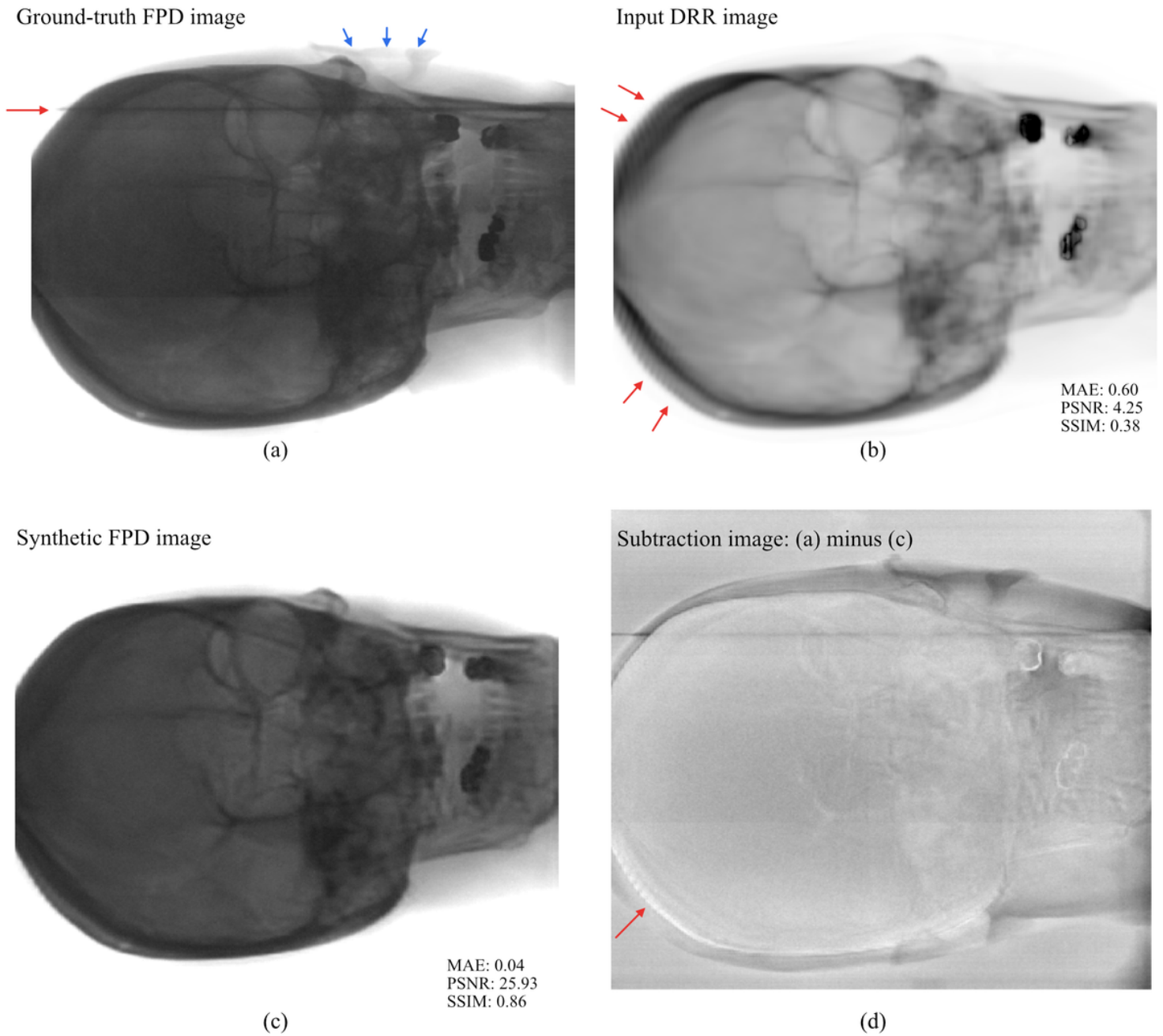
Figure 4

(a) Ground-truth FPD image. The red arrow shows the irradiation port cover edge. The blue arrows point to the earlobe. (b) Input DRR image. (c) Synthetic FPD image with the DNN. (d) Image differences ([a] minus [c]). The red arrow shows an area of misalignment.

Abbreviations: DNN = deep neural network; DRR = digitally reconstructed radiography; FPD = flat panel detector; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.
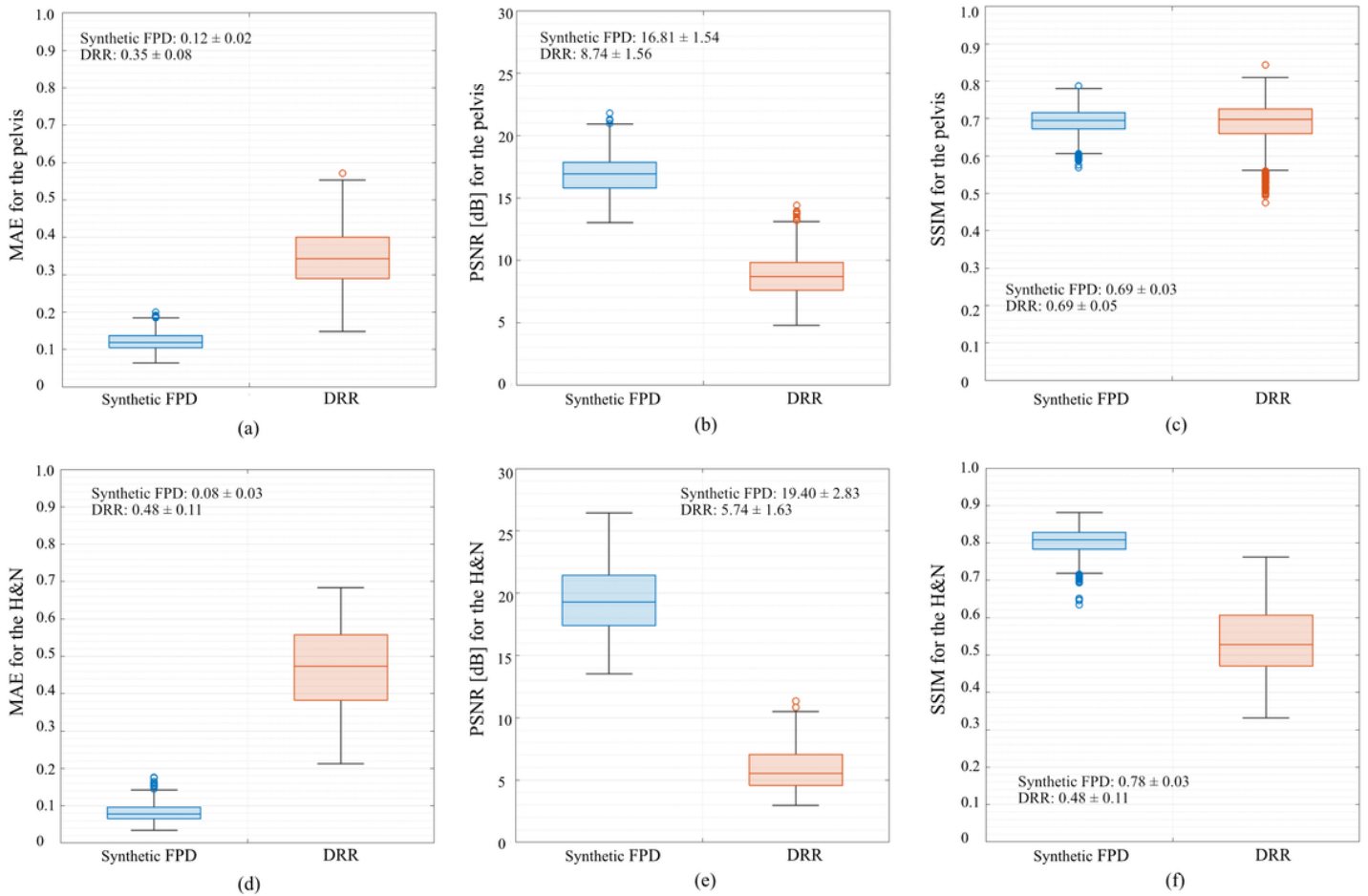
Figure 5

Image quality metrics for the synthetic FPD and DRR with the ground-truth FPD images. (a) MAE, (b) PSNR and (c) SSIM for pelvis. (d) MAE, and (e) PSNR and (f) SSIM for the H&N.

The horizontal line in the center of the box indicates the median, with the box's bottom and top edges indicating the 25th ($q1$) and 75th percentiles ($q3$), respectively. The whiskers extend to the most extreme data points that are not identified as outliers. An outlier (light blue open circle) is identified if it is > $q3$ + ($q3 - q1$) × 1.5 or < $q1 - (q3 - q1)$ × 1.5.

Abbreviations: DRR = digitally reconstructed radiography; FPD = flat panel detector; H&N = head and neck; MAE mean absolute error; PSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.