



## PAPER

## Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy

RECEIVED  
12 September 2019REVISED  
25 November 2019ACCEPTED FOR PUBLICATION  
16 December 2019PUBLISHED  
24 January 2020Nimu Yuan<sup>1,2</sup>, Brandon Dyer<sup>3,4</sup>, Shyam Rao<sup>3</sup>, Quan Chen<sup>5</sup>, Stanley Benedict<sup>3</sup>, Lu Shang<sup>3</sup>, Yan Kang<sup>1</sup>,  
Jinyi Qi<sup>2</sup> and Yi Rong<sup>3</sup><sup>1</sup> Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, Liaoning, People's Republic of China<sup>2</sup> Department of Biomedical Engineering, University of California, Davis, CA, United States of America<sup>3</sup> Department of Radiation Oncology, University of California Davis Medical Center, Sacramento, CA, United States of America<sup>4</sup> Department of Radiation Oncology, University of Washington, Seattle, WA, United States of America<sup>5</sup> Department of Radiation Medicine, University of Kentucky, Lexington, KY, United States of AmericaE-mail: [yrong@ucdavis.edu](mailto:yrong@ucdavis.edu) and [qi@ucdavis.edu](mailto:qi@ucdavis.edu)**Keywords:** fast-scan low-dose, cone-beam CT, image quality enhancement, head and neck, deep learning, convolutional neural network  
Supplementary material for this article is available [online](#)**Abstract**

To improve image quality and CT number accuracy of fast-scan low-dose cone-beam computed tomography (CBCT) through a deep-learning convolutional neural network (CNN) methodology for head-and-neck (HN) radiotherapy.

Fifty-five paired CBCT and CT images from HN patients were retrospectively analysed. Among them, 15 patients underwent adaptive replanning during treatment, thus had same-day CT/CBCT pairs. The remaining 40 patients (post-operative) had paired planning CT and 1st fraction CBCT images with minimal anatomic changes. A 2D U-Net architecture with 27-layers in 5 depths was built for the CNN. CNN training was performed using data from 40 post-operative HN patients with 2080 paired CT/CBCT slices. Validation and test datasets include 5 same-day datasets with 260 slice pairs and 10 same-day datasets with 520 slice pairs, respectively. To examine the impact of differences in training dataset selection and network performance as a function of training data size, additional networks were trained using 30, 40 and 50 datasets. Image quality of enhanced CBCT images were quantitatively compared against the CT image using mean absolute error (MAE) of Hounsfield units (HU), signal-to-noise ratio (SNR) and structural similarity (SSIM).

Enhanced CBCT images reduced artifact distortion and improved soft tissue contrast. Networks trained with 40 datasets had imaging performance comparable to those trained with 50 datasets and outperformed those trained with 30 datasets. Comparison of CBCT and enhanced CBCT images demonstrated improvement in average MAE from 172.73 to 49.28 HU, SNR from 8.27 to 14.25 dB, and SSIM from 0.42 to 0.85. The image processing time is 2 s per patient using a NVIDIA GeForce GTX 1080 Ti GPU.

The proposed deep-learning methodology was fast and effective for image quality enhancement of fast-scan low-dose CBCT. This method has potential to support fast online-adaptive re-planning for HN cancer patients.

**1. Introduction**

Cone-beam CT (CBCT) has been widely used for inter-fractional daily patient positioning and tumor target alignment verification in radiotherapy treatments (Verellen *et al* 2007, Simpson *et al* 2010). Similar to the planning CT, CBCT images provide a 3D representation of patients' organs and tumors. CBCT images further support online evaluation of treatment delivery accuracy and the necessity of treatment plan adaptation during the course of treatment (Jaffray *et al* 2002, Hvid *et al* 2018). CBCT images may also allow early assessment of

treatment response and be a prognostic factor of treatment outcomes (van Timmeren *et al* 2017, Shi *et al* 2019). However, parameters for acquiring CBCT images are often chosen for fast scanning to increase high clinical throughput and low imaging dose to avoid any long term risks (de Gonzalez and Darby 2004). The resultant fast-scan low-dose (FSLD) CBCT images are associated with poor image quality (high noise, low contrast, scatter artifacts, etc) (Sykes *et al* 2005) which precludes the use of these images for adaptive re-planning during the course of radiotherapy (Kurz *et al* 2015).

Adaptive re-planning is particularly important in patients receiving head and neck (HN) radiotherapy as weight loss and tumor regression may result in large anatomic changes that could affect accurate dose delivery to radiotherapy targets, organs at risk (OARs), or both (Simone *et al* 2011). Numerous mathematical algorithms have been proposed to improve FSLD CBCT image quality, including iterative reconstruction (IR) with compressed sensing (CS) algorithms (Sidky *et al* 2006, Tian *et al* 2011, Yu and Wang 2010, Jia *et al* 2011, Chen *et al* 2008, Xu *et al* 2014). CS-based IR algorithms reconstruct CBCT images from under-sampled projections, using *a priori* sparsity properties of the images by total variation (TV) regularization (Sidky *et al* 2006, Tian *et al* 2011), soft-thresholding regularization (Yu and Wang 2010), tight frame (TF) regularization (Jia *et al* 2011), or prior images (Chen *et al* 2008, Xu *et al* 2014). The resulting image quality is improved and may be comparable to images with high imaging dose (Sidky *et al* 2006, Tian *et al* 2011, Yu and Wang 2010, Jia *et al* 2011, Chen *et al* 2008, Xu *et al* 2014). However, IR algorithms for CBCT have high computational complexity, thus have not been commonly implemented for clinical use. Alternatively, conventional analytic reconstruction algorithms, such as filtered back-projection, remain the mainstream due to fast computation. Deep learning approaches, particularly convolution neural networks (CNNs), have emerged as a potential solution to overcome computational complexity of prior reconstruction algorithms and inherent poor image quality of CBCT (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019, Liang *et al* 2019). These approaches have demonstrated promising results for CBCT by applying denoising networks to generate synthetic CT images.

Most CNN-based approaches using supervised training for CBCT mainly focus on prostate cancer (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019), whereas the only one study for HN used an unsupervised network CycleGAN (Liang *et al* 2019). In this work, we present an efficient method to improve FSLD CBCT image quality for HN cancer patients using a U-Net CNN. The scope of the present study is primarily on the network construction, feasibility assessment, and image quality improvement evaluation. Herein, a deep learning U-Net CNN is constructed through direct mapping from the CBCT images to the corresponding CT images in order to generate high quality CBCTs, a.k.a. synthetic CTs. The main challenge of the present study is that the FSLD CBCT images that we used are with low signal-to-noise ratio, low soft tissue contrast, and high scatter artifacts.

## 2. Methods and materials

### 2.1. Network design and architecture

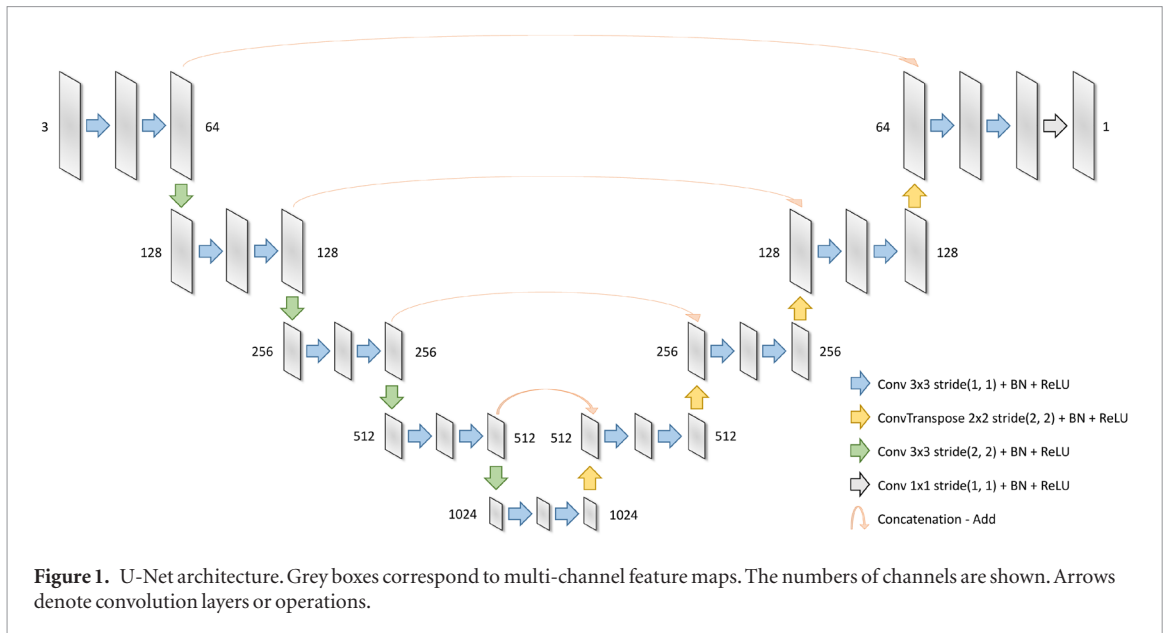
The present study used a U-Net (Ronneberger *et al* 2015) architecture with an encoder-decoder path. As shown in figure 1, the network is built with a series of convolution layers involving stride  $1 \times 1$  convolutional layers, stride  $2 \times 2$  convolutional layers (down-sampling process), and transposed convolutional layers (up-sampling process) with skip connections (add operation). The kernel size is  $3 \times 3$  in all convolutional layers except in the last one, where the kernel size is  $1 \times 1$ . The root filter number is 32 and the filter number is doubled after each down-sampling process. Batch normalization (BN) (Ioffe and Szegedy 2015) and rectified linear units (ReLU) (Nair and Hinton 2010) are used after each convolutional layer.

Compared with the architecture of the original published U-Net (Ronneberger *et al* 2015), our approach differed in several key areas: (1) the max-pooling layers were replaced by convolution kernels of stride 2 to keep more subtle features in the down-sampling process (Springenberg *et al* 2014); (2) the up-convolution  $2 \times 2$  was replaced with bilinear interpolation up-sampling with a  $3 \times 3$  convolution layer to avoid checkerboard artifacts (Odena *et al* 2016); (3) the skip-add operation was used instead of the skip-concatenation to be GPU memory efficient.

### 2.2. Network training

For all experiments the networks were trained using a full-size CBCT slice and two adjacent slices, i.e. 3-channel data were used as the network input. The original CBCT used for the study is denoted as oCBCT, while the network output is denoted as enhanced CBCT (eCBCT). Mean absolute error (MAE) was used as the loss function (Zhao *et al* 2016), which measures the pixel-wise difference between eCBCT images  $I_{eCBCT} \in R^{m \times n}$  and the label CT images  $I_{CT} \in R^{m \times n}$ :

$$Loss_{MAE} = \frac{1}{m \times n} \sum_x \sum_y \|I_{eCBCT}(x, y) - I_{CT}(x, y)\|_1. \quad (1)$$



The adaptive moment estimation (Adam) algorithm (Kingma and Ba 2014) was used as the optimizer. The initial learning rate was 0.001, which was then automatically reduced by 80% once the training loss stopped reducing for 20 epochs. The mini-batch size is 2. The number of total epochs in each training was 1000.

The network was implemented using Tensorflow (Martín *et al* 2015) 1.8.0 and Keras (Chollet and others 2015) 2.2.4, and trained using an NVIDIA GeForce GTX 1080 Ti GPU. Training a single network took approximately 25 h for 1000 epochs.

### 2.3. Metrics for evaluation

In this study, MAE, signal-to-noise ratio (SNR) and structural similarity (SSIM) (Wang *et al* 2004) were used to evaluate the performance of the prediction from different trained networks between eCBCT  $I_{eCBCT}$  and CT  $I_{CT}$  slices. The definition of MAE is given in (1).

SNR is defined as the ratio of signal power to the noise power with a unit of decibels (dB). It is defined as

$$SNR = 10 \cdot \log_{10} \left[ \frac{\sum_x \sum_y [I_{CT}(x, y)]^2}{\sum_x \sum_y [I_{CT}(x, y) - I_{eCBCT}(x, y)]^2} \right]. \quad (2)$$

SSIM is a perceptual metric that quantifies image quality degradation. It is used for measuring the similarity between two images and is calculated as

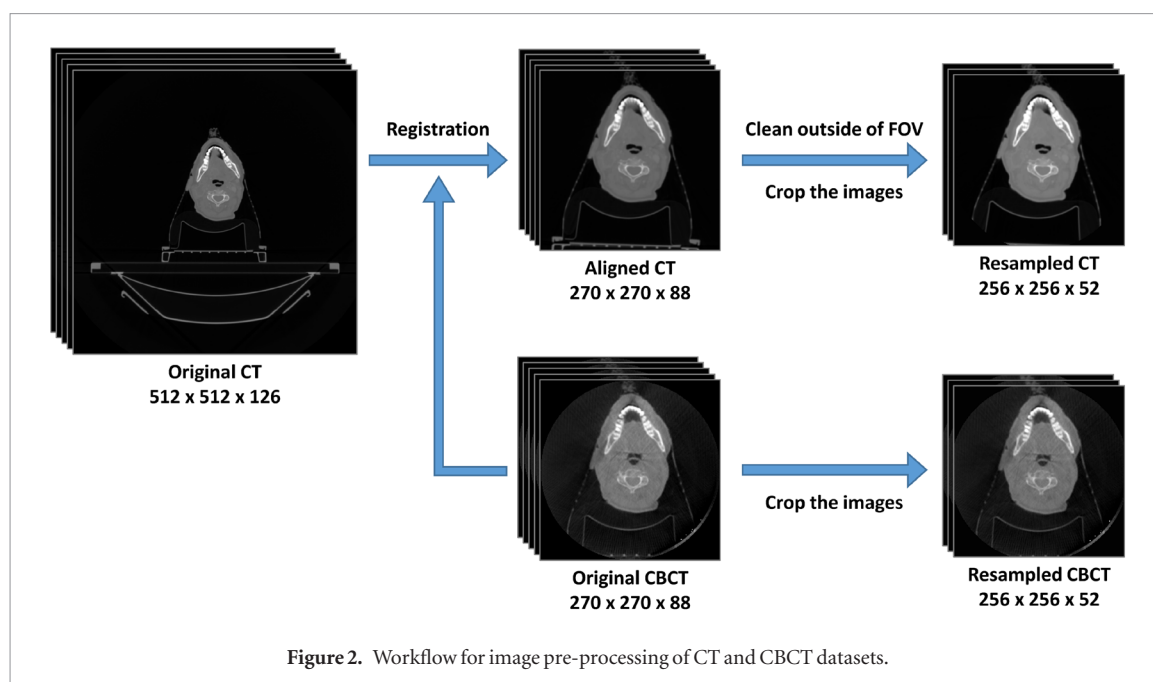
$$SSIM = \frac{(2\mu_{eCBCT}\mu_{CT} + C_1)(2\delta_{eCBCT\&CT} + C_2)}{(\mu_{eCBCT}^2 + \mu_{CT}^2 + C_1)(\delta_{eCBCT}^2 + \delta_{CT}^2 + C_2)} \quad (3)$$

where  $\mu$  denotes the mean value,  $\delta^2$  denotes the variance and the parameters  $C_1 = (k_1 Q)^2$  and  $C_2 = (k_2 Q)^2$  are two variables to stabilize the division with weak denominators, where  $k_1 = 0.01$  and  $k_2 = 0.02$ .  $Q$  is the dynamic range of the pixel-values.

### 2.4. Experimental datasets

#### 2.4.1. CT and CBCT image acquisition

This retrospective study was approved by the institutional review board (IRB) and a total of 55 HN cancer patients who completed radiotherapy treatment with daily CBCT image guidance were identified. Among these 55 patients, the planning-CT and first fraction CBCT (one to three weeks apart) from the 40 post-operative HN patients were rigid registered. Anatomic similarity was visually and quantitatively confirmed based on a criterion that boundary differences between the two images were  $<5$  mm (5 pixels) for the external contour, bony anatomy, and internal cavities (nasal cavity, oral cavity, etc). The 15 remaining CT/CBCT pairs were from patients who received offline adaptive re-planning during the treatment process. In these 15 patients, a new simulation CT (re-sim CT) was obtained during the course of treatment typically due to tumor shrinkage and/or subsequent changes in adjacent organs. Therefore, these patients had the re-sim CT and a CBCT acquired on the same day. The pairing of re-sim CT and same-day CBCT ensured minimal anatomic variation between the two imaging studies. The CBCT and re-sim CT images for these 15 patients were also rigid registered and confirmed  $<5$  mm anatomic deviation.



CT simulation images were obtained on a Brilliance Big Bore CT scanner (Philips Health) using the following parameters: 120 kVp tube voltage and 400–500 mAs exposure. The matrix size of simulation CT images was  $512 \times 512 \times 144$  with a pixel size of  $1.17 \text{ mm} \times 1.17 \text{ mm}$  and slice thickness of 3.00 mm. CBCT images were acquired with the XVI onboard imager system on an Elekta Synergy Linac (Elekta, Sweden) using a tube voltage of 100 kVp and an exposure of 18.2 mAs (10 mA and 10 ms per projection). The number of projections per scan was 182 frames for a  $205^\circ$  ( $-45^\circ$ – $160^\circ$ ) rotation. The mode of reconstruction was set to fast reconstruction ( $<60 \text{ s}$ ). The matrix size of raw CBCT images was  $270 \times 270 \times 88$  with a pixel size of  $1.00 \text{ mm} \times 1.00 \text{ mm}$  and slice thickness of 3.00 mm.

#### 2.4.2. Image pre-processing

The data pre-processing workflow is shown in figure 2. For each patient, the CT and CBCT images were co-registered through grey scale rigid transformation. The matrix size of aligned CT and CBCT images was  $270 \times 270 \times 88$ . Pixel values outside the field of view (FOV) were set to the Hounsfield unit (HU) of air ( $-1000$ ). The images were further center-cropped to  $256 \times 256$  pixels in the transverse plane. In the Z dimension, only the center 52 slices were used and the rest were discarded due to the severe CBCT photon scatter artifacts at the superior and inferior extent of the image stack.

#### 2.4.3. Training, validation, and testing datasets

Data were divided into three subsets for (1) network training, (2) validation, for monitoring the training process and tuning network hyper-parameters and (3) testing, for evaluating the network performance. The evaluation results were not boot-strapped for network re-optimization.

#### 2.4.4. Datasets arrangement

The network was trained using different size datasets and combinations to verify training sufficiency. Specifically, among the 55 datasets, five from the same-day CT/CBCT pairs were used for validation, which include a wide spectrum of variable head positions (as shown in supplementary materials figure S1 ([stacks.iop.org/PMB/65/035003/mmedia](https://stacks.iop.org/PMB/65/035003/mmedia))). The remaining 50 datasets were divided amongst five groups, with Group 1 containing the rest of 10 same-day CT/CBCT datasets and Group 2–5 containing the 40 post-operative patient datasets (10 datasets per group, as shown in the first row in table 1). A total of 15 networks were trained using 30, 40, and 50 datasets formed from various combinations of dataset groups, as indicated by table 1. The five networks trained using five groups differ only by the random initialization of the network weights and have different shuffling of the batches. In order to eliminate the impact of loss fluctuations, the average evaluation loss of the last 100 epochs (No. 900–1000) was calculated.

## 3. Results

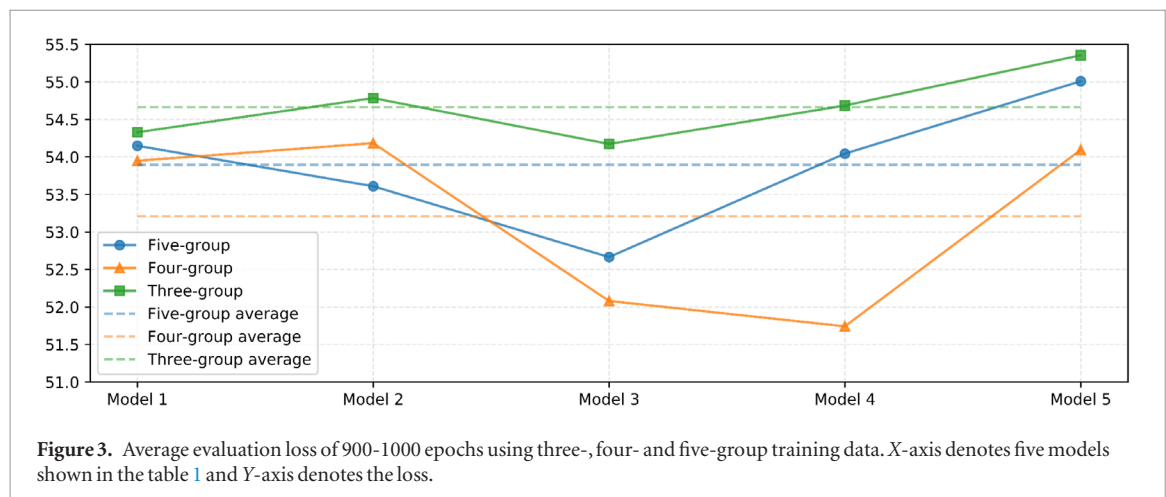
### 3.1. Network optimization

The results of the average evaluation loss are shown in figure 3. A lower evaluation loss value is indicative of improved/superior network training performance. The worst values are seen in the three-group data training,

**Table 1.** The dataset arrangements for cross evaluation. Network 6 (four groups training) is the main result in this work.

|                          | Group 1<br>(same-day) | Group 2 (post-<br>operative) | Group 3 (post-<br>operative) | Group 4 (post-<br>operative) | Group 5 (post-<br>operative) |
|--------------------------|-----------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| Five groups, network 1   | ★                     | ★                            | ★                            | ★                            | ★                            |
| Five groups, network 2   | ★                     | ★                            | ★                            | ★                            | ★                            |
| Five groups, network 3   | ★                     | ★                            | ★                            | ★                            | ★                            |
| Five groups, network 4   | ★                     | ★                            | ★                            | ★                            | ★                            |
| Five groups, network 5   | ★                     | ★                            | ★                            | ★                            | ★                            |
| Four groups, network 6   | —                     | ★                            | ★                            | ★                            | ★                            |
| Four groups, network 7   | ★                     | —                            | ★                            | ★                            | ★                            |
| Four groups, network 8   | ★                     | ★                            | —                            | ★                            | ★                            |
| Four groups, network 9   | ★                     | ★                            | ★                            | —                            | ★                            |
| Four groups, network 10  | ★                     | ★                            | ★                            | ★                            | —                            |
| Three groups, network 11 | —                     | —                            | ★                            | ★                            | ★                            |
| Three groups, network 12 | —                     | ★                            | —                            | ★                            | ★                            |
| Three groups, network 13 | —                     | ★                            | ★                            | —                            | ★                            |
| Three groups, network 14 | —                     | ★                            | ★                            | ★                            | —                            |
| Three groups, network 15 | ★                     | ★                            | ★                            | —                            | —                            |

★: used in the trainings; —: not used in the training.

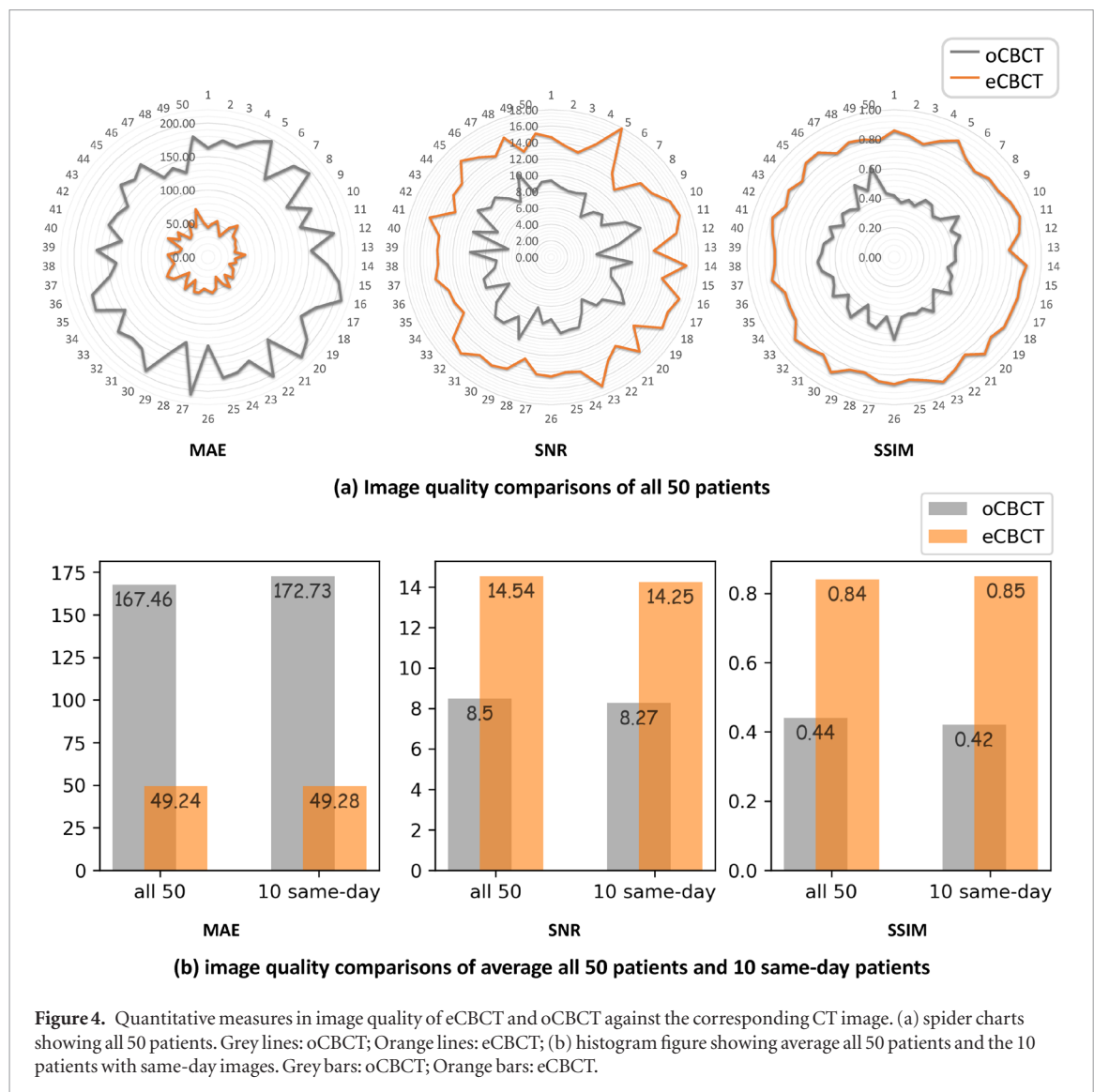
**Figure 3.** Average evaluation loss of 900-1000 epochs using three-, four- and five-group training data. X-axis denotes five models shown in the table 1 and Y-axis denotes the loss.

indicating 30 datasets are insufficient to obtain a well-trained network (paired *t*-test with four-group and five-group,  $p < 0.05$ ). Lower evaluation loss can be achieved by both four-group and five-group datasets, with no significant difference (paired *t*-test,  $p > 0.1$ ). This indicates that an improved network can be obtained using four groups of datasets with less training time. Therefore, in the subsequent study, Network 6 with four-group datasets setup was used for generating the main results unless noted otherwise. In this setting, patient images from 40 post-operative patient datasets (73%) were used for training, 5 same-day CT/CBCT datasets (9%) for validation, and 10 remaining same-day CT/CBCT datasets (18%) for testing.

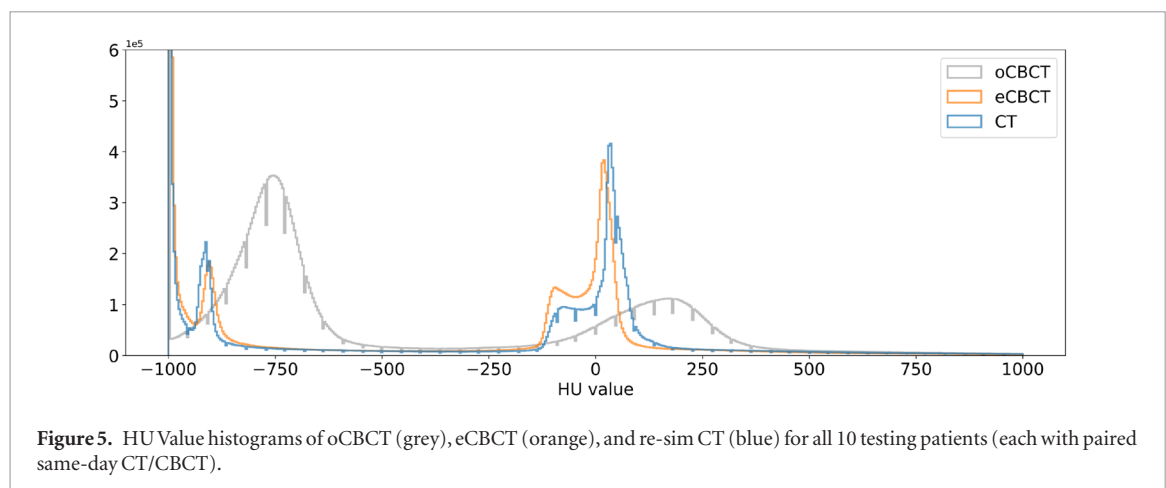
### 3.2. Quantitative assessment for the enhanced CBCT image quality

Figure 4(a) shows the testing results of all 50 patients using five-fold cross-validation (exclude five same-day patients used in network optimization) from networks trained by four groups (Networks 6–10) in table 1. All eCBCT images had significant improvement in MAE, SNR, and SSIM compared with the oCBCT (in all paired *t*-test,  $p \ll 0.01$ ). As shown in figure 4(b) based on 50 patients, the average MAE of HU improved from 167.46 to 49.24 HU, the average SNR improved from 8.50 to 14.54 dB and the average SSIM increased from 0.44 to 0.84. Using the 10 same-day testing patients (main results, network 6), the average MAE of HU improved from 172.73 to 49.28 HU, the average SNR from 8.27 to 14.25 dB, and the average SSIM from 0.42 to 0.85. Figure 5 shows the histograms of oCBCT, eCBCT, and re-sim CT for the 10 testing datasets. The range of the HU value is from  $-1000$  to  $1000$  with 512 bins. The HU histogram of oCBCT displays significant shifts in low density/air and soft tissue/muscle regions, whereas the HU histogram of eCBCT is much closer to that of re-sim CT for both regions. The Chi-squared distance of the histograms (Zhang *et al* 2007) was suppressed from  $9.46 \times 10^6$  (oCBCT) to  $1.26 \times 10^6$  (eCBCT).





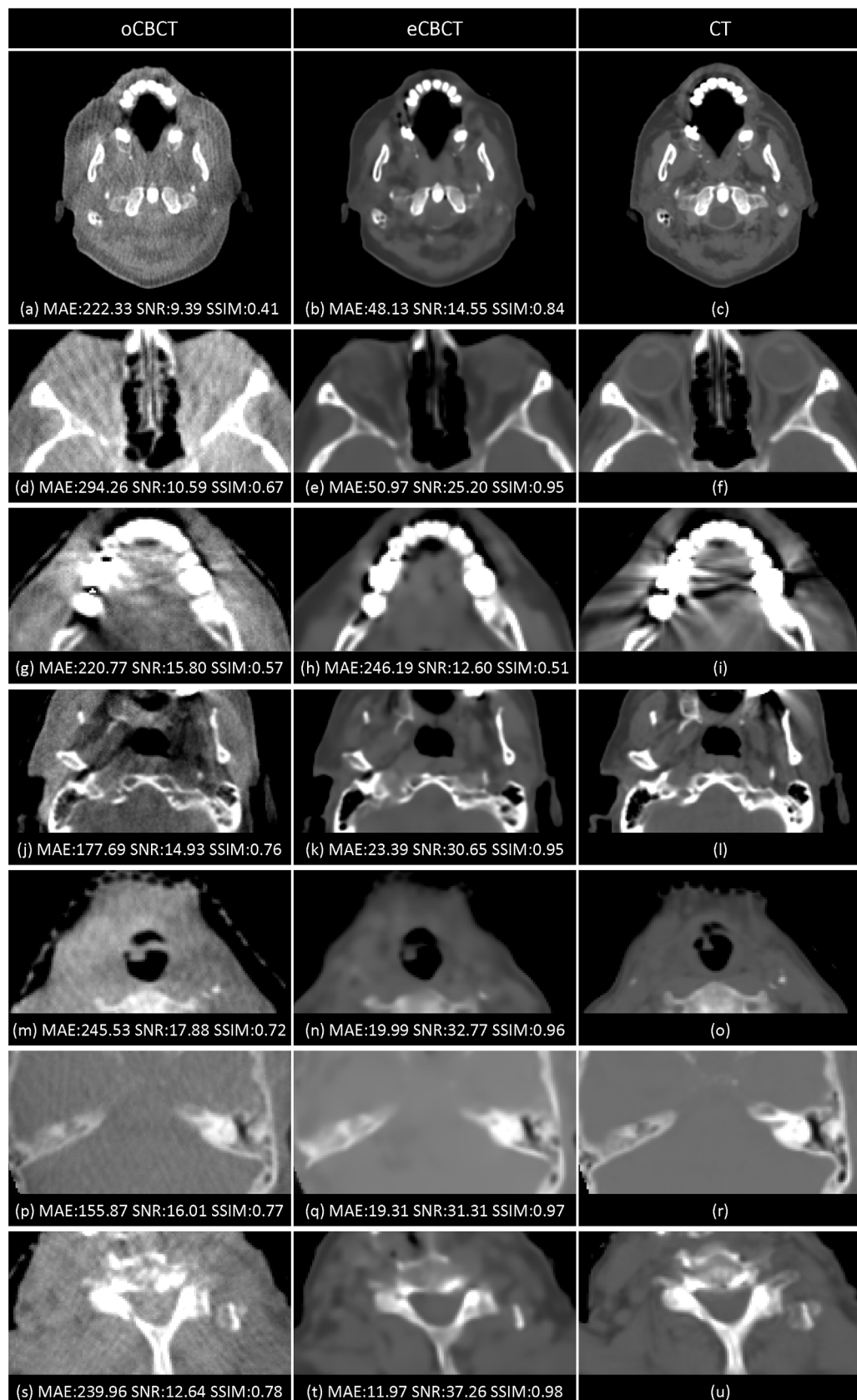
**Figure 4.** Quantitative measures in image quality of eCBCT and oCBCT against the corresponding CT image. (a) spider charts showing all 50 patients. Grey lines: oCBCT; Orange lines: eCBCT; (b) histogram figure showing average all 50 patients and the 10 patients with same-day images. Grey bars: oCBCT; Orange bars: eCBCT.



**Figure 5.** HU Value histograms of oCBCT (grey), eCBCT (orange), and re-sim CT (blue) for all 10 testing patients (each with paired same-day CT/CBCT).

### 3.3. Qualitative assessment for the enhanced CBCT image quality

Seven representative image slices from a testing dataset are shown in figure 6. The display window is  $W = 950$  HU,  $L = 181$  HU. The values of MAE, SNR and SSIM are shown at the bottom of each panel. The trained network significantly improved the SNR and SSIM of eCBCT compared with oCBCT images. The HU difference between eCBCT and CT were reduced to 25% of those values for oCBCTs. Overall image quality seemed much improved from a representative slice shown in figures 6(a)–(c). Small but critical structures, i.e. the optic nerves, were enhanced in eCBCT (figures 6(d)–(f)). High-Z streak artifacts were reduced in the dental regions in eCBCTs (figures 6(g)–(i)). Soft tissue contrast of parotid regions was improved (figures 6(j)–(l)); however, surgical clips



**Figure 6.** The oCBCT images (left column), eCBCT images (middle column) and reference CT images (right column). (a)–(c) The representative image slices which have similar metrics to the average of the full testing data; (d)–(f) optic nerve regions; (g)–(i) dental regions; (j)–(l) parotid regions; (m)–(o) submandibular gland (SMG) regions; (p)–(r) Brainstem regions; (s)–(u) cord regions.

(small white dots in the image) may also be suppressed in the denoising process of the network (figures 6(m)–(o)). Soft tissue contrast was greatly enhanced with corrected HU values for regions of brainstem (figures 6(p)–(r)) and cord (figures 6(s)–(u)). HU-line profiles were taken for figures 6(a)–(c), (p)–(l), and (s)–(u) and shown in figures 7 (a)–(c), respectively. As the network removed most artifacts on the oCBCT images and minimize the differences of HU values between oCBCT and CT images, eCBCT HU profiles are much smoother and closer to the corresponding CT HU profiles, especially in the soft-tissue areas.

### 3.4. Computation time

After training, the average time to enhance a single oCBCT image slice is 13 ms. Complete processing of a CBCT dataset for one patient takes less than 15 s, which includes loading the oCBCT raw data, pre-processing, network computing (13 ms/slice  $\times$  52 slices), and saving the network predictions.

## 4. Discussions

The CNN-based method we developed in this study aims to improve image quality of FSLD CBCTs obtained from routine image-guided radiotherapy and daily treatment verification. The improvement was evaluated for image uniformity, SNR, contrast resolution, and image similarity in comparison with standard-dose CT. Networks were trained to map from oCBCT to CT images, in order to create an eCBCT. Compared with oCBCT images, eCBCT images demonstrated a substantial improvement in HU number accuracy and small anatomical structure integrity. Furthermore, for eCBCT images, the CNN visually suppressed noise and scatter artifacts, as well as streaky beam hardening artifacts in dental areas, which is consistent with the previous publication from our group (Yuan *et al* 2019). Most recently, Liang *et al* published a similar study on CBCT image improvement for HN cancer using CycleGAN (Liang *et al* 2019), which showed CBCT image quality improvement, but was unable to suppress metal and dental artifacts because they were shown in both CT and CBCT images.

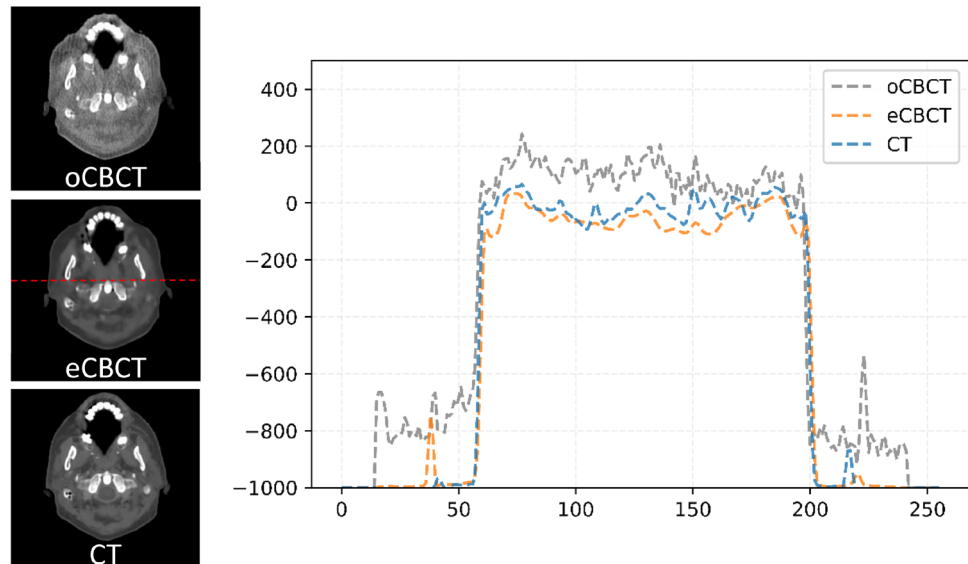
As shown in figure 3, the three-group training data demonstrated suboptimal results with high validation loss. When using the four-group training dataset, evaluation loss decreased significantly (paired *t*-test,  $p < 0.05$ ). When five groups of datasets were used for training, there was no significant improvement (paired *t*-test,  $p > 0.1$ ), with a slight increase in the average loss. Therefore, the four-group dataset including 40 independent sets was an optimal and sufficient size for training in this study.

For this supervised CNN, CT and CBCT image pairs with matching anatomy were used to train the network. Identifying matched CT/CBCT pairs is extremely challenging in a retrospective setting. Table 2 summarized the number of training data and imaging parameters in previous publications with the same aim in improving CBCT image quality based on deep-learning CNN methods (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019, Liang *et al* 2019) and CT-assisted intensity correction methods (Hu *et al* 2008, Park *et al* 2015, Xu *et al* 2015). Our study used the highest number of datasets for training, compared to the other three supervised deep-learning CNN-based studies (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019). In addition, the present study also utilized a relatively large sample size for testing with the same-day CT/CBCT, which includes a wide range of variable treatment positions in terms of head tilt, treatment region size, and dental artifacts. Most importantly, the re-sim CT represents the ground truth of patients' anatomy on the day when the corresponding CBCT was taken, which serves as a precise image verification.

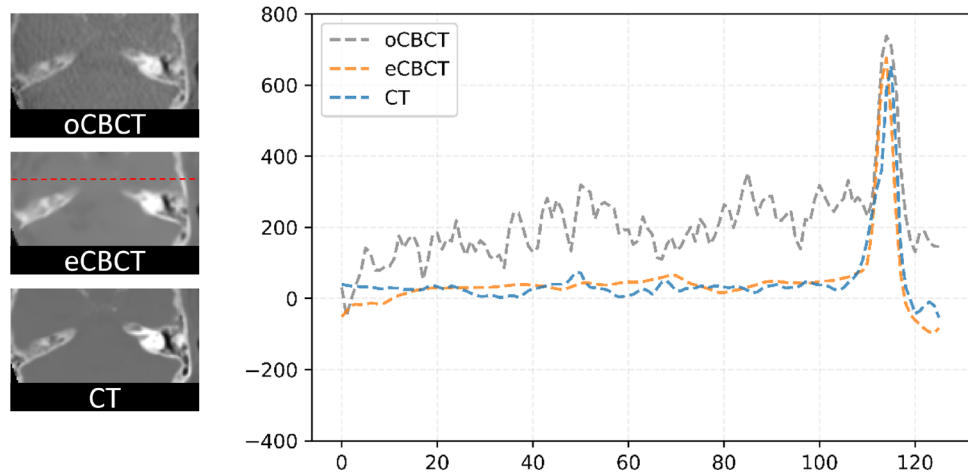
The training data used in this study are from post-operative patients and include the planning CT and CBCT performed on the first day of treatment. Non-operative patients were excluded in order to ensure minimal anatomic differences between the CT and CBCT. As validated in Dinkla *et al* 2019, acceptable results were achieved using up to a maximum of 28 d separation between the CT and MR image pairs, which were aligned by deformable image registration (DIR) to prevent mismatch between the image pairs. For our study, grey-scale-based rigid registration was considered sufficient in registering each image pair, given careful dataset screening as we mentioned above. In addition, we preliminarily evaluated a pixel-value based commercial DIR algorithm in the initial phase of our study, but concluded that it was not superior to the rigid registration due to inherent deformation errors associated with significant image artifacts from the FSLD CBCT images. Furthermore, to assess test dataset performance, Network 6 trained with all post-operative image pairs was compared with Network 7–10 trained with both same-day and post-operative image pairs. Based on the four-group dataset cross validation experiments shown in table 1 and figure 3, no significant difference was noted in terms of the network performance. Therefore, the training dataset obtained from post-operative patient datasets is comparable to the same-day dataset for this study.

The most challenging aspect of this study is the use of FSLD CBCTs. The oCBCT images were acquired with a very low scanning mAs parameters, in order to reduce daily patient imaging dose. Images were acquired for each patient using the XVI S10 filter and  $10 \times 10$  protocol, with the exposures of 10 ms and x-ray tube current of 10 mA per projection for 182 projections. Table 2 compared the CBCT imaging parameters of the present work with previous publications on prostate (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019) and HN (Hu *et al*

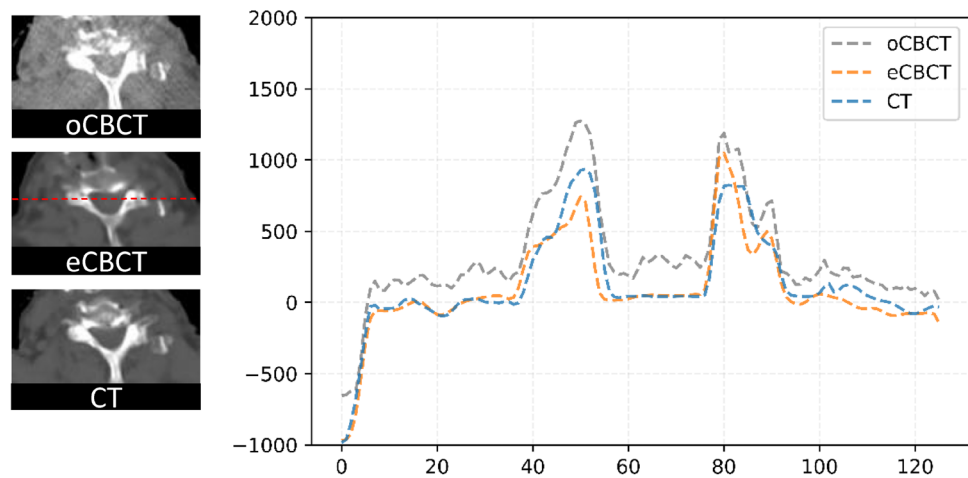




(a) Overall images



(b) Brainstem regions



(c) Cord regions

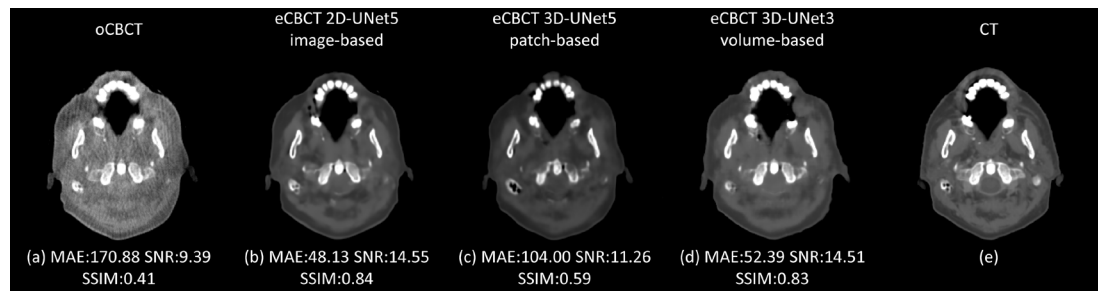
**Figure 7.** HU line profiles taken at three regions at (a) figures 6(a)–(c), (b) figure 6(p)–(l), (c) figures 6(s)–(u). The right column shows HU profiles of the red dashed lines in the left column. Y axes: HU; X axes: pixels.

**Table 2.** The comparison between our method and Ref (Hansen *et al* 2018, Kida *et al* 2018, Landry *et al* 2019, Liang *et al* 2019) and (Hu *et al* 2008, Park *et al* 2015, Xu *et al* 2015).

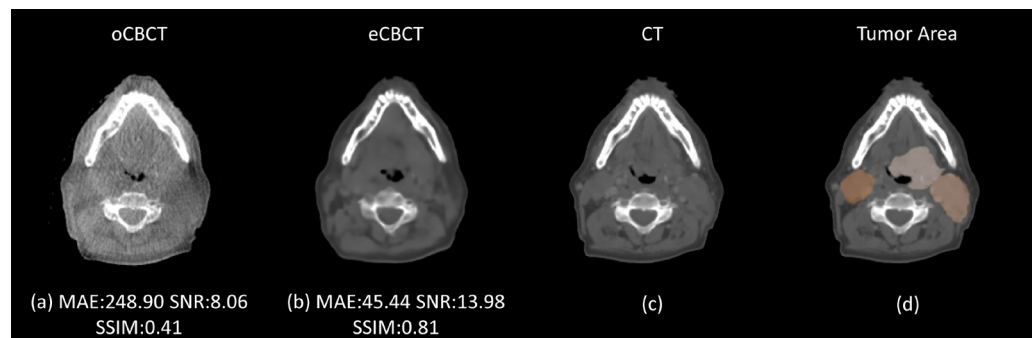
|   | Image processing method      | Training dataset size | Testing dataset size | oCBCT scan dose (mAs) | oCBCT MAE (HU)   | eCBCT MAE (HU) |
|---|------------------------------|-----------------------|----------------------|-----------------------|------------------|----------------|
| Current study (HN)                            | Supervised CNN               | 40                    | 10                   | 18.2                  | 167<br>[130–209] | 49<br>[31–74]  |
| Kida <i>et al</i> (2018) (prostate)           | Supervised CNN               | 16                    | 4                    | 350                   | —                | —              |
| Hansen <i>et al</i> (2018) (prostate)         | Supervised CNN               | 15                    | 8                    | 560–1147.2            | 144              | 46             |
| Net1 in Landry <i>et al</i> (2019) (prostate) | Supervised CNN               | 27                    | 8                    | 138–142               | 104<br>[91–119]  | 51<br>[43–62]  |
| Net2 in Landry <i>et al</i> (2019) (prostate) | Supervised CNN               | 27                    | 8                    | 138–142               | 104<br>[91–119]  | 88<br>[75–105] |
| Net3 in Landry <i>et al</i> (2019) (prostate) | Supervised CNN               | 27                    | 8                    | 138–142               | 104<br>[91–119]  | 58<br>[49–69]  |
| Liang <i>et al</i> (2019) (HN)                | Unsupervised CNN             | 81                    | 20                   | —                     | 69<br>[80–68]    | 30<br>[35–25]  |
| Xu <i>et al</i> (2015) (HN)                   | Monte Carlo simulation-based | —                     | —                    | 145.6                 | —                | —              |
| Park <i>et al</i> (2015) (HN)                 | CT-prior-based               | —                     | —                    | 34–69                 | —                | —              |
| Hu <i>et al</i> (2008) (HN)                   | ROI mapping                  | —                     | —                    | 65                    | —                | —              |

2008, Park *et al* 2015, Xu *et al* 2015) cancer sites. Amongst the three deep-learning CNN-based studies, Landry *et al* (2019) used  $20 \times 20$  protocol for CBCT acquisition and the number of projections was between 346 and 357, whereas Hansen *et al* (2018) used  $40 \times 40$  protocol and the number of projections was between 350 and 717. Amongst the studies using conventional algorithms for HN CBCT (Hu *et al* 2008, Park *et al* 2015, Xu *et al* 2015), Xu *et al* (2015) adopted the CBCT scanning protocol of 0.4 mAs/projection and 364 projections; Park *et al* (2015) used 0.1 mAs/projection and 343 projections; Hu *et al* (2008) acquired CBCT with 0.1 mAs/projection and 650 projections. In all, the FSLD protocol (0.1 mAs, 182 projections) that we used in the present study is beneficial to patients, but resulted in large image quality degradation for oCBCTs (average HU difference in oCBCT versus CT in our study: 167 [130–209] HU; average HU difference in Landry *et al* (2019): 104 [91–119] HU). Nevertheless, the image quality improvement shown in eCBCT is slightly better in HU number corrections compared to Landry *et al* (2019) (the present study: 49 [31–74] HU, Unet1 in Landry *et al* (2019): 51 [43–62] HU, Unet2 in Landry *et al* (2019): 88 [75–105] HU, Unet3 in Landry *et al* (2019): 58 [49–69] HU).

Our present study showed promising results using a U-Net CNN for creating synthetic CT with FSLD CBCTs; however the limitations of the study should be noted. A multi-slice 2D CNN was used in this study due to the limitation of the GPU memory. Given the 3D nature of CBCT images, a 3D CNN may have greater capability to remove noise and streaky artifacts in CBCT images (Çiçek *et al* 2016, Milletari *et al* 2016). One way to reduce the memory requirement of a 3D CNN is to use patch-based training. However, 3D patch-based training may not have equivalent performance with the full-image/volume-based training due to the receptive fields being limited to the patch size. Additionally, long streaky or scatter artifacts can cause long-range correlations, which may not be captured by patch-based training. Therefore, the performance of a patch-based 3D network is not necessarily superior to a 2D network. To demonstrate this, two 3D U-Net networks, (1) a depth-5 U-Net trained using patch-based data ( $32 \times 32 \times 32$ , 50% overlap) and (2) a simpler depth-3 U-Net trained using full-size 3D volume, were evaluated as part of our preliminary assessment. The former has four down-sampling processes and the latter only has two. As shown in the figure 8, 3D patch-based training has a bias issue as the intensity of the entire image can be mismatched. This mismatch may result in degraded image quality compared to 2D-UNet results. In comparison, volume-based training shows in general equivalent image quality metrics compared to that of the 2D-UNet with slightly inferior soft-tissue contrast and spatial resolution due to its simpler architecture. Our future study will include setting up multi-GPUs in order to further explore 3D volume-based training network for improving FSLD CBCTs. The second limitation is that very small structures, such as surgical clips, were not well preserved in eCBCT. As shown in figure 6, the spatial resolution in eCBCT is inferior compared to the reference CT. One possible solution is to modify the loss functions to focus on very small structures, image boundary, and soft tissue. Finally, soft tissue contrast improvement of the tumor areas in the test datasets is limited due to the use of post-operative patients (tumor regions removed after surgery) for network training. Figures 9(a)–(d) show an example from the testing dataset with tumors present. Our trained CNN provided little enhancement in the gross tumor areas compared to the original CBCT. This could be a limitation of the FSLD CBCT. For those oCBCT from non-operative HN cancer patients, tumor area enhancement would require matching training data and more optimized imaging scanning parameters. A prospective clinical trial is underway to include



**Figure 8.** (a) oCBCT and processed CBCT images using (b) image-based 2D depth-5 U-Net, (c) patch-based 3D depth-5 U-Net and (d) volume-based 3D depth-3 U-Net, in comparison with (e) the reference CT.



**Figure 9.** Tumor area comparison on a testing dataset with same-day CBCT/CT: (a) oCBCT, (b) eCBCT, (c) re-sim CT, and (d) mask areas for gross tumor volumes.

non-operative HN cancer patients for network training and to explore the clinical feasibility of applying the proposed method to facilitate online evaluation of treatment delivery accuracy for plan adaptation during treatment. Future work will also include testing the feasibility and accuracy of direct dose calculation on eCBCT given improved image quality in tumor area and normal tissues.

## 5. Conclusions

The proposed deep-learning based method is demonstrated to be fast and effective for enhancing FSLD CBCT. Quantitative and qualitative comparisons showed improved image quality in terms of soft tissue contrast, SNR, and HU number accuracy with the eCBCT datasets. This study has its potential utility for fast online-dose verification and adaptive re-planning of radiotherapy for HN cancer patients.

## Acknowledgments

This work is supported in part by a scholarship (N Yuan) from China Scholarship Council (CSC) under the Grant CSC No. 201706080096, and by NIH Grant R43EB027523 (Q Chen).

## ORCID iDs

Quan Chen <https://orcid.org/0000-0001-5570-2462>

Jinyi Qi <https://orcid.org/0000-0002-5428-0322>

Yi Rong <https://orcid.org/0000-0002-2620-1893>

## References

- Chen G H, Tang J and Leng S 2008 Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets *Med. Phys.* **35** 660–3
- Chollet F *et al* 2015 Keras (<https://keras.io>)
- Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O 2016 3D U-Net: learning dense volumetric segmentation from sparse annotation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* pp 424–32
- de Gonzalez A B and Darby S 2004 Risk of cancer from diagnostic x-rays: estimates for the UK and 14 other countries *Lancet* **363** 345–51

- Dinkla A M *et al* 2019 Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based 3D convolutional neural network *Med. Phys.* **46** 4095–104
- Hansen D C, Landry G, Kamp F, Li M, Belka C, Parodi K and Kurz C 2018 ScatterNet: A convolutional neural network for cone-beam CT intensity correction *Med. Phys.* **45** 4916–26
- Hu W, Wang J, Xie J, Ma X, Ye J and Zhang Z 2008 Use of kilovoltage cone beam CT in patient dose calculation for head and neck and partial brain radiation therapy *Int. J. Radiat. Oncol.* **72** S601
- Hvid C A, Elstrøm U V, Jensen K and Grau C 2018 Cone-beam computed tomography (CBCT) for adaptive image guided head and neck radiation therapy *Acta Oncol.* **57** 552–6
- Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *32nd Int. Conf. on Machine Learning, ICML 2015* vol 1 pp 448–56
- Jaffray D A, Siewerdsen J H, Wong J W and Martinez A A 2002 Flat-panel cone-beam computed tomography for image-guided radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 1337–49
- Jia X, Dong B, Lou Y and Jiang S B 2011 GPU-based iterative cone-beam CT reconstruction using tight frame regularization *Phys. Med. Biol.* **56** 3787–807
- Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, Yamashita H and Nakagawa K 2018 Cone beam computed tomography image quality improvement using a deep convolutional neural network *Cureus* **10** e2548
- Kingma D P and Ba J 2014 Adam: a method for stochastic optimization *3rd Int. Conf. on Learning Representations, ICLR 2015—Conf. Track Proc.*
- Kurz C *et al* 2015 Comparing cone-beam CT intensity correction methods for dose recalculation in adaptive intensity-modulated photon and proton therapy for head and neck cancer *Acta Oncol.* **54** 1651–7
- Landry G, Hansen D, Kamp F, Li M, Hoyle B, Weller J, Parodi K, Belka C and Kurz C 2019 Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations *Phys. Med. Biol.* **64** 035011
- Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, Wang J and Jiang S 2019 Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy *Phys. Med. Biol.* **64** 125002
- Martin A *et al* 2015 *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* (<http://tensorflow.org>)
- Milletari F, Navab N and Ahmadi S-A 2016 V-net: fully convolutional neural networks for volumetric medical image segmentation *2016 4th Int. Conf. on 3D Vision* pp 565–71
- Nair V and Hinton G E 2010 Rectified linear units improve restricted boltzmann machines *Proc. of the 27th Int. Conf. on Machine Learning* pp 807–14
- Odena A, Dumoulin V and Olah C 2016 Deconvolution and checkerboard artifacts *Distill* **1** e3
- Park Y K, Sharp G C, Phillips J and Winey B A 2015 Proton dose calculation on scatter-corrected CBCT image: feasibility study for adaptive proton therapy *Med. Phys.* **42** 4449–59
- Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* ed N Navab *et al* (Berlin: Springer) p 8
- Shi L, Rong Y, Daly M, Dyer B A, Benedict S, Qiu J and Yamamoto T 2019 Cone-beam computed tomography-based delta-radiomics for early response assessment in radiotherapy for locally advanced lung cancer *Phys. Med. Biol.* **65** 015009
- Sidky E Y, Kao C-M and Pan X 2006 Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT *J. X-Ray Sci. Technol.* **14** 119–39
- Simone C B, Ly D, Dan T D, Ondos J, Ning H, Belard A, O'Connell J, Miller R W and Simone N L 2011 Comparison of intensity-modulated radiotherapy, adaptive radiotherapy, proton radiotherapy, and adaptive proton radiotherapy for treatment of locally advanced head and neck cancer *Radiother. Oncol.* **101** 376–82
- Simpson D R, Lawson J D, Nath S K, Rose B S, Mundt A J and Mell L K 2010 A survey of image-guided radiation therapy use in the United States *Cancer* **116** 3953–60
- Springenberg J T, Dosovitskiy A, Brox T and Riedmiller M 2014 Striving for simplicity: the all convolutional net *3rd Int. Conf. on Learning Representations, ICLR 2015—Workshop Track Proc.*
- Sykes J R, Amer A, Czajka J and Moore C J 2005 A feasibility study for image guided radiotherapy using low dose, high speed, cone beam x-ray volumetric imaging *Radiother. Oncol.* **77** 45–52
- Tian Z, Jia X, Yuan K, Pan T and Jiang S B 2011 Low-dose CT reconstruction via edge-preserving total variation regularization *Phys. Med. Biol.* **56** 5949–67
- van Timmeren J E, Leijenaar R T H, van Elmpt W, Reymen B and Lambin P 2017 Feature selection methodology for longitudinal cone-beam CT radiomics *Acta Oncol.* **56** 1537–43
- Verellen D, De Ridder M, Linthout N, Tournel K, Soete G and Storme G 2007 Innovations in image-guided radiotherapy *Nat. Rev. Cancer* **7** 949–60
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: From error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12
- Xu Q, Yu H, Wang G and Mou X 2014 Dictionary learning based low-dose x-ray CT reconstruction *Frontiers of Medical Imaging* (Singapore: World Scientific) pp 99–119
- Xu Y, Bai T, Yan H, Ouyang L, Pompos A, Wang J, Zhou L, Jiang S B and Jia X 2015 A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy *Phys. Med. Biol.* **60** 3567–87
- Yu H and Wang G 2010 A soft-threshold filtering approach for reconstruction from a limited number of projections *Phys. Med. Biol.* **55** 3905–16
- Yuan N, Zhou J and Qi J 2019 Low-dose CT image denoising without high-dose reference images *Proc. SPIE* **11072** 110721C
- Zhang J, Marszałek M, Lazebnik S and Schmid C 2007 Local features and kernels for classification of texture and object categories: A comprehensive study *Int. J. Comput. Vis.* **73** 213–38
- Zhao H, Gallo O, Frosio I and Kautz J 2016 Loss functions for image restoration with neural networks *IEEE Trans. Comput. Imaging* **3** 47–57