Optimal hyperparameters of CBCT-based synthetic CT using U-net deep learning to improve image quality for adaptive radiotherapy in the H&N region

ไฮเปอร์พารามิเตอร์ที่เหมาะสมของการสร้างภาพเอกซเรย์คอมพิวเตอร์สังเคราะห์จากภาพเอกซเรย์คอมพิวเตอร์ลำรังสีกรวยด้วยการเรียนรู้เชิงลึกชนิดยูเน็ตเพื่อเพิ่มคุณภาพของภาพสำหรับการรักษาเทคนิครังสีแบบปรับแผนในบริเวณศีรษะและลำคอ

Tipaporn Prakarnpilas[1], Sangutid Thongsawad[1], Kittipol Dachaworakul[1], Aphisara Deeharing[2], Chirasak Khamfongkhruea[1]

[1]Medical Physics Program, Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok

[2]Radiation Oncology Department, Chulabhorn hospital, Chulabhorn Royal Academy, Bangkok

Corresponding author

Chirasak Khamfongkhruea

Medical Physics Program, Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok, Thailand

906 Khamphaengpet 6 Rd., Lak Si, Bangkok Thailand, 10210

Email: chirasak.kha@cra.ac.th

ทิพาพร ปราการพิลาศ[1], แสงอุทิศ ทองสวัสดิ์[1], กิตติพล เดชะวรกุล[1], อภิสรา ดีหะริง[2], จีรศักดิ์ คำฟองเครือ[1]

[1]หลักสูตรฟิสิกส์การแพทย์ วิทยาลัยแพทยศาสตร์ศรีสวางควัฒน ราชวิทยาลัยจุฬาภรณ์ กรุงเทพมหานคร

[2]งานรังสีรักษามะเร็งวิทยา โรงพยาบาลจุฬาภรณ์ ราชวิทยาลัยจุฬาภรณ์ กรุงเทพมหานคร

ผู้นิพนธ์ประสานงาน

จีรศักดิ์ คำฟองเครือ

หลักสูตรฟิสิกส์การแพทย์ วิทยาลัยแพทยศาสตร์ศรีสวางควัฒน ราชวิทยาลัยจุฬาภรณ์

906 ถ.กำแพงเพชร 6 แขวงตลาดบางเขน เขตหลักสี่ กรุงเทพมหานคร 10210

อีเมล: chirasak.kha@cra.ac.th

Submitted: Nov 4, 2022

Revised: Jan 8, 2023

Accepted: Feb 21, 2023

Abstract

Background: Cone-beam CT (CBCT) imaging is used for adaptive radiation therapy (ART) in head and neck cancer (HNC) due to its more convenient image acquisition and no additional dose. However, CBCT limitations in Hounsfield (HU) accuracy and image quality have emerged for treatment planning. Recently, several studies have proposed using deep learning to generate synthetic CT (sCT) images from CBCT images. However, the quality of images depends on the hyperparameter setting.

Objective: To determine the optimal hyperparameters of the U-net deep learning (DL) for generating sCT images for ART in HNC.

Materials and methods: To generate sCT images, U-net DL with a mean absolute error loss function was used in this study. A total of 3491 image pairs from planning CT (pCT) and CBCT datasets from 40 HNC patients were split into 80% (2976 images from 32 patients) and 20% (515 images from 8 patients) for training and testing, respectively. Each parameter for tuning the U-net model, consisting of learning rates, batch sizes, and epochs, was investigated with various hyperparameter settings in a total of 45 conditions. The best model was assessed using four metrics, including a mean absolute error (MAE) and root mean square error (RMSE) for HU accuracy, peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) for image quality between sCT and pCT images, as well as a training time.

Results: For optimal hyperparameters, we found that the learning rate was set to 1e-3, batch size of 8, and epoch of 200. According to this setting, MAE, RMSE, and PSNR improved from 53.15±40.09, 153.99±79.78, and 47.91±4.98 to 41.47±30.59, 130.39±78.06, and 49.93±6.00, respectively, while SSIM remained constant. The learning rate played an essential role in the training model. All models with various hyperparameters enhanced the reduction of artifacts and noise. The edges of the bone and the soft tissue boundary were clearly visible. The average training time of an optimal hyperparameter was 6 hours and 36.6 minutes (398 ms/step), while it took less than 10 seconds to generate sCT images.

Conclusion: Hyperparameter optimization can improve the quality of sCT images for treatment planning. This study demonstrates the potential of U-net to use CBCT images for ART in HNC.

Keywords: synthetic CT, cone-beam CT, deep learning, U-net, hyperparameters

บทคัดย่อ

หลักการและเหตุผล: การนำภาพเอกซเรย์คอมพิวเตอร์ชนิดลำรังสีกรวย (Cone-beam computed tomography: CBCT) มาใช้ในการรักษาด้วยรังสีแบบปรับแผนสำหรับมะเร็งศีรษะและลำคอ สามารถทำได้ง่ายและไม่มีการเพิ่มปริมาณรังสี แต่อย่างไรก็ตามยังมีข้อจำกัดด้านความถูกต้องของค่า HU และคุณภาพของภาพ เมื่อเร็วๆนี้หลายงานวิจัยได้เสนอให้ใช้การเรียนรู้เชิงลึก (Deep learning: DL) เพื่อสร้างภาพเอกซเรย์คอมพิวเตอร์สังเคราะห์ (Synthetic CT: sCT) ซึ่งคุณภาพของภาพจากวิธีการนี้ขึ้นอยู่กับการตั้งค่าไฮเปอร์พารามิเตอร์

วัตถุประสงค์: เพื่อหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของแบบจำลองการเรียนรู้เชิงลึกชนิดยูเน็ต (U-net) สำหรับการสร้างภาพ sCT ในการรักษาด้วยรังสีแบบปรับแผนสำหรับ HNC

วัสดุและวิธีการ: การศึกษานี้ใช้แบบจำลอง U-net โดยใช้ภาพทั้งหมด 3,491 คู่จากชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์วางแผนการรักษา (pCT) และ CBCT ของ HNC 40 ราย โดยแบ่งข้อมูลเป็น 80% และ 20% สำหรับการฝึกและทดสอบตามลำดับ เพื่อหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม ค่าอัตราการเรียนรู้, ขนาดแบทช์, และรอบในการฝึก ได้มีการปรับแต่งทั้งหมด 45 เงื่อนไข และมีการประเมินคุณภาพของภาพระหว่างภาพ pCT และ sCT ดังนี้ ค่าผิดพลาดสัมบูรณ์เฉลี่ย (MAE) และค่าผิดพลาดกำลังสองเฉลี่ยของราก (RMSE), อัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงสุด (PSNR) และดัชนีความคล้ายคลึงกันของโครงสร้าง (SSIM) รวมถึงเวลาในการฝึกแบบจำลอง

ผลการศึกษา: ไฮเปอร์พารามิเตอร์ที่เหมาะสมประกอบด้วยค่าอัตราการเรียนรู้ที่ 1e-3, ขนาดแบทช์ที่ 8, และรอบการฝึกแบบจำลองที่ 200 ค่า MAE, RMSE และ PSNR มีค่าที่ดีขึ้นจาก 53.15±40.09, 153.99±79.78 และ 47.91±4.98 เป็น 41.47±30.59, 130.39±78.06 และ 49.93±6.00 ตามลำดับ ขณะที่ค่า SSIM มีค่าคงที่ งานวิจัยนี้พบว่าค่าอัตราการเรียนรู้คือค่าที่มีความสำคัญในการฝึกแบบจำลอง โดยแบบจำลอง U-net สามารถลดสิ่งแปลกปลอมและสัญญาณรบกวนได้เป็นอย่างดี เวลาการฝึกแบบจำลองมีค่าเฉลี่ยที่ 6 ชั่วโมง 36.6 นาที (398 มิลลิวินาที/ขั้น) โดยใช้เวลาการสร้างภาพ sCT น้อยกว่า 10 วินาที

ข้อสรุป: การปรับไฮเปอร์พารามิเตอร์ที่เหมาะสมสามารถช่วยเพิ่มคุณภาพของภาพ sCT ให้ดีขึ้น การศึกษานี้แสดงให้เห็นว่าภาพ CBCT สามารถนำมาใช้สำหรับการรักษาด้วยรังสีแบบปรับแผนใน HNC ได้

คำสำคัญ: เอกซเรย์คอมพิวเตอร์สังเคราะห์, เอกซเรย์คอมพิวเตอร์ลำรังสีกรวย, การเรียนรู้เชิงลึก, ยูเน็ต, ไฮเปอร์พารามิเตอร์

J Thai Assoc Radiat Oncol 2023; 29(1): R34 - R51

Introduction

In radiation therapy, head and neck cancer (HNC) is generally treated with complex techniques, such as intensity-modulated radiation therapy (IMRT) and volumetric-modulated arc therapy (VMAT) due to numerous organs at risk (OARs) near the treatment volume. However, these techniques are highly sensitive to uncertainties, especially anatomical changes[1–3]. This results in error in the actual delivery that does not correspond to the planned dose. To solve this problem, adaptive radiation therapy (ART) has been proposed to reduce the effect of anatomical changes[4,5].

Generally, the images used for ART can be achieved by re-scanned a patient from a computed tomography (CT) simulator. This approach requires more staffing and time. Recently, several studies have proposed using cone-beam CT (CBCT) images obtained from treatment rooms for ART[6–8]. However, CBCT images cannot be directly used for treatment planning due to the limitations of CBCT image characteristics, including consistency of CT numbers (Hounsfield units: HU) and image quality in CBCT with scattered artifacts and noise[9–11]. Therefore, before using CBCT images in ART, the improvement of HU accuracy and image quality is needed to meet the requirements of clinical treatment.

Currently, many studies use deep learning (DL) to generate synthetic CT (sCT) images from CBCT images[12–14]. One of the most popular ones is the U-net model, the U-shaped convolutional neural network (CNN) architecture. The U-net has a convolution encoder-decoder (CED) network structure where the encoder and decoder parts are directly skip-connected[15,16]. To achieve the best model in the U-net, the parameter setting is an important step for model training. These parameters, the so-called hyperparameters, significantly impact the model training in terms of predictive image quality, training time, and computer memory space. For the U-net model, the hyperparameters include learning rate, batch size, and epoch. Therefore, this study aims to determine the optimal value of the hyperparameters of the U-net model for generating sCT images for ART in HNC.

Materials and methods

Patient selection and image dataset

This study is a retrospective study. The CBCT and planning CT (pCT) images of 3,720 image pairs from 40 paired pCT datasets and CBCT images of HNC treated with VMAT between January 2018 and December 2021 at the Radiation Oncology Department, Chulabhorn Hospital were enrolled in this study. A dedicated 16-slice helical big-bore CT simulator (Phillips Medical Systems, Andover, MA) and a TrueBeam linear accelerator (Varian Medical Systems, Pala Alto, CA) were used to acquire the pCT and CBCT image datasets, respectively. To minimize the difference in anatomical structure, only the first fraction of all patients' CBCT images before treatment were selected. A voxel spacing was 1.00×1.00×3.00 mm3 and

0.51×0.51×2.00 mm3 for pCT and CBCT images, respectively. Both dimensions of the pCT and CBCT images were 512×512.

Image preparation

Rigid registration was performed to align the pCT images with the CBCT images using an open-sourced registration graphical user interface (OpenREGGUI), a MATLAB-based medical image processing software. During image registration, the number of pixels pCT images was resampled to the CBCT images (0.51×0.51×2.00 mm3). Due to the incomplete field of view (FOV) of CBCT images in HNC, images with uncompleted body outlines were not included in the image datasets. Furthermore, a structure outside the body was created and assigned to the air density (-1000 HU) for both pCT and CBCT images.

Model generation

The proposed model in this study is based on the U-net architecture, as shown in **Figure 1**. The network model was developed using Keras and TensorFlow 2.9.0 with Python version 3.8, NVIDIA CUDA® Deep Neural Network library version 8.1, and Compute Unified Device Architecture version 11.2. All experiments were implemented on an NVIDIA Quadro RTX 8000 GPU with 48 GB of memory (training was done in a JetBrains PyCharm anaconda environment).

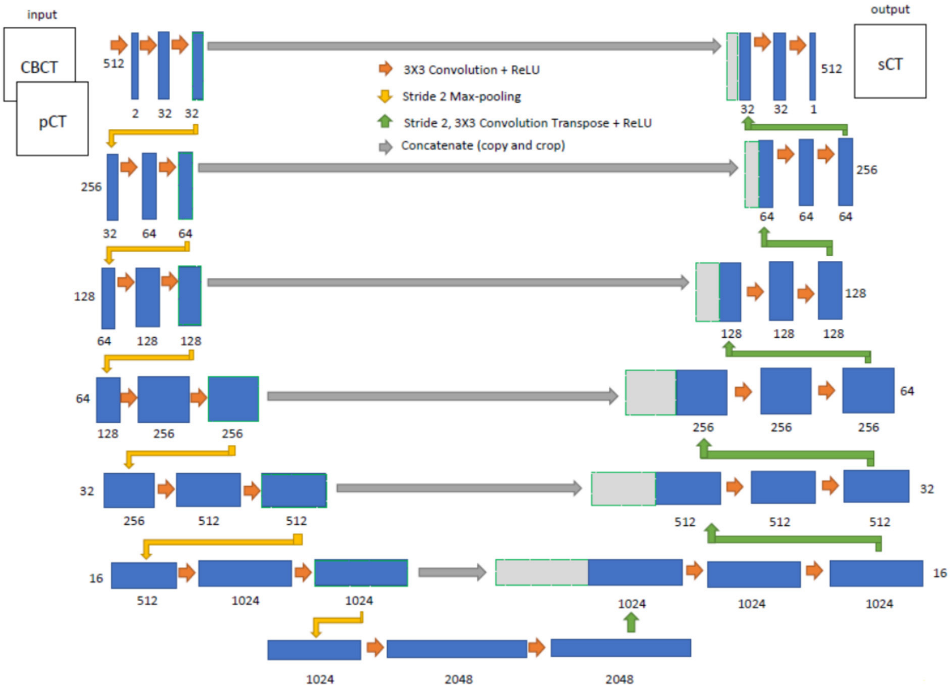


Figure 1 U-net network architecture.

Model architecture

The U-net model **(Figure 1)**, which has a CED network structure, comprises two networks: encoders and decoders[16]. Because of its excellent performance, CED has been widely used in the DL literature. During the encoding phase, low-level feature maps are downsampled to high-level ones. In the decoding phase, the prediction image is constructed by up-sampling, utilizing the transposed convolutional layer, and the high-level feature maps are converted to low-level feature maps. The encoder network employs a set of 2D convolution filters with normalization (batch normalization)[17], a non-linear activation function (rectified linear unit: ReLU), and maximum pooling for identifying image features. The decoder network uses transposed convolutional layers with concatenating layers and convolutional layers with a ReLU[18] for combining features and spatial information.

The inputs of the model are pairs of CBCT and pCT images. A max-pooling layer and two convolutional layers are placed before each of the six down-sampling blocks in the encoder, which is the 3×3 convolution kernel with a ReLU as the activation function. The convolutional layers (down-sampling blocks) have a beginning feature number of 32 and grow by 2 with each following block number. The pooling size in the max-pooling layers is 2×2. As a result, as the encoder continuously increases the depth, the image size decreases from 512×512x2 to 4×4×1024. With a kernel size of 3×3 and feature numbers of both 2048, the final two encoder convolutional layers concentrate the input image information into 4×4×2048.

For the decoder, two convolutional layers follow the up-sampling blocks (convolution transpose) layers of the decoder. One transposed convolutional layer, one concatenate layer, and two convolutional layers are present in each up-sampling block. To obtain more precise location information at the same level, the concatenates layer combines the feature maps from the encoder with the output of the transposed convolution layers. Following the concatenates layer, two convolutional layers let the model develop a more exact output. The transposed convolutional layers have a kernel size of 2×2 and a stride size of 2×2. The kernel size of each convolutional layer in the decoder is 3×3. The convolutional layers (up-sampling blocks) start with a feature number of 2048 and decrease by 2 with each following block number. Then, apply the kernel size of 3x3 and the corresponding feature numbers of 32 and 1 to keep the size of sCT images generated as input images. By reducing the loss error from an average absolute difference between the CBCT to sCT images and the corresponding pCT images, the mean absolute error (MAE) loss function is optimized.

Model training

The 3,491 image pairs of pCT and CBCT image datasets from 40 HNC patients were split into 80% (2,976 images from 32 patients) and 20% (515 images from 8 patients) for training and testing, respectively. For the training stage, 20% of the dataset was used for validation, with a different number of slices depending on each batch size.

To determine the optimal hyperparameter, each parameter for tuning the U-net model, consisting of learning rates, batch sizes, and epochs, was varied, as shown in **Figure 2**. The detail of the parameters are as follows:

1) The learning rates

The learning rate is a hyperparameter that controls how much weight of the neural network is adjusted in one step of the training by setting the learning rate through the Stochastic Gradient Descent Algorithm Adam (Adaptive Moment Estimation) Optimizer[19,20].

2) Batch sizes

Batch size is the sample number of training datasets divided into the number of batches in one epoch[20,21].

3) Epochs

The number of epochs is a hyperparameter specifying how often the learning algorithm will iterate over the training dataset[22].

Model testing

After model training, eight independent CBCT patient datasets were fed to the model to generate sCT images. The performance of predictive sCT images generated with different hyperparameters was evaluated in terms of quality metrics and time. For quality metrics, pCT (ground truth images) and sCT images were measured and compared as follows:

1) HU accuracy

The image intensity was evaluated in terms of HU differences between the pCT and

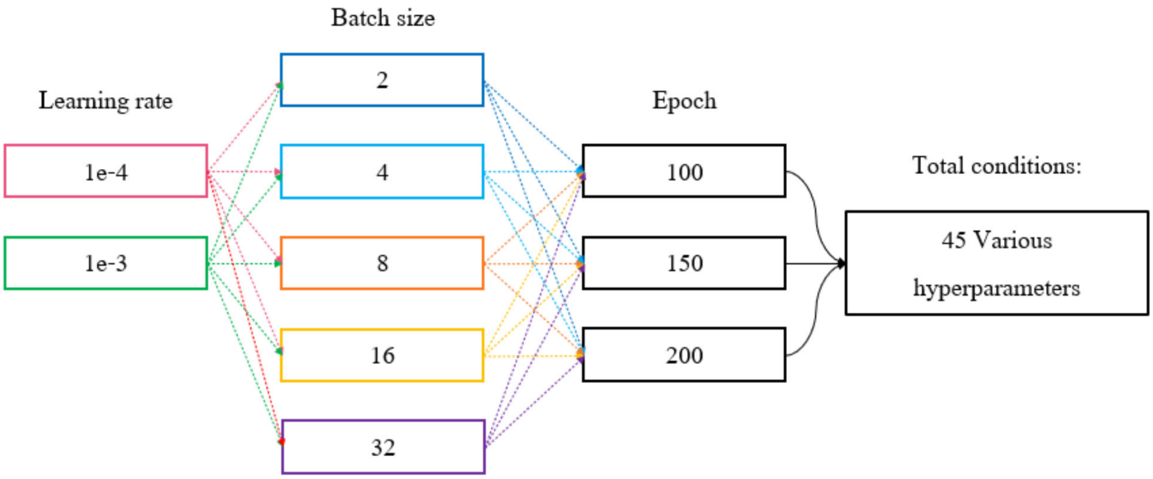


Figure 2 Setting conditions by adjusting various hyperparameters.

the sCT images. Mean absolute error (MAE) and root mean square error (RMSE) were used in this study. Both terms can be described as follow:

MAE: the mean absolute difference of all the pixel values between two images in which the image with lower MAE values shows more accurate pixel-wise HU values.

$$MAE(pCT, sCT) = \frac{1}{N}\sum_{i=1}^{N} |pCT_i - sCT_i|$$

RMSE: the root mean square difference between the two images. Lower RMSE indicates a lower noise level in the image because pCT images are less noisy than CBCT images.

$$RMSE(pCT, sCT) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |pCT_i - sCT_i|^2}$$

Where N = the total number of pixels

pCTi and sCTi = the HU in each pixel of the pCT and sCT images, respectively.

2) Image quality

The performance of the image quality of sCT was determined via two image quality accepted metrics: peak-signal-noise-ratio (PSNR) and structural similarity index (SSIM). The details of each term can be described as follows:

PSNR: the ratio of a maximal pixel value to the noise of the images degrades the representation of its quality. This term can calculate using the ground truth (pCT) image's maximum intensity value and the mean squared error. PSNR is the standard metric used to evaluate image quality for noise reduction.

$$PSNR(pCT, sCT) = 20 \cdot log_{10}\left(\frac{MAX(pCT)}{RMSE(pCT, sCT}\right.$$

Where MAX = the maximum intensity for pCT and sCT images.

SSIM: a standard metric for comparing the structural similarity of two images. The SSIM close to 1 indicates similar images.

$$SSIM(pCT, sCT) = \frac{(2\mu_{sCT}\mu_{pCT} + C_1) + (2\sigma_{sCT,pCT} + C_2)}{(\mu_{sCT}{}^2 + \mu_{pCT}{}^2 + C_1) + (\sigma_{sCT}{}^2 + \sigma_{pCT}{}^2 + C_2)}$$

Where μsCT and μpCT = the mean values of HU of the sCT and pCT images, respectively.

σsCT and σpCT = the variance of HU values of the sCT and pCT images, respectively.

σsCT,pCT = the covariance, the parameters C1 = (k1L)2 and C2 = (k2L)2are two variables to stabilize the division with weak denominators, L is the range of HU values in the CT image. k1 = 0.01, k2 = 0.02.

Results

Table 1 shows the MAE, RMSE, PSNR, SSIM, and training time from 45 different hyperparameter settings. There was no pattern relationship for MAE, RMSE, and PSNR when learning rate and batch size were increased. At the same learning rate and batch size, MAE and RMSE tended to decline, PSNR increased and SSIM remained constant as epoch increased.

Table 1 MAE, RMSE, PSNR, SSIM, and training time for various hyperparameters from eight independent test datasets based on 3D images compared with pCT as ground truth images.

Learning rate	Batch size	Epoch	MAE (HU)	RMSE (HU)	PSNR (dB)	SSIM	Training time (hr.)
	CBCT		53.15 ± 40.09	153.99 ± 79.78	47.91 ± 4.98	0.97 ± 0.02	-
1e-4	2	100	45.82 ± 27.85	142.69 ± 81.54	48.94 ± 5.64	0.98 ± 0.02	5.00
		150	44.75 ± 27.23	141.02 ± 79.60	48.99 ± 5.55	0.98 ± 0.02	7.50
		200	43.90 ± 26.77	141.13 ± 78.74	48.94 ± 5.47	0.98 ± 0.02	10.00
	4	100	43.81 ± 29.20	137.70 ± 80.30	49.34 ± 5.80	0.98 ± 0.02	3.89
		150	44.03 ± 31.03	139.75 ± 83.30	49.31 ± 5.97	0.98 ± 0.02	5.78
		200	44.78 ± 30.30	139.28 ± 85.32	49.47 ± 6.193	0.98 ± 0.02	7.78
	8	100	51.22 ± 36.87	148.31 ± 89.09	48.83 ± 6.03	0.98 ± 0.02	3.25
		150	50.33 ± 35.24	147.83 ± 88.09	48.82 ± 5.96	0.98 ± 0.02	4.88
		200	47.80 ± 33.00	144.38 ± 85.68	49.00 ± 5.93	0.98 ± 0.02	6.50
	16	100	45.26 ± 31.67	133.94 ± 76.57	49.49 ± 5.65	0.98 ± 0.02	2.67
		150	45.91 ± 32.00	130.72 ± 87.18	49.29 ± 5.75	0.98 ± 0.02	4.00
		200	43.52 ± 29.53	130.39 ± 73.25	49.65 ± 5.52	0.98 ± 0.02	5.33
	32	100	48.13 ± 33.17	140.06 ± 84.07	49.32 ± 6.02	0.98 ± 0.02	2.53
		150	47.21 ± 32.04	135.77 ± 80.83	49.50 ± 6.00	0.98 ± 0.02	3.79
		200	46.43 ± 32.39	136.62 ± 81.58	49.51 ± 5.98	0.98 ± 0.02	5.06
1e-3	2	100	46.21 ± 35.03	143.24 ± 87.66	49.23 ± 6.12	0.98 ± 0.02	5.19
		150	47.57 ± 34.79	143.25 ± 86.92	49.18 ± 6.11	0.98 ± 0.02	7.79
		200	44.97 ± 32.15	138.21 ± 80.99	49.33 ± 5.83	0.98 ± 0.02	10.39
	4	100	42.94 ± 31.53	133.71 ± 79.90	49.70 ± 5.99	0.98 ± 0.02	3.92
		150	44.46 ± 31.91	137.05 ± 78.97	49.32 ± 5.70	0.98 ± 0.02	5.88
		200	43.52 ± 31.19	132.40 ± 79.23	49.80 ± 6.00	0.98 ± 0.02	7.83
	8	100	42.81 ± 31.58	135.20 ± 78.57	49.48 ± 5.77	0.98 ± 0.02	3.31
		150	45.21 ± 30.61	133.89 ± 76.10	49.47 ± 5.60	0.98 ± 0.02	4.96
		200	41.47 ± 30.59	130.39 ± 78.06	49.93 ± 6.00	0.98 ± 0.02	6.61
	16	100	42.56 ± 30.54	139.30 ± 73.20	48.83 ± 5.07	0.98 ± 0.02	2.67
		150	44.41 ± 31.21	130.26 ± 86.67	48.80 ± 5.22	0.98 ± 0.02	4.00
		200	46.23 ± 32.79	139.68 ± 74.89	48.88 ± 5.20	0.98 ± 0.02	5.33
	32	100	49.20 ± 33.40	152.28 ± 89.39	48.49 ± 5.85	0.98 ± 0.02	2.56
		150	47.06 ± 29.97	148.45 ± 83.19	48.51 ± 5.50	0.98 ± 0.02	3.38
		200	48.34 ± 28.69	150.97 ± 86.81	48.47 ± 5.69	0.98 ± 0.02	5.11

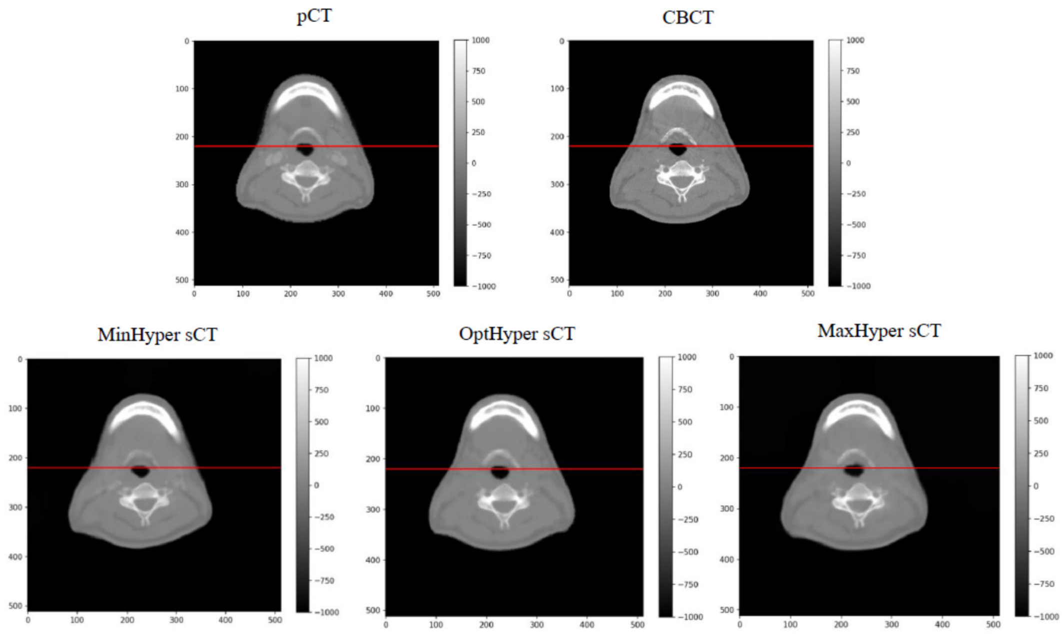
Abbreviations: CBCT=cone-beam computed tomography; MAE = mean absolute error; RMSE = root-mean-square error; SSIM = structural similarity index; PSNR = peak signal-to-noise ratio; HU = Hounsfield units; dB = decibel; hr = hours.

The best sCT images in this study indicated the optimal hyperparameter (OptHyper) when the learning rate was set to 1e-3, batch size was 8, and epoch was 200. When compared to CBCT images, these metrics improved the values of MAE, RMSE, and PSNR, from 53.15 ± 40.09, 153.99 ± 79.78, and 47.91 ± 4.98 to 41.47 ± 30.59, 130.39 ± 78.06, and 49.93 ± 6.00, respectively. On the other hand, sCT was generated by setting the hyperparameter to the lowest value, called the minimal hyperparameter (MinHyper) setting, with a set learning rate of 1e-4, batch size of 2, and epoch of 100. With this setting, the MAE, RMSE, and PSNR were 45.82 ± 27.85, 142.69 ± 81.54, and 48.94 ± 5.64, respectively. The highest value of hyperparameter configuration, called maximal hyperparameter (MaxHyper) setting, with a set learning rate of 1e-3, batch size of 32, and epoch of 200. This condition provided the MAE, RMSE, and PSNR of 48.34 ± 28.69, 150.97 ± 86.81, and 48.47 ± 5.69, respectively. The SSIM of sCT images from all hyperparameter conditions was 0.98 ± 0.02, higher than 0.97 ± 0.02 for CBCT images.

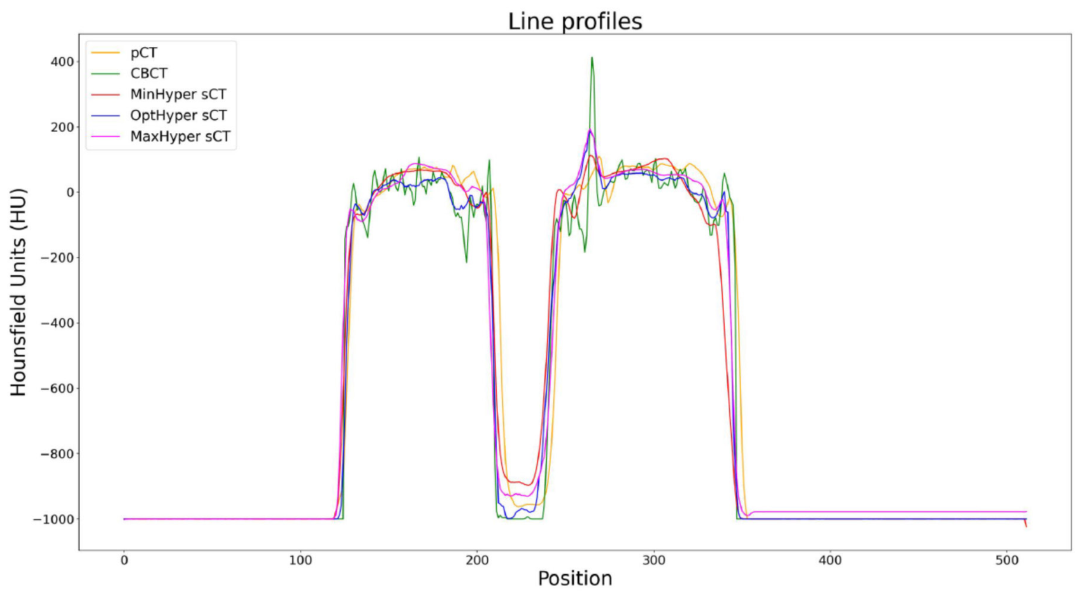
Figure 3 depicts the axial image and HU line profile of pCT, CBCT, and three sCT images obtained from the test datasets (MinHyper, OptHyper, and MaxHyper). This profile crosses through bone and soft tissue structures. The blue line represents the HU profile of OptHyper sCT images, setting optimal hyperparameters as a learning rate of 1e-3, batch size of 8, and epoch 200. The sCT images improved in HU value, with an increase in HU smoothness and accuracy close to pCT images in the area of soft tissue, bone, and closer to the body boundary. In contrast, both MinHyper sCT images, represented by the red line as setting minimal hyperparameters, and the magenta line representing of MaxHyper sCT images as setting maximal hyperparameters, were less close to pCT images in any area. However, all sCT images were of better quality than the HU profile in the CBCT image, which was noisy and inaccurate.

From the results in **Table 1**, the training time for the minimal hyperparameters setting was 5 hours, the training time for the maximal hyperparameters setting was 5 hours and 7 minutes, and the training time for the optimal hyperparameters setting was 6 hours and 36.6 minutes. Most of the training time was reduced to 9%, 8%, and 53%, respectively, when the batch size increased from 2 to 4, 8 to 16, and 16 to 32, while only the batch size increased from 4 to 8, increasing 3% training time when the learning rate had risen from 1e-4 to 1e-3. This demonstrated less training time when the batch size and learning rate were higher.

Three axial slices of pCT, CBCT, and OptHyper sCT are shown in **Figure 4**. The OptHyper sCT images improved HU values close to pCT images while preserving the geometrical information of CBCT images. Moreover, The OptHyper sCT images could also reduce streak artifacts found in CBCT images (Red arrow in Figure 4).



(a)



(b)

Figure 3 The axial images (a) and HU line profiles (b) of pCT, CBCT, and three sCT images obtained test datasets. Comparing of the pCT (orange), CBCT (green), MinHyper sCT (red), OptHyper sCT (blue), and MaxHyper sCT (magenta).

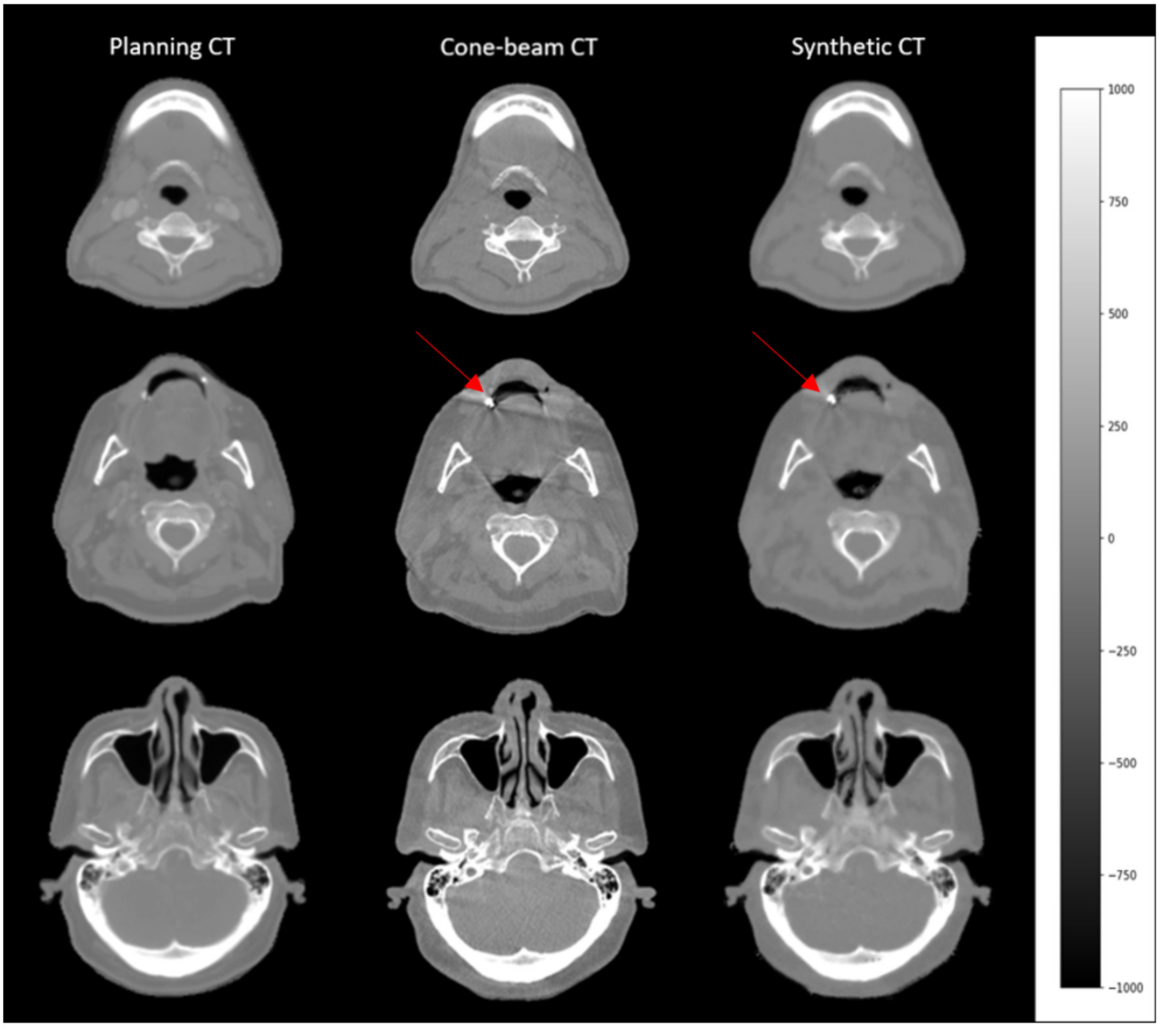
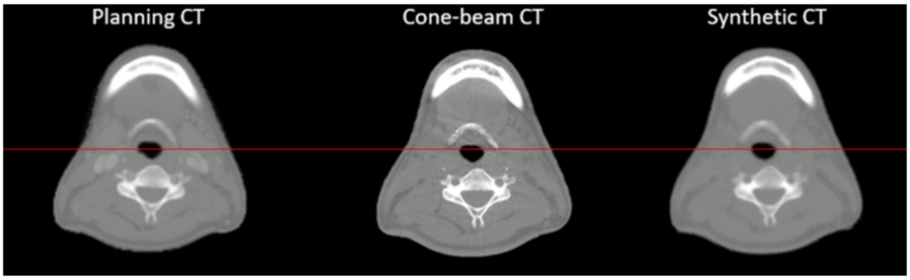
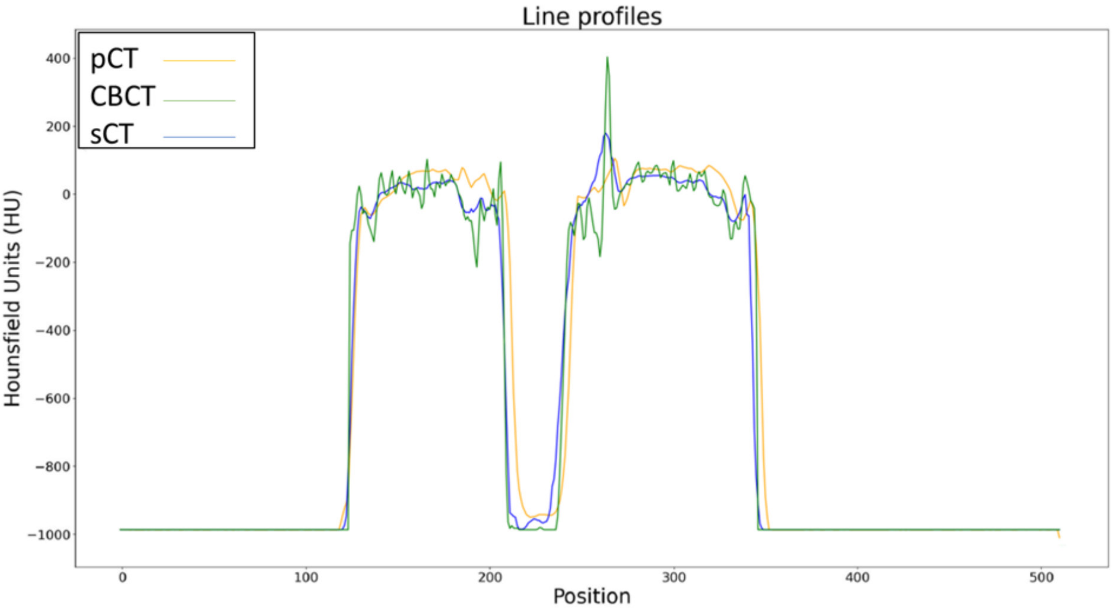


Figure 4 Three axial slices of pCT, CBCT, and OptHyper sCT images with -1000 to 1000 HU window width.

Figure 5 shows the HU line profile through the body of pCT (orange), CBCT (green), and OptHyper sCT (blue) images. According to the line profiles, OptHyper sCT images had an HU line closer to pCT images than the CBCT images, especially at the body boundaries. There was a peak reduction and a smoothness in OptHyper sCT images. At the soft tissue-air interface, the OptHyper sCT images provided a smooth edge than the CBCT images.



(a)



(b)

Figure 5 The axial images (a) and HU line profiles (b) of pCT (orange), CBCT (green), and OptHyper sCT (blue) images obtained from test datasets.

Discussion

The purpose of this study was to determine the optimal value of the hyperparameters of the U-net model for generating sCT images for ART in HNC. The head and neck region are extremely sensitive to anatomical variations, such as weight loss and tumor shrinkage. This study concentrated on three hyperparameters: learning rate, batch size, and epochs. There were four acceptable metrics for evaluating the quality of sCT with a different set of hyperparameters. These evaluation metrics were assessed in terms of MAE and RMSE for HU accuracy and PSNR and SSIM for image quality. Moreover, the time during model training was collected as well. Since acquisition at the same time for pCT and CBCT images was difficult, first-fraction CBCT images of HNC patients were used and then registered to pCT images for generating synthesis CT images from a model based on a U-net architecture for anatomical structure mismatch.

The advantage of the U-net model is reducing global scattering and improving local HU due to applying in the spatial domain of an image both global and local features[13,23]. To resolve this issue, a trained model using MAE as a loss function was used to learn the anatomical structures of CBCT images through image preprocessing and the HU value of pCT images. Furthermore, the U-net can generate an output the same size as its input. However, the disadvantage of the model is the unknown number of optimal depths of an encoder-decoder network based on the task complexity for training. In addition, there is a robust theory for the design of skip connections between encoder-decoder networks operating at the same level. Because these feature maps are semantically different, it is not guaranteed that they are the perfect fit for feature fusion[24].

This study investigated only three hyperparameters, including the learning rate, batch size, and epoch. We did not include weight decay due to less effect on model performance[25]. Furthermore, these agree with Bergstra et al.[26], who compared a careful combination of manual and grid searches of deep belief networks with varied these three hyperparameters.

According to the results of the study, it was found that the learning rate had a significant impact on the image quality in terms of MAE, RMSE, and PSNR, except SSIM, while the batch size and epoch had less effect on the image quality. Where SSIM is invariant, this may be due to the high accuracy of alignment of pCT and CBCT images by rigid image registration and selection of images with FOV encompassing body parts. According to **Table 1**, the learning rate impacts the image quality of almost 45 setting conditions across various hyperparameters for the training model to generate sCT. Obviously, when increasing the learning rate to 10e-3, almost all the results had better image quality than the learning rate of 10e-4.

The conditions with the learning rate of 1e-3, the batch size of 8, and the epoch of 200 were the optimal hyperparameters for the U-net model to generate sCT due to the highest PSNR and lowest MAE and RMSE. All various hyperparameters for sCT generation enhanced the reduction of artifacts and noise. The edges of the bone and the soft tissue boundary were clearly visible. The sCT images can improve HU accuracy and image quality compared to the original CBCT images.

According to the HU line profiles of pCT, CBCT, and sCT images with three different hyperparameter settings, OptHyper sCT images had an HU line closer to pCT images than the others, especially at the body boundaries. At the thyroid cartilage region, all sCT images showed better intensity with the reduction from the peak of CBCT images. The smoothness of soft tissue had improved in all sCT images, the closest to pCT images being MinHyper sCT, MaxHper sCT, and OptHyper sCT images, respectively. Although when the HU line crosses through the surrounding edge of the air region, such as the outside of the body and subglottic larynx, it indicates that sCT images had rounded corners close to pCT images. In contrast, CBCT images had square corners, the closest to pCT images being OptHyper sCT, MinHyper sCT, and MaxHper sCT images, respectively. Furthermore, the MinHyper sCT images had a tail at the bottom of the image. In contrast, the MaxHper sCT images had an inaccurate HU value, especially in air density at the end edge of the body part (intensity higher than -1000 HU). As a result, OptHyper sCT was the best hyperparameter, with an HU profile line that crossed soft tissue, bone, and body boundaries closer to pCT images than the other MinHyper sCT, MaxHper sCT, and the original CBCT images. It also provides OptHyper sCT images with good image detail of bone structure and structure boundaries, soft tissue contrast, and reduces artifacts and noise in CBCT images.

The average training time of an optimal hyperparameter was 6 hours and 36.6 minutes (398 ms/step), while it took less than 10 seconds to generate sCT images. From the results, the consumed training time depends on the number of images and the setting of the hyperparameter value. The model's answer converges slowly when the learning rate is too low. It takes longer to train because each trained model changes the weight and bias incrementally, requiring several epochs. On the other hand, when the learning rate is set too high, the model's answer stays consistent with the correct or expected answer because each training cycle improves weight and bias rapidly, requiring fewer epochs. The batch size determination directly affected the train speed. The larger the batch size, the faster the training model, but the more memory needs to be processed.

This study reported similar results to Kida et al.[12]. They used the same U-net model structure without ReLU in transpose convolution and used MAE as a loss function in the pelvic regions. Furthermore, our results showed better MAE, RMSE, PSNR, SSIM, and train time than Li et al.'s work[27], which developed the U-net architecture with a residual block using MAE as the loss function in the H&N region. However, the study of Chen et al.[13] provided better HU accuracy metrics results than ours because of an improved loss function that combines structure dissimilarity and MAE. Although this study shows better image quality, the training time was shorter.

Although our results showed improvements in HU accuracy and image quality, there are several limitations to keep in mind. Firstly, the FOV of CBCT images was limited to 250 mm, which cannot provide the anatomy information, especially the low-risk target volume of HNC in the shoulder region. Secondly, anatomical changes may occur between a pair of pCT-CBCT images due to the different times of image acquisition (taken on separate days). This could impact the model training due to using a pair image of the U-net algorithm. Finally, this study had a small sample size of image data for training and testing, and a larger sample size might lead to better results.

Further research suggests that these limits could be improved by creating an advanced loss function for optimization, such as extracting image features more precisely. Furthermore, using an unsupervised model or employing another DL algorithm, such as the generative adversarial network (GAN)[14] might be improved the model due to the difference in using the unpair images for model training.

Acknowledgment

The authors would like to thank Dr. Monchai Phonkrai, Dr.Pasit Jarutatsanangkoon, and Mr. Patiparn Kummanee for a consult with python programming. The authors declare no conflict of interest.

References

1. Spencer S. Head and neck cancers: Ad vantages of advanced radiation therapy and importance of supportive care. J Natl Compr Canc Netw 2018;16:666-9.
2. Belshaw L, Agnew CE, Irvine DM, Rooney KP, McGarry CK. Adaptive radiotherapy for head and neck cancer reduces the requirement for rescans during treatment due to spinal cord dose. Rad Oncol 2019;14:1-7.
3. Yeh SA. Radiotherapy for head and neck cancer. Semin Plast Surg 2010;24: 127–136.
4. Castadot P, Lee JA, Geets X, Grégoire V. Adaptive radiotherapy of head and neck cancer. Sem in Rad Oncol 2010;20:84–93.
5. Schwartz DL. Current progress in adaptive radiation therapy for head and neck cancer. Curr Oncol Rep 2012;14: 139–47.
6. Sonke JJ, Aznar M, Rasch C. Adaptive radiotherapy for anatomical changes. Semin Radiat Oncol. 2019;29:245–57.
7. Posiewnik M, Piotrowski T. A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer. Phys Medi 2019;59:13–21.
8. Søvik Å, Rødal J, Skogmo HK, Lervåg C, Eilertsen K, Malinen E. Adaptive radiotherapy based on contrast enhanced cone beam CT imaging. Acta Oncol 2010; 49:972–7.

9. Buchanan A, Mcdavid D. Characterization and correction of cupping effect artefacts in cone beam CT. Dentomaxillofac Radiol. 2012;41:217–23.
10. Jaju P, Jain M, Singh A, Gupta A. Artefacts in cone beam CT. Open J Stomatol 2013;03:292–7.
11. Schulze R, Heil U, Groß D, Bruellmann D, Dranischnikow E, Schwanecke U, et al. Artefacts in CBCT: a review. Dentomaxillofac Radiol 2011;40:265–73.
12. Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, et al. Cone beam computed tomography image quality improvement using a deep convolutional neural network. Cureus 2018;10:e2548.
13. Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. Med Phys 2020;47:1115–25.
14. Barateau A, de Crevoisier R, Largent A, Mylona E, Perichon N, Castelli J, et al. Comparison of CBCT-based dose calculation methods in head and neck cancer radiotherapy: from Hounsfield unit to density calibration curve to deep learning. Med Phys 2020;47:4683–93.
15. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA, Bottou L. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 2010;11:3371-408.
16. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408-19.
17. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Int Conf Mach Learn 2015;37:448-56.
18. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. Int Conf Mach Learn 2010;807–14.
19. Smith LN. A disciplined approach to neural network hyper-parameters: Part 1-- learning rate, batch size, momentum, and weight decay. US Nav Res Lab Tech Rep 2018:1-21
20. Rashmi, Ghose U, Gupta M. Comparative design analysis of optimized learning rate for convolutional neural network. In: Sharma H, Saraswat M, Kumar S, Bansal JC, editors. Intelligent Learning for Computer Vision. Springer Singapore; 2021;339–52.
21. Qian X, Klabjan D. The impact of the mini-batch size on the variance of gradients in stochastic gradient descent. arXiv 2020:1-11.
22. Afaq S, Rao S. Significance of epochs on training a neural network. Int J Scien Tech Res 2020;9:485-8.

23. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Int Conf Med Imag Comput Com-assist Int 2015;234-41.
24. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Img 2020;39:1856–67.
25. Krogh A, Hertz JA. A simple weight decay can improve generalization. Int Conf Neur Inform Pros Syst. 1991;4:950–7.
26. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Lear Res 2012;13:281-305.
27. Li Y, Zhu J, Liu Z, Teng J, Xie Q, Zhang L, et. al. A preliminary study of using a deep convolution neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma. Phys Med Biol 2019;64:145010.