Received: 29 April 2022 | Revised: 1 September 2022 | Accepted: 7 September 2022

DOI: 10.1002/mp.15986

RESEARCH ARTICLE

MEDICAL PHYSICS

SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning

Sheng Zhao[1] | **Changran Geng**[1,2] | **Chang Guo**[3] | **Feng Tian**[1] | **Xiaobin Tang**[1,2]

[1]Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, People's Republic of China

[2]Key Laboratory of Nuclear Technology Application and Radiation Protection in Astronautics (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, Nanjing, People's Republic of China

[3]Department of Radiation Oncology, Jiangsu Cancer Hospital, Nanjing, People's Republic of China

Correspondence
Changran Geng and Xiaobin Tang, Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, People's Republic of China.
Email: gengchr@nuaa.edu.cn; tangxiaobin@nuaa.edu.cn

Funding information
National Key Research and Development Program, Grant/Award Number: 2022YFE0107800; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20220132; Nanjing University of Aeronautics and Astronautics, Grant/Award Number: xcxjh20210615

Abstract

Purpose: Despite the significant physical differences between magnetic resonance imaging (MRI) and computed tomography (CT), the high entropy of MRI data indicates the existence of a surjective transformation from MRI to CT image. However, there is no specific optimization of the network itself in previous MRI/CT translation works, resulting in mistakes in details such as the skull margin and cavity edge. These errors might have moderate effect on conventional radiotherapy, but for boron neutron capture therapy (BNCT), the skin dose will be a critical part of the dose composition. Thus, the purpose of this work is to create a self-attention network that could directly transfer MRI to synthetical computerized tomography (sCT) images with lower inaccuracy at the skin edge and examine the viability of magnetic resonance (MR)-guided BNCT.

Methods: A retrospective analysis was undertaken on 104 patients with brain malignancies who had both CT and MRI as part of their radiation treatment plan. The CT images were deformably registered to the MRI. In the U-shaped generation network, we introduced spatial and channel attention modules, as well as a versatile "Attentional ResBlock," which reduce the parameters while maintaining high performance. We employed five-fold cross-validation to test all patients, compared the proposed network to those used in earlier studies, and used Monte Carlo software to simulate the BNCT process for dosimetric evaluation in test set.

Results: Compared with UNet, Pix2Pix, and ResNet, the mean absolute error (MAE) of self-attention ResUNet (SARU) is reduced by 12.91, 17.48, and 9.50 HU, respectively. The "two one-sided tests" show no significant difference in dose-volume histogram (DVH) results. And for all tested cases, the average 2%/2 mm gamma index of UNet, ResNet, Pix2Pix, and SARU were 0.96 ± 0.03, 0.96 ± 0.03, 0.95 ± 0.03, and 0.98 ± 0.01, respectively. The error of skin dose from SARU is much less than the results from other methods.

Conclusions: We have developed a residual U-shape network with an attention mechanism to generate sCT images from MRI for BNCT treatment planning with lower MAE in six organs. There is no significant difference between the dose distribution calculated by sCT and real CT. This solution may greatly simplify the BNCT treatment planning process, lower the BNCT treatment dose, and minimize image feature mismatch.

KEYWORDS
boron neutron capture therapy, deep learning, synthetical computerized tomography

1 | INTRODUCTION

Boron neutron capture therapy (BNCT) is a kind of binary radiotherapy modality, which is featured on the cell targeting by the tumor-targeting drugs (i.e., ^{10}B drug) and the irradiation of thermal/epithermal neutrons.[1,2] Treatment planning is essential for BNCT. Typically, the planning system uses computed tomography (CT) images for radiation dose calculation. However, taking CT pictures exposes patients to ionizing radiation to some extent. Multiple exposures considerably raise the risk of cancer, with particularly dangerous consequences for pregnant women, patients with hyperthyroidism, and those with impaired liver and kidney functions. On the other hand, the acquisition of boron concentration distribution is one of the crucial phases in BNCT, as boron dose is the main component of BNCT dose composition. Magnetic resonance imaging (MRI) has some advantages over other imaging modalities; for example, less invasiveness, more versatility with fewer restrictions of a boron–gadolinium compound, and can also provide functional and morphological information without using radiation, which is helpful for treatment planning.[3–8]

However, MRI does not provide direct information for dose calculation as CT imaging before BNCT therapy. Despite the significant physical differences between MRI and CT, the high entropy of MRI data indicates the existence of a surjective transformation from MRI to CT image. The existing methods to generate synthetical computerized tomography (sCT) by MRI can be divided into three categories: density-based partitioning, atlas-based, and machine learning (ML) methods (including classical methods and deep learning methods). If sCT images can be synthesized based on MRI to calculate dose deposition, CT scanning can be omitted, the clinical workflow of BNCT can be simplified, and the mismatch between boron drug distribution images and CT images can be avoided, which can significantly reduce the system uncertainty.

The UNet proposed by Ronneberge and coworkers has been used for tasks such as brain, head and neck, abdomen, and so forth, and has achieved encouraging results.[9–16] Another image conversion network based on conditional GAN (cGAN) named Pix2Pix is also commonly utilized.[17–20] Because cGAN may employ condition information to guide image generation, the input image can be used as a condition in image transformation to learn the mapping between the input and output samples. It has been widely used in background removal, image repair, style transfer, and other fields.

However, there is no specific optimization of the network itself in previous MRI/CT translation works, resulting in huge mistakes in details such as the skull margin and cavity edge. These errors might have moderate effect on conventional radiotherapy. But for BNCT, the skin dose will be a critical component of the dose composition.[21] Therefore, in this work, we proposed a new deep attention mechanism network using the deep learning method to generate CT images directly from T1-weighted MRI, to obtain a better edge prediction effect. Furthermore, we evaluated the performance impact of the synthesized CT on BNCT from dosimetry using Monte Carlo simulations.

2 | MATERIALS AND METHODS

2.1 | Data acquisition and preprocess

We collected MRI and CT data from 104 patients (male: 43, female: 61) with brain malignancies at the Jiangsu Cancer Hospital from March 2021 to March 2022. The ages range from 18 to 83 years old, the median age of all patients is 55. All the data have been authorized by the Ethics Committee of Jiangsu Cancer Hospital (ID: 2022011). The interval between CT and MR scans for each patient is within 1 week. Simultaneously, pretraining data from the Retrospective Image Registration Evaluation project (RIRE)[22] is used. The turning angle of the MRI excitation pulse is 10°, TE is 4 ms and TR is 10 ms. The MRI image dimension ranged from $22 \times 256 \times 256$ to $117 \times 256 \times 256$, and the corresponding pixel size from $6 \times 0.4 \times 0.4$ to $0.3 \times 0.4 \times 0.4$ mm. The CT images are obtained with tube voltage of 120 kVp and pixel size of $1 \times 0.4 \times 0.4$ mm.

We used T1-weighted MRI and CT images as the input and output of the neural network, respectively, because T1-weighted MRI shows the difference in tissue T1 relaxation (longitudinal relaxation), which is more conducive to observing the anatomical structure.[23] The training set and verification set consisted of 94 patients, and the test set consisted of the remaining 10 patients. To get more patients into the validation phase, we separated the original data (excluding the test set) into five groups, utilized each subset data as validation set, and the remaining four sets of subset data as training set, and trained five times in total to get the five-fold cross-validated errors.

As the original MRI and CT images are not registered, the Insight Segmentation and Registration Toolkit (ITK) is used to convert the original MHD files into NIfTI format.[24] We performed symmetric normalization[25] (rigid + affine + deformable transformation, with mutual information as optimization metric) for MRI and CT, and the registration is completed when the maximum number of iterative steps or convergence conditions are met. To avoid the intensity inhomogeneity of MRI affecting the prediction performance of the neural network, we use N4 bias field correction algorithm to further correct MRI. At the same time, the Otsu threshold segmentation algorithm is used to extract the mask of the head MRI and separate the head region from other nonanatomical background regions in MRI and CT images. In the

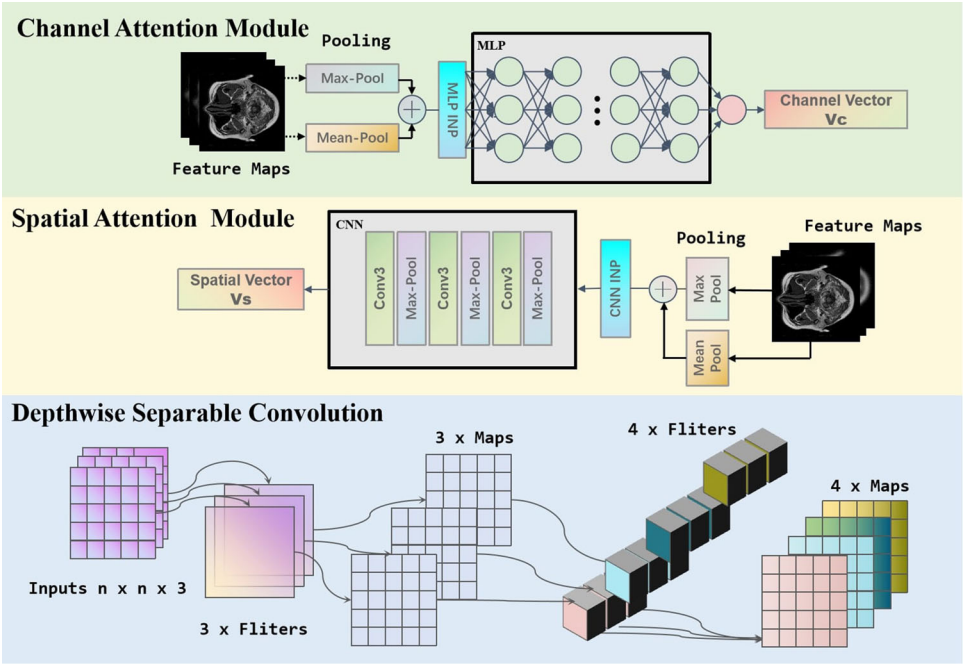

FIGURE 1 Schematic diagram of the channel attention module, spatial attention module, and depth separable convolution (from top to bottom)

CT image, set all pixels outside the masked area to air (CT value equal to −1000 HU) to avoid the adverse effect of treatment bed, bandage, and other irrelevant components on model training.

We additionally setup image operations like random rotation (−45° to 45°), random flip (horizontal flip or vertical flip), stretching, and cropping when loading the dataset. The benefit of this is that the neural network's input training data are not exactly same in different epoch, which considerably reduces overfitting, as evidenced by later results. Finally, the preprocessed MRI and CT files are converted into *Numpy* format and fed into Pytorch.[26] To facilitate network training and save GPU memory, 512×512 CT images are resampled to 256×256.

2.2 | Network architectures

To better realize the synthesis task from MRI to sCT, we introduce channel attention mechanism and dimensional attention mechanism module to U-shaped network architecture. A neural network module called "Attentional ResBlock" is proposed to optimize edge prediction while reducing the network parameters. To validate the superiority of the proposed network, we compare it with several popular frameworks (UNet, Pix2Pix,[27] ResNet[28]). All training and testing are implemented on two NVIDIA TITAN V 12GB graphics cards using the deep learning framework Pytorch.

2.2.1 | Attentional mechanism

Attention mechanism is a special structure embedded in the ML model, which is used to automatically learn and calculate the contribution of input data to output data. The weight of the attention mechanism is determined by the weighted average of the concealed states of all the encoder's time steps. The decoder adjusts these weights, known as Attention Weights, at each time step, allowing it to focus on different parts of the input sequence at different time steps. Inspired by CBAM,[29] we introduced the attention mechanism into U-shaped neural network for MRI/CT translation. As shown in Figure 1, the attention mechanism module designed for this work is mainly divided into two parts: channel attention module (CAM) and spatial attention module (SAM).

2.3 | Channel attention module

Global average pooling and maximum pooling are used to utilize the high-level and low-level information, respectively. Figure 1 contains the schematic diagram of the channel attention module. When the input is a feature map of $H \times W \times C$, two-channel images of $1 \times 1 \times C$ are first obtained through global average pooling and maximum pooling of space, respectively. They are then fed into a multilayer perceptron (MLP). Then, the weight coefficient is obtained by adding the

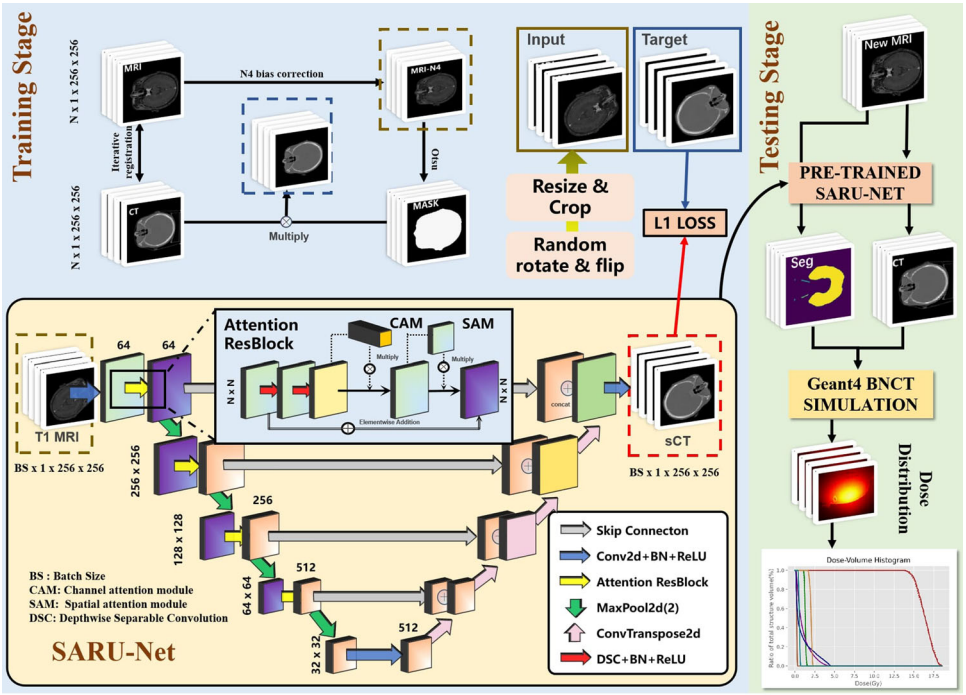

FIGURE 2 Schematic diagram of the training and testing of the model: left training stage, right testing stage. (the yellow part is the SARU network structure diagram, grey arrow: jump connection, blue arrow: basic convolution layer and batch normalization and LeakyReLU, yellow arrow: attention ResBlock, green arrow: Max pooling, pink arrow: ConvTranspose3d and batch normalization and LeakyReLU, red arrow: 1 × 1 convolution, CAM: channel attention module, SAM: spatial attention module, DSC: depthwise separable convolution)

two features through a Sigmoid activation function. Finally, we multiply the weight coefficient V_c and the original feature F to get the new feature map after scaling. Equation (1) shows the full stages of CAM in detail.

$$V_c(F) = \sigma(\mathbf{MLP}(\text{MeanPool}(F)) + \mathbf{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F^C_{\text{mean}})) + (W_1(W_0(F^C_{\text{max}}))) \quad (1)$$

2.4 | Spatial attention module

After CAM, we introduce SAM to let the network learn which features are meaningful. Figure 1 also contains the schematic diagram of SAM. Similar to channel attention, given a feature map of $H \times W \times C$, two $H \times W \times 1$ images are obtained by average pooling and maximum pooling of one channel dimension, respectively, and the two images are spliced together according to the channel. Then, the weight coefficients are obtained through a tiny convolutional neural network (CNN). Finally, the new feature can be obtained by multiplying the weight coefficient versus the original feature graph F, as given by Equation (2):

$$V_s(F) = \sigma(f([\text{MeanPool}(F), \text{MaxPool}(F)])) = \sigma(f([F^S_{\text{mean}}; F^S_{\text{max}}])) \quad (2)$$

2.4.1 | Attentional ResBlock

We introduced the CAM and SAM into U-shaped neural network. We also added additional residual links, which will make the attention block learn the residuals between distinct feature maps instead of the entire feature map, to avoid overfitting as much as possible while speeding up the network's convergence. Finally, we employ depthwise separable convolution (DSC) to reduce the number of parameters in the network, as illustrated by the red arrow in Figure 2.

The DSC consists of a combination of depthwise (DW) and pointwise (PW) components to extract the feature map, which greatly reduces the number of parameters compared to the conventional convolution operation. The principle of deep separable convolution is shown in Figure 1. This results in fewer mathematical operations and fewer parameters than nonseparable convolution.[30] Trebing et al. created a UNet using DSC instead of regular convolution and their model had eight times fewer parameters than the original UNet implementation, and they showed that their model was able to have similar performance to the UNet on segmentation tasks.[31]

2.4.2 | Self-attention ResUNet

Self-attention ResUNet (SARU) is formed by incorporating the Attentional ResBlock into a U-shaped

TABLE 1 Hyperparameters of the networks

Hyperparameter	Value
Batch size	8
Loss	torch.nn.L1Loss
Optimizer	torch.optim.Adam
scheduler	ReduceLROnPlateau
Learning rate	0.0002
Momentum term of adam	0.5
num_threads	8
Image preprocess	rotate and resize and crop
init type	kaiming

neural network; Figure 2 depicts the network structure. The whole generator network contains a total of 16.212 M trained parameters to realize the complex end-to-end mapping function and converts the input MRI into the corresponding sCT image. Learning the end-to-end mapping function requires optimization of network parameters by minimizing losses or prediction errors between the predicted image and the corresponding real CT image. In order to quantitatively reflect the advantages of SARU, we simultaneously train three other network models, which are UNet, ResNet, and Pix2Pix. Pix2Pix uses UNet as the backbone and PatchGAN as discriminator.

We employ the mean absolute error (MAE) as the loss function for generator networks.[32] For Pix2Pix, the loss function consists of generator loss and discriminator loss.[27] Due to GPU memory limitations and limited training data in this study, 3D network training is not extended considering that 2D slices already contain rich context information. Except for the network framework, all hyperparameters are the same for all networks in Table 1. The batch size is set to eight, Adam is the optimizer, and the momentum term is set to 0.5. And the initial learning rate is set to 0.0002, and is reduced by a factor of 2 when the error on the validation set stops falling.

2.5 | Monte Carlo simulation and dose calculation for BNCT

Monte Carlo toolkit Geant4[33] is used to carry out simulation and quantitatively evaluate the bias of sCT generated by the neural network in dose calculation. Based on the CT data of selected patients, 24 different tissue materials are divided according to CT values based on the Schneider method.[34] The voxel size is set as $1 \times 1 \times 1\text{mm}^3$, and the boron concentration in the skin of the body model is set as 25 ppm. The default neutron irradiation method used in the simulation is top–down (or left–right), the distance between source and skin is from 5 to 10 cm, and the neutron source has a radius of 6 cm. The ratio between the total neutron current and the total neutron flux is 0.7,[35] and the energy spectrum of the neutron source is set to be the same as the neutron source of the Harvard-MIT reactor.[36] The neutron source spectrum and the schematic diagram of neutron irradiation mode are shown in Figure 3. All parameter settings are related to the location and size of tumors in different cases.

Considering the influence of calculation time, the number of neutrons simulated of each patient is 2×10^8 to ensure enough uncertainties as in previous studies.[37] We simulated the treatment process of one of the patient 10 times to calculate the standard deviation of the 10 dose results, and the uncertainty was 1.72% (head area only) and 0.74% (>10% maximum dose). The doses analyzed here are all 30-minute RBE doses.[38] The average MAE results of various organs in different networks are shown in Figure 3. Compared to other networks, the MAE of SARU dropped significantly.

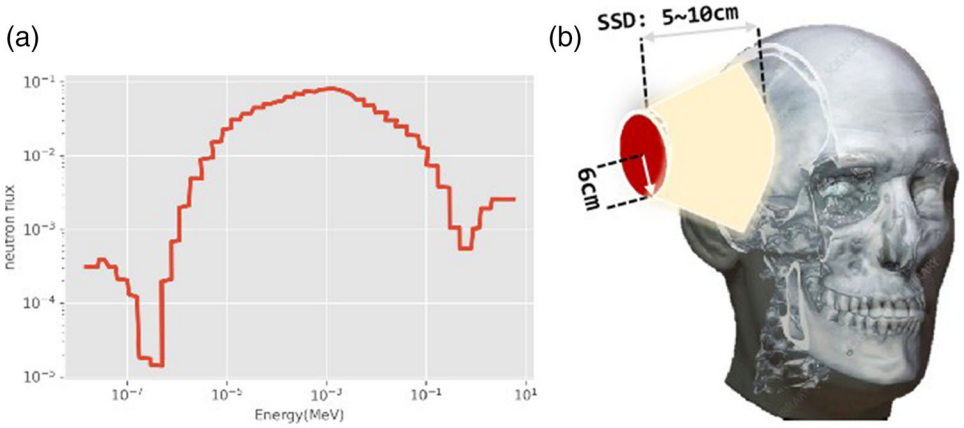

FIGURE 3 Neutron source spectrum (a) and the schematic diagram of neutron irradiation mode (b)

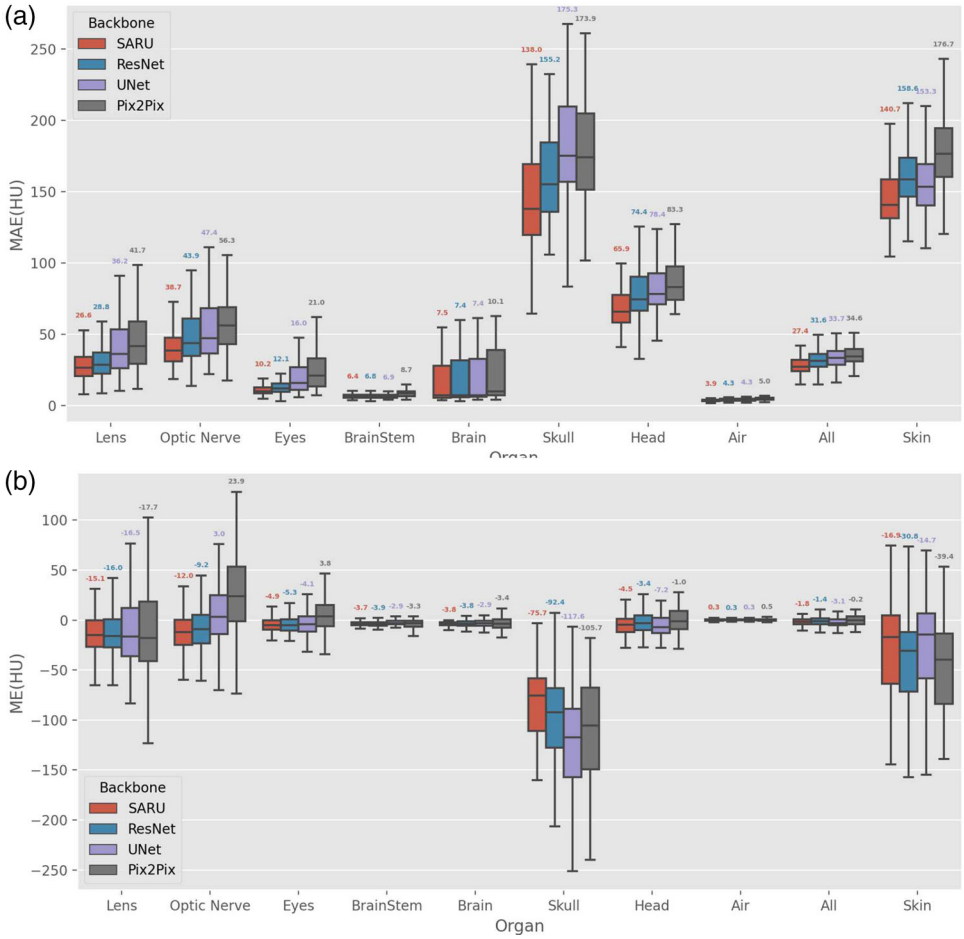

FIGURE 4 Average mean absolute error (MAE) (a) and ME (b) results of various organs in different networks

3 | RESULT

3.1 | Image quality assessment of sCT images

The MAE and mean error (ME) distributions of each organ are given in Figure 4. As shown in Table 2, the MAE of SARU results in the head region is 67.81 ± 24.31 for all 104 cases (five-fold cross-validation), while UNet, ResNet, and Pix2Pix results are 80.73 ± 24.68, 77.31 ± 26.73, and 85.29 ± 23.23, respectively. With the datasets and model parameters in this study, UNet and Pix2Pix had slightly lower or similar MAE than previous studies.[15,18,39–44] At the same time, Table 2 shows that SARU generally had a better prediction effect for the seven organs at risk (i.e., eye, lens, optic nerve, brain, brain stem, and skull) that were studied than other models with the same datasets. The "previous research" in Table 2 represents the results in previous studies. Due to the differences in the original data and preprocessing methods, the results are not exactly the same. The more intuitive results are shown in Figure 5.

3.2 | Dose assessment

The neutron source is set up and BNCT simulation is conducted on the test cases with different irradiation fields to obtain the three-dimensional distribution of boron dose. Detailed dose results are given in Table 3. For GTV and skin, the dose errors calculated by sCT produced by SARU network are smaller than UNet, ResNet, and Pix2Pix. The dose errors of BNCT from different networks are given in Table 4. The TOST-P (two one-sided test p-value)[45,46] shows no significant difference in DVH results. And for all tested cases, the average 2%/2 mm gamma index of UNet, ResNet, Pix2Pix, and SARU were 0.96 ± 0.03, 0.96 ± 0.03, 0.95 ± 0.03, and 0.98 ± 0.01, respectively. DVH images of test cases are presented in Figure 6. And as shown in Figure 7, SARU's DVH curve is closer to the ground truth than other networks.

TABLE 2 Mean absolute error (MAE) results of head, brain, and skull for different backbones in 104 patients

Backbone	Head		Skull		Brain	
	MAE (mean ± std/HU)	ME (mean ± std/HU)	MAE (mean ± std/HU)	ME (mean ± std/HU)	MAE (mean ± std/HU)	ME (mean ± std/HU)
UNet	80.73 ± 24.68	−2.84 ± 21.20	179.96 ± 47.29	−123.02 ± 54.62	17.62 ± 25.51	−3.79 ± 6.99
ResNet	77.31 ± 26.73	0.34 ± 21.73	158.93 ± 48.08	−95.86 ± 49.86	17.06 ± 25.52	−3.87 ± 6.02
Pix2Pix	85.29 ± 23.23	1.48 ± 21.43	174.24 ± 41.61	−110.41 ± 51.32	20.30 ± 26.11	−3.97 ± 8.49
SARU (ours)	**67.81 ± 24.31**	**−1.45 ± 21.72**	**143.97 ± 45.83**	**−84.63 ± 48.57**	**14.90 ± 21.21**	**−4.29 ± 5.48**
Previous research						
UNet Alvarez Andres et al.[15]	90 ± 21	–	199 ± 54	–	–	–
Pix2Pix Hemsley et al.[18]	94.60 ± 17.20	–	237 ± 31	–	–	–
CNN Alvarez Andres et al.[15]	81 ± 22	–	228 ± 63	–	–	–
CNN Han et al.[14]	84.8 ± 17.3	–	–	–	–	–
cGAN Kazemifar et al.[20]	73 ± 17	–	–	–	–	–

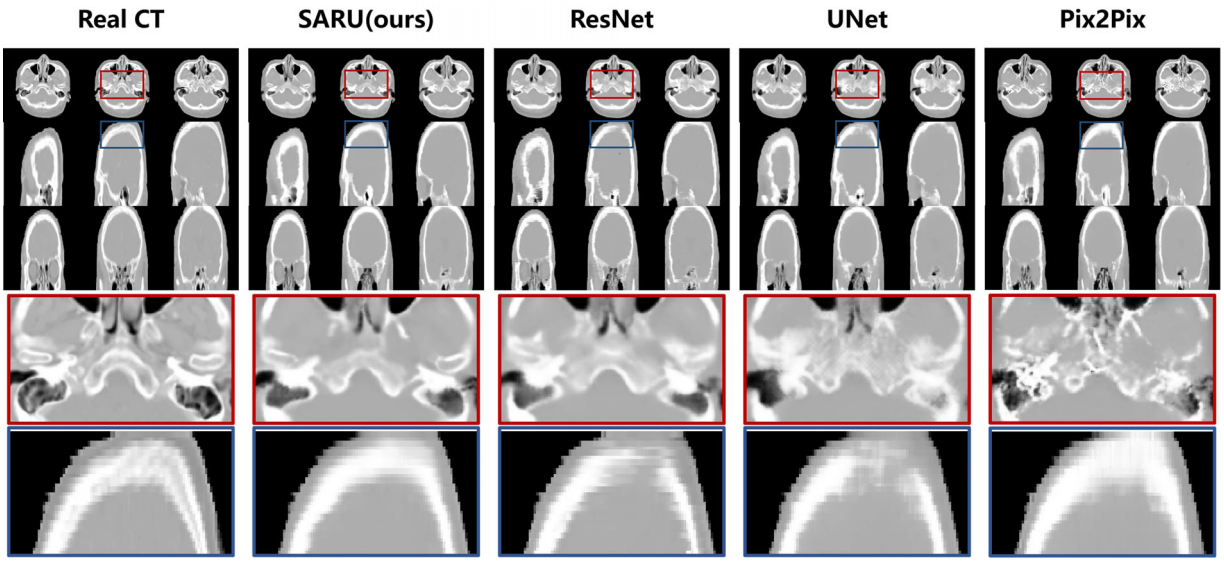

FIGURE 5 The prediction results of the transverse, sagittal, and coronal planes. Blue and red frame lines are the proportionally enlarged images of some areas.

4 | DISCUSSION

MRI has the advantages of no ionizing radiation, clear imaging of soft tissue, and real-time imaging. It can be used for boron concentration estimation and pharmacokinetic evaluation before the treatment of BNCT. MRI is increasingly being incorporated into the planning and implementation of radiotherapy. In this work, we provide a CNN called "SARU" with an attention mechanism that can be trained end-to-end for MRI/CT translation in MR-only BNCT treatment planning. To prevent the network from overfitting, deep separable convolutions and specialized residual connections are used. By introducing the attentional mechanism module, SARU demonstrates a good prediction effect at the edges of images and the junction of distinct organs. The quality of sCT prediction is considerably improved when compared to other networks such as ResNet, UNet, and Pix2Pix, and MAE is lowered.

Network parameters should not be very large for medical image translation tasks, as bloated networks are prone to major overfitting. But an overly simple network is difficult to anticipate accurately. Compared with other methods to generate sCT, the model based on the deep attention mechanism proposed in this paper has superior performance. The viability of T1w-weighted MRI

TABLE 3 Comparison of dose-volume histogram (DVH) parameters of plans on synthetic and planning computed tomography (CT) for 10 testing patients

DVH parameter		Real CT Mean ± SD (95% confidence interval)	sCT (SARU) Mean ± SD (95% confidence interval)	TOST_P(−0.5%, 0.5% equival interval)
GTV	D95 (Gy)	27.466 ± 1.432	27.576 ± 1.424	5.120E-12
		(24.659–30.272)	(24.785–30.367)	
	Dmean (Gy)	29.906 ± 1.149	29.987 ± 1.144	5.630E-13
		(27.654–32.158)	(27.744–32.230)	
	D2 (Gy)	32.282 ± 2.716	32.351 ± 2.663	4.341E-11
		(26.959–37.605)	(27.132–37.570)	
Len L	Dmean (Gy)	1.603 ± 0.403	1.608 ± 0.405	8.140E-14
		(0.812–2.393)	(0.814–2.402)	
	D2 (Gy)	1.992 ± 0.428	1.993 ± 0.437	2.653E-10
		(1.153–2.830)	(1.137–2.849)	
Len R	Dmean (Gy)	1.740 ± 0.355	1.737 ± 0.355	6.930E-13
		(1.043–2.436)	(1.042–2.432)	
	D2 (Gy)	2.080 ± 0.616	2.077 ± 0.602	8.846E-11
		(0.872–3.287)	(0.897–3.257)	
Optic nerve L	Dmean (Gy)	2.492 ± 0.456	2.495 ± 0.456	2.950E-15
		(1.599–3.386)	(1.600–3.389)	
	D2 (Gy)	3.067 ± 1.500	3.059 ± 1.500	4.417E-12
		(0.127–6.007)	(0.119–5.999)	
Optic nerve R	Dmean (Gy)	2.681 ± 0.369	2.674 ± 0.365	1.230E-11
		(1.957–3.405)	(1.958–3.390)	
	D2 (Gy)	3.293 ± 1.267	3.279 ± 1.259	4.238E-07
		(0.810–5.776)	(0.811–5.747)	
Skull	Dmean (Gy)	2.557 ± 0.186	2.561 ± 0.186	2.490E-16
		(2.191–2.922)	(2.196–2.927)	
	D2 (Gy)	6.493 ± 0.497	6.504 ± 0.491	5.283E-11
		(5.519–7.467)	(5.541–7.467)	
Skin	Dmean (Gy)	2.387 ± 0.171	2.389 ± 0.171	3.550E-16
		(2.052–2.723)	(2.053–2.724)	
	D2 (Gy)	7.400 ± 0.536	7.416 ± 0.522	6.134E-10
		(6.349–8.451)	(6.393–8.839)	
Brain	Dmean (Gy)	3.336 ± 0.347	3.336 ± 0.346	4.680E-16
		(2.655–4.016)	(2.658–4.014)	
	D2 (Gy)	6.220 ± 0.875	6.218 ± 0.885	1.431E-11
		(4.505–7.935)	(4.483–7.952)	

imaging for BNCT dose calculation has been confirmed. However, secondary fine-tuning may still be necessary for T1w with drugs. Since MRI has no dose risk, two MRIs are also an option for patients with BNCT. The registration between two MRI images of the same machine can also be much simpler and more precise than the registration between MRI and CT.

For 10 patients with brain tumors evaluated using SARU, there is no significant dosimetric difference between the doses obtained from sCT and those calculated based on real CT. D95 of GTV showed a high level of performance, and all sCT plans met clinical planned dose limits. The calculation deviation of the sCT dose is less than 1% compared with the real CT dose, and the gamma index of 2%/2 mm is greater than 98%. We believe this is related to the attentional mechanism module's automatic seeking capacity. It can learn the regions of interest of various feature maps automatically. At the same time, we apply random deformation and rotation to the training data to increase the number

TABLE 4 Dose error of boron neutron capture therapy (BNCT) from different networks

Dose error		SARU Mean ± SD (95% confidence interval)	UNet Mean ± SD (95% confidence interval)	Pix2Pix Mean ± SD (95% confidence interval)	ResNet Mean ± SD (95% confidence interval)
GTV	Dmean (%)	0.350 ± 0.098	1.018 ± 0.192	1.067 ± 0.255	0.683 ± 0.135
		(0.158–0.542)	(0.642–1.393)	(0.567–1.566)	(0.419–0.947)
	D2 (%)	0.321 ± 0.133	0.995 ± 0.294	0.970 ± 0.346	0.733 ± 0.219
		(0.067–0.671)	(0.430–1.520)	(0.305–2.116)	(0.313–1.076)
Skin	Dmean (%)	0.510 ± 0.157	0.883 ± 0.218	0.909 ± 0.275	0.713 ± 0.100
		(0.202–0.819)	(0.465–1.310)	(0.370–1.449)	(0.517–0.909)
	D2 (%)	0.416 ± 0.118	0.955 ± 0.438	1.451 ± 0.465	0.656 ± 0.327
		(0.190–1.643)	(0.114–1.796)	(0.559–2.343)	(0.028–1.285)

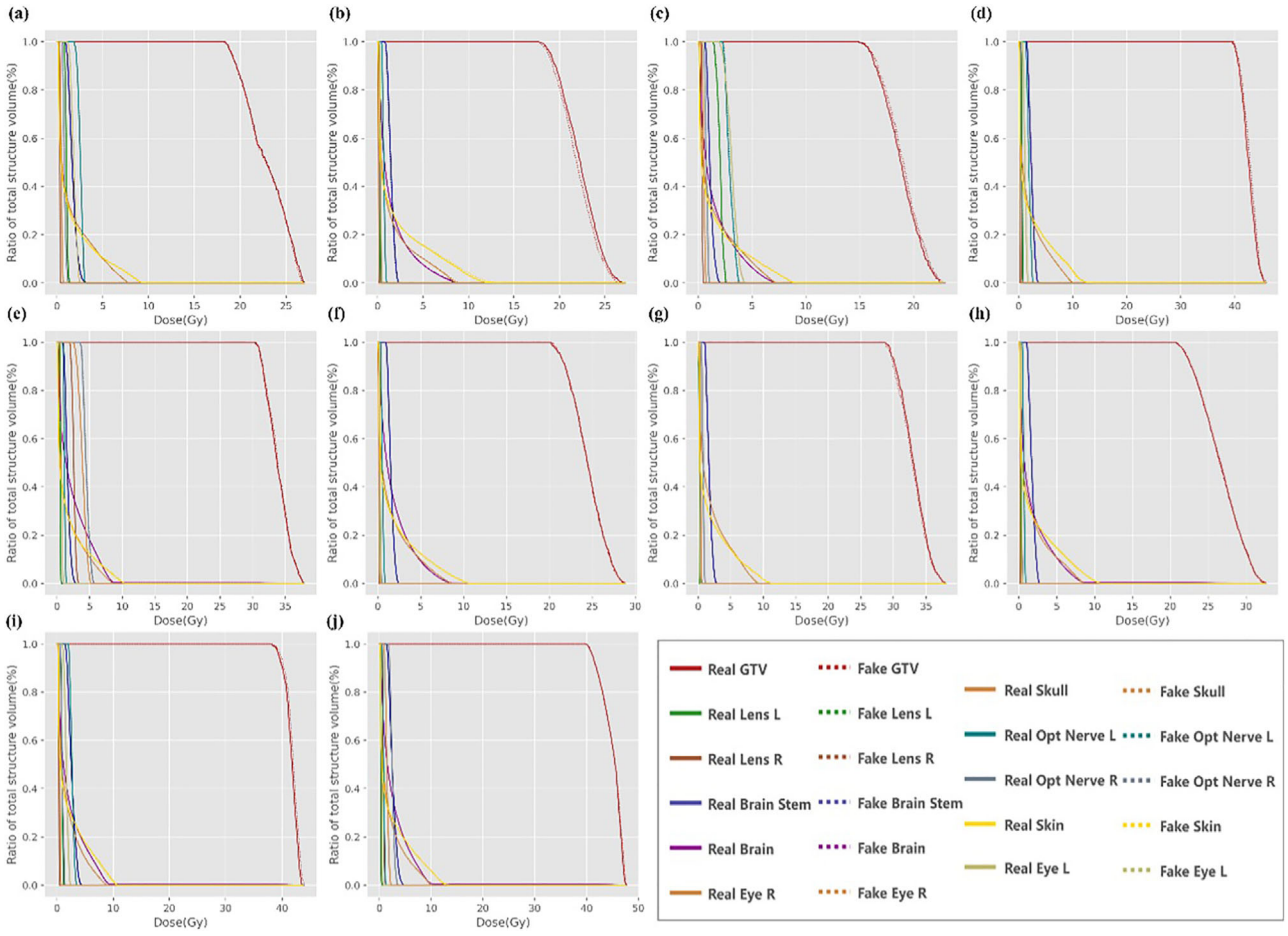

FIGURE 6 Dose-volume histogram (DVH) result of test patients 1–10 (a–j)

of training examples. This method improves the model's resistance to soft tissue and organ deformation while also reducing overfitting.

However, the output results will be worthless if the input data are too different from the data in the training set, hence the datasets should cover as many categories as feasible. Simultaneously, we must guarantee that the MRI and CT images in the training set are appropriately registered, as registration quality can have a major impact on the findings. Other limitations of this approach have not been identified in a large number of clinical cases. Furthermore, the T1-weighted image is taken as input and outputs the corresponding CT image, using a single-in-single-out mode. If images from other modalities are available, such as T2-weighted and ultra-short-term echo (UTE),

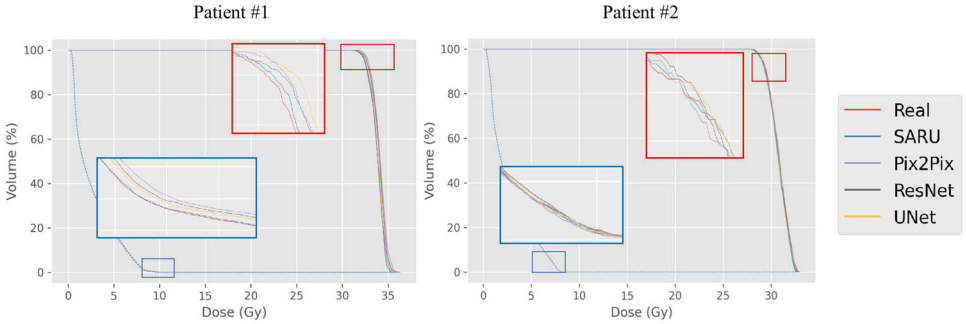

FIGURE 7 Dose-volume histogram (DVH) result of two patients from different networks

we can employ a multiple-input-single-output mode to improve the results. Additionally, 3D networks will be an interesting topic if more datasets are available, as the introduction of interlayer information can also improve edge artifacts in images to some extent.

5 | CONCLUSIONS

Dose calculation is a key step in BNCT treatment planning. In radiotherapy, CT images are critical for dose computation. However, cross-modality image registration is often not good enough, and additional CT scans complicate treatment procedures and increase dose risk. MRI has the advantages of no ionizing radiation, clear imaging of soft tissue, and real-time imaging, and has been proposed to be potentially used for boron concentration estimation before BNCT treatment. In this work, a new neural network structure named SARU with an attentional mechanism is proposed to directly map MRI to CT images, which can greatly simplify the pretreatment workflow and reduce the dose received by patients. The head MAE of SARU is 67.81 ± 24.31 HU, which is better than UNet (80.73 ± 24.68 HU), ResNet (77.31 ± 26.73 HU), and Pix2Pix (85.29 ± 23.23 HU) for the same time, and the sCT image has been registered with the MRI, so the problem of image feature mismatch has also been eliminated. Meanwhile, Monte Carlo dose calculation results show that the gamma index of 2%/2 mm is greater than 98%, which meets the dose calculation requirements. This method can greatly simplify the treatment planning process of BNCT and reduce the dose risk of BNCT therapy.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program (2022YFE0107800), the Natural Science Foundation of Jiangsu Province (BK20220132) and the Nanjing University of Aeronautics and Astronautics (Grant No. xcxjh20210615).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Moss RL. Critical review, with an optimistic outlook, on boron neutron capture therapy (BNCT). *Appl Radiat Isot*. 2014;88:2-11.
2. Hirose K, Konno A, Yoshimoto S, et al. Safety and antitumor activity of accelerator-based boron neutron capture therapy in patients with inoperable recurrent and locally advanced head and neck cancer: a phase II study. *J Clin Oncol*. 2019;37:6028-6028.
3. Kabalka GW, Davis M, Bendel P. Boron-11 MRI and MRS of intact animals infused with a boron neutron capture agent. *Magn Reson Med*. 1988;8:231-237.
4. Aime S, Barge A, Crivello A, et al. Synthesis of Gd(III)-C-palmitamidomethyl-C′-DOTAMA-C 6-o-carborane: a new dual agent for innovative MRI/BNCT applications. *Org Biomol Chem*. 2008;6:4460-4466.
5. Alberti D, Deagostino A, Toppino A, et al. An innovative therapeutic approach for malignant mesothelioma treatment based on the use of Gd/boron multimodal probes for MRI guided BNCT. *J Control Release*. 2018;280:31-38.
6. Alberti D, Toppino A, Geninatti Crich S, et al. Synthesis of a carborane-containing cholesterol derivative and evaluation as a potential dual agent for MRI/BNCT applications. *Org Biomol Chem*. 2014;12:2457-2467.
7. Wu CY, Hsieh HH, Chang TY, et al. Development of mri-detectable boron-containing gold nanoparticle-encapsulated biodegradable polymeric matrix for boron neutron capture therapy (BNCT). *Int J Mol Sci*. 2021;22:8050.
8. Skwierawska D, López-Valverde JA, Balcerzyk M, Leal A. Clinical viability of boron neutron capture therapy for personalized radiation treatment. *Cancers (Basel)*. 2022;14:2865.
9. Cusumano D, Lenkowicz J, Votta C, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol*. 2020;153:205-212.
10. Dinkla AM, Florkow MC, Maspero M, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. *Med Phys*. 2019;46:4095-4104.
11. Largent A, Barateau A, Nunes JC, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int J Radiat Oncol Biol Phys*. 2019;105:1137-1150.
12. Liu F, Yadav P, Baschnagel AM, Mcmillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. *J Appl Clin Med Phys*. 2019;20:105-114.

13. Qi M, Li Y, Wu A, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med Phys*. 2020;47:1880-1894.
14. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408-1419.
15. Alvarez Andres E, Fidon L, Vakalopoulou M, et al. Dosimetry-driven quality measure of brain pseudo computed tomography generated from deep learning for MRI-only radiation therapy treatment planning. *Int J Radiat Oncol Biol Phys*. 2020;108:813-823.
16. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: *Lecture Notes in Computer Science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Vol. 9901. LNCS; 2016:424-432.
17. Bourbonne V, Jaouen V, Hognon C, et al. Dosimetric validation of a gan-based pseudo-ct generation for mri-only stereotactic brain radiotherapy. *Cancers (Basel)*. 2021;13(5):1082.
18. Hemsley M, Chugh B, Ruschin M, et al. Deep generative model for synthetic-CT generation with uncertainty predictions. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. In: *Lecture Notes in Computer Science*. Vol. 12261. Springer; 2016:834-844.
19. Maspero M, Bentvelzen LG, Savenije MHF, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. *Radiother Oncol*. 2020;153:197-204.
20. Kazemifar S, Mcguire S, Timmerman R, et al. MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother Oncol*. 2019;136:56-63.
21. Fukuda H, Hiratsuka J, Kobayashi T, et al. Boron neutron capture therapy (BNCT) for malignant melanoma with special reference to absorbed doses to the normal skin and tumor. *Australas Phys Eng Sci Med*. 2003;26:97-103.
22. West J, Fitzpatrick JM, Wang MY, et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr*. 1997;21:554-568.
23. Stokking R, Vincken KL, Viergever MA. Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data. *Neuroimage*. 2000;12:726-738.
24. Larobina M, Murino L. Medical image file formats. *J Digit Imaging*. 2014;27:200-206. https://doi.org/10.1007/s10278-013-9657-9
25. Avants B, Epstein C, Grossman M, Gee J. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12:26-41.
26. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32:8026-8037.
27. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 21–26 July 2017. IEEE; 2017:5967-5976.
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, December 2016. IEEE; 2016:770-778.
29. Woo S, Park J, Lee JY, Kweon ISC. Convolutional block attention module. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11211 LNCS*. Springer; 2018:3-19.
30. Chollet F. Xception: deep learning with depthwise separable convolutions: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. IEEE; 2017:1800-1807.
31. Trebing K, Stanczyk T, Mehrkanoon S. Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognit Lett*. 2021;145:178-186.
32. Wang C, Xu C, Wang C, Tao D. Perceptual adversarial networks for image-to-image transformation. *IEEE Trans Image Process*. 2018;27:4066-4079.
33. Agostinelli S, Allison J, Amako K, et al. Geant4—a simulation toolkit. *Nucl Instrum Methods Phys Res A*. 2003;506:250-303.
34. Schneider W, Bortfeld T, Schlegel W. Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. *Phys Med Biol*. 2000;45:459-478.
35. Yanch JC. Current status of neutron capture therapy in the United States. *Radiother Oncol*. 1995;37:S35.
36. Busse PM, Zamenhof RG, Harling OK, et al. The Harvard-MIT BNCT Program. In: *Frontiers in Neutron Capture Therapy*. Springer; 2001:37-60. https://doi.org/10.1007/978-1-4615-1285-1_4
37. Tian F, Geng CR, Tang XB, et al. Analysis of influencing factors on the method for determining boron concentration and dose through dual prompt gamma detection. *Nucl Sci Tech*. 2021;32:35.
38. Wu S, Geng C, Tang X, et al. Dosimetric impact of respiratory motion during boron neutron capture therapy for lung cancer. *Radiat Phys Chem*. 2020;168:108527.
39. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408-1419.
40. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI 2017. In: *Lecture Notes in Computer Science*. Vol. 10435. Springer; 2017:417-425.
41. Wolterink JM, Dinkla AM, Savenije MHF, et al. Deep MR to CT synthesis using unpaired data. In: Tsaftaris S, Gooya A, Frangi A, Prince J, eds. *Simulation and Synthesis in Medical Imaging. SASHIMI 2017. Lecture Notes in Computer Science*. Vol. 10557. Springer; 2017:1423.
42. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. 2018;45:3627-3636.
43. Xiang L, Wang Q, Nie D, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med Image Anal*. 2018;47:31-44.
44. Stoyanov D, v D, v D, et al. Deep learning in medical image analysis and multimodal learning for clinical decision support: Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018. Vol. 11045. Springer; 2018.
45. Mara CA, Cribbie RA. Paired-samples tests of equivalence. *Commun Stat Simul Comput*. 2012;41:1928-1943.
46. Cusumano D, Lenkowicz J, Votta C, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol*. 2020;153:205-212.

How to cite this article: Zhao S, Geng C, Guo C, Tian F, Tang X. SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning. *Med Phys*. 2022;1-11. https://doi.org/10.1002/mp.15986