IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2019, accepted November 5, 2019, date of publication November 18, 2019, date of current version December 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953923

Stepwise Local Synthetic Pseudo-CT Imaging Based on Anatomical Semantic Guidance

HONGFEI SUN[1], KUN ZHANG[2], RONGBO FAN[1], WENJUN XIONG[1], AND JIANHUA YANG[1]

[1]School of Automation, Northwestern Polytechnical University, Xi'an 710129, China
[2]Department of Radiotherapy, Shaanxi Provincial Tumor Hospital, Xi'an 710061, China

Corresponding author: Jianhua Yang (yangjianhua@nwpu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC211500.

ABSTRACT In this study, an anatomic semantic guided neural style transfer (ASGNST) algorithm was developed and pseudo-computed tomography (CT) images synthesized in steps. CT images and ultrasound (US) images of 20 cervical cancer patients to be treated were selected. The foreground (FG) and background (BG) regions of the US images were segmented by the region growth method, and three objective functions for content, style, and contour loss were defined. Based on the two types of regions, a local pseudo-CT image synthesis model based on a convolution neural network was established. Then, global 2D pseudo-CT images were obtained using the weighted average fusing algorithm, and the final pseudo-CT images were obtained through 3D reconstruction. US phantom and data of five additional cervical cancer patients were used for prediction. Furthermore, three image synthesis algorithms—global deformation field (GDF), stepwise local deformation field (SLDF), and neural style transfer (NST)—were selected for comparative verification. The pseudo-CT images synthesized by the four algorithms were compared with the ground-truth CT images obtained during treatment. The structural similarity index between the ground-truth CT and pseudo-CT synthesized by the improved algorithm significantly differed from those synthesized by the other three algorithms ($t_{GDF\_bg} = 7.175$, $t_{SLDF\_bg} = 4.513$, $t_{NST\_bg} = 3.228$, $t_{GDF\_fg} = 10.518$, $t_{SLDF\_fg} = 5.522$, $t_{NST\_fg} = 2.869$, $p < 0.05$). Further, the mean absolute error and peak signal-to-noise ratio values prove that the pseudo-CT synthesized by the ASGNST algorithm is similar to the ground-truth CT. The improved algorithm can obtain pseudo-CT images with high precision and provides a novel direction for image guidance in cervical cancer brachytherapy.

INDEX TERMS Neural style transfer, pseudo-CT, radiotherapy, ultrasound.

I. INTRODUCTION

Brachytherapy based on the afterloader system is an important part of cervical cancer radiotherapy. Compared with external radiotherapy, the dose distribution of brachytherapy is uneven, the dose is multiplied from the position of the applicator, and the dose-drop gradient is large. Therefore, dose control of organs at risk (OARs) such as the bladder, cervix, and rectum is particularly important during brachytherapy [1]. Image-guided technology can adjust the relative position between the applicator and the OARs to ensure accurate treatment. Computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) are regularly used in brachytherapy [2], [3]. Radiotherapy planning and dose assessment in patients are performed using CT images, thus CT images are the necessary image modality for brachytherapy. However, artifacts exist in the metal material applicator after CT scanning, which affect the imaging quality at planning target volume (PTV) and the accuracy of dose implementation [4]. Repeated CT scans also increase the patient's additional radiation dose during the calibration of the application position for fractional therapy, resulting in toxic side effects [5]. Compared with CT images, MRI provides a high resolution of soft tissue and does not involve radiation. However, the scanning time of MRI is longer, and the position of the patients can change during scanning, resulting in errors [6]. Compared with the first two guiding imaging devices, US has the advantages of portability, non-radiation, and real-time performance while determining the position of

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

the applicator [7]. Tagger *et al.* found in an image-guided device for cervical cancer brachytherapy that about 70% of patients require US image guidance when implanting the applicator [8]. At present, US can only assist CT images to perform brachytherapy and cannot be directly used to design and modify a radiotherapy plan. Moreover, the ability of radiotherapists to read US images is weaker than their ability to read CT images. Based on the anatomical semantic guided neural style transfer (ASGNST) algorithm, a matching relationship between ultrasound and CT images can be established in the foreground (FG) and background (BG) regions, and partial pseudo-CT images can be synthesized stepwise and finally merged into global pseudo images. This is a novel idea to solve the above problems.

The acquisition of pseudo-CT images based on US imaging is mainly divided into two categories: the deformation field registration method and the image synthesis method. The former refers to virtual CT images formed by applying a deformation field obtained by registration between the US image acquired at the simulation stage (US_{sim}) and the US image acquired at the treatment stage (US_{tx}) to the CT image acquired at the simulation stage (CT_{sim}) [9]. Van der Meer et al. first obtained pseudo-CT images based on the US image deformation field and applied it to clinical radiotherapy. They proved that pseudo-CT images can represent the anatomical structure information of patients better than CT_{sim} images [10]. Based on the research of Van der Meer et al., we defined different regions of interest (ROIs) for the US_{sim} and US_{tx} and then performed local registration to obtain several deformation fields. These fields were then applied to the CT_{sim} to obtain pseudo-CT. We verified that the pseudo-CT images obtained by US deformation based on stepwise local registration were similar to cone beam computed tomography (CBCT) images [11]. However, the deformation fields based on multiple ROIs are discontinuous, resulting in inaccurate connection of tissues and organs at the critical positions of different ROIs, which can be solved by smoothing, which reduces image clarity. The image synthesis method can overcome the limitation of acquiring pseudo-CT images based on deformation fields. Image synthesis involves extracting the features of CT images and other modal images using methods such as machine learning or deep learning, matching the features of the same region, and using matching results as the training set. By adding images of other corresponding modalities in the test stage, the corresponding pseudo-CT images can be obtained. At present, most scholars have synthesized pseudo-CT images through MRI and PET, while in-depth research on the synthesis of pseudo-CT based on US images is lacking [12]–[14].

Neural style transfer (NST) algorithms have been applied to image synthesis and mainly include image iterative methods [15]–[17] and model iterative methods [18], [19]. However, NST has not been applied yet to study medical image synthesis. Traditional NST methods include semantic mismatching problems, such as migrating the BG texture of the style image to the FG of the target image. Semantic consistency is not maintained during the migration process, and there is no global linear matching relationship between US images and CT images [20]. Champanard et al. solved the above mismatching problem by manually preparing semantic graph information, but the high labor cost and low efficiency limited the promotion of the algorithm [21]. In this study, an ASGNST algorithm was proposed to synthesize pseudo-CT images stepwise. We defined three objective functions: content loss, style loss, and contour loss. The regional growth method was used to segment the bladder and cervix in the US image as the FG region, and the remaining region as the BG region. Convolutional neural networks (CNNs) were established based on the two types of regions. The algorithm can automatically extract the semantic information of the reference CT images and the contour information of the US images and can synthesize a pseudo-CT image with two types of information as constraints. Mismatch of content and style between different regions is avoided, while the accuracy of the FG region contour is ensured. We used the ultrasound phantom and real patient data to synthesize pseudo-CT images separately. The synthesized pseudo-CT images were compared with the ground-truth CT images to verify the feasibility of the method applied to brachytherapy. Ground-truth CT refers to the CT images acquired during the treatment phase. The process of pseudo-CT image synthesis is shown in Fig. 1. After the model was trained, only the US images were added to the network in the test phase to obtain the corresponding pseudo-CT images. The results of this study will provide a novel direction for image guidance in cervical cancer brachytherapy.

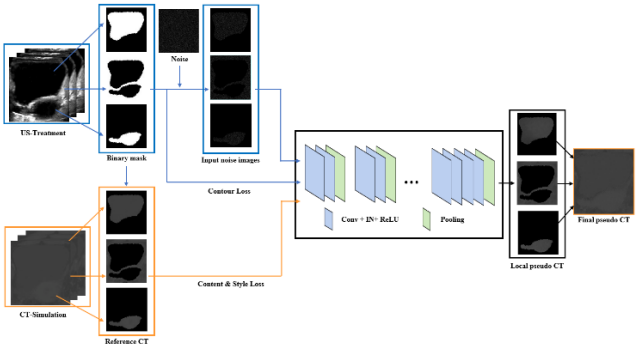


FIGURE 1. Flowchart of pseudo-CT images synthesis.

II. METHODS

A. DATA ACQUISITION

In this study, 20 patients to be treated for cervical cancer were randomly selected to obtain CT images at the simulation stage and US images at the treatment stage. The original CT images were obtained from the CT scanner (Philips Company, Netherlands) with a size of $512 \times 512 \times (76–101)$ and a resolution of $0.97 \times 0.97 \times 1$ mm^3. The original US images were obtained from the Clarity US scanning device

(Elekta Company, Sweden) with a size of 256 × 256 × (86–120) and a resolution of 1.17 × 1.17 × 1 mm^3. Before image synthesis, we preprocessed the experimental data with affine registration to correct the shift, rotation, and scale difference between the two modal images. The size of the image data of all cervical cancer patients was unified as 120 × 130 × 60, and the resolution was uniformly adjusted to 1 × 1 × 1 mm^3, which was used as the training data set of the CNN. In addition, image data of five cervical cancer patients and US phantoms were selected, and the model was tested after preprocessing. In the CNN, we removed the full connection layer in the classic VGG Net to obtain the image results. To ensure that the image size after the output network is the same as the image size before the input network, we set the padding in the convolution layer and the pooling layer to SAME. The CNN has 16 layers, including four convolution groups and four pooling layers. The first two convolution groups have two network layers. Each layer contains convolution, instance normalization (IN), and ReLU activation; the last two convolution groups have four network layers, which also contain convolution, IN, and ReLU. In the next layer of each convolution group, there is a pooling layer; the size of convolution kernel in the convolution layer is 3 × 3, and the size of the pooling kernel in the pooling layer is 2 × 2. The number of filters are 64, 64, 256, 256, 256, 256, 512, 512, 512, and 512, respectively, for the individual layers.

B. IMAGE PREPROCESSING OF US AND CT

Because of the different scanning conditions and imaging principles of ultrasonic and CT equipment, direct synthesis using the original images obtained by these machines would result in image distortion. Therefore, the original ultrasound and CT images had to be preprocessed before they could be used for pseudo-CT synthesis. The image preprocessing procedure is illustrated in Fig. 2.

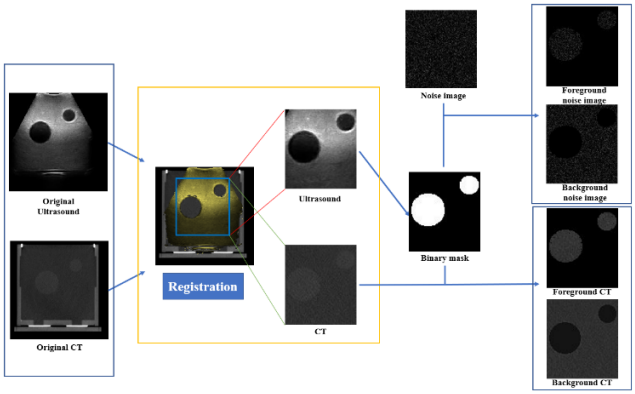


FIGURE 2. Flowchart of image preprocessing.

The process can be divided into the following four steps. I: Affine registration was performed on the CT images acquired during treatment and the simulation stage to ensure that the size and resolution of the two modal images were consistent. II: Compared with the CT images, US images have a small imaging range. Therefore, to ensure the accuracy of the image synthesis regions, a common region of the two modal images needs to be extracted. Based on the results of image registration and resampling, we extracted the local imaging region from each of the 2D slices in the global original US and CT images. III: A binary mask of US images is made by the region-growing image-segmentation method, and the FG and BG regions of the US image are separated. The binary image is also used as a reference image for contour loss in subsequent image synthesis operations. In this study, the Sobel edge detection method was used to obtain the contour of the binary image. IV: White noise is added to the binary mask image based on the FG and BG regions, and the FG and BG noise images are obtained after pixel operation. The two types of images were taken as the initial images for stepwise acquisition of the pseudo-CT images based on CNN. In addition, FG CT and BG CT images were obtained based on the FG and BG regions distinguished by the binary mask, which were used as the reference images of content loss and style loss in the stepwise image synthesis operation.

C. STEPWISE ACQUISITION OF PSEUDO-CT IMAGES BASED ON CNN

As a classic CNN structure, VGG Net, has been widely used in image segmentation and image recognition [22]–[24]. Among the many deep neural networks, VGG Net has a simple structure, and the feature extraction effect obtained using multiple filters of smaller size is better than that obtained using larger filters. Multiple nonlinear network layers increase the depth of the network, so the network can learn more complex features and the calculation cost can be reduced [25]. In this study, a part of the network layer in VGG Net19 was used to synthesize the pseudo-CT images, which were synthesized on phantom and real patients. The framework of the proposed algorithm is shown in Fig. 3. The algorithm model is composed of two parts: FG-based and BG-based image generation models. The local imaging regions were synthesized stepwise to avoid the problem of semantic mismatching style transfer, and the final global pseudo-CT images were obtained using the weighted average fusion algorithm. Because the resolution of the abdominal CT images is lower than that of MRI and US images, the texture features and grayscale differences between different tissues are not very obvious. Therefore, to approximate the similarity

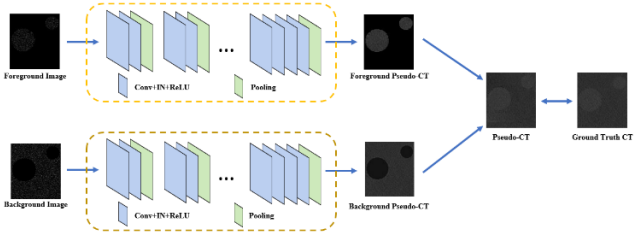


FIGURE 3. Stepwise acquisition of pseudo-CT images based on convolution neural network.

of style and content between pseudo-CT images and real CT images, we reserved the pooling layer in the network. Based on the convolution operation of the CNN, a network layer with N_h filters provides N_h feature maps. The size of the feature map is the product of its length H_h and width W_h, and the feature maps of the network layer h can be represented by the matrix $X^h \in R^{H_h * W_h}$. $\vec{x}$ and $\vec{y}$ are the images to be generated and CT images, respectively, and X and Y are the corresponding feature maps. Then, the function of content loss is defined using (1):

$$L_{content}(\vec{x}, \vec{y}, h) = \frac{1}{2H_h W_h} \sum_{i,j} (X_{ij}^h - Y_{ij}^h)^2 \quad (1)$$

X_{ij}^h is the ReLU activation value of the ith filter in the h network layer of the image to be generated at position j, and Y_{ij}^h is the ReLU activation value of the ith filter in the network layer h of the CT image at position j. H_h and W_h are the length and width of the h-layer feature maps, respectively. The Gram matrix method is used to establish the response relationship between different filters in the same h network layer [26], which is expressed by G_{ij}^h. The Gram matrix can be regarded as an eccentric covariance matrix between features (that is, a covariance matrix without subtracting the mean). In the feature maps, each value is obtained by convolution of a specific filter at a specific location; thus, they represent the intensity of the feature. Gram actually calculates the correlation between the two features. The diagonal elements of the Gram matrix also reflect the number of each type of feature in the image. Furthermore, the Gram matrix is obtained by calculating the inner product between the i-th and j-th feature maps of the h network layer, as shown in (2):

$$G_{ij}^h = \frac{1}{H_h W_h} \sum F_i^h F_j^h \quad (2)$$

By calculating the Euclidean distance of the statistical features of the Gram matrix between the CT image and the image to be generated, we can minimize the difference in the Gram matrix between the two images as the optimization objective, and the Gram style feature information of the image to be generated is constructed based on the feature maps of each network layer. $\vec{x}$ and $\vec{y}$ are the images to be generated and CT images, respectively, and Gx and Gy are the Gram style features corresponding to the images. Thus, the formula for calculating the function of style loss is shown in (3):

$$L_{style}(\vec{x}, \vec{y}, h) = \sum_{i,j} \sum_{h} \frac{\alpha_h}{2H_h W_h} (Gx_{ij}^h - Gy_{ij}^h)^2 \quad (3)$$

where α_h represents the weight of the image style loss of the h network layer, and the final loss value can be obtained by accumulating the style loss of each network layer. Gx_{ij}^h represents the Gram matrix obtained by the inner product of the i-th and the j-th feature maps in the h network layer of the images to be generated. Gy_{ij}^h is the Gram matrix obtained by the inner product of the i-th and the j-th feature maps in the h network layer of the CT images.H_h and W_h are the length and width of the h-layer feature maps, respectively.

Because the US and CT images used in image synthesis are acquired in different stages, the contours of the organs presented by them differ. In the process of image synthesis, the contour of the FG region should be considered to ensure that the contour of the bladder, cervix, and other OARs in the generated CT image is the same as that in the original US image. In this study, a Dice coefficient was introduced, and x and z were set as pseudo-CT images and ground-truth CT images, respectively. The binary pseudo-CT images and the binary original US images were compared, and the contour loss function was calculated using (4):

$$L_{contour}(\vec{x}, \vec{z}) = \frac{2|L_{\vec{x}} \cap L_{\vec{z}}|}{|L_{\vec{x}}| + |L_{\vec{z}}|} \quad (4)$$

where $L_{\vec{x}}$ and $L_{\vec{z}}$ are the binary segmentation results of the FG region in the pseudo-CT images and the original US images, respectively.

In summary, to introduce the content and style of the CT image into the pseudo-CT image and to obtain the contour of the FG region accurately, the content loss, style loss, and maximum contour loss functions need to be minimized. The total loss function is defined as

$$L_{total}(\vec{x}, \vec{y}, \vec{z}, h) = \omega_1 L_{content}(\vec{x}, \vec{y}, h) + \omega_2 L_{style}(\vec{x}, \vec{y}, h) + \omega_3 L_{contour}(\vec{x}, \vec{z}) \quad (5)$$

where ω_1, ω_2, and ω_3 are the weights of content loss, style loss, and contour loss, respectively. Considering the results obtained by Gatys *et al.* [30], we set $\omega_1/\omega_2 = 10^{-3}$ and $\omega_3/\omega_2 = 10^{-2}$.

III. EVALUATION

In this study, three quantitative measures, structural similarity index (SSIM) [27], mean absolute error (MAE), and peak signal-to-noise ratio (PSNR), were used to evaluate the accuracy of the image synthesis algorithm. The mathematical definition of SSIM is as follows:

$$SSIM = \frac{(2\mu_{I_g}\mu_{I_p} + c_1)(2\sigma_{I_g I_p} + c_2)}{(\mu_{I_g}^2 + \mu_{I_p}^2 + c_1)(\sigma_{I_g}^2 + \sigma_{I_p}^2)} \quad (6)$$

$$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2 \quad (7)$$

where μ_{Ig} and μ_{Ip} are the mean gray values of the ground-truth CT images and the pseudo-CT images, respectively.σ_{Ig} and σ_{Ip} are the variances of the gray values of the ground-truth CT images and the pseudo-CT images, respectively. σ_{IgIp} is the covariance. c_1 and c_2 are constants used to maintain stability. L refers to the range of gray values of the CT image. The values of k_1 and k_2 are 0.01 and 0.03, respectively. The value range of SSIM is [−1, 1]. The larger the value is, the greater is the similarity between the two images.

The mathematical definitions of the other two measurement methods are as follows:

$$MAE = \frac{1}{XYZ} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \sum_{z=0}^{Z-1} |I_g(x, y, z) - I_p(x, y, z)| \quad (8)$$

$$PSNR = 20\log_{10}\left(\frac{MAX_I}{\sqrt{\frac{1}{XYZ}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}\sum_{z=0}^{Z-1}|I_g(x,y,z)-I_p(x,y,z)|^2}}\right) \quad (9)$$

In (8) and (9), I_g and I_p represent the ground-truth CT images and the synthetic pseudo-CT images, respectively. X, Y, and Z represent the size of the image. MAX_I represents the maximum gray value of the CT images. The lower the MAE value is and the higher the PSNR value is, the closer the pseudo-CT image is to the ground-truth CT image. The calculation of the above three measurement methods is related to the gray values of the CT image, and the gray range of CT image is unified as [0,4000] [28].

IV. RESULTS

Figs. 4 and 5 show the error maps of ground-truth CT and pseudo-CT images obtained by four different synthetic methods based on the data of US phantom and cervical cancer patients, respectively. Because the US mode is a rigid structure, no deformation occurs. Therefore, to obtain the US deformation field and subsequently the pseudo-CT based on the deformation field method, this study artificially deformed the US and CT images. Based on the global US deformation field (GDF), the soft tissue area in the pseudo-CT images was deformed, which was not ideal. The pseudo-CT images obtained from the stepwise local US deformation field (SLDF) were better than the former, but the continuity of the images was poor. The pseudo-CT image could be roughly synthesized based on the NST algorithms. However, because

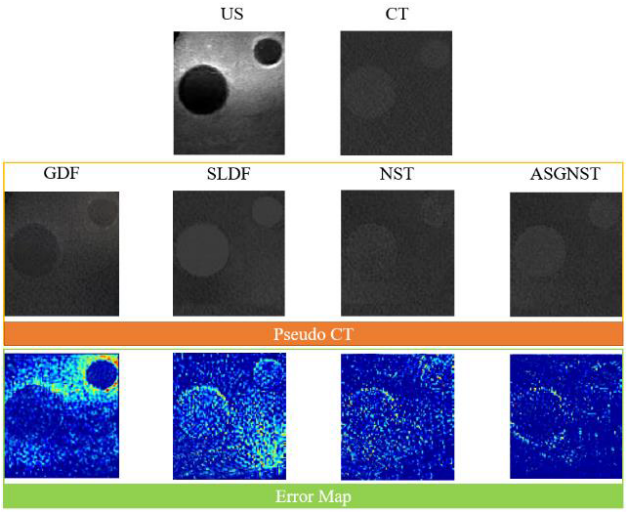


FIGURE 4. Error maps of ground-truth CT and pseudo-CT images obtained by four different synthetic methods based on US phantom data. GDF: global US deformation field. SLDF: stepwise local US deformation field. NST: neural style transfer. ASGNST: anatomic semantic guided neural style transfer.

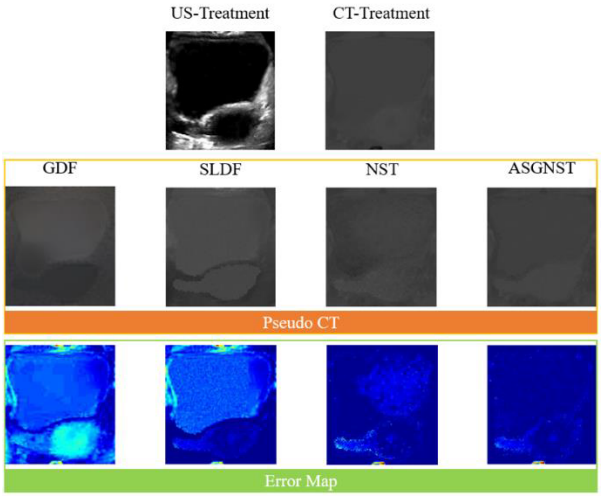


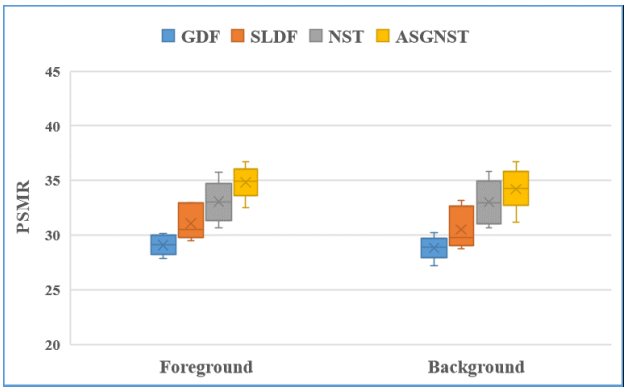
FIGURE 5. Error maps of ground-truth CT and pseudo-CT images obtained by four different synthetic methods based on cervical cancer patients data. GDF: global US deformation field. SLDF: stepwise local US deformation field. NST: neural style transfer. ASGNST: anatomic semantic guided neural style transfer.

of the mismatch between the style and the content in different FG and BG regions, the gray values of the pseudo-CT images were inaccurate and the noise was obvious. The pseudo-CT images synthesized by the ASGNST based on the FG and BG regions had better effect. The contour of each organ was similar to the ground-truth CT image, the grayscale of the image was relatively accurate, and the noise of the image was not obvious. Compared with the previous methods, the images obtained by the proposed algorithm were better.

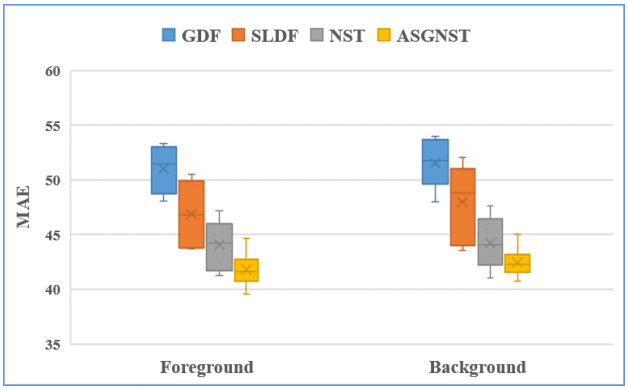
TABLE 1. SSIM measurement results of pseudo-CT images and ground truth CT images based on four different image synthesis algorithms.

		SSIM (%)			
		GDF	SLDF	CNN	S-CNN
Model	Foreground	88.1	94.3	97.1	**97.9**
	Background	93.7	96.9	97.7	**97.8**
Patient1	Foreground	87.6	90.7	92.9	**93.5**
	Background	85.2	85.7	90.5	**91.5**
Patient2	Foreground	82.4	87.4	89.3	**91.7**
	Background	80.2	86.3	89.2	**91.9**
Patient3	Foreground	83.0	87.9	91.5	**94.3**
	Background	82.9	85.7	90.3	**92.9**
Patient4	Foreground	86.5	89.5	93.6	**93.5**
	Background	82.8	88.1	92.1	**93.3**
Patient5	Foreground	85.1	91.5	91.6	**92.9**
	Background	84.6	90.3	91.7	**92.4**

Table 1 shows the SSIM measurement results between the ground-truth CT images and the pseudo-CT images synthesized by the four different image synthesis algorithms. Based on the comparison results of ground-truth CT and pseudo-CT images synthesized by the proposed algorithm, the paired



(a)



(b)

FIGURE 6. Results of similarity measure evaluation. (a) MAE measure evaluation results. (b) PSNR measure evaluation results.

T test was performed using the results of other three algorithms. The results showed that there were significant differences between the FG and BG regions ($t_{GDF\_bg} = 7.175$, $t_{SLDF\_bg} = 4.513$, $t_{NST\_bg} = 3.228$, $t_{GDF\_fg} = 10.518$, $t_{SLDF\_fg} = 5.522$, $t_{NST\_fg} = 2.869$, $p < 0.05$). Fig. 6a and b respectively show the results of two measures of MAE and PSNR on the FG and BG regions of ground-truth CT and pseudo-CT synthesized by the four methods. Compared with the other three algorithms, the pseudo-CT images obtained by the image synthesis algorithm proposed in this study were more similar to the ground-truth CT images in anatomical structure.

V. DISCUSSION

In this study, the FG and BG regions were segmented based on US images, the real CT images were used as the reference images, and a style transfer deep CNN was used to establish a matching relationship between US and CT images in the FG and BG regions. Further, local pseudo-CT images were synthesized stepwise and finally fused into complete pseudo-CT images. Without the guidance of semantic information, synthesis of pseudo-images based on global images would result in semantic mismatching, which would cause serious deviation between the Hounsfield Unit (HU) values in the FG regions of the pseudo-CT images and those in the FG regions of real CT images, thus affecting the accuracy of radiotherapy planning. Different OARs were segmented as different FG regions based on the US images, and the image synthesis progress in those regions was restricted by adding style loss and content loss functions.

Meanwhile, to ensure that the pseudo-CT image synthesized in the FG region was as similar as the reference CT image in the contours, a contour loss function based on the Dice coefficient was added to constrain the anatomical structure of the pseudo-CT FG regions. According to the experimental results, compared with the methods based on US deformation field and neural style transfer, the pseudo-CT images obtained by the method proposed in this study are more similar to the ground-truth CT images in anatomical structure. The SSIM value was approximately one, the MAE value was low, and the PSNR was high.

According to Ulyanov et al., standardizing a single image instead of a group of images by the example standardization method can significantly improve the quality of image migration, which is equivalent to the batch normalization of batch size 1 [29]. Gatys *et al.* showed that replacement of the max-pooling layer in VGG Net19 into an average-pooling layer could synthesize images with better visual effects [30]. Therefore, we improved the original VGG Net network and added IN before the corresponding ReLU activation layer. Therefore, our study improved the original VGG Net network, replacing the largest pooling layer with the average-pooling layer, and adding IN before the corresponding ReLU activation layer of the network.

Compared with MRI and US images, CT images have a low soft tissue resolution, and the differentiation of soft tissues in the images is not obvious. Therefore, a fully convolutional network was not adopted, and the average-pooling layer was added after the corresponding convolution layer in network to ensure the invariance of features as much as possible and reduce the error of feature extraction [31]. Three-dimensional reconstruction was performed based on the marching cubes algorithm according to the sequence of 2D pseudo-CT images synthesized by the CNN to form 3D pseudo-CT images.

The method of acquiring pseudo images based on US images has a disadvantage in that because US images and CT images are acquired in different periods, the position of the bladder and uterus in the patient's body may change to different degrees, and the pressing of the probe may cause the surface of the patient's body to deform, thereby causing deformation of the bladder. The above unavoidable factors may interfere with the synthesis of pseudo-CT images [32]. To solve this error problem, US robotic arm will be used to replace manual acquisition of US images in subsequent experiments. The US probe will be attached to the robotic arm that performs constant pressure scans over a predetermined orbit. Although neural style transfer can be used to synthesize images in real time on the GPU, owing to the obvious differences in the styles of different modal images, a separate model needs to be trained for each style in advance. A large number of manual operations have great limitations. This study is a preliminary study on the method of neural style transfer to synthesize pseudo-CT images based on

US images. The next study will utilize generative adversarial networks (CycleGAN, etc.) to generate pseudo-CT images.

VI. CONCLUSION

In this study, an ASGNST algorithm was developed, and pseudo-CT images were synthesized in steps. In this algorithm, the CT images obtained in the simulation stage were used as the reference image of the style loss function and the content loss function, and the binary US segmentation images obtained in the treatment stage were used as the reference image of the contour loss function. A matching relationship between US and CT images was established with the segmented FG and BG regions. The local pseudo-CT images were synthesized stepwise and finally merged into global pseudo images, which could be used for brachytherapy for cervical cancer patients. The experimental results show that the proposed algorithm is more accurate than those based on US deformation field or traditional neural style transfer. Pseudo-CT images synthesized based on US images have good application prospects for replacing traditional CT images as guiding images in cervical cancer brachytherapy.

REFERENCES

[1] W. Majewski, I. Wesolowska, H. Urbanczyk, L. Hawrylewicz, B. Schwierczok, and L. Miszczyk, "Dose distribution in bladder and surrounding normal tissues in relation to bladder volume in conformal radiotherapy for bladder cancer," *Int. J. Radiat. Oncol. Biol. Phys.* vol. 75, no. 5, pp. 1371–1378, Dec. 2009.

[2] N. Nesvacil, R. Pötter, A. Sturdza, N. Hegazy, M. Federico, "Adaptive image guided brachytherapy for cervical cancer: A combined MRI-/CT-planning technique with MRI only at first fraction," *Radiotherapy Oncol.*, vol. 107, no. 1, pp. 75–81, Apr. 2013.

[3] S. V. Dyk, "OC-0265: Using MRI and integrated ultrasound to guide brachytherapy for cervix cancer," *Radiotherapy Oncol.*, vol. 111, no. 1, pp. 102–103, Jan. 2014.

[4] Y. Wang, W.-J. Ye, L.-H. Du, A.-J. Li, Y.-F. Ren, and X.-P. Cao, "Dose-volume parameters and clinical outcome of CT-guided free-hand high-dose-rate interstitial brachytherapy for cervical cancer," *Chin. J. Cancer* vol. 31, no. 12, pp. 598–604, Dec. 2012.

[5] T. M. Shepherd, C. P. Hess, C. T. Chin, R. Gould, and W. P. Dillon "Reducing patient radiation dose during CT-guided procedures: Demonstration in spinal injections for pain," *Amer. J. Neuroradiol.*, vol. 32, no. 10, pp. 1776–1782, Nov. 2011.

[6] F. Wang, Q. Tang, G. Lv, F. Zhao, X. Jiang, X. Zhu, X. Li, and S. Yan, "Comparison of computed tomography and magnetic resonance imaging in cervical cancer brachytherapy: A systematic review," *Brachytherapy*, vol. 16, no. 2, pp. 353–365, Mar./Apr. 2017.

[7] L. Su, S. K. Ng, Y. Zhang, T. Ji, I. Iordachita, J. Herman, J. Wong, H. T. Sen, P. Kazanzides, M. A. L. Bell, and K. Ding, "MO-FG-CAMPUS-JeP3-04: Feasibility study of real-time ultrasound monitoring for abdominal stereotactic body radiation therapy," *Med. Phys.*, vol. 43, no. 6, pp. 3727–3728, Jun. 2016.

[8] A. Taggar, T. Phan, L. Traptow, R. Banerjee, and C. M. Doll, "Cervical cancer brachytherapy in canada: A focus on interstitial brachytherapy utilization," *Int. J. Radiation Oncol. Biol. Phys.*, vol. 96, no. 2, pp. 309–310, 2016.

[9] S. Camps, S. van der Meer, F. Verhaegen, and D. Fontanarosa, "Various approaches for pseudo-CT scan creation based on ultrasound to ultrasound deformable image registration between different treatment time points for radiotherapy treatment plan adaptation in prostate cancer patients," *Biomed. Phys. Eng. Express*, vol. 2, no. 3, Jun. 2016, Art. no. 035018.

[10] S. van der Meer, S. M. Camps, W. J. C. van Elmpt, M. Podesta, P. G. Sanches, B. G. L. Vanneste, D. Fontanarosa, and F. Verhaegen, "Simulation of pseudo-CT images based on deformable image registration of ultrasound images: A proof of concept for transabdominal ultrasound imaging of the prostate during radiotherapy," *Med. Phys.*, vol. 43, no. 4, pp. 1913–1920, 2016.

[11] H. Sun, "Imaging study of pseudo-CT images of superposed ultrasound deformation fields acquired in radiotherapy based on step-by-step local registration," *Med. Biol. Eng. Comput.*, vol. 57, no. 3, pp. 643–651, Mar. 2018.

[12] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.

[13] X. Cao, J. Yang, Y. Gao, Q. Wang, and D. Shen, "Region-adaptive deformable registration of CT/MRI pelvic images via learning-based image synthesis," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3500–3512, Jul. 2018.

[14] D. Andreasen, K. Van Leemput, and J. M. Edmund, "A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis," *Med. Phys.*, vol. 43, no. 8, pp. 4742–4752, Aug. 2016.

[15] A. Selim, M. Elgharib, and L. Doyle, "Painting style transfer for head portraits using convolutional neural networks," *Int. Conf. Comput. Graph. Interact. Techn.*, vol. 35, no. 4, Jul. 2016, Art. no. 129.

[16] C. Li and M. Wand, "Combining Markov random fields and convolutional neural networks for image synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2479–2486.

[17] J. Liao, "Visual attribute transfer through deep image analogy," *ACM Trans. Graph.*, vol. 36, no. 4, Jul. 2017, Art. no. 120.

[18] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[19] H. Huang, H. Wang, W. Luo, L. Ma, W. Jiang, X. Zhu, Z. Li, and W. Liu, "Real-time neural style transfer for videos," in *Proc. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7044–7052.

[20] H. Zhao, P. L. Rosin, and Y. K. Lai, "Automatic semantic style transfer using deep convolutional neural networks and soft masks," in *Proc. Visual Comput.*, 2017, pp. 1–18.

[21] A. J. Champandard, "Semantic style transfer and turning two-bit doodles into fine artworks," Mar. 2016, *arXiv:1603.01768*. [Online]. Available: https://arxiv.org/abs/1603.01768

[22] Y. Tang and X. Wu, "Scene text detection and segmentation based on cascaded convolution neural networks," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1509–1520, Mar. 2017.

[23] P. M. Cheng and H. S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," *J. Digit. Imag.*, vol. 30, no. 2, pp. 234–243, Apr. 2017.

[24] Z. Chen-Mccaig, R. Hoseinnezhad, and A. Bab-Hadiashar, "Convolutional neural networks for texture recognition using transfer learning," in *Proc. Int. Conf. Control, Automat. Inf. Sci. (ICCAIS)*, Oct./Nov. 2017, pp. 187–192.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[26] S. Xiang and L. Hao, "Anime style space exploration using metric learning and generative adversarial networks," May 2018, *arXiv:1805.07997*. [Online]. Available: https://arxiv.org/abs/1805.07997

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] X. Cao, "Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis," *Med. Image Anal.*, vol. 41, pp. 18–31, Oct. 2017.

[29] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: Feed-forward Synthesis of textures and stylized images," Mar. 2016, *arXiv:1603.03417*. [Online]. Available: https://arxiv.org/abs/1603.03417

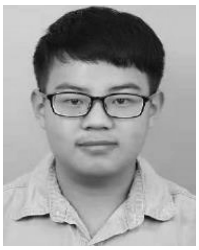
[30] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," Aug. 2015, *arXiv:1508.06576*. [Online]. Available: https://arxiv.org/abs/1508.06576

[31] Y. Boureau, "Learning mid-level features for recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2559–2566.

[32] V. D. M. Skadi, E. B.-van Gurp, J. Hermans, R. Voncken, D. Heuvelmans, C. Gubbels, D. Fontanarosa, P. Visser, L. Lutgens, F. van Gils, and F. Verhaegen, "Critical assessment of intramodality 3D ultrasound imaging for prostate IGRT compared to fiducial markers," *Med. Phys.*, vol. 40, no. 7, Jul. 2013, Art. no. 071707.

HONGFEI SUN received the bachelor's degree from Taishan Medical University and the master's degree from Nanjing Medical University. He is currently pursuing the Ph.D. degree with the School of Automation, Northwestern Polytechnical University. His research interests include deep learning and medical image analysis.

KUN ZHANG received the bachelor's degree from Taishan Medical University and the master's degree from Tsinghua University. His research interests include medical physics and radiotherapy.



RONGBO FAN received the bachelor's degree from Harbin Engineering University. He is currently pursuing the M.S. degree with Northwestern Polytechnical University, China. His research interests include deep learning and image processing.

WENJUN XIONG received the bachelor's degree from Northwestern Polytechnical University, China, where he is currently pursuing the M.S. degree. His research interests include video processing and machine learning.

JIANHUA YANG received the bachelor's degree from Xidian University and the master's and Ph.D. degrees from Northwestern Polytechnical University. She is currently a Professor with the School of Automation, Northwestern Polytechnical University. Her research interests include biomedical image processing, detection, and control technology.