Received: 20 May 2024 | Revised: 27 June 2024 | Accepted: 27 June 2024

DOI: 10.1002/mp.17294

RESEARCH ARTICLE

MEDICAL PHYSICS

Patient-specific deep learning for 3D protoacoustic image reconstruction and dose verification in proton therapy

Yankun Lang[1] | **Zhuoran Jiang**[2] | **Leshan Sun**[3] | **Phuoc Tran**[1] | **Sina Mossahebi**[1] | **Liangzhong Xiang**[3] | **Lei Ren**[1]

[1]Department of Radiation Oncology Physics, University of Maryland, Baltimore, Maryland, USA

[2]Department of Radiation Oncology, Duke University, Durham, North Carolina, USA

[3]Department of Biomedical Engineering and Radiology, University of California, Irvine, California, USA

Correspondence
Lei Ren, Department of Radiation Oncology, University of Maryland, Baltimore, MD 21201, USA.
Email: lren@som.umaryland.edu

Funding information
National Institutes of Health, Grant/Award Numbers: R01 EB032680, R01 EB028324

Abstract

Background: Protoacoustic (PA) imaging has the potential to provide real-time 3D dose verification of proton therapy. However, PA images are susceptible to severe distortion due to limited angle acquisition. Our previous studies showed the potential of using deep learning to enhance PA images. As the model was trained using a limited number of patients' data, its efficacy was limited when applied to individual patients.

Purpose: In this study, we developed a patient-specific deep learning method for protoacoustic imaging to improve the reconstruction quality of protoacoustic imaging and the accuracy of dose verification for individual patients.

Methods: Our method consists of two stages: in the first stage, a group model is trained from a diverse training set containing all patients, where a novel deep learning network is employed to directly reconstruct the initial pressure maps from the radiofrequency (RF) signals; in the second stage, we apply transfer learning on the pre-trained group model using patient-specific dataset derived from a novel data augmentation method to tune it into a patient-specific model. Raw PA signals were simulated based on computed tomography (CT) images and the pressure map derived from the planned dose. The reconstructed PA images were evaluated against the ground truth by using the root mean squared errors (RMSE), structural similarity index measure (SSIM) and gamma index on 10 specific prostate cancer patients. The significance level was evaluated by *t*-test with the *p*-value threshold of 0.05 compared with the results from the group model.

Results: The patient-specific model achieved an average RMSE of 0.014 ($p < 0.05$), and an average SSIM of 0.981 ($p < 0.05$), out-performing the group model. Qualitative results also demonstrated that our patient-specific approach acquired better imaging quality with more details reconstructed when comparing with the group model. Dose verification achieved an average RMSE of 0.011 ($p < 0.05$), and an average SSIM of 0.995 ($p < 0.05$). Gamma index evaluation demonstrated a high agreement (97.4% [$p < 0.05$] and 97.9% [$p < 0.05$] for 1%/3 and 1%/5 mm) between the predicted and the ground truth dose maps. Our approach approximately took 6 s to reconstruct PA images for each patient, demonstrating its feasibility for online 3D dose verification for prostate proton therapy.

Conclusions: Our method demonstrated the feasibility of achieving 3D high-precision PA-based dose verification using patient-specific deep-learning approaches, which can potentially be used to guide the treatment to mitigate

the impact of range uncertainty and improve the precision. Further studies are needed to validate the clinical impact of the technique.

KEYWORDS
dose verification, patient specific modeling, protoacoustic reconstruction

1 | INTRODUCTION

Proton therapy is an advanced radiation treatment that uses protons to disrupt and destroy tumor cells. Once the protons reach the tumor, they deposit most of their energy at a specific depth, known as the Bragg peak, before dropping off sharply. This characteristic allows for better dose distribution and reduced radiation exposure beyond the tumor, making proton therapy particularly beneficial for treating tumors located near critical structures or in pediatric patients.[1] However, the precision of proton therapy is subject to the influence of range uncertainties caused by patient positioning variations, anatomical changes, and uncertainties in dose calculation.[2] Due to the sharp dose falloff at Bragg Peak, even minor deviations in proton dose delivery can lead to substantial underdosing of the intended target and simultaneous overdosing of adjacent healthy tissues. Consequently, the integration of online 3D dose verification methods within the treatment process is regarded as essential in proton therapy to ensure the accurate delivery of radiation and optimize treatment efficacy.[3] In recent years, various in-vivo dose verification methods have been developed to validate the range of proton dose delivery. These methods encompass the utilization of wireless implantable dosimeters to measure dose or fluence,[4–6] as well as the delivery and imaging of separate proton beams.[7–9] An alternative approach involves the measurement of surrogate data generated during proton irradiation to verify the deposition of proton doses. Notably, positron emission tomography (PET)[10–13] and prompt gamma (PG) imaging[14–17] have been employed for this purpose, detecting the gamma rays produced along the path of the proton beam. However, these techniques possess certain limitations. PET imaging techniques do not offer real-time online dose verification during treatment, as they necessitate post-treatment data analysis. Additionally, while PG imaging provides information about the gamma rays generated during proton beam irradiation, its accuracy is compromised due to low signal intensity and the absence of comprehensive 3D volumetric data.[16,18] These limitations underscore the need for further advancements in online dose verification methods for improved accuracy and real-time monitoring during proton therapy treatment.

In recent years, protoacoustic (PA) imaging has emerged as a promising approach for dose verification in proton therapy. PA imaging involves the detection of radiofrequency (RF) signals generated by the proton beam during dose deposition. Various studies have been conducted to explore the potential of PA imaging for dose verification, employing simulation-based approaches using 2D computed tomography (CT) images,[19–21] as well as experimental investigations utilizing matrix array transducers,[22,23] where the initial pressure map is reconstructed from the RF signals, and then related to the dose deposition. Traditional algorithms for reconstructing pressure maps from PA signals have been proposed, such as Time Reversal (TR)[24,25] and the Universal Back Projection (UBP) method.[26] However, reconstructed PA pressure maps still suffer from distortion and artifacts due to limited-angle views of the matrix array detectors, which limits their accuracy for dose verification.[27]

To address these challenges in image reconstruction, deep learning-based methods have been introduced. For instance, Zhu et al. used fully connected (FC) layers to map the dual-domain correlations.[28] This method faces memory limitations when dealing with high-resolution protoacoustic images. Häggström et al. utilized convolutional layers instead of FC layers to reduce memory consumption.[29] Zhang et al. proposed a self-supervised learning method for ultrasound image reconstruction, minimizing the difference between the sinogram projected from the reconstructed image and the initially measured sinogram.[30] Jiang et al. utilized a 3D U-net to enhance an initial pressure map reconstructed by the TR method to solve the limited-angle view issue.[31] However, the efficacy of deep learning enhancement depends heavily on the quality of the initial reconstruction, which is hindered by severe distortion and the loss of detailed anatomical structures. Moreover, the TR method is time-consuming, rendering this approach impractical for real-time online dose verification. More recently, Lang et al.[32] proposed a hybrid supervised deep learning method, where a Recon-Enhance two-stage strategy is applied to achieve improved reconstruction quality and efficiency. Specifically, in the Recon-stage, a transformer-based network is proposed to directly reconstruct the initial map from RF signals. Transfer learning is employed to ensure consistency in data fidelity. Subsequently, in the Enhance-stage, a 3D U-net is applied to improve the reconstruction quality further. The reconstructed dose distributions obtained through this approach exhibited improved accuracy and efficiency, offering promising prospects for real-time dose verification during treatment.

Despite the advancements in enhancing the quality and efficiency of PA image reconstruction, a notable limitation in previous studies lies in the potential suboptimal performance of deep learning models trained on a limited number of patients. This limitation can result in errors when applying the models to specific patients, primarily due to inter-patient anatomical variations, thereby compromising the accuracy of derived dose deposition. Furthermore, for PA imaging, due to the limited angle view issues, some essential information may be lost when applying a model trained using a limited group of patients on a specific patient, limiting the efficacy of the model. By leveraging the pre-knowledge and data of each patient, a patient-specific model can better account for the individual variations and optimize the accuracy of dose verification for each patient. Also, the information lost due to limited angle view issues can be relieved from the patient-specific data. Recently, there has been a growing interest in patient-specific training[33–38] within the field of medical physics as it offers potential benefits in improving the accuracy and applicability of models for individual patients. Elmahdy et al.[35] introduced a patient-specific method aimed at improving the segmentation accuracy for prostate cancer treatment. The approach applied transfer learning by fine-tuning the last layer of a pre-trained baseline Convolutional Neural Network (CNN) model using patient-specific data derived from previous treatment fractions. Jansen et al.[36] also proposed the same patient-specific method for lesion quantification and the segmentation of brain white matter hyperintensities. Both methods fine-tuned the group model using the initial scan of multiple patients and evaluated the patient-specific model on the remaining scans without any augmentation. However, these methods are not suitable for our study due to the limited number of CT scans available per patient (1–9). Additionally, they are tailored to a subgroup of patients rather than individual patients. Notably, Chun et al.[37] proposed a patient-specific framework named Intentional Deep Overfit Learning (IDOL) and applied it on three tasks: auto contouring on radiation pneumonitis CT (rpCT), super-resolution (SR)-based magnetic resonance image (MRI) reconstruction and synthetic CT reconstruction. Specifically, a generalized model with a diverse training dataset was trained, then a patient-specific dataset was derived by augmentation and subsequently used to fine-tune the entire network weights. Results showed great improvement for the three tasks, indicating the efficacy of the approach when applied to a specific patient. Jiang et al.[38] proposed a patient-specific deep learning method applied for enhancing digital tomosynthesis(DTS) for lung radiotherapy. For each specific patient, the deformation vector fields were calculated based on the patient's scans taken on various days. Subsequently, the first three principal motion components were estimated using the principal component analysis (PCA) technique. These components were then assigned varying weights to create a dataset used for training the patient-specific model.

Inspired by the previous research, in this study, we develop a patient-specific deep learning method, aiming at improving the reconstruction quality of PA imaging, as well as the accuracy of the dose verification. Specifically, as shown in Figure 1, our method consists of two stages: in the first stage, a group model is trained using a population based training set, where a novel deep learning network[32] is employed to directly reconstruct the initial pressure maps from the RF signals; in the second stage, we apply transfer learning to tune the pre-trained group model into a patient-specific model using a patient-specific dataset derived by a novel data augmentation method. The patient-specific model will then be tested for individual patients. In contrast to the approach described in reference,[37] which focuses on image-to-image translation tasks, our work addresses the unique challenge of signal-to-image transformation. As such, conventional augmentation techniques such as random cropping are not applicable in our work due to inconsistency between signal-image domains. Moreover, deformation with random displacement vector fields (DVFs) may compromise the integrity of anatomical structures specific to individual patients, which are unsuitable for our patient-specific task. The main contributions of our article are multi-fold: (1) Extending the group model[32] to a patient-specific deep learning approach for PA imaging; (2) Introduction of a pioneering intra-patient augmentation method, facilitating data expansion while preserving intra-feature consistency; (3) Evaluation of the proposed methodology utilizing protoacoustic data derived from CT images and clinical treatment plans of patients with prostate cancer, thereby demonstrating the feasibility of high-precision 3D dose verification in proton therapy for individual patients.

2 | METHODS

2.1 | Domain transfer reconstruction network

We employ Domain Transfer Reconstruction Network (DTR-Net) in ref. [32] as shown in Figure 2 to directly the reconstruct initial pressure map from RF signals. Same as U-Net, DTR-net utilizes a contracting-expanding architecture. By taking the 3D RF image and its corresponding first-order derivative image as input, the DTR-Net utilizes residual transformer[39–42] blocks (RTBs) to capture long-range dependencies in the input data. Each RTB consists of several 3D Swin transformer (ST) layers. The number of STs used in each RTB increases progressively, starting from 2 and doubling at each

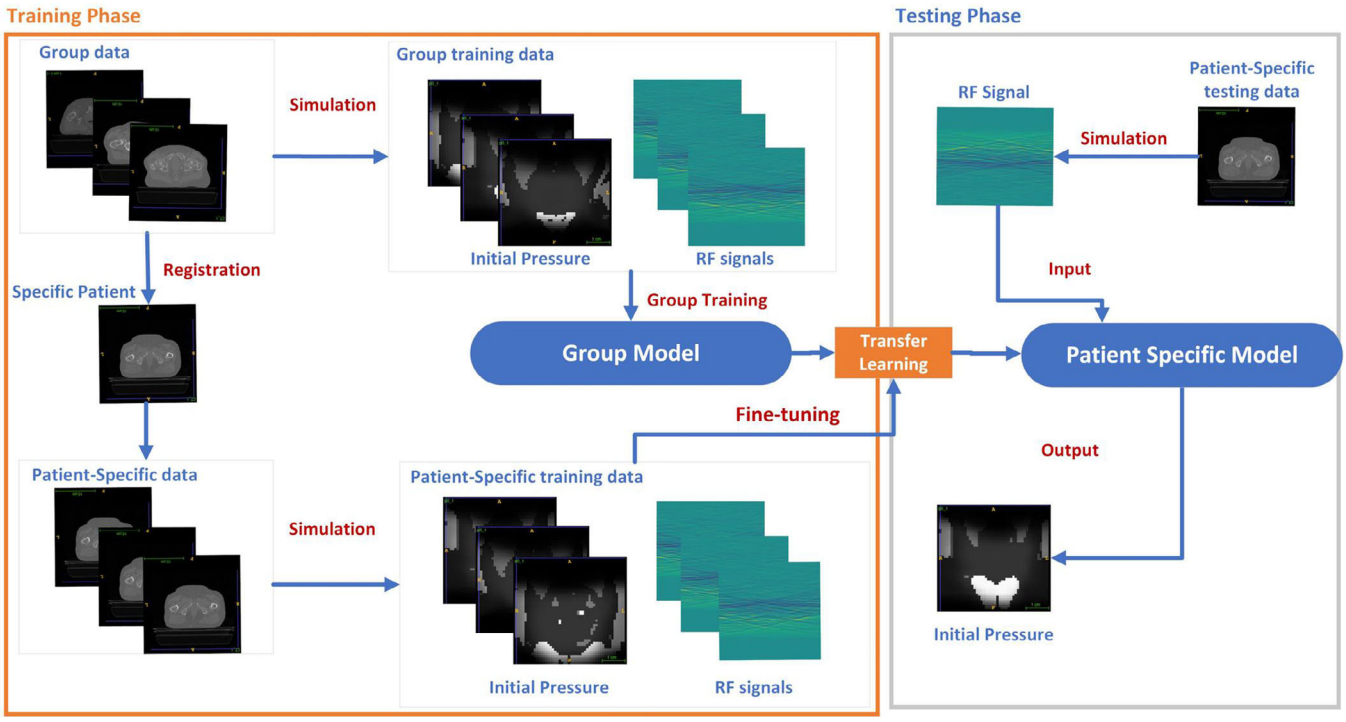

FIGURE 1 The overview of the proposed patient-specific method.

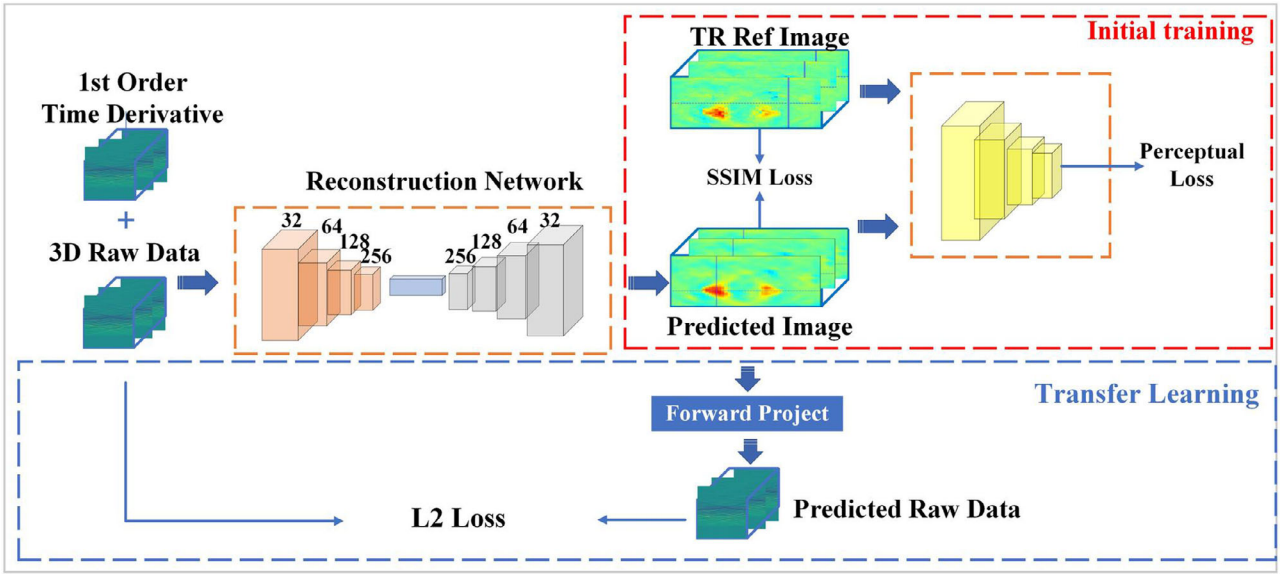

FIGURE 2 The architecture of DTR network for the initial pressure reconstruction from RF signals. DTR, domain transfer reconstruction; RF, radiofrequency.

subsequent RTB (i.e., 4, 8, and 16 STs). The extracted features are transformed from the signal domain to the image domain using a convolution layer. This step helps in transferring the learned features to a format suitable for image reconstruction. The expanding path of the network consists of four residual blocks. Each block includes an up-sampling layer followed by two consistent 3D convolution layers with a $3 \times 3 \times 3$ kernel. ReLU activation and Group normalization layers are applied after the convolution layers to introduce non-linearity and normalize the feature maps. Finally, the reconstructed initial pressure map is obtained by applying a convolution layer with a $1 \times 1 \times 1$ kernel to the features obtained from the expanding path.

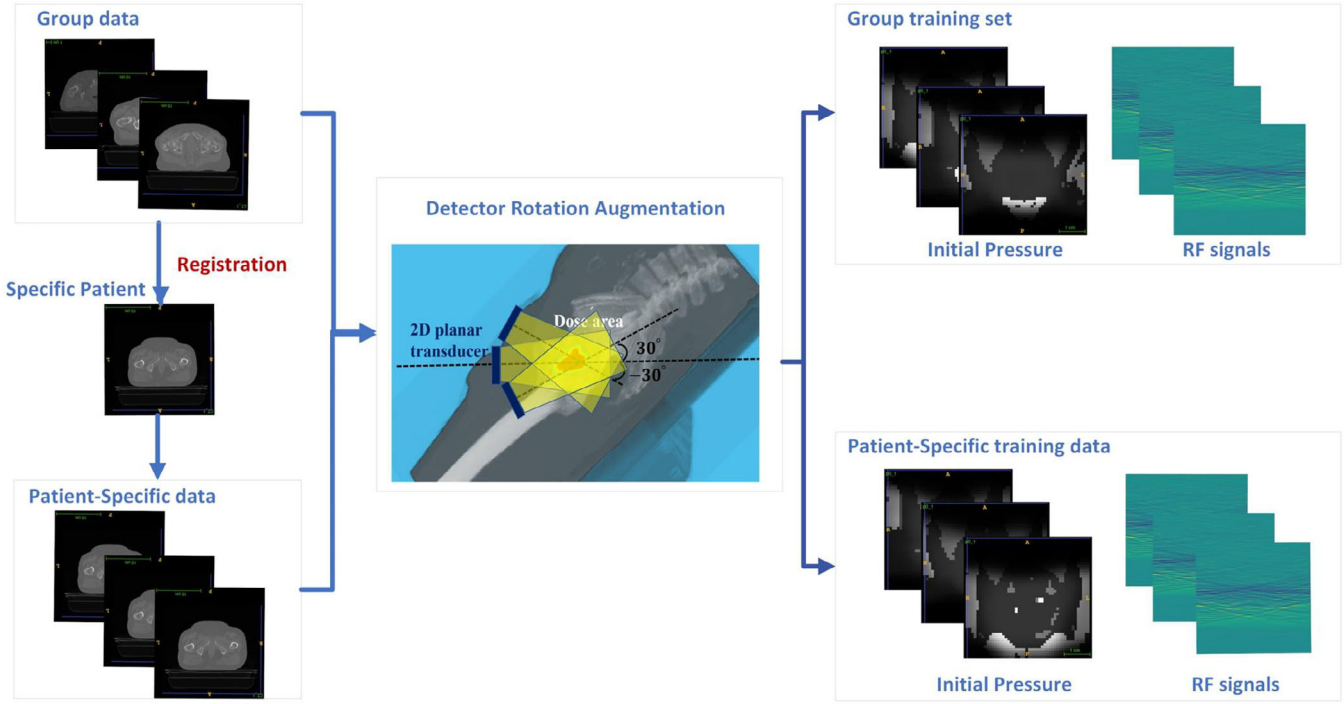

FIGURE 3 The overview of the augmentation of group data and patient-specific data.

In the Recon-stage, the network is trained with a hybrid supervision method: (1) During the initial training, we use the pressure map reconstructed by iterative TR method as ground truth to initially train the model by using SSIM loss and perceptual loss. This step enables DTR-Net to focus on discovering the most representative features for fast reconstruction. (2) After the initial training, the network is fine-tuned using transfer learning and self-supervision based on data fidelity constraint. In this step, the reconstructed pressure maps are projected back to the signal domain to generate RF signals, which are then compared with the input RF signals using the *l*2 loss function. Specifically, we leverage the k-wave toolbox[43] to simulate the forward procedure. Subsequently, during backpropagation, TR reconstruction is utilized, mirroring the procedural steps employed in generating the TR reconstruction outcomes. The integration of the forward and the backward MATLAB functions into our Python framework is enabled through the "transplant" interface. By fine-tuning the network based on the data fidelity constraint, the focus shifts to solely reconstructing images based on the raw data without relying on the imperfect supervision provided by the TR reconstructed images in the initial training. This fine-tuning step helps to improve the overall reconstruction quality by aligning the reconstructed RF signals more closely with the input RF signals.

In the Enhance-stage, a 3D U-net architecture was employed to further improve the quality of the reconstructed images, where the initial pressure map is used as ground truth. Once the enhanced images are obtained, they are further converted into dose values for proton verification purposes.

2.2 | Patient-specific intra-data augmentation

During the patient-specific training process, each patient has 2–10 diagnostic scans that can be leveraged as prior knowledge to improve the performance of patient-specific models. However, training a patient-specific model using a small dataset can lead to overfitting and limited generalization over patient day-to-day anatomical variations when applying the model to reconstruct patient images on different days.

To address the aforementioned challenge, numerous studies have employed data augmentation techniques to expand the size of their datasets. However, in our specific work, the conventional data augmentation methods such as rotation, translation, and rescaling are not adequate since they do not represent the daily deformations commonly seen in patient data. In the following, we used two data augmentation methods to avoid overfitting. Initially, as depicted in Figure 3, we conduct the registration of all other patients' data to the first scan of each specific patient. An alternative approach involves registering the first scan to all other patients' data. In this scenario, the intensity of the image remains the same, however, the anatomical features of specific patients are subject to deformation to conform to the anatomical characteristics of other patients, resulting

in a lack of specificity within the generated dataset. Consequently, the data generated using this approach has a variety of "inter-patient variations" rather than "intra-patient variations". In our method, aligning the patients' data with the first scan of the target patient ensures a shared anatomical structure (SSIM >0.95) while adjusting the intensity. To introduce some variation and enhance the robustness of the trained model, we allow for slight deformations in the registered data, which helps to account for natural anatomical variations and potential discrepancies in positioning or patient-specific characteristics between different individuals. By including such variations in the augmented dataset, the model is able to learn to recognize and adapt to different scenarios, such as diverse anatomical structures. By exposing the model to a more extensive range of examples through data augmentation, it becomes more robust and better equipped to handle unseen data during inference.[44] In this work, we employ affine registration as the initial step to align all other patients' data to the target scan, ensuring a coarse correspondence between anatomical structures. Then we utilize non-rigid deformation using a 3D B-spline transformation[45] to perform local deformations and fine adjustments to enhance the alignment accuracy. We empirically choose a B-spline interpolation order of 5, as it provides the necessary flexibility to model complex deformations while ensuring smoothness in the transformed data. By applying the deformation map, we can achieve a more precise alignment of anatomical structures, accounting for patient-specific variations. Additionally, the same deformation map derived from the B-spline transformation is applied to the dose maps for dose verification. This ensures that the dose maps are consistent with the registered patient data, enabling accurate verification of the applied treatment dose.

Also, we utilized simulation methods to generate additional RF-P0 (radiofrequency and initial pressure map) pairs, effectively expanding the training set. An additional augmentation process was employed to vary the positions of the PA (photoacoustic) detector in the perineum area. Specifically, the PA detector was initially positioned below the prostate and near the perineum area, with an initial tilt angle of $\frac{\pi}{6}$ degrees. Subsequently, the detector was rotated along the lateral axis, covering the entire prostate area, with different angles uniformly sampled within the range of $[-\frac{\pi}{6}, \frac{\pi}{6}]$. This process is visually represented in Figure 3 of the study. For each sampled angle, protoacoustic simulation procedures were conducted to generate the corresponding RF signals from P0 maps. The augmentation process was repeated 20 times, with equally spaced angles selected for each patient. This approach ensured a diverse range of simulated detector positions and captured a wider array of acoustic responses.

2.3 | Training implementation and inference

2.3.1 | Group model training

We utilize the same training strategy as proposed in ref. [32] to train the group model. The models in both the Recon-Enhance stages were trained by the ADAM optimizer with an initial learning rate of 0.001. The learning rate was reduced by a factor of 5 after every 500 000 epochs. During this phase, the reconstructed initial pressure map was compared to the corresponding TR-reconstructed reference by calculating SSIM loss and perceptual loss. Then the transfer learning process commenced for another 1 000 000 epochs by calculating the SSIM loss between the predicted RF signal and the input RF signal. Following the transfer learning phase, the Enhance-stage focused on training the enhancement network for another 1 00 0000 epochs. This stage aimed to further refine the model's performance and optimize the enhancement capabilities. The network was implemented based on PyTorch framework version 1.7.1, a 40 GB Nvidia A100 server GPU with CUDA Toolkit 12.0 and a 64 GB RAM.

2.3.2 | Patient-specific model training

To train the patient-specific model, we utilize the same training strategy as applied for the group model with all training weights being frozen except for the last layer. Based on transfer learning, only the last layer in the network is fine-tuned since it capture more task-specific features, while the early layers of the model tend to capture more general features representing correlation from signal domain to image domain. By only fine-tuning the last layer,[35,36,46] the group model can be adapted to the patient-specific task without overfitting or losing the valuable general features learned from the group model. Besides, we skip the initial training since the model was already trained from the group data. Specifically, we reduce the learning rates to 0.0005 and fine-tune the model using the augmented patient-specific dataset for 1 000 000 epochs where only the SSIM loss between the predicted RF signals and the input RF signals are calculated. In the Enhance-stage, the last layer of the enhancement network is fine-tuned for additional 500 000 epochs to calculate the L2 loss between the reconstructed initial pressure map and the corresponding ground truth, with the parameters of other layers kept frozen.[35,36]

By employing this training approach for the patient-specific model, we aim to leverage the pre-existing knowledge of the group model while tailoring the model's parameters to the specific characteristics of

TABLE 1 Tissue-specific parameter setting for RF signal simulation. v, ρ and Γ refer to the speed of sound, tissue density and the Grüneisen parameter, respectively.

Tissue	HU value	v (m/s)	ρ (kg/m^3)	Γ	$\rho \times \Gamma$ (kg/m^3)	α (dB/cm/MHz)
Air	[−1000, −200)	-	-	-	-	-
Water	air overwritten	1500	1000	0.11	110	0.0022
Fat	[−200, −50)	1480	920	0.80	736	0.5
Soft tissue	[−50,100)	1540	1040	0.30	312	1
Bone	[100, max)	2000	1900	0.80	1520	10

Note: α denotes the attenuation coefficient.
Abbreviation: RF, radiofrequency.

individual patients. The fine-tuning process, guided by SSIM loss and L2 loss calculations, enables the model to adapt to the patient-specific dataset and enhances its performance in terms of RF signal prediction and initial pressure map reconstruction.

2.3.3 | Dose conversion

We utilize a 3D U-net model with the same architecture and training setting as described in ref. [32] for deriving dose maps from the reconstructed initial pressure maps. An initial dose map is computed by dividing the reconstructed pressure map by a dose conversion coefficient map derived from patient CT images. By leveraging the U-net model and its ability to capture complex relationships in 3D data, the final dose map is generated by combining the initial dose map with the predicted residual errors. This process ensures a precise representation of the actual dose distribution, accounting for any discrepancies or errors that may have been present in the initial estimation. A group dose verification model is first trained from the group dataset for 1 000 000 epochs by ADAM optimizer with an initial learning rate of 0.001. Then the last layer of this model is fine-tuned using the patient-specific dataset for another 500 000 epochs with a learning rate of 0.0005.

2.3.4 | Inference

During the inference phase, the trained patient-specific model is utilized to reconstruct the pressure map from RF data acquired through limited angle PA imaging. Subsequently, the reconstructed pressure map is further enhanced by the enhancement network, resulting in the generation of the final PA images. This recon-enhance approach exhibits efficient processing capabilities, with each 3D RF signal image, sized at $32 \times 32 \times 112$, being processed in less than 6 s. Finally, dose map is directly derived from the trained dose verification model by taking the reconstructed initial pressure map as input. The total procedures only takes less than 6 s under an environment of 40 GB Nvidia A100 GPU and 64 GB of RAM.

2.4 | Experiment configuration

In this study, we collected a dataset comprising 126 anonymized patients diagnosed with prostate cancer under the approval of an Institutional Review Board (IRB) protocol, ensuring compliance with ethical guidelines and regulations. Each patient's data in the dataset consists of a planning CT scan and the corresponding clinical treatment plan. The dose map for each treatment plan was obtained using a commercial software named RayStation (RaySearch Laboratories, Stockholm/Sweden), then normalized to the maximum dose, providing a standardized representation of the delivered radiation dose. Prior to further analysis, each CT scan in the dataset underwent a segmentation process to categorize the different tissue types present in the scan, including air, fat, soft tissue, and bone according to the predefined HU value thresholding empirically.[31,32,47] For the purposes of our study, various tissue-specific parameters were predefined and utilized. The tissue-specific parameters including density, speed of sound, and the Grüneisen parameter[31,32,48] are given in Table 1.

By multiplying the dose map with the tissue density and the Grüneisen parameter, the initial pressure (P0) was computed as:

$$P0 = dose\_map \times \rho \times \Gamma, \quad (1)$$

Then, the acoustic simulation process for generating RF signals was conducted using the open-source k-wave toolbox within the Matlab environment. Specifically, a planar detector with dimensions of 8 cm×,8 cm and a 64 × 64 ultrasound transducer array was simulated. The detector was positioned below the prostate and near the perineum area, with a tilt angle of 30 degrees. This configuration was chosen to cover the prostate region while avoiding interference from the pelvic bones. The transducer array was designed with a central frequency of 500 kHz, a bandwidth of 100, and

TABLE 2 Quantitative analysis of the reconstruction results of initial pressure maps and dose verification.

Modality	Method	MSE	PSNR (dB)	SSIM	Speed (s)
PA image	Group model[32]	0.029 ± 0.008	30.37 ± 0.076	0.962 ± 0.013	6
	Patient-specific	**0.014 ± 0.005**	**35.37 ± 0.036**	**0.981 ± 0.021**	6
Dose Verification	Group model[32]	0.018 ± 0.009	34.86 ± 0.27	0.989 ± 0.007	6
	Patient-specific	**0.011 ± 0.004**	**36.26 ± 0.057**	**0.995 ± 0.003**	6

Note: All our results achieved a significance level $p < 0.05$.
Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

a sampling rate of 5 MHz. These settings determined the frequency characteristics and sampling resolution of the acquired RF signals. During the simulation, tissue-specific heterogeneity and attenuation were considered in the propagation of the acoustic signals. To mimic real-world conditions, a Gaussian white noise with a signal-to-noise ratio (SNR) of 10 dB was added to the acquired RF signals.

For the training of the DTR-Net, the time-reversal (TR) method was employed. It involved 10 iterations to reconstruct the initial pressure maps from the simulated RF signals. These reconstructed pressure maps served as the ground truth for the initial training of the group model. Both the initial pressure map (P0) and the dose map were used as ground truth for training the pressure map enhancement network and the dose conversion network, respectively. To ensure consistency and comparability, both the pressure map and dose map were resampled to a resolution of $2.50 \times 2.50 \times 1.25\,\text{mm}^3$, with a size of $48 \times 48 \times 112$. Additionally, the simulated RF signal was resampled to a size of $32 \times 32 \times 112$ to reduce memory consumption while still retaining essential information.

A total of 116 patients were selected for training in the group model. The CT scan of each patient was augmented by positioning the detector with 20 different angles uniformly sampled within the range of $[-\frac{\pi}{6}, \frac{\pi}{6}]$ during the simulation. Finally, a group dataset consisting of 2320 pairs of the generated RF signals with the corresponding P0 maps and the derived dose maps are used for training the group model with the training/validation splitting of 5:1, where the number of training/validation is 1856 and 464, respectively. We use the remaining 10 patients, each with 2–9 CT scans from different time phases, to generate the patient-specific dataset. For each patient, we perform data augmentation as mentioned in Section 2.2 on the CT scan from the first time phase, resulting in a total of 2744 training cases, where the split of training/validation is 5:1, with the number of 2195 and 548. Once the patient-specific model was trained on this dataset, it was evaluated and tested on the remaining CT scans from different days that were not used during the training process. These CT scans are Quality Assurance CTs (QACTs) that are acquired during the middle of treatment course to verify anatomy changes. In this work, we trained one group model, and fine-tuned it on 10 patient-specific datasets, separately.

3 | RESULTS

3.1 | Pressure map reconstruction results

To evaluate the performance of our patient-specific model, we compared the reconstructed initial pressure map with the ground truth by calculating the root mean squared errors (RMSE). The RMSE metric provided an indication of the average difference between the corresponding pixel values in the two maps. Lower RMSE values indicated a higher level of agreement and better reconstruction accuracy. Additionally, we also employed peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) metrics to further investigate the reconstruction quality on details and basic structure reconstruction. SSIM assessed the similarity between the predicted and ground truth maps based on both structural information and luminance. It compared the structural organization of the two maps and provided a measure of the similarity in terms of texture, edges, and overall structural patterns.

The overall quantitative results of pressure map reconstruction are summarized in Table 2. The qualitative results are also shown in Figure 4. The group model, trained without employing patient-specific augmentation method, serves as a baseline for comparison. It achieved an average RMSE value of 0.029 ($p < 0.05$), indicating a significant decrease in the discrepancy between the predicted and ground truth pressure maps. The SSIM was 0.962 ($p < 0.05$), showing a high similarity of anatomic structure compared with the ground truth. The observed improvement in the reconstruction quality can be attributed to the utilization of the transform network for quality enhancement, as well as the implementation of transfer learning to maintain data fidelity consistency. However, some challenging details were still not reconstructed, while blur effect still existed on the whole structure. Our patient-specific model further improved the reconstruction accuracy by reducing the RMSE value to 0.014 ($p < 0.05$), indicating a significant improvement in the accuracy of pressure map reconstruction compared to the group model. As depicted in Figure 4, the patient-specific model successfully reconstructed most of the challenging details, while also eliminating the blurring effect observed in the group

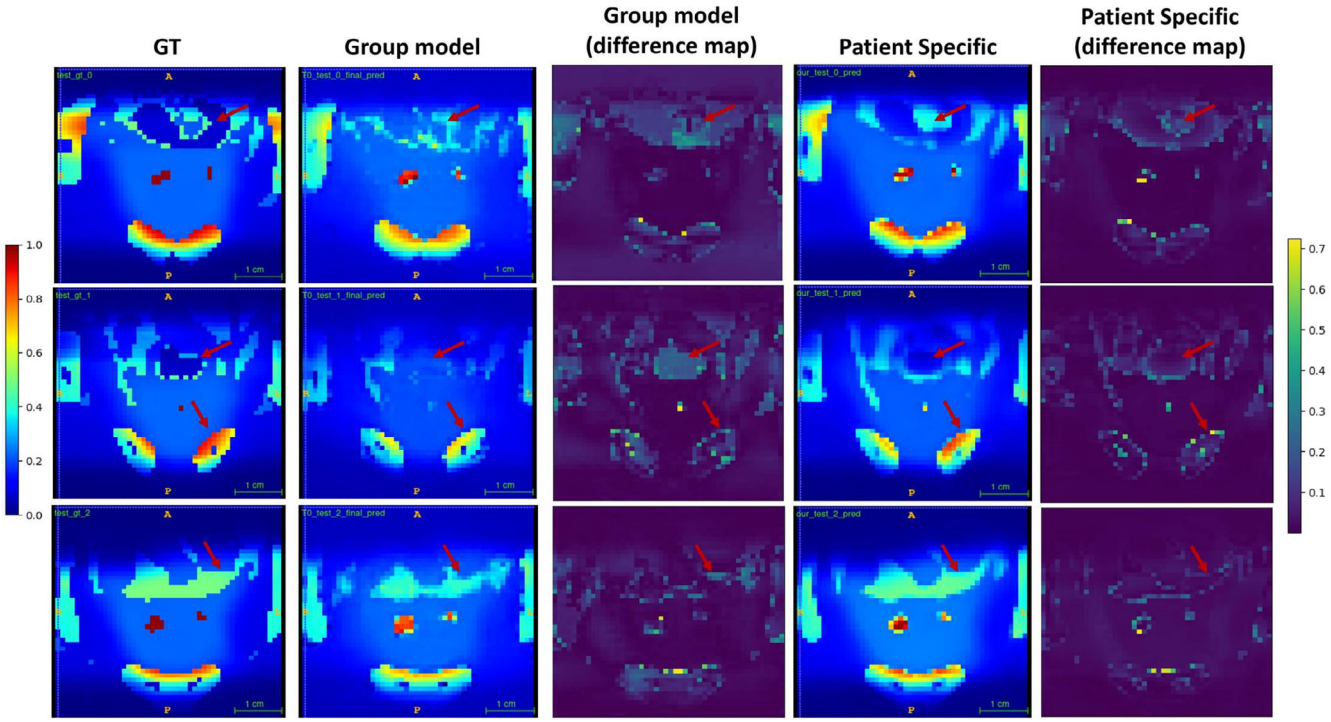

FIGURE 4 Example P0 map reconstruction results (normalized). From left to right: Color bar of the P0 map, ground truth, results by group model, difference maps between ground truth and results from group model, results by our approach, difference maps between ground truth and our results, and colorbar of the difference maps.

TABLE 3 Quantitative analysis (Gamma index) of the reconstruction results of dose maps.

Modality	Metric	Group model[32]	Patient-specific
Dose	Gamma index (3%/3 mm)	99.3% ± 0.4%	**99.4% ± 0.4%**
	Gamma index (3%/5 mm)	99.6% ± 0.3%	**99.6% ± 0.2%**
	Gamma index (2%/3 mm)	97.1% ± 1.9%	**98.6% ± 1.2%**
	Gamma index (2%/5 mm)	97.8% ± 2.0%	**98.8% ± 0.8%**
	Gamma index (1%/3 mm)	94.7% ± 1.7%	**97.4% ± 1.5%**
	Gamma index (1%/5 mm)	95.7% ± 1.4%	**97.9% ± 1.3%**

Note: All our results achieved a significance level $p < 0.05$.

model. Furthermore, the SSIM value increased to 0.981 ($p < 0.05$), indicating a higher level of similarity in the anatomical structures compared to the ground truth. These findings confirm the effectiveness of utilizing transfer learning on patient-specific data to enhance the reconstruction accuracy. Our approach also achieved a testing speed of approximately 6 s. This rapid processing time makes our method suitable for online dose verification in proton therapy, enabling real-time assessment and monitoring of treatment accuracy.

3.2 | Dose conversion results

We compared the dose maps that were predicted from the pressure maps reconstructed by group model and the patient-specific model, in terms of RMSE, PSNR and SSIM. Table 2 also gives the quantitative results. Specifically, the patient-specific model reduces the RMSE value from 0.018 in the group model to 0.011 ($p < 0.05$), indicating a decrease in the discrepancy between the predicted and ground truth dose maps. Additionally, the SSIM value increases from 0.989 to 0.995 ($p < 0.05$), highlighting a high level of similarity between the predicted and the ground truth 3D dose maps. These findings confirm the superior accuracy of the dose maps restored by our method. Our approach achieved significant improvements in the gamma index values compared to the ground truth as shown in Table 3. Specifically, for the 3%/3 mm criterion, the gamma index increased from 99.3% to 99.4% ($p < 0.05$). Similarly, for the 3%/5 mm criterion, the gamma index improved from 99.6% to 99.6% ($p < 0.05$). For the 2%/3 mm and 2%/5 mm criteria, our approach increased the gamma

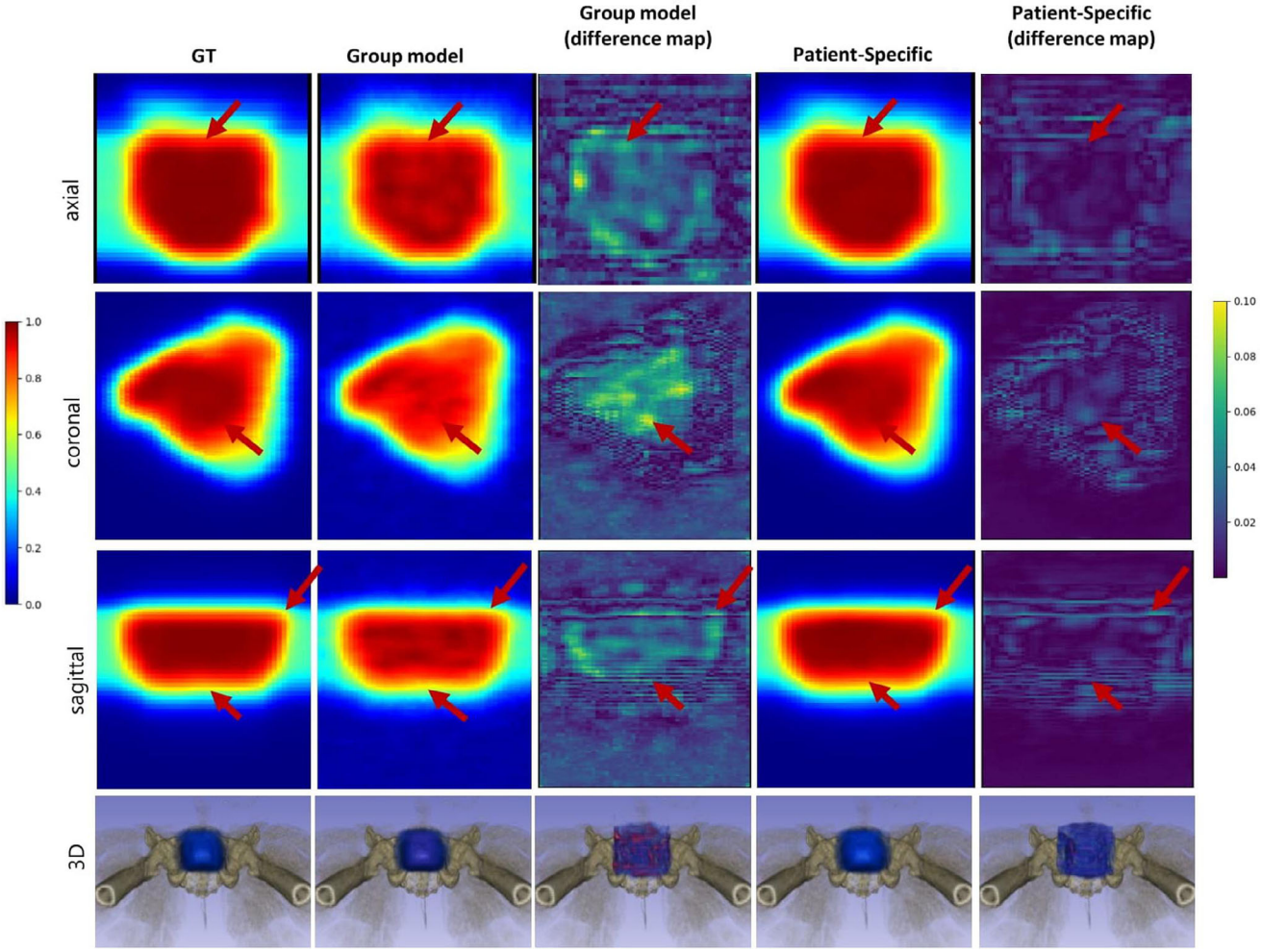

FIGURE 5 Example dose map reconstruction results (normalized). From left to right: color bar of the dose map, ground truth, results by group model, difference maps between ground truth and results from group model, results by our approach, difference maps between ground truth an our results, and colorbar of the difference maps.

index from 97.1% to 98.6% ($p < 0.05$) and from 97.8% to 98.8% ($p < 0.05$), respectively. Notably, our approach demonstrated excellent performance in terms of the gamma index for the 1%/3 mm and 1%/5 mm criteria, achieving rates of 97.4% ($p < 0.05$) and 97.9% ($p < 0.05$), respectively. These high gamma index rates indicate a strong agreement between the predicted and ground truth dose maps, further highlighting the effectiveness and accuracy of our approach.

Qualitative results of several challenging cases are shown in Figure 5, where the dose maps reconstructed using our approach exhibit higher accuracy compared to alternative methods. This improved accuracy can be attributed to the high quality of the input pressure maps obtained from our proposed method. The qualitative analysis further supports the effectiveness and reliability of our approach in predicting dose maps accurately.

We compared the dose volume histogram (DVH) of organ at risks (OARs) of the predicted dose map with the ground truth. The DVHs of the target organ (prostate) and the OARs including rectum, bladder and small bowel are plotted in Figure 6 with different colors. Notably, DVHs derived from the ground truth dose map were represented by solid lines, while those derived from the predicted dose map were depicted using dotted lines. Our analysis revealed a noteworthy consistency between the DVH curves obtained from the predicted and ground truth datasets. Notably, the DVH for the prostate demonstrated that approximately 95% of the organ's volume received the maximum dose. Conversely, for the rectum, bladder, and small bowel, approximately 35%, 34%, and 3.9% of their volumes respectively received doses surpassing 50% of the maximum prescribed dose. These findings serve to confirm the efficacy of our approach in predicting dose distributions.

3.3 | Statistic analysis

To investigate the significance of the results for both PA reconstruction and dose verification, we conduct statistic analysis with two null hypotheses, under the assumption

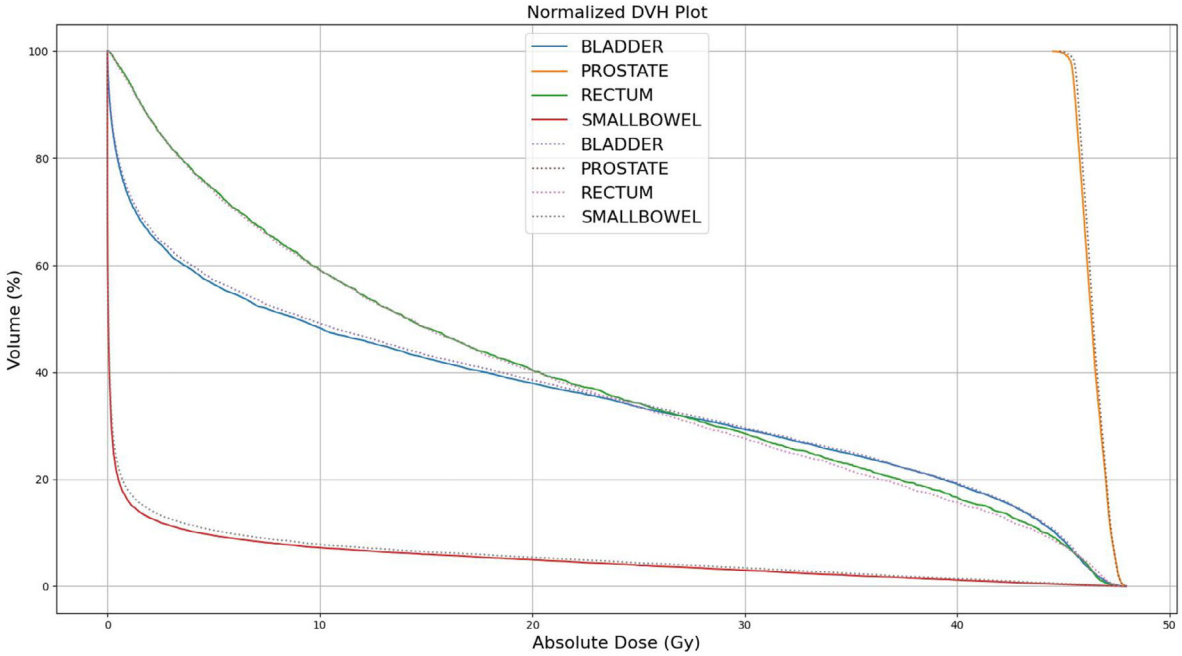

FIGURE 6 DVHs of ROI (prostate) and OARs (rectum, baldder and small bowel). DVHs, dose volume histograms; OARs, organ at risks.

TABLE 4 *p*-Value calculated by *t*-testing for each quantitative results.

Test index	Number of cases	P_1-RMSE	P_1-PSNR	P_1-SSIM	P_2-RMSE	P_2-PSNR	P_2-SSIM
test 1	4	1.14e-4	5.54e-5	2.13e-5	5.7e-6	3.84e-4	3.17e-5
test 2	5	3.69e-5	2.68e-5	2.60e-5	1.85e-5	3.9e-4	1.06e-4
test 3	4	2.24e-5	2.89e-5	2.78e-5	1.12e-5	1.3e-4	1.30e-4
test 4	4	2.55e-4	5.13e-5	4.24e-5	1.27e-6	1.44e-5	1.39e-4
test 5	7	2.78e-4	6.90e-5	2.52e-5	1.06e-5	2.56e-4	2.12e-5
test 6	3	2.12e-4	2.58e-5	1.24e-5	1.41e-5	3.45e-5	1.31e-4
test 7	4	2.82e-5	3.90e-5	8.50e-5	1.26e-5	1.29e-4	6.22e-6
test 8	4	2.53e-4	3.81e-5	1.81e-5	1.23e-4	1.95e-4	4.25e-5
test 9	9	2.47e-4	1.24e-5	8.59e-5	1.49e-5	6.21e-4	9.08e-5
test 10	4	9.8e-6	7.52e-6	6.34e-5	2.23e-5	3.76e-4	4.29e-7
entire	48	2.03e-7	4.03e-6	1.41e-6	1.01e-6	2.01e-5	7.09e-6

Note: P_1 corresponds to PA imaging reconstruction results, and P_2 corresponds to dose verification results.
Abbreviations: PA, protoacoustic; PSNR, peak signal-to-noise ratio; RMSE, root mean squared errors; SSIM, structural similarity index measure.

that the results of both methods are random variables following normal distribution: (1) the quantitative results of the proposed method in terms of RMSE, PSNR, and SSIM has no difference with the results of the group model (benchmark method); (2) the quantitative results of the proposed method in terms of RMSE, PSNR, and SSIM has worse performance than those of the group model. The significance level was empirically set to 0.05. The *p*-value was calculated by *t*-testing, and then corrected by Bonferroni multi-comparison correction method.[49] Specifically, we performed the analysis on each testing patient (totally 10 patients), and calculated their corresponding *p*-values using *t*-testing. Then, according to Bonferroni multi-comparison correction, each *p*-value was multiplied by the number of the testing sets. The corrected *p*-value of each testing set is shown in Table 4. We did not list the *p*-values under the first hypothesis since they are much smaller than the significance level $p = 0.05$. For the second null hypothesis, we can see that all corrected *p*-value are less than 0.05, without any unfavorable results or non-significant results as shown in the table (all *p*-values < alpha) indicating that our model achieved a significantly improved performance.

4 | DISCUSSION

4.1 | Pressure and dose reconstruction for protoacoustic imaging

The deep learning network that is used in our research as the group model has demonstrated high reconstruction accuracy while mitigating the limited angle view issues that exist in traditional reconstruction methods. However, due to inner-patient variation, the group model trained from a group of patients is not optimal for specific patients. Specifically, the inability to reconstruct details in challenging locations may be attributed to the inherent complexity and variability of the data in those regions. Factors such as acoustic attenuation, tissue heterogeneity, or limitations in the imaging technique itself can contribute to difficulties in accurately capturing and representing fine details. Furthermore, the blurring effect observed in the overall structure could be attributed to the inherent trade-off between preserving fine details and maintaining the group's structural consistency. While efforts were made to enhance the quality of the reconstructed pressure maps, there may be limitations in striking the perfect balance between capturing fine details and preserving the overall structure.

The results of our study indicate that the patient-specific model outperformed the group model in terms of pressure map reconstruction and dose prediction accuracy. Specifically, our method employed transfer learning by fine-tuning the last layer of a pre-trained model using the patient-specific dataset. The transfer learning allowed the model to leverage the knowledge gained from the group model while adapting to the specific characteristics and variations present in the patient-specific data. By fine-tuning the model on the patient-specific dataset, we were able to capture and incorporate the specific nuances and variations present in the individual patient's data, leading to improved accuracy and personalized predictions. This method enables us to leverage both group knowledge and patient-specific information, resulting in superior performance compared to the group model.

4.2 | Limitations and future work

It is important to acknowledge that our approach was evaluated using simulated data instead of real patient data due to the lack of real patient experiments for this new imaging technique that is still in the development stage. While simulated data offers the advantage of providing ground truth information for evaluation purposes, it may not fully capture the complexity and variability present in real patient scenarios. The simulation parameters were set based on empirical knowledge to approximate real data characteristics as closely as possible. To validate the clinical efficacy of our technique, further studies using real patient data and experimental setups are warranted. Real patient studies can provide insights into the performance and generalizability of the approach in diverse clinical scenarios and patient populations. In the future, clinical trials will be carried out to acquire actual patient data to further validate the feasibility, accuracy, and potential benefits of our approach in real-world applications.

Finally, we plan to explore different methods of deformation registration to introduce more anatomical variations into the patient-specific dataset. This will help to enlarge the dataset and capture a wider range of anatomical characteristics while maintaining anatomical consistency. Additionally, we are interested in incorporating cone-beam computed tomography (CBCT) images into the patient-specific dataset. Since CBCT (daily information) images are more routinely acquired than CT scans, using CBCT images for data augmentation can be more effective and will be investigated in our future work.

In this work, we use prostate cancer treatment as an example to demonstrate the feasibility and advantages of the technique. Furthermore, we aim to extend the application of our approach to other imaging modalities. This will involve evaluating the generalization capability of the proposed network by applying it to different image modalities, such as magnetic resonance imaging (MRI) or positron emission tomography (PET). Verifying the performance of our approach on different imaging modalities will demonstrate its versatility and potential for broader clinical applications.

5 | CONCLUSION

In this work, we have proposed a patient-specific deep learning method to reconstruct PA images for proton therapy dose verification. A group model is trained from a diverse patient set, then fine-tuned by a patient-specific dataset through transfer learning. The results show that our patient-specific method achieve better performance for PA imaging reconstruction and dose verification compared to the group model. Meanwhile, our approach achieved superior performance on reconstructing 3D dose with a fast processing speed, making it very practical for online 3D dose verification in proton therapy.

ACKNOWLEDGEMENTS

This work was sponsored in part by National Institutes of Health grants R01 EB032680 and R01 EB028324.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data can be made available upon request.

REFERENCES

1. Knopf A-C, Antony L. In vivo proton range verification: a review. *Phys Med Biol*. 2013;58(15):R131-R160.
2. Liebl J, Paganetti H, Zhu M, Winey BA. The influence of patient positioning uncertainties in proton radiotherapy on proton range and dose distributions. *Med Phys*. 2014;41(9):091711.
3. Paganetti H, Botas P, Sharp GC, Winey B. Adaptive proton therapy. *Phys Med Biol*. 2021;66(22):22TR01.
4. Lu H, Mann G, Cascio E. Investigation of an implantable dosimeter for single-point water equivalent path length verification in proton therapy. *Med Phys*. 2010;37(11):5858-5866.
5. Bentefour EH, Shikui T, Prieels D, Lu HM. Effect of tissue heterogeneity on an in vivo range verification technique for proton therapy. *Phys Med Biol*. 2012;57(17):5473-5484.
6. Telsemeyer J, Jäkel O, Martišíková M. Quantitative carbon ion beam radiography and tomography with a flat-panel detector. *Phys Med Biol*. 2012;57(23):7957-7971.
7. Schneider U, Besserer J, Pemler P, et al. First proton radiography of an animal patient. *Med Phys*. 2004;31(5):1046-1051.
8. Schneider U, Pedroni E, Hartmann M, Besserer J, Lomax T. Spatial resolution of proton tomography: methods, initial phase space and object thickness. *Z Med Phys*. 2012;22(2):100-108.
9. Penfold SN, Rosenfeld AB, Schulte RW, Schubert KE. A more accurate reconstruction system matrix for quantitative proton computed tomography. *Med Phys*. 2009;36(10):4511-4518.
10. Fiedler F, Priegnitz M, Jülich R, et al. In-beam PET measurements of biological half-lives of 12C irradiation induced $\beta \pm$ activity. *Acta Oncol*. 2008;47(6):1077-1086.
11. Fiedler F, Shakirin G, Skowron J, et al. On the effectiveness of ion range determination from in-beam PET data. *Phys Med Biol*. 2010;55(7):1989-1998.
12. Nishio T, Miyatake A, Ogino T, Nakagawa K, Saijo N, Esumi H. The development and clinical use of a beam ON-LINE PET system mounted on a rotating gantry port in proton therapy. *Int J Radiat Oncol Biol Phys*. 2010;76(1):277-286.
13. Miyatake A, Nishio T, Ogino T, Saijo N, Esumi H, Uesaka M. Measurement and verification of positron emitter nuclei generated at each treatment site by target nuclear fragment reactions in proton therapy. *Med Phys*. 2010;37(8):4445-4455.
14. Min CH, Lee HR, Kim CH, Lee SB. Development of array-type prompt gamma measurement system for in vivo range verification in proton therapy. *Med Phys*. 2012;39(4):2100-2107.
15. Polf JC, Peterson S, Ciangaru G, Gillin M, Beddar S. Prompt gamma-ray emission from biological tissues during proton irradiation: a preliminary study. *Phys Med Biol*. 2009;54(3):731-743.
16. Draeger E, Mackin D, Peterson S, et al. 3D prompt gamma imaging for proton beam range verification. *Phys Med Biol*. 2018;63(3):035019.
17. Kormoll T, Fiedler F, Schöne S, Wüstemann J, Zuber K, Enghardt W. A Compton imager for in-vivo dosimetry of proton beams–A design study. *Nucl Instrum Methods Phys Res, Sect A*. 2011;626-627:114-119.
18. Polf JC, Avery S, Mackin DS, Beddar S. Imaging of prompt gamma rays emitted during delivery of clinical proton beams with a Compton camera: feasibility studies for range verification. *Phys Med Biol*. 2015;60(18):7085-7099.
19. Freijo C, Herraiz JL, Sanchez-Parcerisa D, Udias JM. Dictionary-based protoacoustic dose map imaging for proton range verification. *Photoacoustics*. 2021;21:100240.
20. Yao S, Hu Z, Xie Q, Yang Y, Peng H. Further investigation of 3D dose verification in proton therapy utilizing acoustic signal, wavelet decomposition and machine learning. *Biomed Phys Eng Express*. 2021;8(1):015008.
21. Yu Y, Li Z, Zhang D, Xing L, Peng H. Simulation studies of time reversal-based protoacoustic reconstruction for range and dose verification in proton therapy. *Med Phys*. 2019;46(8):3649-3662.
22. Yu J, Yoon H, Khalifa YM, Emelianov SY. Design of a volumetric imaging sequence using a vantage-256 ultrasound research platform multiplexed with a 1024-element fully sampled matrix array. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2019;67(2):248-257.
23. Wang M, Samant P, Wang S, et al. Toward in vivo dosimetry for prostate radiotherapy with a transperineal ultrasound array: a simulation study. *IEEE Trans Radiat Plasma Med Sci*. 2020;5(3):373-382.
24. Treeby BE, Zhang EZ, Cox BT. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Prob*. 2010;26(11):115003.
25. Hristova Y, Kuchment P, Nguyen L. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Prob*. 2008;24(5):055006.
26. Xu M, Wang LV. Universal back-projection algorithm for photoacoustic computed tomography. *Phys Rev EStat Nonlin Soft Matter Phys*. 2005;71(1):016706.
27. Parodi K, Polf JC. In vivo range verification in particle therapy. *Med Phys*. 2018;45(11):e1036-e1050.
28. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555(7697):487-492.
29. Häggström I, Schmidtlein CR, Campanella G, Fuchs TJ. DeepPET: a deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med Imag Anal*. 2019;54:253-262.
30. Zhang J, He Q, Xiao Y, Zheng H, Wang C, Luo J. Ultrasound image reconstruction from plane wave radio-frequency data by self-supervised deep neural network. *Med Image Anal*. 2021;70:102018.
31. Jiang Z, Sun L, Yao W, Wu QJ, Xiang L, Ren L. 3D in vivo dose verification in prostate proton therapy with deep learning-based proton-acoustic imaging. *Phys Med Biol*. 2022;67(21):215012.
32. Lang Y, Jiang Z, Sun L, Xiang L, Ren L. Hybrid-supervised deep learning for domain transfer 3D protoacoustic Image reconstruction. *Phys Med Biol*. 2024;69(8):085007.
33. Zhang Z, Huang M, Jiang Z, et al. Patient-specific deep learning model to enhance 4D-CBCT image for radiomics analysis. *Phys Med Biol*. 2022;67(8):085003.
34. Zhang Z, Huang M, Jiang Z, et al. Patient-specific synthetic magnetic resonance imaging generation from cone beam computed tomography for image guidance in liver stereotactic body radiation therapy. *Precis Radiat Oncol*. 2022;6(2):110-118.
35. Elmahdy MS, Ahuja T, van der Heide UA, Staring M. Patient-specific finetuning of deep learning models for adaptive radiotherapy in prostate CT. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2020.
36. Jansen MJ, Kuijf HJ, Dhara AK, et al. Patient-specific fine-tuning of convolutional neural networks for follow-up lesion quantification. *J Med Imaging*. 2020;7(6):064003.
37. Chun J, Park JC, Olberg S, et al. Intentional deep overfit learning (IDOL): a novel deep learning strategy for adaptive radiation therapy. *Med Phys*. 2022;49(1):488-496.
38. Jiang Z, Yin F-F, Ge Y, Ren L. Enhancing digital tomosynthesis (DTS) for lung radiotherapy guidance using patient-specific deep learning model. *Phys Med Biol*. 2021;66(3):035009.
39. Parmar N, Vaswani A, Uszkoreit J, et al. Image transformer. In: *International Conference on Machine Learning*. PMLR; 2018:4055-4064.
40. Matsoukas C, Haslum JF, Söderberg M, Smith K. Is it time to replace cnns with transformers for medical images? arXiv preprint arXiv:2108.09038. 2021.
41. Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of*

the IEEE/CVF International Conference on Computer Vision. 2021:9992-10002.
42. Huang J, Wu Y, Wu H, Yang G. Fast MRI reconstruction: How powerful transformers are? In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2022:2066-2070.
43. Treeby BE, Cox BT. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *J Biomed Opt*. 2010;15(2):021314.
44. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6. doi:10.1186/s40537-019-0197-0
45. Zhong H, Jinkoo K, Chetty IJ. Analysis of deformable image registration accuracy using computational modeling. *Med Phys*. 2010;37(3):970-979.
46. Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. In: *International Conference on Machine Learning*. PMLR; 2015.
47. Mast TD. Empirical relationships between acoustic parameters in human soft tissues. *Acoust Res Lett Online*. 2000;1(2):37-42.
48. Yao D-K, Zhang C, Maslov K, Wang LV. Photoacoustic measurement of the Grüneisen parameter of tissue. *J Biomed Opt*. 2014;19(1):017007-017007.
49. Bland JM, Douglas GA. Multiple significance tests: the Bonferroni method. *Bmj*. 1995;310(6973):170.

How to cite this article: Lang Y, Jiang Z, Sun L, et al. Patient-specific deep learning for 3D protoacoustic image reconstruction and dose verification in proton therapy. *Med Phys*. 2024;51:7425–7438.
https://doi.org/10.1002/mp.17294