



Learning-Based Stopping Power Mapping on Dual-Energy CT for Proton Radiation Therapy

Tonghe Wang, PhD; Yang Lei, PhD; Joseph Harms, PhD; Beth Ghavidel, MS; Liyong Lin, PhD; Jonathan J. Beitler, MD; Mark McDonald, MD; Walter J. Curran, MD; Tian Liu, PhD; Jun Zhou, PhD; Xiaofeng Yang, PhD

Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA, USA

Abstract

Purpose: Dual-energy computed tomography (DECT) has been used to derive relative stopping power (RSP) maps by obtaining the energy dependence of photon interactions. The DECT-derived RSP maps could potentially be compromised by image noise levels and the severity of artifacts when using physics-based mapping techniques. This work presents a noise-robust learning-based method to predict RSP maps from DECT for proton radiation therapy.

Materials and Methods: The proposed method uses a residual attention cycle-consistent generative adversarial network to bring DECT-to-RSP mapping close to a 1-to-1 mapping by introducing an inverse RSP-to-DECT mapping. To evaluate the proposed method, we retrospectively investigated 20 head-and-neck cancer patients with DECT scans in proton radiation therapy simulation. Ground truth RSP values were assigned by calculation based on chemical compositions and acted as learning targets in the training process for DECT datasets; they were evaluated against results from the proposed method using a leave-one-out cross-validation strategy.

Results: The predicted RSP maps showed an average normalized mean square error of 2.83% across the whole body volume and an average mean error less than 3% in all volumes of interest. With additional simulated noise added in DECT datasets, the proposed method still maintained a comparable performance, while the physics-based stoichiometric method suffered degraded inaccuracy from increased noise level. The average differences from ground truth in dose volume histogram metrics for clinical target volumes were less than 0.2 Gy for $D_{95\%}$ and D_{max} with no statistical significance. Maximum difference in dose volume histogram metrics of organs at risk was around 1 Gy on average.

Conclusion: These results strongly indicate the high accuracy of RSP maps predicted by our machine-learning-based method and show its potential feasibility for proton treatment planning and dose calculation.

Keywords: stopping power; dual-energy CT; proton therapy; machine learning

Introduction

Proton radiation therapy has been one of the emerging treatment modalities that may have better clinical outcomes for a wide range of patients due to favorable dosimetric properties related to the Bragg peak and virtually no exit dose compared with photon

Submitted 25 Mar 2020
Accepted 04 Dec 2020
Published 12 Feb 2021

Corresponding Author:

Xiaofeng Yang, PhD
Department of Radiation
Oncology
Emory University School of
Medicine
1365 Clifton Road NE
Atlanta, GA 30322, USA
Phone: +1 (404) 778-8622
Fax: +1 (404) 778-4139
xiaofeng.yang@emory.edu

Original Article

DOI
10.14338/IJPT-D-20-00020.1

© Copyright
2021 The Author(s)

Distributed under
Creative Commons CC-BY

OPEN ACCESS

<http://theijpt.org>

radiation therapy [1–4]. The calculation of proton dose based on computed tomography (CT) simulation images requires the conversion from the Hounsfield unit (HU) numbers to the relative stopping power (RSP) for different materials [5–8].

One of the currently implemented methods is to calibrate RSP based on the HU number on tissue characterization phantoms with known atomic compositions and electron densities. A 1-to-1 relationship between HU and RSP can then be established by a piecewise linear function. However, a direct HU-RSP calibration may introduce ambiguity since tissues with different combinations of atomic composition and electron density may have the same attenuation, which may cause inaccuracy in determining radiation absorption properties for proton dose calculations in treatment planning [9]. Moreover, the approximation of real tissue with tissue substitutes in phantom also introduces error due to the differences in chemical composition [10].

Recently, dual-energy CT (DECT) has been introduced to radiation therapy simulation for its ability to provide material-specific information by differentiating the energy dependence of photoelectric and Compton interactions of different materials [11–14]. Parametric maps, such as RSP, electron density (ρ_e), effective atomic number Z_{eff} , and mean excitation energy (I), can be derived from DECT images in a voxel-wise manner using physical equations [15, 16]. However, the physics-based method can be very sensitive to the noise and artifacts on the DECT images since the overlapping of the 2 energy spectra would turn the system into an ill-posed problem. On the other hand, in addition to the noise and artifacts present in single-energy CT, DECT suffers extra noise and artifacts caused by non-ideal dual-energy dataset acquisition schemes, such as patient motion artifacts in 2-sequential-scan DECT [17, 18] and cross scatter artifacts in dual-source CT [19, 20]. Meanwhile, twin-beam DECT (TBCT) has been introduced into radiation therapy simulation for its good temporal coherence, full field of view, and low hardware complexity and cost, while it has poorer energy spectra separation compared with other DECT modalities [17, 21, 22]. The strong overlapping of energy spectra of linear attenuation coefficient among different materials would lead to significant noise magnification from the acquired projection to the results of material differentiation [12, 23–25]. The physical derivation does not accommodate these non-idealities and would magnify the noise and artifacts on the DECT images to the derived parametric maps that directly lead to uncertainty and inaccuracy.

Inspired by the success of the development of machine learning in recent years, novel methods have been developed to convert between images presenting similar anatomy but different modalities [26–35]. Due to the data-driven properties, learning-based methods are expected to be more robust to noise. These algorithms have also been introduced for parametric map generation from DECT [9]. In the study by Su et al [9], models are trained by a large number of pairs of DECT images and corresponding parametric maps using different algorithms, and then predict parametric maps from a new DECT image input. However, the training model is based on phantom of tissue substitutes but is used for predicting clinical patient datasets. The approximation of real tissue by tissue substitutes still exists, and the learning on phantom with piecewise values on simple geometry may neglect the heterogeneities and complexities present in real patient datasets. Thus, an appropriate training model with ground truth based on human scans is of interest.

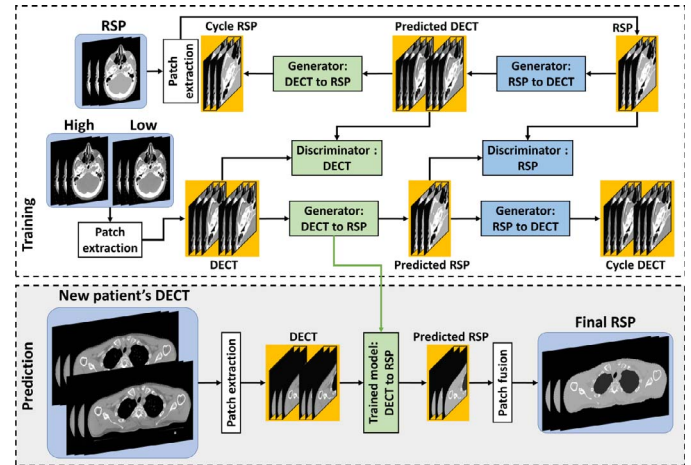
In this work, we propose a novel machine-learning-based method to predict RSP maps from DECT for proton radiation therapy. The aim of this study was to provide an alternative approach to the physics-based method with more noise robustness. In our method, we integrated a residual attention architecture into cycle-consistent generative adversarial networks (cycle-GANs). Compared with other machine-learning-based methods, the advantages include automatic extraction of deep features from DECT images and use of residual blocks to force the model to focus on the differences between RSP and DECT. Moreover, a deep attention strategy was integrated into the network architecture to highlight the informative features that would well represent the difference between RSP and DECT images. Third, cycle-GANs were used to let the DECT-to-RSP mapping be close to a 1-to-1 mapping by introducing an inverse RSP-to-DECT mapping. We retrospectively investigated 20 head-and-neck cancer patients with twin-beam DECT scan acquired during CT simulation and treated with proton radiation therapy. Ground-truth RSP values were assigned by calculation based on chemical compositions [9, 10, 36]. These ground-truth RSP maps acted as learning targets in the training process for DECT datasets and were evaluated against results from the proposed method using a leave-one-out cross-validation strategy. The accuracy of predicted RSP maps by the proposed method was quantified with multiple quality and dosimetric metrics. All results generated by the proposed method were compared with the physics-based stoichiometric method.

Materials and Methods

Workflow

Figure 1 outlines the schematic workflow of our prediction method. The DECT is able to provide 2 CT image datasets acquired with low- and high-energy spectra. For a given series of low and high DECT images and their corresponding RSP maps, the

Figure 1. The schematic flowchart of the proposed method. The first row shows the training stage. The second row shows the prediction stage. DECT, dual-energy computed tomography; RSP, relative stopping power.

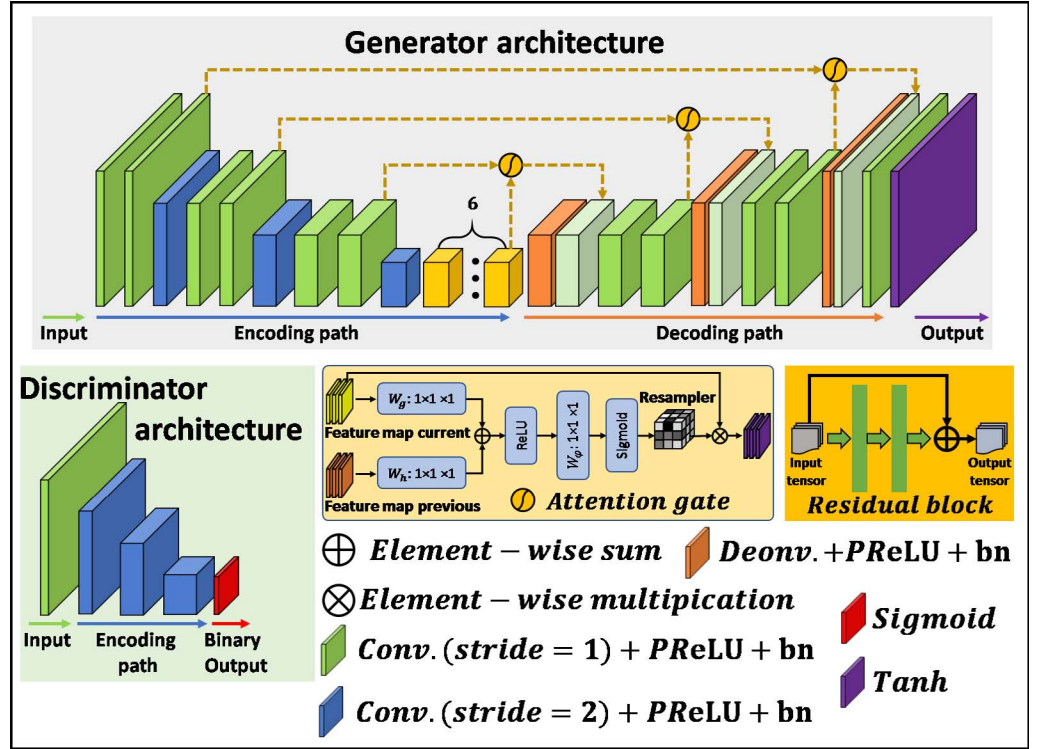


RSP maps were used as the regression target. The use of both energy images gives the model higher fidelity since the CT values at 2 energy levels demonstrate 2 different energy dependences of material attenuation, which indicates material information. During training, $96 \times 96 \times 32$ voxel patches were extracted from low- and high-energy images by a sliding window with an overlap of $80 \times 80 \times 22$ between any 2 neighboring patches. Furthermore, data augmentations were used to enlarge the training data variation. Then, prepared patches were fed into a 3-dimensional (3D) deep learning framework as a 2-channel input and mapped to a 1-channel output patch, which was correlated to training target, that is, RSP image. To enforce a 1-to-1 mapping, a cycle-GAN-based framework was used to introduce an inverse mapping, which took the RSP map as 1-channel input and mapped it to low- and high-energy CT images as 2-channel output. In order to learn the specific differences among these 3 datasets, 6 residual blocks were used as short skip connections in the end-to-end U-Net-based generator architecture. To further highlight significant features that can fully represent these 3 image datasets, 3 attention gates were used for the other 3 long skip connections as shown in the generator architecture in **Figure 2**. To judge the realism of the predicted maps, a fully convolutional network architecture-based discriminator was used to discriminate the true image from a predicted image. After training, the paired patches of low and high DECT images were extracted from a new arrival patient's DECT dataset and are fed into the trained networks to obtain the predicted RSP patches. Finally, by using patch fusion, the RSP map of a new arrival patient's DECT was predicted.

Network Architecture

Figure 2 shows the network architecture of both generators and discriminators used in the proposed method. Both generators share same architecture with different trainable weights. Both discriminators (discriminator of RSP and discriminator of DECT) share same architecture with different trainable weights. We used a traditional fully convolutional network to implement the discriminators [34]. The generator architecture (of both DECT-to-RSP and RSP-to-DECT) is an end-to-end U-Net, including encoding and decoding paths. The encoding path is composed of 3 convolution layers with a stride size of 2 to reduce the feature maps size and several further convolution layers with stride size of 1. The decoding path is composed of 3 deconvolution layers to enlarge the feature map size and match the size of input image (end-to-end), several further convolution layers with stride size of 1, and a tanh layer to perform the regression. To collect features extracted from encoding path and decoding path with same feature map size, 1 short skip connection and 3 long skip connections were used to concatenate equal-sized feature map that is extracted from encoding and decoding convolution layers, respectively. The short connection included 6 residual blocks. The aim of using residual block is to lead the network focusing on the difference between source and target images. The details of implementation can be shown in Harm et al [37]. As shown in **Figure 2**, the long skip connection was implemented by concatenating these feature maps extracted from the layer of encoding path with same-sized feature maps extracted from the layer of decoding path. Attention gate could capture the most relevant semantic (segment) information without enlarging the receptive field [31]. Since in this work, the target RSP image is close to a semantic image, we proposed integrating attention gate into the long skip connection to highlight the semantic features from feature maps extracted from the previous layer of encoding path. The details of the implementation of attention gate can be found in our previous work [31].

Figure 2. The generator and discriminator network. W_g and W_h are two linear transformations in attention gates. Deconv, Deconvolution; PReLU, Parametric Rectified Linear Unit; bn, Batch normalization; Conv, Convolution.



Loss Function

By using the specific design of loss functions for generators and discriminators, the learnable parameters of the proposed network can be learned iteratively via Adam gradient descent minimization of these losses. Because generators and discriminators have different goals in this work (the generator's goal is to generate a synthetic image that can fool the discriminators; the discriminator's goal is to discriminate synthetic from real), in each iteration, the generators and discriminators were optimized in an alternative manner. For generator loss, in order to fool the discriminators, adversarial loss was used. Adversarial loss relies on the output of the discriminators, that is, the distribution of feeding synthetic RSP image (generated from DECT-to-RSP generator $G_{\text{DECT-RSP}}$) into the discriminator of RSP and the distribution of feeding synthetic DECT image (generated from RSP-to-DECT generator $G_{\text{RSP-DECT}}$) into the discriminator of DECT. For clarity, we present only formulation for $G_{\text{DECT-RSP}}$.

$$(1) \quad L_{\text{adv}}(G_{\text{DECT-RSP}}, D_{\text{RSP}}, I_{\text{DECT}}, I_{\text{RSP}}) = \text{SCE} \left[D_{\text{RSP}} \left(G_{\text{DECT-RSP}}(I_{\text{DECT}}) \right), 1 \right],$$

where I_{DECT} is the DECT 2-channel image and $G_{\text{DECT-RSP}}(I_{\text{DECT}})$ the output of the DECT-to-RSP generator, that is, the predicted RSP. D_{RSP} is the RSP discriminator, which is designed to return a distribution indicating whether a pixel region is real (from RSP) or fake (from predicted RSP). The function $\text{SCE}(\cdot, 1)$ is the sigmoid cross entropy between the discriminator map of the predicted RSP and a unit mask.

In addition, to guide the generators to learn a mapping that could be closed to a 1-to-1 mapping, cycle consistency loss was used. In this work, mean absolute error (MAE) and gradient difference error (GDE) were used as a compound loss to calculate the cycle consistency loss of generator [34]. The MAE loss focuses on the accuracy of voxel intensity level. The GDE losses focus on the accuracy of image structure,

$$(2) \quad L_{\text{cyc}}(G_{\text{DECT-RSP}}, G_{\text{RSP-DECT}}, I_{\text{DECT}}) = \text{MAE} \left[G_{\text{RSP-DECT}} \left(G_{\text{DECT-RSP}}(I_{\text{DECT}}) \right), I_{\text{DECT}} \right] + \lambda \cdot \text{GDE} \left[G_{\text{RSP-DECT}} \left(G_{\text{DECT-RSP}}(I_{\text{DECT}}) \right), I_{\text{DECT}} \right],$$

where λ is a balancing parameter that controls the MAE and GDE loss for cycle consistency. $G_{\text{RSP-DECT}}(G_{\text{DECT-RSP}}(I_{\text{DECT}}))$ is the output of first feeding I_{DECT} into the generator $G_{\text{DECT-RSP}}$ and then feeding the output into the generator $G_{\text{RSP-DECT}}$; namely, the output of this term denotes the cycle RSP.

Table 1. Ground-truth relative stopping power of different materials calculated based on chemical compositions published in ICRP Publication 23 [36].

	Relative stopping power
Air	0
Lung	0.26
Adipose	0.96
Muscle	1.04
Brain	1.06
Skeleton: spongiosa	1.16
Skeleton: cortical bone	1.63

Finally, the optimization of generator is obtained by

$$(3) \quad G_{\text{DECT-RSP}}, G_{\text{RSP-DECT}} = \underset{G_{\text{DECT-RSP}}, G_{\text{RSP-DECT}}}{\operatorname{argmin}} \left\{ \begin{array}{l} L_{\text{adv}}(G_{\text{DECT-RSP}}, D_{\text{RSP}}, I_{\text{DECT}}, I_{\text{RSP}}) + L_{\text{adv}}(G_{\text{RSP-DECT}}, D_{\text{DECT}}, I_{\text{RSP}}, I_{\text{DECT}}) \\ L_{\text{cyc}}(G_{\text{DECT-RSP}}, G_{\text{RSP-DECT}}, I_{\text{DECT}}) + L_{\text{cyc}}(G_{\text{DECT-RSP}}, G_{\text{RSP-DECT}}, I_{\text{DECT}}) \end{array} \right\}.$$

The optimization of discriminator is obtained by

$$(4) \quad (D_{\text{RSP}}, D_{\text{DECT}}) = \underset{D_{\text{RSP}}, D_{\text{DECT}}}{\operatorname{argmin}} \left\{ \begin{array}{l} SCE[D_{\text{RSP}}(G_{\text{DECT-RSP}}(I_{\text{DECT}})), 0] + SCE[D_{\text{RSP}}(I_{\text{RSP}}), 1] + \\ SCE[D_{\text{DECT}}(G_{\text{RSP-DECT}}(I_{\text{RSP}})), 0] + SCE[D_{\text{DECT}}(I_{\text{DECT}}), 1] \end{array} \right\}.$$

Data Acquisition

In this study, we retrospectively analyzed the images from 20 patients with squamous cell carcinoma in the head and neck region. The 20 patients included 13 men and 7 women ranging in age from 25 to 89 years. The locations of tumors in these 20 patients included larynx, buccal mucosa, tongue, etc, and 12 of them underwent excisions. Each patient had CT simulation by TBCT in DECT mode with a 110-second delay after 100 mL Omnipaque 300 iodine contrast injected at 2.5 mL/s. Their targets and organs-at-risk (OARs) were delineated by physicians. Institutional review board approval was obtained with no informed consent required for this Health Insurance Portability and Accountability Act-compliant retrospective analysis.

The DECT images were acquired by a Siemens SOMATOM Definition Edge twin-beam CT scanner (Erlangen, Germany) at 120 kVp with the patient in treatment position (pitch: 0.45, rotation time: 0.5 seconds, scan time: around 30 seconds, CTDIvol: around 20 mGy, reconstruction algorithm: SAFIRE, reconstruction kernel: Q30f, tube current ranges from 500 to 650 mA, and metal artifact correction was in use). The 120 kVp x-rays were split into high and low spectra by 0.6 mm tin and 0.05 mm gold filters, yielding high-energy and low-energy scans for reconstruction, respectively. Composed images were also reconstructed from raw projection dataset by disregarding spectral differences. All the above images were reconstructed by Siemens Syngo CT VA48A with spacing $0.98 \times 0.98 \times 1.5 \text{ mm}^3$ and 512×512 pixels of each slice.

The ground truth RSP map of each patient was created by manually classifying the composed DECT images into different materials and assigning corresponding RSPs. In this study, we differentiated 7 types of material as listed in **Table 1**, each of which has RSP calculated using the chemical compositions published in Su et al [9] and ICRP Publication 23 [36]. Equations and details of calculation were done as previously reported [10]. Note that although the learning target is bulk-assigned RSP maps, the predicted RSP maps are still patient-specific with continuous values.

The cohort of 20 patients was used to evaluate our method using leave-one-out cross-validation. For each test patient, the model was trained by the remaining 19 patients. The model is initialized and retrained for next test patient by training another group of 19 patients. The training datasets and testing datasets were separated and independent during each study. For our training, we used data augmentation and 3D patch-based method to increase training data variation. Flipping, rotation, scaling, and rigid warping were used to enlarge the data size by 72 times. Patch size was set to $96 \times 96 \times 32$.

For comparison, a physics-based dual-energy stoichiometric method was implemented to generate RSP maps as well. We used the Gammex RMI 467 electron density phantom (Sun Nuclear, Melbourne, FL) with the chemical compositions of inserted materials specified by the manufacturer. The RSP value of each rod was calculated from the known chemical compositions using the Bethe equations as reported [10]. The rods were randomly placed throughout the phantom for 5 different scans so an average HU value, relatively independent of positioning, could be obtained. The same CT scanning

protocol as used for patient scans was used during the phantom scan. The HU was measured on each rod at both high- and low-energy images and calibrated with the calculated RSP values using equations reported [9]. The calibration was then applied on patient DECT images to generate physics-based RSP maps.

To investigate the performance of the proposed method at different noise levels, we simulated additional noise on the original DECT image datasets. On each pixel, random noise was added with a probability distribution of Gaussian with mean = 0 and standard deviation = 0.5%, 1%, 2%, and 5% of the value of that pixel + 1000 HU, respectively.

To further demonstrate the feasibility of dose calculation using the predicted RSP by the proposed method in proton treatment planning, we compared the difference between dose maps calculated on the ground truth RSP and the predicted RSP using the same plan parameters. Of the 20 patients, 19 had clinical proton treatment plans using pencil beam scanning and optimized with multi-field optimization technique. All plans had 4 to 5 beams and were robustly optimized with 3-mm setup uncertainty in all directions, which was covered by the 3.5% range uncertainty. The average prescription dose to clinical target volumes (CTVs) was 60.2 Gy among all patients. The structures and original treatment plans were duplicated onto the RSP maps of ground truth or prediction for dose calculation using the same algorithm (proton convolution superposition) in Eclipse 13.7 (Varian Medical Systems, Palo Alto, California). For each plan, we visually checked the similarity of the dose distributions calculated on ground-truth RSP and predicted RSP. Quantitatively, clinically relevant dose volume histogram (DVH) metrics were extracted for comparison of dose to CTVs and relevant OARs. For plans with multiple CTVs, the DVH of each CTV was counted separately.

In addition to the patient study, we also included a phantom study to evaluate our method in a controlled condition. We scanned an anthropomorphic phantom (CIRS Adult Male 701) with same scanning protocol for patients with head and neck cancer. Half of the slices (about 150 slices) were trained, and the rest were tested in an alternating pattern. The training slices and testing slices were alternating. The same data augmentation and 3D patch-based method was used to increase training data variation. The phantom mainly contained 3 materials, that is, soft tissue, lung, and bone, and the ground-truth RSP of each material was calculated based on the material information found in the manual.

Image-Quality Metrics

In this study, we evaluated the accuracy of generated RSP maps by comparing it with ground-truth RSP maps. We quantitatively characterized the overall accuracy by normalized mean square error (NMSE) within the patient body. The NMSE measures the average error among all pixels, which can be described as

$$(5) \quad \text{NMSE} = \frac{\| \text{RSP} - \text{RSP}_0 \|_2^2}{\| \text{RSP} \|_2^2},$$

where RSP and RSP₀ are generated and ground-truth RSP maps, respectively, and $\| \cdot \|_2$ is the L2 norm. To quantify the accuracy of each material, we calculated the mean error (ME) and NMSE on each material VOI except air. ME quantifies the error of averaged RSP in certain materials, which is defined as,

$$(6) \quad \text{ME} = \frac{\sum_{i \in \text{VOI}_j} \text{RSP}(i) - \text{RSP}_0(i)}{\sum_{i \in \text{VOI}_j} \text{RSP}_0(i)},$$

where i indicates the index of the i th pixel in the VOIs of the j th material.

Results

Figure 3 shows the anthropomorphic phantom results of our proposed method as well as ground truth. Most RSP differences were in lung and at the boundary of 2 materials. In such case with simple geometry and piece-wise constant RSP distribution, the proposed method achieved NMSE of 0.60% within the patient body, and ME/NMSE of -2.20%/4.08%, 0.31%/0.55%, and -0.10%/0.43% for lung, soft tissue, and bone, respectively.

In **Figure 4**, the quality of RSP maps is shown using a side-by-side comparison with ground-truth and physics-based results from 1 patient without additional simulated noise. By the proposed method, the RSP maps maintained comparable resolution, contrast, and most of the details as the ground truth. Errors were observed due to misclassification between adipose and muscle, as well as between spongiosa bone and cortical bone, in addition to boundaries between different materials (column



Figure 3. Axial views of phantom on (a) high-energy computed tomography images, (b) low-energy tomography images, and (c) its ground-truth relative stopping power (RSP) maps, and (d) RSP maps produced by the proposed method, respectively. (e) The difference maps of (d)–(c). The window level of RSP maps (c and d) is [0, 2].

g). Compared with the results from the proposed method, physics-based results showed a larger error in lungs (e3). It also showed degraded quality caused by noise and artifacts. For example, a large discrepancy in (e3) can be seen around the corresponding region of streaking artifacts in (a3) and (b3).

Similarly, an exemplary result with additional 5% simulated noise is also shown in **Figure 5**. RSP maps generated by the proposed method and their corresponding profiles (**Figure 6**, C2) remained close to the results without additional noise, while those generated by the physics method were highly affected by the noise and demonstrated severe inaccuracies (column b). The averages of measured NMSE within the patient body at different noise levels among all 20 patients are summarized in **Table 2**. The proposed method maintained a comparable accuracy when noise increased from 0% to 5%, while the physics-based method showed severe degradation in performance.

The averages of measured NMSE and ME within different VOIs at different noise levels among all 20 patients are summarized in **Table 3**. With several exceptions of lower but comparable NMSEs in adipose and muscle at the lowest noise level, overall, our method outperformed the physics-based method in NMSE and ME. When the noise level increased, physics-based results were more affected in accuracy, while the accuracy of the proposed method was successfully maintained. Thus, the advantage of the proposed method over the physics-based method was more prominent at higher noise levels, which indicates greater robustness to noise.

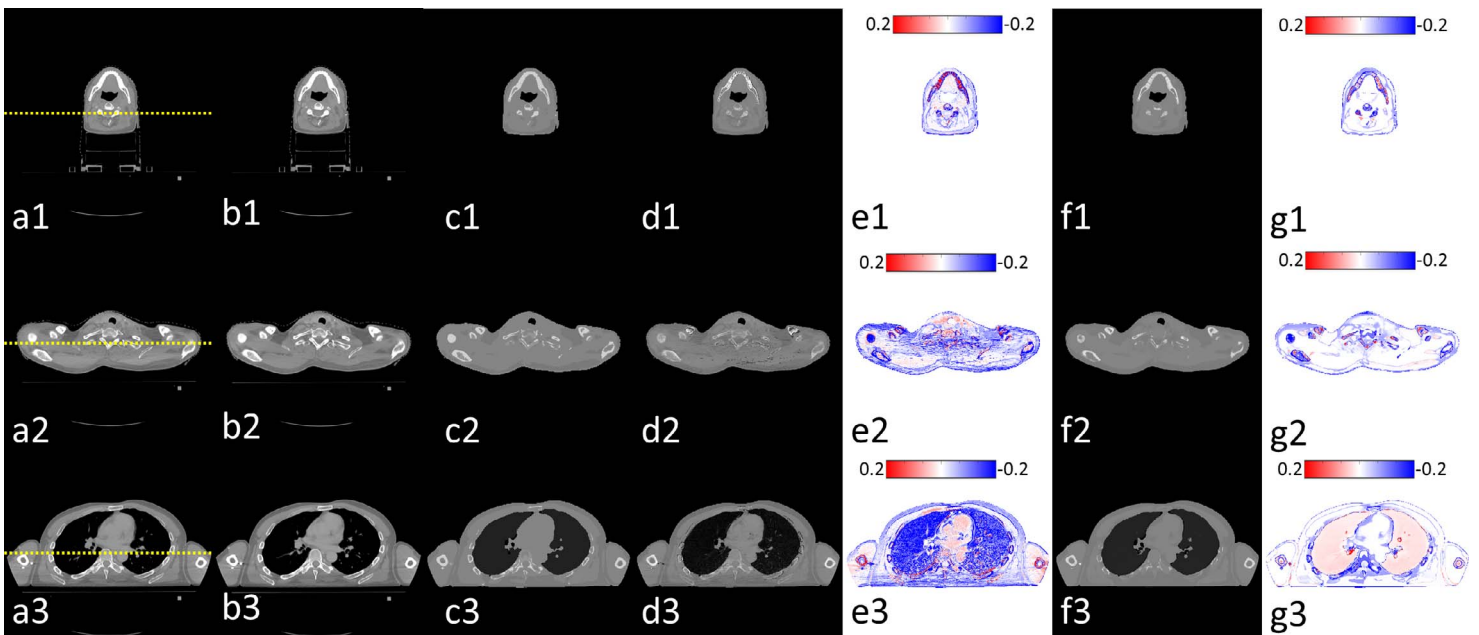


Figure 4. Axial views of 1 patient without simulated noise. Rows (1), (2), and (3) are 3 different slices. Columns (a), (b), (c), (d), and (f) are the high-energy computed tomography images, low-energy computed tomography images, ground-truth relative stopping power (RSP) maps, physics-based RSP map, and maps produced by the proposed method, respectively. Columns (e) and (g) are the difference maps of (d)–(c) and (f)–(c), respectively. The yellow dotted lines on (a) indicate the positions of profiles displayed in Figure 6. The window level of RSP maps (a, b, c, d, and f) is [0, 2].

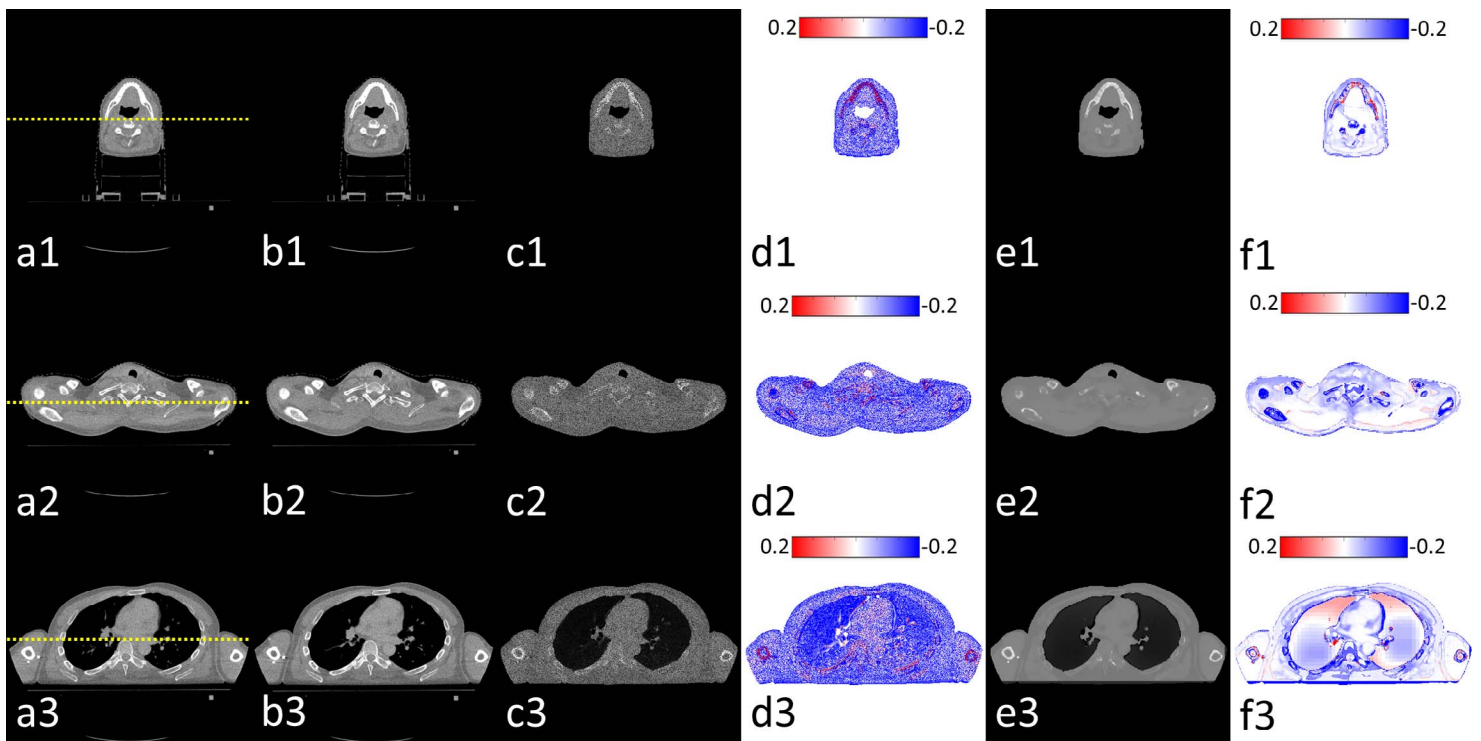


Figure 5. Axial views of 1 patient with 5% additional simulated noise. Rows (1), (2), and (3) are 3 different slices. Columns (a), (b), (c), and (e) are the high-energy computed tomography images, low-energy computed tomography images, Relative stopping power maps by the physics method and by the proposed method, respectively. Columns (d) and (f) are the difference maps of (c) and (e) from ground truth in **Figure 4** column (c), respectively. The yellow dotted lines on (a) indicate the positions of profiles displayed in **Figure 6**. The window level of relative stopping power maps (a, b, c, and e) is [0, 2].

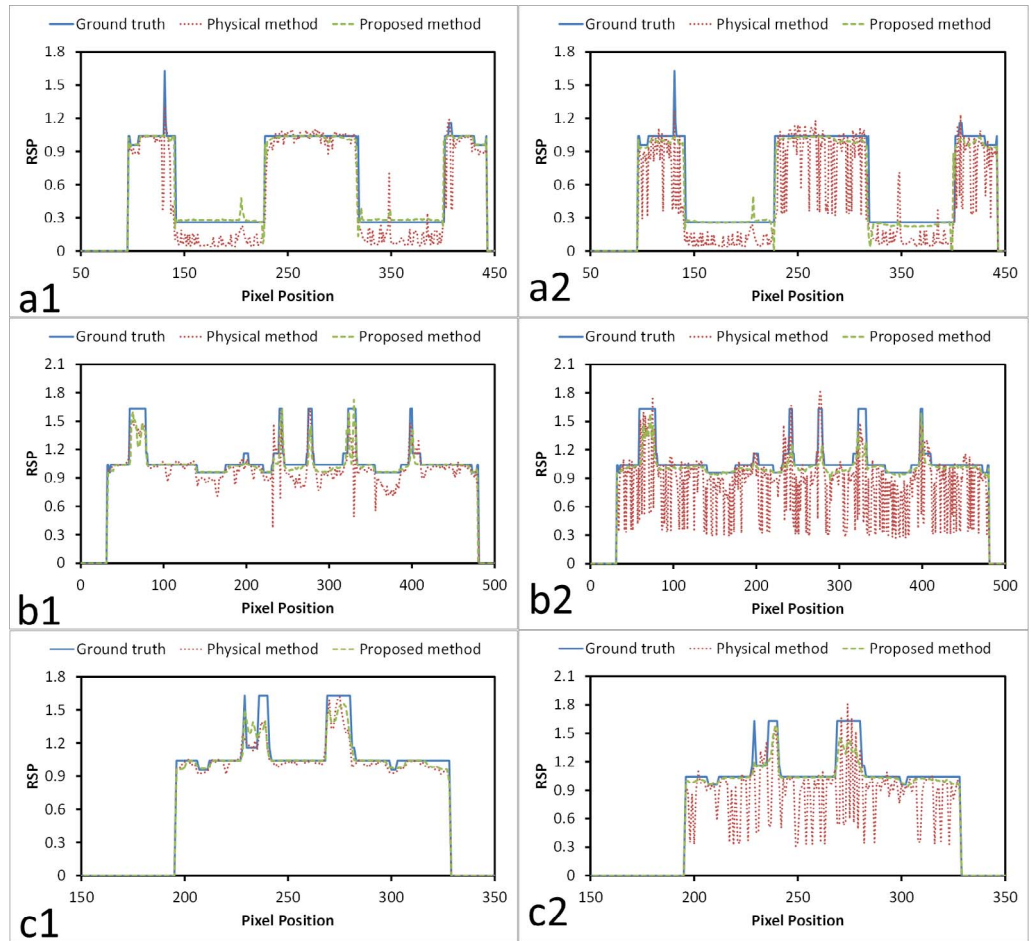
Figure 7 compares the calculated proton dose maps at selected axial, sagittal, and coronal planes of a patient as an example in the case of no additional noise simulated. The dose maps calculated on the ground truth qualitatively appeared to be more similar on those of the predicted RSP maps by the proposed method than by the physics-based method. Most dose errors of the proposed method were around the distal end of the beams, while that of the physics-based method happened at all the high dose-gradient regions to a much wider and severe extent. The gamma index maps (3%/3 mm) also showed more failed areas in physics-based results than our results. The gamma passing rates (threshold = 10%) for the physics-based method were 79.67%, 83.03%, and 91.43% for the axial, coronal, and sagittal planes at iso-center, and 97.81%, 97.75%, and 98.92% for the proposed method at the corresponding planes.

The differences in DVH metrics of clinically relevant OARs between ground truth and predicted RSP among all patients are summarized in **Table 4**. For the proposed method, the mean differences were less than 0.2 Gy for CTVs with no statistical significance. A 2-tailed paired *t*-test was performed to compare the mean values of each DVH metric between ground truth and predicted RSP maps. It was shown that DVH metrics had no significant difference in all OARs except esophagus and brainstem with overestimation less than 1.2 Gy. Compared with the results by the physics-based method, the results by our method had less variation in most of the DVH metrics.

Discussion

In this study, we proposed a novel machine-learning based method to predict RSP maps from DECT for proton radiation therapy. As an alternative to the current physics-based method, our proposed method aims to provide accurate RSP values with more resistance to noise. We evaluated the accuracy of predicted RSP maps using our method in the context of patients with head-and-neck cancer. The RSP maps showed an average NMSE of 2.83% across the whole body volume and average ME less than 3% in all VOIs. With additional simulated noise added in DECT datasets, the proposed method still maintained comparable performance, while the physics-based method suffered from degraded inaccuracy with increased noise level. Based on the statistical analysis of comparative DVH metrics of dose maps calculated on ground truth

Figure 6. Comparison of relative stopping power (RSP) map profiles. The positions of profiles (a), (b), and (c) are indicated by yellow dotted lines in **Figure 3** or **Figure 4** (a1), (a2), and (a3), respectively. Rows (1) and (2) are the profiles of results in **Figure 3** (without simulated noise) and 4 (5% additional simulated noise), respectively.



and predicted RSP maps among 19 pencil beam scanning proton treatment plans, we showed that the average differences in DVH metrics for CTVs were less than 0.2 Gy for D95% and Dmax with no statistical significance. Maximum difference in DVH metrics of OARs was around 1 Gy on average. These results strongly indicated the high accuracy of RSP maps predicted by our machine-learning-based method and showed potential feasibility for proton treatment planning and dose calculation.

In this study, we proposed a method to generate RSP maps by learning from DECT image datasets and the corresponding ground-truth RSP. Note that the patients' true RSP values were unavailable. The ground truth of this study was then assumed by assigning calculated RSP values to manually segmented regions of materials on patient DECT images, where the calculation was based on the known chemical composition. Such a material assignment method is used in current proton and photon studies when the required material information cannot be readily or accurately derived from CT simulation images [9, 38, 39]. Although the learning target was bulk-assigned RSP maps, the predicted RSP maps were still patient-specific with continuous values, preserved fine image textures and contrasts. Potential error may have been caused by the inaccuracy in segmentation and inter- and inpatient variation in composition, which could be a limitation in the implementation of this study. Another limitation is that the training and testing images in the anthropomorphic phantom study, although from different slices, were still very similar in terms of geometry and material.

Table 2. Mean ± standard deviations of normalized mean square error within patient body at different noise levels among all 20 patients.

	Additional noise level				
	0%	0.5%	1%	2%	5%
Physics	-3.00 ± 0.51	-3.06 ± 0.51	-3.25 ± 0.51	-4.22 ± 0.54	-13.18 ± 0.47
Proposed	-2.83 ± 0.52	-2.97 ± 0.65	-2.99 ± 0.65	-3.03 ± 0.52	-3.35 ± 1.00

Table 3. Mean \pm standard deviation of mean error and normalized mean square error within different VOIs at different noise level among all 20 patients.

	Additional noise levels				
	0%	0.50%	1%	2%	5%
Mean error (%)					
Lungs					
Physics	-47.31 \pm 14.10	-47.33 \pm 14.09	-47.40 \pm 14.06	-47.67 \pm 13.98	-49.18 \pm 13.47
Proposed	0.18 \pm 3.98	1.81 \pm 7.13	1.82 \pm 7.10	1.77 \pm 7.11	1.68 \pm 9.35
Adipose					
Physics	-5.46 \pm 1.87	-5.62 \pm 1.87	-6.08 \pm 1.89	-8.05 \pm 2.07	-22.10 \pm 1.97
Proposed	0.15 \pm 1.34	0.74 \pm 3.94	0.77 \pm 3.93	0.87 \pm 4.19	-0.01 \pm 4.02
Muscle					
Physics	-3.73 \pm 1.62	-3.90 \pm 1.62	-4.38 \pm 1.63	-6.30 \pm 1.73	-20.41 \pm 1.41
Proposed	-1.24 \pm 1.00	-0.74 \pm 3.84	-0.76 \pm 3.83	-0.80 \pm 4.00	-0.76 \pm 3.90
Brain					
Physics	-3.11 \pm 0.19	-3.32 \pm 0.19	-3.89 \pm 0.19	-5.71 \pm 0.19	-20.53 \pm 0.18
Proposed	2.91 \pm 1.33	4.16 \pm 4.41	4.18 \pm 4.40	4.22 \pm 4.38	4.76 \pm 4.49
Spongiosa bone					
Physics	-6.45 \pm 1.69	-6.55 \pm 1.70	-6.85 \pm 1.70	-8.45 \pm 1.66	-19.64 \pm 1.20
Proposed	1.03 \pm 2.12	1.51 \pm 4.37	1.50 \pm 4.36	1.45 \pm 4.44	2.02 \pm 4.75
Cortical bone					
Physics	-13.66 \pm 2.37	-13.92 \pm 2.32	-14.73 \pm 2.20	-17.14 \pm 1.97	-24.54 \pm 1.55
Proposed	-1.71 \pm 1.96	-3.53 \pm 5.24	-3.55 \pm 5.24	-3.63 \pm 5.28	-2.61 \pm 5.70
Normalized mean square error (%)					
Lung					
Physics	39.51 \pm 7.28	39.53 \pm 7.28	39.57 \pm 7.28	39.68 \pm 7.28	40.37 \pm 7.33
Proposed	5.04 \pm 2.54	9.11 \pm 4.19	9.11 \pm 4.18	9.07 \pm 4.21	10.39 \pm 4.57
Adipose					
Physics	1.73 \pm 0.48	1.78 \pm 0.49	1.93 \pm 0.51	2.87 \pm 0.62	12.46 \pm 0.64
Proposed	2.45 \pm 0.97	2.75 \pm 1.07	2.77 \pm 1.06	2.80 \pm 0.92	3.10 \pm 1.23
Muscle					
Physics	2.32 \pm 0.71	2.36 \pm 0.73	2.49 \pm 0.76	2.87 \pm 0.62	12.46 \pm 0.77
Proposed	3.01 \pm 0.62	2.84 \pm 0.66	2.88 \pm 0.68	2.79 \pm 0.92	3.18 \pm 1.05
Brain					
Physics	0.15 \pm 0.01	0.17 \pm 0.02	0.26 \pm 0.01	0.75 \pm 0.01	10.71 \pm 0.01
Proposed	0.12 \pm 0.01	0.37 \pm 0.51	0.38 \pm 0.51	0.38 \pm 0.51	0.44 \pm 0.60
Spongiosa bone					
Physics	3.87 \pm 1.15	3.97 \pm 1.15	4.27 \pm 1.14	5.58 \pm 1.01	13.46 \pm 0.67
Proposed	2.80 \pm 0.56	2.61 \pm 0.87	2.60 \pm 0.86	2.59 \pm 0.77	3.07 \pm 1.37
Cortical bone					
Physics	7.94 \pm 1.14	8.13 \pm 1.12	8.74 \pm 1.07	10.70 \pm 0.92	17.14 \pm 0.56
Proposed	4.81 \pm 0.45	5.58 \pm 0.85	5.89 \pm 0.86	5.62 \pm 0.77	6.30 \pm 1.29

Testing on a different anthropomorphic phantom with different materials and reasonable geometry variation would be helpful in investigating the fundamental correlation between learning algorithm and physical artifacts in DECT images under a well-controlled condition.

However, it is worth noting that this study did not aim to evaluate the absolute error of the predicted RSP from patients' true RSP values which were unavailable. Instead, we demonstrated the performance of the proposed method by evaluating its discrepancy of the prediction with its training dataset. Such performance should remain at a similar level when the training dataset is selected differently. With the prediction performance maintained, a better knowledge of RSP in the training stage, such as more and finer types of materials segmented and assigned on a patient, differentiation of the same type of tissue among diverse demographics [40], or estimation from animal tissue models [41, 42], would directly lead the predicted results to be closer to physical reality.

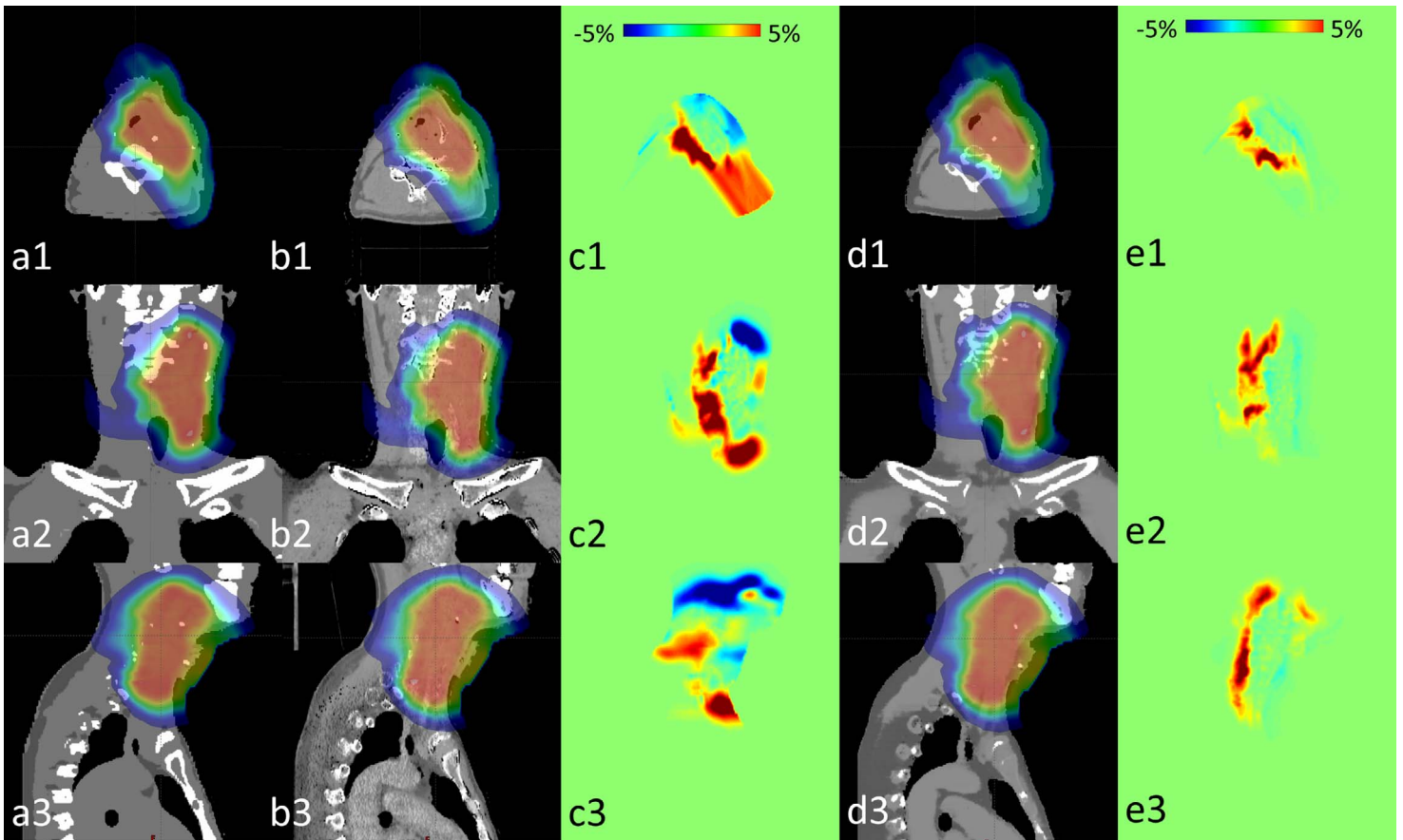


Figure 7. Dose distribution calculated on (a) ground-truth and predicted relative stopping power by the (b) physics-based method and (d) proposed method from dual-energy computed tomography datasets of 1 patient without additional simulated noise presented in 3 orthogonal views (1), (2), and (3). The dose difference maps of (a) versus (b) and (a) versus (d) are shown in (c) and (e), respectively.

Machine-learning-based methods are relatively new for DECT RSP map prediction. Only a few existing studies have relied on patient data, rather than phantom data. The RSP prediction accuracy by our method was competitive to others. Su et al [9] investigated the performance of historical centroid, random forest, and artificial neural networks in solving this problem. Models were trained on tissue substitutes and tested on patient data. Results showed that MEs were around 5% for all VOIs in the abdominal region compared with calculated RSP values as ground truth. Our study moved a step further in integrating a

Table 4. Mean \pm standard deviation of differences of clinically relevant dose volume histogram between dose maps calculated on ground truth and predicted relative stopping power maps.

	Clinical target volume		Brainstem	Cochlea	Esophagus	Mandible	Oral cavity	Parotid
	D _{95%}	D _{max}	D _{max}	D _{max}	D _{mean}	D _{max}	D _{mean}	D _{mean}
Physics								
Mean \pm standard deviation (Gy)	0.074 \pm 0.681	-0.662 \pm 1.630	-0.482 \pm 1.379	0.016 \pm 0.498	-0.751 \pm 0.933	-0.046 \pm 1.305	-0.559 \pm 1.075	-0.165 \pm 1.439
P value versus ground truth	.535	.026	.156	.884	.013	.893	.116	.537
Proposed								
Mean \pm standard deviation (Gy)	0.103 \pm 0.357	0.175 \pm 0.792	0.741 \pm 1.291	-0.023 \pm 0.567	1.173 \pm 1.152	0.015 \pm 0.329	0.116 \pm 0.208	-0.009 \pm 0.064
P value versus ground truth	.107	.218	.037	.850	.005	.859	.095	.437

deep attention strategy and cycle-GAN into the mapping from DECT to RSP, directly training on patient datasets, and demonstrating the dosimetric feasibility.

In this study, we used a twin-beam scanner for DECT acquisition. Note that the proposed method does not specify the scan scheme; thus, it is applicable to other DECT modalities. We presented it in the context of TBCT in order to demonstrate the noise robustness of the proposed method. TBCT has inferior energy separation compared with other DECT modalities, which leads to a higher sensitivity to artifacts and noise on DECT image datasets [22]. It potentially explains why the results by the physics-based method demonstrated larger error and higher noise level than previously reported [15], where DECT images were acquired by 2 sequential scans at 2 different energy levels, which have a larger separation between the 2 energy spectra than the twin-beam scan scheme used in this study. However, TBCT has unique advantages over other DECT modalities for radiation therapy simulation in good temporal coherence, full field of view, and low hardware complexity and cost. Thus, our method overcomes the drawback of TBCT in RSP map generation by its insensitivity to noise, thus facilitating the clinical use of TBCT in radiation therapy workflow. Moreover, the recently developed multispectral CT, which allows concurrent identification of multiple materials with increased accuracy [43], is also compatible with the proposed method by adding more input channels for CT images at multiple energy levels.

We investigated the performance of the proposed method at different noise levels and compared with the physics-based method. The DECT images without any simulated noise represents the most likely noise level in clinical practice. With simulated noise, the DECT images may be much noisier than usual, while it may still happen when the patient is large or a low-mAs scanning protocol is preferred. In such cases, the advantage in noise robustness of the proposed method is desired. In addition to the noise robustness investigated in this study, deep learning-based methods may have an advantages over physics-based methods in other aspects. For example, a major struggle in physics-based methods is the systematic uncertainty in reconstructed images, which includes, but is not limited to, beam-hardening artifact, scatter artifact, CT number variations on reconstruction algorithms, and patient size [44–47]. Deep learning-based methods, on the other hand, are expected to be less affected by these uncertainties, as long as the training datasets are representative and the corresponding training target (RSP maps) reflects the true RSP distributions. It can be attributed to the data-driven approach used in deep learning that focuses on images features, unlike the sole dependence on pixel values in the physics-based method. Given these potential merits, deep learning-based methods may overcome the current obstacles in physics-based proton RSP estimation, while the specific improvement needs further investigations.

In the present study, we found that the dose difference in columns (c) and (f) of **Figure 7** mostly happened at the distal end of the beam. It is consistent with current studies about range uncertainties of proton beams caused by the error in RSP maps [39]. The overestimation in dose could have resulted from the underestimation of muscle RSP with less compensation from the overestimation of adipose RSP. Future studies could include investigations in determining the range deviations to the ground truth from the systematic prediction error on each material.

Deep learning-based methods usually involve large computational costs to train a model. We implemented the proposed algorithm with Python 3.7 (Python Software Foundation, Wilmington, DE) and TensorFlow (Company, City, State) as in-house software on a NVIDIA Tesla V100 GPU (Santa Clara, CA) with 32 GB of memory. The Adam gradient optimizer with a learning rate of $2e-4$ was used for optimization. In the present study, the training stage required ~ 30 GB and ~ 16 hours for the training datasets of 19 patients and ~ 2 minutes for each patient in the testing stage.

We limited our study to the head-and-neck region. Patients with head-and-neck cancer have high anatomical complexity and variability between patients. The tumor shape, size, and location can vary greatly for different patients, and it is common to see the tumor changing the exterior body shape, which is challenging for the learning-based method. A comprehensive evaluation with a larger population of patients with diverse anatomical abnormalities should be included in future studies during the model training. Different testing and training datasets from other centers would also be necessary to validate the clinical utility of our method. Moreover, the proposed method can be applied to other treatment sites of clinical importance for proton therapy, which would be of great interest for expanding this work to the clinic.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflicts of Interest: The authors have no relevant conflicts of interest to disclose.

Funding: This research was supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718, and Emory Winship Pilot Grant.

Ethical Approval: All patient data have been collected under internal review board approved protocol.

References

1. Sheets NC, Goldin GH, Meyer A-M, Wu Y, Chang Y, Stürmer T, Holmes JA, Reeve BB, Godley PA, Carpenter WR, Chen RC. Intensity-modulated radiation therapy, proton therapy, or conformal radiation therapy and morbidity and disease control in localized prostate cancer. *JAMA*. 2012;307:1611–20.
2. MacDonald SM, Safai S, Trofimov A, Wolfgang J, Fullerton B, Yeap BY, Bortfeld T, Tarbell NJ, Yock T. Proton radiotherapy for childhood ependymoma: initial clinical outcomes and dose comparisons. *Int J Radiat Oncol Biol Phys*. 2008;71:979–86.
3. Yock T, Schneider R, Friedmann A, Adams J, Fullerton B, Tarbell N. Proton radiotherapy for orbital rhabdomyosarcoma: clinical outcome and a dosimetric comparison with photons. *Int J Radiat Oncol Biol Phys*. 2005;63:1161–8.
4. Zietman AL, Bae K, Slater JD, Shipley WU, Efstathiou JA, Coen JJ, Bush DA, Lunt M, Spiegel DY, Skowronski R, Jabola BR, Rossi CJ. Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/American College of Radiology 95-09. *J Clin Oncol*. 2010;28:1106–11.
5. Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, Patel P, Jani AB, McDonald M, Curran WJ, Liu T, Zhou J, Yang X. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. *Phys Med Biol*. 2019. doi: 10.1088/1361-6560/ab41af.
6. Liu Y, Lei Y, Wang Y, Wang T, Ren L, Lin L, McDonald M, Curran WJ, Liu T, Zhou J, Yang X. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. *Phys Med Biol*. 2019;64:145015.
7. Shafai-Erfani G, Lei Y, Liu Y, Wang Y, Wang T, Zhong J, Liu T, McDonald M, Curran WJ, Zhou J. MRI-based proton treatment planning for base of skull tumors. *Int J Part Ther*. 2019;6:12–25.
8. Harms J, Lei Y, Wang T, McDonald M, Ghavidel B, Stokes W, Curran WJ, Zhou J, Liu T, Yang X. Cone-beam CT-derived relative stopping power map generation via deep learning for proton radiotherapy. *Med Phys*. 2020;47:4416–4427.
9. Su KH, Kuo JW, Jordan DW, Van Hedent S, Klahr P, Wei Z, Al Helo R, Liang F, Qian P, Pereira GC, Rassouli N, Gilkeson RC, Traughber BJ, Cheng CW, Muzic RF. Machine learning-based dual-energy CT parametric mapping. *Phys Med Biol*. 2018;63:125001.
10. Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys Med Biol*. 1996;41:111–24.
11. Wang T, Zhu L. Dual energy CT with one full scan and a second sparse-view scan using structure preserving iterative reconstruction (SPIR). *Phys Med Biol*. 2016;61:6684.
12. Harms J, Wang T, Petrongolo M, Niu T, Zhu L. Noise suppression for dual-energy CT via penalized weighted least-square optimization with similarity-based regularization. *Med Phys*. 2016;43:2676–86.
13. Harms J, Wang T, Petrongolo M, Zhu L. Noise suppression for energy-resolved CT using similarity-based non-local filtration. Paper presented at: SPIE Medical Imaging 2016, February 27–March 3, San Diego, CA.
14. van Elmpt W, Landry G, Das M, Verhaegen F. Dual energy CT in radiotherapy: current applications and future outlook. *Radiother Oncol*. 2016;119:137–44.
15. Zhu J, Penfold SN. Dosimetric comparison of stopping power calibration with dual-energy CT and single-energy CT in proton therapy treatment planning. *Med Phys*. 2016;43:2845–54.
16. Yang M, Virshup G, Clayton J, Zhu XR, Mohan R, Dong L. Theoretical variance analysis of single- and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues. *Phys Med Biol*. 2010;55:1343–62.jpg
17. McCollough CH, Leng S, Yu L, Fletcher JG. Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology*. 2015;276:637–53.
18. Johnson TRC. Dual-energy CT: General principles. *AJR Am J Roentgenol*. 2012;199(suppl):S3–S8.
19. Engel KJ, Herrmann C, Zeitler G. X-ray scattering in single- and dual-source CT. *Med Phys*. 2008;35:318–32.
20. Petersilka M, Stierstorfer K, Bruder H, Flohr T. Strategies for scatter correction in dual source CT. *Med Phys*. 2010;37:5971–92.
21. Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 1. *Neuroimaging Clin N Am*. 2017;27:371–84.

22. Wang T, Ghavidel BB, Beitler JJ, Tang X, Lei Y, Curran WJ, Liu T, Yang X. Optimal virtual monoenergetic image in “TwinBeam” dual-energy CT for organs-at-risk delineation based on contrast-noise-ratio in head-and-neck radiotherapy. *J Appl Clin Med Phys*. 2019;20:121–8.
23. Niu T, Dong X, Petrongolo M, Zhu L. Iterative image-domain decomposition for dual-energy CT. *Med Phys*. 2014;41:041901. DOI:https://doi.org/10.1118/1.4866386
24. Petrongolo M, Zhu L. Noise suppression for dual-energy CT through entropy minimization. *IEEE Trans Med Imaging*. 2015;34:2286–97.
25. Petrongolo M, Dong X, Zhu L. A general framework of noise suppression in material decomposition for dual-energy CT. *Med Phys*. 2015;42:4848–62.
26. Aouadi S, Vasic A, Paloor S, Hammoud RW, Torfeh T, Petric P, Al-Hammadi N. Sparse patch-based method applied to mri-only radiotherapy planning. *Phys Med*. 2016;32:309.
27. Wang T, Manohar N, Lei Y, Dhabaan A, Shu H-K, Liu T, Curran WJ, Yang X. MRI-based treatment planning for brain stereotactic radiosurgery: dosimetric validation of a learning-based pseudo-CT generation method. *Med Dosim*. 2019;44:199–204.
28. Wang T, Lei Y, Manohar N, Tian S, Jani AB, Shu H-K, Higgins K, Dhabaan A, Patel P, Tang X, Liu T, Curran WJ, Yang X. Dosimetric study on learning-based cone-beam CT correction in adaptive radiation therapy. *Med Dosim*. 2019;44:e71–9.
29. Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, Shen D. Estimating CT image from MRI data using structured random forest and auto-context model. *IEEE Trans Med Imaging*. 2016;35:174–83.
30. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408–19.
31. Dong X, Lei Y, Tian S, Wang T, Patel P, Curran WJ, Jani AB, Liu T, Yang X. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother Oncol*. 2019;141:192–9.
32. Dong X, Wang T, Lei Y, Higgins K, Liu T, Curran WJ, Mao H, Nye JA, Yang X. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. *Phys Med Biol*. 2019;64:215016.
33. Lei Y, Shu H-K, Tian S, Jeong JJ, Liu T, Shim H, Mao H, Wang T, Jani A, Curran W, Yang X. Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning. *J Med Imaging*. 2018;5:034001.
34. Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, Curran WJ, Mao H, Liu T, Yang X. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46:3565–81.
35. Lei Y, Harms J, Wang T, Tian S, Zhou J, Shu H-K, Zhong J, Mao H, Curran WJ, Liu T, Yang X. MRI-based synthetic CT generation using semantic random forest with iterative refinement. *Phys Med Biol*. 2019;64:085001.
36. Snyder W, Cook M, Nasset E, Karhausen L, Howells G, Tipton I. *ICRP Publication 23: Report of the Task Group on Reference Man*. Elmsford, NY: International. Commission on Radiological Protection; 1975.
37. Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, Curran WJ, Liu T, Yang X. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. 2019;46:3998–4009.
38. Beaulieu L, Carlsson Tedgren Å, Carrier J-F, Davis SD, Mourtada F, Rivard MJ, Thomson RM, Verhaegen F, Wareing TA, Williamson JF. Report of the Task Group 186 on model-based dose calculation methods in brachytherapy beyond the TG-43 formalism: current status and recommendations for clinical implementation. *Med Phys*. 2012;39:6208–36.
39. Wohlfahrt P, Mohler C, Richter C, Greilich S. Evaluation of stopping-power prediction by dual- and single-energy computed tomography in an anthropomorphic ground-truth phantom. *Int J Radiat Oncol Biol Phys*. 2018;100:244–53.
40. White DR, Booz J, Griffith RV, Spokas JJ, Wilson IJ. Report 44. *J ICRU*. 2016;os23(1).
41. Xie Y, Ainsley C, Yin L, Zou W, McDonough J, Solberg TD, Lin A, Teo BK. Ex vivo validation of a stoichiometric dual energy CT proton stopping power ratio calibration. *Phys Med Biol*. 2018;63:055016.
42. Schaffner B, Pedroni E. The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power. *Phys Med Biol*. 1998;43:1579–92.
43. Fornaro J, Leschka S, Hibbeln D, Butler A, Anderson N, Pache G, Scheffel H, Wildermuth S, Alkadhi H, Stolzmann P. Dual- and multi-energy CT: approach to functional imaging. *Insights Imaging*. 2011;2:149–59.

44. Han D, Siebers JV, Williamson JF. A linear, separable two-parameter model for dual energy CT imaging of proton stopping power computation. *Med Phys*. 2016;43:600–12.
45. Zhang S, Han D, Williamson JF, Zhao T, Politte DG, Whiting BR, O'Sullivan JA. Experimental implementation of a joint statistical image reconstruction method for proton stopping power mapping from dual-energy CT data. *Med Phys*. 2019;46:273–85.
46. Zhang S, Han D, Politte DG, Williamson JF, O'Sullivan JA. Impact of joint statistical dual-energy CT reconstruction of proton stopping power images: comparison to image- and sinogram-domain material decomposition approaches. *Med Phys*. 2018;45:2129–42.
47. Li B, Lee HC, Duan X, Shen C, Zhou L, Jia X, Yang M. Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging. *Phys Med Biol*. 2017;62:7056–74.