DblurDoseNet: A deep residual learning network for voxel radionuclide dosimetry compensating for single-photon emission computerized tomography imaging resolution

Zongyu Li[1] | **Jeffrey A. Fessler**[1] | **Justin K. Mikell**[2] | **Scott J. Wilderman**[3] | **Yuni K. Dewaraja**[4]

[1] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, USA

[2] Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA

[3] Department of Nuclear Engineering and Radiologic Sciences, University of Michigan, Ann Arbor, Michigan, USA

[4] Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Zongyu Li, Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Catherine, 2276 Medical Science I/5610, Ann Arbor, MI 48109, USA.
Email: zonyul@umich.edu

Funding information
National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number: R01 EB022075; National Cancer Institute; NIH, Grant/Award Number: R01 CA240706

Abstract

Purpose: Current methods for patient-specific voxel-level dosimetry in radionuclide therapy suffer from a trade-off between accuracy and computational efficiency. Monte Carlo (MC) radiation transport algorithms are considered the gold standard for voxel-level dosimetry but can be computationally expensive, whereas faster dose voxel kernel (DVK) convolution can be suboptimal in the presence of tissue heterogeneities. Furthermore, the accuracies of both these methods are limited by the spatial resolution of the reconstructed emission image. To overcome these limitations, this paper considers a single deep convolutional neural network (CNN) with residual learning (named DblurDoseNet) that learns to produce dose-rate maps while compensating for the limited resolution of SPECT images.

Methods: We trained our CNN using MC-generated dose-rate maps that directly corresponded to the true activity maps in virtual patient phantoms. Residual learning was applied such that our CNN learned only the difference between the true dose-rate map and DVK dose-rate map with density scaling. Our CNN consists of a 3D depth feature extractor followed by a 2D U-Net, where the input was 11 slices (3.3 cm) of a given Lu-177 SPECT/CT image and density map, and the output was the dose-rate map corresponding to the center slice. The CNN was trained with nine virtual patient phantoms and tested on five different phantoms plus 42 SPECT/CT scans of patients who underwent Lu-177 DOTATATE therapy.

Results: When testing on virtual patient phantoms, the lesion/organ mean dose-rate error and the normalized root mean square error (NRMSE) relative to the ground truth of the CNN method was consistently lower than DVK and MC, when applied to SPECT images. Compared to DVK/MC, the average improvement for the CNN in mean dose-rate error was 55%/53% and 66%/56%; and in NRMSE was 18%/17% and 10%/11% for lesion and kidney regions, respectively. Line profiles and dose–volume histograms demonstrated compensation for SPECT resolution effects in the CNN-generated dose-rate maps. The ensemble noise standard deviation, determined from multiple Poisson realizations, was improved by 21%/27% compared to DVK/MC. In patients, potential improvements from CNN dose-rate maps compared to DVK/MC were illustrated qualitatively, due to the absence of ground truth. The trained residual CNN took about 30 s on a single GPU (Tesla V100) to generate a $512 \times 512 \times 130$ dose-rate map for a patient.

Conclusion: The proposed residual CNN, trained using phantoms generated from patient images, has potential for real-time patient-specific dosimetry in

clinical treatment planning due to its demonstrated improvement in accuracy, resolution, noise, and speed over the DVK/MC approaches.

1 | INTRODUCTION

Accurate and computationally efficient methods for patient-specific absorbed dose estimation are essential for clinical implementation of dosimetry-guided treatment planning in radionuclide therapy. For example, current Lu-177 DOTATATE therapy for neuroendocrine tumors uses a fixed activity basis (four cycles of 7.4 GBq), whereas SPECT/CT imaging-based dosimetry after one cycle can be used to individualize the next administration to potentially enhance tumor response while keeping toxicity to critical organs like kidney at an acceptable level.[1] Traditionally, the mean absorbed doses in volumes of interest (VOIs) are the reported quantity. However, voxel-level calculation enables consideration of multiple alternative dose metrics, such as statistics from dose-rate volume histogram (DRVH) analyses that are potentially more relevant to treatment planning. Explicit Monte Carlo (MC) radiation transport using the patient's emission (PET or SPECT) and anatomical images (CT) as input is broadly accepted as the gold standard for voxel-level patient-specific dosimetry. However, it is computationally expensive to generate estimates with low statistical uncertainty. In contrast, faster and simpler dose voxel kernel (DVK) convolution methods[2] can be inaccurate in the presence of heterogeneous tissues, for example, at the liver–lung or bone–marrow interfaces. Moreover, even though MC is theoretically accurate, the dose accuracies of both MC and DVK methods are degraded by reconstruction artifacts and the limited spatial resolution of SPECT and PET images.

Over the past few years, deep learning methods have been broadly used in many fields of medical imaging.[3–7] For example, one of the most popular deep neural networks, the U-Net,[8] achieved state-of-the-art accuracy on the 2012 International Symposium on Biomedical Imaging challenge for segmentation of neuronal structures in electron microscopic stacks. Recently, there is increased interest in studies that apply deep neural networks in nuclear medicine applications.[9–13] However, deep learning applications in radionuclide therapy dosimetry are limited.[14–17] Akhavanallaf et al.[14] employed a modified ResNet[18] that represented voxel S-values kernels[2] to predict the distribution of the deposited energy in whole-body organ-level dosimetry and demonstrated comparable performance to the direct MC approach. Lee et al.[15] implemented a 3D U-Net[8] that used PET and CT-based density image patches to predict 3D voxel-level dose-rate maps. Götz et al.[16] proposed a hybrid method based on a combination of a modified U-Net and an empirical mode decomposition of density maps to enhance the accuracy/reliability of radiation dose estimation. Götz et al.[17] also trained a neural network to predict dose voxel kernels (DVK) for dosimetry in Lu-177 targeted radionuclide therapies. Despite promising results, a limitation of the training approaches in these prior studies[14–17] is that they used MC-generated dose-rate maps derived from each patient's measured SPECT or PET images as the training label, which are degraded by the camera spatial resolution and reconstruction artifacts. Moreover, the concept of residual learning can be adopted in a convolutional neural network (CNN) dosimetry model by exploiting a fast DVK convolution dose-rate map as an initial estimate. Residual learning for image denoising was first proposed to improve the effectiveness and efficiency of a denoising CNN[19] and was further applied to low-dose PET and CT reconstruction.[20,21]

The aim of this study was to develop a deep learning-based absorbed dose-rate estimation method that can overcome the accuracy-efficiency trade-off associated with current voxel dosimetry methods and attempt to learn to reduce the degrading effects of spatial resolution and reconstruction artifacts. Specifically, we used dose-rate estimates directly corresponding to phantom (virtual patient) activity maps as the training label, instead of the patient SPECT-derived dose-rate images (Figure 1). Furthermore, unlike prior studies where a CNN was trained to directly estimate the dose-rate map or S-value kernels, we first used the approximate physics-based fast Fourier transform (FFT) DVK convolution method (with density scaling) to produce initial estimates, and then trained the CNN to learn the subtle residual differences between the initial estimate and the true dose-rate maps. We trained and tested the proposed CNN for SPECT/CT imaging-based dosimetry following Lu-177 DOTATATE therapy of neuroendocrine tumors (NETs).

2 | MATERIALS AND METHODS

2.1 | Virtual patient phantom generation for training and testing

Figure 2 gives an overview of our data generation and training process. To define the true activity
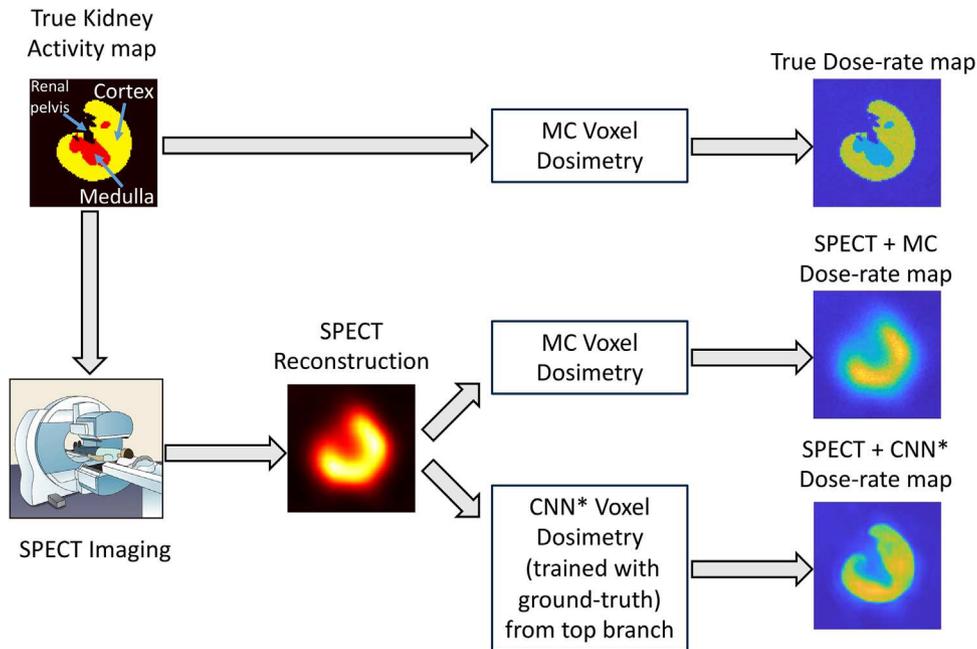
FIGURE 1 Illustration of blurring of dose-rate maps due to the limited resolution of the SPECT-based input activity map and the potential for a learning-based method to outperform MC, the current gold-standard. The CNN* used in this illustration was trained and tested on different XCAT[22] phantoms
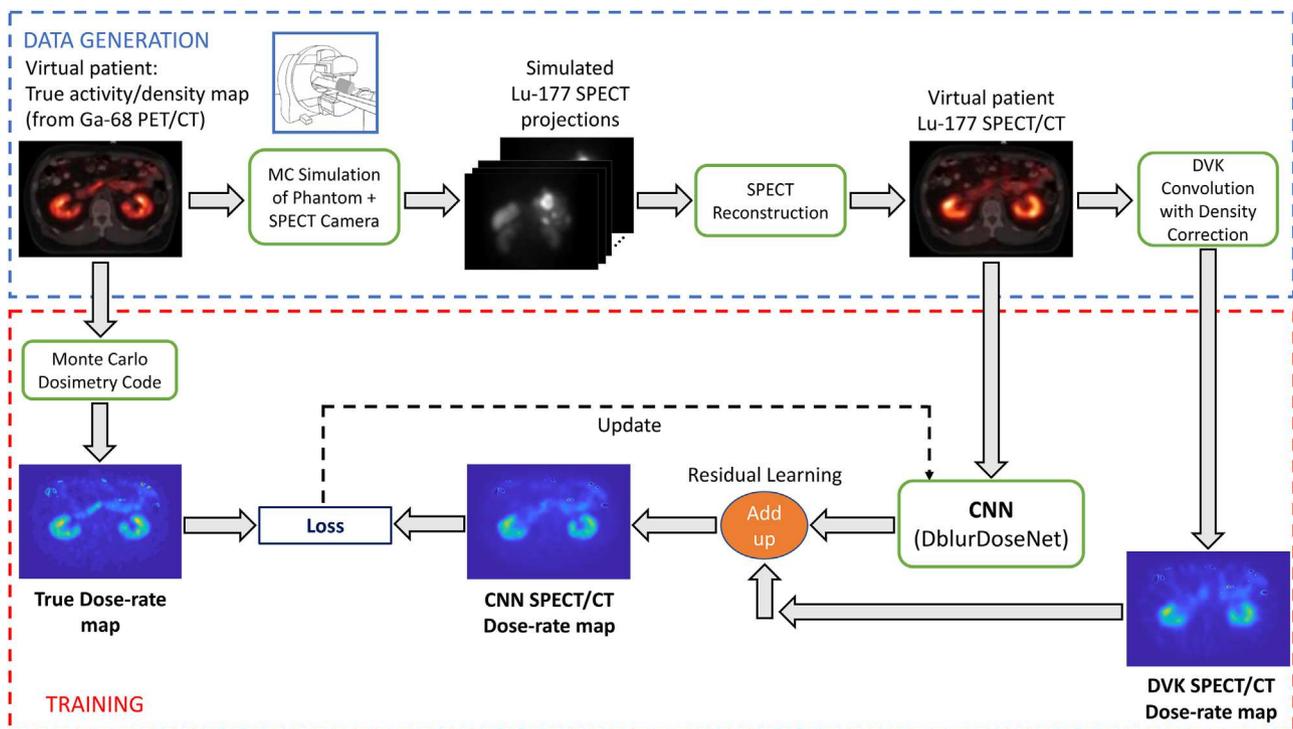
FIGURE 2 Overview of phantom data generation for training/testing and the network training process

maps of virtual patient phantoms, we chose to use PET instead of SPECT-based activity maps because PET offers substantially higher spatial resolution than SPECT as evident in the top branch of Figure 2.

These images were readily available because, prior to Lu-177 DOTATATE, patients underwent diagnostic Ga-68 DOTATATE PET/CT imaging (Siemens Biograph mCT) to determine eligibility for therapy. The Ga-68

DOTATATE distribution in patients is expected to be similar to the Lu-177 DOTATATE distribution and hence our virtual patient phantoms can provide a reasonable approximation to the activity distribution of Lu-177 patients. The PET images (of size $200 \times 200 \times 577$, voxel size is $4.073 \times 4.073 \times 2 mm^3$) were obtained from our Siemens mCT (resolution is 5–6 mm FWHM[23]) and reconstructed using the standard clinic protocol: 3D ordered subset expectation maximization (OSEM) with three iterations, 21 subsets that included resolution recovery, time-of-flight, and a 5 mm (FWHM) Gaussian post-reconstruction filter. We selected 14 such PET images from our clinic database to generate anthropomorphic phantoms for training and testing, with University of Michigan Institutional Review Board (IRB) approval for retrospective analysis. The selected cases covered a diverse range with regard to sex (nine males and five females), age (35–88 years), weight (49–100 kg), and lesions of different sizes and locations (within and outside the liver). The PET/CT images were first extracted into 195 slices with 0.2 cm slice width that covered the SPECT field-of-view (39 cm) with the liver and kidney centered, which is the typical region imaged following Lu-177 DOTATATE. Meanwhile, the corresponding density maps were generated using an experimentally derived CT-to-density calibration curve.

Next, Lu-177 SPECT projections corresponding to each phantom's activity/density maps were generated using the SIMIND MC code[24] (Figure 2, top branch) simulating approximately 2 billion histories per projection. The SIMIND model parameters were based on Lu-177 patient imaging in our clinic (Siemens Intevo with medium energy collimators, a 5/8'' crystal, a 20% photopeak window at 208 keV, and two adjacent 10% scatter windows). Poisson noise was simulated after the 128 projection views were scaled to a count-level in the range of 3–20 million total counts, corresponding to the range in post-therapy imaging. SPECT reconstruction used an in-house 3D OSEM algorithm with CT-based attenuation correction, triple energy window scatter correction and collimator-detector response modeling (four subsets and 16 iterations, $128 \times 128 \times 81$ matrix with voxel size $4.8 \times 4.8 \times 4.8$ mm^3, no Gaussian smoothing). All images were finally registered into CT image space ($512 \times 512 \times 130$ with voxel size $0.98 \times 0.98 \times 3$ mm^3).

Out of 14 virtual patient phantoms, we randomly selected nine for training and five for testing. Out of the training dataset, to assess under/over-fitting, we randomly selected 20% of the total slices to serve as a validation dataset.

2.2 | Patient data

In addition to the above virtual patients, our testing data included a total of 42 scans from 12 patients imaged at up to 4 time points during the first week following cycle 1 of standard Lu-177 DOTATATE (7.4 GBq). The images were acquired as part of an ongoing University of Michigan IRB approved research study, where all subjects signed an informed consent form. SPECT acquisition time was 25 min and all other SPECT imaging reconstruction parameters were as described above for the phantom simulation. The CT was performed in low-dose mode (120 kVp; 15–80 mAs) with free breathing.

2.3 | MC dosimetry and dose voxel kernel convolution

2.3.1 | Monte Carlo

The MC code that we used, called dose planning method (DPM), was originally developed and validated for fast dose-rate estimation in external beam radiotherapy.[25] Previously, we adapted and benchmarked DPM for internal radionuclide therapy applications.[26] Because DPM was optimized specifically for voxel-level electron/photon dose computations with full radiation transport, it is faster than using general-purpose MC codes for voxel-level dose estimation. We used DPM to generate the ground truth training labels (Figure 2) by simulating ~1 billion histories to generate dose-rate maps with reasonably low statistical uncertainty. For example, with 1 billion histories for the phantom results shown in Figures 4 and 5, the average statistical uncertainty across the kidney and lesions was less than 0.1% for both the ground truth MC run and the SPECT+MC run (obtained from the uncertainty images available from DPM).

2.3.2 | DVK convolution with density scaling

To provide DVK dose-rate maps for residual learning, Lu-177 soft tissue (1.04 g/cm^3) voxel kernels were generated using DPM. The beta particle kernel size was $9 \times 9 \times 9$ and the photon kernel size was $99 \times 99 \times 99$ (both with voxel size $0.98 \times 0.98 \times 3 mm^3$). We convolved the SPECT image with the DVKs using FFT-convolution. Since using homogeneous soft tissue kernels neglects tissue inhomogeneities, we applied density scaling that has been shown to be a reasonable correction in past reports.[27] Here, after convolution, each voxel was scaled by 1.04 (g/cm^3) and divided by the local voxel density value (g/cm^3) derived from the CT scan. Because our goal was to generate a reasonably accurate and quick initial estimate for the residual learning process, we did not pursue other more sophisticated approaches[28,29] that account for tissue heterogeneities. To address the very high dose-rate estimate in extra low-density regions, for example, air gaps, we set the
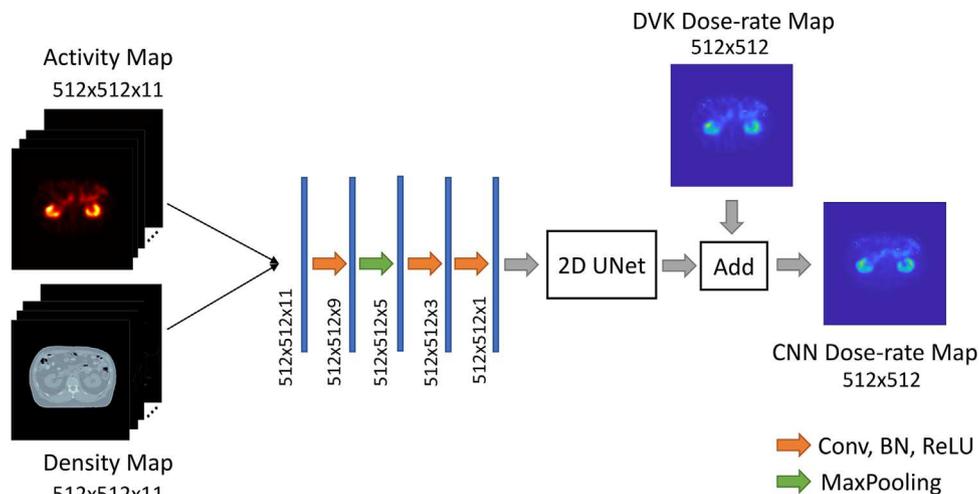
FIGURE 3 The architecture of our DblurDoseNet

dose-rate in regions where the density is less than 0.1 g/cm³ to 0.

2.4 | Network: DblurDoseNet

Our network design considers the decay properties of Lu-177 and the physics of beta/photon interaction in tissue. The mean energy of the emitted electrons in the beta decay of Lu-177 is 134 keV and the maximum energy is 497 keV, and the corresponding continuous slowing down approximation ranges (in water) are 0.3 and 1.8 mm, respectively.[30] The gamma-rays associated with Lu-177 are low in intensity (113 (6.2%) and 208 keV (10.4%)), and hence, the absorbed dose is dominated by the beta component.

The input to the DVK method was an entire 3D SPECT image volume and its output was a 3D dose-rate map. In principle, a CNN could be designed similarly. However, for Lu-177 considering the short beta particle range in tissue and the low photon contribution, we designed a more memory efficient CNN that used a pack of 11 adjacent slices of the SPECT and density images at a time to produce one output dose-rate map corresponding to the middle slice of that pack. The CNN was applied with an 11-slice sliding window to all axial slices using padding that replicated the first and last slices at the top and bottom boundaries, respectively. Thus, the input to CNN was two arrays of size $512 \times 512 \times 11$ (with voxel size $0.98 \times 0.98 \times 3$ mm³) and the output was an array of size 512×512 that corresponded to the dosimetry of the middle slice in the input arrays. During training and testing, these $512 \times 512 \times 11$ packs could be processed sequentially, but GPU devices could accelerate the processing by parallel computation.

As shown in Figure 3, we first concatenated the input activity/density maps along the channel dimension, and then applied three 3D convolutional layers (with kernel size $7 \times 7 \times 5$, $7 \times 7 \times 3$, $7 \times 7 \times 3$, respectively) to extract depth features. Next, we implemented a 2D U-Net that had four down-sample and up-sample layers, where the first convolutional layer in the 2D U-Net had 16 filters. After each down-sample layer, the number of filters at the next convolutional layer was increased by a factor of two until it reached 128. We added the DVK dose-rate map to the 2D U-Net output, as in the common residual learning approach. Finally, we obtained the CNN dose-rate map estimate after setting the dose-rate value in very low-density voxels ($\rho < 0.1$g/cm³) to 0. As discussed in Section 3, the residual CNN produced consistently better dose-rate estimation accuracy than a CNN without residual learning.

The CNN was trained by minimizing the mean square error between the ground-truth and CNN dose-rate maps using a batch size of 32. We used the Adam optimizer[31] with a dynamic learning rate (an initial value 0.001 with ReduceOnPlateau management strategy) and trained our CNN for 200 epochs on two Nvidia Tesla V100 GPUs. The training/validation loss converged visually to 288/410 after 4 h of training (Figure S1). To cover different input count levels, we normalized each SPECT activity map so that all its voxels summed to one, and then inversely scaled the dose-rate map estimate accordingly. To potentially improve convergence during training, we also scaled the normalized SPECT and dose-rate maps with a constant value so that they have a similar range as the density maps.

2.5 | Evaluation in test phantoms

In test phantoms, we used MC with the phantom activity and density maps to calculate the ground truth dose-rate maps for performance evaluation. The estimated

dose-rate maps generated from SPECT/CT with the DVK (with density scaling), MC (with 1 billion histories), and CNN methods were evaluated qualitatively by visual comparison of images, line profiles, and DRVHs with those corresponding to the ground truth. For quantitative evaluations, we used the following metrics.

2.5.1 | Dose-rate error

For each VOI, the dose-rate error is the absolute error across the whole VOI calculated relative to the ground truth. This error was calculated for the mean absorbed dose and DRVH statistics (DR10, DR30, DR70, DR90), corresponding to the minimum dose-rate to 10%, 30%, 70%, and 90% of the VOI, respectively.

2.5.2 | Normalized root mean square error

The normalized root mean square error (NRMSE) is defined as

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n_p} \sum_{j=1}^{n_p} \left(\hat{x}_j - x_j\right)^2}}{\sqrt{\frac{1}{n_p} \sum_{j=1}^{n_p} x_j^2}},$$

where n_p denotes the total number of voxels in the VOI. Subscript j, for example, x_j, denotes the jth voxel in the image. The true and estimated dose-rate image are denoted by x and $\hat{x}$, respectively.

2.5.3 | Ensemble noise

The ensemble noise in spherical VOIs defined in non-tumoral liver or spleen was calculated across 3 ($M = 3$) Poisson noise realizations as:

$$\text{Noise} = \frac{\sqrt{\frac{1}{n_p} \sum_{j \in \text{VOI}} \left(\frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{x}_m[j] - \mu_j\right)^2\right)}}{\frac{1}{n_p} \sum_{j \in \text{VOI}} \mu_j}$$
$$\times 100\%,$$

where $\mu_j = \frac{1}{M} \sum_{m=1}^{M} \hat{x}_m[j]$, n_p is the total number of voxels in the VOI, and $\hat{x}_m[j]$ denotes the jth voxel in the estimated dose-rate image of the mth Poisson noise realization.

The lesion VOIs for these quantitative evaluations were defined manually on CT of SPECT/CT guided by baseline diagnostic CT or MRI by a radiologist with abdomen imaging expertise. Organ contours were defined using semiautomatic CT segmentation tools.

The healthy liver was defined as liver minus lesions in the liver.

3 | RESULTS

3.1 | Virtual patient phantom test results

3.1.1 | Qualitative assessment

Generally, there was a better visual agreement between CNN dose-rate maps and the ground truth than between DVK/MC dose-rate maps and the ground truth. The example images and line profiles in Figures 4 and 5 and the DRVHs in Figure 6 provide qualitative evidence of the superior performance of the CNN across multiple regions (kidney, abdominal lesion, lung lesion).

3.1.2 | Quantitative assessment

Table S1 reports the mean dose-rate values for organs/lesions across five test phantoms. Figure 7 compares the mean dose-rate error and NRMSE in lesions and organs across all test phantoms. Similar to the results of the qualitative assessment (Figures 4–6), the CNN also consistently showed superior results compared to DVK and MC in quantitative evaluations (Figure 7). For instance, compared to DVK and MC, the CNN estimates showed an average improvement (in mean dose-rate error) of 52%/20%, 55%/53%, 66%/50%, 66%/62%, 48%/49%, and 58%/39% in healthy liver, lesion, left kidney, right kidney, spleen, and lumbar vertebra, respectively. The NRMSE was also substantially lower for the CNN than for DVK and MC across all VOIs (Figure 7). The average improvement (in NRMSE) demonstrated by the CNN compared to DVK/MC was 10%/9%, 18%/17%, 11%/12%, 9%/10%, 26%/27%, and 18%/10% in healthy liver, lesion, left kidney, right kidney, spleen, and lumbar vertebra, respectively. In addition to the improvement in the average values, the maximum errors (denoted by the error bars in Figure 7) were also consistently lower with CNN compared to DVK and MC. In Figure 7, all three methods showed the highest errors for lesion and lumbar vertebra regions. This was attributed to the smaller size of these VOIs compared to other organs and the corresponding increase in partial volume effects. In the case of lumbar vertebra, relevant to bone marrow dosimetry, the very low uptake in these regions also contributed to higher dose-rate errors. For lesions and lumbar vertebra that had a relatively large sample size (15 and 18), a paired t-test demonstrated that the differences of mean dose-rate error and NRMSE between CNN and MC (DVK), as shown in Figure 7, were statistically significant (Table S2 shows p-values). Moreover, DRVHs statistics (DR10, DR30, DR70, DR90) as demonstrated in Figures 8
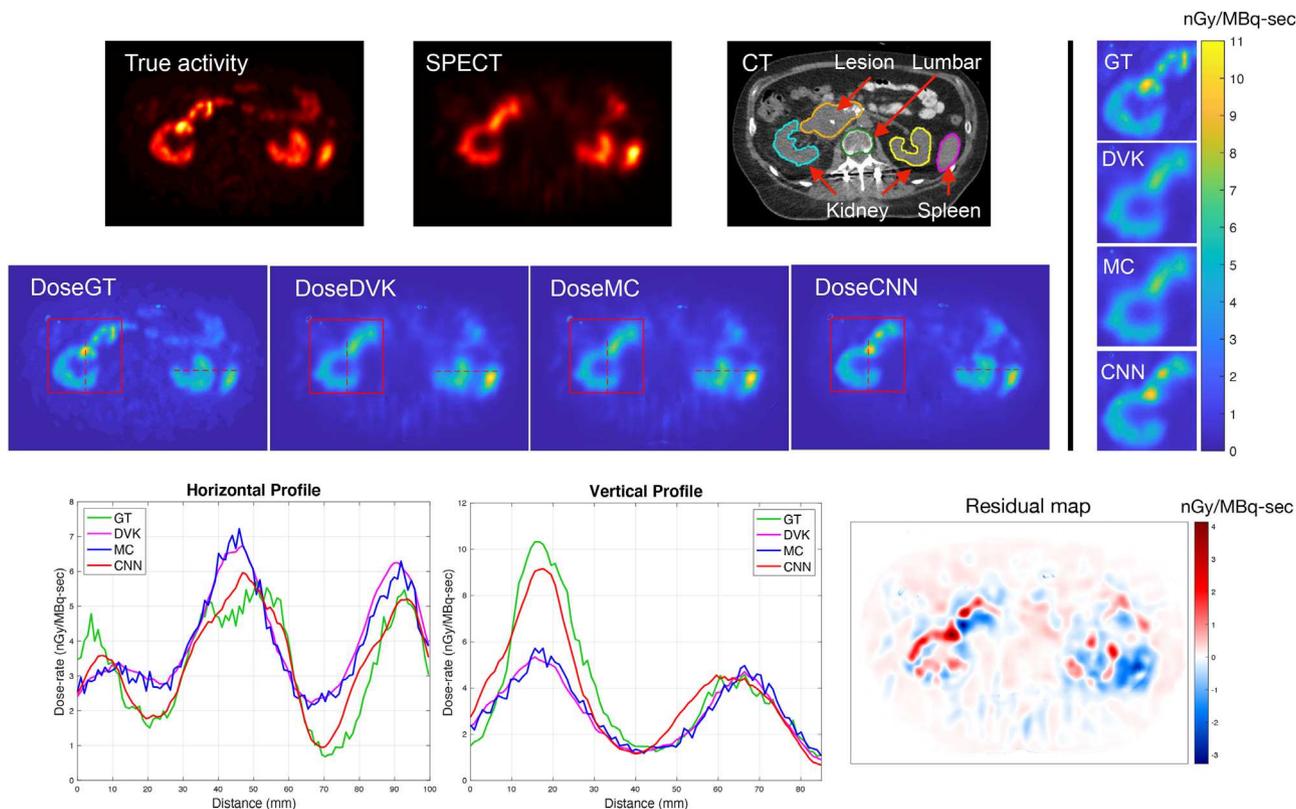
FIGURE 4 One slice of the test virtual patient phantom #2. The top two branches show the true activity map defined based on Ga-68 PET, SPECT, and CT images, the ground truth dose-rate map, and the dose-rate images from the different methods (DVK, MC, and CNN). The bottom branch shows line profiles across the kidney and the residual map (the difference between CNN and DVK dose-rate map). The dose-rate units were normalized to 1 MBq in the field-of-view in all figures

and 9 also show the superiority of the CNN compared to DVK.

3.1.3 | Noise evaluation

Table 1 shows a consistent reduction of ensemble noise in background VOIs with an average of 21% and 27% improvement demonstrated by the CNN compared to DVK and MC (running 1 billion histories), where MC had the highest level of noise due to its statistical nature.

3.2 | Patient results

In patients where there was no known ground truth, results were instead compared visually. Figures 8 and 9 show examples of dose-rate maps corresponding to high-count (day 1 post-therapy) and low-count (day 7 post-therapy) imaging conditions post-Lu-177 DOTATATE. Although concrete conclusions could not be drawn, as there was no known ground truth; visual inspections implied potential reduction of SPECT spatial resolution effects on dose-rate accuracy by our DblurDoseNet. For instance, with the CNN, the

enlarged kidney map and line profiles of Figure 8 show a larger decrease in dose-rate in the medulla and renal pelvis areas, which could be due to the expected lower physiological Lu-177 uptake in this part of the kidney compared to the cortex region. In addition, in Figure 9, the lesion with a necrotic center demonstrated a larger drop in dose-rate at the center with the CNN compared to DVK or MC, which could be due to the expected lower uptake associated with necrosis. Moreover, to demonstrate the generalizability of our CNN on patient data, we tested our CNN using 42 SPECT/CT scans of 12 patients and then compared with DVK and MC dose-rate maps in terms of the mean dose-rate and DRVH statistics (DR10, DR30, DR70, and DR90) across lesions and kidneys. As demonstrated in Table 2, there was a strong agreement between CNN and MC for mean dose-rate in kidneys; and for mean dose-rate in lesions, CNN showed higher values than MC, which could be partially due to the compensation of SPECT resolution effects. For DRVH statistics shown in Figure 6, the CNN and MC results also agreed well in kidney; but for lesions, the CNN measurement showed a lower dose-rate value in DR70 (DR90) and a higher dose-rate value in DR30 (DR10), compared to MC and DVK. The DRVHs in the lesions might be improved because the blurring effects caused by the limited
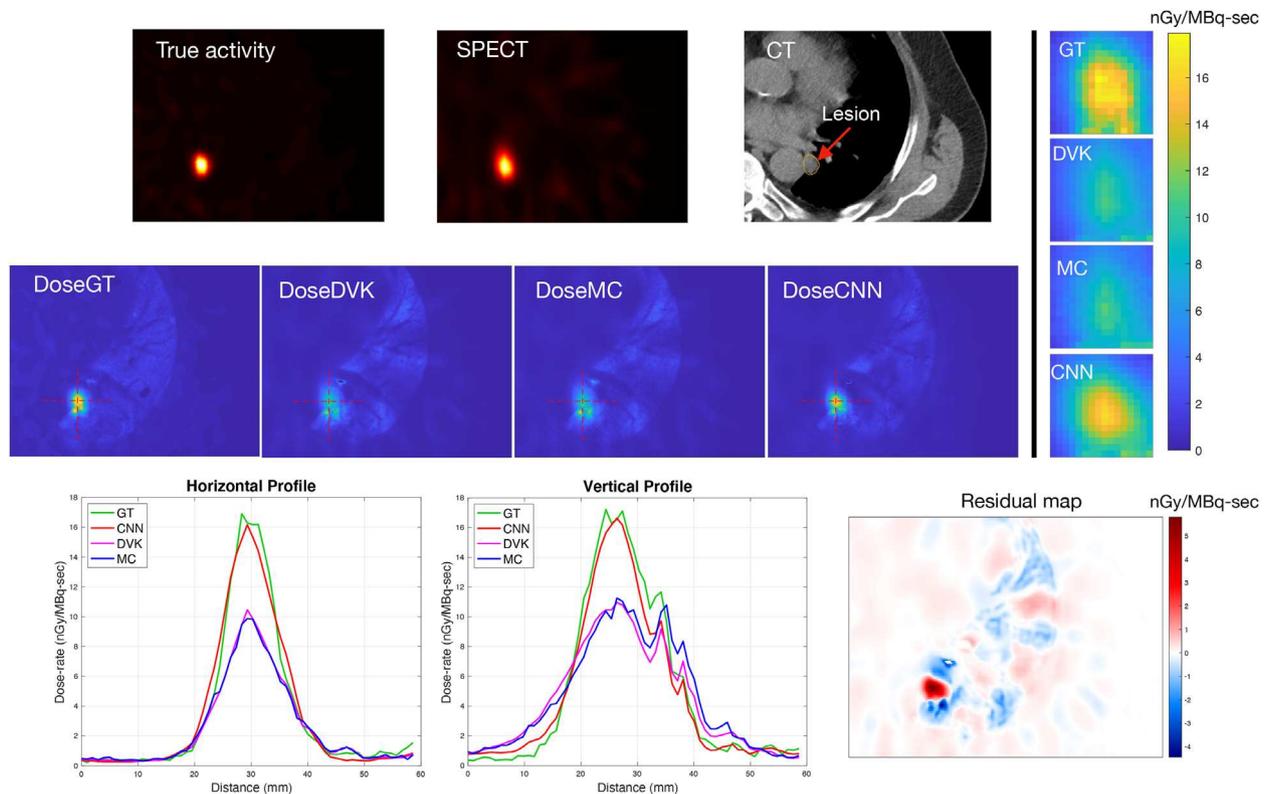
FIGURE 5 One slice of the test virtual patient phantom #5. The top two branches show the true activity map defined based on Ga-68 PET, SPECT, and CT images, the ground truth dose-rate map, and the dose-rate images from the different methods (DVK, MC, and CNN). The bottom branch shows line profiles across the kidney and the residual map (the difference between the CNN and DVK dose-rate maps)

SPECT camera resolution would lead to a higher DR70 (DR90) and a lower DR30 (DR10), but concrete conclusions could not be made due to the absence of ground truth.

3.3 | Comparing performance with a nonresidual network and a 2D network

To demonstrate the effectiveness of residual learning and the 3D convolutional feature extractor that we implemented, we also compared our proposed CNN with a CNN that had the same architecture but without residual learning (not adding the DVK dose-rate map to the output of 2D U-Net), and to a CNN without 3D feature extractor (a purely 2D U-Net where we treated the depth dimension of input as channels). The nonresidual CNN and the 2D CNN were trained using the same hyperparameters and the same training data as for the proposed CNN. All the testing used the same test phantoms demonstrated in the previous section. As shown in Table 3, quantitative comparisons across all test phantoms showed superior results of our proposed CNN (DblurDoseNet) for almost all VOIs except for some cases where all the networks show comparable results. Based on these promising results, we believed the idea of residual learning was effective and

it was beneficial to include a few 3D convolutional layers to extract 3D information rather than using only 2D convolutions.

3.4 | Time cost

We compared the computation times of the different methods for generating a dose-rate map corresponding to the typical $512 \times 512 \times 130$ patient SPECT/CT image size on CPU (Intel Core i9 @2.3 GHz) or GPU (Tesla V100). DVK with density scaling took ~20 s on the CPU and ~10 s on the GPU. DPM MC code took ~60 min simulating 1 billion histories (for both ground truth and test phantoms/patients) on the CPU while running DPM on a GPU is not an option at this time (we are unaware of any MC code for internal therapy running on a GPU). The CNN took ~20 min on the CPU and ~20 s using the GPU. After considering the DVK precomputation time for the residual learning network, the total GPU time cost for the CNN with residual learning is ~20 + 10 s.

4 | DISCUSSION

Reliable voxel-level dosimetry requires reliable dose-rate images at multiple timepoints as well as dependable
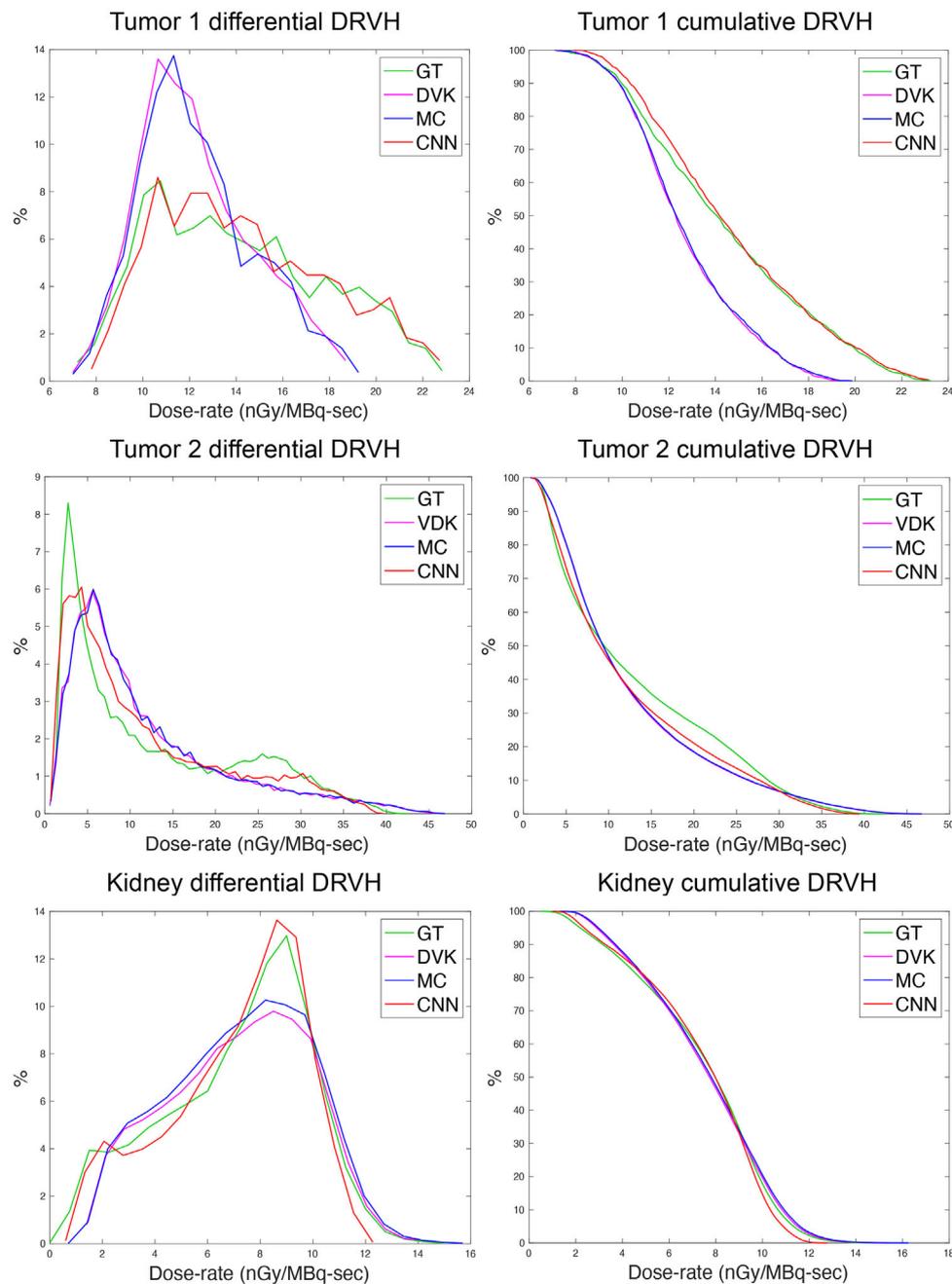
FIGURE 6 Tumor and kidney differential and cumulative dose-rate volume histograms corresponding to DVK, MC, CNN and the ground-truth dose-rate maps of virtual patient phantoms. The sizes of tumors 1 and 2 are 4 and 65 ml, respectively

co-registration and fitting of the dose-rate versus time data estimated at the voxel level. Performing reliable voxel-level co-registration and fitting to generate dose maps can be challenging, but the feasibility has been demonstrated.[32,33] In this work, we focused on generating reliable dose-rate maps. With evaluation both on virtual patient phantoms that covered clinically relevant conditions and patients who underwent Lu-177 DOTATATE therapy in our clinic, we demonstrated that our CNN using residual learning framework could provide fast and accurate dose-rate estimation. Despite using only a moderate amount of training data, Dblur-

DoseNet provided consistently superior performance over conventional voxel dosimetry in terms of resolution, accuracy, and noise across multiple regions including kidneys, lumbar vertebra, and lesions in soft tissue and lung. Importantly, for clinical implementation, the CNN voxel dose-rate map for a 512 × 512 × 130 patient image could be generated in ∼30 s, which was a fraction of the time associated with running MC, the current gold standard. Although generating the ground-truth labels for training by MC was computationally expensive, this effort was needed only once at training time, for a given SPECT imaging system.
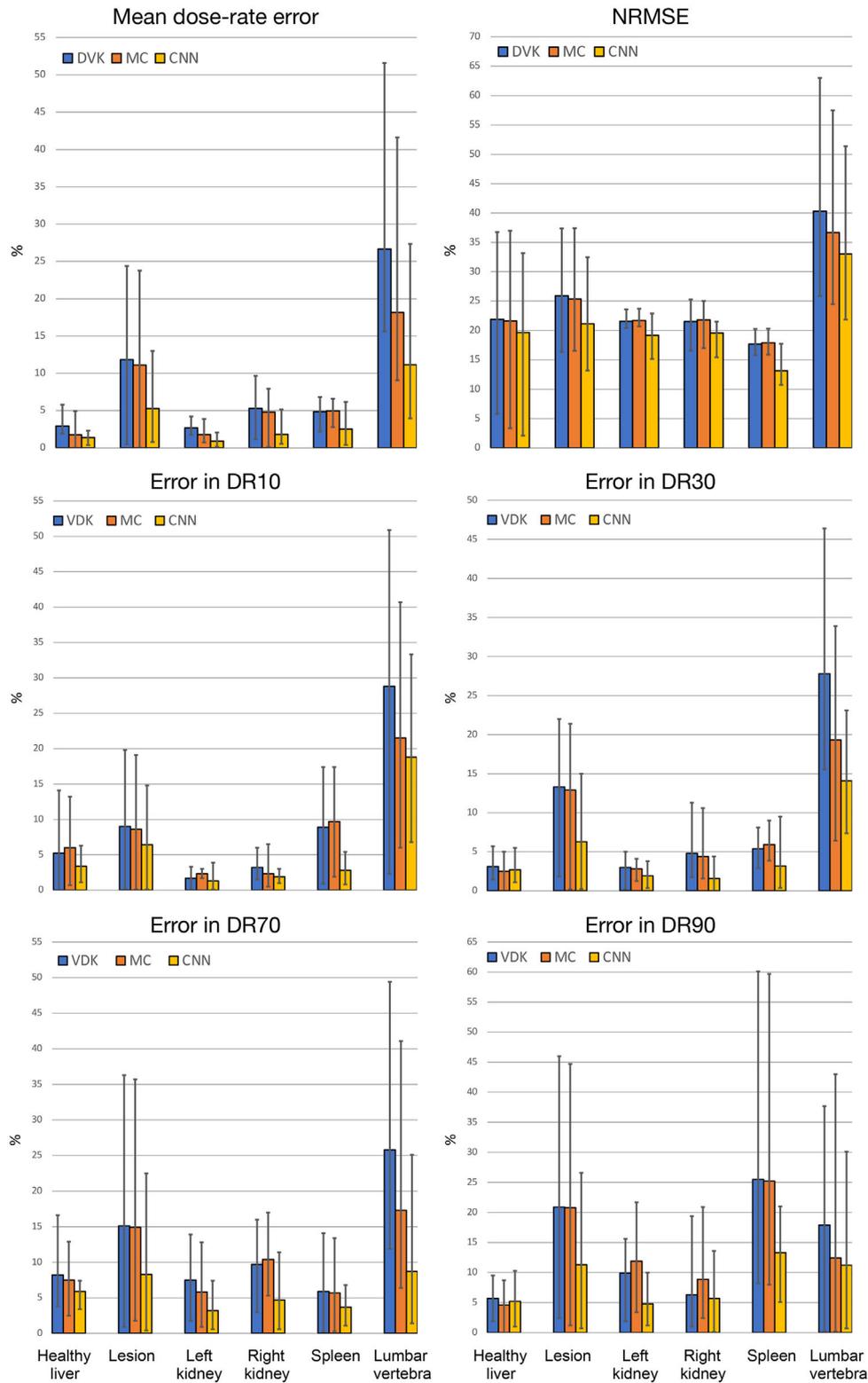
FIGURE 7 Mean dose-rate error, NRMSE, and error in DRVH statistics (DR10, DR30, DR70, DR90) comparison of DVK, MC, and CNN relative to ground-truth dose-rate map across all test phantoms. Median (range) VOI volumes are as follows: healthy liver (liver minus lesions): 1607 ml (1164–2262 ml); lesion: 16 ml (4–181 ml); left kidney: 177 ml (98–211 ml); right kidney: 156 ml (76–249 ml); spleen: 191 ml (131–467 ml); lumbar vertebra L2 to L5: 54 ml (34–68 ml)
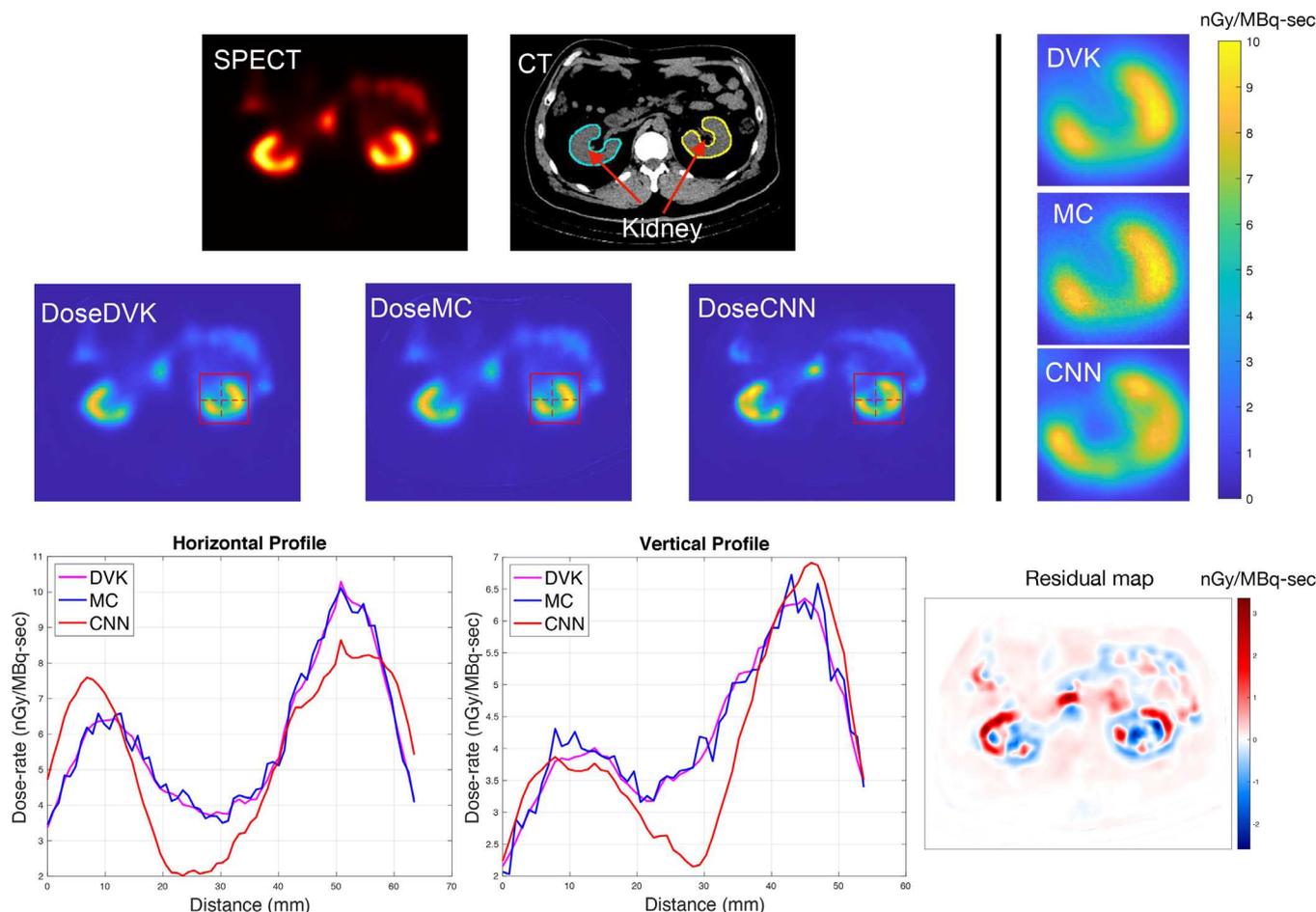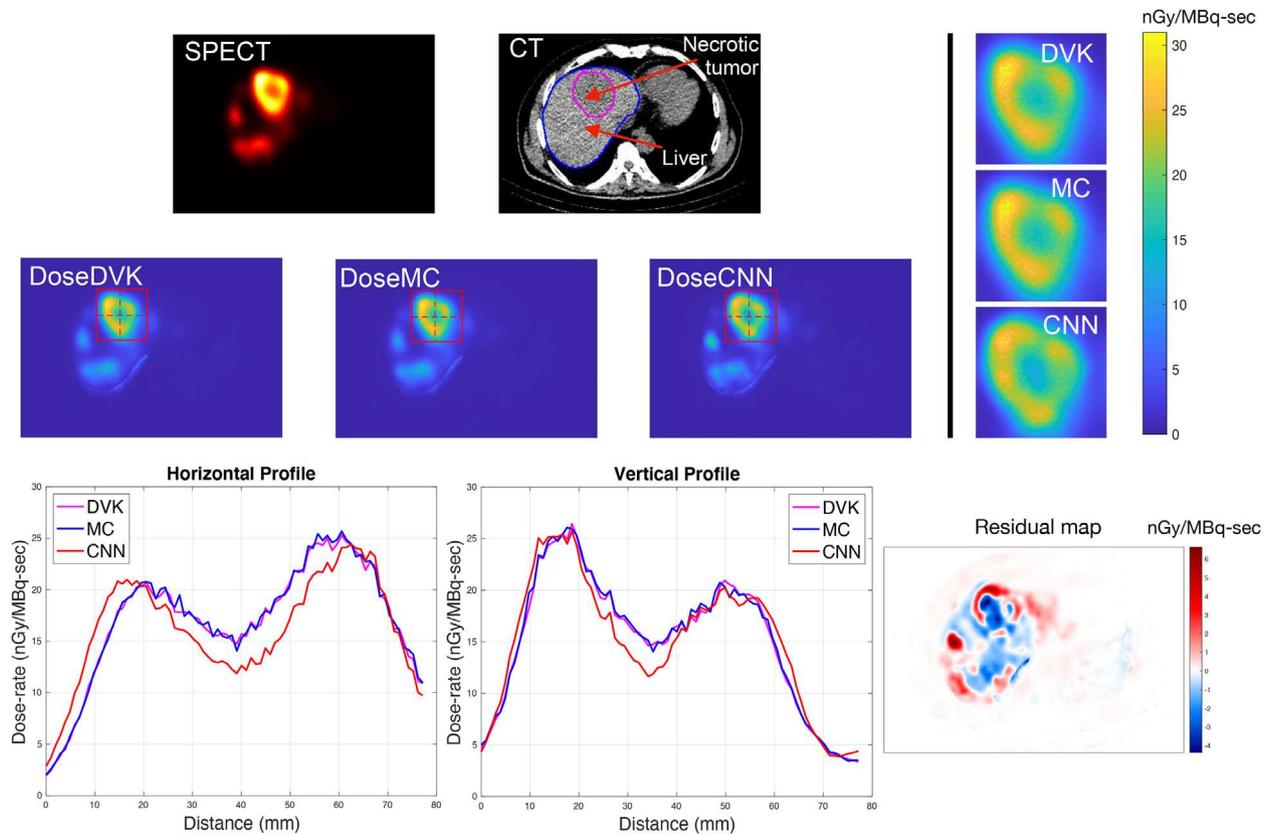
FIGURE 8 One slice across kidney of the input images (SPECT, CT) and output DVK, MC, and CNN dose-rate maps and line profiles for a patient imaged after Lu-177 DOTATATE (at day 1 post-therapy). The residual map is the difference between CNN and DVK dose-rate map

The main limitation to accurate voxel-level patient-specific dose-rate estimation with nonlearning-based methods is the poor spatial resolution associated with the input SPECT (or PET) images. This issue was evident in our results where the theoretically accurate MC-based calculation only slightly outperformed DVK with density scaling. In contrast, by using the true activity map-based dose-rate estimates for training, our CNN has the ability to "learn" the physics of dose deposition and to compensate for the SPECT resolution effects that both lead to blurring of the conventional (nonlearning-based) dose-rate maps, as demonstrated in the phantom results (Figures 4–7, Table 1). In patient studies, potential mitigation of SPECT resolution effects was demonstrated empirically. In Figure 8, the CNN-based estimates show sharper line profiles and larger drops in dose-rate over the medulla area of the kidney, analogous to the illustration of Figure 1. In Figure 9, the larger drop of dose-rate in the necrotic center of a tumor may reflect what is expected based on physiology. Although test results were promising over 42 scans originating from 12 patients, further testing is planned as more patient images become available. We did not investigate training with more virtual patients, because simulating Lu-177 SPECT projections by full MC simulation was computationally expensive. Furthermore, we found that our CNN, trained by nine virtual patient phantoms, was able to generate promising dose-rate estimates across a diverse range of test cases. We expect that applying self/weakly supervised training may address the computational inefficiency of simulating Lu-177 SPECT projections in the future. In addition, due to the lack of ground truth, we were unable to make concrete conclusions about the performance of our CNN on test patient data. But the uncertainty of our CNN can be quantified by generating confidence maps[34–36] using Bayesian networks,[37] an ensemble of multiple networks,[38] or an extension of the probabilistic U-Net,[39] which can be one direction to investigate in the future.

The mean dose-rate errors shown in Figure 7, especially for lesions, were generally lower than one would expect based on reported activity recovery in quantitative Lu-177 SPECT phantom studies. For example, for 72 OSEM updates, activity recovery of only 80% was

FIGURE 9 One slice across lesion of the input images (SPECT, CT) and output DVK, MC, and CNN dose-rate maps and line profiles for a patient imaged after Lu-177 DOTATATE (at day 7 post-therapy). The residual map is the difference between CNN and DVK dose-rate map

TABLE 1 Ensemble noise from three realizations for DVK, MC, and CNN across all test phantoms. The number of voxels ranged from 2527 to 23,411

Ensemble noise	Background region	DVK	MC	CNN
Phantom #1	Liver and spleen	4.6%	6.1%	3.4%
Phantom #2	Liver	12.6%	13.3%	9.2%
Phantom #3	Liver	14.0%	14.6%	12.9%
Phantom #4	Liver	20.3%	19.6%	14.8%
Phantom #5	Spleen	7.1%	7.6%	5.8%

reported for a 26.5 ml volume "hot" sphere in a "warm" background region.[40] The results of the current study showed lower errors because, unlike in a physical phantom, the assigned "true" activity values at the boundary of the structures in our PET-based virtual patients did not drop off sharply, and instead, were blurred out. Moreover, in Figure 7, all three methods showed the largest mean dose-rate error for lesions and lumbar vertebra, as expected due to the relatively smaller sizes of these structures compared to other organs, and hence partial volume effects associated with SPECT resolution were higher. The large error for the lumbar region with DVK (~25%) was likely to be due to the heterogeneous tissue within this region, which includes cortical bone, trabecula bone, and yellow and red marrow. Regarding DVK,

the simple density scaling that was performed in our study was potentially inadequate for this region. Furthermore, the Lu-177 uptake in a lumbar region was very low, so the cross-dose contribution to dose-rate there, including the photon cross-dose, could be significant. Our $99 \times 99 \times 99$ photon kernel may have been insufficient to capture the full photon cross dose contribution to the lumbar vertebra. Our study did not include standard partial volume correction using volume-dependent recovery coefficients (RCs) because such methods provide only a mean dose, not a voxel-level correction. Furthermore, the limitations of standard RC methods due to dependence on object shape, activity distribution and target-to-background ratios are well known. Voxel-level partial volume correction is much more challenging[41] and their applications in SPECT are not well established. Our results demonstrated that training using true dose-rate maps could reduce the need for such corrections to compensate for resolution effects.

To define our virtual patient activity maps, we chose to use Ga-68 DOTATATE PET/CT to exploit the availability of these images that had finer resolution than SPECT and showed similar uptake patterns as Lu-177 DOTATATE. Despite the standard practice of using Ga-68 PET or Lu-177 SPECT as a theranostic pair, some differences between the two distributions were to be expected, but we did not expect this to impact our CNN

TABLE 2 Dose-rate values (mean dose-rate and DRVH statistics) for DVK, MC, and CNN methods averaged across all 42 scans from 12 patients

| | | Dose-rate* (nGy/MBq-sec) | | |
		DVK	MC	CNN
Lesion	Mean dose-rate	13.7 (0.2–87.9)	13.9 (0.3–88.9)	14.4 (0.3–88.8)
	DR10	25.1 (0.4–177)	25.4 (0.4–179)	27.8 (0.5–188)
	DR30	15.8 (0.3–122)	16.1 (0.3–123)	16.9 (0.3–127)
	DR70	8.2 (0.2–48.1)	8.3 (0.3–48.5)	8.1 (0.3–48.8)
	DR90	5.4 (0.1–17.1)	5.6 (0.1–19.3)	5.0 (0.1–22.9)
Left kidney	Mean dose-rate	3.7 (0.7–8.6)	3.8 (0.7–8.7)	3.8 (0.7–8.3)
	DR10	6.0 (1.4–12.7)	6.1 (1.4–12.8)	5.8 (1.4–12.3)
	DR30	4.6 (1.0–10.7)	4.7 (1.0–11.0)	4.6 (1.0–10.7)
	DR70	2.7 (0.2–7.0)	2.7 (0.2–7.0)	2.8 (0.2–7.0)
	DR90	1.6 (0.1–3.8)	1.7 (0.1–3.8)	1.6 (0.1–3.6)
Right kidney	Mean dose-rate	4.2 (0.7–9.1)	4.3 (0.8–9.2)	4.2 (0.8–9.1)
	DR10	7.1 (1.5–17.2)	7.2 (1.6–17.2)	7.0 (1.8–15.2)
	DR30	5.2 (1.0–11.8)	5.3 (1.1–11.8)	5.3 (1.1–12.2)
	DR70	2.8 (0.2–7.3)	2.9 (0.2–7.3)	2.8 (0.2–7.7)
	DR90	1.6 (0.1–4.4)	1.7 (0.1–4.5)	1.5 (0.1–4.0)

Note: Minimum and maximum values are shown in parenthesis. The medians (ranges) for the VOI volumes are as follows: Lesion: 15 ml (2.3–582 ml); left kidney: 192 ml (105–275 ml); right kidney: 180 ml (122–259 ml).
*Reported dose-rates are normalized to 1 MBq in field-of-view.

TABLE 3 Mean (maximum) dose-rate error and NRMSE comparison between CNN with and without residual learning and with 2D and 3D networks evaluated across VOIs in all test phantoms

| | Mean dose-rate error | | | NRMSE | | |
	3D w/ res (DblurDoseNet)	3D w/o res	2D w/ res	3D w/ res (DblurDoseNet)	3D w/o res	2D w/ res
Healthy liver	1.4% (2.3%)	5.5% (7.0%)	1.2% (3.1%)	19.6% (33.2%)	21.6% (35.1%)	23.2% (33.3%)
Lesion	5.3% (13.0%)	6.9% (12.5%)	6.0% (13.9%)	21.2% (32.5%)	21.4% (31.5%)	21.8% (38.0%)
Liver	1.9% (3.5%)	5.7% (7.6%)	1.9% (4.8%)	20.6% (26.3%)	21.6% (27.6%)	22.8% (26.6%)
Left kidney	0.9% (2.1%)	5.2% (6.5%)	1.8% (3.8%)	19.2% (22.9%)	20.1% (22.0%)	19.0% (20.8%)
Right kidney	1.8% (5.1%)	5.8% (12.6%)	2.6% (7.5%)	19.6% (21.5%)	20.5% (24.3%)	20.0% (23.6%)
Spleen	2.5% (6.2%)	6.3% (9.5%)	2.2% (6.4%)	13.1% (17.7%)	14.4% (19.9%)	13.2% (18.2%)
Lumbar vertebra	11.1% (27.4%)	10.5% (27.2%)	12.1% (30.6%)	33.0% (51.4%)	32.9% (49.1%)	32.7% (50.2%)

performance because the PET images were used only to define the virtual patient phantoms and not in the training process itself, as proposed in another study.[42] Ideally, however, images of higher resolution than clinical PET should be considered as the true representation of the activity map of patients when generating the virtual patient training set, but usually they are not readily available. To circumvent this issue, we also investigated using phantoms with piece-wise uniform uptake in CT-defined organs/lesions for training (such as XCAT[22] in Figure 1), but we found that such training led to unnaturally uniform dose-rate maps when tested on patient images. We also fed our CNN with an all-zero activity map, as a sanity check to our proposed framework as well as implementation. The output dose-rate map, as expected, was all zeros. This illustrates that if there is no apparent sig-

nal in the reconstructed SPECT, then there will not be any unexpected nonzero values in the dose-rate map. A possible alternative to our PET-based virtual patient activity maps is to assign distributions based on high-resolution animal models, for instance, ex-vivo autoradiography showing uptake distribution of DOTATATE in kidney.[43]

Our results also demonstrate the advantage of residual learning framework exploiting the fast DVK approach as an initial estimate, which was not utilized in the prior studies.[14–17] We also conjectured that incorporating residual learning could not only improve performance on the test data, but also accelerate the training process. As shown in Figure S1, after 200 epochs of training, the training/validation loss of residual CNN went down to 288/410 at the last 50 epochs, compared to 902/1250

without residual learning, which means fewer epochs are needed to train the residual CNN. Other than using a fast DVK approach for residual learning, an alternative was to generate a quick MC (low number of histories) estimate, which was not explored here. Another advantage of our network is that we first implemented a couple of depth feature extractor layers that shrink the 3D input into 2D at the beginning of our network. Compared to fully 3D approaches, this approach leads to a network having fewer parameters (because 2D kernels have fewer parameters than 3D kernels), so it is less likely to overfit the training data, avoiding a common problem in deep learning applications for medical imaging, where only moderate amount of training data is available. Another option that we did not investigate is to use 2.5D CNN architectures.[44] A potential drawback of our proposed CNN is possible discontinuity of pixel values in coronal slices; however, we did not observe such discontinuity as evident in Figure S2, presumably due to the 11-slice sliding window.

We expect that training a single CNN, as we did in the current study, is simpler than training two separate CNNs to learn the dosimetry and SPECT resolution effects. Typically, there will be three stages needed to train two separate CNNs; stage 1: training CNN-A for SPECT resolution; stage 2: training CNN-B for dosimetry; stage 3: jointly fine-tuning CNN-A and CNN-B. Compared to our proposed end-to-end network (DblurDoseNet), which only involves one training stage, such 3-stage of training will be more complex and potentially inefficient. However, only through comprehensive comparisons can one draw definite conclusions between these two approaches, which we expect to undertake in the future. Although our study only investigated Lu-177 dosimetry, we expect that by changing the training dataset and making minor modifications to the architecture, our CNN approach can be extended to other radionuclides including Y-90 that is a pure-beta emitter and I-131 that has significant beta and gamma contributions to the dose-rate.

5 | CONCLUSION

We constructed and tested a residual CNN that was trained on virtual patient phantom images to learn the mapping from SPECT/CT images to the corresponding dose-rate maps. We took the novel approach of using a single CNN to learn not only the dose-rate estimation but also to compensate for blurring of the dose-rate map due to poor SPECT resolution. Across multiple regions such as kidney, lumbar vertebra and lesions in both soft tissue and lung, the proposed residual DblurDoseNet was able to outperform conventional voxel-level dosimetry methods, including the current "gold standard" MC, in terms of accuracy, noise, and speed. Patient-specific voxel-level dose rate maps can be generated in ∼30 s on GPU; hence the CNN approach has much promise for real-time clinical use in radionuclide therapy dosimetry for treatment planning.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

INFORMED CONSENT
Informed consent was obtained from all individual participants included in the study according to the University of Michigan Institutional Review Board (IRB) criteria. The current study only involved retrospective analysis of existing SPECT/CT data.

DATA AVAILABILITY STATEMENT
The training/testing virtual patient phantom images are available at: https://deepblue.lib.umich.edu/data/concern/data_sets/s1784m12w?locale=en. The CNN was implemented in PyTorch and is available at: https://github.com/ZongyuLi-umich/DblurDoseNet. The Image reconstruction algorithm (OSEM) was implemented in MATLAB and is available at: https://github.com/JeffFessler/mirt. DOI for the dataset: https://doi.org/10.7302/ykz6-cn05.

REFERENCES

1. Eberlein U, Cremonesi M, Lassmann M. Individualized dosimetry for theragnostic: necessary, nice to have, or counterproductive? *J Nuc Med*. 2017;58(2):97-103.
2. Bolch WE, Bouchet LG, Robertson JS, et al. MIRD pamphlet No. 17: the dosimetry of nonuniform activity distributions–radionuclide S values at the voxel level. *J Nuc Med*. 1999;40:11-36.
3. Kortesniemi M, Tsapaki V, Trianni A, et al. The European Federation of Organizations for Medical Physics (EFOMP) white paper: big data and deep learning in medical imaging and in relation to medical physics profession. *Phys Med*. 2018;56:90-93.
4. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Imag Anal*. 2017;42:60-88.
5. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift Gür Medizinische Physik*. 2019;29(2):102-127.
6. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46(1):1-36.
7. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imag Rad Sci*. 2019;50(4):477-487.
8. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. 2015;9351:234-241.
9. Uribe CF, Mathotaarachchi S, Gaudet V, et al. Machine learning in nuclear medicine: part 1—introduction. *J Nuc Med*. 2019;60(4):451-458.
10. Zukotynski K, Gaudet V, Uribe CF, et al. Machine learning in nuclear medicine: part 2-neural networks and clinical aspects. *J Nuc Med*. 2021;62(1):22-29.
11. Seifert R, Weber M, Kocakavuk E, Rischpler C, Kersting D. Artificial intelligence and machine learning in nuclear medicine: future perspectives. *Semi Nuc Med*. 2021;51(2):170-177.

12. Veit-Haibach P, Buvat I, Herrmann K. EJNMMI supplement: bringing AI and radiomics to nuclear medicine. *Euro J Nuc Med Mol Imag*. 2019;46(13):2627-2629.

13. Arabi H, Zaidi H. Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy. *Euro J Hybr Imag*. 2020;4(1):1-23.

14. Akhavanallaf A, Shiri I, Arabi H, Zaidi H. Whole-body voxel-based internal dosimetry using deep learning. *Euro J Nuc Med Mol Imag*. 2021;48(3):670-682.

15. Lee MS, Hwang D, Kim JH, Lee JS. Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry. *Sci Rep*. 2019;9(1):10308-10309.

16. Götz TI, Schmidkonz C, Chen S, Al-Baddai S, Kuwert T, Lang EW. A deep learning approach to radiation dose estimation. *Phys Med Bio*. 2020;65(3) 035007.

17. Götz TI, Lang EW, Schmidkonz C, Kuwert T, Ludwig B. Dose voxel kernel prediction with neural networks for radiation dose estimation. *Zeitschrift Für Medizinische Physik*. 2021;31(1):23-36.

18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016. https://doi.org/10.1109/CVPR.2016.90

19. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Img Proc*. 2017;26(7):3142-3155.

20. Xu J, Gong E, Pauly J, Zaharchuk G. 200x low-dose PET reconstruction using deep learning. arXiv:1712.04119; 2017.

21. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imag*. 2017; 36(12): 2524-2535.

22. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010;37(9):4902-4915.

23. Soderlund AT, Chaal J, Tjio G, Totman JJ, Conti M, Townsend DW. Beyond 18F-FDG: characterization of PET/CT and PET/MR scanners for a comprehensive set of positron emitters of growing application—18F, 11C, 89Zr, 124I, 68Ga, and 90Y. *J Nuc Med*. 2015;56(8):1285-1291.

24. Ljungberg M. The SIMIND Monte Carlo program. *Monte Carlo Calculations in Nuclear Medicine*. CRC Press; 2012. https://doi.org/10.1201/b13073-8

25. Sempau J, Wilderman SJ, Bielajew AF. DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations. *Phys Med Bio*. 2000;45(8):2263-2291.

26. Wilderman SJ, Dewaraja YK. Method for fast CT/SPECT-based 3D Monte Carlo absorbed dose computations in internal emitter therapy. *IEEE Trans Nuc Sci*. 2007;54(1):146-151.

27. Dieudonné A, Hobbs RF, Lebtahi R, et al. Study of the impact of tissue density heterogeneities on 3-dimensional abdominal dosimetry: comparison between dose kernel convolution and direct Monte Carlo methods. *J Nuc Med*. 2013;54(2):236-243.

28. Götz T, Schmidkonz C, Lang EW, Maier A, Kuwert T, Ritt P. A comparison of methods for adapting 177Lu dose-voxel-kernels to tissue inhomogeneities. *Phys Med Biol*. 2019;64(24) 245011.

29. Lee MS, Kim JH, Paeng JC, et al. Whole-body voxel-based personalized dosimetry: the multiple voxel s-value approach for heterogeneous media with nonuniform activity distributions. *J Nuc Med*. 2018;59(7):1133-1139.

30. Brice DK. Stopping powers for electrons and positrons, ICRU report 37. *Nuc Inst Meth Phys Res*. 1985;12(1):187-188.

31. Kingma DP & Ba J Adam: a method for stochastic optimization. IEEE International Conference on Learning Representations (ICLR); San Diego, CA, USA, May 7-9, 2015.

32. Jackson P, Beauregard J-M, Hofman M, Kron T, Hogg A, Hicks R. An automated voxelized dosimetry tool for radionuclide ther-

33. apy based on serial quantitative SPECT/CT imaging. *Med Phys*. 2013;40:112503.

34. Sarrut D, Halty A, Badel JN, Ferrer L, Bardiès M. Voxel-based multimodel fitting method for modeling time activity curves in SPECT images. *Med Phys*. 2017;44(12):6280-6288.

35. Laves M-H, Tölle M, Ortmaier T. Uncertainty estimation in medical image denoising with Bayesian deep image prior. arXiv:2008.08837; 2020.

36. Baumgartner CF, Tezcan KC, Chaitanya K, et al. PHiSeg: capturing uncertainty in medical image segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Springer; 2019:119-127.

37. Yasarla R, Patel VM. Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15-20 June 2019.

38. Kendall A, Badrinarayanan V, Cipolla R. Bayesian SegNet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. Proceedings of the British Machine Vision Conference (BMVC), 2017, 57.1–57.12. BMVA Press.

39. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, Long Beach, CA, 2017.

40. Kohl SAA, Romera-Paredes B, Meyer C, et al. *A probabilistic U-Net for segmentation of ambiguous images*. NIPS'18. Curran Associates Inc.; 2018: 6965-6975.

41. Tran-Gia J, Lassmann M. Characterization of noise and resolution for quantitative 177 Lu SPECT/CT with XSPECT quant. *J Nuc Med*. 2019;60(1):50-59.

42. Thomas BA, Cuplov V, Bousse A, et al. PETPVC: a toolbox for performing partial volume correction techniques in positron emission tomography. *Phys Medic Biol*. 2016;61(22):7975-7993.

43. Xue S, Gafita A, Afshar-Oromieh A, Eiber M, Rominger A, Shi K. Voxel-wise prediction of post-therapy dosimetry for 177Lu-PSMA I&T therapy using deep learning. *J Nuc Med*. 2020;61(Suppl 1):1424-1424.

44. Melis M, Krenning EP, Bernard BF, Barone R, Visser TJ, de Jong M. Localisation and mechanism of renal retention of radiolabelled somatostatin analogues. *Euro J Nuc Med Mol Imag*. 2005;32(10):1136-1143.

45. Ziabari A, Ye DH, Srivastava S, Sauer KD, Thibault J-B, Bouman CA. 2.5D deep learning for CT image reconstruction using a multi-GPU implementation. *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE; 2018:2044-2049. https://doi.org/10.1109/ACSSC.2018.8645364

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

SUPPLEMENT

This supplemental material is for the paper "A Deep Residual Learning Network for Voxel Radionuclide Dosimetry Compensating for SPECT Imaging Resolution" by Zongyu Li, Jeffrey A. Fessler, Justin K. Mikell, Scott J. Wilderman and Yuni K. Dewaraja. The paper is submitted to Medical Physics.

Fig. S.1 compares the training/validation loss curve (200 epochs, MSE loss) of residual learning with non-residual learning. Fig. S.2 shows discontinuity of pixels along axial direction was not observed in our CNN method. Table I shows the ground truth dose-rate values of test virtual patient phantoms across multiple volumes of interest (VOI) including liver, lesion and kidneys, and corresponding values estimated by DVK, MC and CNN methods. Table II shows p-values of 3 paired t-tests (DVK-MC, DVK-CNN, MC-CNN) using mean dose-rate error and NRMSE on VOIs of test virtual patient phantoms.
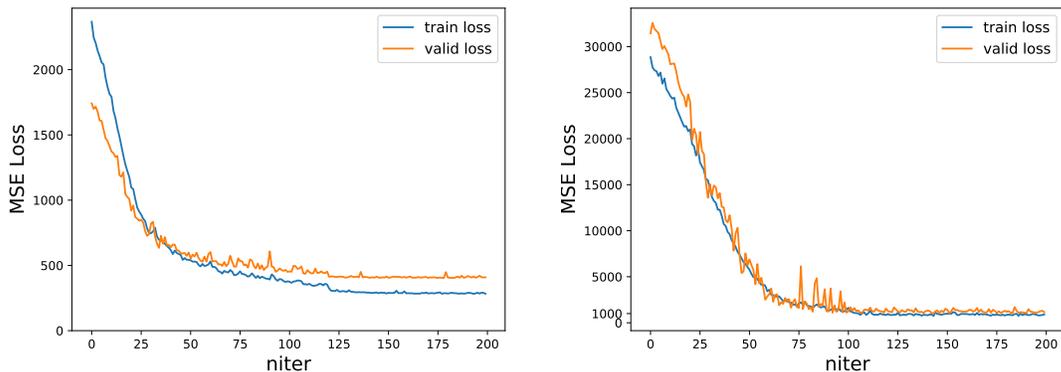
(a) Training with residual learning. Averaged train/valid loss for last 50 epochs are **288/410**.

(b) Training without residual learning. Averaged train/valid loss for last 50 epochs are 902/1250.

Fig. S.1: Train/valid loss curve of residual CNN vs. non-residual CNN during 200 epochs of training.

(a) DoseGT

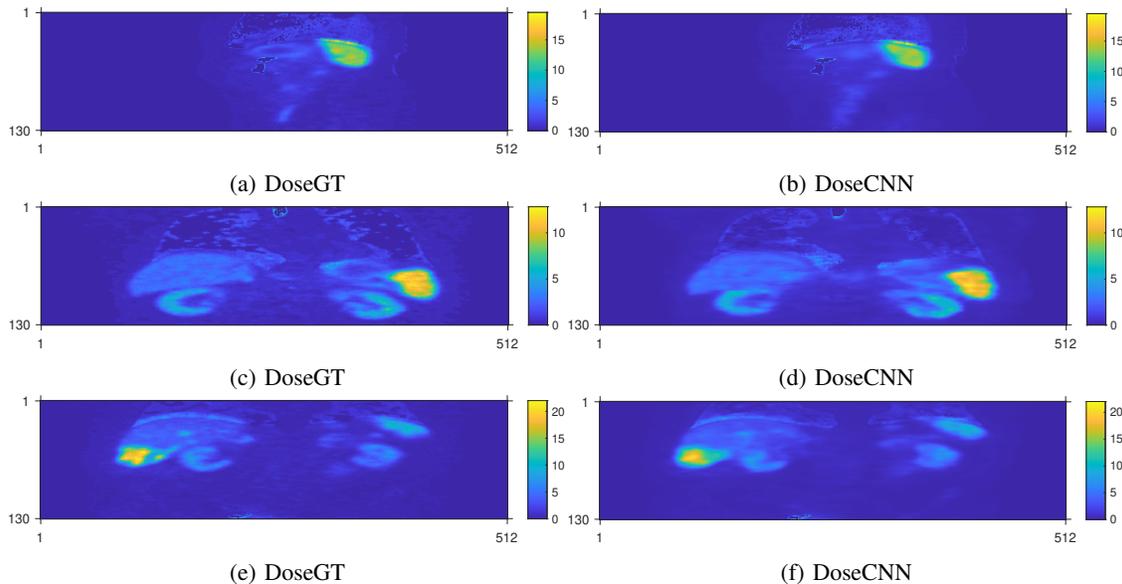(b) DoseCNN

(c) DoseGT

(d) DoseCNN

(e) DoseGT

(f) DoseCNN

Fig. S.2: GT and CNN dose-rate maps along coronal directions. The CNN maps are reasonably smooth (discontinuities are not observed) along the axial direction, presumably due to the 11-slice sliding window approach. Unit in all colorbars is nGy/MBq-sec.

TABLE I: Dose-rate values comparing GT, DVK, MC and CNN in VOIs across all 5 test phantoms. The medians (ranges) for the VOI volumes are: healthy liver: 1607mL (1164mL – 2262mL); lesion: 16mL (4mL – 181mL); left kidney: 177mL (98mL – 211mL); right kidney: 156mL (76mL – 249mL); spleen: 191mL (131mL – 467mL); lumbar: 54mL (34mL – 68mL). Healthy liver is defined as liver minus lesions. Lumbar vertebra covers L2 to L5. *Reported dose-rates are normalized to 1 MBq in field-of-view.

Organ/Dose-rate* (nGy/MBq-sec)	GT	DVK	MC	CNN
Healthy liver	3.4 (2.3 − 4.4)	3.3 (2.1 − 4.4)	**3.4 (2.2 − 4.4)**	**3.4 (2.2 − 4.4)**
Lesion	12.4 (2.8 − 15.2)	10.9 (2.6 − 14.8)	10.9 (2.7 − 14.9)	**11.8 (2.6 − 14.7)**
Liver	4.0 (2.3 − 5.2)	3.9 (2.1 − 5.1)	**4.0 (2.2 − 5.1)**	**4.0 (2.2 − 5.1)**
Left kidney	5.5 (3.7 − 7.2)	5.3 (3.6 − 7.0)	5.4 (3.6 − 7.1)	**5.5 (3.7 − 7.0)**
Right kidney	5.7 (3.7 − 7.8)	5.6 (3.5 − 8.2)	5.6 (3.5 − 8.2)	**5.7 (3.7 − 8.0)**
Spleen	8.1 (5.3 − 11.4)	8.3 (5.0 − 11.7)	8.3 (5.0 − 11.7)	**8.0 (5.0 − 11.2)**
Lumbar vertebra	0.88 (0.35 − 3.4)	0.67 (0.17 − 2.7)	0.73 (0.21 − 2.7)	**0.79 (0.26 − 3.0)**

TABLE II: P-value of the paired t-test using mean dose-rate error (NRMSE) of VOIs in 5 virtual phantoms.

Organ/p-value	sample size	DVK-MC	DVK-CNN	MC-CNN
Healthy liver	5	0.271 (**0.002**)	0.064 (**0.02**)	0.736 (**0.01**)
Lesion	15	**0.016 (0.048)**	**4.0E-4 (2.1E-5)**	**0.002 (2.1E-4)**
Left kidney	5	**0.003 (0.036)**	0.051 (0.084)	0.304 (0.075)
Right kidney	5	0.347 (0.122)	**0.015 (0.016)**	**0.038 (0.003)**
Liver	5	0.177 (0.187)	0.261 (**0.004**)	0.767 (**0.003**)
Spleen	5	0.781 (**0.049**)	**0.049 (0.004)**	0.066 (**0.004**)
Lumbar vertebra	18	**3.4E-10 (9.3E-9)**	**1.5E-7 (1.8E-5)**	**1.0E-4 (3.1E-4)**