A Deep Residual Learning Network for Practical Voxel Dosimetry in Radionuclide Therapy

Zongyu Li, *Student Member, IEEE,* Jeffrey A. Fessler, *Fellow, IEEE,* Justin K. Mikell, *Member, IEEE,* Scott J. Wilderman and Yuni K. Dewaraja, *Member, IEEE*

2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) | 978-1-7281-7693-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/NSS/MIC42677.2020.9507764

Abstract*—Current standard methods for voxel-level dosimetry in radionuclide therapy suffers from a tradeoff between accuracy and computational efficiency. Monte Carlo (MC) radiation transport algorithms are considered as the gold standard, but are associated with long computation time, while fast voxel dose kernel (VDK) based methods can be inaccurate in the presence of tissue density heterogeneities. This paper investigates a deep residual Convolutional Neural Networks (CNN) approach that learns the difference between the MC and the VDK dose-rate maps to address the speed-accuracy trade-off issue. As with MC and VDK-based dosimetry, the input to the CNN was the patient's SPECT activity map and CT-based density map. MC dosimetry was used only during the training process to generate ground truth training labels. Furthermore, to potentially account for the degradation of dose-rate maps due to poor SPECT spatial resolution, we trained the CNN using dose-rate maps directly corresponding to phantom activity/density maps that were generated from patient's PET scans. The test data consisted of phantom simulations and one patient who underwent ^{177}Lu DOTATATE therapy for neuroendocrine tumors. In phantom cases, the lesion/organ mean dose-rates from ground truth (GT) agreed better with the CNN dose-rates compared to VDK with density scaling, with an average of 60% improvement for lesions and 55%, 63% improvement for left/right kidney, respectively. For all regions, the normalized root mean square error (NRMSE) relative to GT was substantially lower with CNN than with VDK and MC, i.e., an average of 23%, 22% improvement for lesion, respectively. Using a GPU, the CNN took only about 2.0 seconds to generate a patient's 512×512×130 absorbed dose-rate map while the same calculation took about 40 minutes using our fast in-house Dose Planning Method (DPM) MC algorithm that runs on a CPU. In conclusion, the proposed CNN approach demonstrated consistently higher accuracy than VDK-density scaling and comparable accuracy versus MC and is fast enough to be used clinically.

I. Introduction

In dosimetry guided treatment planning of internal radionuclide therapy, it is important to have accurate and computationally efficient methods for voxel-level dosimetry estimation. Coupling a patient's own images with full Monte Carlo radiation transport is broadly accepted to be the gold standard for patient specific voxel-level dosimetry. However, running enough histories of MC is very computationally expensive and time consuming thus it can be impractical for the clinical use. Faster voxel dose kernel (VDK) methods such as those based on Medical Internal Radiation Committee (MIRD [1]) voxel S-values, can be inaccurate in the presence of heterogeneous tissues, e.g., at the liver-lung interface. Moreover, the accuracy of dose estimation by both MC and VDK methods can be significantly degraded by the poor SPECT camera spatial resolution.

Recently, deep learning has achieved much success in medical imaging. For example, U-Net [2] showed the state-of-the-art performance on challenging medical imaging segmentation task. In nuclear imaging, deep learning methods were also been pursued in nuclear medicine [3] including for dosimetry tasks. Lee *et al*. [4] applied a 3D U-Net with positron emission tomography (PET) and CT-based density image patches as input to accurately and efficiently predict 3D voxel-level dose-rate maps. Götz *et al*. [5] proposed a hybrid Deep Neural Network - Empirical Mode Decomposition (DNN-EMD) method that could enhance the precision and reliability of dose-rate estimations. Akhavanallaf *et al*. [6] employed a CNN to represent specific S-value kernels to predict the distribution of deposited energy; that method had comparable performance to direct MC approach. A limitation of CNN training in these prior studies [4, 5, 6] is that the ground truth dose distribution was derived from SPECT-based activity maps where the poor SPECT resolution could limit the voxel-level dose-rate accuracy. To overcome this limitation, we proposed a CNN training strategy that had the following features:

1) We adopted images from phantoms instead of patients for training so that the CNN can be trained by the ground truth without SPECT resolution effect.
2) We used patient ^{68}Ga PET scans to generate phantoms for training/testing to make simulated activity distributions that were realistic yet in higher resolution than possible with SPECT-based phantoms.

Furthermore, we applied residual learning by training the CNN to only learn the difference between the MC ground truth and the VDK dose-rate map so that the CNN model could potentially be simplified. Our proposed method is theoretically applicable to other radionuclides for internal therapy dosimetry in general; in this work, the implemented neural network was for ^{177}Lu DOTATATE therapy of neuroendocrine tumors.

Manuscript received December 20, 2020. This work was supported by R01 EB022075 awarded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and R01 CA240706 awarded by the National Cancer Institute (NCI), NIH.

Z. Li and J. A. Fessler are with Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122 (e-mail: zonyul@umich.edu, fessler@umich.edu).

J. K. Mikell is with Radiation Oncology, University of Michigan, Ann Arbor, MI 48109-5010 (e-mail: jusmikel@med.umich.edu).

S. J. Wilderman is with Department of Nuclear Engineering and Radiologic Sciences, University of Michigan, Ann Arbor, MI 48109-2104 (email: sjwnc@umich.edu).

Y. K. Dewaraja is with Department of Radiology, University of Michigan, Ann Arbor, MI 48109-5667 (e-mail: yuni@med.umich.edu).

978-1-7281-7693-2/20/$31.00 ©2020 IEEE

II. Method

A. Virtual Patient Phantom Generation & Data Set

Prior to ^{177}Lu DOTATATE therapy, patients undergo diagnostic ^{68}Ga DOTATATE PET/CT imaging to determine the eligibility for ^{177}Lu therapy. To generate phantoms for training/testing, we selected thirteen such ^{68}Ga PET images (nine were used for training and the rest for testing) from our clinic patient database. The selected PET activity maps were then registered into CT image space ($512 \times 512 \times 130$ with voxel size $0.98 \times 0.98 \times 3\text{mm}^3$) followed by extracting 130 slices that covered the field of view of SPECT camera with liver and kidney centered. Meanwhile, the corresponding CT images were converted into density maps according to the following formula that was determined from a prior calibration measurement using a phantom with 16 tissue equivalent rods:

$$\rho(HU) = \begin{cases} 0.00108 \cdot HU + 1.02351, & HU \leq 0 \\ 0.00069 \cdot HU + 1.03107, & HU > 0 \end{cases} . \quad (1)$$

Next, ^{177}Lu SPECT/CT projections with 128 views were generated using the SIMIND MC code [7] where the inputs were activity maps and density maps generated above. We simulated $\sim$2 billion photon histories to generate projections with high statistics. The projection counts were then scaled to 4 count levels (1.9e7, 5.5e6, 4.4e6, 3.0e6) that were chosen based on the counts of SPECT/CT that typically imaged at 4 time points within 7 days after therapy corresponding to one of our ^{177}Lu patients. An in-house developed OSEM SPECT reconstruction algorithm (16 iterations, 4 subsets, scatter and attenuation correction, collimator-detector response modeling) was applied after adding Poisson noise to the scaled projections. To prevent under/over-fitting, we randomly selected 20% of training slices to serve as validation dataset. In addition, a patient's SPECT/CT scan at day 0 post ^{177}Lu DOTATATE was also included in our test dataset.

B. Monte Carlo Dosimetry & Voxel Dose Kernels

To provide the ground truth for training/testing, true activity and density maps were directly input to our previously developed Dose Planning Method (DPM) Monte Carlo code [8] to generate absorbed dose-rate maps with full radiation transport simulating $\sim$1 billion histories.

To provide the VDK dose-rate maps for residual learning, soft tissue voxel kernels (beta particle kernel of size $9 \times 9 \times 9$ and photon kernel of size $99 \times 99 \times 99$, both have voxel size $0.98 \times 0.98 \times 3\text{mm}^3$) were also generated by the same DPM MC code. After convolving the VDKs with the normalized 3D SPECT activity maps, each voxel was multiplied by 1.04 (g/cm^3) and divided by CT-derived local voxel density (g/cm^3) as a simple correction to account for differences in density.

C. CNN Training

Different from [4] and [5] where a 3D U-Net was applied, we first extracted the depth features of the concatenated inputs using three 3D convolutional layers (with kernel size $7\times7\times5, 7\times7\times3, 7\times7\times3$, respectively). Next, we implemented a 2D U-Net that had 4 down-sample and up-sample layers with 16 initial filters and added the VDK dose-rate map to the 2D U-Net output to obtain the final CNN dose-rate map. Thus, the parameters of CNN were updated only according to the difference between GT dose-rate maps and VDK dose-rate maps. Fig. 1 shows our CNN model. To cover the range of ^{177}Lu beta particles (maximum CSDA range in water = 1.8mm [9]) and part of emitted photons that are low in intensity, we resized the 3D input SPECT activity/CT-derived density maps into packs containing 5 adjacent slices for each side. Thus there were two 3D input arrays of size ($512 \times 512 \times 11$ with slice thickness 3mm) and one 2D output of size (512×512) corresponding to the dose distribution in the middle slice of the input arrays. Together with phantom generation, Fig. 2 summarizes our entire workflow.

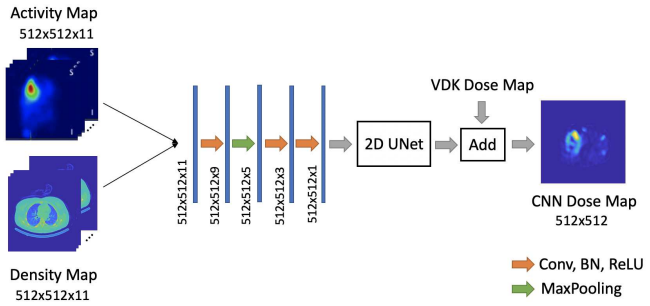

Fig. 1. Our CNN Architecture.

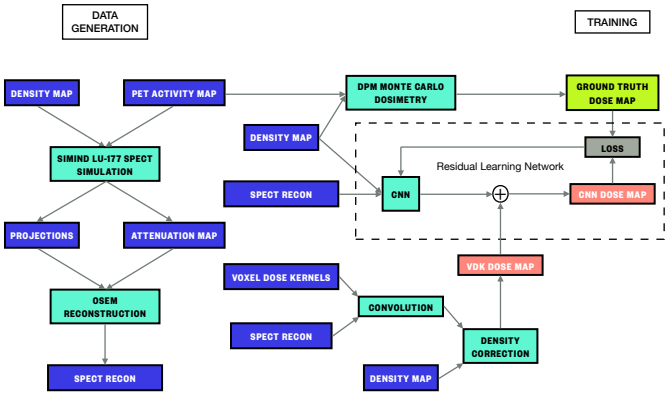

Fig. 2. Entire Workflow of our proposed method.

D. Evaluation Metrics

1) Mean dose error (MDE). For each organ/lesion, MDE is defined as the absolute error of the mean dose-rate relative to GT across that organ/lesion.
2) Normalized root mean square error (NRMSE). The NRMSE was calculated by:

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

where n_p is the total number of voxels, x is the ground truth (GT) image, $\hat{x}$ denotes the estimated image and the

subscript j indicates the jth voxel of the corresponding image.

III. Experiments

A. Implementation Details

To cover different input count levels, we normalized (divided by the sum) each SPECT activity map put into the CNN. Then we scaled the normalized SPECT along with the VDK and GT dose-rate maps with a constant value so that they have similar range with the density maps. The CNN was trained using ADAM optimizer to minimize the mean square error (MSE) with dynamic learning rate (an initial value 0.001 with ReduceOnPlateau management strategy) and batch size 32 for 200 epochs on two Nvidia Tesla V100 GPUs. The training/validation loss converged to 288/410 after 4 hours of training for 200 epochs, as demonstrated in Fig. 3.

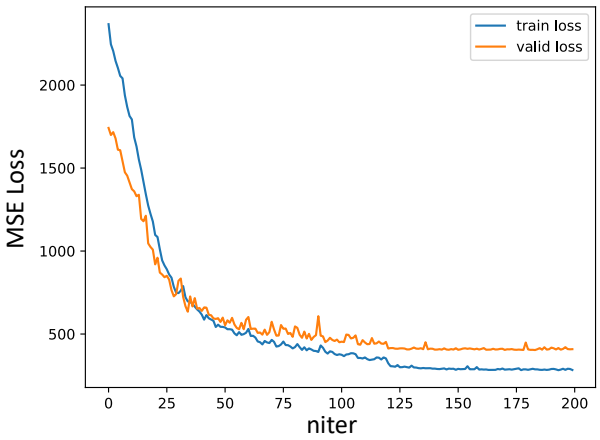

Fig. 3. Training/Validation MSE loss versus number of epochs.

B. Virtual Patient Phantom Results

Visually, there was much better agreement between GT and dose-rate maps generated by CNN than those generated by VDK with density scaling and MC as evident in Fig. 4. In Fig. 4(A), CNN revealed clearly separated small lesion and kidney. In Fig. 4(B), CNN showed improved resolution of kidney than VDK and MC relative to GT. For quantitative results, Table. I and Table. II compared the average MDE and NRMSE across all test phantoms with the corresponding range shown in the parenthesis. For all organs in I, the GT agreed better with CNN than with VDK. In particular, the error of mean dose-rate had an average of 50%, 60%, 40%, 55% and 63% improvement in healthy liver, tumor, spleen and left/right kidney, respectively. The NRMSE relative to GT was also substantially lower for CNN than for VDK, for all regions evaluated in Table. II, where CNN improved NRMSE by 7%, 23%, 27%, 15% and 11% in healthy liver, tumor, spleen and left/right kidney compared to VDK with density correction. Besides, in all of the cases except the MDE in healthy liver, CNN also showed consistent improvement compared to the MC. For example, in tumors, spleen and left/right kidney, CNN demonstrated an average of 58%, 44%, 24%, 56% improvement for MDE, and an average of 22%, 28%, 16%, 11% improvement for NRMSE, respectively.

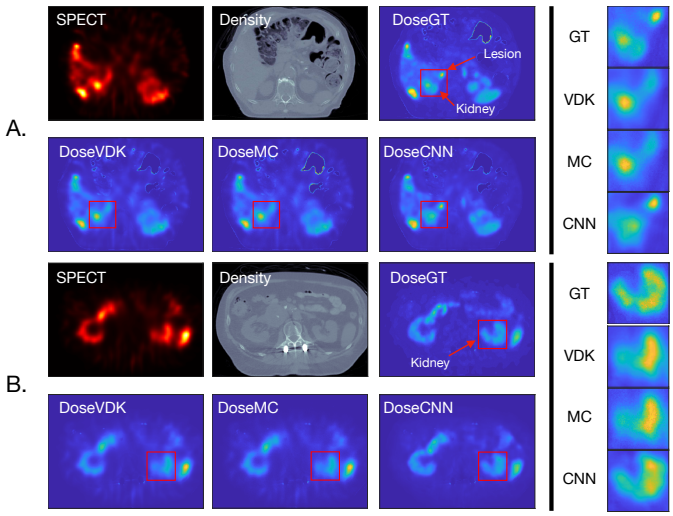

Fig. 4. VDK, MC and CNN dose-rate maps for test phantoms relative to GT.

TABLE I
Comparisons of VDK, MC and CNN in mean dose error across all test phantoms.

Organ	VDK w/ density	MC	CNN
Healthy liver	3.2% (1.9-6.4)	**1.0% (0.1-3.7)**	1.6% (0.4-2.7)
Liver	4.0% (1.8-6.4)	2.1% (0.1-4.2)	**1.9% (0.4-3.5)**
Tumor	13.7% (0.5-24.4)	13.2% (0.1-23.8)	**5.5% (0.8-13.0)**
Spleen	4.5% (2.2-6.8)	4.8% (2.8-6.6)	**2.7% (0.4-6.2)**
Left kidney	2.9% (1.9-4.2)	1.7% (0.9-2.6)	**1.3% (0.2-2.1)**
Right kidney	5.6% (1.2-9.7)	4.8% (2.8-6.6)	**2.1% (0.6-5.1)**

TABLE II
Comparisons of VDK, MC and CNN in NRMSE across all test phantoms.

Organ	VDK w/ density	MC	CNN
Healthy liver	22.3% (18.9-27.1)	22.7% (19.2-27.4)	**20.8% (17.0-26.1)**
Liver	21.6% (17.8-24.9)	21.8% (18.1-25.0)	**19.4% (15.4-22.1)**
Tumor	22.4% (16.3-32.6)	22.3% (16.5-32.3)	**17.3% (13.2-28.8)**
Spleen	17.7% (16.2-20.3)	18.0% (16.7-20.3)	**12.9% (10.7-17.7)**
Left kidney	21.5% (20.4-23.7)	21.7% (20.7-23.8)	**18.3% (15.1-21.9)**
Right kidney	21.6% (16.5-25.3)	21.7% (17.0-25.0)	**19.3% (15.4-21.5)**

C. Patient Study

Although the GT dose-rate map was unknown for patients, potential improvement could be seen in CNN dose-rate maps compared to VDK and MC dose-rate maps. Fig. 5 visualizes the first scan that was taken at day 0 after ^{177}Lu DOTATATE for a test patient, where the necrotic lesion and the small lesion at the left side of the liver showed a potentially better dose-rate recovery.

D. Time cost

Table. III shows the computation time (for image size $512\times512\times130$) of VDK, MC and CNN running on CPU/GPU. Calculating VDK dose-rate map took about 20 seconds using a single CPU processor (Intel Core i9 @2.3 GHz) and 10 seconds

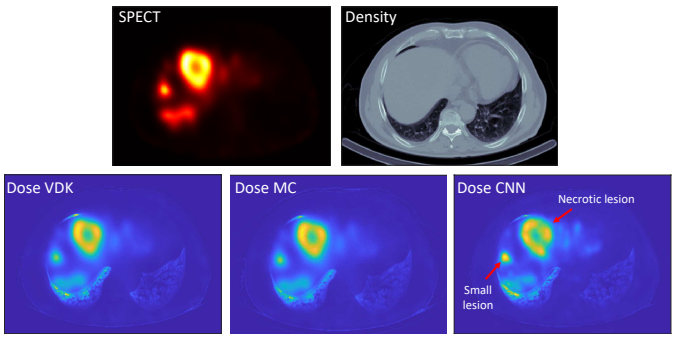

Fig. 5. CNN dose-rate maps for test patients compared to VDK and MC.

using a Tesla V100 GPU. Running DPM MC code took about 40 min for the same patient simulating 1 billion histories on a single CPU processor (Intel Xeon @3.2 GHz), while we are unaware of if any radionuclide therapy dosimetry MC code for GPU is available. Note that our in-house DPM MC code is substantially faster than general purpose MC codes typically used in internal dosimetry calculations. The CNN network dose-rate map prediction for the same patient took about 20 min using a single CPU processor (Intel Core i9 @2.3 GHz), and took 2 seconds using a Tesla V100 GPU. If considering the VDK pre-computation time for the residual learning network, the total GPU time cost for CNN will be about 12 seconds.

TABLE III
TIME COST COMPARISON AMONG VDK, MC AND CNN.

	VDK w/density	MC	CNN
CPU	∼20 sec	∼40 min	∼20 min
GPU	∼10 sec	?	∼10+2 sec

IV. Discussion

With test virtual patient phantoms that covered clinically relevant conditions, we demonstrated our CNN using residual learning framework can be applied for the fast, accurate dosimetry estimation. Despite using only moderate amount of training data, our CNN provided highly promising results across all test phantoms and a patient. Although generating the ground truth labels of the training data by running DPM MC code with full radiation transport algorithms is computationally expensive, these procedures are applied only once at training time, for a given SPECT imaging system. Results on phantoms revealed that the main limitation to the voxel-level dose-rate estimation was the poor resolution of the SPECT camera, because the theoretically accurate MC algorithm only slightly outperformed VDK, and resulted in relatively large error on test lesions/organs. However, due to the use of true-activity map based dose estimates instead of SPECT-based dose estimates as training labels, our residual CNN had the ability to partially compensate for the SPECT resolution effects. Thus, the CNN can even outperform SPECT-based MC that is considered to be the clinical gold standard for voxel-level dose-rate accuracy, as evident in Table. I and Table. II. In the patient example, although the true dose-rate distribution is unknown, potential recovery and resolution improvement can be concretely seen in Fig. 5. While our CNN provided consistently promising results in terms of MDE and NRMSE, more evaluations such as joint histograms, dose volume histograms are needed. Besides, a larger set of phantoms and patients should also be used to further test the CNN performance.

V. Conclusion

In this study, we constructed a residual CNN that was trained on phantoms derived from a set of patient ^{68}Ga DOTATATE PET/CT scans to learn the mapping from SPECT emission and density maps to the corresponding dose-rate distributions in ^{177}Lu DOTATATE radionuclide therapy. The performance of the proposed method, evaluated by MDE and NRMSE, was consistently superior to VDK with density correction and mostly better than MC. In patient studies, there was also potential resolution improvement. The residual CNN possesses much promise for real-time clinical use because of the time efficiency (∼12 sec on GPU, compared to ∼40 min of for MC using our relatively fast DPM code) while achieving high voxel-level dose-rate accuracy by potentially reducing the degradation caused by the poor SPECT resolution. Further works include tuning the CNN to further optimize the time-performance trade off, evaluating noise, expanding the training/testing data set and extending to other radionuclides used in therapy. Testing on a larger set of phantoms and patient studies is essentially prior to considering clinical use of the proposed CNN approach.

References

[1] W. E. Bolch, L. G. Bouchet, J. S. Robertson, et al. Mird pamphlet no. 17: the dosimetry of nonuniform activity distributions—radionuclide s-values at the voxel level. *Nucl. Med.*, 40:11S–36S, 1999.
[2] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, volume 9351 of *LNCS*, pages 234–241, 2015.
[3] H. Choi. Deep learning in nuclear medicine and molecular imaging: Current perspectives and future directions. *Nucl. Med. Mol. Imaging*, 52(2):109–118, 2018.
[4] M. S. Lee, D. Hwang, J. H. Kim, and J. S. Lee. Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry. *Sci. Rep.*, 9:10308, 2019.
[5] Th I Götz, C. Schmidkonz, S. Chen, S. AI-Baddai, and T. Kuwert. A deep learning approach to radiation dose estimation. *Phys. in Med. & Bio.*, 65(3):035007, 2020.
[6] A. Akhavanallaf, I. Shiri, H. Arabi, and H. Zaidi. Whole-body voxel-based internal dosimetry using deep learning. *Eur. J. Nuc. Med. Mol. Im.*, 2020.
[7] M. Ljungberg. The simind monte carlo program, 2012.
[8] S. J. Wilderman and Y. K. Dewaraja. Method for fast ct/spect-based 3d monte carlo absorbed dose computations in internal emitter therapy. *IEEE Trans. Nucl. Sci.*, 54(1):146–151, 2007.
[9] M. J. Berger, M. Inokuti, H. H. Anderson, et al. Stopping power for electrons and positrons. *Journal of the IRCU*, os19, 1984.