Friedrich-Alexander-Universität
Technische Fakultät

Deep Learning-Based Partial Volume Correction in Lu-177 SPECT/CT Imaging

Research Internship Project Report

at the Friedrich-Alexander-Universität Erlangen-Nürnberg
Pattern Recognition Lab

Principal Supervisor:	Prof. Dr.-Ing. habil. Andreas Maier
Associate Supervisor:	Merlin Nau, Andreas Grings, Hinz Ravn
Presented by:	Boncy Elias boncy.elias@fau.de 23211736
Submission:	3rd July 2026

Abstract

This research internship investigated whether a deep learning (DL) model can compensate for partial volume effects in Lu-177 single-photon emission computed tomography (SPECT) images. The project focused on a reproducible pipeline for training and evaluating residual 2.5D U-shaped convolutional neural network (U-Net) models that predict corrected activity distributions from low-resolution reconstructions. A central part of the work was lesion-level evaluation, because clinically relevant performance depends not only on image similarity but also on recovered lesion activity.

The implemented workflow includes patient-level train/validation/test splitting, lesion-aware sampling, residual prediction, activity-weighted losses, final hold-out evaluation, and analysis of lesion activity recovery. A separate resampling study quantified the information loss introduced when 512 matrix lesion segmentations are represented in 128 matrix space. On the calibrated final test cohort, the model increased the mean lesion recovery coefficient (RC) from 0.363 for the forward-projected reconstruction to 0.586, and reduced mean relative lesion activity error from 0.679 to 0.542 across 388 lesions. The results show that deep learning-based partial volume correction (PVC) can recover a substantial part of the missing lesion activity, while also revealing remaining challenges in calibration, small-lesion stability, and timepoint-dependent performance.

Contents

1 Introduction

Quantitative Lu-177 SPECT/computed tomography (CT) is important for patient-specific dosimetry in radiopharmaceutical therapy. A major limitation is the partial volume effect (PVE), which arises from the limited spatial resolution of the imaging system. Activity originating from a structure is blurred into the surrounding tissue, while activity from neighbouring regions may also spill into the structure. The PVE affects both lesions and normal organs but is particularly pronounced in small structures whose dimensions are comparable to or smaller than the spatial resolution of the SPECT system. Consequently, the activity concentration in small lesions or organs can be strongly underestimated. This affects absorbed dose estimation, treatment planning, and response assessment [1, 2].

The spatial resolution of a SPECT system is governed by the photon energy of the emitting isotope, the collimator geometry, detector response, reconstruction method, and object-to-collimator distance. For parallel-hole collimator imaging, this leads to a depth-dependent point-spread function, so activity recovery varies with the size, shape, and position of the activity distribution [3]. Classical PVC methods often require anatomical masks, explicit assumptions about the point-spread function, or access to reconstruction details that are not always available in routine clinical workflows.

Deep learning offers an alternative by learning a correction mapping from degraded reconstructions to target activity distributions. The research question of this report was based on previous work by the research group, including a patient-derived dataset curated at the Department of Nuclear Medicine, University Hospital Erlangen, and on the method proposed by Leube et al. [4–6]. These studies motivate PVC models trained on realistic anatomy and synthetic lymph-node metastases, with the aim of improving quantitative Lu-177 SPECT/CT without requiring manual lesion segmentation at inference time.

1.1 Aim and Scope

The aim was to develop and evaluate a reproducible deep learning pipeline for Lu-177 SPECT/CT PVC. The central question was whether a model trained on paired forward-projected reconstructions and target activity maps can improve lesion-level activity recovery compared with the uncorrected reconstruction.

The work focused on three concrete contributions. First, data loading, training, checkpointing, and evaluation scripts were implemented or extended for systematic model development and assessment. Second, lesion-level activity recovery was evaluated

using probabilistic lesion masks in 128 matrix space. Third, an additional analysis investigated the effect of downsampling the activity maps and corresponding segmentations from 512 to 128 matrix resolution on lesion representation.

2 Materials and Methods

2.1 Data and Preprocessing

The study used preprocessed Lu-177 SPECT/CT activity data from patient-derived digital phantoms. The broader project context follows the idea of using realistic patient anatomy and synthetic lymph-node metastases to train PVC models on paired degraded and target activity distributions [5]. Synthetic activity maps and corresponding segmentations were available at the original high-resolution matrix size and were downsampled to the lower matrix size used for network training because the reconstructed SPECT data were represented in 128 matrix space.

The forward model generated projection data from the activity maps using a rotation-based SPECT simulation and reconstruction framework. Poisson noise was applied to the projection data before reconstruction, resulting in degraded activity images that served as model inputs. Three anatomical regions were considered: abdomen, pelvis, and thorax. Five timepoints were included where available: 0h, 4h, 24h, 48h, and 72h.

Two lesion representations were used. The original 512 matrix lesion masks were treated as binary three-dimensional volume segmentations. After resampling to 128 matrix space, masks were treated as probabilistic masks with values between 0 and 1. This avoids losing boundary information from small lesions. Lesion volume was therefore calculated as a weighted physical volume rather than by hard thresholding.

2.2 Pipeline and model

The implemented pipeline loads paired reconstructed and target activity volumes, samples training slices, trains a neural network, stores checkpoints, evaluates image-level and lesion-level metrics, and aggregates the resulting tables and figures. Phantom fabrication, acquisition settings, and point-spread-function estimation used in the broader project are described in the project material [5].

A 3D U-Net baseline was considered first. A 3D U-Net processes volumetric input and can use through-plane context directly, but it also has substantially higher memory demand. In the available experiments, the initial 3D results were not visually suitable for lesion-level correction, so the final experiment used a residual 2.5D U-Net. The 2.5D input consists of five neighboring slices, which preserves limited through-plane context while keeping the training problem smaller than full-volume 3D learning.

Instead of predicting a complete corrected image, the residual model predicted a correction term that was added to the input reconstruction. This formulation was chosen because the reconstruction already contains the global activity structure, while the network should mainly learn a local correction for partial-volume underestimation. Training used lesion-aware sampling, AdamW optimization with learning rate 10^{-4}, batch size 32, max normalization, mixed precision where available, and early stopping with patience four. The loss combined image-level error, lesion-weighted voxel error, lesion activity error, global activity error, and residual regularization. This combination was used to balance global image fidelity with the primary endpoint of lesion activity recovery and to discourage unstable residual corrections. The best checkpoint was selected using validation lesion RC error.

2.3 Evaluation

The main endpoint was lesion activity recovery. For each lesion, the recovery coefficient was computed as

$$\mathrm{RC} = \frac{A_{\mathrm{estimate}}}{A_{\mathrm{GT}}}, \tag{2.1}$$

where A_{estimate} is the activity from the reconstruction or model prediction, and A_{GT} is the target activity. A value of one indicates perfect recovery. Absolute and relative lesion activity errors were also computed. For the final test evaluation, a calibrated residual output was used to reduce overcorrection observed in the uncalibrated residual model. The validation-based calibration scaled the residual by timepoint, with a strong reduction at 0h and nearly full correction at later timepoints.

3 Results and Discussion

3.1 Training and Resampling Results

The final residual 2.5D training run selected its best checkpoint after the first epoch, with a validation lesion RC error of 0.545. Although the maximum number of epochs was 20, early stopping ended the run after five epochs because later epochs did not improve the validation selection metric. The fact that the best validation performance was reached after the first epoch and did not improve during subsequent epochs suggests that the current training configuration had limited capacity for further optimization. This may indicate that the selected hyperparameters, loss weighting, sampling strategy, or model configuration were not yet optimally tuned for the lesion-level recovery objective. On the calibrated final test cohort, the residual 2.5D model achieved an overall MAE of 13023.1, RMSE of 41107.8, and PSNR of 11.77 dB. The forward-projected reconstruction underestimated lesion activity, with a mean RC of 0.363. The calibrated deep learning prediction increased the mean RC to 0.586 and improved 296 of 388 lesions, corresponding to 76.3%.

The resampling analysis included 3517 lesions. Mean lesion volume was similar between the 512 and 128 matrix representations, with 1.820 ml and 1.805 ml, respectively. This close agreement indicates that the probabilistic 128 matrix masks largely preserved lesion volume during resampling. The effect of resampling on integrated lesion activity requires separate evaluation with consistent treatment of image scaling and voxel-volume differences between both matrix representations.

Region	Lesions	Volume 512 ml	Volume 128 ml	Activity loss %
Abdomen	1255	1.717	1.707	94.10
Pelvis	1998	1.975	1.955	93.98
Thorax	264	1.143	1.136	94.41
Overall	3517	1.820	1.805	94.06

Table 1 Mean lesion volume and activity loss after resampling from 512 to 128 matrix space.

3.2 Final Test Performance

On the calibrated final test cohort, the residual 2.5D model achieved an overall MAE of 13023.1, RMSE of 41107.8, and PSNR of 11.77 dB. The forward-projected reconstruction underestimated lesion activity, with a mean RC of 0.363. The calibrated deep learning prediction increased the mean RC to 0.586 and improved 296 of the 388 evaluated lesions, corresponding to 76.3%. Mean relative error decreased from 0.679 to 0.542, and mean absolute activity error decreased from 6.24×10^6 to 4.44×10^6.

The strongest regional improvement was observed in the pelvis, where mean RC increased from 0.402 to 0.639 and 85.4% of lesions improved. Abdomen and thorax also improved, although thorax had fewer lesions and should be interpreted more cautiously. Timepoint analysis showed why calibration was needed: the uncalibrated residual model boosted 0h lesions too strongly, increasing mean 0h RC to 1.887. Calibration reduced this value to 0.964 while preserving the stronger corrections at later timepoints.

Region	Lesions	RC recon	RC prediction	Improved %
Abdomen	139	0.297	0.486	65.5
Pelvis	212	0.402	0.639	85.4
Thorax	37	0.382	0.657	64.9
Overall	388	0.363	0.586	76.3

Table 2 Mean lesion recovery on the calibrated final test cohort.

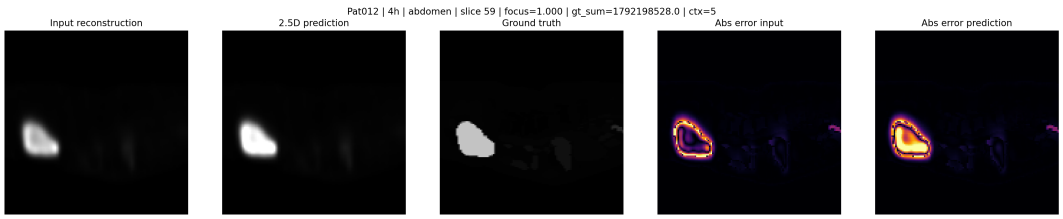

Figure 1 Example from the final test cohort.

3.3 Discussion and Outlook

The results show that the implemented DL-based PVC pipeline can recover a meaningful fraction of lesion activity that is missing in the forward-projected reconstruction. This agrees with the local project motivation of training on patient-derived phantoms and synthetic lymph-node metastases, and with Leube et al. [5, 6], who showed that U-Net-based PVC can improve quantitative Lu-177 SPECT/CT. The work also meets the main goals of the proposal: a reproducible training and evaluation workflow was created and preliminary quantitative results were generated.

Despite the overall improvement, further optimization is still needed. The mean lesion RC after correction was 0.586, which remains below the ideal value of 1 and therefore indicates persistent under-recovery on average. Regional differences were also observed, with mean RC values of 0.486 in the abdomen, 0.639 in the pelvis, and 0.657 in the thorax. In addition, 92 of the 388 evaluated lesions did not improve relative to the forward-projected reconstruction. Timepoint-dependent behavior remained another limitation, as the uncalibrated model strongly overcorrected 0h lesions, increasing their mean RC to 1.887 before calibration. These findings show that the current model does not yet provide consistently accurate recovery across lesions, anatomical regions, and timepoints.

4 Conclusion

This research internship produced a working pipeline for deep learning-based PVC in Lu-177 SPECT/CT. The final calibrated residual 2.5D U-Net improved 76.3% of final-test lesions, increased mean lesion RC from 0.363 to 0.586, and reduced mean relative lesion activity error from 0.679 to 0.542. These findings demonstrate the potential of deep learning-based correction to improve lesion-level activity recovery, while also highlighting the need for further work on robust model training, calibration, and validation across lesions, anatomical regions, and timepoints.

References

[1] M. Soret, S. L. Bacharach, and I. Buvat, “Partial-volume effect in pet tumor imaging”, *Journal of Nuclear Medicine*, vol. 48, no. 6, pp. 932–945, 2007.

[2] K. Erlandsson, I. Buvat, P. H. Pretorius, B. A. Thomas, and B. F. Hutton, “A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology”, *Physics in Medicine and Biology*, vol. 57, no. 21, R119–R159, 2012.

[3] J. Tran-Gia and M. Lassmann, “Characterization of noise and resolution for quantitative 177lu spect/ct with xspect quant”, *Journal of Nuclear Medicine*, vol. 60, no. 1, pp. 50–59, 2019.

[4] J. R. Hinz, T. Kuwert, M. Beck, P. Ritt, and A. Grings, “Estimation and correction of the partial volume effect using personalized phantoms of lymph node metastases in lutetium 177 spect/ct dosimetry”, *Molecular Imaging and Biology*, 2026.

[5] J. R. Hinz, M. Nau, T. Kuwert, M. Beck, and A. Grings, “Deep learning for partial volume correction in 177lu spect using patient-derived phantoms and synthetic lymph node metastasis”, by the research group, 2026.

[6] J. Leube, J. Gustafsson, M. Lassmann, M. Salas-Ramirez, and J. Tran-Gia, “A deep-learning-based partial-volume correction method for quantitative 177lu spect/ct imaging”, *Journal of Nuclear Medicine*, vol. 65, no. 6, pp. 980–987, 2024.