3D PET-DIP Reconstruction with Relative Difference Prior Using a SIRF-Based Objective

Imraj RD. Singh[1,2], Riccardo Barbano[1], Željko Kereta[1], Bangti Jin[1], Kris Thielemans[2], and Simon Arridge[1]

[1]Department of Computer Science, UCL, UK
[2]Institute of Nuclear Medicine & Centre for Medical Image Computing, UCL, UK

Abstract: Deep Image Prior (DIP) is an unsupervised deep learning technique that does not require ground truth images. For the first time, 3D PET reconstruction with DIP is cast as a single optimisation via penalised maximum likelihood estimation, with a log-likelihood data-fit and an optional Relative Difference Prior term. Experimental results show that although unpenalised DIP optimisation trajectory performs well in high count data, it can fail to adequately resolve lesions in lower count settings. Introducing the Relative Difference Prior into the objective function the DIP trajectory can yield notable improvements.

1 Introduction

Deep Image Prior (DIP) [1] is a state-of-the-art unsupervised deep learning method for image reconstruction. It leverages the inductive bias of Convolutional Neural Networks (CNNs) to fit to natural signals faster than to noise, allowing regularisation via early stopping along the optimisation trajectory.

Gong et al. [2] were the first to apply DIP to PET reconstruction by splitting the reconstruction into Expectation Maximisation (EM) and DIP denoising. Splitting the optimisation was necessary as the PET forward model was not integrated into a deep learning framework. This was subsequently done by Hashimoto et al. [3] and DIP was implemented as a single optimisation problem, thus reducing the number of hyperparameters and the computational overhead, and simplifying implementation. But, the forward model was stored as a sparse matrix which had an excessive GPU memory overhead. Furthermore, mean-squared-error was used as the data-fidelity. Their work was recently extended to 3D PET [4] through slicing the forward operator and solving with a subset-based block iterative approach.

In this work we implement DIP for 3D PET as a single optimisation problem with log-likelihood data-fit and an optional penalisation term. Our implementation uses the wrapper developed in [5]. Projectors utilised are implicit, thus alleviating the large GPU overhead associated with explicit projection and allowing full gradient updates for 3D PET data. The implementation is tested on realistic simulated data with two count levels. Results are compared to solutions from the provably convergent Block-Sequential Regularised Expectation Maximisation (BSREM) algorithm.

2 Preliminaries

2.1 Penalised Maximum Likelihood

Penalised maximum likelihood methods for PET image reconstruction aim to solve the following optimisation problem:

$$\text{argmin}_{\mathbf{x} \geq 0}\{\Phi(\mathbf{x}) = -L(\mathbf{y}|\mathbf{x}) + \beta R(\mathbf{x})\}, \qquad (1)$$

where $L(\mathbf{y}|\mathbf{x})$ is the Poisson log-likelihood describing the goodness of fit of the reconstructed image $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$ to the measurements $\mathbf{y} \in \mathbb{R}_{\geq 0}^M$; M and N denote the number of projection bins and image voxels, respectively; $R(\mathbf{x})$ is the penalty, and $\beta > 0$ balances the data-fit and penalty. Up to an additive constant, the Poisson log-likelihood is given by: $L(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{M} y_i \log(\bar{y}_i(\mathbf{x})) - \bar{y}_i(\mathbf{x})$. The mean of the measurements $\bar{\mathbf{y}}$ is obtained by projecting the reconstructed image with an affine PET forward model, defined by $\bar{\mathbf{y}}(\mathbf{x}) = \mathbb{E}[\mathbf{y}] = \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}$. The system matrix $\mathbf{A}$ models the PET scanner characteristics as well as physical phenomena, e.g., attenuation and positron range. The expected background events $\bar{\mathbf{b}}$ include both scatter and randoms.

In this work we consider the Relative Difference Prior (RDP) [6], defined by: $R(\mathbf{x}) = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} w_{ij} \frac{(x_i - x_j)^2}{x_i + x_j + \gamma|x_i - x_j|}$, where $\mathcal{N}_i$ is a $3 \times 3 \times 3$ neighbourhood of the i-th image voxel, w_{ij} are the neighbourhood weights. The edge preservation parameter is set as $\gamma = 2$, as is standard in a clinical setting.

2.2 Block-Sequential Regularised Expectation Maximisation

BSREM [7] is a provably convergent subset algorithm for PET image reconstruction, with an iterative update given by:

$$\mathbf{x}_{k+1} = P_{\mathbf{x} \geq 0}\left[\mathbf{x}_k - \alpha_{k,n} D(\mathbf{x}_{k,n}) \nabla \Phi_{m_k}(\mathbf{x}_k)\right], \quad k \geq 0. \qquad (2)$$

Here Φ_{m_k} is a subset gradient and $m_k \in \{1, \ldots, n_{\text{subsets}}\}$ is the index of a subset chosen at image update k, out of n_{subsets} subsets. The epoch number $n \geq 0$ is incremented after every n_{subsets} image updates. The step-size is given by α_n, preconditioner by $D(\cdot)$, and $P_{\mathbf{x} \geq 0}[\cdot]$ is a non-negativity projection. The EM preconditioner is used; $D(\mathbf{x}_{k,n}) = \text{diag}\left\{(\mathbf{x}_{k,n} + \delta)/\mathbf{A}^\top \mathbf{1}\right\}$, where $\delta = 1\text{e-}9$ ensures positive definiteness, $\mathbf{A}^\top \mathbf{1}$ is the sensitivity image, and $\mathbf{x}_{k,n}$ is the reconstruction at epoch n. The step-size is computed with $\alpha_{k,n} = \alpha_0/(\eta n + 1)$, where $\alpha_0 = 1$ is the initial step size and η is a relaxation coefficient.

2.3 Deep Image Prior

DIP [1] represents $\mathbf{x}$ through learnable parameters $\theta \in \mathbb{R}^p$ of a CNN $\mathbf{f}(\mathbf{z}; \theta)$ with a fixed random input $\mathbf{z}$. The optimisation problem (1) is then recast as: $\theta^\star \in \text{argmin}_{\theta \in \mathbb{R}^p} \Phi(\mathbf{f}(\mathbf{z}; \theta))$, where the reconstructed image is obtained from $\mathbf{x}^\star = \mathbf{f}(\mathbf{z}; \theta^\star)$. In the original DIP work [1, 2] the objective function Φ consists solely of the likelihood term where the regularisation is imparted through network architecture and stopping criteria. A penalisation can be included to alleviate the lack of robust stopping criteria, which is critical to prevent overfitting to noise. The utility of additional penalisation was

first investigated for CT [8] and was included for PET in [4], although the latter did not compare against traditional penalised maximum likelihood solutions.

3 Methods

3.1 Wrapping the SIRF-Objective

SIRF is a multi-modality synergistic reconstruction framework providing access to several well-established reconstruction engines. For advanced PET and SPECT reconstruction the Software for Tomographic Image Reconstruction (STIR) engine is used [9]. In this work we utilise various features of STIR through SIRF such as the parallelised C++ backend, access to 3D GPU-based projectors, and access to clinically relevant PET penalties (e.g. RDP).

The wrapper integrates SIRF into PyTorch, via exposure of `sirf.STIR.ObjectiveFunction.value` and `sirf.STIR.ObjectiveFunction.get_gradient` methods in a custom autograd function that subclasses `torch.autograd.Function`.

3.2 Synthetic Data Generation and System Modelling

A Monte-Carlo photon emission simulation of a voxelised XCAT torso [10] with GE Discovery 690 scanner modelled acquisition was performed using OpenGATE [11], STIR [9] and STIR-GATE-Connection [12]. The distribution of back-to-back 511 keV photon emissions is representative of activity concentrations from a ^{18}F-FDG tracer study. Cylindrical hot lesions of dimensions 1cm diameter by 1cm length were inserted into the abdominal wall (1.6:1), liver (1.3:1), lung (2:1), and spine (1.6:1). The lesion to associated background ratio is indicated by (lesion:background).

Projection data sets containing 250 ("lower") and 1200 ("higher") million coincidence events were acquired. A true-to-background ratio of 0.93:1 was maintained for the datasets. The resulting list-mode data were re-binned into sinograms with 288 projection angles; all ring differences were used as is typical in clinical practice. The reconstruction volume had dimensions of $47 \times 128 \times 128$ with voxel-size $3.27 \times 4.0 \times 4.0$ mm. Normalisation, randoms and scatter were estimated from the Monte-Carlo data and incorporated within the forward model, see [12] for details.

3.3 BSREM and DIP Implementation

For BSREM the objective $\Phi(\mathbf{x})$ was split into 32 ordered subsets, accessed in accordance to the Herman-Meyer order [13]. The initial image was set as a reconstruction with ordered subset EM with 24 subsets after 1 epoch. The maximum epoch n_{epochs} and relaxation coefficient η were found through a grid search for both datasets: for higher count data $\eta = 0.02$ and $n_{\text{epochs}} = 1000$, for lower count $\eta = 0.04$ and $n_{\text{epochs}} = 500$. The grid search was assessed to ensure fast convergence and small step-sizes at n_{epochs}.

For DIP a three-scale 3D U-Net [14] was implemented in PyTorch (1.13.0), in-line with previous work [2, 4]. Trilinear upsampling and strided convolutions were used to change scale, with the number of features compensating for the increase/decrease of spatial dimensionality. Batch normalisa-

tion and Leaky ReLU were included after each convolution. Skip connections were also present between encoding and decoding paths of the network. ReLU was used on the network output as a non-negativity constraint. An ADAM optimiser was used for training, with initial learning rate of 1.0 and cosine annealing tending to 0 over 20,000 iterations. Two configurations of DIP were implemented; DIP with only Poisson negative log-likelihood objective (referred to as"DIP"), and with RDP in the objective ("DIP+RDP").

3.4 Quality Metrics

Standard metrics for quantification and detectability of lesions are used to assess image quality. Contrast Recovery Coefficient (CRC) values are calculated between the lesions and associated background Regions of Interest (ROIs) by: $\text{CRC} = \left(\frac{\bar{a}}{\bar{b}} - 1\right) / \left(\frac{a_{\text{t}}}{b_{\text{t}}} - 1\right)$, where $\bar{a}$ and $\bar{b}$ are average emissions over lesion and associated background ROIs, respectively. The subscript t denotes ground truth emission values. Standard Deviation (STDEV) was calculated on each of the background ROIs according to: $\text{STDEV} = \left(N_{\text{ROI}}^{-1} \sum (b_i - \bar{b})^2\right)^{1/2}$.

4 Results and Discussion

For BSREM results, a set of eleven regularisation values β were used for each count level. For higher count $\beta \in [3.125, 31.25] \cdot \text{e-3}$; for lower count $\beta \in [1.5, 15] \cdot \text{e-2}$. The largest and smallest values in the range represent over-penalised and under-penalised solutions respectively. DIP results are shown at different epochs. For DIP+RDP, four regularisation values β were used for each count level: higher count $\beta \in \{3.125, 12.5, 21.875, 31.25\} \cdot \text{e-3}$; lower count $\beta \in \{1.5, 6.0, 10.5, 15.0\} \cdot \text{e-2}$. DIP+RDP worked best when β was lowest (but non-zero), results are shown in Figs. 1c and 1d. Qualitative visual comparisons of the lower count reconstructions are given in Fig. 2.

In the higher count regime, see Fig. 1a, unpenalised DIP is able to considerably out-perform BSREM across all ROIs. In the lower count data this is not the case, see Fig. 1b. Through the inclusion of RDP in the objective function, the trajectory of DIP is improved significantly such that improved image quality metrics are observed in both lower and higher count data. However, the improvement is not consistent across all lesions. From Figs. 1a and 1c, the CRC of the abdominal wall lesion decreased markedly with the inclusion of RDP for high count data. This could be due to the abdominal wall lesions' location at the edge of the axial field-of-view, where noise is higher as sensitivity is lower. These issues of lesion dependence on local sensitivity, contrast and surrounding activity have been observed and investigated with non-DIP reconstruction [15]. Extension of such work to DIP remains for the future.

A single NVIDIA RTX 3090 with 24GB of dedicated memory (VRAM) was used in this study. PARALLELPROJ [16] was used for the projection operator, both the forward and adjoint are implemented in CUDA (GPU-specific language). One full gradient 3D PET DIP iteration took ≈ 2.4s, therefore 13.3 hours for the 20,000 iterations. This included
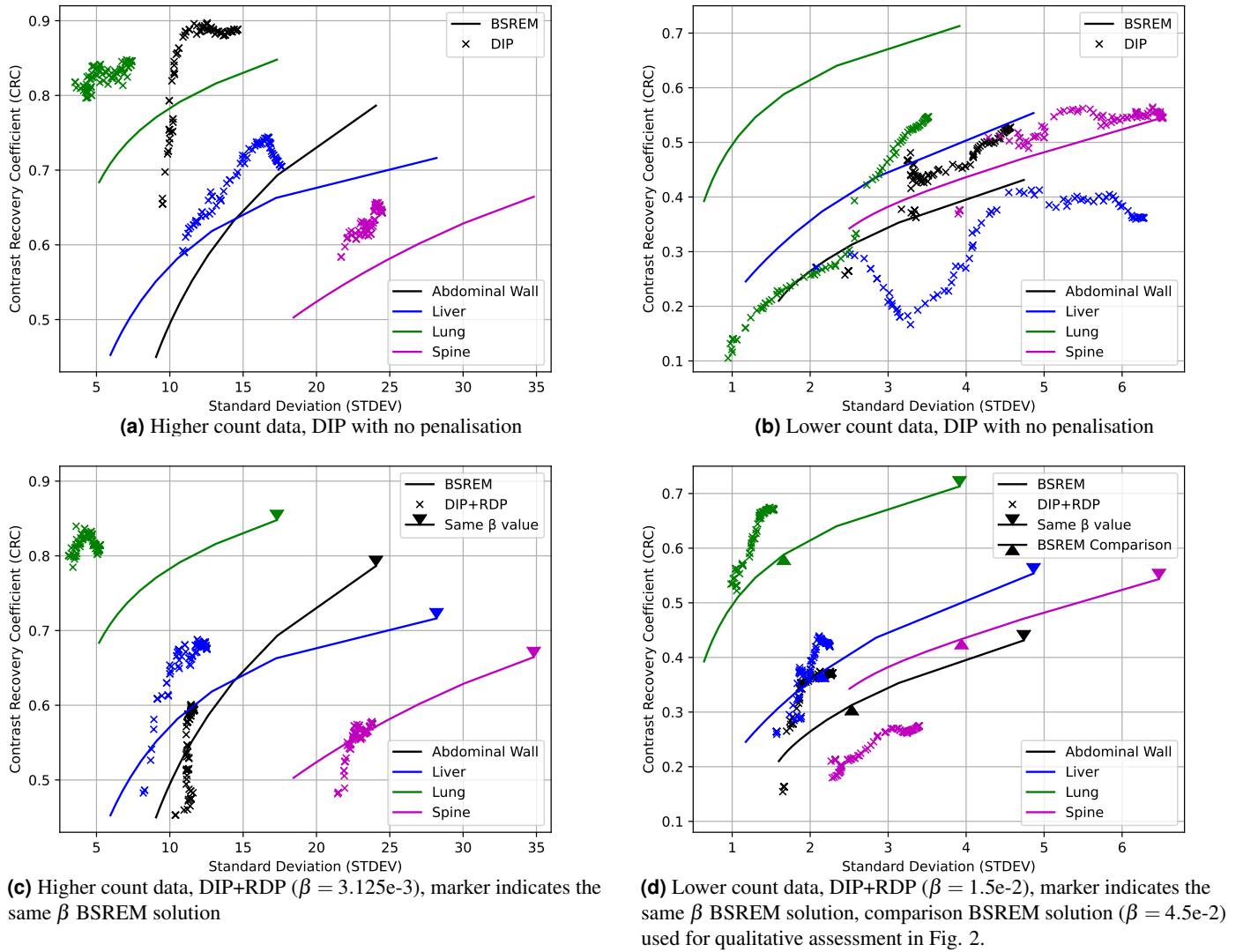
(a) Higher count data, DIP with no penalisation

(b) Lower count data, DIP with no penalisation

(c) Higher count data, DIP+RDP ($\beta = 3.125e\text{-}3$), marker indicates the same β BSREM solution

(d) Lower count data, DIP+RDP ($\beta = 1.5e\text{-}2$), marker indicates the same β BSREM solution, comparison BSREM solution ($\beta = 4.5e\text{-}2$) used for qualitative assessment in Fig. 2.

Figure 1: Contrast Recovery Coefficient between lesion and background regions of interest against the standard deviation of the background. Results closer to the top left are best. Solid lines correspond to the BSREM solution with relative difference prior with different penalty strengths. Cross markers represent the minimum-loss DIP solution (at fixed penalty strength) taken every 100 epochs after an initial 10,000 epochs up to 20,000 epochs.

Table 1: GPU memory requirements on tested data for explicit vs. our implicit projector; estimated from sinogram/image sizes as well as 8-byte sparse element-size, and observing GPU memory usage respectively. Memory requirements for the 3D U-Net (see Sect. 3.3) used in forward and backward modes, and maximum image volume allowable on a 24 GB GPU.

Projector		3D U-Net		
Explicit matrix	Implicit (ours)	Forward	Backward	Maximum Volume
$> 100\,\text{GB}$	$< 1\,\text{GB}$	0.65 GB	0.88 GB	300^3

costly copying to-and-from the GPU which is currently necessary for integration with SIRF. The wrapper could be developed further by interfacing directly with the projector through a CUDA-based PyTorch wrapper which would keep operations on the GPU and arrays saved in VRAM; speeding up computation. Run-time could also be reduced by the use of subsets in DIP+RDP. This will be pursued in the future as it would be an important step in developing efficient deep learning techniques for PET reconstruction.

5   Conclusion

This is the first single optimisation implementation of 3D PET reconstruction via penalised maximum likelihood with DIP. The implementation utilises a wrapper integrating a well-established reconstruction framework (SIRF) with PyTorch. The application of DIP on high count data was able to significantly increase quality metrics, whereas on lower count data this was not observed. Introducing RDP into the objective function significantly improved the DIP trajectory for lower count data. Results indicate that further investigation is needed as conclusive and consistent improvements are not observed across count levels and lesions.
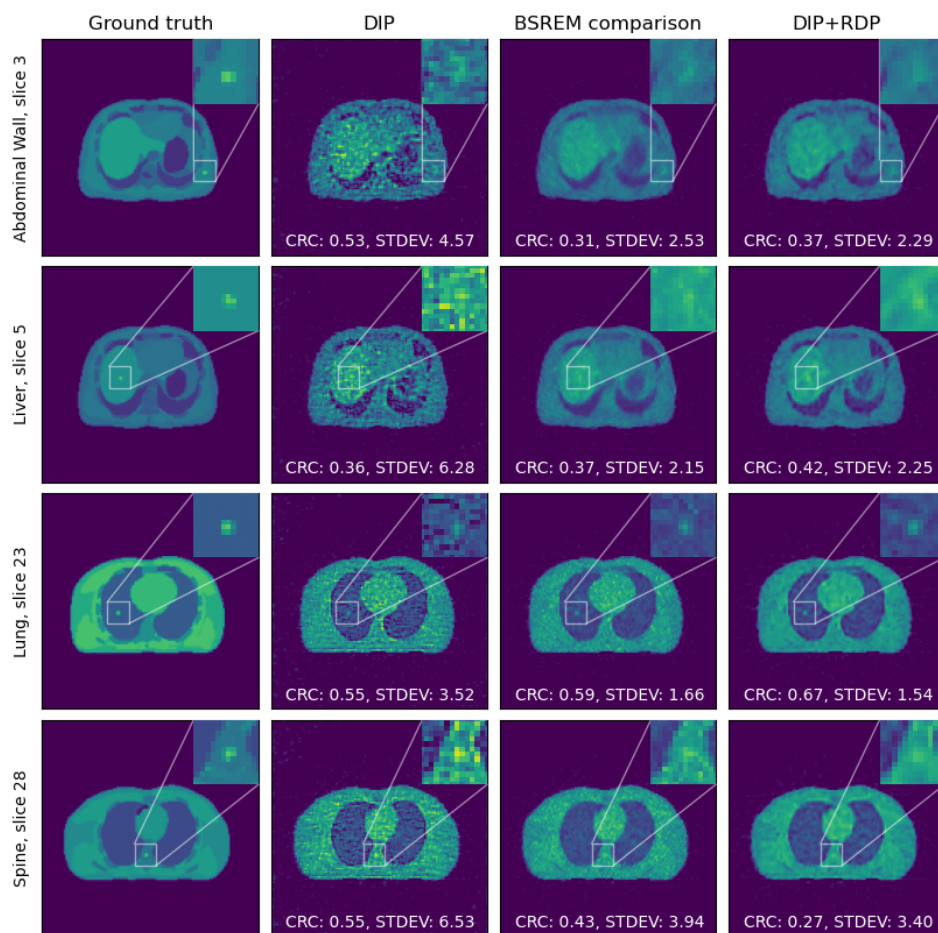
Figure 2: Axial slices taken through the center of lesions. Slices of ground truth emission and lower count data reconstructions with DIP, BSREM ($\beta = 4.5e - 2$), and DIP+RDP ($\beta = 1.5e - 2$). DIP reconstructions are the minimum-loss solutions over 20,000 epochs. CRC and STDEV values quotes are for the lesion shown in the slice. Colour-scales between reconstructed image slices are kept constant.

References

[1] D. Ulyanov et al. "Deep Image Prior". *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2018, pp. 9446–9454.

[2] K. Gong et al. "PET Image Reconstruction Using Deep Image Prior". *IEEE Trans. Med. Imag.* 38.7 (2018), pp. 1655–1665. DOI: 10.1109/tmi.2018.2888491.

[3] F. Hashimoto et al. "PET Image Reconstruction Incorporating Deep Image Prior and a Forward Projection Model". *IEEE Trans. Radiat. Plasma Med. Sci.* 6.8 (2022), pp. 841–846. DOI: 10.1109/trpms.2022.3161569.

[4] F. Hashimoto et al. "Fully 3D Implementation of the End-to-end Deep Image Prior-based PET Image Reconstruction Using Block Iterative Algorithm" (2022). DOI: 10.48550/ARXIV.2212.11844.

[5] I. R. D. Singh et al. "Deep Image Prior PET Reconstruction using a SIRF-Based Objective". *IEEE Nuclear Sci. Symp. and Med. Imag.* (2022).

[6] J. Nuyts et al. "A Concave Prior Penalizing Relative Differences for Maximum-A-Posteriori Reconstruction in Emission Tomography". *IEEE Trans. Nucl. Sci.* 49.1 (2002), pp. 56–60. DOI: 10.1109/tns.2002.998681.

[7] A. D. Pierro and M. Yamagishi. "Fast EM-Like Methods for Maximum "A Posteriori" Estimates in Emission Tomography". *IEEE Trans. Med. Imag.* 20.4 (2001), pp. 280–288. DOI: 10.1109/42.921477.

[8] D. O. Baguer et al. "Computed Tomography Reconstruction Using Deep Image Prior and Learned Reconstruction Methods". *Inverse Probl.* 36.9 (2020), pp. 1–24. DOI: 10.1088/1361-6420/aba415.

[9] K. Thielemans et al. "STIR: Software for Tomographic Image Reconstruction Release 2". *Phys. Med. Biol.* 57.4 (2012), pp. 867–883. DOI: 10.1088/0031-9155/57/4/867.

[10] W. P. Segars et al. "4D XCAT Phantom for Multimodality Imaging Research". *Med. Phys.* 37.9 (2010), pp. 4902–4915. DOI: 10.1118/1.3480985.

[11] S. Jan et al. "Gate V6: A Major Enhancement of the Gate Simulation Platform Enabling Modelling of CT and Radiotherapy". *Phys. Med. Biol.* 56.4 (2011), pp. 881–901. DOI: 10.1088/0031-9155/56/4/001.

[12] R. Twyman et al. "A Demonstration of STIR-GATE Connection". *IEEE Nuclear Sci. Symp. and Med. Imag.* (2021), pp. 1–3. DOI: 10.1109/nss/mic44867.2021.9875442.

[13] G. Herman and L. Meyer. "Algebraic Reconstruction Techniques Can Be Made Computationally Efficient". *IEEE Trans. Med. Imag.* 12.3 (1993), pp. 600–609. DOI: 10.1109/42.241889.

[14] O. Ronneberger et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation". *Med. Imag. Comput. and Comput. Assist Interv.* 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[15] Y. Tsai et al. "Benefits of Using a Spatially-Variant Penalty Strength With Anatomical Priors in PET Reconstruction". *IEEE Trans. Med. Imag.* 39.1 (2020), pp. 11–22. DOI: 10.1109/tmi.2019.2913889.

[16] G. Schramm. "PARALLELPROJ – An Open-Source Framework for Fast Calculation of Projections in Tomography" (2022). DOI: 10.48550/ARXIV.2212.12519.