



OPEN

DATA DESCRIPTOR

# LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction

Johannes Leuschner<sup>1,2</sup> <sup>✉</sup>, Maximilian Schmidt<sup>1,2</sup> <sup>✉</sup>, Daniel Otero Baguer<sup>1</sup> <sup>1</sup> & Peter Maass<sup>1</sup>

Deep learning approaches for tomographic image reconstruction have become very effective and have been demonstrated to be competitive in the field. Comparing these approaches is a challenging task as they rely to a great extent on the data and setup used for training. With the Low-Dose Parallel Beam (LoDoPaB)-CT dataset, we provide a comprehensive, open-access database of computed tomography images and simulated low photon count measurements. It is suitable for training and comparing deep learning methods as well as classical reconstruction approaches. The dataset contains over 40000 scan slices from around 800 patients selected from the LIDC/IDRI database. The data selection and simulation setup are described in detail, and the generating script is publicly accessible. In addition, we provide a Python library for simplified access to the dataset and an online reconstruction challenge. Furthermore, the dataset can also be used for transfer learning as well as sparse and limited-angle reconstruction scenarios.

## Background & Summary

Tomographic image reconstruction is an extensively studied field. One popular imaging modality in clinical and industrial applications is computed tomography (CT). It allows for the non-invasive acquisition of the inside of an object or the human body. The measurements are based on the attenuation of X-ray beams. To obtain the internal distribution of the body from these measurements, an inverse problem must be solved. Traditionally, analytical methods, like filtered back-projection (FBP) or iterative reconstruction (IR) techniques, are used for this task. These methods are the gold standard in the presence of enough high-dose/low-noise measurements. However, as high doses of applied radiation are potentially harmful to the patients, modern scanners aim at reducing the radiation dose. There exist several strategies, but all introduce specific challenges for the reconstruction algorithm, e.g. undersampling or increased noise levels, which require more sophisticated reconstruction methods. The higher the noise or undersampling, the more prior knowledge about the target reconstructions is needed to improve the final quality<sup>1</sup>. Analytical methods are only able to use very limited prior information. Alternatively, machine learning approaches are able to learn underlying distributions and typical image features, which constitute a much larger and flexible prior. Recent image reconstruction approaches involving machine learning, in particular deep learning (DL), have been developed and demonstrated to be very competitive<sup>2-8</sup>.

DL-based approaches benefit strongly from the availability of comprehensive datasets. In the last years, a wide variety of CT data has been published, covering different body parts and scan scenarios. For the training of reconstruction models, the projections (measured data) are crucial but are rarely made available. Recently, Low Dose CT Image and Projection Data (LDCT-and-Projection-data)<sup>9</sup> was published by investigators from the Mayo Clinic, which include measured normal-dose projection data of 299 patients in the new open DICOM-CT-PD format. The AAPM Low Dose CT Grand Challenge data<sup>10</sup> includes simulated measurements, featuring 30 different patients. The Finish Inverse Problems Society (FIPS) provides multiple measurements of a walnut<sup>11</sup> and a lotus root<sup>12</sup> aimed at sparse data tomography. Recently, Der Sarkissian *et al.*<sup>13</sup> published cone-beam CT projection data and reconstructions of 42 walnuts. Their dataset is directly aimed at the training and comparison of machine learning methods. In magnetic resonance imaging, fastMRI<sup>14</sup> with 1600 scans of humans knees is another prominent example.

<sup>1</sup>Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 5, 28359, Bremen, Germany. <sup>2</sup>These authors contributed equally: Johannes Leuschner, Maximilian Schmidt. ✉e-mail: [jleuschn@uni-bremen.de](mailto:jleuschn@uni-bremen.de); [maximilian.schmidt@uni-bremen.de](mailto:maximilian.schmidt@uni-bremen.de)

Other CT datasets focus on the detection and segmentation of special structures like lesions in the reconstructions for the development of computer-aided diagnostic (CAD) methods<sup>15–20</sup>. Therefore, they do not include the projection data. The LIDC/IDRI database<sup>15</sup>, which we use for the ground truth of our dataset (cf. section “Methods”), targets lung nodule detection. FUMPE<sup>16</sup> contains CT angiography images of 35 subjects for the detection of pulmonary embolisms. KiTS2019<sup>17</sup> is built around the segmentation of kidney tumours in CT images. The Japanese Society of Radiology Technology (JSRT) database<sup>18</sup> and the National Lung Screening Trial (NLST) in cooperation with the CT Image Library (CTIL)<sup>19,20</sup> each contain scans of the lung. These datasets can also be used for the investigation of reconstruction methods by simulating the missing measurements.

Different learned methods have been successfully applied to the task of low-dose reconstruction<sup>7</sup>. However, comparing these approaches is a challenging task since they highly rely on the data and the setup that is used for training. The main goal of this work is to provide a standard dataset that can be used to train and benchmark learned low-dose CT reconstruction methods. To this end, we introduce the Low-Dose Parallel Beam (LoDoPaB)-CT dataset, which uses the public LIDC/IDRI database<sup>15,21,22</sup> of human chest CT reconstructions. We consider these, in the form of 2D images, to be the so-called ground truth. The projections are created by simulating low photon count CT measurements with a parallel beam scan geometry. Due to the slice-based 2D setup, each of the generated measurements corresponds directly to a ground truth slice. Thus, the reconstruction process can be carried out slice-wise without rebinning<sup>23</sup>, which would have to be applied to the measurements for 3D helical cone-beam geometries commonly used in modern scanners<sup>9</sup> to allow for the slice-wise use of a 2D reconstruction algorithm. In order to generalise from our dataset to the clinical 3D setup, the effect of rebinning needs to be evaluated. Also, learned algorithms directly targeted at 3D reconstruction should be considered in this case, which at the moment are barely computationally feasible<sup>24</sup>, but presumably outperform 2D reconstruction algorithms applied to rebinned measurements. Despite the generalisation to the 3D case not being straight-forward, our dataset allows to train and compare a large number of approaches applicable to the 2D scenario, which we expect to yield insights for the design of 3D algorithms as well.

Paired samples constitute the most complete training data and could be used for all kinds of learning. In particular, methods that require independent samples from the distributions of images and measurements, or only from one of these distributions, can still make use of the dataset. In total, the dataset features more than 40000 sample pairs from over 800 different patients. This amount of data and variability can be necessary to successfully train deep neural networks<sup>25</sup>. It also qualifies the dataset for transfer learning. In addition, the included measurements can be easily modified for sparse and limited angle scan scenarios.

## Methods

In this section, the considered mathematical model of CT is stated first, followed by a detailed description of the dataset generation. This starts with the LIDC/IDRI database<sup>15</sup>, from which we extract the ground truth reconstructions. Finally, the data processing steps are described, which are also summarised in a semi-formal manner at the end of the section. As a technical reference, the script<sup>26</sup> used for generation is available online ([https://github.com/jleuschn/lodopab\\_tech\\_ref](https://github.com/jleuschn/lodopab_tech_ref)).

**Parallel beam CT model.** We consider the inverse problem of computed tomography given by

$$\mathcal{A}x + \varepsilon(\mathcal{A}x) = y^\delta \quad (1)$$

with:

- $\mathcal{A}$  the linear ray transform defined by the scan geometry,
- $x$  the unknown interior distribution of the X-ray attenuation coefficient in the body, also called image,
- $\varepsilon$  a sample from a noise distribution that may depend on the ideal measurement  $\mathcal{A}x$ ,
- $y^\delta$  the noisy CT measurement, also called projections or sinogram.

More specifically, we choose a two-dimensional parallel beam geometry, for which the ray transform  $\mathcal{A}$  is the Radon transform<sup>27</sup>. It integrates the values of  $x: \mathbb{R}^2 \rightarrow \mathbb{R}$  fulfilling some regularity conditions (cf. Radon<sup>27</sup>) along the X-ray lines

$$L_{s,\varphi}(t) := s\omega(\varphi) + t\omega^\perp(\varphi), \quad \omega(\varphi) := \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix}, \quad \omega^\perp(\varphi) := \begin{bmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{bmatrix}, \quad (2)$$

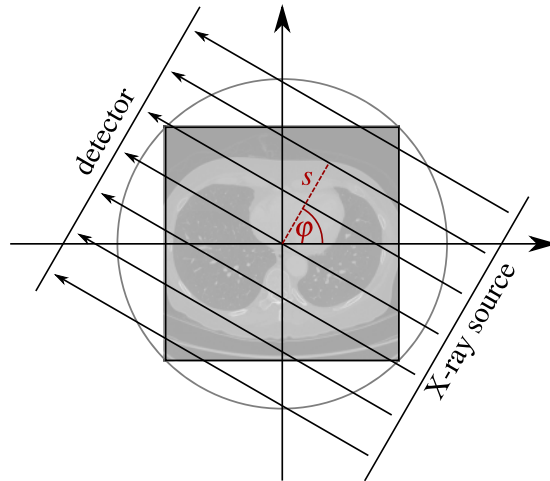
for all parameters  $s \in \mathbb{R}$  and  $\varphi \in [0, \pi)$ , which denote the distance from the origin and the angle, respectively (cf. Figure 1). In mathematical terms, the image is transformed into a function of  $(s, \varphi)$ ,

$$\mathcal{A}x(s, \varphi) := \int_{\mathbb{R}} x(L_{s,\varphi}(t)) dt, \quad (3)$$

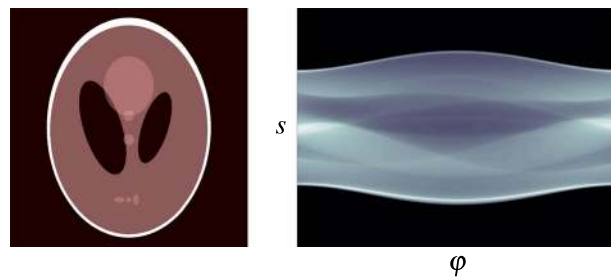
which is called projection, since for each fixed angle  $\varphi$  the 2D image  $x$  is projected onto a line parameterised by  $s$ , namely the detector. Visualisations of projections as images themselves are called sinograms (cf. Figure 2). The projection relates to the ideal intensity measurements  $I_1(s, \varphi)$  at the detector according to Beer-Lambert’s law by

$$\mathcal{A}(s, \varphi) = -\ln \left( \frac{I_1(s, \varphi)}{I_0} \right) = y(s, \varphi), \quad (4)$$

where  $I_0$  is the intensity of an unattenuated beam.



**Fig. 1** Visualisation<sup>25</sup> of the parallel beam geometry.



**Fig. 2** The Shepp-Logan phantom (left) and its corresponding sinogram (right).

In practice, the measured intensities are noisy. The noise can be classified into *quantum noise* and *detector noise*. Quantum noise stems from the process of photon generation, attenuation and detection, which as a whole can be modelled by a Poisson distribution<sup>28</sup>. The detector noise stems from the electronic data acquisition system and is usually assumed to be Gaussian. It would play an important role in ultra-low-dose CT with very small numbers of detected photons<sup>29</sup> but is neglected in our case. Thus we model the number of detected photons and, by this, the measured intensity ratio with

$$\tilde{N}_1(s, \varphi) \sim \text{Pois}(N_0 \exp(-\mathcal{A}x(s, \varphi))), \quad \frac{\tilde{I}_1(s, \varphi)}{I_0} = \frac{\tilde{N}_1(s, \varphi)}{N_0}, \tag{5}$$

where  $N_0$  is the mean photon count without attenuation and  $\text{Pois}(\lambda)$  denotes the probability distribution defined by

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N}_0. \tag{6}$$

For practical application, the model needs to be discretised. The forward operator is then a finite-dimensional linear map  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , where  $n$  is the number of image pixels and  $m$  is the product of the number of detector pixels and the number of angles for which measurements are obtained. The discrete model reads

$$Ax + \varepsilon(Ax) = y^\delta, \quad \varepsilon(Ax) = -Ax - \ln(\tilde{N}_1/N_0), \quad \tilde{N}_1 \sim \text{Pois}(N_0 \exp(-Ax)). \tag{7}$$

Here,  $\text{Pois}(\lambda)$  denotes the joint distribution of  $m$  Poisson distributed observations with parameters  $\lambda_1, \dots, \lambda_m$ , respectively. Note that since the negative logarithm is applied to the observations, the noisy post-log values  $y^\delta$  do not follow a Poisson distribution but the distribution resulting from this log-transformation. However, taking the negative logarithm is required to obtain the linear model and therefore is most commonly applied as a preprocessing step. For our dataset, we consider post-log values by default.

The Radon transform is a linear and compact operator. Therefore, the continuous inverse problem of CT is mildly ill-posed in the sense of Nashed<sup>30,31</sup>. This means that small variations in the measurements can lead to significant differences in the reconstruction (unstable inversion). While the discretised inverse problem is not ill-posed, it is typically ill-conditioned<sup>28</sup>, which leads to artefacts in reconstructions obtained by direct inversion from noisy measurements.



**Fig. 3** Scans from the LIDC/IDRI database<sup>15</sup> with poor quality, good quality and an artefact. The shown HU window is  $[-1024, 1023]$ .

For our discrete simulation setting, we use the described model with following dimensions and parameters:

- Image resolution of  $362 \text{ px} \times 362 \text{ px}$  on a domain of size  $26 \text{ cm} \times 26 \text{ cm}$ .
- 513 equidistant detector bins  $s$  spanning the image diameter.
- 1000 equidistant angles  $\varphi$  between  $0$  and  $\pi$ .
- Mean photon count per detector bin without attenuation  $N_0 = 4096$ .

**LIDC/IDRI database and data selection.** The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) published the LIDC/IDRI database<sup>15,21,22</sup> to support the development of CAD methods for the detection of lung nodules. The dataset consists of 1018 helical thoracic CT scans of 1010 individuals. Seven academic centres and eight medical imaging companies collaborated for the creation of the database. As a result, the data is heterogeneous with respect to the technical parameters and scanner models.

Both standard-dose and lower-dose scans are part of the dataset. Tube peak voltages range from 120 kV to 140 kV and tube current from 40 mA to 627 mA with a mean of 222.1 mA. Labels for the lung nodules were created by a group of 12 radiologists in a two-phase process. The image reconstruction was performed with different filters, depending on the manufacturer of the scanner. Figure 3 shows examples of the provided reconstructions. The LIDC/IDRI database is freely available from The Cancer Imaging Archive (TCIA)<sup>22</sup>. It is published under the Creative Commons Attribution 3.0 Unported License (<https://creativecommons.org/licenses/by/3.0/>).

The LoDoPaB-CT dataset is based on the LIDC/IDRI scans. Our dataset is intended for the evaluation of reconstruction methods in a low-dose setting. Therefore, we simulate the projection data, which is not included in the LIDC/IDRI database. In order to enable a fair comparison with good ground truth, scans that are too noisy were removed in a manual selection process (cf. section “Technical Validation”). Additional scans were excluded due to their geometric properties, namely an image size different from  $512 \text{ px} \times 512 \text{ px}$ , a too small area of valid pixel values (cf. subsection “Ground truth image extraction” below), or a different patient orientation. The complete lists of excluded scan series are given in file [series\\_list.json](#) in the technical reference repository<sup>26</sup>. In the end, 812 patients remain in the LoDoPaB-CT dataset.

The dataset is split into four parts: three parts for training, validation and testing, respectively, and a “challenge” part reserved for the LoDoPaB-CT Challenge (<https://lodopab.grand-challenge.org/>). Each part contains scans from a distinct set of patients, as we want to study the case of learned reconstructors being applied to patients that are not known from training. The training set features scans from 632 patients, while the other parts contain scans from 60 patients each. Every scan contains multiple slices (2D images) for different  $z$ -positions, of which only a subset is included. The amount of extracted slices depends on the slice thickness obtained from the metadata. As slices with small distances are similar, they may not provide much additional information while increasing the chances to overfit. The distances of the extracted slices are larger than 5.0 mm for >45% and larger than 2.5 mm for >75% of the slices. In total, the dataset contains 35820 training images, 3522 validation images, 3553 test images and 3678 challenge images.

*Remark.* We propose to use our default dataset split, as it allows for a fair comparison with other methods that use the same split. However, users are free to remix or re-split the dataset parts. For this purpose, randomised patient IDs are provided, i.e., the same random ID is given for all slices obtained from one patient. Thus, when creating custom splits it can be regulated whether—and to what extent—data from the same patients are contained in different splits.

**Ground truth image extraction.** First, each image is cropped to the central rectangle of  $362 \text{ px} \times 362 \text{ px}$ . This is done because most of the images contain (approximately) circle-shaped reconstructions with a diameter of 512 px (cf. Figure 3). After the crop, the image only contains pixels that lie inside this circle, which avoids value jumps occurring at the border of the circle. While this yields natural ground truth images, we need to point out that the cropped images, in general, do not show the full subject but some interior part. Hence, it is unlikely for methods trained with this dataset to perform well on full-subject measurements.

```

1:  $(\mu_{\text{air}}, \mu_{\text{water}}, \mu_{\text{max}}) \leftarrow (0.02, 20.0, 81.35858)$ 
2: procedure EXTRACTGROUNDTRUTH( $z$ ) ▷  $z$  is an image from the LIDC dataset in the DICOM data format
3:    $\bar{z} \leftarrow \text{transpose}(\text{center\_crop}(z, 362 \times 362))$  ▷ Take center crop of size  $362 \times 362$ 
4:    $\text{image\_hu} \leftarrow \bar{z} * z.\text{RescaleSlope} + z.\text{RescaleIntercept}$  ▷ Transform pixels to HU values
5:    $\text{image\_hu} \leftarrow \text{image\_hu} + \text{noise}$  ▷ Add dequantisation noise uniformly distributed in  $[0, 1]$  for each pixel
6:    $\text{image\_mu} \leftarrow \text{image\_hu} * (\mu_{\text{water}} - \mu_{\text{air}}) / 1000 + \mu_{\text{water}}$  ▷ Convert HU values to linear attenuation
7:   return  $\text{clip}(\text{image\_mu} / \mu_{\text{max}}, \text{min} = 0, \text{max} = 1)$  ▷ Clip all pixel values to be within  $[0, 1]$ 

8: procedure GENERATEPROJECTION( $x$ ) ▷  $x$  is a ground truth generated by the previous procedure
9:    $\bar{x} \leftarrow \text{bilinear\_interpolation}(\mu_{\text{max}} * x, 1000 \times 1000)$  ▷ Interpolate ground truth to avoid inverse crime
10:   $y \leftarrow \text{ray\_trafo}(\bar{x})$  ▷ ODL RayTransform with 1000 angles, 513 detector pixels
11:  and image domain  $[-0.13, 0.13]^2$ 
12:   $\text{photons} \leftarrow \max(\text{poisson}(\exp(-y) * 4096), 0.1)$  ▷ Scale and add Poisson noise. Max is taken pixel-wise.
13:  return  $-\log(\text{photons} / 4096) / \mu_{\text{max}}$  ▷ Apply log-transform and normalisation constant

```

**Fig. 4** Data generation algorithm.

For some scan series, the circle is subject to a geometric transformation either shrinking or expanding the circle in some directions. In particular, for a few scan series, the circle is shrunk such that it is smaller than the cropped rectangle. We exclude these series, i.e. those with patient IDs 0004, 0032, 0102, 0116, 0120, 0289, 0368, 0418, 0541, 0798, 0926, 0972 and 1000, from our dataset, which allows to crop all included images consistently to  $362 \text{ px} \times 362 \text{ px}$ .

The integer Hounsfield unit (HU) values obtained from the DICOM files are dequantised by adding uniform noise from the interval  $[0, 1)$ . By adding this noise, the discrete distribution of stored values is transformed into a continuous distribution (up to the floating-point precision), which is a common assumption of image models. For example, the meaningful evaluation of densities learned by generative networks requires dequantization<sup>32</sup>, which in some works<sup>33</sup> is more refined than the uniform dequantization applied to the HU values in our dataset.

In the next step, the linear attenuations  $\mu$  are computed from the dequantised HU values using the definition of the HU,

$$\text{HU} = 1000 \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \Leftrightarrow \mu = \text{HU} \frac{\mu_{\text{water}} - \mu_{\text{air}}}{1000} + \mu_{\text{water}}, \quad (8)$$

where we use the linear attenuation coefficients

$$\mu_{\text{water}} = 20/\text{m}, \quad \mu_{\text{air}} = 0.02/\text{m}, \quad (9)$$

which approximately correspond to an X-ray energy of  $60 \text{ keV}$ <sup>34</sup>. Finally, the  $\mu$ -values are normalised into  $[0, 1]$  by dividing by

$$\mu_{\text{max}} = 3071 \frac{\mu_{\text{water}} - \mu_{\text{air}}}{1000} + \mu_{\text{water}} = 81.35858/\text{m}, \quad (10)$$

which corresponds to the largest HU value that can be represented with the standard 12-bit encoding, i.e.  $(2^{12}-1-1024)\text{HU} = 3071 \text{ HU}$ , followed by the clipping of all values into the range  $[0, 1]$ ,

$$\hat{\mu} = \text{clip}(\mu / \mu_{\text{max}}, [0, 1]) = \begin{cases} 0 & , \mu / \mu_{\text{max}} \leq 0 \\ \mu / \mu_{\text{max}} & , 0 < \mu / \mu_{\text{max}} \leq 1. \\ 1 & , 1 < \mu / \mu_{\text{max}} \end{cases} \quad (11)$$

The Eqs. (8) and (11) are applied pixel-wise to the images.

**Projection data generation.** To simulate the measurements based on the virtual ground truth images, the main step is to apply the forward operator, which is the ray transform (Radon transform in 2D) for CT. For this task we utilise the Operator Discretization Library<sup>35</sup> (ODL) with the ‘astra\_cpu’ backend<sup>36</sup>.

*Remark.* We choose ‘astra\_cpu’ over the usually favoured ‘astra\_cuda’ because of small inaccuracies observed in the sinograms when using ‘astra\_cuda’, specifically at angles  $0, \frac{\pi}{2}$  and  $\pi$  and detector positions  $-1/\sqrt{2} \frac{l}{2}$  and  $1/\sqrt{2} \frac{l}{2}$  with  $l$  being the length of the detector. The used version is `astra-toolbox==1.8.3` on Python 3.6. The tested CUDA version is 9.1 combined with `cupy==8.0`.

In order to avoid “committing the inverse crime”<sup>37</sup>, which, in our scenario, would be to use the same discrete model both for simulation and reconstruction, we use a higher resolution for the simulation. Otherwise, good performance of reconstructors for the specific resolution of this dataset ( $362 \text{ px} \times 362 \text{ px}$ ) could also stem from the properties of the specific discretised problem, rather than from good inversion of the analytical model. We use bilinear interpolation for the upscaling of the virtual ground truth from  $362 \text{ px} \times 362 \text{ px}$  to  $1000 \text{ px} \times 1000 \text{ px}$ .



The non-normalised, upscaled image is projected by the ray transform. Based on this projection,  $Ax$ , the measured photon counts  $\tilde{N}_1$  are sampled according to Eq. (7). The sampling in some cases yields photon counts of zero, which we then replace by photon counts of 0.1. Hereby strictly positive values are ensured, which is a prerequisite for the log-transform in the next step (cf. Wang *et al.*<sup>38</sup>). The negative logarithm of the photon counts quotient  $\max(0, 1, \tilde{N}_1)N_0$  is taken, resulting in the post-log measurements  $y^\delta$  according to Eq. (7) (up to the 0.1 photon count approximation). Finally,  $y^\delta$  is divided by  $\mu_{\max}$  to match the normalised ground truth images. A summary of all steps can be found in Fig. 4 (Data generation algorithm).

**Remark.** Although the linear model obtained by the log-transform is easier to study, in some cases pre-log models are more accurate. See Fu *et al.*<sup>29</sup> for a detailed comparison. For applying a pre-log method, the stored observation data  $\hat{y} = y^\delta / \mu_{\max}$  must be back-transformed by  $\exp(-\mu_{\max} \cdot \hat{y})$ . To create physically consistent data pairs, the ground truth images should then be multiplied with  $\mu_{\max}$ , too.

**Remark.** Note that the minimum photon count of 0.1 can be adapted subsequently. This is most easily done by filtering out the highest observation values and replacing them with  $-\log(\varepsilon_0/4096)/\mu_{\max}$ , where  $\varepsilon_0$  is the new minimum photon count.

## Data Records

The LoDoPaB-CT dataset is published as open access on Zenodo (<https://zenodo.org>) in two repositories. The main data repository<sup>39</sup> (<https://doi.org/10.5281/zenodo.3384092>) has a size of around 55GB and contains observations and ground truth data of the train, validation and test set. For each subset, represented by \*, the following files are included:

- CSV files `patient_ids_rand_*.csv` include randomised patient IDs of the samples. The patient IDs of the train, validation and test parts are integers in the range of 0–631, 632–691 and 692–751, respectively. The ID of each sample is stored in a single row.
- Zip archives `ground_truth_*.zip` contain HDF5<sup>40</sup> files of the ground truth reconstructions.
- Zip archives `observation_*.zip` contain HDF5 files of the simulated low-dose measurements.
- Each HDF5 file contains one HDF5 dataset named `data`, that provides several samples (128 except for the last file in each ZIP file). For example, the  $n$ -th training sample pair is stored in the HDF5 files `observation_train_%03d.hdf5` and `ground_truth_train_%03d.hdf5` where the placeholder %03d is floor( $n/128$ ). Within these HDF5 files, the observation or ground truth is stored at entry  $(n \bmod 128)$  of the HDF5 dataset `data`.

The second repository<sup>41</sup> for the challenge data (<https://doi.org/10.5281/zenodo.3874937>) consists of a single zip archive:

- `observation_challenge.zip` contains HDF5 files of the simulated low-dose measurements.

The structure inside the HDF5 files is the same as in the main repository.

## Technical Validation

**Ground truth & data selection.** Creating high-quality ground truth images for tomographic image reconstruction is a challenging and time-consuming task. In computed tomography, one option is to cut open the object after the scan or use 3D printing<sup>42</sup>, whereby the digital template of the object is the reference. In general, this also involves high radiation doses and many scanning angles. This combination makes it even harder to generate ground truth images for medical applications.

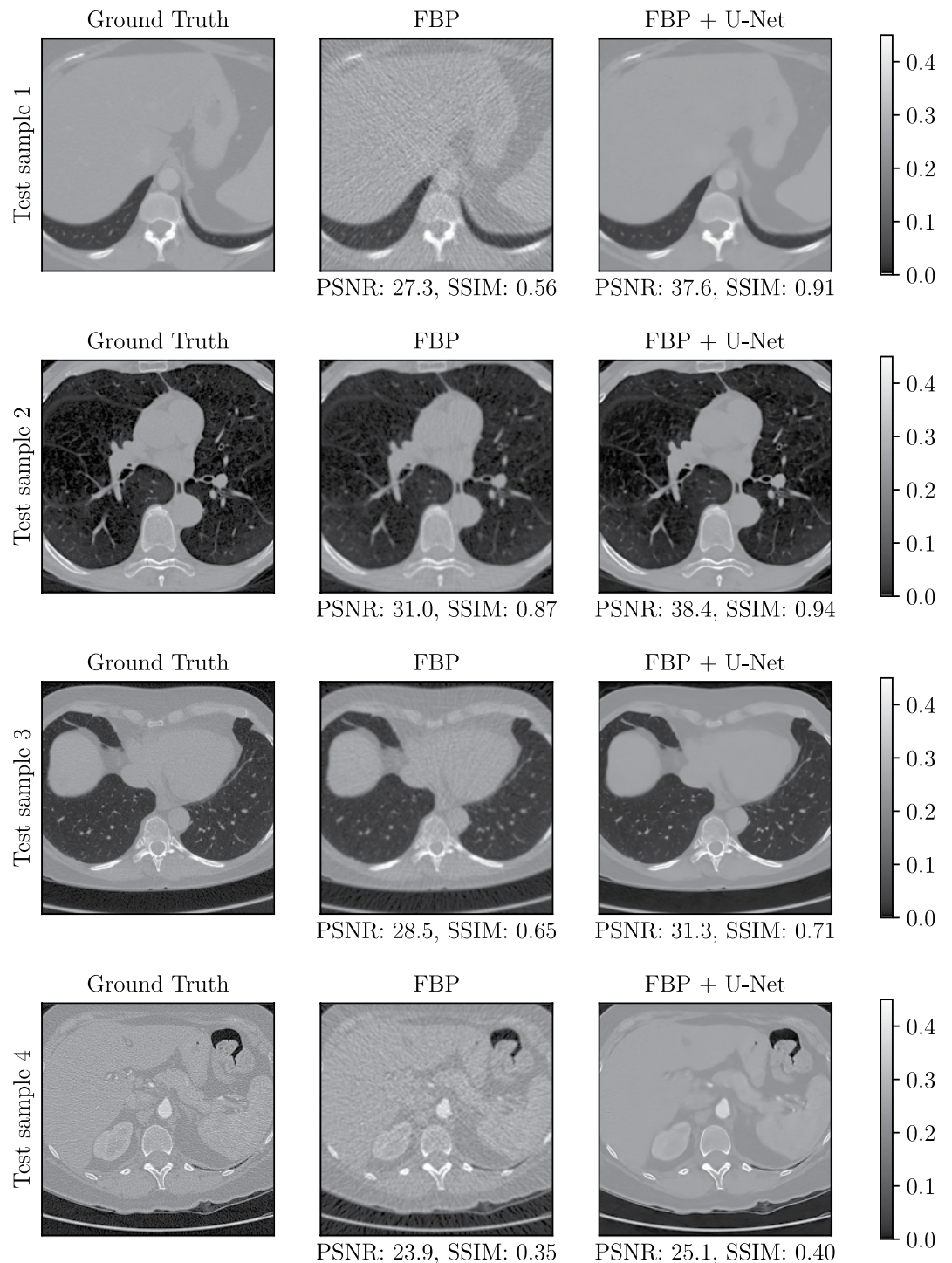
For low-dose CT reconstruction models, the primary goal is to match the normal-dose reconstruction quality of methods currently in use. Therefore, normal-dose reconstructions from classical methods, e.g. filtered back-projection, are an adequate choice as ground truth. This simplifies the process considerably.

The ground truth CT reconstructions of LoDoPaB-CT are taken from the established and well-documented LIDC/IDRI database. An independent visual inspection of one 2D slice per scan was performed by three of the authors. Figure 3 shows three examples of such slices. A five-star rating system was used to evaluate the image quality and remove noisy ground truth data, like the first slice in Fig. 3. Scans with artefacts, e.g. from photon starvation due to dense material (cf. Figure 3 (right)), were in general not removed, as the artefacts only affect a few slices of the whole scan. The slice in the middle of Fig. 3 represents an ideal ground truth. The following procedure was then used to exclude scans based on their rating:

1. Centring of the ratings from each evaluator around the value 3.
2. Calculation of the mean rating and the variance for each looked at 2D slice.
3. For a variance  $< 1$ , the mean was used as the rating score. Otherwise, the scan is evaluated by all three authors together.
4. All scans with a rating  $\leq 2$  are excluded from the dataset.

These excluded scans are listed at key “`series_excluded_manual_low_q_filter`” in file [series\\_list.json](#) in the technical reference repository<sup>26</sup>.

**Reference reconstructions & quantitative results.** To validate the usability of the proposed dataset for machine learning approaches, we provide reference reconstructions and quantitative results for the standard



**Fig. 5** Different baseline reconstructions from the FBP and FBP + U-Net methods. The ground truth images are part of the LoDoPaB-CT test set. The window  $[0, 0.45]$  corresponds to a HU range of  $\approx[-1001, 831]$ .

filtered back-projection (FBP) and a learned post-processing method (FBP + U-Net). FBP is a widely used analytical reconstruction technique (cf. Buzug<sup>28</sup> for an introduction). If the measurements are noisy (due to the low dose), FBP reconstructions tend to include streaking artefacts. A typical approach to overcome this problem is to apply some post-processing such as denoising. Recent works<sup>3,4,8</sup> have successfully used convolutional neural networks, such as the U-Net<sup>43</sup>. The idea is to train a neural network to create clean reconstructions out of the noisy FBP results.

In this initial study, for the FBP, we used the Hann filter with a frequency scaling of 0.641. We selected these parameters based on the performance over the first 100 samples of the validation dataset. For the post-processing approach (FBP + U-Net), we used a U-Net-like architecture with 5 scales. We trained it using the proposed

	training set		validation set		test set	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
FBP	30.45 ± 2.65	0.7415 ± 0.1314	30.75 ± 2.52	0.7577 ± 0.1231	30.52 ± 3.10	0.7372 ± 0.1467
FBP + U-Net	36.17 ± 3.75	0.8623 ± 0.1228	36.74 ± 3.28	0.8819 ± 0.1017	35.84 ± 4.59	0.8443 ± 0.1501

**Table 1.** Baseline performance. Values are the mean and standard deviation over all samples.

dataset by minimising the mean squared error loss with the Adam algorithm<sup>44</sup> for a maximum of 250 epochs with batch size 32. Additionally, we used an initial learning rate of  $10^{-3}$ , decayed using cosine annealing until  $10^{-4}$ . The model with the highest mean peak signal-to-noise ratio (PSNR) on the validation set was selected from the models obtained during training. Sample reconstructions are shown in Fig. 5.

Table 1 depicts the obtained results in terms of the peak signal-to-noise ratio (PSNR) and structural similarity<sup>45</sup> (SSIM) metrics (cf. “Evaluation practice” in the next section for a detailed explanation). As it can be observed, the post-processing approach, which was trained using the proposed dataset, outperforms the classical FBP reconstructions by a margin of 5 dB. This demonstrates that the dataset indeed contains valuable data ready to be used for training machine learning methods to obtain CT reconstructions with higher quality than the standard methods.

### Usage Notes

**Download & Easy access.** The whole LoDoPaB-CT dataset<sup>39,41</sup> can be downloaded directly from the Zenodo website. However, we recommend the Python library *DIV $\alpha$ l*<sup>46</sup> (<https://github.com/jleuschn/dival>) for easy access of the dataset. The library includes specific functionalities for the interaction with the provided dataset.

*Remark.* Access to the dataset on Zenodo might be restricted or slow in some regions of the world. In this case please contact one of the corresponding authors to get an alternative download option.

*DIV $\alpha$ l* is also available through the package index PyPI (<https://pypi.org/project/dival>). With the library, the dataset is automatically downloaded, checked for corruption and ready for use within two lines of Python code:

```
from dival import get_standard_dataset
dataset = get_standard_dataset('lodopab').
```

*Remark.* When loading the dataset using *DIV $\alpha$ l*, an ODL<sup>35</sup> *RayTransform* implementing the forward operator is created. This requires a backend, the default being ‘astra\_cuda’, which requires both the astra toolbox<sup>36</sup> and CUDA to be available. If either is unavailable, a different backend (‘astra\_cpu’ or ‘skimage’) must be selected by keyword argument *impl*.

In addition, *DIV $\alpha$ l* offers multiple options to work with the LoDoPaB-CT dataset:

- Access the train, validation and test subset and draw a specific number of samples.
- Sort the data by the patient ids.
- Use the pre-log or post-log data (cf. projection data generation in the “Methods” section).
- Evaluate the reconstruction performance.
- Compare with pre-trained standard reconstruction models.

**Evaluation practice.** Since ground truth data is provided in the dataset, we recommend using so-called full-reference methods for the evaluation. The peak signal-to-noise ratio (PSNR) and the structural similarity<sup>45</sup> (SSIM) are two standard image quality metrics often used in CT applications<sup>42,47</sup>. While the PSNR calculates pixel-wise intensity comparisons between ground truth and reconstruction, SSIM captures structural distortions.

*Peak signal-to-noise ratio.* The PSNR expresses the ratio between the maximum possible image intensity and the distorting noise, measured by the mean squared error (MSE),

$$\text{PSNR}(\bar{x}, x) := 10 \log_{10} \left( \frac{\max_x}{\text{MSE}(\bar{x}, x)} \right), \quad \text{MSE}(\bar{x}, x) := \frac{1}{n} \sum_{i=1}^n |\bar{x}_i - x_i|^2. \quad (12)$$

Here  $x$  is the ground truth image and  $\bar{x}$  the reconstruction. Higher PSNR values are an indication of a better reconstruction. We recommend choosing  $\max_x = \max(x) - \min(x)$ , i.e. the difference between the highest and lowest entry in  $x$ , instead of the maximum possible intensity, since the reference value of 3071HU is far from the most common values. Otherwise, the results can often be too optimistic.

*Structural similarity.* Based on assumptions about the human visual perception, SSIM compares the overall image structure of ground truth and reconstruction. Results lie in the range  $[0, 1]$ , with higher values being better. The SSIM is computed through a sliding window at  $M$  locations

$$\text{SSIM}(\bar{x}, x) := \frac{1}{M} \sum_{j=1}^M \frac{(2\bar{\mu}_j \mu_j + C_1)(2\sigma_j + C_2)}{(\bar{\mu}_j^2 + \mu_j^2 + C_1)(\bar{\sigma}_j^2 + \sigma_j^2 + C_2)}, \quad (13)$$



where  $\tilde{\mu}_j$  and  $\mu_j$  are the average pixel intensities,  $\tilde{\sigma}_j$  and  $\sigma_j$  the variances and  $\Sigma_j$  the covariance of  $\tilde{x}$  and  $x$  at the  $j$ -th local window. Constants  $C_1 = (K_1L)^2$  and  $C_2 = (K_2L)^2$  stabilise the division. Following Wang *et al.*<sup>45</sup> we choose  $K_1 = 0.01$  and  $K_2 = 0.03$  for the technical validation in this paper. The window size is  $7 \times 7$  and  $L = \max(x) - \min(x)$ .

**Test & challenge set.** The test data is the advised subset for offline model evaluation. To guarantee a fair comparison, the data should be in no way involved in the training process or hyperparameter selection of the model. We recommend using the whole test set and select the above-mentioned parameters for PSNR and SSIM. Deviations from this setting should be mentioned.

In addition, a challenge set without ground truth images is provided. We encourage users to submit their challenge reconstructions to the evaluation website (<https://lodopab.grand-challenge.org/>). All methods are assessed under the same conditions and with the same metrics. The performance can be directly compared with other methods on a public leaderboard. Therefore, we recommend to report performance measures on the challenge set for publications that use the LoDoPaB-CT dataset without modifications, in addition to any evaluations on the test set. In accordance with the Biomedical Image Analysis (BIAS) guidelines<sup>48</sup>, more information about the challenge can be found on the aforementioned website.

**Further usage. Scan scenarios.** The provided measurements and simulation scripts can easily be modified to cover different scan scenarios:

- Limited and sparse-angle problems can be created by loading a subset of the projection data, e.g. a sparser setup with 200 angles was already used by Baguer *et al.*<sup>25</sup>.
- Super-resolution experiments can be mimicked, by artificially binning the projection data into larger pixels.
- To study lower or higher photon counts, the dataset can be re-simulated with a different value of  $N_0$  (e.g. using `resimulate_observations.py`<sup>26</sup> by changing the value of `PHOTONS_PER_PIXEL`).

The provided reconstructions can still be used as ground truth for all listed scenarios.

**Transfer learning.** Transfer learning is a popular approach to boost the performance of machine learning models on smaller datasets. The idea is to first train the model on a different, comprehensive data collection. Afterwards, the determined parameters are used as an initial guess for fine-tuning the model on the smaller one. In general, the goal is to learn to process low-level features, e.g. edges in images, from the comprehensive dataset. The adaption to specific high-level features is then performed on the smaller dataset. For imaging applications, the ImageNet database<sup>49</sup>, with over 14 million natural images, is frequently used in this role. The applications range from image classification<sup>50</sup> to other domains like audio data<sup>51</sup>.

Transfer learning has also been successfully applied to CT reconstruction tasks. This includes training on different scan scenarios<sup>52,53</sup>, e.g. a different number of angles, as well as first training on 2D data and continuing on 3D data<sup>54</sup>. He *et al.*<sup>55</sup> simulated parallel beam measurements on some of the natural images contained in ImageNet. Subsequently, the training was continued on CT images from the Mayo Clinic<sup>10</sup>. LoDoPaB-CT, or parts of the dataset, can be used in similar roles for transfer learning. Additionally, the ground truth data from real thoracic CT scans may be advantageous for similar CT reconstruction tasks compared to random natural images from ImageNet<sup>56</sup>.

Nonetheless, we advise the user to check the applicability for their specific use case and reconstruction model. Re-simulation or other changes to the LoDoPaB-CT dataset might be needed, especially for datasets with different scan geometries. Additionally, simulated data can not capture all aspects of real-world measurements and therefore cause reconstruction errors. For a comprehensive study on the benefits and challenges of transfer learning for medical imaging, we refer the reader to the publication by Raghu *et al.*<sup>56</sup>.

**Remark.** An example for a simulation script with a fan beam geometry on the ground truth data can be found in the `DIVα`<sup>46</sup> library: `dival/examples/ct_simulate_fan_beam_from_lodopab_ground_truth.py`.

**Limits of the dataset.** The LoDoPaB-CT dataset is designed for a methodological comparison of CT reconstruction methods on a simulated low-dose parallel beam setting. The focus is on how a model deals with the challenges that arise from low photon count measurements to match the quality of normal-dose images. Of course, this represents only one aspect of many for the application in real-world scenarios. Therefore, results achieved on LoDoPaB-CT might not completely reflect the performance on real medical data. The following limits of the dataset should be considered when evaluating and comparing results:

- The simulation uses the Radon transform and Poisson noise. Real measurements can be influenced by additional physical effects, like scattering.
- Modern CT machines use advanced scanning geometries, like helical fan beam or cone beam. Specific challenges for the reconstruction can arise compared to parallel beam measurements (cf. Buzug<sup>28</sup>).
- In general, the goal is to reconstruct a whole 3D subject and not just a single 2D slice. Reconstruction methods might benefit from additional spacial information. On the other hand, requirements on memory and compute power can be higher for methods that reconstruct 3D volumes directly.

- Image metrics, e.g. PSNR and SSIM, cannot express and cover all aspects of high-quality CT reconstruction. An additional assessment by experts in the field can be beneficial.
- The ground truth images are based on reconstructions from normal-dose medical scans. As such, they can contain noise and artefacts. The measurements are created from this “noisy” ground truth. Therefore, a perfect reconstruction model would re-create the imperfections. Approaches that are designed to remove them can score lower PSNR and SSIM values, although their reconstruction quality might be higher.
- A crop to a region of interest is used for the ground truth images (cf. “Ground truth image extraction”). Hence, the results for full-subject measurements can be different.

### Code availability

Python scripts<sup>26</sup> for the simulation setup and the creation of the dataset are publicly available on Github ([https://github.com/jleuschn/lodopab\\_tech\\_ref](https://github.com/jleuschn/lodopab_tech_ref)). They make use of the ASTRA Toolbox<sup>36</sup> (version 1.8.3) and the Operator Discretization Library<sup>35</sup> (ODL, version  $\geq 0.7.0$ ). In addition, the ground truth reconstructions from the LIDC/IDRI database<sup>21</sup> are needed for the simulation process. A sample data split into training, validation, test and challenge part is also provided. It differs from the one used for the creation of this dataset in order to keep the ground truth data of the challenge set undisclosed. The random seeds used in the scripts are modified for the same reason. The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health, and their critical role in the creation of the free publicly available LIDC/IDRI database used in this study.

Received: 12 November 2020; Accepted: 4 March 2021;

Published online: 16 April 2021

### References

1. Benning, M. & Burger, M. Modern regularization methods for inverse problems. *Acta Numerica* **27**, <https://doi.org/10.1017/S0962492918000016> (2018).
2. Adler, J. & Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems* **33**, 124007, <https://doi.org/10.1088/1361-6420/aa9581> (2017).
3. Chen, H. *et al.* Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging* **36**, 2524–2535, <https://doi.org/10.1109/TMI.2017.2715284> (2017).
4. Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **26**, 4509–4522, <https://doi.org/10.1109/TIP.2017.2713099> (2017).
5. Li, H., Schwab, J., Antholzer, S. & Haltmeier, M. NETT: solving inverse problems with deep neural networks. *Inverse Problems* **36**, 065005, <https://doi.org/10.1088/1361-6420/ab6d57> (2020).
6. Shan, H. *et al.* Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nature Machine Intelligence* **1**, 269–276, <https://doi.org/10.1038/s42256-019-0057-9> (2019).
7. Wang, G., Ye, J. C., Mueller, K. & Fessler, J. A. Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging* **37**, 1289–1296, <https://doi.org/10.1109/TMI.2018.2833635> (2018).
8. Yang, Q. *et al.* Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging* **37**, 1348–1357, <https://doi.org/10.1109/TMI.2018.2827462> (2018).
9. McCollough, C. *et al.* Data from Low Dose CT Image and Projection Data. *The Cancer Imaging Archive* <https://doi.org/10.7937/9npb-2637> (2020).
10. McCollough, C. TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge. *Medical Physics* **43**, 3759–3760, <https://doi.org/10.1118/1.4957556> (2016).
11. Hämäläinen, K. *et al.* Tomographic X-ray data of a walnut. Preprint at <https://arxiv.org/abs/1502.04064> (2015).
12. Bubba, T. A., Hauptmann, A., Huotari, S., Rimpeläinen, J. & Siltanen, S. Tomographic X-ray data of a lotus root filled with attenuating objects. Preprint at <https://arxiv.org/abs/1609.07299> (2016).
13. Der Sarkissian, H. *et al.* A cone-beam X-ray computed tomography data collection designed for machine learning. *Scientific Data* **6**, 215, <https://doi.org/10.1038/s41597-019-0235-y> (2019).
14. Knoll, F. *et al.* fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiology: Artificial Intelligence* **2**, e190007, <https://doi.org/10.1148/ryai.2020190007> (2020).
15. Armato, S. G. III *et al.* The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* **38**, 915–931, <https://doi.org/10.1118/1.3528204> (2011).
16. Masoudi, M. *et al.* A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism. *Scientific Data* **5**, 180180 EP, <https://doi.org/10.1038/sdata.2018.180> (2018).
17. Heller, N. *et al.* The KiTS19 Challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. Preprint at <https://arxiv.org/abs/1904.00445> (2019).
18. Shiraiishi, J. *et al.* Development of a digital image database for chest radiographs with and without a lung nodule. *American Journal of Roentgenology* **174**, 71–74, <https://doi.org/10.2214/ajr.174.1.1740071> (2000).
19. Clark, K. W. *et al.* Creation of a CT image library for the lung screening study of the National Lung Screening Trial. *Journal of Digital Imaging* **20**, 23–31, <https://doi.org/10.1007/s10278-006-0589-5> (2007).
20. Cody, D. D. *et al.* Normalized CT dose index of the CT scanners used in the national lung screening trial. *American Journal of Roentgenology* **194**, 1539–1546, <https://doi.org/10.2214/AJR.09.3268> (2010).
21. Armato, S. G. III *et al.* Data from LIDC-IDRI. *The Cancer Imaging Archive* <https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX> (2015).
22. Clark, K. *et al.* The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging* **26**, 1045–1057, <https://doi.org/10.1007/s10278-013-9622-7> (2013).
23. Defrise, M., Noo, F. & Kudo, H. Rebinning-based algorithms for helical cone-beam CT. *Physics in Medicine and Biology* **46**, 2911–2937, <https://doi.org/10.1088/0031-9155/46/11/311> (2001).
24. Etmann, C., Ke, R. & Schönlieb, C. iUNets: Learnable invertible up- and downsampling for large-scale inverse problems. In *30th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2020, Espoo, Finland, September 21–24, 2020*, 1–6, <https://doi.org/10.1109/MLSP49062.2020.9231874> (IEEE, 2020).
25. Bagger, D. O., Leuschner, J. & Schmidt, M. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems* **36**, 094004, <https://doi.org/10.1088/1361-6420/aba415> (2020).
26. Leuschner, J., Schmidt, M. & Bagger, D. O. LoDoPaB-CT Generation Technical Reference ( $\geq v1.2$ ). *Zenodo* <https://doi.org/10.5281/zenodo.3957743> (2020).

27. Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging* **5**, 170–176, <https://doi.org/10.1109/TMI.1986.4307775> (1986).
28. Buzug, T. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT* (Springer Berlin Heidelberg, 2008).
29. Fu, L. *et al.* Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction. *IEEE Transactions on Medical Imaging* **36**, 707–720, <https://doi.org/10.1109/TMI.2016.2627004> (2017).
30. Nashed, M. A new approach to classification and regularization of ill-posed operator equations. In Engl, H. W. & Groetsch, C. (eds.) *Inverse and Ill-Posed Problems*, 53–75, <https://doi.org/10.1016/B978-0-12-239040-1.50009-0> (Academic Press, 1987).
31. Natterer, F. *The mathematics of computerized tomography*. No. 32 in *Classics in applied mathematics* (Society for Industrial and Applied Mathematics, Philadelphia, 2001).
32. Theis, L., van den Oord, A. & Bethge, M. A note on the evaluation of generative models. In Bengio, Y. & LeCun, Y. (eds.) *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings* (2016).
33. Ho, J., Chen, X., Srinivas, A., Duan, Y. & Abbeel, P. Flow++: Improving flow-based generative models with variational dequantization and architecture design. vol. 97 of *Proceedings of Machine Learning Research*, 2722–2730 (PMLR, Long Beach, California, USA, 2019).
34. Hubbell, J. & Seltzer, S. Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements  $z = 1$  to 92 and 48 additional substances of dosimetric interest. Tech. Rep. PB-95-220539/XAB; NISTIR-5632; TRN: 51812148, National Inst. of Standards and Technology - PL, Gaithersburg, MD (United States). Ionizing Radiation Div. <https://doi.org/10.18434/T4D01F> (1995).
35. Adler, J. *et al.* odlgroup/odl: ODL 0.7.0. *Zenodo* <https://doi.org/10.5281/zenodo.592765> (2018).
36. van Aarle, W. *et al.* The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **157**, 35–47, <https://doi.org/10.1016/j.ultramic.2015.05.002> (2015).
37. Wirgin, A. The inverse crime. Preprint at <https://arxiv.org/abs/math-ph/0401050> (2004).
38. Wang, G., Zhou, J., Yu, Z., Wang, W. & Qi, J. Hybrid pre-log and post-log image reconstruction for computed tomography. *IEEE Transactions on Medical Imaging* **36**, 2457–2465, <https://doi.org/10.1109/TMI.2017.2751679> (2017).
39. Leuschner, J., Schmidt, M. & Baguer, D. O. LoDoPaB-CT dataset (v1.0.0). *Zenodo* <https://doi.org/10.5281/zenodo.3384092> (2019).
40. The HDF Group. Hierarchical Data Format, version 5 (1997). <https://www.hdfgroup.org/HDF5/>.
41. Leuschner, J., Schmidt, M. & Baguer, D. O. LoDoPaB-CT challenge set (v1.0.0). *Zenodo* <https://doi.org/10.5281/zenodo.3874937> (2020).
42. Joemai, R. M. S. & Geleijns, J. Assessment of structural similarity in CT using filtered backprojection and iterative reconstruction: a phantom study with 3D printed lung vessels. *The British Journal of Radiology* **90**, 20160519, <https://doi.org/10.1259/bjr.20160519> (2017).
43. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28) (Springer International Publishing, Cham, 2015).
44. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).
45. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600–612, <https://doi.org/10.1109/TIP.2003.819861> (2004).
46. Leuschner, J., Schmidt, M., Baguer, D. O. & Erzmänn, D. DIVAL library. *Zenodo* <https://doi.org/10.5281/zenodo.3970516> (2021).
47. Adler, J. & Öktem, O. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging* **37**, 1322–1332, <https://doi.org/10.1109/TMI.2018.2799231> (2018).
48. Maier-Hein, L. *et al.* BIAS: Transparent reporting of biomedical image analysis challenges. *Medical Image Analysis* **66**, 101796, <https://doi.org/10.1016/j.media.2020.101796> (2020).
49. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255, <https://doi.org/10.1109/CVPR.2009.5206848> (2009).
50. Shermin, T. *et al.* Enhanced transfer learning with ImageNet trained classification layer. In Lee, C., Su, Z. & Sugimoto, A. (eds.) *Image and Video Technology*, 142–155, [https://doi.org/10.1007/978-3-030-34879-3\\_12](https://doi.org/10.1007/978-3-030-34879-3_12) (Springer International Publishing, Cham, 2019).
51. Grzywczak, D. & Gwardys, G. Deep image features in music information retrieval. *International Journal of Electronics and Telecommunications* **60**, 187–199, [https://doi.org/10.1007/978-3-319-09912-5\\_16](https://doi.org/10.1007/978-3-319-09912-5_16) (2014).
52. Wu, Z., Yang, T., Li, L. & Zhu, Y. Hierarchical convolutional network for sparse-view X-ray CT reconstruction. In Mahalanobis, A., Tian, L. & Petrucci, J. C. (eds.) *Computational Imaging IV*, vol. 10990, 141–146, <https://doi.org/10.1117/12.2521239>. International Society for Optics and Photonics (SPIE, 2019).
53. Kalare, K. W. & Bajpai, M. K. RecDNN: Deep neural network for image reconstruction from limited view projection data. *Soft Comput.* **24**, 17205–17220, <https://doi.org/10.1007/s00500-020-05013-4> (2020).
54. Shan, H. *et al.* 3-d convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-d trained network. *IEEE Transactions on Medical Imaging* **37**, 1522–1534, <https://doi.org/10.1109/TMI.2018.2832217> (2018).
55. He, J., Wang, Y. & Ma, J. Radon inversion via deep learning. *IEEE Transactions on Medical Imaging* **39**, 2076–2087, <https://doi.org/10.1109/TMI.2020.2964266> (2020).
56. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding transfer learning for medical imaging. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 32, 3347–3357 (Curran Associates, Inc., 2019).

## Acknowledgements

Johannes Leuschner, Maximilian Schmidt and Daniel Otero Baguer acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) within the framework of GRK 2224/1 “ $\pi^3$ : Parameter Identification – Analysis, Algorithms, Applications”. We thank Simon Arridge, Ozan Öktem, Carola-Bibiane Schönlieb and Christian Etmann for the fruitful discussion about the procedure, and Felix Lucka and Jonas Adler for their ideas and helpful feedback on the simulation setup. Open Access funding enabled and organized by Projekt DEAL.

## Author contributions

All authors worked on the concept and simulation setup of the dataset. J.L. and M.S. wrote the simulation scripts and the main parts of the manuscript. D.O. performed and documented the model-based technical validation of the dataset. All authors reviewed, finalised and approved the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to J.L. or M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021