 doi:10.20944/preprints202107.0265.v1

*Article*

# A green prospective for learned post-processing in sparse-view tomographic reconstruction

**Elena Morotti[1,‡,*] and Davide Evangelista [2,‡] and Elena Loli Piccolomini [3,‡]**

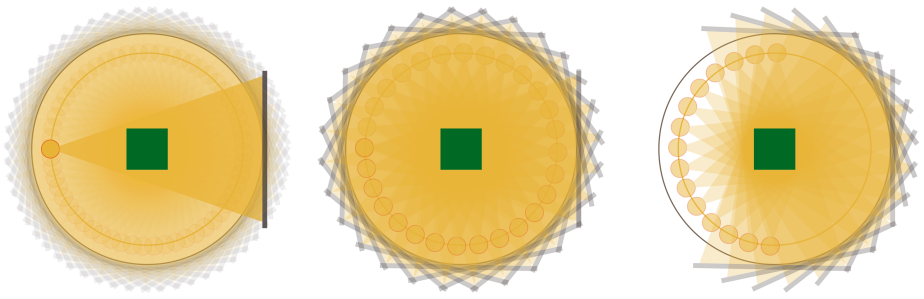
[1] Department of Political and Social Sciences, University of Bologna
[2] Department of Mathematics, University of Bologna
[3] Department of Computer Science and Engineering, University of Bologna
* Correspondence: elena.loli@unibo.it
‡ These authors contributed equally to this work.

**Abstract:** Deep Learning is developing interesting tools which are of great interest for inverse imaging applications. In this work, we consider a medical imaging reconstruction task from subsampled measurements, which is an active research field where Convolutional Neural Networks have already revealed their great potential. However, the commonly used architectures are very deep and, hence, prone to overfitting and unfeasible for clinical usages. Inspired by the ideas of the green-AI literature, we here propose a shallow neural network to perform an efficient Learned Post-Processing on images roughly reconstructed by the filtered backprojection algorithm. The results obtained on images from the training set and on unseen images, using both the non-expensive network and the widely used very deep ResUNet show that the proposed network computes images of comparable or higher quality in about one fourth of time.

**Keywords:** Green AI; Sparse-views tomography; Learned Post Processing; CNN; UNet; Tomographic reconstruction

## 1. Introduction

Convolutional Neural Networks (CNNs), with their remarkable capacity of learning with multiple levels of abstraction, are giving new impetus to researchers working on inverse problems and the imaging sector is one of the most involved field [1]. In fact, researchers have begun to tackle inverse imaging applications such as denoising, deconvolution, in-painting, superresolution, and medical image reconstruction with CNNs and they all report significant improvements over state-of-the-art techniques, encompassing sparsity-based models derived from compressed sensing approaches [2,3].



**(a)** Traditional CT **(b)** Sparse-view CT **(c)** Limited-angle CT

**Figure 1.** The three different CT geometric protocols.

In this paper, we focus on the X-ray Computed Tomography (CT) image reconstruction as a representative field of study of challenging inverse problem tasks for imaging. CT,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



in fact, is one of the most exploited diagnostic modality in medical imaging, but the high radiation exposure per patient is unhealthy and may cause cancers Hence the definition and implementation of new safer protocols has become an active and interesting area of research in the inverse imaging community. In particular, the so-called Sparse-view CT (SpCT, or few-view CT) technique is a recent and popular proposal that lowers the radiation dose by reducing the number of X-ray projection views. In traditional CT (Figure 1a) about one thousand projections are executed over the 360-degree trajectory, whereas in the SpCT protocols (Figure 1b) the angular step between two adjacent scans is enlarged. Common sparse protocols consider scanning angular interval of one degree approximately. Furthermore, due to limitations of human anatomy or equipment manufactory, in special cases the X-ray source only walks a semi-circular or C-shape path (as depicted in Figure 1c) and the SpCT configuration is labeled as *limited angle CT*.

Such low-dose tomographic approaches lead to incomplete CT projection data and these subsampled measurements usually produce severe streaking artifacts on the Filtered Back-Projection (FBP) reconstructions. To address this, compressed sensing-based approaches have been investigated in literature, minimizing the Total Variation (TV) or other sparsity-promoting priors combined with data fidelity terms [4–11]. Although the very accurate achievable results, the optimization approach has not been widely adopted yet in clinical setting, for its high computational cost.

As anticipated, the advent of deep learning is revolutionizing how researchers address CT (and, in particular, SpCT) image reconstruction and a number of works have already been published, trying to exploit the deep learning data-adaptivity for reaching high quality medical images [12,13]. To this aim, we now focus on the paradigm sometimes referred to as *Learned Post Processing* (LPP) or *Deep Artifact Correction*, which employs deep neural networks to suppress artifacts on roughly reconstructed images. This framework is graphically represented in Figure 2 for the specific context of SpCT, where the FBP algorithm is typically used to transform the sub-sampled sinogram data into the 2D medical image and the LPP is performed at the end of whole reconstruction workflow, to remove streaking artifacts and noise. To the best of our knowledge, the first proposal of an LPP scheme for

e CNN.png

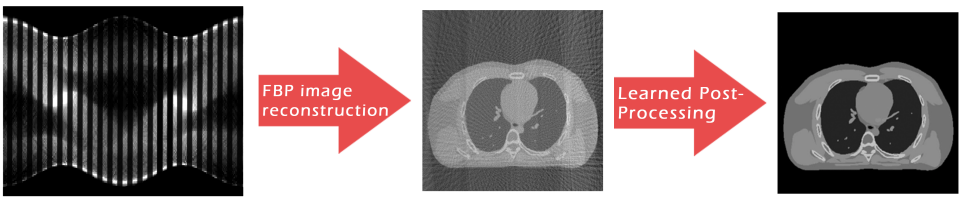



**Figure 2.** Graphical draft of the considered two-step workflow for tomographic reconstruction from sparse-view data.

sparse-view CT dates back to 2016, with the pioneering paper [14] where chest images were restored by an end-to-end CNN, pre-trained to learn the mapping between the FBP and artifact-free images. Later, many works have comprised UNet [15] architectures to fully take advantage from the down-sampling operations in the contracting path, since the FBP reconstruction from subsampled measurements are characterised by streaking (and hence global) artifacts that may be better addressed by CNNs equipped with large receptive fields [16–18]. In addition, residual learning strategies have been proposed to better preserve high texture details which are important as well as difficult to recover during the expansive path of the UNet[3,19,20]. Interestingly, the studies by Han et al. have already demonstrated the superiority of LPP strategies over some TV-based iterative algorithms for sparse-view CT reconstructions [17,21].

On the other hand, two main disadvantages of neural networks limit the effectiveness of LPP aproach. On one side, as highlighted in [22,23], the robustness of neural networks for medical applications is still a concern as they are vulnerable to unseen patterns. For instance, whenever the network takes as input an out-of-domain image, the noise- and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



artifact-free output may contain anatomical structures placed at wrong positions or even fake organ-like structures in the background. On the other side, the very deep structure of UNet requires a very expensive training in terms of time and consumed energy. To handle this constraint, the common choice adopted in the aforementioned papers consists in training the neural networks on small size bi-dimensional images: real medical 2D and 3D images are often too large for nowadays training possibilities.
Intertwined to these drawbacks, the green AI (Artificial Intelligence) line of thought is currently offering a new perspective and an interesting prospective [24–27] which fit also the inverse medical imaging community. In fact, managing and reducing the energy cost of infrastructure and keeping a balance between model accuracy and sustainable computational costs, green AI is in line with medical requirements of real clinical settings.

*Aim and contribution of the paper*

Aim of this paper is to propose a "green" (but still accurate) alternative to the widely used UNet scheme for the LPP reconstruction of CT images. Such choice may have many positive sides. First, looking for solutions saving time and energy is an essential prospective in our society. Secondly, lowering computational times can also reduce the cost of the hardware necessary to training the algorithms, making CT clinical exams and research more accessible. At last, due to the ongoing development of 3D CT imaging and the clinical requirement of almost real-time reconstructions, the forward pass in the LPP scheme must be as fast as possible.

In this scenario, the main contributions of this paper can be resumed as follows. On one side, we propose to use a very light convolutional network to correct artifacts on CT reconstructions from sparse views. The considered CNN allows for a very fast training, that can be adapted to large 2D images and 3D volumes. In particular, differently from the UNet, the considered architecture is composed by only three inner layers and acts in single-scale modality on the input image. Due to its extreme light structure, it is expected not to overfit on the training set. On the other side, we validate the robustness and vulnerability of the proposed learned post-processing not only on a test set but even on out-of-domain cases, i.e. on images with slightly different patterns or statistics than the training samples. Such analysis is unusual in literature, although it is well known that it is important to investigate whether a neural network is vulnerable to perturbations on its input with respect to the training images, to assess the CNN stability or overfitting.

The paper is organized as follows. In section 2 we describe the LPP workflow for CT image reconstruction and we illustrate the networks architecture; in section 3 the numerical experiments are presented and discussed and finally section 4 reports some conclusions.

## 2. Methods and materials

In this section, we present and compare the two neural network architectures we test for artifact removal on tomographic image reconstructions from sparse views. The first one is a residual UNet, labeled as ResUNet in the following; the second scheme is a very simple CNN composed by three layers and working in Single-Scale on the input image, hence it is denoted as 3L-SSNet.

As already mentioned, each proposed CNN is applied on the FBP reconstruction image to correct its artifacts (Figure 2). Formally, if we denote as $y$ the artifact-corrupted image achieved by the FBP and the network output as $\bar{x}$, the Learned Post Processing task can be formulated as:

$$\bar{x} = F_\theta(y) \tag{1}$$

where $F_\theta$ describes the neural network action on the input image $y$ for the final restored image $\bar{x}$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



### 2.1. The ResUNet architecture

State-of-art results in the image processing field have demonstrated that the popular UNet architecture by [15] works particularly well in situations where the input image is corrupted with global artifacts. As a matter of facts, it is known [28] that the pooling/unpooling strategy does permit to enlarge the receptive field of convolutional filters in such a way that, in the lowest inner layers, it is possible to capture global information about the image, whereas in the higher part only the local information are processed. As a consequence, the UNet structure has been elected as the standard/baseline architecture even for sparse-view tomographic imaging tasks, where the streaking artifacts are globally distributed on the image. In such a scenario, in fact, the UNet scheme has been already applied, working on the image [3,17,18] or on a wavelet-based image transformation [29–31].

As observed in the theoretical work [21] by Han et al., in the case of sparse-view CT with FBP reconstruction, it can be proved that the residual manifold which contains the artifacts is easier to learn than the true image manifold. In other words, it could be more effective to learn a residual map

$$\mathcal{M}_R : y \longmapsto y + \bar{x} \tag{2}$$

than the correction map

$$\mathcal{M}_C : y \longmapsto \bar{x} \tag{3}$$

for artifact suppression tasks. For this reason, we consider a residual-learning UNet scheme (labeled as *ResUNet* in the following) for the LPP step. To define it, a skip connection between the input and the output of the network is added to the UNet [32], modifying the image restoration model 1 into the following one:

$$\bar{x} = y + R_\theta(y) \tag{4}$$

which implies that the residual neural network $R_\theta(\cdot)$ must learn the artifacts manifold from $y$.

Figure 3 schematically represents the considered architecture. The ResUNet is a fully convolutional neural network with a symmetric encoder-decoder structure and pooling-unpooling operators to enlarge the receptive field. The pooling operations in the encoder naturally divides the network in distinct levels of resolution, to which we will refer as $l$, $l = 0, \dots, L$, where $L + 1$ is the total number of levels in the network. At each level, a fixed number $n_l$ of convolutional filters is applied, each one with the same number of channels $c_l$, which is constant along the level. Given a baseline convolutional channels number $c_0$ that corresponds to the number of channels in the first level, we will compute $c_l$ for the next levels with the recursive formula $c_{l+1} = 2c_l$, $l = 0, \dots, L-1$. In our specific implementation, we decided to fix $L = 4$, $n_0 = \cdots = n_3 = 3$ and $c_0 = 64$. As already said, the decoder is symmetric to the encoder, with upsampling layers instead of the pooling ones. Moreover, to maintain high frequency information, skip connections are added between the last layer at each level of the encoder and the first layer at the correspondent level in the decoder. To lower the number of parameters, we implement the skip connections as additions instead of the largely used concatenations. A residual connection is added between the input layer and the output layer too, following the implementation described in [3]. Each convolutional layer is composed by a Conv2D + BatchNormalization + ReLU structure, as it is common in literature, except for the last layer, where we used a tanh activation function (as it is necessary to learn a residual map). As it is clear, the ResUNet must learn a high number of parameters, during its training.

### 2.2. The 3L-SSNet architecture

Inspired by green AI ideas, we now consider a very simple architecture inspired by the one used in [33]. It is a three-Layered fully Convolutional Neural Network without

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



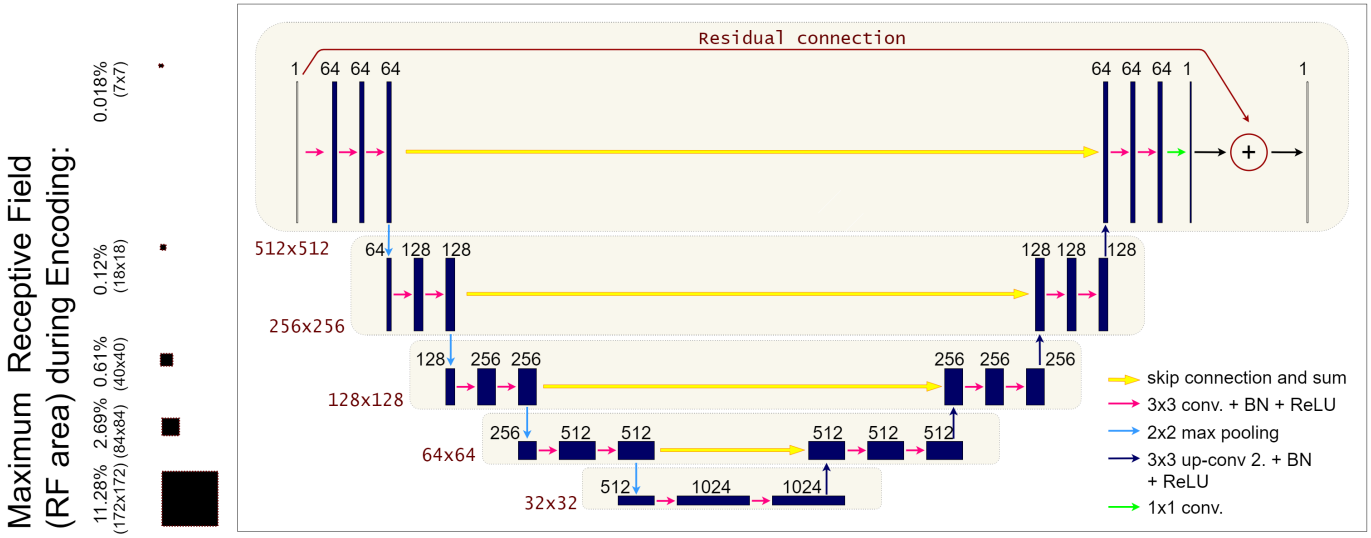



**Figure 3.** On the right: graphical representation of ResUNet architecture; On the left: details on the maximum receptive fields for each of the five levels of the network encoder (RF percentage respect to the input 512x512 image and size of RF).

pooling and unpooling operators, with constant channel number equals to 128 and filter size of dimension $\{9,5,3\}$. Each layer is the common Conv2D + BatchNormalization + ReLU block. A draft of the 3L-SSNet structure is reported in Figure 4.

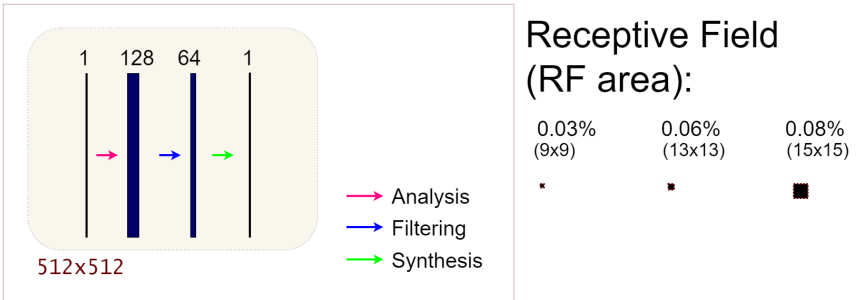



**Figure 4.** On the left: graphical representation of 3L-SSNet architecture; on the right: details on the receptive fields for each of the three layers of the network (RF percentage respect to the input 512x512 image and size of RF). The name of the three layers follows the notation in [33].

### *2.3. Receptive Field*

The portion of the input image $y$ that is captured by each filter at a certain depth in the network is named *receptive field*. Formally, the receptive field of a CNN at a fixed layer $t$ is defined as the portion of the input image $y$ that produces a certain pixel of the feature map at the $t$-th layer [34,35]. Since we are interested in comparing our neural network architectures in term of their receptive field, we need to derive a formula to compute it for a given network.

For each layer $t$, let $k_t$ and $s_t$ be its kernel dimension and the stride, respectively. Moreover, let $r_t$ be the receptive field; the receptive field of the input layer is $r_0 = 1$. The value of $r_t$ can be computed with the recursive formula [34]:

$$r_t = r_{t-1} + A_t \tag{5}$$

where $A_t$ is the non-overlapping area between subsequent filter applications. Note that $A_t$ can be simply computed as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



$$A_t = (k_t - 1) \prod_{i=1}^{t} s_i \tag{6}$$

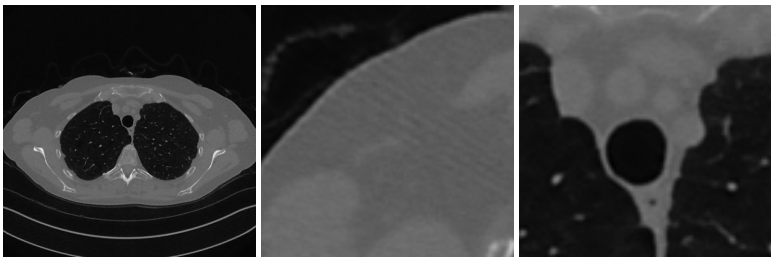
which implies that the receptive field at each $t$-th layer is:

$$\begin{cases} r_0 = 1 \\ r_t = r_{t-1} + (k_t - 1) \prod_{i=1}^{t} s_i \end{cases} \tag{7}$$

Equation (7) shows that the receptive field scales linearly with the depth of the network if the kernel dimension is fixed, while it is exponentially related with the strides. For this reason, reducing the image dimension with pooling operators while processing the image, exponentially enlarges the receptive field of the convolutions.
As reported in Figure 3, ResUNet has the maximum receptive field of $172 \times 172$ pixels, corresponding to the 11.28% of the input image. On the contrary, the 3L-SSNet is significantly smaller, since its receptive field is at most of $15 \times 15$ pixels in its last layer.

### *2.4. Training of the networks*

To train the networks we have used numerical simulations generated from full-sampling CT images provided by the AAPM Low Dose CT Grand Challenge [36]. The downloaded images are $512 \times 512$ pixels and are chest reconstructions from full-dose acquisition data, thus we used them as ground truth images, after scaling them in the interval $[0, 1]$. We used ten patients (3306 images) for the training phase and one patient (357 images) for testing. As visible from Figure 5 where a slice of the test patient is reported in its ground-truth (GT) original version, the considered samples are not noise- nor artifact-free at all. We remark that this feature may lead to some little corruptions on the CNN restored images.



**(a)** **(b)** **(c)**

**Figure 5.** Ground truth image (a) and the considered zooms-in (b) and (c).

Given the training set $\mathbb{D} = \{(y_i, x_i^{GT}\}_{i=1,\dots,N_{\mathbb{D}}}$, where $y_i$ are the input samples to the network, and $x_i^{GT}$ are the correct labels, we train the parameters $\theta$ such that, if $\bar{x}_i = F_\theta(y_i)$ is the restored image given $y_i$, we have

$$\theta^* = \arg\min_\theta \frac{1}{N} \sum_{i=1}^{N} \ell(\bar{x}_i, x_i^{GT}). \tag{8}$$

In our implementation, $\ell(\bar{x}_i, x_i^{GT}) = ||\bar{x}_i - x_i^{GT}||_2^2$. In ResUNet, training is performed by running Stochastic Gradient Descent (SGD) for 50 epochs, with batch size equal to 8 and Nesterov Acceleration with Momentum equal to 0.99. The step size for SGD decreases with polynomial decay, going from $10^{-2}$ to $10^{-5}$ during training. To increase the stability over the first iterations, we clipped the gradient to 5.
In 3L-CNN, the training parameters are exactly the same as for ResUNet, except for the fact that we runned Adam instead of SGD, as we noticed that in that situation SGD got

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



| | Parameters | FLOPs | Training Time |
|---|---|---|---|
| ResUNet | $34.5 \cdot 10^6$ | $406 \cdot 10^9$ | 209 |
| 3L-SSNet | $85 \cdot 10^3$ | $44 \cdot 10^9$ | 53 |

Table 1: Comparison of the cost of the considered networks. The training time is expressed in sec/epoch in the third column.

stucked in a local minimum after a bunch of iterations. The training was performed on two Nvidia GeForce RTX 2080Ti.

### *2.5. Network comparison*

To complete the comparison between the considered ResUNet and 3L-SSNet, we report in Table 1 further useful details. Focusing on the number of parameters and the time for training of each structure, we observe that 3L-SSNet has only 85000 parameters and it requires less than one minute to complete an epoch (corresponding to a quarter of the ResUNet time). The Green AI FLOPs index reflects the faster performance of the 3L-SSNet even in the forward execution to process new images in real-time. The higher computational advantage of the 3L-SSNet network is clearly visible.

## 3. Experimental results and discussion

In this section, we report and discuss the representative experiments performed to test the effectiveness of the considered networks.

We developed our workflow in Python and the code is available at https://github.com/loibo/3LSSNet.

To build the training and testing data sets, we computed the synthetic projection data using ASTRA toolbox [37], providing routines for the forward 2D projections of the ground truth images. To simulate the sparse-view geometry, we considered two different protocols: a full angular acquisition with 1-degree spaced projections (denoted as *full-range* in the following), and a reduced scanning trajectory limited to 180 degrees with 180 projections (denoted as *half-range* in the following). We added to the sinograms white Gaussian noise with $10^{-2}$ noise level and finally we computed the FBP reconstruction by ASTRA routine.

### *3.1. Metrics for image quality assessment*

To evaluate the efficiency and the accuracy of the reconstructed images we consider the following widely used metrics. Given a reconstructed image $x$ of $n$ pixels, we compute its relative error (RE)

$$RE = \frac{\|x - x^{GT}\|_2^2}{\|x^{GT}\|_2^2} \tag{9}$$

and the Peak Signal-to-Noise Ratio (PSNR) index

$$PSNR = 20 \log_{10}\left(\frac{n \cdot max(x^{GT})}{\|x - x^{GT}\|_2}\right). \tag{10}$$

To better evaluate the visual appearance of an image, we also compute the well-known Structural Similarity (SSIM) index [38], measuring the perceptual difference between two similar images, and the Feature Similarity (FSIM) index [39] which should better interpret the low-level features conveying the most crucial information for the human visual system. We remark that $x^{GT}$ has values in $[0, 1]$, whereas the output images $x$ are visualized in the interval $[x_{min}, x_{max}]$.

### *3.2. Results on the test set*

In this paragraph we discuss the results obtained on the test set. We analyse in particular the reconstructions of the slice in Figure 5a, considering the projections acquired with both the full-range and half-range geometries described before. We compare the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



| | RE | PSNR | SSIM | FSIM |
|---|---|---|---|---|
| FBP | 0.9966 | 86.42 (33.89) | 0.2924 | 0.5456 |
| ResUNet | 0.0942 | 106.99 (41.95) | 0.9262 | 0.9709 |
| 3L-SSNet | 0.0840 | 107.92 (42.32) | 0.9480 | 0.9627 |

Table 2: Average of the full-reference metrics on the test set, in case of full-range geometry.

| | RE | PSNR | SSIM | FSIM |
|---|---|---|---|---|
| FBP | 0.9932 | 86.45 (33.90) | 0.2962 | 0.6819 |
| ResUNet | 0.1016 | 106.38 (41.71) | 0.9324 | 0.9478 |
| 3L-SSNet | 0.1309 | 104.34 (40.91) | 0.9021 | 0.9474 |

Table 3: Average of the full-reference metrics on the test set, in case of half-range geometry.

results computed by the FBP algorithm and the LPP images with ResUNet and 3L-SSNet networks.

In Table 2 and 3 we report the average values of the considered metrics in the full-range and half-range cases, respectively. We first remark the very poor values achieved by the FBP, that are motivated by the its difficulty in recovering the actual intensities of the ground truth images. Nevertheless, both the LPP images enhance such quality indices significantly. The 3L-SSNet exhibits better values (except for the SSIM parameter) with full-range, whereas for half-range the ResUNet is outperforming.

In Figure 6 (full-range geometry) and Figure 7 (half-range geometry), we focus our visual inspection on the reconstructions of the slice in Figure 5a. From the crops of Figure 6 we observe that the images learned by the two networks are very similar. The striking artifacts of the FBP reconstruction (Figure 6d) are not completely removed in either image 6e and 6f. The area shown in 6g is mainly corrupted by noise which is well cleaned especially in the reconstruction with ResUNet (Figure 6h).

In Figure 7 we depict the reconstructions obtained with the half-range geometry. In this case, the 3L-SSNet network produces more accurate images. In Figure 7f the striking artifacts are less visible than in Figure 7e. Moreover, the low contrast objects, pointed by the arrows, are more clearly distinguishable and have sharper contours in Figure 7i than in Figure 7h.

### *3.3. Tests on out-of-domain data*

Since it is well known that one critical drawback of neural networks is their performance on unseen data, in this paragraph we test the considered networks on *out-of-domain* data. We apply the algorithms to two different projection sets: the first one is from the patient test data with increased noise with respect to the training set (section 3.3.1); the second one is obtained from a digital image of the XCAT phantom [40], elsewhere used in literature to test neural networks on Xray images [41] (section 3.3.2). In this case, the test problem has been built as for the training set.

#### 3.3.1. Test on unseen noise

We analyse the results of the algorithms on the test set obtained by adding noise with level $2 \cdot 10^{-2}$ to the projections of the test set images.

In Table 4 we report the average RE and SSIM for both the geometries. We observe that with full-range geometry the 3L-SSNet performs better, whereas the ResUNet shows

| | FBP | | ResUNet | | 3L-SSNet | |
|---|---|---|---|---|---|---|
| | RE | SSIM | RE | SSIM | RE | SSIM |
| Full-range | 0.9966 | 0.2526 | 0.0966 | 0.9172 | 0.0896 | 0.9295 |
| Half-range | 0.9932 | 0.0.2567 | 0.0986 | 0.9212 | 0.1162 | 0.8866 |

Table 4: Average values of the full-reference metrics on the test patient image with unseen noise, in the full-range and half-range cases.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



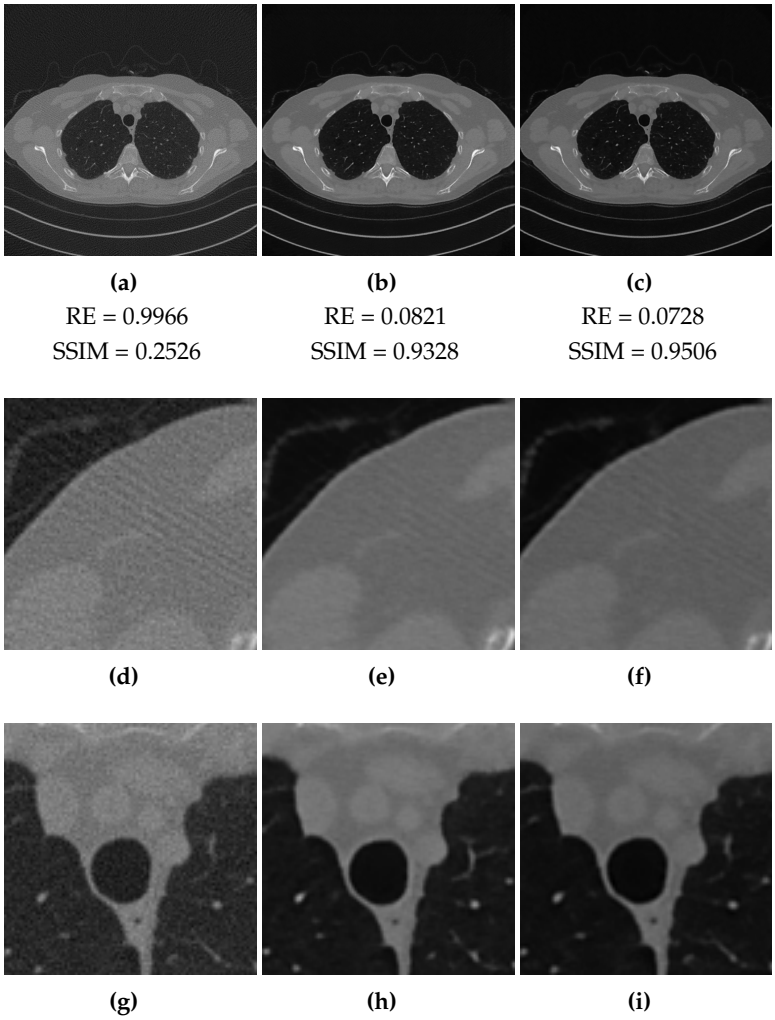



**Figure 6.** Full-range geometry reconstructions. Results obtained with FPB (left column), ResUNet (central column) and 3l-SSNet (right column). Below each image, the values of its RE and SSIM metrics.

superior values in the half-range case. However, even in this case, the visual inspection is not fully consistent with the metrics, since the reconstruction obtained by 3L-SSNet with half-range geometry shows the highest quality. In Figure 8 (half-range geometry), the image 8b learned by 3L-SSNet is less noisy than the crop 8a from ResUNet; in the second zoom, the low contrast objects pointed by the arrow are far more contrasted in the 3L-SSNet reconstruction 8c than in the ResUNet 8d. Moreover, a noisy pattern is still visible inside the dark background of the lungs in Figure 8c, reflecting the difficulty of the residual deep network in handling unseen noise.

### 3.3.2. Test on unseen image

We finally analyse the LPP reconstructions of the XCAT digital image, which is not in the AAPM Low Dose CT Grand Challenge data set. The ground truth image is displayed in Figure 9a, together with two zooms-in in Figure 9e and 9i. We observe that it has different features with respect to training images, since it is completely noise-free and it is constituted by flat regions containing small sharp objects of interest.

We do not report the images obtained considering full-range case, where the 3L-SSNet metrics (RE = 0.0506, SSIM = 0.9213) outperform the ResUNet (RE = 0.0567, SSIM = 0.8503). We focus our analysis on the half-range case, whose results are depicted in Figure 9. In the first crop (images 9e-9h) it is evident that the noise is better suppressed by the 3L-SSNet,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



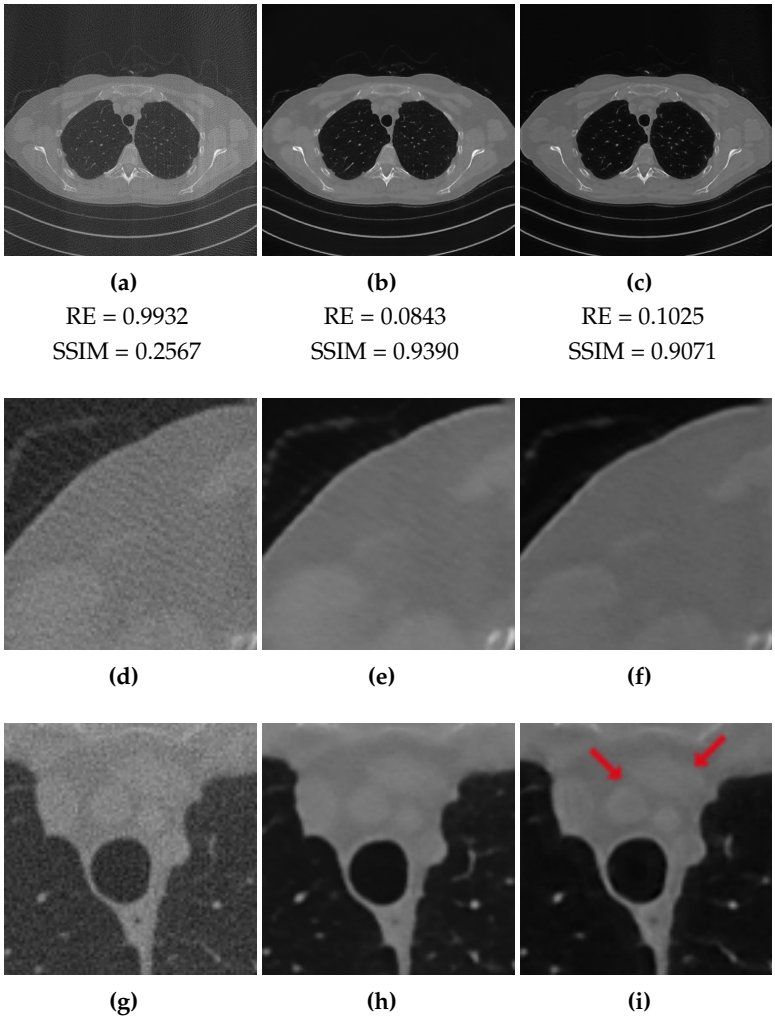



**Figure 7.** Half-range geometry reconstructions. Results obtained with FPB (left column), ResUNet (central column) and 3L-SSNet (right column). Below each image, the values of its RE and SSIM metrics.

which gives images with more uniform areas similar to the GT. Concerning the second zoom-in (images 9i-9l), the artifact (pointed by the arrow in 9k), surprisingly introduced in the ResUNet reconstruction, catches our attention. The darker contour following the border of the chest is not present neither in the Ground Truth nor in the FBP reconstruction.

### *3.4. Discussion*

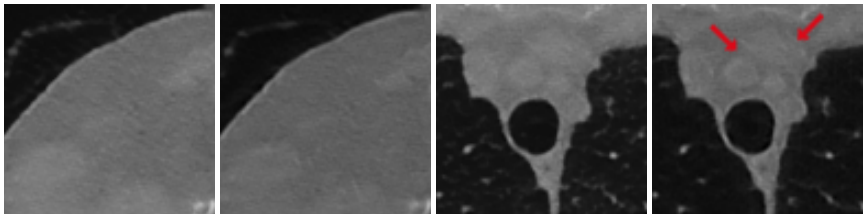
Our numerical results demonstrate the potential of both the ResUNet and the 3L-SSNet in correcting the FBP reconstructions, which are affected by severe corrupting effects and lack of contrast. In particular, the two networks give comparable results in terms of metrics and image quality, when applied on test data coherent to the training samples. The comparison between Table 2 and 3 highlights the ResUNet superiority in case of half-range geometry, where less projections are available, but the visual inspection of Figures 6 and 7 reveals very similar reconstructions in all the shown images and zooms.

The artifact correction by ResUNet becomes less effective than 3L-SSNet when processing out-of-domain data characterized by features different from those learned from the training set, such as the image dealt with in Section 3.3. It reflects the trend of very deep neural networks to overfit on the learned image patterns.

In general, even if the receptive field of the 3L-SSNet is extremely smaller than the ResUNet one, its 15x15 RF area is big enough to discern the SpCT artifacts (due to the FBP

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1





(a) (b) (c) (d)

**Figure 8.** Crops of the reconstructions of the test patient with unseen noise and half-range geometry. ResUNet in (a) and (c), 3L-SSNet in (b) and (d).

reconstruction) from the specific patterns of the ground truth images. We think that this explains why the 3L-SSNet post-processing has comparable effects to the ResUNet ones.

## 4. Conclusions

In this paper we proposed 3L-SSNet, a non-intensive computation neural network for a Learned Post Processing reconstruction algorithm in CT. The proposal fits with the Green AI research, which studies computationally cheap algorithms to save energy and to be inclusive. Moreover, in CT setting, reducing time as far as possible is important to make the algorithms usable in clinics.

The results obtained by 3L-SSNet on in-domain images (i.e. images belonging to the data set used for network training) are comparable, for metrics and visual inspection, to the output of ResUNet, a widely used very deep architecture.

We alsoi tested both networks on out-of-domain images, i.e. images with noise or features not seen in the training phase, and we surprisingly got reconstructions from 3L-SSNet sometimes more accurate than with ResUNet. The deep ResUNet, besides requiring computational time four times greater than 3L-SSNet, does not handle the unseen features efficiently.

Motivated by these very good results, we intend to test other Green networks for reducing CT artifacts in possibly different reconstruction frameworks. Moreover, a 3L-SSNet shallow-like network can be tested for artifact correction in other inverse problems in imaging, such as deblur or super resolution.

**Funding:** The research was funded by the Indam GNCS grant 2020 *Ottimizzazione per l'apprendimento automatico e apprendimento automatico per l'ottimizzazione.*

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** The authors would like to thank Prof. G. Marfia, head of the VARLAB of the University of Bologna, for kindly allowed to use the computers to execute the numerical experiments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arridge, S.; Maass, P.; Öktem, O.; Schönlieb, C.B. Solving inverse problems using data-driven models. *Acta Numerica* **2019**, *28*, 1–174. doi:10.1017/S0962492919000059.
2. McCann, M.T.; Jin, K.H.; Unser, M. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine* **2017**, *34*, 85–95.
3. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **2017**, *26*, 4509–4522.
4. Graff, C.; Sidky, E. Compressive sensing in medical imaging. *Appl. Opt.* **2015**, *54*, C23–C44.
5. Tian, Z.; Jia, X.; Yuan, K.; Pan, T.; Jiang, S.B. Low Dose CT Reconstruction via Edge-preserving Total Variation Regularization. *Phys. Med. Biol.* **2011**, *56*, 5949–5967. doi:10.1088/0031-9155/56/18/011.Low.
6. Jensen, T.L.; Jørgensen, J.H.; Hansen, P.C.; Jensen, S.H. Implementation of an optimal first-order method for strongly convex total variation regularization. *BIT Numer. Math.* **2012**, *52*, 329–356.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2021 doi:10.20944/preprints202107.0265.v1



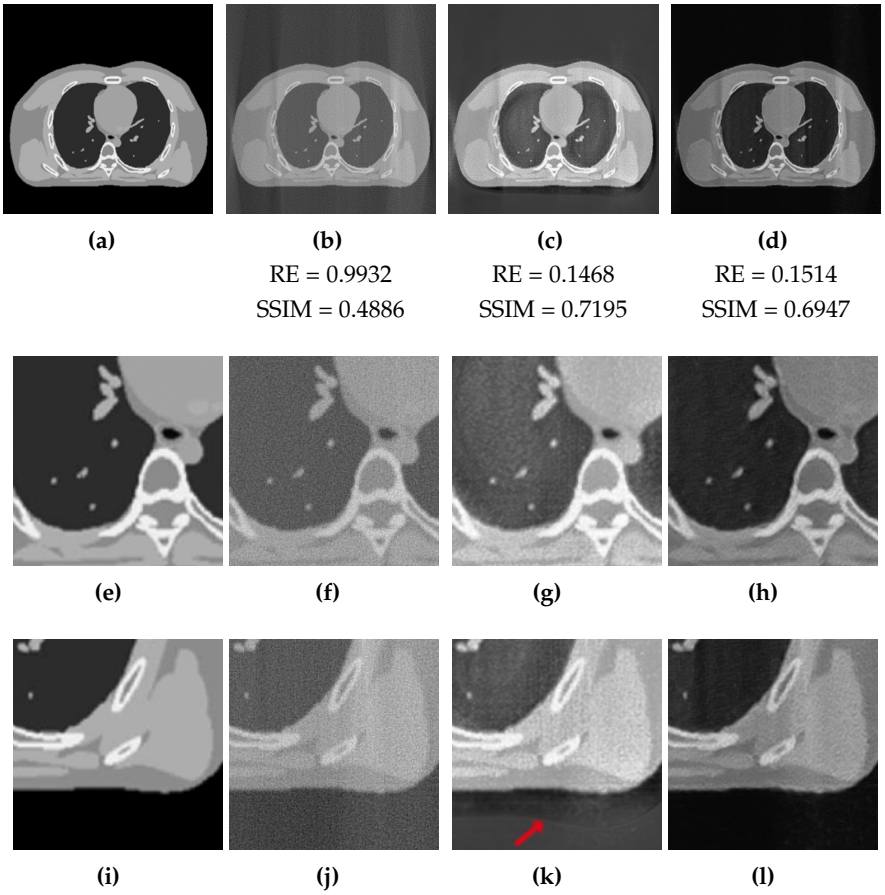



**Figure 9.** XCAT phantom test image with half-range geometry. From the left to the right. First column: Ground Truth image (a) and the considered zooms-in (e) and (i). Reconstructions from half-range geometry with FBP (second column), ResUNet (third column) and 3L-SSNet (fourth column).


7. Sidky, E.; Chartrand, R.; Boone, J.; Pan, X. Constrained T p V-minimization for enhanced exploitation of gradient sparsity: application to CT image reconstruction **2013**. pp. 1–15.
8. Liu, L. Model-based Iterative Reconstruction: A Promising Algorithm for Today's Computed Tomography Imaging. *Journal of Medical Imaging and Radiation Sciences* **2014**, *45*, 131 – 136. doi: https://doi.org/10.1016/j.jmir.2014.02.002.
9. Loli Piccolomini, E.; Morotti, E. A Model-Based Optimization Framework for Iterative Digital Breast Tomosynthesis Image Reconstruction. *Journal of Imaging* **2021**, *7*. doi:10.3390/jimaging7020036.
10. Rantala, M.; Vanska, S.; Jarvenpaa, S.; Kalke, M.; Lassas, M.; Moberg, J.; Siltanen, S. Wavelet-based reconstruction for limited-angle X-ray tomography. *IEEE transactions on medical imaging* **2006**, *25*, 210–217.
11. Purisha, Z.; Rimpeläinen, J.; Bubba, T.; Siltanen, S. Controlled wavelet domain sparsity for x-ray tomography. *Measurement Science and Technology* **2017**, *29*, 014002. doi:10.1088/1361-6501/aa9260.
12. Zhang, H.M.; Dong, B. A review on deep learning in medical image reconstruction. *Journal of the Operations Research Society of China* **2020**, pp. 1–30.
13. Ahishakiye, E.; Van Gijzen, M.B.; Tumwiine, J.; Wario, R.; Obungoloch, J. A survey on deep learning in medical image reconstruction. *Intelligent Medicine* **2021**. doi: https://doi.org/10.1016/j.imed.2021.03.003.
14. Zhang, H.; Li, L.; Qiao, K.; Wang, L.; Yan, B.; Li, L.; Hu, G. Image prediction for limited-angle tomography via deep learning with convolutional neural network. *arXiv preprint arXiv:1607.08707* **2016**.

15. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
16. Li, H.; Mueller, K. Low-dose CT streak artifacts removal using deep residual neural network. Proceedings of Fully 3D conference, 2017, Vol. 2017, pp. 191–194.
17. Han, Y.; Ye, J.C. Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE transactions on medical imaging* **2018**, *37*, 1418–1429.
18. Zhang, T.; Gao, H.; Xing, Y.; Chen, Z.; Zhang, L. DualRes-UNet: Limited Angle Artifact Reduction for Computed Tomography. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, pp. 1–3.
19. Han, Y.; Ye, J.C. Deep residual learning approach for sparse-view CT reconstruction. Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. Fully3D conference organization, 2017.
20. Schnurr, A.K.; Chung, K.; Russ, T.; Schad, L.R.; Zöllner, F.G. Simulation-based deep artifact correction with convolutional neural networks for limited angle artifacts. *Zeitschrift für Medizinische Physik* **2019**, *29*, 150–161.
21. Han, Y.S.; Yoo, J.; Ye, J.C. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis. *arXiv preprint arXiv:1611.06391* **2016**.
22. Huang, Y.; Würfl, T.; Breininger, K.; Liu, L.; Lauritsch, G.; Maier, A. Some investigations on robustness of deep learning in limited angle tomography. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 145–153.
23. Liu, C.; Huang, Y.; Maier, J.; Klein, L.; Kachelrieß, M.; Maier, A. Robustness Investigation on Deep Learning CT Reconstruction for Real-Time Dose Optimization. *arXiv preprint arXiv:2012.03579* **2020**.
24. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green ai. *Communications of the ACM* **2020**, *63*, 54–63.
25. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243* **2019**.
26. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Modern Deep Learning Research. *Proceedings of the AAAI Conference on Artificial Intelligence* **2020**, *34*, 13693–13696. doi:10.1609/aaai.v34i09.7123.
27. Asperti, A.; Evangelista, D.; Piccolomini, E.L. A Survey on Variational Autoencoders from a Green AI Perspective. *SN Computer Science* **2021**, *2*, 1–23.
28. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT press, 2016.
29. Kang, E.; Min, J.; Ye, J.C. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical physics* **2017**, *44*, e360–e375.
30. Ye, J.C.; Han, Y.; Cha, E. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences* **2018**, *11*, 991–1048.
31. Bubba, T.A.; Kutyniok, G.; Lassas, M.; Maerz, M.; Samek, W.; Siltanen, S.; Srinivasan, V. Learning the invisible: a hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Problems* **2019**, *35*, 064002.
32. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* **2017**, *26*, 3142–3155. doi:10.1109/TIP.2017.2662206.
33. Chen, H.; Zhang, Y.; Zhang, W.; Liao, P.; Li, K.; Zhou, J.; Wang, G. Low-dose CT via convolutional neural network. *Biomedical optics express* **2017**, *8*, 679–694.
34. Le, H.; Borji, A. What are the Receptive, Effective Receptive, and Projective Fields of Neurons in Convolutional Neural Networks? *arXiv e-prints* **2017**, p. arXiv:1705.07049, [arXiv:cs.CV/1705.07049].
35. Araujo, A.; Norris, W.; Sim, J. Computing Receptive Fields of Convolutional Neural Networks. *Distill* **2019**. https://distill.pub/2019/computing-receptive-fields, doi:10.23915/distill.00021.
36. McCollough, C. TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge. *Medical physics* **2016**, *43*, 3759–3760.
37. Van Aarle, W.; Palenstijn, W.J.; De Beenhouwer, J.; Altantzis, T.; Bals, S.; Batenburg, K.J.; Sijbers, J. The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **2015**, *157*, 35–47.
38. 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* **2004**, *13*, 600–612.

39. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing* **2011**, *20*, 2378–2386. doi: 10.1109/TIP.2011.2109730.
40. Segars, W.P.; Sturgeon, G.; Mendonca, S.; Grimes, J.; Tsui, B.M. 4D XCAT phantom for multi-modality imaging research. *Medical physics* **2010**, *37*, 4902–4915.
41. Russ, T.; Goerttler, S.; Schnurr, A.K.; Bauer, D.F.; Hatamikia, S.; Schad, L.R.; Zöllner, F.G.; Chung, K. Synthesis of CT images from digital body phantoms using CycleGAN. *International journal of computer assisted radiology and surgery* **2019**, *14*, 1741–1750.