Mattias P. Heinrich, Maik Stille, and Thorsten M. Buzug

Residual U-Net Convolutional Neural Network Architecture for Low-Dose CT Denoising

https://doi.org/10.1515/cdbme-2018-0072

Abstract: Low-dose CT has received increasing attention in the recent years and is considered a promising method to reduce the risk of cancer in patients. However, the reduction of the dosage leads to quantum noise in the raw data, which is carried on in the reconstructed images. Two different multilayer convolutional neural network (CNN) architectures for the denoising of CT images are investigated. ResFCN is based on a fully-convolutional network that consists of three blocks of 5×5 convolutions filters and a ResUNet that is trained with 10 convolutional blocks that are arranged in a multi-scale fashion. Both architectures feature a residual connection of the input image to ease learning. Training images are based on realistic simulations by using the XCAT phantom. The ResUNet approach shows the most promising results with a peak signal to noise ratio of 44.00 compared to ResFCN with 41.79.

Keywords: Denoising, Low Dose CT, Phantom Simulation, Deep Learning, CNN.

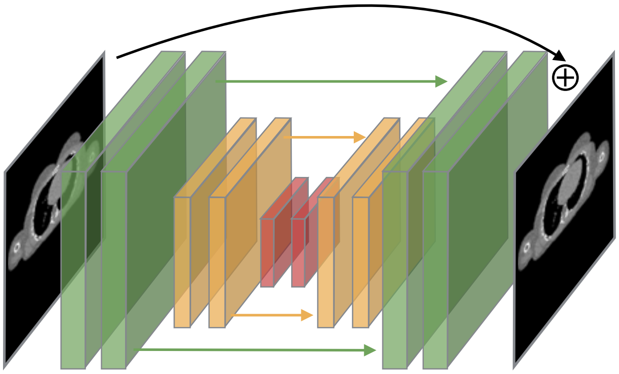

Fig. 1: The employed architecture of residual U-Net for CT denoising. The network learns a powerful medical image prior and derives features using three resolution scales: green (high-resolution, local details), yellow (mid-resolution, regional context), red (low-resolution, global information). Skip connections and a residual forward carrying of the input data improves learning and reconstruction quality.

1 Motivation and Related Work

In order to reduce the risk of cancer through radiation in CT imaging, a reduction of the dose is considered [19]. Dose reduction is generally achieved by lowering the operating current of the X-ray tube. However, the reduced dosage results in quantum noise due to the limited number of photons that are collected by the detector. The noise in the measured projection values thus leads to noise in the reconstructed images [17]. Consequently, the diagnostic value of the images is reduced.

Noise reduction in CT images is a highly active field of research. In low-dose CT, the insufficient number of photons in the projection domain causes noise that does not obey a uniform distribution [10]. Current approaches are based on an iterative reconstruction scheme, sinogram filtering, or image processing [11, 12, 16, 20]. Commonly, the limited amount of data poses an ill-posed problem for denoising algorithms. The assumption of a certain sparsity in natural and medical images leads to the use of edge-preserving filtering methods. Yet, automatically differentiating between small but relevant anatomical structures may lead to insufficient results in traditional methods that are based on mathematical models and neighbourhood information. Deep learning has undergone enormous development in recent years. In particular, in the field of image recognition astonishing performance level that surpasses human raters have been reached using dozens of learned 3×3 filter layers interconnected with residual connections [7].

In this work, we explore two convolutional neural network (CNN) architectures capable of learning a nonlinear mapping that produces high-quality multi-layer denoising results. The training is based on very realistic simulations that are performed based on the XCAT phantom [14, 15] with different photon counts. In contrast to simple additive Gaussian noise in image space, our approach that disturbs the measurements' sinograms can be trained to denoise images that are affected globally. The paper is organised as follows: previous and related work in the field of image denoising for natural and medical images is discussed in the next section. In Section 2, we describe our method for generating ground truth training data based on realistic simulations and the two compared deep CNN architectures used to learn a denoising filtering algorithm. We compare experimental results for a hold-out val-

Mattias P. Heinrich, Institute of Medical Informatics, University of Lübeck, Lübeck, Germany, e-mail: heinrich@imi.uni-luebeck.de
Maik Stille, Thorsten M. Buzug, Institute of Medical Engineering, University of Lübeck, Lübeck, Germany
Mattias P. Heinrich, Maik Stille, these authors contributed equally.

idation set using either a straightforward fully-convolutional architecture that only performs local 5×5 convolutions and the multi-scale U-Net architecture (see Fig. 1) that has demonstrated advantages for contextual learning for semantic segmentations [13]. Next, we analyse the influence of residual connections that have been proposed recently for improved learning for the related image reconstruction tasks of single-image superresolution [9].

1.1 Image Denoising

Several algorithms have been proposed in recent years to deal with problems of noise in image processing. Early approaches were motivated by a simple yet powerful mathematical model that penalises strong gradients in the denoised signal by its squared norm through a diffusion regularisation [6]. In order to preserve edges, a local or non-local neighbourhood weighting based on intensity differences can be included. Relying on single pixels for estimating regularisation weights is unstable for strong image noise, newer work employs patch distances, as e.g. done in [4] and [1], which can already be seen as a precursor to convolution filters. In [3] a nonlinear diffusion model is proposed that is able to learn all relevant parameters, including filters and activation functions, directly from training data by minimising a loss function. Despite following almost all conventions of convolutional neural networks, [3] was not considered directly as deep learning approach. Zhang et al. [18] consequently presented a deep fully-convolutional network with a residual connection that will serve as the baseline in our experiments. A similar architecture with additional residual connections and formulated as encoder/decoder architecture has been used in [2] for the denoising of low-dose CT. Another approach that is closely related to our work is [8], which use a similar U-Net architecture for iterative sparse-view reconstruction.

2 Methods

In the following, we will describe the preparation of training data, using the XCAT phantom and details on the simulated noise will be given. Subsequently, we describe the two considered deep convolutional neural network architectures, their parameterisation, and the employed training process.

2.1 Anatomical CT Phantom

Four different software phantoms were generated using the XCAT software [14, 15]. Each phantom consists of 151 slices with a slice thickness of 0.2 cm, pixel width of 0.25 cm and an array size of 256×256. The phantoms show the upper abdomen of a female body and differ in various anatomical details such as the size of the heart, the radial scale of the arms, the size and transversal angle of the breasts and the thickness of the small and large intestine wall. Furthermore, the thickness of scapula, ribs, humerus, and backbone is varied in each phantom. The simulated radionuclide energy is equal to 120 keV. In order to gain reasonable ground truth images, a forward projection of the generated image data is calculated with 1472 detector elements and 1152 views over 360°. Afterwards, images were reconstructed using the filtered back projection and used as ground truth. Ethical approval: The conducted research is not related to either human or animals use.

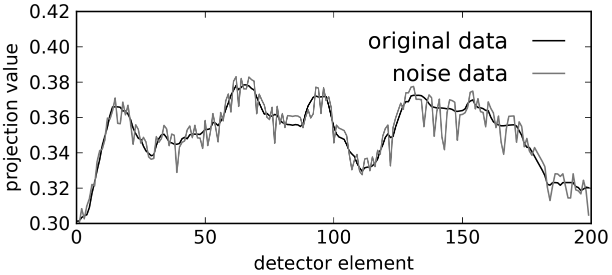

Fig. 2: Line profile through generated forward projection with simulated acquisitions affected by Poisson noise.

2.2 Simulated Noise

In order to simulate noise in the CT images, forward projections (1472 detector elements and 1152 views over 360°) of the generated slices are calculated. The resulting projection values are converted into intensity values according to

$$I_i = e^{-p_i}, i \in \mathcal{M}, \tag{1}$$

where $\mathcal{M}$ is the set of projection indices, p is the simulated projection value and I is the intensity value. Based on the intensity values, Poisson noise is generated and added to the values. Here, the degree of error is a non-linear function of the total attenuation p whereas the signal-to-noise ratio (SNR) is given by [5]

$$\mathrm{SNR}_I \sim \sqrt{I}. \tag{2}$$

The final projection values were generated by calculating the negative logarithm of the noise affected intensity values. The resulting values were used in a filtered backprojection in order to reconstruct the noisy images. An example of the generated noise profile in comparison to the underlying ground truth data is shown in Fig. 2.

2.3 Deep Learning Architectures for CT Denoising

We implemented two deep CNN architectures with residual connections for denoising CT scans in image domain.

ResFCN: First, a straightforward fully-convolutional network similar to [18] that consists of three blocks of 32 5×5 convolutions filter kernels each, followed by batch normalisation and rectified linear unit (ReLU) activations. The final layer uses a 1×1 convolution to map the 32 features to an output image that is formed in addition to the input (residual connection). Therefore, the whole network comprises approx. 50’000 trainable parameters.

ResUNet: Second, we explore a more powerful U-Net architecture [13] that is designed to contain 10 convolutional blocks (again including batch norm and ReLU) that are arranged in a multi-scale fashion as shown in Fig. 1. All convolutions are specified as 3×3 kernels, except the final one and the ones in the lowest resolution (red blocks), which are 1×1 to reduce the parameter count. Every second convolutional block is followed by a 2×2 average pooling that reduces the resolution in the contracting part or a bilinear upsampling operation that doubles the resolution in the expanding path (none of these layers contain free parameters). We start with 16 feature channels that are doubled in each scale. Detail information is conserved within the network by connecting the output of each scale in the left part as concatenation to the corresponding block in the right part (shown as arrows in Fig. 1). At the same time, a large receptive field that processes regional or even global context can be realised within the red convolution blocks. Again a residual connection is added to the last layer to ease learning. In total, our **ResUNet** contains 150’000 learned parameters.

Tab. 1: Numerical evaluation on hold-out validation dataset. The PSNR (peak signal to noise ratio) shows a clear advantage for the **ResUNet** architecture.

Method	Root Mean Squared Error	PSNR
without denoising	4.25×10^{-3}	35.03
ResFCN (50k params.)	1.95×10^{-3}	41.79
ResUNet (150k params.)	1.52×10^{-3}	44.00

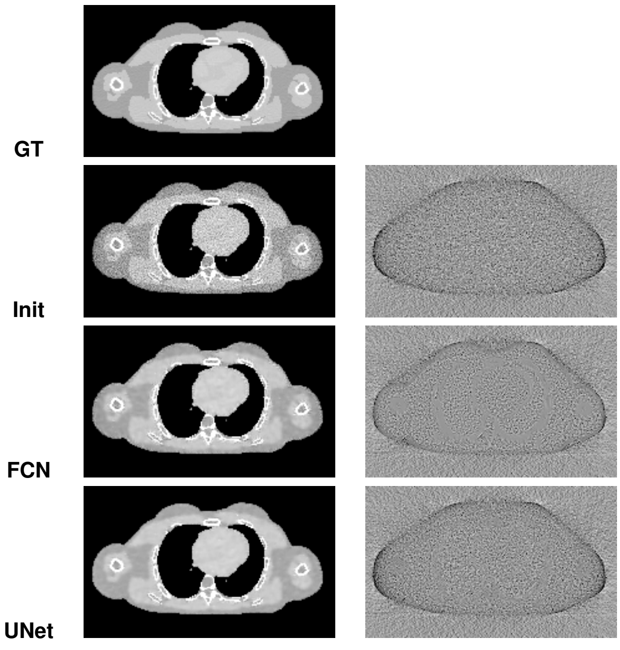

Fig. 3: Example axial slice of ground truth validation data (GT). Simulated reconstruction and initial noise level (Init). Denoising result and removed noise using ResFCN. A visually much better outcome is achieved using the ResUNet.

3 Experiments and Results

We train our model using randomly sampled patches of size 64×64 and a batch-size of 32. The input to the network is based on the simulated noisy slices as detailed in Sec. 2.2. The output is compared to the corresponding noise-free ground truth patch using an L1-norm loss that sums the absolute norm of pixelwise differences. The Adam optimiser with a learning rate of 0.001 and no weight or rate decay is used for 30 epochs with 256 iterations each. We evaluate the performance of the model on a hold-out validation set also simulated using the XCAT phantom software. The results are shown in Table 1 using the average RMSE (root mean squared error) and the PSNR (peak signal to noise ratio) that is based on the RMSE. The intensity range of our data that is represented by attenuation correction values falls between 0.0 and 0.24, resulting therefore also in small numerical errors, hence the PSNR may be more expressive.

Despite containing several thousand adjustable parameters, our convolutional networks can be trained in few minutes using less than 500 MBytes of graphics RAM. A full 3D scan can be processed using the learned model in few second. When examining the numerical, quantitative results in Table 1 it becomes clear, that the multi-scale **ResUNet** achieves better noise reduction reducing the RMSE to 1.93×10^{-3}, more than 50% less than the simpler fully-convolutional model. The visual comparison between **ResFCN** and **ResUNet** in Fig. 3 shows less noticeable structural information in the difference image of the latter. At the same time, the **ResUNet** result is able to preserve more edge details and small lung vessels.

4 Conclusion

Two convolutional neural network architectures that are capable of learning a nonlinear mapping for inverse problems in CT imaging are investigated. Multi-layer CNN denoising models are learned from on a set of training images, which are simulated using the XCAT phantom. In order to simulate a realistic noise model, Poisson noise is added in the sinogram domain. Both implemented architectures, a straightforward fully-convolutional network **ResFCN** and a more powerful U-Net architecture **ResUNet**, feature a residual connection for improved denoising. Results show that **ResUNet** is able to outperform **ResFCN** with a peak signal to noise ratio of 44.00 and 41.79, respectively.

Author Statement

Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The conducted research is not related to either human or animals use

References

[1] A Buades, B Coll, and J-M Morel. A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, volume 2, pages 60–65. IEEE, 2005.

[2] H Chen, Y Zhang, M K Kalra, F Lin, Y Chen, P Liao, J Zhou, and G Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on Medical Imaging*, 36(12):2524–2535, 2017.

[3] Y Chen and T Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2017.

[4] K Dabov, A Foi, V Katkovnik, and K Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[5] B De Man. *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. PhD thesis, 2001.

[6] G Gilboa and S Osher. Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling & Simulation*, 6(2):595–630, 2007.

[7] K He, X Zhang, S Ren, and J Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] KH Jin, MT McCann, E Froustey, and M Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[9] J Kim, Jung Kwon L, and K Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[10] Y Liu, Q Zhang, and ZG Gui. Noise reduction for low-dose CT sinogram based on fuzzy entropy. *Journal of Electronics & Information Technology*, 35(6):1421–1427, feb 2014.

[11] A Manduca, L Yu, JD. Trzasko, N Khaylova, JM. Kofler, CM. McCollough, and JG. Fletcher. Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Medical Physics*, 36(11):4911–4919, oct 2009.

[12] P J. La Rivière. Penalized-likelihood sinogram smoothing for low-dose CT. *Medical Physics*, 32(6Part1):1676–1683, may 2005.

[13] O Ronneberger, P Fischer, and T Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4d XCAT phantom. *Medical Physics*, 35(8):3800–3808, jul 2008.

[15] W. P Segars and B M. W. Tsui. MCAT to XCAT: The evolution of 4-d computerized phantoms for imaging research. *Proceedings of the IEEE*, 97(12):1954–1968, dec 2009.

[16] J Wang, Z Liang, and H Lu. Multiscale penalized weighted least-squares sinogram restoration for low-dose X-Ray computed tomography. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, aug 2006.

[17] Y Yamada, M Jinzaki, T Hosokawa, Y Tanami, H Sugiura, T Abe, and S Kuribayashi. Dose reduction in chest CT: Comparison of the adaptive iterative dose reduction 3D, adaptive iterative dose reduction, and filtered back projection reconstruction techniques. *European Journal of Radiology*, 81(12):4185–4195, dec 2012.

[18] K Zhang, W Zuo, Y Chen, D Meng, and L Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[19] X Zhu, J Yu, and Z Huang. Low-dose chest CT: Optimizing radiation protection for patients. *American Journal of Roentgenology*, 183(3):809–816, sep 2004.

[20] Y Zhu, M Zhao, Y Zhao, H Li, and P Zhang. Noise reduction with low dose CT data based on a modified ROF model. *Optics Express*, 20(16):17987, jul 2012.