Denoising of X-ray projections and computed tomography images using convolutional neural networks without clean data

Paolo Gnudi, Bernd Schweizer, Marc Kachelrieß, and Yannick Berker

***Abstract*—In X-ray projections and computed tomography (CT) imaging, image quality depends on the patient's exposure during the scan. Reducing the exposure reduces patients' health risks, but also reduces image quality due to higher noise in the image. A denoising technique preserving image features enables the acquisition of low-dose images without losing too much information. In this work, the performance of convolutional neural networks in denoising X-ray projections and CT images was evaluated. The peculiarity of the network is that it does not require the usage of clean data for training: instead, two or more independent noise realizations of each image are input during training. The network output was compared with that of a network trained using clean images, as well as images filtered using three conventional filters, namely Gaussian, Median and Bilateral. The results of this work indicate that the absence of clean images during training does not prevent the network from learning a good denoising model. In all examined cases, the performance of the deep-learning approaches provided better results than the selected filters. It is also shown that, whether clean images were used during training or not, each model has a denoising validity range related to the noise level of the training data.**

***Index Terms*—Convolutional neural networks, CT denoising, X-ray denoising, medical image denoising.**

I. Introduction

NOISE is present in virtually all types of acquired images, regardless of whether they are photographs or medical images, reducing the visibility of low-contrast objects. This loss of information can become a serious issue, especially in medical imaging. Some effort has been targeted at reducing noise by optimizing acquisition parameters: for example, using longer acquisition times usually increases the signal-to-noise ratio, but is not always practical in medical imaging due to issues such as patient comfort and the higher operating costs. In some medical imaging techniques, such as X-ray and CT imaging, a longer acquisition time also exposes the patient to a higher quantity of radiation, which should be avoided.

Various software techniques, such as image filtering, have been developed to reduce noise in medical imaging, so that noise removal has become an essential practice in medical imaging applications. Among all denoising techniques, the most recent ones are based on machine learning (ML), having shown very promising performance in image denoising, often superior to image filtering approaches [1, 2]. In this work, we compare the performance of different denoising approaches when applied to X-ray projections and CT images. These approaches comprise two ML techniques as well as three conventional denoising filters (Gaussian, Bilateral and Median). The first machine learning approach tested in this work follows the so-called "Noise2Clean" approach, requiring pairs of noisy and clean images during training in order to accomplish the denoising task. The second machine-learning approach ("Noise2Noise") was recently proposed by Lehtinen et al. [2]. The innovation of this approach is that it does not require clean images to learn how to restore noisy images.

The aim of the work is to evaluate which of these denoising tools is most suitable for X-ray applications. We investigated if, as expected, the ML approaches show superior performance than image filtering. The results of the noise-to-noise and noise-to-clean approaches were compared, and the impact of two loss functions (MSE and MAE) on the noise-to-noise results was investigated. Moreover, the range of validity of the learned denoising model, in dependence of the noise level of the training data, was investigated.

II. Theoretical Background

A. U-Net architecture

The machine-learning approaches used in this work are based on the U-Net, a particular kind of convolutional neural network (CNNs) that has been effective in noise removal [2, 3] although originally proposed for biomedical (semantic) image segmentation [4]. The architecture can be divided in an encoding and a decoding part.

In the encoding part, the information is encoded via spatial dimension reduction, while in the decoding part the information is decoded via spatial dimension increase. This structure produces levels where the inputs of the encoder part have the same dimension as the outputs of the decoder part.

One peculiarity of U-Nets are the skip connections between the encoding outputs and the decoding inputs of the same level. The concatenation between inputs and outputs gives U-

Manuscript received May 1, 2020.

P. Gnudi was, and B. Schweizer is with the Faculty of Engineering, RheinMain University of Applied Sciences, Am Brückweg 26, 65428 Rüsselsheim, Germany (e-mail: Bernd.Schweizer@hs-rm.de).

P. Gnudi and Y. Berker were, and M. Kachelrieß is with the Division of X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany (e-mail: marc.kachelriess@dkfz.de).

Y. Berker is with the Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany (e-mail: yannick.berker@dkfz.de).

Nets the ability to maintain high-level as well as low-level features at the same time.

B. Noise2Clean and Noise2Noise approaches

A denoising model can be obtained by training a CNN regression model with a large number of training pairs $(\hat{x}; x)$ of noisy inputs $\hat{x}$ and clean images x. In such cases, the training inputs are the noisy images and the training targets are the clean images (Noise2Clean). The network is trained to output an image matching the clean input image as well as possible, as measured by a loss function, learning to denoise a noisy image by looking at both noisy and clean images. One of the problems in this approach is to find a sufficient number of clean targets to train the model.

Lehtinen et al. demonstrated that this denoising task could be accomplished also without clean images [2]. They trained a U-Net with noisy images as training instances and another noisy realization of the training instances as targets (Noise2Noise). In such a situation, the network is forced to learn to reproduce the second noisy realization, which is an impossible task because of the random behavior of the noise. Still, through a large number of training samples, the network learns to reproduce an average representation instead, which yields the desired denoising effect. In their work, the authors showed that, for noise removal in magnetic resonance imaging, the results produced with Noise2Noise are comparable with the ones produced with Noise2Clean.

The Noise2Noise technique is particularly suitable to train a denoising model for medical applications, because sometimes it can be challenging to obtain enough clean training targets. The Noise2Noise approach solves this problem not least because with three noisy realizations of the same image, six training pairs can be produced. In general, if N is the number of noisy realizations per image and I is the total number of available images, the number of possible training pairs is $I \cdot (N \cdot (N - 1))$.

III. Materials And Methods

For both X-ray and CT imaging data sets described in the following, we trained different Noise2Noise and Noise2Clean models by changing the loss function used during training and the targets in the training set, all the while using the same U-Net structure.

We used the peak signal-to-noise ratio (PSNR) as a metric to compare the output images to their clean references. On each test set, we also applied Gaussian, Bilateral and Median filters in order to compare their denoising capabilities with those of the ML approaches. For each filter, the best setting was determined image by image in order to ensure the best filter performance. To test the limits of the models, for both imaging techniques, we investigated the output noise level as a function of various input noise levels different from the noise level of the training data.

A. Simulated X-ray dataset

Clean images of the X-ray dataset were simulated projections, produced using the Extended Cardiac Torso (XCAT) simulation framework, a command line application capable of producing realistic digital human phantoms [5]. Noisy images were obtained from the simulated clean projections by computing two independent Poisson noise realizations. In total, we simulated 100 digital human phantoms with varying organ geometry and, for each phantom, projections at 63 different projection angles. In order to avoid different projections of the same phantom to be in the training set and the test set, the data were divided by phantom into 80 training and 20 test phantoms. This produced a training set composed by 10,080 images and a test set of 1,000 images. Table I shows the trained networks together with the adopted loss function for this dataset.

TABLE I
Networks Trained on X-ray Projections

Trained network	Loss function
Noise2Noise	MSE
Noise2Noise	MAE
Noise2Clean	MSE

Before being fed to the network, all noisy and clean slices were rescaled roughly into the [0, 1] range: clean slices were rescaled by dividing all pixels in an image by the maximum pixel value of the image itself. The noisy images were rescaled by dividing each pixel by the maximum of the respective clean image. This produced noisy images where some pixel values were greater than 1; negative pixel values were not present in this dataset due to the non-negativity of the Poisson distribution. For the construction of the training set to be used with the Noise2Noise approach, two different noisy realizations per clean slice were used. The baseline PSNR of this dataset was 29.91 dB.

B. Real CT dataset

Approximately clean CT slices were obtained from part of the DeepLesion dataset released by the National Institutes of Health (NIH) Clinical Center [6], which consists of 32,120 axial CT slices from 10,594 studies of 4,427 unique patients. The clean slices were rescaled by dividing all images by the highest pixel value of the entire dataset.

Noise was added after rescaling the clean images. Again, noisy images were artificially obtained starting from the clean slices. In CT images, noise is similar to Gaussian noise, because in back-projection of a Poisson-corrupted sinogram, all noise contributions are added over the entire image plane. We decided to add an approximation of this type of noise, namely, additive white Gaussian noise (AWGN) before feeding images to the network.

Two training and test sets at two different baseline PSNRs were created: one at 24.44 dB and another at 18.66 dB. Each training set consisted of 10,000 images, and each test set of 1,000 images. In addition, using the MSE loss function, we trained a Noise2Noise network for the training set at 18.66 dB. Table II summarizes the trained networks together with the adopted loss function for the CT dataset.

TABLE II
Networks Trained on CT Slices

Trained network	Noise level [dB]	Loss function
Noise2Noise	24.44	MSE
Noise2Noise	24.44	MAE
Noise2Clean	24.44	MSE
Noise2Noise	18.66	MSE

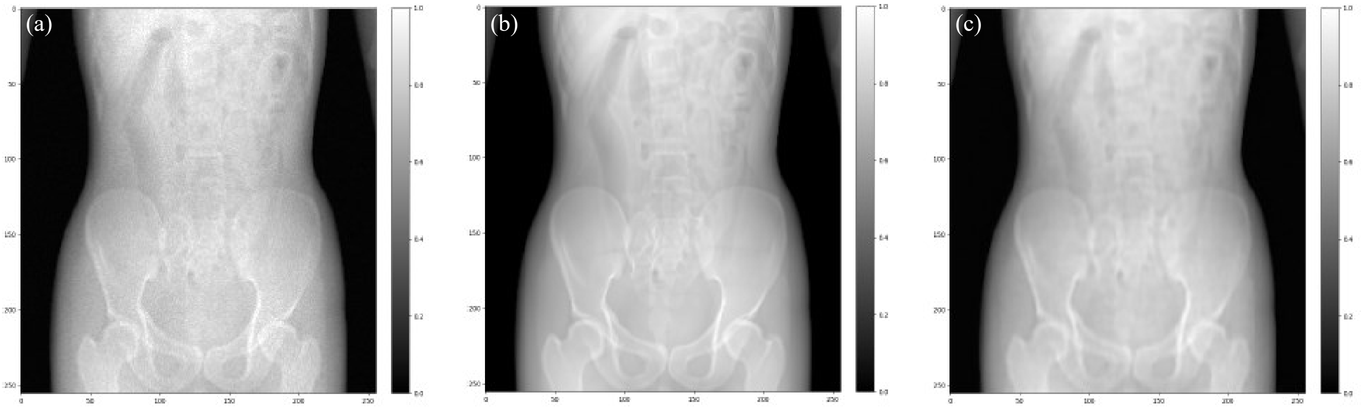

Fig. 1: Comparison between (a) noisy input (31.57 dB), (b) clean reference and (c) prediction of the Noise2Noise model trained with the MSE loss (38.36 dB).

C. Network architecture

The network used in this work is directly inspired by the one proposed previously [2], a U-Net with six levels. In the encoding part, except for the first layer, the dimensional reduction is performed by a series of 2D convolutions with 48 feature maps each and max pooling. In the decoding part, up-sampling is performed by an up-sampling layer, plus two consecutive 2D convolutions with 96 feature maps each. Each level of the encoding part is concatenated to the same level of the decoding part. Leaky ReLUs ($\alpha = 0.1$) were used as activation functions for all layers except for the output layer, where a linear activation function was adopted. The initial learning rate was set to 0.001 with a reduction to its half after two epochs in which the PSNR did not improve during testing.

IV. Results

A. Denoised X-ray projections

Table III shows the results in terms of PSNR obtained by applying different denoising tools on the X-ray projection test set. The experimental results show that the Noise2Noise network showed better performance when trained with the MSE loss instead of the MAE loss. Among all ML approaches the Noise2Clean showed the best performance. It is important to notice that independently of the approach and the adopted loss function, the ML methods are superior to all tested filters. Figure 1 shows the comparison between the noisy input, the clean reference and the prediction of the Noise2Noise model trained with the MSE loss.

The Noise2Noise and Noise2Clean trained networks were applied to different noisy realizations of the same input image in a PSNR range between 14.80 dB to 50.06 dB. Figure 2 shows the PSNR of the predictions by both models over the input PSNR, as well as the model gain (output – input PSNR).

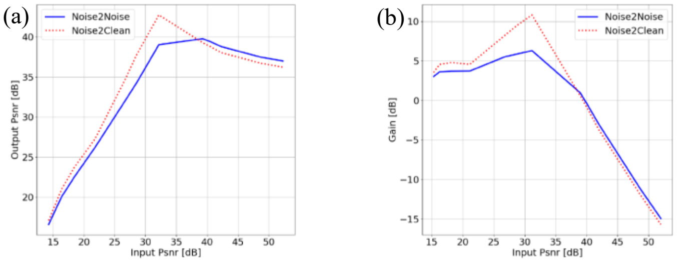

Fig. 2: Behavior of the Noise2Noise (solid blue) and Noise2Clean (dotted red) models trained with the MSE loss function for different noise levels of the same input image. (a) Output PSNR and (b) PSNR gain over input PSNR. Both models were trained on a baseline PSNR of 29.91 dB.

Both models show a similar behavior. The output PSNR increases until the input PSNR is around 35 dB, which is when it starts to decrease (Fig. 2a), indicating that the network's learned denoising model is valid in a high-to-moderate range of noise. When the input image PSNR is higher than 40 dB, both learned models start to deteriorate the image instead of restoring it. Figure 2b reveals a non-linear gain behavior in the region of very high-moderate noise level ([15, 32] dB). For an input PSNR in the range [15, 22] dB, the gain of both networks starts to saturate, suggesting that the learned model has the same denoising effect on very-high-level-noise images, regardless of the value of the input PSNR.

B. Denoised CT slices

Tables IV and V show the results of the various tools used to denoise the CT slices of the two test sets with the two different noise levels. Also in case of high-moderate noise level, the Noise2Noise model performs better compared to the three conventional filters. In case of moderate noise level, the mean predicted PSNR of the Noise2Noise model trained with the

TABLE III
Experimental Results on the X-ray Dataset (Baseline PSNR 29.91 dB)

Tool	Mean PSNR [dB]
Noise2Noise MAE loss	37.66
Noise2Noise MSE loss	39.34
Noise2Clean MSE loss	40.54
Gaussian filter	35.09
Bilateral filter	35.22
Median filter	35.71

TABLE IV
Experimental Results on the CT Dataset (Baseline PSNR 24.44 dB)

Tool	Mean PSNR [dB]
Noise2Noise MAE loss	41.80
Noise2Noise MSE loss	42.09
Noise2Clean MSE loss	40.63
Gaussian filter	34.25
Bilateral filter	35.09
Median filter	33.48

TABLE V
EXPERIMENTAL RESULTS ON THE CT DATASET (BASELINE PSNR 18.66 dB)

Tool	Mean PSNR [dB]
Noise2Noise MSE loss	30.01
Gaussian filter	26.94
Bilateral filter	27.22
Median filter	25.30

MAE loss is 41.80 dB, while the predicted PSNR of the model trained with the MSE loss is 42.09 dB. The performance of the Noise2Noise model, independently from the adopted loss function during training, is better than the performance of the Noise2Clean model.

Figure 3 shows the comparison between the noisy input, the clean reference and the prediction of the Noise2Noise model trained with the MSE loss. It can be noticed that the network was able to restore the image preserving the patient's anatomical structures without the introduction of artifacts in the predicted image. Figure 4 shows the behavior of the two Noise2Noise models trained with the MSE loss function at two different noise levels (24.44 dB and 18.66 dB) when applied to different noise realizations of the same clean image in a PSNR range between 3.61 dB and 46.51 dB.

Both models present a similar behavior. While the input PSNR is in the range [3, 30] dB (Fig. 4a), the output PSNR of both models increases, indicating a denoising effect. For input PSNRs lower than 16 dB, the model trained at 18.66 dB performs better than the model trained at 24.44 dB; the inverse is true for the input PSNR range [16, 30] dB. Around 30 dB the output PSNR of both models starts to saturate, leading to lower output than input PSNRs starting from 35 and 45 dB, respectively.

V. CONCLUSIONS

An ML approach is capable of denoising X-ray and CT images with superior performance compared to the selected conventional filters. In the case of the CT dataset, we found that the results of the Noise2Noise approach are even better than the ones of the more conventional Noise2Clean. Since we used an approximated noise model in CT images, a new training on CT slices with high noise is required to confirm the superiority of the Noise2Noise approach. The best performance of the Noise2Noise approach was reached using the MSE loss function. We have observed that the validity of each trained model is restricted to a range around the amount of noise in the training data. Results should be confirmed using additional metrics beyond PSNR.

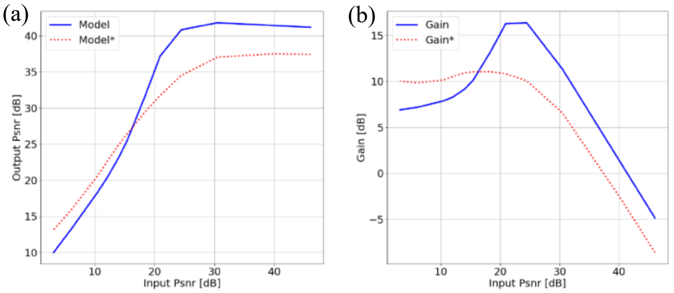

Fig. 4: Behavior of the Noise2Noise network trained at 24.44 dB (solid blue) and trained at 18.66 dB (dotted red) for different noise levels of the same input image. (a) Output PSNR and (b) PSNR gain over input PSNR.

More generally, the experimental results, especially for the CT dataset, suggest a possible application of ML methods in reducing patient exposure during image acquisition without loss of important information. The Noise2Noise method is particularly suitable to train denoising algorithms for all those imaging modalities which are inherently noisy (e.g. positron emission tomography, ultrasound) and where clean targets are not available. Since each scanner and reconstruction mode is characterized by their own noise, a different denoising model trained on images of each combination may be required.

REFERENCES

[1] A.S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI". Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2019. https://doi.org/10.1016/j.zemedi.2018.11.002

[2] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. "Noise2Noise: Learning Image Restoration without Clean Data". arXiv:1803.04189v3, 2018. https://arxiv.org/abs/1803.04189v3

[3] M.G. Asante-Mensah and A. Cichocki, „Medical Image De-noising Using Deep Networks". IEEE International Conference on Data Mining Workshops, 2018. https://doi.org/10.1109/ICDMW.2018.00052

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation". Medical Image Computing and Computer-Assisted Intervention, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[5] P.W. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B.M. Tsui, "4D XCAT phantom for multimodality imaging research". Medical Physics, vol. 37, no. 9, pp. 4902-4915, Sep. 2010. https://doi.org/10.1118/1.3480985

[6] K. Yan, X. Wang, L. Lu, and R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning". Journal of Medical Imaging, vol. 5, no. 3, 036501, 2018. https://doi.org/10.1117/1.JMI.5.3.036501

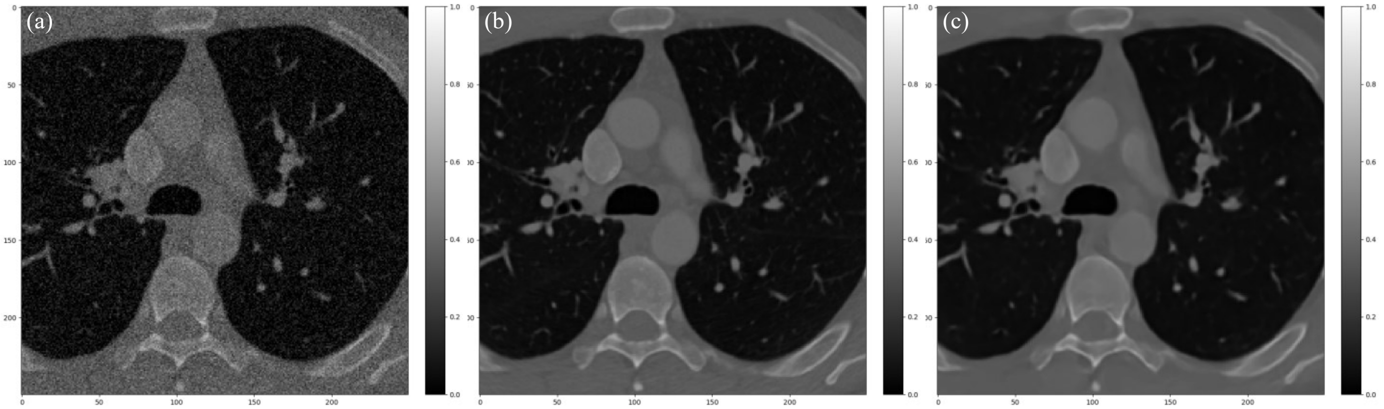

Fig. 3: Comparison between (a) noisy input (24.44 dB), (b) clean reference, and (c) prediction of the Noise2Noise model trained with the MSE loss (40.22 dB).