Optimizing Image Denoising Neural Network for highly noisy Images

Melanie Brtan, Stefano van Gogh, Marco Stampanoni

***Abstract*—Breast cancer is the most common cancer type in women and an early detection can improve the chances of treatment as well as increase the survival of the patient. Therefore, better imaging technologies are desired to ensure a non-invasive diagnosis. In the last years, Grating Interferometry-based phase contrast X-ray Computed Tomography has shown to be a promising application for imaging. Unfortunately, creating high-quality images comes with certain difficulties. Especially the grating fabrication and the following reconstruction can add high noise amplitudes to the measured data. For this reason, a good denoising network is necessary. In medical applications an interpretable algorithm is preferred. The presented neural network is based on the INSIDEnet and has been modified to yield an improved and more general performance. Regularization strategies, like changing the loss function, or changing the layering structure have been performed. The results have shown, that the network has been slightly improved compared to the baseline model. The goal is to apply the INSIDEnet as a proximal mapping within an iterative phase contrast CT reconstruction pipeline.**

I. Introduction

Only in the year 2020 2.3 million women have been diagnosed with breast cancer [7]. Fortunately, early tumor detection due to diagnostic technologies can increase the chances for a successful treatment [3]. Different breast imaging techniques like breast ultrasound, breast MRI, mammography, digital tomosynthesis or absorption-based computed tomography (CT). Absorption-based X-ray imaging methods in general suffer from the disadvantage, that due to low absorption contrast in soft tissue types, it leads to low contrast images [8]. Especially denser breast tissue can make it difficult to see a tumor due to tissue overlap [4] and lead to high false positive rates in diagnoses [3].

In the last years, it has become increasingly attractive to use phase contrast-based imaging, because it could solve those issues. Many different applications have been tested to implement the phase contrast imaging technology. One of them is grating interferometry (GI), which uses gratings in its setup to exploit the differential phase contrast (DPC) signal [8]. Combined with CT it creates a new imaging application.

For this reason, our group has designed a Grating Interferometry Breast Computed Tomography (GI-BCT) prototype (1). This technology allows to obtain three dimensional images of the breast, while the detector and the source rotate around it without any uncomfortable breast compression. The setup of our GI-BCT scanner will be explained in detail in section II. DPC images are more sensitive to the high frequencies and the

Fig. 1: GI-BCT prototype

reconstruction algorithm enhances low frequency noise due to the integration step, which acts as a low-pass filter. For this reason a powerful denoising network must suppress the noise to create valuable images.

Deep learning creates networks, which learn how a denoised image should look like and then apply this knowledge on the unseen noisy images. Today this is mainly done by convolutional neural networks (CNNs), which perform many convolutions until the image is denoised [9]. The downside here is that sometimes a CNN adds or removes structures, i.e. it hallucinates. To solve this problem van Gogh et al. proposed a hybrid algorithm called the 'Interpretable NonexpanSIve Data-Efficient network' (INSIDEnet) [11] which uses the interpretability of classical filters. It has a defined forward and backward path which always goes back into the image space, so that the filtering steps can be followed. [11] implemented the network with orthogonality, which gives it more stability but also limits its possibilities. This leads to the idea that there is still room for improvement. In this paper the INSIDEnet has been modified and improved in order to make it less restrictive and non-orthogonal compared to the baseline model. The results have been obtained on phantoms and show which approaches

worked the best. The goal is to use this network as a denoising regularizer during iterative phase contrast reconstruction as proposed in [10].

II. Materials and Methods

A. Grating Interferometry Breast Computed Tomography (GI-BCT)

When sending X-rays through matter, the electromagnetic wave propagates through it as described in the index of refraction n, as follows

$$n = 1 - \delta + i\beta. \tag{1}$$

Here, the real part δ determines the shift of the phase and the imaginary part β the X-ray attenuation [8]. Especially in soft tissues the X-ray phase shift is stronger than in the attenuation. Therefore, it is reasonable to use the phase shift for imaging techniques. The change of the X-ray's wavefront with the beam's direction l is defined by

$$\Phi = \int \delta \; dl. \tag{2}$$

Because of the phase shift, the refraction of the beam is induced under the angle α with λ as the wavelength of the X-ray. The phase shift Φ in (3) must be integrated.

$$\alpha = \frac{\lambda}{2\pi} \frac{\partial \Phi}{\partial x} \tag{3}$$

But not only the real part δ is being used, also the imaginary part β has its function. It determines the attenuation of the beam with the Beer-Lambert law, where the attenuation coefficient μ is needed

$$\mu = \frac{4\pi\beta}{\lambda}. \tag{4}$$

Especially in soft tissues the phase shift is more significant than in its attenuation. One promising facility is the GI-CT. Its setup is very robust with a large field-of-view and works with limited monochromaticity and spatial coherence. Thus, it makes it adaptable for a clinical environment. [12] For the setup, gratings are placed between the source and the detector. Depending whether the GI-CT is placed at a Synchrotron beamline or in a laboratory with a conventional X-ray tube, the setup needs two or three gratings. For the latter option generally three gratings are placed. As shown in Figure 2, the first grating G0, also called source grating, lies right behind the source and increases the beams coherence. The second grating G1, called phase grating, lies in front of the sample. It creates an interference pattern called Talbot carpet, which

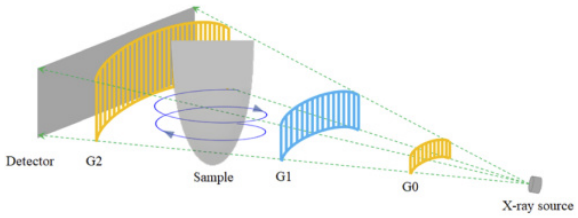

Fig. 2: Schematics of the GI-BCT set up, which consists of a X-ray source, three gratings and a detector [13].

is necessary to detect the phase shift induced by the sample. The last grating G2, the analyzer grating, is placed between the sample and the detector. It is highly absorbing and allows to resolve the fringes in the shifted interference pattern. [1] If the setup is at a Synchrotron, the source grating can be left out. When one of the gratings is moved in x-direction with respect to the other gratings, a phase stepping curve in (5) can be obtained based on α

$$I_k = I_0 T \cdot [1 + V_0 D \cdot \cos(k + \Phi_0 - \varphi)]. \tag{5}$$

The variables indexed with zero represent the flat field measurements. k is the k-th phase step. T stands for the transmission signal, D for the dark-field signal and φ for the differential phase signal [13]. They can be computed by

$$T = \exp\left[-\int \mu \; dl\right] \tag{6}$$

$$D = \exp\left[-\int \epsilon \; dl\right] \tag{7}$$

$$\varphi = \frac{\lambda d_2}{g_2} \frac{\partial}{\partial x} \int \delta \; dl \tag{8}$$

In (8), d_2 is the distance between the sample and G2 and g_2 is the pitch of G2. From the phase stepping curve of the flat-field data and the sample data, the DPC signals can be obtained and therefore also its phase shift with simple Fourier analysis. For the reconstruction the phase shift has to be integrated in order to convert the phase contrast tomograms from signals into images. The algorithm used for the reconstruction was taken from [10] and will not be further explained.

B. Simulated breast phantoms

1) Clean and noisy breast phantoms for network training: To train a neural network, a training dataset is necessary. For this reason, the INSIDEnet requires a noisy image as well as a clean image in order to fit the parameters. The phantoms were simulated as described in [11]. The only difference is that here

a photon count of 120000 instead of 200000 has been used, where the photon number indicates the number of photons leaving the source. Since the goal is to apply the INSIDEnet in the reconstruction pipeline, our trained phantoms need to have a certain level of noise to get as realistic results as possible. In the end, the aim is to get as close as possible to the clinical setup. Currently, the tested dose is still higher than the allowed dose. All approaches during the development were tested with the photon count of 120000. Later, to train the INSIDEnet for the reconstruction, phantoms with photons of 80000 and 100000 were used. Here, 30 in-silico breast phantoms of 44 $\times$ 1536 $\times$ 1536 voxels with a voxel size of 100 μm were created. Ten phantoms are used for training, ten for validation and ten for testing. The phantoms represent the main tissue types seen in real data, i.e. adipose, glandular and skin tissue. In order to create a more realistic effect, two masks were added. To this clean phantom, which acts as a base, noise has been added.

C. INSIDEnet structure

During the signal retrieval, noise is being added due to quantum noise in the detector, which is being amplified by imperfect gratings. This must compensated by a powerful denoising algorithm. A very common strategy is to use deep learning approaches. It creates neural networks, which work by using so-called weights. Those are updated based on the training data [2]. In this case it receives the noisy images and compares it with its ground truth image, for the purpose of learning how to denoise the image correctly. Some networks exist already, as for example the U-net, a convolutional neural network. Like its name reveals, the network has a shape of a U. This means, the image is first downsampled in the contracting path and later upsampled in the expansive path. This is being performed, while applying convolutions and rectified linear unit (ReLU) as well as max pooling operations in the layers [6]. The nonlinear mapping structure makes it possible to try out many different possibilities of combinations but it has a higher expressive power than linear mapping. This could be the reason for the strong denoising effect it has, but it has also has no interpretability and might lead to hallucinations on the images. This could lead to serious consequences, when implemented in medical imaging applications. To avoid such implications, van Gogh et al. created the INSIDEnet [11]. It has a structure, which also includes a up- and downsampling path, but for the filtering step, it contains orthogonal filters. For the forward path the filters are multiplied with the input images and a thresholding is applied. In order to return to the image space, the inverse, which is also the transpose, is multiplied. The exact terminology and further informations regarding the initial INSIDEnet can be found in [11]. With the presented various regularization strategies, we attempted to improve this network's performance.

D. Nonorthogonality

In this paper the same terminology as in [11] will be used. Its baseline model uses orthogonal filters to perform the filtering path. The orthogonality provides stability but also restricts the network. Therefore, our approach was to remove the orthogonality entirely hoping to increase performance. When removing the orthogonality, it is possible for the filters to arrange themselves in any direction. More precisely, we have removed the matrix Q from the forward path as well as from the backward path and replaced it with the trainable matrix B. The backward path, which enables the interpretability of the network, was previously performed by using the transpose of Q. Due to nonorthogonality, this is no option anymore and therefore we now use the inverse for the backward path. Important to mention here is, that at some point we tried to increase the parameter number by using rectangular filters with the shape of $m \times n$ with $m > n$ for B instead of $n \times n$ filters. This means, that the inverse could not be used anymore and therefore the pseudoinverse was implemented.

E. Loss function and regularizers

By building a better loss function and adding soft constraints, we believe, that the performance can be improved. Especially since the previous hard constraint (the orthogonality) has been removed. Various initializers, constraints and regularizers have been tested, among them, these were the most promising ones. In the end, only (10) and (11) have been definitely implemented into the network.

1) Determinant regularizer: Based on the idea of [5], a determinant regularizer has been added because of the missing orthogonality, to avoid collinearity. This was achieved by computing the the logarithm of the determinant of the weights. If the filters are too similar, the loss function will be penalized, because the determinant of a singular matrix is zero and therefore the logarithm of zero is infinitely big.

$$-\lambda \log\det B \tag{9}$$

Unfortunately, this regularizer did not work well enough. It could have been because the filters were still too orthogonal and therefore the determinant regularizer was zero or the learning dynamics were not optimal.

2) Dissimilarity loss: To replace the determinant regularizer, which did not have a big influence, the dissimilarity loss is proposed.

$$+\lambda \sum_{i=1}^{n} \sigma(B^n)^2 \tag{10}$$

Here, σ calculates the angle between the rows (the vectors) of the non-orthogonal filter matrices, which should not be collinear or too similar in this context. It should penalize the filters for the same reason as for the determinant regularizer. The dissimilarity loss should also not be perpendicular, because then the matrices are orthogonal again. The matrices should be as dissimilar as possible to increase their learning ability. When utilized in the network, λ has been fine-tuned to have an constant dissimilarity value.

3) Variance loss: Additionally, a second regularizer has been used to keep the threshold variance small. This avoids that filtering strength distribution fluctuates too much and that there are some filters, which learned too little while some learned too much. With the parameter tuning of γ the weighting of the variance loss can be tuned.

$$+\gamma var(Threshold) \tag{11}$$

For both regularizers, hyperparameter tuning of λ and γ has been performed.

4) Loss function: the loss function calculates the distance between the current output and the expected output during a training. Therefore, a small loss value is welcome and should be decreasing. Before the loss function was only defined by the mean squared error (MSE), whereas now it is a combination of the MSE and the two regularizers

$$\mathcal{L} = MSE + \lambda \sum_{i=1}^{n} \sigma(B^n)^2 + \gamma var(Threshold). \tag{12}$$

F. Nonlinear Filter Structure

A new filter structure has been implemented into the INSIDEnet, where the filter sizes differ in each forward path and the filtering order is not linear, because it performs all forward steps one after another and likewise for the backward path. This had the aim to increase the parameter number and its expressiveness. The filter sizes start by 64×16 increasing to 256×256 like shown in 3 until performing five filter multiplications and then taking the same backward path but with the inverse. This was the maximal filter size without any memory issues.

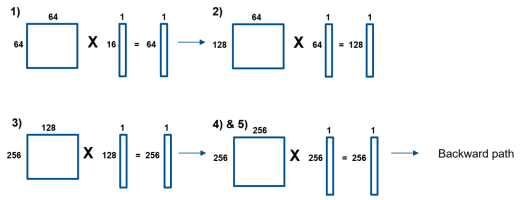

Fig. 3: Nonlinear filter size illustration of the forward.

G. Convolutions

In the baseline model the noisy input image is divided into patches which are being averaged in the filtering step. This means that by using small patches the blurring should not be as strong as in bigger patches. Unfortunately, the blurring was still too high even by using smaller patches, which led us to the approach to use convolutions. In [11] the multiplications are performed by the Einstein sum, we replaced it with convolutions. Instead of inverses, now deconvolutions are performed. We also tried to create a linear (same staying filter size) and nonlinear (increasing filter size) but due to memory issues the convolutional nonlinear worked but only with two filters. This did not lead to sufficient results.

H. Application in reconstruction pipeline

The reconstruction pipeline uses the iterative phase contrast reconstruction which is described here [10]. Because the backprojection adds a lot of noise to the images, it needs a good denoising algorithm. Here, van Gogh et al. used a non-expansive and biasless variant of the U-net, which is being used iteratively after every ten data update steps to denoise the image. The goal was to replace the U-net by the INSIDEnet, but due to time reasons it was not possible to fulfill this approach and implement it correctly.

1) New phantom generation: To improve the performance of the INSIDEnet in the reconstruction, the idea was to add an additional parameter in the filter matrices. This new component would teach to the network also the reconstruction steps. To put this into practice, we generated new datasets of phantoms with a different noise patterns by using the L-BFGS algorithm from [10]. The data generation includes three steps like represented in algorithms 1, 2 and 3. The first part inputs an empty image δ with only zero entries and the sinograms φ_{clean} of the clean phantoms in the L-BFGS optimization algorithm and iterates 200 times where only every 10th reconstruction is saved as $\delta_{clean,i}$. This forms our clean base for the new phantoms. The second part produces the noise by generating one noisy image of a clean phantom reconstruction $\delta*$ and a noisy sinograms

φ_{noisy} after the 10th L-BFGS optimization. Then we subtract the noisy image δ_{noisy} by the initial clean phantom δ_*, which leaves us only with the noise pattern. In the third step we combine the new clean images $\delta_{clean,i}$ of step one with the $noise$ of second step. With the new data, the networks also learns the noise on every iteration and different noise levels. Unfortunately, this step has not been implemented, but it is planned to implement this approach with the non-expansive U-net in the next months. Especially, since it has not been presented in any literature yet.

Algorithm 1: New phantom generation; Clean data

input : $i = 0$; $n_{\max} = 20$; $\delta = 0$; $\varphi = \varphi_{clean}$; $k = 0$; $k_{\max} = 10$
while $i <$ $n_{\max}$ **do**
  while $k < k_{\max}$ **do**
    $\delta_k = \mathrm{LBFGS}(\delta, \phi, k)$; $k = k + 1$;
  end
  $\delta_i = \delta_k$; save δ_i as $\delta_{clean,i}$; $i = i + 1$;
end
output : $\delta_{clean,i}$

Algorithm 2: New phantom generation; Noise

input : $i = 0$; $n_{\max} = 10$; $\delta = \delta*$; $\varphi = \varphi_{noisy}$
while $i <$ $n_{\max}$ **do**
  $\delta_i = \mathrm{LBFGS}(\delta, \phi, i)$; $i = i + 1$;
end
save δ_i as δ_{noisy}; $noise = \delta_{noisy} - \delta*$ **output :** $noise$

Algorithm 3: New phantom generation; Noisy data

input : $i = 0$; $n_{\max} = 20$; $\delta_{clean,i}$; $noise$
while $i <$ $n_{\max}$ **do**
  $\delta_{noisy,i} = \delta_{clean,i} + noise$; $i = i + 1$;
end
output : $\delta_{noisy,i}$

III. Results

For the purpose of determining the performance of the modified INSIDEnets, we compared it to the previous INSIDEnet of [11]. All trainings and predictions were obtained on images with 120000 photons. For the parameter number, the baseline model uses 166403 parameters while the linear nonorthogonal INSIDEnet uses only 43523 parameters. The linear convolutional INSIDEnet counts 348033. The nonlinear nonorthogonal INSIDEnet, which had the highest parameter count of 696067, performed the worst in the quantification. But here we used the nonlinearity of the filters to increase the parameter number, which worked with our proposed method. The linear nonorthogonal INSIDEnet has been trained with many different filter sizes, where 64×16 filters turned out to be the best option, which is now also used for the comparison. Important to mention is, that we used a very small training set because real-world clinical data sets are also rather small and to get as close to this scenario. The training has then be interrupted, when the loss was remaining the same and not improving anymore.

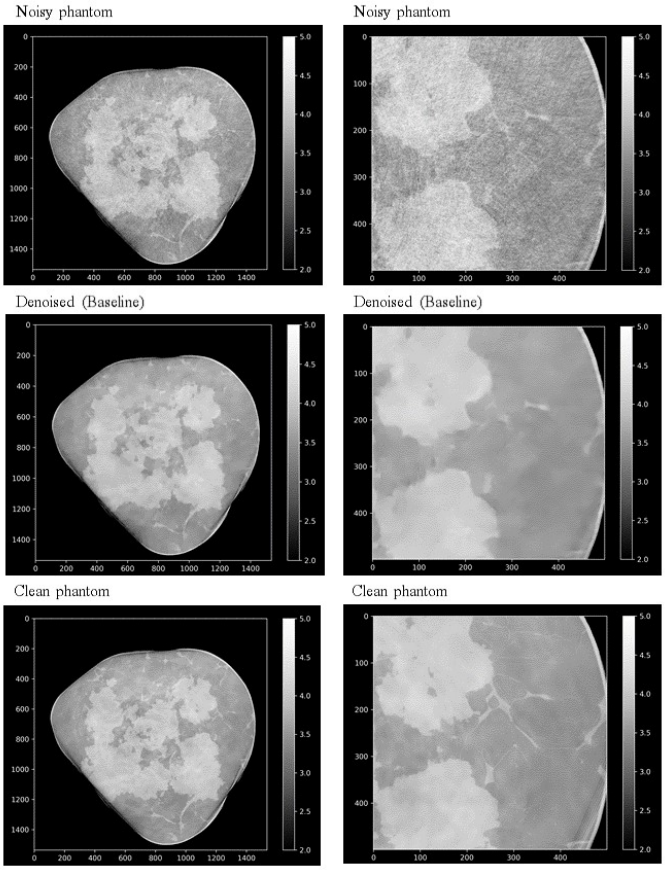

Fig. 4: Denoising results of the baseline model on simulated phantoms with 120000 photons. First row: Noisy phantom. Middle row: Denoised phantom by the baseline model. Third row: Ground truth phantom. On the left a full slice is presented and on the right the zoomed-in section.

A. Quantification

For comparing the different networks quantitaively, the structural similarity index measure (SSIM), mean squared error (MSE), signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) were calculated in table I. To compute the SNR and the CNR in regions of interest, where the grey level values are approximately constant in the ground truth.

Network	SSIM	MSE	SNR	CNR
Baseline	0.8915	0.0093	79.2496	9.0093
NonOrthogonal (lin.)	0.9459	0.0059	85.3241	9.6135
NonOrth. (nonlin.)	0.6430	2.5192	44.4358	0.5182
Convolutional (lin.)	0.9257	0.0081	60.6345	6.7618

TABLE I: Quantitative results of the compared networks

B. Simulated Data

The denoising results are shown below with a zoomed-in image for better visualization. The first two rows represent the noisy and the clean phantoms, followed by the baseline model, the linear nonorthogonal INSIDEnet and the linear convolutional INSIDEnet. By studying the results in table I, we would think that the linear nonorthogoanal model outperforms both the baseline as well as the linear convolutional model.

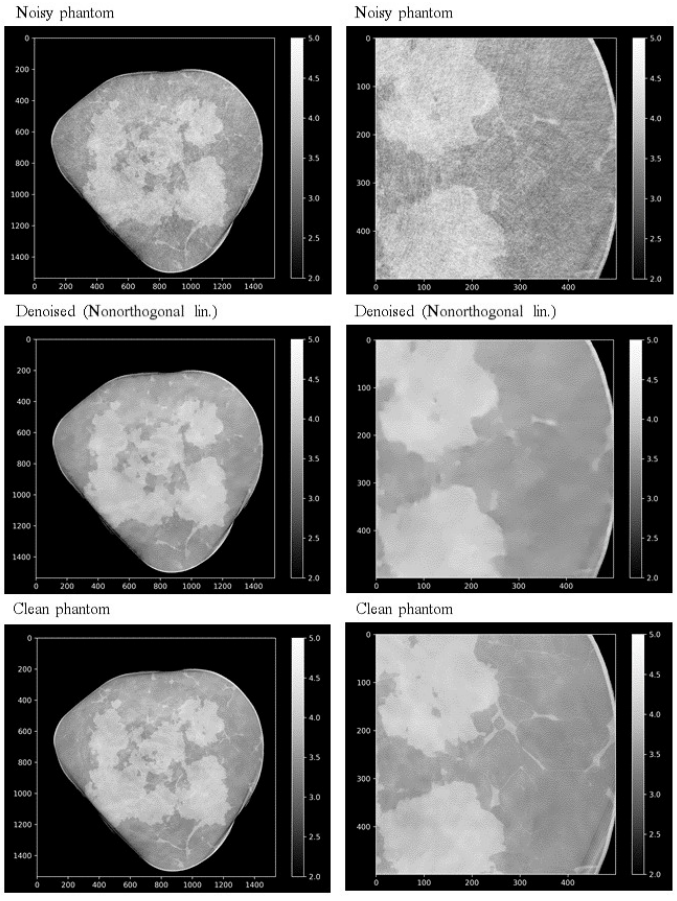

Fig. 5: Denoising results of the linear nonorthogonal INSIDEnet on simulated phantoms with 120000 photons. First row: Noisy phantom. Middle row: Denoised phantom by the baseline model. Third row: Ground truth phantom. On the left a full slice is presented and on the right the zoomed-in section.

Surprisingly, we asserted, that the baseline model in fig. 4 and the linear nonorthogonal model in fig. 5 look very much alike. Both seem to have a blurring. When deciding by eye, which image is more beneficial, we would claim, that the convolutional INSIDEnet performs better (6). A possible explanation is, that the structure is much more comparable with the ground truth. A reason for this could be because the both the baseline and the nonorthogonal model use manual generated patches. Exactly this was the goal in the convolutional network to intentionally remove the blurring by removing the patches, which did work. We can see that both networks perform either better or on the same level as the initial INSIDEnet. Unfortunately, the non linear structure did not perform well enough to present reasonable images. We do not know whether the implementation failed or for any other reason. We believed, that by increasing the parameter number the performance will increase. Therefore, we decided to not present its predictions.

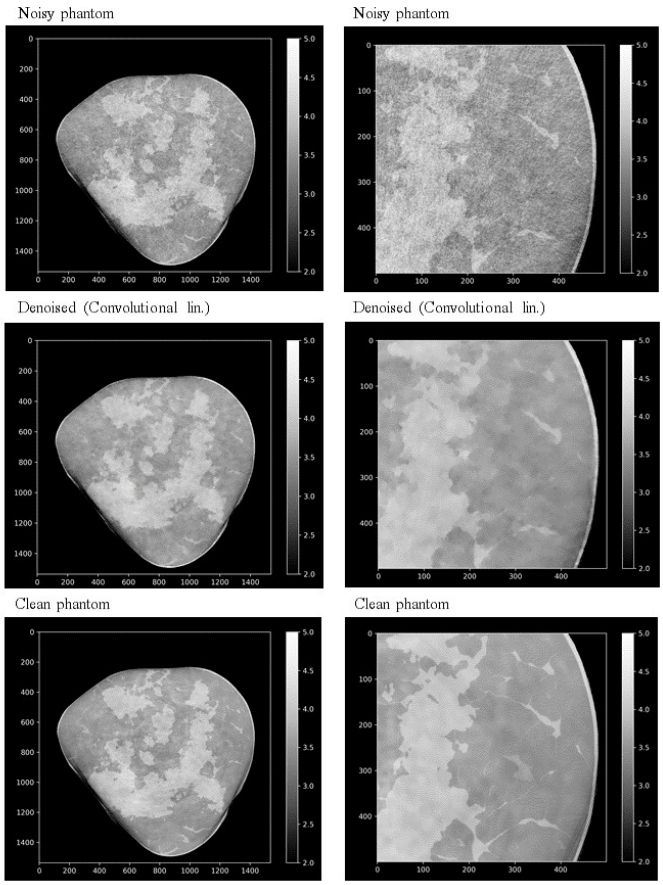

Fig. 6: Denoising results of the linear convolutional INSIDEnet on simulated phantoms with 120000 photons. First row: Noisy phantom. Middle row: Denoised phantom by the baseline model. Third row: Ground truth phantom. On the left a full slice is presented and on the right the zoomed-in section.

IV. Conclusion

The improved INSIDEnet was initially supposed to perform much better, with the goal to use it later in the reconstruction for

the GI-BCT in a clinical setup. There is this trade-off between interpretability and performance. But after trying many different approaches, especially by approximating the structure of the U-net as close as possible, we could not achieve the U-net performance with the interpretability. The results show, that for now this denoiser can not be used in the reconstruction.It remains an open question whether it was due to intrinsic limitations in the network's architecture or if there is some type of trade-off. Denoising can be defined as a solved problem if the network is big enough and the interpretability is not important. If the aim is to know exactly what is happening in the denoising steps and the interpretability plays an important role, the INSIDEnet can be used, but not with the same performance on high noise images as the U-net. In the future there might be a chance to improve the denoising algorithm, by enlarging the INSIDEnet or to use nonlinear mapping, which are both ideas to approach the same structure as the U-net. It is also possible that manufacturing problems like imperfect gratings will be solved, so the denoiser could perform better on higher quality images.

References

[1] Lorenz Birnbacher, Marian Willner, Astrid Velroyen, Mathias Marschner, Alexander Hipp, Jan Meiser, Frieder Koch, Tobias Schröter, Danays Kunka, Jürgen Mohr, et al. Experimental realisation of high-sensitivity laboratory x-ray grating-based phase-contrast computed tomography. *Scientific reports*, 6(1):1–8, 2016.

[2] Nikolaus Kriegeskorte and Tal Golan. Neural network models and deep learning. *Current Biology*, 29(7):R231–R236, 2019.

[3] Magnus Løberg, Mette Lise Lousdal, Michael Bretthauer, and Mette Kalager. Benefits and harms of mammography screening. *Breast Cancer Research*, 17(1):1–12, 2015.

[4] Ritse M Mann, Alexandra Athanasiou, Pascal AT Baltzer, Julia Camps-Herrero, Paola Clauser, Eva M Fallenberg, Gabor Forrai, Michael H Fuchsjäger, Thomas H Helbich, Fleur Killburn-Toppin, et al. Breast cancer screening in women with extremely dense breasts recommendations of the european society of breast imaging (eusobi). *European Radiology*, pages 1–10, 2022.

[5] Saiprasad Ravishankar and Yoram Bresler. Learning sparsifying transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, 2012.

[6] Olaf Ronneberger. Invited talk: U-net convolutional networks for biomedical image segmentation. In *Bildverarbeitung für die Medizin 2017*, pages 3–3. Springer, 2017.

[7] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

[8] Seyedamir Tavakoli Taba, Timur E Gureyev, Maram Alakhras, Sarah Lewis, Darren Lockie, and Patrick C Brennan. X-ray phase-contrast technology in breast imaging: principles, options, and clinical application. *American Journal of Roentgenology*, 211(1):133–145, 2018.

[9] Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, 2020.

[10] Stefano van Gogh, Subhadip Mukherjee, Jinqiu Xu, Zhentian Wang, Michał Rawlik, Zsuzsanna Varga, Rima Alaifari, Carola-Bibiane Schonlieb, and Marco Stampanoni. Iterative phase contrast ct reconstruction with novel tomographic operator and data-driven prior. *PLOS ONE, under review*, 2022.

[11] Stefano van Gogh, Zhentian Wang, Michał Rawlik, Christian Etmann, Subhadip Mukherjee, Carola-Bibiane Schönlieb, Florian Angst, Andreas Boss, and Marco Stampanoni. Insidenet: Interpretable nonexpansive data-efficient network for denoising in grating interferometry breast ct. *Medical physics*, 2022.

[12] Timm Weitkamp, Ana Diaz, Christian David, Franz Pfeiffer, Marco Stampanoni, Peter Cloetens, and Eric Ziegler. X-ray phase imaging with a grating interferometer. *Optics express*, 13(16):6296–6304, 2005.

[13] Jinqiu Xu, Zhentian Wang, Stefano van Gogh, Michał Rawlik, Simon Spindler, and Marco Stampanoni. Intensity-based iterative reconstruction for helical grating interferometry breast ct with static grating configuration. *Optics Express*, 30(8):13847–13863, 2022.