Denoising Grating-Based Phase-Contrast Computed Tomography with Self-Supervised Deep Learning: A Comparative Study

Sami Wirtensohn, Clemens Schmid, Daniel Berthe, Dominik John, Lisa Heck, Silja Flenner, and Julia Herzen

1Technical University of Munich
2Affiliation not available

October 31, 2023

Abstract

To increase the detectability of breast cancer, recent efforts have been made in the field of grating-based phase-contrast computed tomography, which provides additional information about the refractive index decrement introduced by the sample. Whereas this signal shows an increased soft-tissue contrast, the implementation for polychromatic sources is still tied to higher doses, preventing medical application on patients. To lower the dose, we introduce the self-supervised deep learning network Noise2Inverse into the grating-based phase-contrast computed tomography field and compare its results with other denoising methods, namely the Statistical Iterative Reconstruction, Block Matching 3D, and Patchwise Phase Retrieval. Before the comparison, the behavior of Noise2Inverse parameters on the phase-contrast results are evaluated. We show that Noise2Inverse delivers superior denoising results with respect to the investigated image quality metrics. Its application allows the increase of resolution while maintaining a good contrast-to-noise ratio. This could enable grating-based phase-contrast computed tomography to outperform conventional absorption-based computed tomography in dose-relevant applications.
Denoising Grating-Based Phase-Contrast Computed Tomography with Self-Supervised Deep Learning: A Comparative Study

Sami Wirtensohn, Clemens Schmid, Daniel Berthe, Dominik John, Lisa Heck, Silja Flenner and Julia Herzen

Abstract—To increase the detectability of breast cancer, recent efforts have been made in the field of grating-based phase-contrast computed tomography, which provides additional information about the refractive index decrement introduced by the sample. Whereas this signal shows an increased soft-tissue contrast, the implementation for polychromatic sources is still tied to higher doses, preventing medical application on patients. To lower the dose, we introduce the self-supervised deep learning network Noise2Inverse into the grating-based phase-contrast computed tomography field and compare its results with other denoising methods, namely the Statistical Iterative Reconstruction, Block Matching 3D, and Patchwise Phase Retrieval. Before the comparison, the behavior of Noise2Inverse parameters on the phase-contrast results are evaluated. We show that Noise2Inverse delivers superior denoising results with respect to the investigated image quality metrics. Its application allows the increase of resolution while maintaining a good contrast-to-noise ratio. This could enable grating-based phase-contrast computed tomography to outperform conventional absorption-based computed tomography in dose-relevant applications.

Index Terms—Computed tomography, X-ray imaging, Deep learning, Noise reduction, Self-supervised learning, Breast

I. INTRODUCTION

Breast cancer remains a major public health concern, with 2.26 million new cases diagnosed globally in 2020 alone [1]. While mammography has played a significant role in reducing breast cancer mortality [2], its low specificity and limitations in women with dense breast tissue [3], [4] have led to the development of alternative imaging techniques such as dedicated breast computed tomography (B-CT). It overcomes the problems of anatomical structures overlapping while maintaining a low radiation dose [5] and therefore shows a superior visualization and characterization of tumor masses [6]. There are several B-CT devices which are used in clinical applications showing promising results [5]–[7]. However, since B-CT is an attenuation-based imaging method, it still suffers from low soft-tissue contrast like mammography, which sometimes makes it impossible to detect palpable masses or suspicious findings like cancerous lesions or cysts [8].

This can be improved by the use of phase-sensitive imaging techniques as demonstrated at high brilliant X-ray sources [9], [10]. Due to the high cost and spatial requirements of synchrotron facilities, they are not feasible for medical applications. Therefore, the grating-based phase-contrast imaging (gbPCI) approach was developed to bring phase-contrast imaging to low brilliant X-ray sources [11]. The method shows promising results in improving soft-tissue contrast [12], and delivers information about the differential phase and the small angle scattering of X-rays in the sample in addition to the attenuation. First studies using ex-vivo samples showed improved visualization of lesions and microcalcifications for 2D grating-based phase-contrast mammography [13]–[16]. To exploit the contrast advantages of phase-sensitive imaging and the 3D information of B-CT, recent research is advancing in the direction of grating-based phase-contrast computed tomography (gbPC-CT).

However the gbPC-CT suffers from a main drawback preventing its medical application. Due to the integration of the differential phase, the reconstructed volume suffers from low frequency noise which leads to a resolution-dependent performance relative to attenuation-based CT (abCT) [17], [18]. Since many medical applications like full body CTs are performed at lower spatial resolution, it has been shown that the contrast advantages of the phase information can be overcompensated, leading to an equal or even worse Contrast-to-Noise Ratio (CNR) per dose [19]. Additionally, the low-frequency noise of the gbPC-CT reduces the observer performance even at equal CNRs since it is more likely to interfere with the sample structure than the high-frequency noise of abCT. [19].
This is less of a concern for grating-based phase-contrast B-CT since it is generally performed at higher resolutions like mammography. Nevertheless, its performance is still limited for the same reasons. An increase in spatial resolution is required to take full advantage of the enhanced soft tissue contrast of the phase information. However, a higher spatial resolution increases the dose since a certain CNR is necessary for clinical application. Hence, a denoiser is needed, which can drastically increase the CNR of high-resolution low-dose data and which is also able to effectively handle low-frequency noise to compensate for its impact on the observer performance.

In recent years, denoising techniques based on deep learning networks have been developed, showing promising results in medical applications [20]–[22]. Most of these networks need to be trained on a set of data containing noise-free images, but such data is not yet available for gbPC-CT. Therefore, we introduce the self-supervised deep learning network Noise2Inverse (N2I) into the field of gbPC-CT, which can be trained on noisy data alone. We also investigate the effects of its hyper parameter on the image quality. The performance of N2I is compared to other denoisers, namely Statistical Iterative Reconstruction (SIR), Block Matching 3D (BM3D), and Patchwise Phase Retrieval (PPR). To quantify the results, this work also introduces a qualitative evaluation system based on several Image Quality Assessment (IQA) methods. Furthermore, the potential impact of N2I on the field of gbPC-CT is investigated by measuring changes in the noise power spectrum.

II. METHODS

A. Data Acquisition

For clinical application, the most common X-ray source is a polychromatic X-ray tube. Therefore, all experiments are conducted at a rotating anode.

1) Experimental Setup: As source served a MicroMax HF007 rotating anode with a molybdenum target from the Rigaku Corporation. At the preferred utilization of the CT with an accelerating voltage of 50 kVp and a current of 24 mA, the setup reaches a measured visibility of 11.5%. For all measurements, the Santis 0808 prototype gallium arsenide photon counting detector of the Dectris AG is used. It has two vertically stacked 500 µm thick modules with a pixel size of 75 µm x 75 µm. Since commercial breast CTs such as the nu:view of AB-CT GmbH use resolutions of 300 µm for screening [5] all measurements are 5x5 binned, resulting in an effective pixel size of 281.5 µm.

2) Sample: For the investigation of denoising behavior, a breast phantom based on a marbled piece of pork neck is used to simulate the varying densities of adipose and soft-tissue in the female breast. The phantom has a size 2.5x2.5x3 cm and is placed in a Falcon tube with a 70% ethanol concentration after fixation in a 4% formaldehyde solution for one week. A polymethylmethacrylate (PMMA) rod is added for possible energy calibration. The whole tube is placed into a 3 cm thick water bath to avoid strong phase changes at the tube-air boundary.

B. Denoising Methods

A central part of this paper is the application of N2I to gbPCI and comparing its performance with conventional denoising techniques already used in this field. In the following, all used denoising methods are briefly introduced. Fig. 1 shows the general processing pipeline, including the interaction points of the used denoising methods.

1) Patchwise Phase Retrieval: PPR takes advantage of the signal-retrieving process of gbPCI to denoise the outputs. This is achieved by considering the information of the neighboring pixels. Since the phase retrieval algorithm fits a sine curve to the intensity values of all grating steps for each pixel individually, the additional information increases the statistic by reducing the variance. It therefore improves the fitting accuracy resulting in a reduction of noise in the final image, if the sample is homogeneous in the neighboring region. But if the sample changes significantly in the neighborhood of a pixel, the PPR leads to blurring artifacts, which reduce the image quality. Hence, the included number of neighboring pixels must be chosen carefully and adjusted to the properties of the sample [23]. This method should not be used together with N2I since the resulting noise in the projections is no longer pixel-wise independent, leading to a non-vanishing factor (see (2)) that reduces the quality of the final result [24]. This needs to be considered for all denoising methods applied directly to the projections.

2) Block Matching 3D: BM3D is an image denoising algorithm that uses collaborative filters [25]. This is done by grouping similar 2D fragments of an image and stacking them into a 3D array. Since high similarity of small image fragments at different spatial locations is common in natural images, the transformation achieves a high sparse representation of the true signal and the noise can therefore be separated during shrinkage [25], [26]. The sparsity increases with the number of grouped blocks. BM3D is proven to be an effective denoiser in the field of X-ray imaging [20], [27], [28]. We use a BM3D implementation based on the work of Y. Mäkinen et al. [29].

3) Statistical Iterative Reconstruction: The SIR algorithm is a statistical reconstruction approach to improve the image quality of CT data. This is done by introducing a statistical model that describes the fluctuations in the measurement due to noise, by taking advantage of known information about the volume, and by using an iterative algorithm that approaches the correct image in multiple steps. Hence it can also decrease the impact of sparse data and missing projection artifacts and can handle complex geometries. In this paper we use the python implementation X-AID from MITOS GmbH [30].

4) Noise2Inverse: N2I is a self-supervised CNN-based denoising method, which is in no need of additional clean or noisy data for training [24]. It is therefore able to learn from one noisy measurement itself. N2I achieves that by dividing the noisy data x into multiple subsets in which the noise is element-wise independent and zero-mean. It is analytically shown, that training a network on complementary noisy subsets x_J and x_J,C as input and target is as good as training the network on the ground truth z_J up to a certain factor, which corresponds to the variance of the noise (see (3))
C. Image Quality Assessment

To quantitatively compare the impact of different denoising algorithms, reliable image quality assessment metrics are required. In the following, the five considered methods are explained.

1) Normalized Root Mean Squared Error: We use the normalized root mean squared error (NRMSE) as a voxel-based quality estimator. It is based on the root-mean-square error (RMSE) which is given by

\[
\text{RMSE} = \sqrt{\frac{1}{NM} \sum_{i=0}^{M} \sum_{j=0}^{N} (x_{ij} - y_{ij})^2},
\]

where the image \( x_{ij} \) and the reference image \( y_{ij} \) both have the size \( M \) times \( N \). The NRMSE is then defined as

\[
\text{NRMSE}(\theta) = \text{RMSE} \sqrt{\frac{MN}{\| y \|}}.
\]

The NRMSE therefore returns the difference between the values of an estimator and the true value [33].

2) Peak Signal to Noise Ratio: Hence the quality of images is often mainly impacted through the presence of noise; we also use the peak signal-to-noise ratio (PSNR). It measures the ratio between the maximum possible signal power and the noise power. It is defined as

\[
\text{PSNR} = 20 \log_{10} \left( \frac{\| y \|_{\infty}}{\text{RMSE}} \right),
\]

where \( \| \cdot \|_{\infty} \) denotes the supremum norm.

3) Structural Similarity Index: Since human vision is highly adapted to extract structural information, perceptual quality is closely related to it. The structural similarity (SSIM) index exploits this fact by comparing the structural degradation of an image to its reference to determine the perceptual quality change [34]. It is defined as:

\[
\text{SSIM}(x, y) = \frac{(2\sigma_x \sigma_y + C_1)(2\mu_x \mu_y + C_2)}{\mu_x^2 + \mu_y^2 + C_2},
\]

with the local means \( \mu_x \) and \( \mu_y \), the standard deviations \( \sigma_x \) and \( \sigma_y \), and the cross covariance \( \sigma_{xy} \) for the test image \( x \) and the reference image \( y \). The variables \( C_1 \) and \( C_2 \) stabilize the division with weak denominators. It delivers values between 0 for a low and 1 for a high structural similarity [33].

4) Edge Preservation Ratio: The edge preservation ratio (EPR) is an image sharpness assessment method based on the ratio of the extracted edges of a reference and a distorted image [35]. The detection of the edges can be performed through any edge detector. As suggested by [35], this work uses the Canny filter \( C \) to determine the edge map based
on local maxima of the gradient magnitude of a Gaussian-smoothed image [36]

\[ R = C(y, p) \] (8)
\[ D = C(x, p). \] (9)

\( R \) is the set of pixels representing edges in the reference image \( y \) and \( D \) the set of edge pixels in the distorted image \( x \). While \( p \) describes the width of the Gaussian filter and the low and high hysteresis threshold. They are kept identical for the reference and distorted image. The EPR is subdivided into two sub-metrics, the EPR accuracy (EPRa) and the EPR robustness (EPRr). They are calculated by

\[ \text{EPRa} = \frac{|R \cap D|}{|R|}, \] (10)
\[ \text{EPRr} = \frac{|R \cap D|}{|D|}, \] (11)

where \( | \cdot | \) denotes the number of edge pixels. The EPRa is a metric for edge preservation and is close to one for good edge conservation. The EPRr measures falsely introduced structures and is closer to one for less false structures [35].

5) Edge Fit Ratio: The edge fit ratio (EFR) measures the sharpness of a chosen edge. It requires a manual selection of an edge \( k \), to which an error function erf is fitted

\[ y = a_0 \cdot \text{erf}(a_1 + k \cdot a_2) + a_3. \] (12)

This fit is performed at the same edge once for the reference and once for the distorted image. The \( a_2 \) value determines the steepness of the edge and is therefore dependent on the sharpness. The ratio between \( a_{2\text{dis}} \) of the distorted image and \( a_{2\text{ref}} \) of the reference image is the EFR

\[ \text{ERF} = \frac{a_{2\text{dis}}}{a_{2\text{ref}}}. \] (13)

An EFR equal to one means a preservation of the edge sharpness, while a ratio smaller than one indicates a decrease in edge sharpness. A ratio larger than one implies a sharpening of the edge and is not to be expected in this comparison.

III. EXPERIMENTAL RESULTS

A. Noise2Inverse for Phase-Contrast Imaging

In the initial N2I paper, Hendriksen et al. [24] showed that two splits deliver good results in general, but the PSNR of conventional abCT can be improved by choosing four splits if the projection angles are not under-sampled. To find out whether this is also valid for the electron density of gbPC-CT, N2I is applied to a measurement with an angular sampling factor of two at 20 mGy, thus using two times as many projection angles as recommended by the Nyquist theorem. The projections are then split once into two and once into four subsets, which are reconstructed independently. N2I is trained on each full set of subsets and for each signal type separately. The outputs are compared to a reference measurement performed at a high dose of 1231 mGy. The results are listed in Table I and displayed in Fig. 3.

For the electron density, the IQA values do not show a noticeable improvement by using four instead of two splits. The right column in Fig. 3 shows the results for the electron density. Here, the application with two splits retrieves more details in the muscle structure, which is visible in the left and right fibers (red and blue arrows). Therefore, separating the projections into two splits is preferred for the electron density.

On the contrary, for the attenuation coefficient, all IQA values are higher in the case of four splits. Upon closer examination of Fig. 3, it reveals that N2I can retrieve small structures better using four splits. In the bottom left of the sample (red arrow), a small fiber in the muscle tissue is nearly indistinguishable in the case of two splits but can be seen in the result of four splits. This can also be observed in the bottom right region (blue arrow). The structure in the central muscle region (purple arrow) is also better recovered using four splits. For this reason, it is recommended to use four splits for the
It is able to improve the image quality significantly compared to the normal filtered back projection (FBP). While BM3D delivers similar PSNR values for the attenuation coefficient, N2I shows better SSIM and NRMSE values. The sharpness values of Table III indicate that N2I also delivers the best EPRr, meaning it can effectively reduce the impact of false structures introduced through the noise. However, compared to BM3D, it shows a lower edge accuracy and, for the electron density, a smaller EFR value, which is only lower in the case of SIR. This is noticeable in a visible reduction of sharpness (see Fig. 4). The sharpness of the N2I result increases with a higher number of epochs but causes overfitting of noise, and therefore the image quality decreases. This might improve for larger data sets since the network is only trained on the available 49 slices in this comparison.

BM3D shows good results but cannot compete with the IQA values of N2I. While the PSNR values are similar for the attenuation coefficient, the SSIM is still higher for N2I. In the case of the electron density, the differences are more significant, and the result shows remaining low-frequency noise, which still visually interferes with the sample structure. BM3D can maintain a decent sharpness for the attenuation coefficient and the electron density, while the EFR drops when applied to the electron density.

SIR can improve the quality beyond PPR and FBP but at the expense of image sharpness and edge accuracy, resulting in the lowest EFR and EPRa values.

The use of PPR improves image quality while decreasing the sharpness only slightly for the electron density. Hence, the use of PPR is recommended when the data is reconstructed by a FBP and no other denoiser is used.

### IV. Impact on Medical Imaging

Raupach and Flohr [19] developed a mathematical framework to analyze the maximal theoretical performance of gbPC-CT by evaluating the ratio between phase and absorption contrast. They mention that gbPC-CT has a fundamentally different noise power spectrum (NPS) than the conventional pbCT, which results in a resolution dependency of its performance [17]–[19]. For low-resolution imaging at full-body CT, the higher contrast of the phase can therefore be overcompensated by the low coherence lengths of gbPC-CT setups with low brilliant sources, resulting in a worse dose efficiency than...
abCT. The relative dose efficiency improves with increasing resolution and is less of a concern for grating-based phase-contrast breast CT, where a higher spatial resolution is used. Nevertheless, a further decrease in dose is still desirable. An established method to further reduce the total noise power is the usage of smoothing filters, which decrease the spatial resolution. Such methods are less effective for gbPC-CT, which is mostly impacted by low-frequency noise instead of high-frequency noise in abCT, as Raupach and Flohr [19] showed on simulated data. This fact is replicated with experimental data, where a measurement without a sample is performed and reconstructed. The result is seen in Fig. 5. The left images show the FBP. Next to it is the FBP with a subsequent Gaussian blur for noise reduction. While the noise pattern in the attenuation coefficient nearly disappears, there are still noticeable remaining noise structures in the electron density. However, this changes with the use of N2I. Since N2I is trained on the corresponding data, it also manages to effectively reduce the low-frequency noise in the electron density, resulting in a smooth image. The change of the NPS due to the application of N2I can be seen in Fig. 6. The right images show the PPR. Following the argumentation of Raupach and Flohr, a decrease of the low-frequency noise and therefore an increase of the CNR of gbPC-CT leads to a lower break-even point of gbPC-CT and abCT performance, which could lead to a higher dose efficiency of gbPC-CT compared to abCT. Put differently, using N2I allows the use of higher resolution while maintaining a CNR that is necessary for diagnosis, where gbPC-CT is able to outperform the conventional abCT. This is supported by a recent study by Rawlik et al. [18]. They showed, based on experimental data, that the break-even point of gbPC-CT and abCT for a fixed CNR of five layers at a resolution of 214\(\mu\)m with a dose of 65\(\mu\)Gy. The use of N2I allows either to lower the dose, maintaining the CNR and resolution, or increasing the resolution, maintaining the CNR and dose. In both cases, N2I shifts the break-even point in favor of gbPC-CT.

### V. Discussion

In our paper we trained the N2I model on each measurement and applied it afterwards. However, this results in high computational efforts and is not suited for the application in medical facilities. Therefore, the next step should be to build a database that can be used to train a more general model. Based on this model, N2I would be applicable to new data without the lengthy learning process and could deliver superior results with lower computational expense than SIR. Two separate models should be created, one for the attenuation coefficient and one for the electron density.
for the electron density, as the noise power spectrum of the two signals differs considerably.

Furthermore, a deeper understanding of the interaction between N2I and small structures with varying contrast is relevant for medical applications. This is particularly interesting in the case of gbPC B-CT since detecting small lesions and microcalcifications can help identify cancerous tissue in its early stages, an essential factor for successful treatment. To further investigate the properties of N2I, simulations containing objects of different sizes, compositions, and shapes could be run. Various noise models, with high- and low-frequency noise, could be applied to these simulations while observing the behavior of N2I. This should give a better understanding of its impact on the resolution of small structures. Experimental measurements with phantoms of known structure sizes and material composition could afterward validate these results.

The quality of N2I could be further improved by investigating the influence of other hyper parameters. One of them is the learning strategy. It describes how many of the splits are used as input during training. The X:1 strategy uses all but one split as input. N2I applied on abCT data achieves the highest image quality when trained with the X:1 strategy and four splits [24]. But since the electron density images show different behavior for the X:1 strategy, it seems reasonable to test the 1:X strategy on electron density data.

Another hyper parameter is the choice of the neural net-
work architecture. We utilized the MS-D network with small parameter numbers as suggested by [24] since it decreases the risk of over-fitting and only small samples with a limited number of pixels and slices were measured. However, there might be more sophisticated networks that can achieve better results for gbPC-CT measurements since CNNs have inductive biases. One example is the size of the convolutional kernels which are set by the chosen structure of the network and not trained on the data itself. While this helps the CNN to archive high performance with minimal data, it limits the model when trained on large quantity of data. In such cases, transformers with minimal inductive biases can outperform CNNs [37] and should be considered when building a persistent model.

N2I can also reduce the scanning time of CT applications since it can maintain image quality while reducing the exposure time per projection [38]. This could be especially useful in time-resolved applications like cryoablation CT which visualizes the freezing process of tissue.

VI. CONCLUSION

In this paper we introduced N2I into gbPC-CT and investigated the denoising performance of N2I, SIR, BM3D, and PPR. In addition, for N2I, the influence of the number of splits, in which the data is divided for training and application, was studied.

We showed that for the application of N2I on the electron density, the choice of two splits is preferred since small structures in the muscle tissue are easier to detect. While for its application on the attenuation coefficient, all IQA values show better numbers when used with four splits.

BM3D delivers great performance when applied to the attenuation coefficient but struggles with the low-frequency noise in the electron density slices. It achieves better results in quality and sharpness than SIR but lags behind the results of N2I, especially when applied to the electron density.

We showed that PPR increases the image quality without decreasing the sharpness significantly but falls in performance behind the other denoising techniques. PPR should not be used with N2I; otherwise, the noise is no longer pixel-wise independent, leading to a performance decrease of N2I.

Overall, N2I showed superior denoising capabilities compared to SIR, BM3D, and PPR without requiring high-dose reference measurements. However, when applied to the electron density, it lacks sharpness, which might be mitigated by training the model on larger data sets. Nevertheless, N2I is capable of efficiently denoising high- and low-frequency noise. Therefore, it significantly increases the image quality, pushing towards higher resolutions and lowering the break-even point of gbPC-CT and conventional abCT, which might ultimately lead to a higher dose efficiency of gbPC-CT, thus bringing it one step closer to medical application.

REFERENCES


