Impact of cone-beam CT noise correlation on self-supervised denoising strategies for low dose breast CT imaging

A. Sisniega,[a] B. Íñigo,[a] A. M. Hernandez,[b] J. McGraw,[c] Y. Achkire,[c] J. H. Siewerdsen,[a,d] J. M. Boone,[b]
[a]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD USA
[b]Department of Radiology, University of California, Davis, CA USA
[c]Izotropic Corp., Surrey, BC Canada
[d]Departments of Imaging Physics, Neurosurgery, and Radiation Physics, The University of Texas M.D. Anderson Cancer Center, TX USA

ABSTRACT

Purpose: Cone-beam breast CT (bCT) provides volumetric images of the uncompressed breast but present higher noise than 2D mammography. Deep Learning (DL) denoising with supervised training has shown successful CBCT noise reduction but requires matched low-dose and high-dose images. Self-supervised training removes that requirement but often assume locally independent noise. This work studies the impact of bCT noise correlation on self-supervised denoising methods.

Methods: The self-supervised training strategies included two blind spot methods – Noise2Self, enforcing local similarity with independent image noise; and Noise2Sim, enforcing image similarity in presence of correlated noise – and two Noisier2Noise approaches: i) noise injection in the image domain; and, ii) noise injection in projection domain with a model of noise correlation. Self-supervised training was performed on bCT images generated from 150 voxelized models with a high-fidelity forward projector, including models of the x-ray spectrum, polychromatic attenuation, and detector signal and noise propagation. Denoised images were assessed with respect to high-dose references and supervised denoising, using RMSE, SSIM, and noise power spectrum (NPS).

Results: Noise2Sim and Noisier2Noise with noise injection in the projection domain showed good performance in presence of correlated noise, achieving RMSE of 0.21 and 0.18 (SSIM of 0.9 and 0.94), respectively, compared to RMSE of 0.17 (SSIM of 0.93) for supervised training. The independent noise assumption in Noise2Self and Noisier2Noise with image domain noise injection resulted in significantly diminished performance, yielding RMSE of 0.23 and 0.37 (SSIM of 0.86 and 0.84). The NPS measurements revealed a shift towards low frequency components for Noise2Sim, arising from blurring of tissue boundaries and residual image transfer induced by the masking of dissimilar regions in the loss function. Noisier2Noise showed a frequency distribution of noise closer to the high-dose reference. Such performance was slightly degraded for non-matched noise injection models inducing shorter correlation kernels than the nominal detector noise correlation, but models inducing longer correlation showed negligible impact in the denoising results.

Conclusion: Self-supervised denoising in presence of correlated noise was proved feasible. Among the evaluated models, Noisier2Noise strategies with projection domain noise injection showed denoising performance comparable to supervised training and noise spectral distribution comparable to high-dose bCT.

Keywords: Breast CT, cone-beam CT, Deep Learning denoising, unsupervised learning, noise models.

1. INTRODUCTION

Breast CT (bCT) systems, based on cone-beam CT (CBCT) geometries with high-resolution flat-panel detectors, have been developed over the last ~15 years. Clinical bCT studies showed that the true volumetric information in bCT provided excellent visualization of soft tissue features and improved capability for detection and characterization of cancerous lesions, compared to 2D mammography or to digital breast tomosynthesis, at comparable radiation dose. However, at equidose conditions, bCT results in increased image noise from lower exposure in individual projection views and from amplification of high frequencies caused by the filtering stage in filtered backprojection reconstruction algorithms, commonly used for bCT image reconstruction.

Recent developments in deep convolutional neural networks (CNNs) for data-driven image restoration, have shown great potential for denoising in photographic[1] and in x-ray CT and CBCT imaging.[2] Deep learning (DL) denoising networks have been conventionally trained using supervised approaches that minimize the mean squared error (MSE) loss, computed between the CNN inference, with a normal (or low) dose (ND) image as input, and a matched high-dose (HD) target, within the Noise2Clean paradigm.[2] However, such matched datasets are rarely available. The dual noise level requirement was relaxed with the introduction of the Noise2Noise[3] strategy, that proved that minimization of MSE across pairs of noisy images with zero-mean independent and identically distributed (i.i.d.) noise realizations, and matched image content,

Medical Imaging 2023: Physics of Medical Imaging, edited by Lifeng Yu, Rebecca Fahrig, John M. Sabol, Proc. of SPIE Vol. 12463, 124630L
 · 1605-7422 · doi: 10.1117/12.2654447

converged to an equivalent solution. While more attainable, acquisition of two matched scans of the same patient is seldom feasible in clinical bCT. Self-supervised training approaches aim at alleviating the need for paired noise realizations by building loss functions employing targets derived from the noisy input that act surrogates of the clean signal.
Self-supervised strategies demonstrated denoising performance comparable to Noise2Clean in photographic imaging, but their performance is impacted by the correlated nature of noise in CT (and bCT), arising from correlation between the noise in neighboring detector pixels and from correlations introduced by the backprojection operator. In this work we characterize the impact of bCT noise correlation on self-supervised DL denoising leveraging high-fidelity models of the bCT imaging chain.

2. METHODS

2.1. High-fidelity bCT simulation with correlated noise models.

Realistic bCT projection data were generated from a collection of 150 voxelized digital breast models, derived from real bCT images of uncompressed breasts, scanned in pendant geometry, and segmented into skin, adipose, and fibroglandular tissue.[4] The voxel size in the bCT models ranged from 0.194 mm to 0.427 mm in the coronal plane (0.191 mm to 0.238 mm in the axial plane). A high-fidelity CBCT simulator[5] was used to generate realistic bCT data from the digital breast models. The simulator integrates accurate models of the bCT imaging chain physics (system geometry, incident spectrum, focal spot blur, and detector response) into a polyenergetic forward projection engine. The detector response featured an energy-dependent model of correlated quantum noise in the projection domain, derived from cascaded systems analysis.[6] The level of noise and correlation kernel (k) was dependent on the input exposure and incident spectrum, calculated after polychromatic attenuation for each line integral. Detector electronic noise was simulated by injecting uncorrelated Gaussian noise to the final projection.[5]
BCT datasets with 500 projections were simulated using a system geometry pertinent to current generation bCT scanners,[7] with source-to-axis distance (SAD) of 500 mm, source-to-detector distance (SDD) of 700 mm, and central ray placed 20 mm below the chest wall. The simulation included a 1024x1024 pixels flat-panel detector with a 250 mg/cm^2 CsI scintillator, and isotropic pixel size of 0.25 mm. The x-ray spectrum was simulated as a tungsten anode x-ray source, with 60 kV and 0.2 mm Gd added filtration. A total of 150 bCT volumes (one per voxelized digital model) were generated at two dose levels: i) a normal-dose setting with source current of 0.8 mAs (denoted ND); and, ii) a reference high-dose setting obtained with 5.0 mAs (denoted HD). The HD and ND projection datasets were then reconstructed with FBP and a raised cosine filter with Hamming apodization (cutoff = 0.9). Two independent noise realizations were computed for each dose level, to allow quantification of noise in matched difference images.

2.2. Self-supervised DL denoising strategies.

Following the Bayesian formulation for DL denoising, an image contaminated by noise, $\boldsymbol{x_i}$, can be expressed as a combination of a noise-free signal ($\boldsymbol{s_i}$) and a noise component ($\boldsymbol{n_i}$), sampled from a joint probability distribution $p(\boldsymbol{s}, \boldsymbol{n}) = p(\boldsymbol{s})p(\boldsymbol{s}|\boldsymbol{n})$. DL denoising methods aim at learning a function f, implemented as a neural network with tunable parameters $\boldsymbol{\theta}$, that estimates a sample $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ of the noise-free signal from a noisy input, such that $\hat{\boldsymbol{y}}_{\boldsymbol{i}} = f(\boldsymbol{x_i}; \boldsymbol{\theta})$. Conventional supervised training uses noise-free (or very low noise, as in the HD dataset in this work) images $\boldsymbol{y}$ to find the $\boldsymbol{\theta}$ that minimizes the MSE loss, given by $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{i}}\|f(\boldsymbol{x_i}, \boldsymbol{\theta}) - \boldsymbol{y_i}\|^2$.
When no clean reference targets are available, self-supervised training approaches replace the noise-free instances $\boldsymbol{y_i}$ in $\mathcal{L}(\boldsymbol{\theta})$ with a surrogate signal derived from the noisy input datasets. As stated above, two main strategies are explored in this work: blind-spot training and Noisier2Noise training.

2.2.1. Blind-spot training with independent noise assumptions (Noise2Self).[8]
In blind-spot training, $\boldsymbol{\theta}$ is obtained by minimizing the MSE loss on a subset J of the dimensions of the input $\boldsymbol{x}$, with $\boldsymbol{x_j}$ defined as $\boldsymbol{x}$ restricted to J. By setting a function $\mathcal{J}$ that define a partition of the dimensions of $\boldsymbol{x}$, the MSE supervised loss $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{i}}\left\|f_J(\boldsymbol{x_i}, \boldsymbol{\theta}) - \boldsymbol{y_{i,J}}\right\|^2$, computed over $\mathcal{J}$, converges to the optimal denoiser, as long as $\mathcal{J}$ is an invariant function (preventing learning the identity) and the noise in each J in $\mathcal{J}$ is independent from the noise in its complement J^c. This last condition is not strictly fulfilled in the presence of correlated noise. In this work, $\mathcal{J}$ was a set of grid masking functions, with masking interval of L pixels. We set L = 32 pixels in our experiments.

2.2.2. Blind-spot training with correlated noise assumptions (Noise2Sim).

To avoid the impact of noise correlations in the *Noise2Self* loss, the *Noise2Sim*[9] strategy proposed to compute the MSE loss using a distinct but similar image $\boldsymbol{x_j}$ as target, and replace $\mathcal{J}$ by a function that estimates the pixel-wise dissimilarity of the input and target images, defining the following cost function:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{i,j}\left\| \left(f(\boldsymbol{x_i}, \boldsymbol{\theta}) - \boldsymbol{x_{ij}}\right) \circ \boldsymbol{m_{ij}} \right\|^2 \quad (1)$$

$$d_{ij}(u,v) = \sqrt{1/_{S^2} \sum_{p,q \in S} (x_i(p,q) - x_j(p,q))^2} \quad (2)$$

where $\boldsymbol{m_{ij}}$ is a mask function set to zero for $d_{ij}(u,v) < d_{th}$, and d_{ij} is a distance function computed over a path S of size s. For the *Noise2Sim* experiments, $\boldsymbol{x_j}$ was set to a randomly selected coronal slice within 1 cm of the noisy input. The values of d_{th} and s were optimized as a function of the training performance (s = 3 pixels and d_{th} = 0.004 mm^{-1}).

2.2.3. Noisier2Noise training with noise injection in the image domain (Noisier2Noise$_{img}$).

The *Noisier2Noise*[10] approach attempts to minimize the MSE loss computed using a synthetic version of the noisy image, $\boldsymbol{z_i}$, obtained by adding a noise realization obtained from a known noise distribution, $\mathcal{A}$, that matches the noise in $\boldsymbol{x_i}$. Thus, the input synthetic image is obtained as $\boldsymbol{z_i} = \boldsymbol{x_i} + \boldsymbol{n_{synth}} = \boldsymbol{s_i} + \boldsymbol{n_i} + \boldsymbol{n_{synth}}$, and the final loss function is computed as $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{i}}\left\| f(\boldsymbol{x_i} + \boldsymbol{n_{synth}}, \boldsymbol{\theta}) - \boldsymbol{x_i} \right\|^2$.

As shown in Ref. 10, under the assumption that $\boldsymbol{n_i}$ and $\boldsymbol{n_{synth}}$ are i.i.d., $\mathbb{E}[\boldsymbol{y}|\boldsymbol{z}] = 2\mathbb{E}[\boldsymbol{x}|\boldsymbol{z}] - \boldsymbol{z}$, which enables estimation of a denoised image ($\hat{\boldsymbol{y}}$) by doubling the network output and subtracting its input. In our *Noisier2Noise$_{img}$* approach we followed the simplest CT noise model and compute $\boldsymbol{n_{synth}}$ as a per pixel independent realization of a Poisson distribution, added directly in the reconstructed image domain, disregarding correlations introduced by the flat-panel detector and the backprojection operator.

2.2.4. Noisier2Noise training with noise injection in the projection domain (Noisier2Noise$_{proj}$).

The *Noisier2Noise$_{proj}$* explores a physically principled noise model for injection of noise in the projection domain, prior to image reconstruction. The generated synthetic input, $\boldsymbol{x_{synth}}$, included realistic models of signal formation, and correlations introduced by the detector and backprojection.

The noise model is described in Refs. 5, 6, yielding the following expressions for $\boldsymbol{x_{synth}}$ and the MSE loss:

$$\boldsymbol{x_{synth}} = FBP\left(\alpha \boldsymbol{p} + \left(\boldsymbol{w} \circ \sqrt{Q(\alpha \boldsymbol{p}) - \alpha^2 Q(\boldsymbol{p})}\right) * k\right) \quad (3)$$

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{i}}\left\| f\left(\boldsymbol{x_{synth_i}}, \boldsymbol{\theta}\right) - \boldsymbol{x_i} \right\|^2 \quad (4)$$

Where $\boldsymbol{p}$ is the set of projections images contributing to $\boldsymbol{x}$, $FBP()$ is the tomographic reconstruction operator, α is the dose reduction factor, set to assume addition of two equidose independent realizations, $\boldsymbol{w}$ is a vector of noise sampled from a normal distribution $[\mathcal{N}(0,1)]$, Q is a function operator relating the variance of the quantum noise to the mean signal,[5,6] and k is the quantum noise correlation kernel, estimated from cascaded systems analysis. To investigate the impact of the mismatch in the noise correlation model, we scaled the width of k by a factor of 0x, 0.5x, 2x, and 3x its nominal width, thus, ranging from a mismatched model with no detector correlation to a mismatched model that triples the detector correlation length.

2.2.5. Training setup and experimental validation.

The 150 bCT volumes were split into 120 for training, 10 for validation, and 20 for testing. For each volume 10 coronal slices were extracted at random positions (total of 1200 training instances (100 validation, 200 testing). All denoising methods used a common U-NET architecture and the same augmentation methods: random crop into 128 x 128 pixels patches, and random rotation at discrete values of 0, 90, 180, and 270 deg. All networks were trained with the Adam optimizer with learning rate decay and initial learning rate of 0.005. A Noise2Clean denoising was obtained to serve as upper bound for the self-supervised methods.

The denoising results were evaluated in terms of root mean squared error (RMSE) and structural similarity (SSIM) computed against the HD reference. Denoising inferences in two noise instances of the same anatomy were used to evaluate the NPS, measured on the difference between the two independent inferences. NPS was measured on 9 non-overlapping regions of interest (ROIs) with 50 x 50 pixels, fully contained within the breast region.

3. RESULTS

Fig. 1 shows RMSE and SSIM in the 200 test images, and Fig.2 show example image results for all self-supervised denoising methods. Quantitative evaluation confirmed that correlations in bCT noise pose a significant challenge to blind-spot denoising with pixel-wise independent noise assumptions, as in *Noise2Self* that yielded median RMSE = 0.0023 mm^{-1} (SSIM = 0.86), compared to RMSE = 0.0017 mm^{-1} (SSIM = 0.93) for the supervised approach, used as upper-bound performance level. The poor match between the bCT noise model and the image-based noise addition in *Noisier2Noise*$_{img}$ resulted in lower RMSE = 0.0037 mm^{-1} (SSIM = 0.84), comparable to the input ND image. Blind-spot denoising with *Noise2Sim* successfully accommodated noise correlations and provided superior denoising results, with RMSE = 0.0021 mm^{-1} (SSIM = 0.9). The slight lower performance compared to *Noise2Clean* arises from slight blurring around tissue boundaries and introduction of noise texture associated with the removal of dissimilarities from the loss function in Eq. 1 that tend to align with boundaries between adipose and fibroglandular tissue. Finally, *Noisier2Noise*$_{proj}$, with a perfectly matched noise correlation model, yielded the best performance with RMSE = 0.18 mm^{-1} (SSIM = 0.94), comparable to supervised denoising.

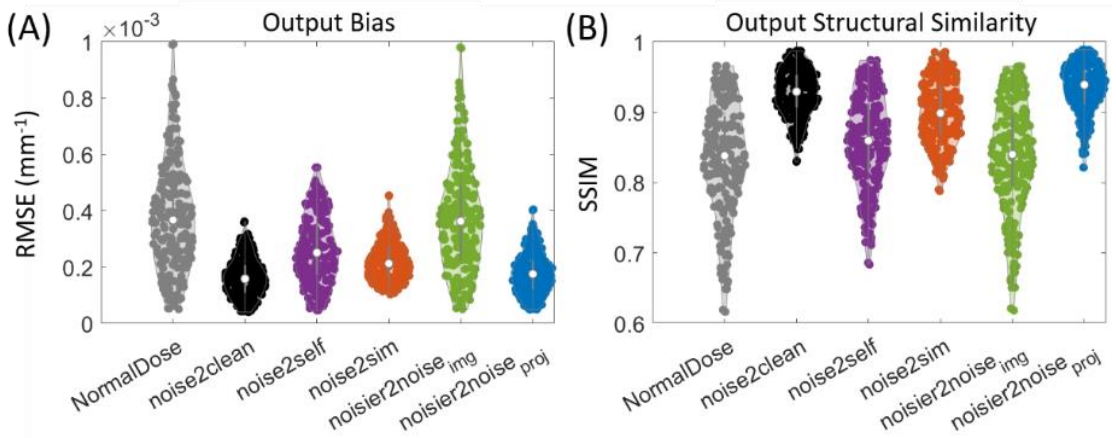

Figure 1. RMSE and SSIM as a function of the denoising approach. Noise correlation diminished the performance of *Noise2Self* and *Noisier2Noise*$_{img}$ while *Noise2Sim* and *Noisier2Noise*$_{proj}$ achieved results comparable with supervised training.

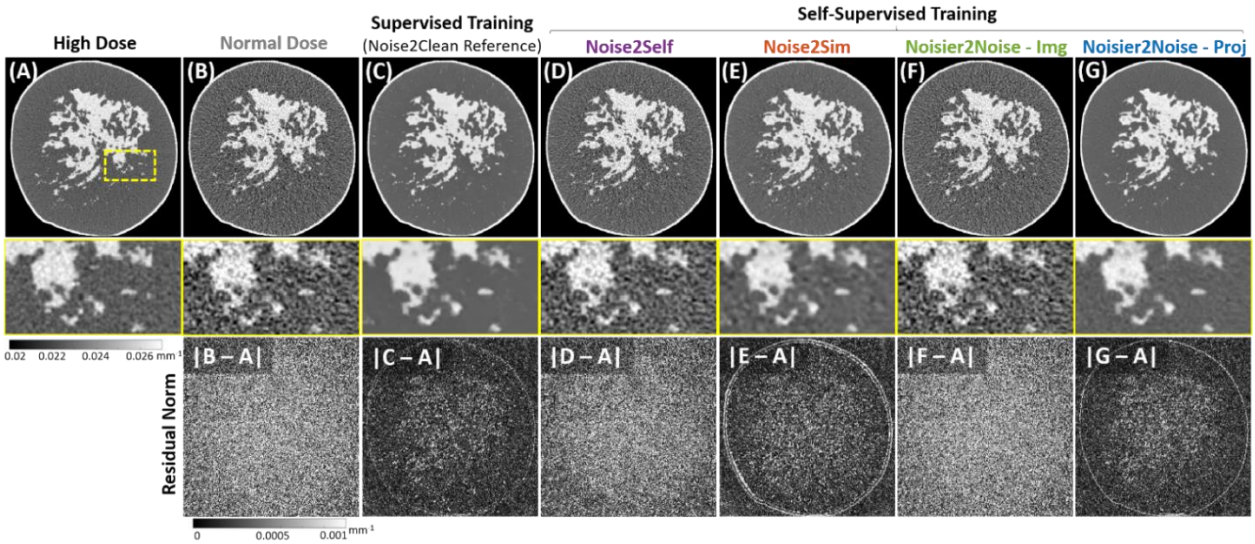

Figure 2. Image results for the HD reference (A), and noisier ND counterpart (B) in comparison with the reference supervised denoising *Noise2Clean* (C). Self-supervised denoising results for *Noise2Self* (D), *Noise2Sim* (E), *Noise2Noise*$_{img}$ (F), and *Noise2Noise*$_{proj}$ (G). Images in the bottom row quantify the residual norm computed as the difference image with respet to the HD reference.

The effects of denoising on noise texture are illustrated in Fig. 3, that shows example difference images from two input images with different noise realizations. Visual inspection of the difference images and NPS showed the inability of *Noise2Noise*$_{img}$ and *Noise2Self* to achieve successful denoising. In the case of *Noise2Sim* and *Noisier2Noise*$_{proj}$ both achieved a reduction in mean NPS, and showed some propagation of image structure into the noise, resulting in a shift of the NPS towards low frequency bands. This effect is more conspicuous for *Noise2Sim*, from instabilities at tissue boundaries induced by the masking approach. Quantification of noise power in low, mid and high frequency bands (Fig. 4) corroborated the better agreement in noise texture between the HD reference and *Noisier2Noise*$_{proj}$.

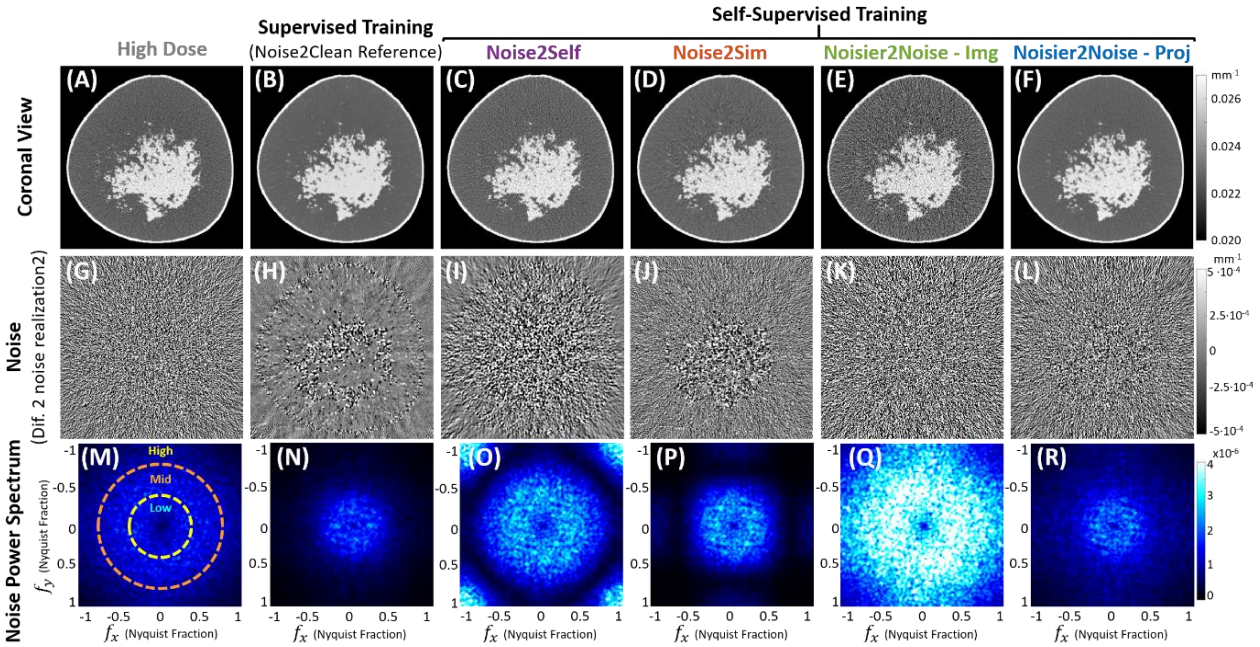

Figure 3. Example HD reference (A) and denoising results for one of the testing cases (B-F). Noise images, computed as the difference in output between two independent realizations quantify the pixel-wise noise level of the denoised output, as well as the propagation of anatomical structures into the noise pattern. (M-R) NPS for the ND reference and the denoised images show overall noise reduction as well as shift in the noise frequency content and, therefore, noise texture for the different apporaches. Both *Noise2Sim* and *Noisier2Noise*$_{proj}$ achieved noise levels comparable to *Noise2Clean*, with *Noisier2Noise*$_{proj}$ NPS more closely resembling the HD noise texture.

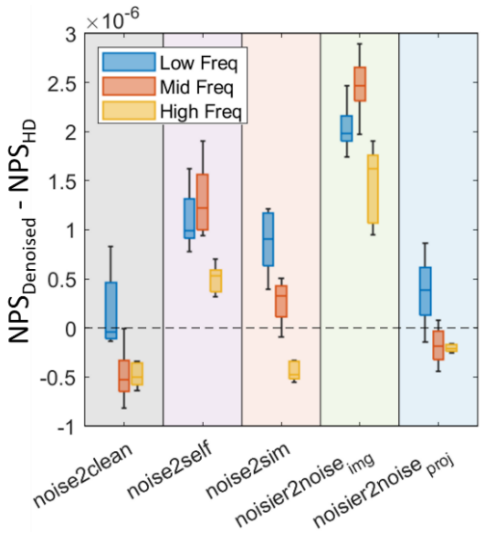

Figure 4. Quantification of the difference in total noise for 3 diferent frequency bands (low, mid, amd high), illustrated in Fig. 3. *Noise2Sim* showed slight shift towards low-frequency noise, while *Noisier2Noise*$_{proj}$ resulted in better agreement with the noise frequency distribution of HD and *Noise2Clean* across bands.

The results in Fig. 3 and Fig. 4 were obtained with a perfectly matched noise model that might not be always available. Fig. 5 illustrates the impact of deviations on the width of the detector noise correlation kernel. When no correlation is applied and only correlations induced by the backprojection operator are included, a decrease in denoising performance is observed, resulting in slight blurring at tissue boundaries and shift of NPS towards low frequencies. Effects of models imparting longer correlation range than the ideal model resulted in much lower impact and showed similar denoising capabilities and noise frequency distribution than those obtained with a perfectly matched model.

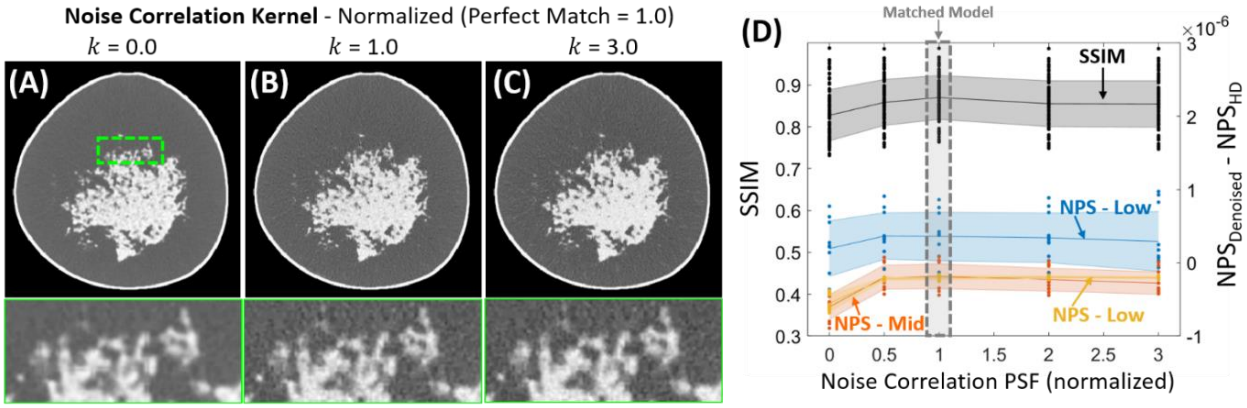

Figure 5. Results for Noisier2Noise with mismatched detector noise correlation kernel, ranging from no correlation for $k = 0$ (A), to perfect correlation matching $k = 1$ (B), to a much longer correlation kernel $k = 3$ (C). (D) Structural similarity and NPS difference with the HD reference for the low, mid, and high frequency band. Mismatches in noise correlation resulted in moderate degradation of SSIM and increased blurring, as well as shift of noise texture towards low frequencies. Mismatched models with long correlation kernels showed more robust performance, with similar SSIM and NPS distriubution as the one obtained with the perfectly matche noise model.

4. CONCLUSIONS

This works presents a comprehensive study of the performance of self-supervised DL denoising in bCT in the presence of noise correlation. A representative set of training approaches were used with synthetic bCT data including accurate models of signal and propagation and noise spatial correlation. Noisier2Noise models with noise injection in the projection domain showed denoising performance comparable to supervised training while yielding a noise frequency distribution similar to high-dose data.

ACKNOWLEDGEMENTS

Work supported by academic-industry collaboration with Izotropic Corp. and NIH (R01-EB030547).

REFERENCES

[1] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, C. Lin, "Deep Learning on Image Denoising: An overview." arXiv:1912.13171v4 (2020).

[2] V. Wang, A. Wei, J. Tan, S. Lu, W. Cao, Y. Gao, "A comparison study of deep learning designs for improving low-dose CT denoising." Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115962N (2021); DOI: 10.1117/12.2582322

[3] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data." arXiv:1803.04189 (2018).

[4] A. Sarno, G. Mettivier, F. di Franco, A. Varallo, K. Bliznakova, A. M. Hernandez, J. M. Boone, P. Russo, "Dataset of patient-derived digital breast phantoms for in silico studies in breast computed tomography, digital breast tomosynthesis, and digital mammography." Medical Physics, 48(5):2682 (2021); DOI: 10.1002/mp.14826

[5] P. Wu, A. Sisniega, A. Uneri, R. Han, C. Jones, P. Vagdargi, X. Zhang, M. Luciano, W. Anderson, J. H. Siewerdsen, "Using Uncertainty in Deep Learning Reconstruction for Cone-Beam CT of the Brain." arXiv:2108.09229 (2021).

[6] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, A. J. Khanna, G. L. Gallia, J. H. Siewerdsen, "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT." Medica Physics, 41(7):071915 (2014); DOI: 10.1118/1.4884039

[7] P. M. Gazi, K. Yang, G. W. Burkett Jr, S. Aminololama-Shakeri, J. A. Seibert, J. M. Boone, "Evolution of spatial resolution in breast CT at UC Davis." Medical Physics, 42(4):1973 (2015); DOI: 10.1118/1.4915079

[8] J. Batson, L. Royer, “Noise2Self: Blind Denoising by Self-Supervision.” arXiv:1901.11365 (2019).

[9] C. Niu, M. Li, F. Fan, W. Wu, X. Guo, Q. Lyu, G. Wang, “Suppression of Correlated Noise with Similarity-based Unsupervised Deep Learning.” arXiv:2011.03384 (2020).

[10] N. Moran, D. Schmidt, Y. Zhong, P. Coady, “Noisier2Noise: Learning to Denoise From Unpaired Noisy Data.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 12064 (2020).