

Deep learning ring artifact correction in photon-counting spectral CT with perceptual loss

Dennis Hein^{a,c}, Konstantinos Liappis^{a,c}, Alma Eguizabal^{b,c}, and Mats Persson^{a,c}

^aDepartment of Physics, KTH Royal Institute of Technology, Stockholm, Sweden

^bDepartment of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden

^cMedTechLabs, BioClinicum, Karolinska University Hospital, Solna, Sweden

ABSTRACT

Photon-counting spectral CT is a novel technology with a lot of promise. However, one common issue is detector inhomogeneity which results in streak artifacts in the sinogram domain and ring artifacts in the image domain. These rings are very conspicuous and limit the clinical usefulness of the images. We propose a deep learning based image processing technique for ring artifact correction in the sinogram domain. In particular, we train a UNet using a perceptual loss function with VGG16 as feature extractor to remove streak artifacts in the basis sinograms. Our results show that this method can successfully produce ring-corrected virtual monoenergetic images at a range of energy levels.

Keywords: Deep learning, perceptual loss, photon-counting CT, ring artifacts.

1. INTRODUCTION

Photon-counting spectral computed tomography (CT) is a promising novel technology for next-generation CT scanners.¹⁻³ Advantages of photon-counting detectors, compared to standard energy-integrating detectors, include higher contrast-to-noise ratio and spatial resolution, and improved low-dose imaging. One common issue in photon-counting CT is detector inhomogeneity, which results in energy threshold variation across detector elements, and, if not corrected for, leads to streak artifacts in the sinogram domain and ring artifacts in the image domain. This type of inhomogeneity can emerge due to an insufficiently calibrated forward model, temperature differences, and defective pixels.⁴ Many methods have been suggested for artifact and noise reduction in CT imaging and lately there has been a shift towards deep learning as a way to tackle these problems.⁴⁻⁹ In this work, we add to this literature by training a deep neural network for ring artifact correction in the sinogram domain and demonstrating its effectiveness in reducing ring artifacts for virtual monoenergetic images at a range of energy levels in photon-counting spectral CT.

2. METHOD

2.1 Photon-counting spectral CT

2.1.1 Material decomposition

Consider a multi-bin system with $B > 2$ energy bins and, for simplicity, a 2-dimensional image space. The material decomposition starts with the ansatz that the X-ray linear attenuation coefficient $\mu(x, y; E)$ can be approximated by a linear combination of M basis materials

$$\mu(x, y; E) \approx \sum_{m=1}^M a_m(x, y) \tau_m(E), \quad (1)$$

Further author information: (Send correspondence to D. Hein)

D. Hein: E-mail: dhein@kth.se

where a_m and $\tau_m(E)$ are the basis coefficients and basis functions, respectively. It is usually resolved in the sinogram domain and thus the target variables are the material line integrals

$$A_m(\ell) = \int_{\ell} a_m(x, y) ds = \mathcal{R}(a_m), \quad (2)$$

where \mathcal{R} denotes the Radon transform operator. The expected number of photons in energy bin j follows the polychromatic Beer-Lambert law

$$\lambda_j(\mathbf{A}) = \int_0^{\infty} \omega_j(E) \exp\left(-\sum_{m=1}^M A_m \tau_m(E)\right) dE. \quad (3)$$

This is our forward model. Finally, the measured data is the vector $\mathbf{y} := [y_1, \dots, y_B]$ where for each j we assume that

$$y_j \sim \text{Poisson}(\lambda_j(\mathbf{A})). \quad (4)$$

Hence, the (non-linear) inverse problem is to map the photon counts \mathbf{y} to the material line integrals $\mathbf{A} := [A_1, \dots, A_M]$. The most common approach to this problem is maximum likelihood.¹⁰⁻¹² Setting up the objective as the log likelihood and simplifying yields

$$\begin{aligned} \min_{\mathbf{A}} \quad & \sum_{j=1}^B (\lambda_j(\mathbf{A}) - y_j \log(\lambda_j(\mathbf{A}))) \\ \text{s.t.} \quad & A_i \geq 0 \quad \forall i = 1, \dots, M. \end{aligned} \quad (5)$$

This is subsequently solved using some iterative algorithm, e.g., the logarithmic barrier method.¹³

2.1.2 Data generation

After generating numerical basis material phantoms (soft tissue, bone and iodine) by segmenting CT images from the KiTS19 dataset,⁹ photon-counting imaging was simulated using the `fanbeam` function in Matlab and a spectral response model of a photon-counting silicon detector¹⁴ with 0.5×0.5 mm² pixels. The simulation was performed for 120 kVp and 200 mAs with 1579 detector pixels and 1600 view angles. After simulating Poisson noise, the maximum likelihood method was used for material decomposition of the simulated sinograms into bone and soft tissue basis sinograms, which were then reconstructed on a 583×583 pixel grid. To avoid streak artifacts due to photon starvation, a logarithmic barrier function was used to penalize large negative basis projection values. To simulate the effect of threshold variations, the simulation was performed with a random threshold shift ($\sigma = 0.5$ keV) applied independently to each of the eight thresholds of each detector pixel, and two material decompositions were performed: one with the actual bin thresholds that were used in the simulations, including the random shift, and one with the nominal bin thresholds, where the latter configuration yields images with ring artifacts.

2.2 Deep learning

2.2.1 Problem statement

We propose an image processing technique for ring artifact correction based on deep neural networks. More formally, let $\mathbf{x} \in \mathbb{R}^{M \times H \times W}$ denote the streak corrupted basis sinograms and $\mathbf{y} \in \mathbb{R}^{M \times H \times W}$ their streak artifact free counterpart, where M, H and W are the number of basis materials, view angles and detector pixels, respectively. Then our objective is to learn

$$f : \mathbf{x} \rightarrow \mathbf{y}. \quad (6)$$

We let f_{θ} be a neural network and learn the map (6) by learning parameters θ .

2.2.2 Network architecture

UNet is an widely utilized architecture for a range of different tasks in biomedical imaging. The defining feature is the encoder-decoder structure. We use a version of the original UNet¹⁵ shown in Fig. 1.

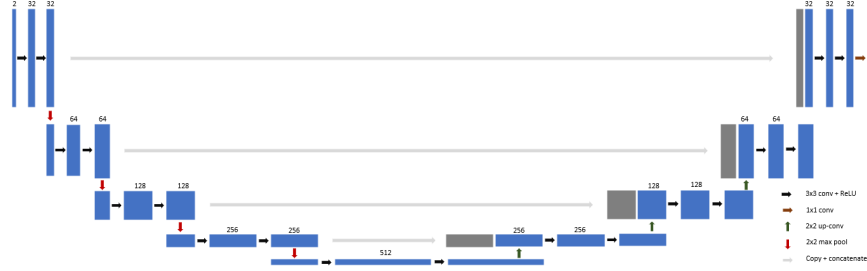


Figure 1. Illustration of UNet.¹⁵

2.2.3 Loss functions

Mean square error (MSE) is perhaps the most commonly used loss function for applications of deep learning in biomedical imaging

$$\ell_{mse} = \frac{1}{MHW} \|f_{\theta}(\mathbf{x}) - \mathbf{y}\|_2^2. \quad (7)$$

Using MSE loss will encourage the output to match the target pixel-by-pixel. This low-level per-pixel comparison is well known to produce output that is overly-smooth and lacking fine details that affect the perceptual quality.^{16,17} For several image transformation tasks, it has proved useful to instead employ a perceptual loss function which, instead of comparing pixel-by-pixel, compares high-level feature representations between the output and target. These feature representations are extracted from a pretrained convolutional neural network. We follow¹⁶ and use VGG16¹⁸ pretrained on ImageNet¹⁹ as feature extractor, or loss network. Let ϕ_j denote the j -th layer of VGG16, then our perceptual loss is defined as

$$\ell_{vgg} = \frac{1}{C_j H_j W_j} \|\phi_j(f_{\theta}(\mathbf{x})) - \phi_j(\mathbf{y})\|_2^2, \quad (8)$$

where C_j is the number of channels in layer j . We will set $j = 9$ which corresponds to “relu2_2” in.¹⁶

3. TRAINING DETAILS

From each of the 1600×1579 basis sinograms we extract 20 256×256 patches. A total of 630 samples, yielding 12600 patches, are split 70/30 into a training and validation set. The network is trained using Adam²⁰ with $\beta_1 = 0.5$, $\beta_2 = 0.9$, and learning rate $\gamma = 1 \times 10^{-4}$ for 100 epochs with a batch size of 16 on one NVIDIA GeForce RTX 3070 Laptop GPU. We standardize the input by dividing by the channel-wise standard deviation. We can obtain ring corrupted data with a range of different artifact magnitudes by taking a linear combination of streak corrupted and streak free basis sinograms. In this work, we are mainly concerned with the case when the rings are barely perceptible. Let w denote the weight given to the ring corrupted basis sinogram and $(1 - w)$ the weight given to its ring free counterpart. We found that $w = 0.4$ produces a realistic level for the artifacts and $w = 1$ a suitable level to train on.

4. RESULTS

4.1 Qualitative results

Qualitative results are available in Fig. 2 and 3. First, in Fig. 2, we have the results from the sinogram domain. Here, a pair of streak corrupted basis sinograms are passed through the network to produce the corresponding predicted pair. Note that despite training on 256×256 patches the network generalizes sufficiently to be able to deal with the entire 1600×1579 basis sinograms. The network does a fairly good job at removing the streaks. We subsequently reconstruct basis images from these sinograms and form virtual monoenergetic images at 40, 70, and 100 keV displayed in Fig. 3. Streak correction in the sinogram domain translates well into ring correction in the image domain. However, some residual rings are still visible. Note that, somewhat surprisingly, there is no significant difference in the performance of the network trained using MSE loss and the network trained using the perceptual loss.

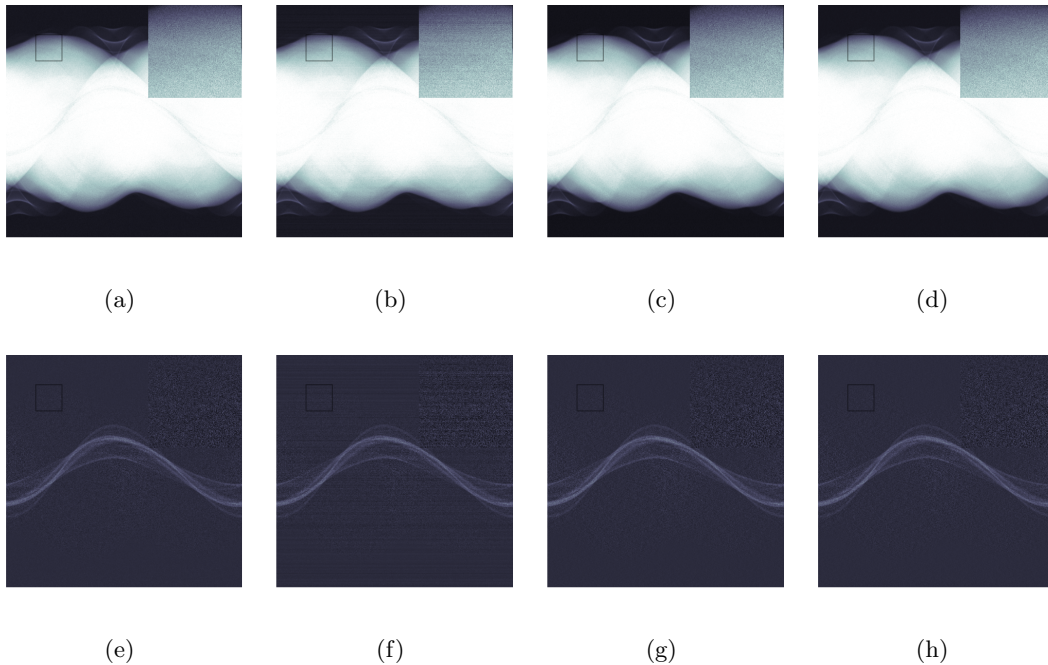


Figure 2. Basis sinograms. The square in top right corner shows a magnification of the indicated ROI. (a-d) soft tissue, (a) truth, (b) observed, (c) observed + UNet_mse, (d) observed + UNet_vgg, (e-h) bone, (e) truth, (f) observed, (g) observed + UNet_mse, (h) observed + UNet_vgg.

Table 1. Quantitative results

Network	SSIM	PSNR	Resolution (mm)
Truth	NA	NA	0.46
Observed	0.69	45.96	NA
UNet_mse	0.88	49.30	0.62
UNet_vgg	0.82	50.17	0.61

4.2 Quantitative results

Quantitative results are available in table 1. We employ the standard metrics used in this type of literature. Namely, structural similarity index measure (SSIM)²¹ and peak signal-to-noise ratio (PSNR). However, we appreciate the fact that these are not necessarily great metrics of perceptual quality* and instead stress our qualitative results. Note that, surprisingly, the network trained with a perceptual loss achieves higher PSNR than the network trained with MSE loss. However, this difference is sufficiently small to reasonably be attributed to stochastic variation in the optimization procedure. We also investigate the resolution by adding a central circular insert in the KiTS19 phantoms and retrieving the edge spread function as an average over radial profiles in the region of interest (ROI). We then fit a Gaussian error function to estimate the resolution as its standard deviation. Both networks produce a slight decrease in resolution.

5. CONCLUSION

Detector inhomogeneity, a common issue in photon-counting spectral CT, results in streak artifacts in the sinogram domain and ring artifacts in the image domain. In this work, we propose a deep learning image processing technique for ring artifact correction in the sinogram domain. Artifact corrupted data is generated by solving the material decomposition problem with a correctly and an incorrectly calibrated forward model. We trained a deep neural network to remove the streaks in the basis sinograms, which are subsequently reconstructed

*See e.g.,¹⁶ for a brief discussion.

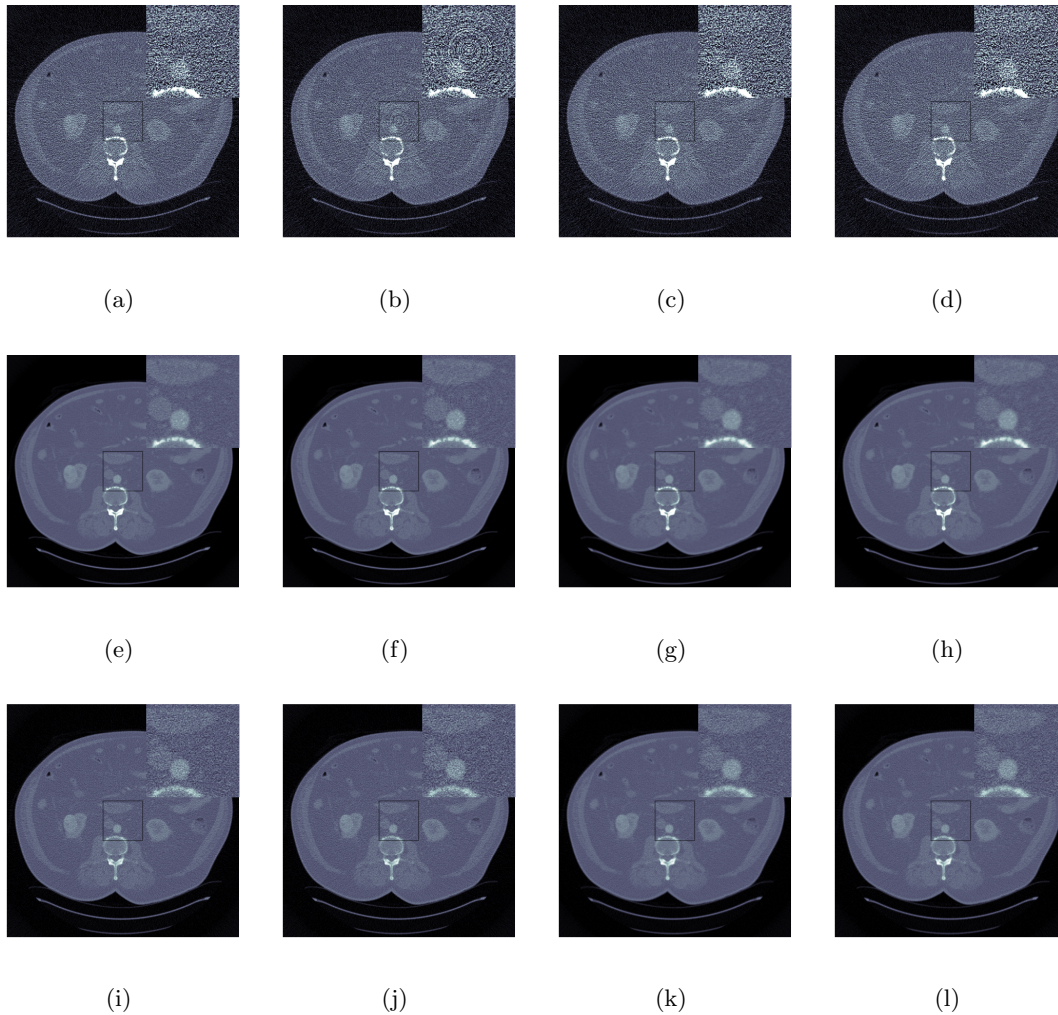


Figure 3. Virtual monoenergetic images at 40, 70 and 100 keV. The square in top right corner shows a magnification of the indicated ROI. (a-d) 40 keV, (a) truth, (b) observed, (c) observed + UNet_mse, (d) observed + UNet_vgg, (e-h) 70 keV, (e) truth, (f) observed, (g) observed + UNet_mse, (h) observed + UNet_vgg, (i-l) 100 keV, (i) truth, (j) observed, (k) observed + UNet_mse, (l) observed + UNet_vgg.

to produce ring corrected basis images and virtual monoenergetic images. Instead of training a network to produce output that is similar to target pixel-by-pixel, we use a perceptual loss function that encourages the feature representation of the output to be similar to that of the target. Unexpectedly, we found that the network trained using the standard MSE loss essentially performs on par with the network trained using the perceptual loss. Future research will address the slight degradation in resolution caused by the networks, investigate why the networks perform so similarly, and further develop this method on a larger and more diverse dataset.

ACKNOWLEDGEMENTS

This study was financially supported by MedTechLabs and the Göran Gustafsson foundation. Mats Persson and Dennis Hein disclose research collaboration with GE Healthcare. Alma Eguizabal discloses consultancy with GE Healthcare.

REFERENCES

- [1] Roessl, E. and Proksa, R., “K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors,” *Physics in Medicine and Biology* **52**, 4679–4696 (jul 2007).
- [2] Willemink, M. J., Persson, M., Pourmorteza, A., Pelc, N. J., and Fleischmann, D., “Photon-counting ct: Technical principles and clinical prospects,” *Radiology* **289**(2), 293–312 (2018). PMID: 30179101.
- [3] Danielsson, M., Persson, M., and Sjölin, M., “Photon-counting x-ray detectors for CT,” *Physics in Medicine & Biology* **66**, 03TR01 (jan 2021).
- [4] Fang, W., Li, L., and Chen, Z., “Removing ring artefacts for photon-counting detectors using neural networks in different domains,” *IEEE Access* **8**, 42447–42457 (2020).
- [5] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G., “Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss,” *IEEE transactions on medical imaging* **37**(6), 1348–1357 (2018).
- [6] Kim, B., Han, M., Shim, H., and Baek, J., “A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images,” *Medical physics (Lancaster)* **46**(9), 3906–3923 (2019).
- [7] Wang, Z., Li, J., and Enoch, M., “Removing ring artifacts in cbct images via generative adversarial networks with unidirectional relative total variation loss,” *Neural computing and applications* **31**(9), 5147–5158 (2019).
- [8] Nauwynck, M., Bazrafkan, S., Heteren, A. V., Beenhouwer, J. D., and Sijbers, J., “Ring artifact reduction in sinogram space using deep learning,” in [6th International Conference on Image Formation in X-Ray Computed Tomography], (2020).
- [9] Eguizabal, A., Persson, M. U., and Grönberg, F., “A deep learning post-processing to enhance the maximum likelihood estimate of three material decomposition in photon counting spectral CT,” in [Medical Imaging 2021: Physics of Medical Imaging], Bosmans, H., Zhao, W., and Yu, L., eds., **11595**, 1080 – 1089, International Society for Optics and Photonics, SPIE (2021).
- [10] Grönberg, F., Lundberg, J., Sjölin, M., Persson, M., Bujila, R., Bornefalk, H., Almqvist, H., Holmin, S., and Danielsson, M., “Feasibility of unconstrained three-material decomposition: imaging an excised human heart using a prototype silicon photon-counting ct detector,” *European radiology* **30**(11), 5904–5912 (2020).
- [11] Alvarez, R. E., “Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis,” *Medical Physics* **38**(5), 2324–2334 (2011).
- [12] Ducros, N., Abascal, J. F. P.-J., Sixou, B., Rit, S., and Peyrin, F., “Regularization of nonlinear decomposition of spectral x-ray projection images,” *Medical Physics* **44**(9), e174–e187 (2017).
- [13] Boyd, S. and Vandenberghe, L., [Convex optimization], Cambridge university press (2004).
- [14] Persson, M., Wang, A., and Pelc, N. J., “Detective quantum efficiency of photon-counting CdTe and Si detectors for computed tomography: a simulation study,” *Journal of Medical Imaging* **7**(4), 1 – 28 (2020).
- [15] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in [Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015], *Lecture Notes in Computer Science*, 234–241, Springer International Publishing, Cham (2015).
- [16] Johnson, J., Alahi, A., and Fei-Fei, L., “Perceptual losses for real-time style transfer and super-resolution,” in [European Conference on Computer Vision], (2016).
- [17] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W., “Photo-realistic single image super-resolution using a generative adversarial network,” in [2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)], 105–114 (2017).
- [18] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” in [3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings], Bengio, Y. and LeCun, Y., eds. (2015).
- [19] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L., “Imagenet large scale visual recognition challenge,” *International journal of computer vision* **115**(3), 211–252 (2015).

- [20] Kingma, D. P. and Ba, J., “Adam: A method for stochastic optimization,” in [*3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*], Bengio, Y. and LeCun, Y., eds. (2015).
- [21] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).