Ringing Artefact Removal From Sparse View Tomosynthesis using Deep Neural Networks

Shabab Bazrafkan, Vincent Van Nieuwenhove, Joris Soons, Jan De Beenhouwer, and Jan Sijbers,

***Abstract*—Tomosynthesis imaging is a special type of X-ray-based image acquisition wherein several projections are acquired from a limited angular range around the object or in a line above it. Subsequently, from these projection images, cross-section images parallel to the detector are reconstructed. However, if only a small number of projection images is acquired on a limited angular range, the reconstruction images suffer from ringing artefacts. This issue has not been addressed in the literature due to the lack of any model-based approach explaining this phenomenon. In the clinical setting, these artefacts can hinder a correct diagnosis. In this work, a deep learning-based ringing artifact reduction algorithm is proposed. The deep learning network was trained on 45786 medical images, resulting in a substantial reduction of ringing artefacts in the tomosynthesis reconstructions. Based on the numerical and visual evaluations, a conclusion is made on the positive effect of a deblurring Deep Neural Network in getting higher quality outputs.**

***Index Terms*—Deep Neural Networks, Deep Learning, Tomosynthesis, Ring Artefact.**

I. Introduction

Tomosynthesis involves the reconstruction of a 3D object imaged with a relatively small number of views acquired over a limited angular range, resulting in typical image artifacts. Indeed, reconstructions with (filtered) backprojection ((F)BP) algorithms generate out-of-plane artifacts that appear as blurred versions of the in-plane objects. Under certain conditions, this blurring changes into ringing, in particular when the perpendicular distance from a ringing source to the imaging plane increases above a certain threshold. Indeed, ringing occurs by a similar mechanism as blurring but is caused by the small number of projections in a sweep. It is caused by a high-contrast structure (e.g. rib structure) that is outside the plane in focus and whose contribution to that imaging plane is not sufficiently blurred. As a result, each interval between two successive projections in the imaging plane is wider than the tomographic blurring, so that separate instances of the ringing source appear in the imaging plane.
In order to avoid the FPB reconstruction disadvantages, other numerical methods known as iterative reconstruction algorithms such as SIRT (Simultaneous Iterative Reconstruction Technique) and CGLS with total variation minimization are used to generate the tomosynthesis image. However the main problem of ringing artefacts persists even with these reconstruction methods. Fig 1 illustrates this issue wherein using a low number of projections induces a disturbing ringing artefact in the final reconstruction which is not describable by analytical models. Hence, there are no successful model-based algorithms that are able to suppress these type of artefacts. In the current study, a data-driven machine learning method known as Deep Learning is utilized to restore the tomosynthesis image. To the best of our knowledge, there are no other data-driven or model-based methods addressing this issue.
Deep learning approaches have been successfully applied to the medical image processing. One of the widely investigated applications of the DNNs in medical image processing is known as image restoration. In image restoration, the input and the output of the network are from the same class of data. They include the same content and represent the same object. Noise removal is an example of image restoration. Other work focussed on the removal of reconstruction artefacts in limited angle or few-view CT with Deep learning techniques as in [6], [7]. In the current study, a similar approach is applied to the tomosynthesis reconstructions in order to remove the ringing artefact caused by a limited number of projections.

In the next section, the Neural Network model used in the current study is explained followed by a discussion on the database and training procedure in section III. In section IV, the numerical and visual results are presented and conclusions and future works is discussed in section V.

II. Deep Neural Networks

A. Network Design

Convolutional Deep Neural Networks consist of consecutive processing units also known as layers which can be fully connected, convolution, deconvolution, pooling, or un-pooling operations [1]. Based on the application, the network designs could take advantage of different regularization units such as batch normalization [2], weight normalization and/or drop-out [3] techniques. One of the widely used network designs in image processing is known as Fully Convolutional Deep Neural Networks (FCDNN). These are DNNs wherein no Fully Connected operation is applied throughout the network design. All the layers are convolution, deconvolution, pooling, and un-pooling units. FCDNNs are used in applications in which the input and the output of the network are images. Restoration problems are good examples of the uses cases of FCDNNs. In our work, two FCDNNs known as Mixed Scale Dense (MSD) Deep Neural Network [4] and a deblurring DNN [5] are utilized to remove ringing artefacts in tomosynthesis reconstruction. These network architectures are explained in the following sections.

S. Bazrafkan, J. De Beenhouwer and J. Sijbers are with imec-Visionlab, Department of Physics, University of Antwerp, Antwerp, Belgium e-mail: {shabab.bazrafkan},{jan.debeenhouwer},{jan.sijbers}@uantwerpen.be.

V. Van Nieuwenhove and J. Soons are with Agfa NV, Mortsel, Belgium email: {vincent.vannieuwenhove},{joris.soons}@agfa.com.

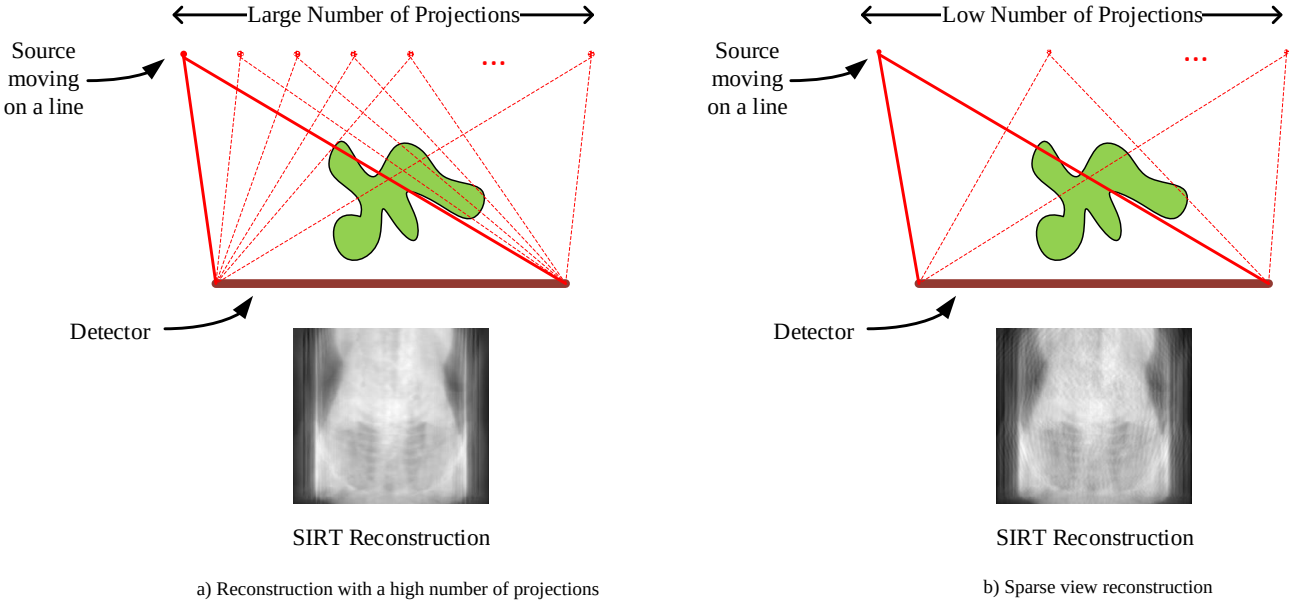


Fig. 1. a) A high number of projections returns a high quality reconstruction. b) Sparse view acquisition induces ring artefact in the final reconstruction.

1) Mixed Scale Dense (MSD) DNN: In order to explain the MSD network, one needs to understand the dilation property of the convolution operation. Dilation allows the convolution kernel to cover a larger area without using a higher number of parameters in larger kernels and/or applying a pooling operation. This is computationally more efficient because using bigger non-dilated kernels leads to a higher number of parameters which is susceptible to over-fitting, and using the pooling operation causes blurring in the final outcome. The dilation idea was first introduced in [4] and is illustrated in Fig. 4a. The operation is simply implemented by dilating the convolution kernel and filling the gaps with zero value.

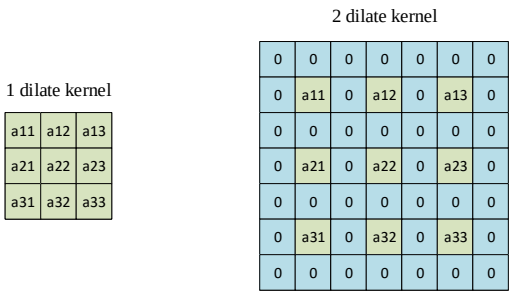


Fig. 2. 3×3 kernels. Left: 1 dilate. Right: 2 dilate.

The Mixed Scale Dense DNN as introduced in [4], is taking advantage of the dilation operation followed by concatenating the output of each layer with its input in the channel dimension. The network architecture is shown in Fig. 4a. The kernel dilation value is chosen by the layer number. This network has been already used to remove low dose CT artefacts from reconstructed images in [6], [7]. In the current study, this architecture is utilized to remove the ringing artefacts from low-dose tomosynthesis reconstructions.

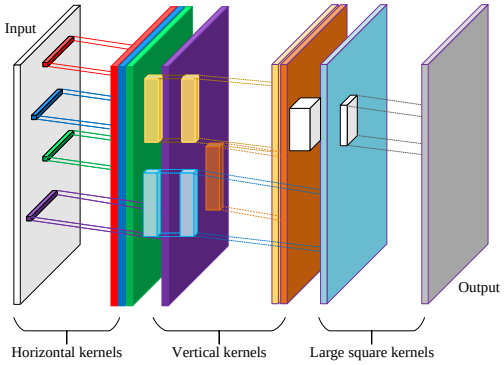
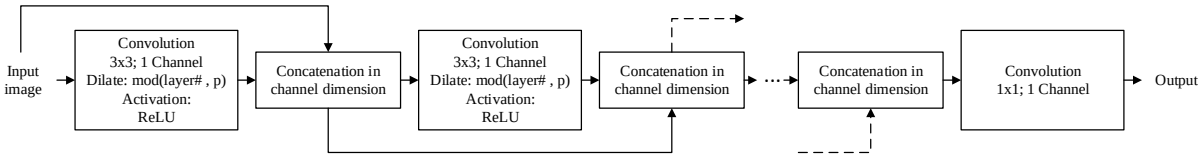
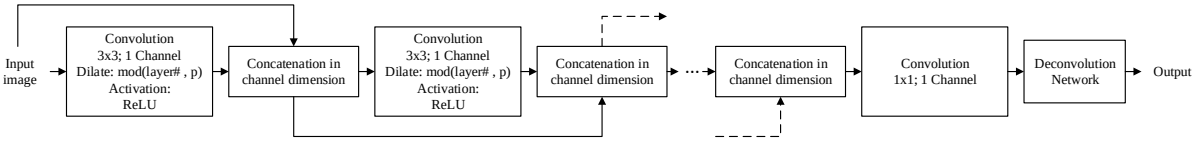


Fig. 3. Deconvolution network presented in [5].

2) Deblurring Convolutional Neural Network: Convolutional Deep Neural Networks tend to introduce a certain amount of blurring into the output especially when used with the Mean Squared Error as the loss function. However, Deblurring or Deconvolutional Convolutional Neural Networks have been proposed that are designed to reduce the blurring artefact in their input image. These networks are widely used to increase the quality of the image in super-resolution applications. In this article, a Deblurring network is used after the MSD network to improve the sharpness of the output. In [5], the authors proposed a network architecture designed for image deblurring, which is shown in Fig. 3. The first two layers consist of horizontal and vertical kernels and the last layer is a large square kernel. This design resembles the Singular Value Decomposition technique used in conventional deblurring methods, with the difference that here these filters are learned during training.



(a) MSD



(b) MSD+Deconv

Fig. 4. a) Mixed Scale Dense network. b) Mixed Scale Dense network followed by deblurring network.

B. Network Architectures

In the current study, two main architectures have been evaluated, and compared for the ring removal application.

1) **MSD**: The network is shown in Fig 4a. This is an MSD network with 25 layers and dilate rage p=5.
2) **MSD+Deconv**: The network is shown in Fig. 4b. This is an MSD network with 25 layers and dilate rage p=5 followed by a deblurring network shown in Fig. 3. The deblurring network uses the horizontal kernel size (1,129), vertical kernel size (21,1), and square kernel size (17,17). The number of channels in the first and second layers is 128 and 32, respectively.

III. Database and Training

In order to simulate the ringing artefacts and the corresponding ground truth, two different scanning geometries were defined, similar to Fig. 1. The first geometry takes 10 projections which eventually leads to reconstructions with the ringing artefacts and the second geometry uses 100 projections and the reconstruction is used as the ground truth. All simulations are accomplished using the ASTRA toolbox [8]. The reconstruction volumes were calculated with 25 iterations of SIRT. Two separate databases are used for training and testing purposes. Using different datasets is essential in evaluating the generality of the solution [9]. The training set is divided into Training and Evaluation subsets with the ratio 4 to 1, and the network is blind to the test set during the training process. Note that the networks are trained on the 2D samples in the transverse plane and results are shown in coronal plane where the ring artefact is visible.

A. Training Database

CPTAC-PDA: National Cancer Institutes Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma (CPTAC-PDA)[1] is a publicly available database containing 45786 Pancreas images from CPTAC phase 3 patients. It consists of 45 radiology and 77 pathology subjects. This database contains several modalities including CT, Computed Radiography (CR) and MRI samples. The images are from different sizes but in the current work, they were resized to 30×512 (transverse plane).

B. Test Database

Visible Human Project CT Datasets: Visible Human Project CT Datasets[2] contains 2989 images from 10 CT imaging cases. This dataset is publicly available. The images are 512×512 while in the current study they were all resized to 30×512 (transverse plane).

C. Training

The Mean Squared Error between the output of the network and the ground truth has been used as the loss function at the training stage. An ADAM optimizer has been utilized to update the parameters with learning rate, β_1, β_2 and ϵ equal to 0.0001, 0.9, 0.999, and 10^{-8} respectively. The MXNET 1.3.0[3] [10] framework has been used to train the network on top of python 2.7. The training and validation losses are shown in Fig 5. As it is shown in this figure, the deblurring network introduces a large gap between the loss functions which means that the MSD+Deconv network converges to a better solution at the training stage. In the next section, both networks are evaluated on the test database.

IV. Results

Table I shows the Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM) for both networks tested on 10 subsets of the test database. The column labeled "Input" corresponds to the comparisons between input and the ground truth. The MSD network returns outputs with up to 5dB improvement compared to the input signal and, the numerical results show 2 to 3 dB improvement for the MSD+Deconv network compared to MSD. This amount of improvement in the ranges higher than 40 dB implies that the deblurring network has a significant effect on the output quality.
Fig. 6 illustrates visual results from the Shoulder M subset and clearly shows that the output of the MSD+Deconv network is sharper and more consistent with the ground truth image.

[1] https://wiki.cancerimagingarchive.net/display/Public/CPTAC-PD

[2] https://mri.radiology.uiowa.edu/visible_human_datasets.html

[3] https://mxnet.apache.org/

	Input				MSD				MSD+Deconv			
	PSNR		SSIM		PSNR		SSIM		PSNR		SSIM	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Hip M	38.88	0.88	0.944	7.1e-5	44.05	0.97	0.980	2.5e-3	**47.08**	**0.95**	**0.989**	**1.5e-3**
Pelvis M	37.73	0.97	0.930	9.5e-3	41.71	1.13	0.969	6.2e-3	**44.25**	**1.0**	**0.982**	**2.5e-3**
Shoulder M	36.60	1.30	0.921	1.0e-2	40.25	**1.43**	0.960	7.1e-3	**43.65**	1.88	**0.980**	**7.1e-3**
Head M	43.98	1.60	0.969	7.0e-3	48.18	1.36	0.980	3.2e-3	**50.78**	**1.17**	**0.991**	**8.8e-4**
Hip F	40.97	2.57	0.957	1.4e-2	45.51	2.16	0.981	5.7e-3	**48.42**	**1.82**	**0.989**	**2.9e-3**
Pelvis F	38.31	0.20	0.935	4.4e-3	42.28	0.59	0.969	2.8e-3	**44.99**	**0.51**	**0.983**	**1.3e-3**
Shoulder F	36.98	1.42	0.929	9.9e-3	40.89	**1.67**	0.964	7.3e-3	**43.70**	2.16	**0.979**	**6.5e-3**
Head F	41.80	1.90	0.960	9.4e-3	46.44	1.65	0.979	3.9e-3	**49.41**	**1.64**	**0.991**	**1.4e-3**
Ankle F	42.60	2.30	0.964	8.1e-3	46.63	1.70	0.980	4.6e-3	**49.11**	**1.66**	**0.989**	**1.7e-3**
Knee F	42.38	0.74	0.967	4.3e-3	46.94	0.50	0.986	1.5e-3	**49.42**	**0.42**	**0.990**	**1.1e-3**

TABLE I
TEST RESULTS FOR BOTH MSD AND MSD+DECONV MODELS ON SUBSETS OF VISIBLE HUMAN PROJECT CT DATASETS

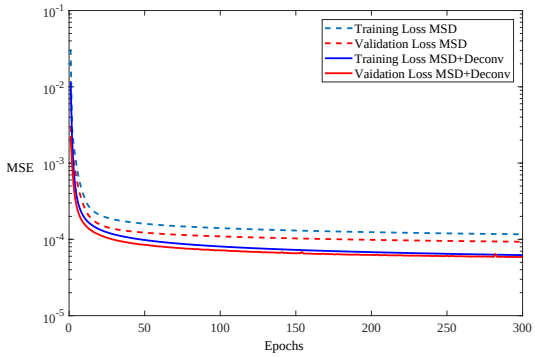


Fig. 5. Training and validation loss for both MSD and MSD+Deconv models.

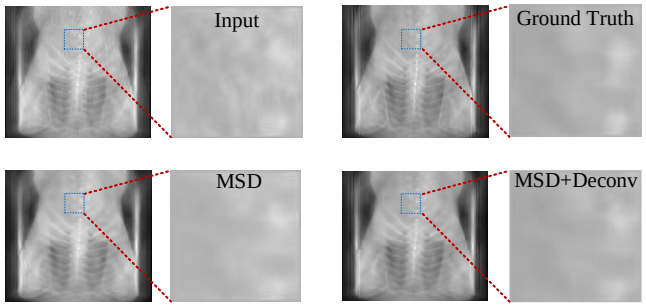


Fig. 6. Top left: The input to each network. Top right: Ground truth. Bottom left : The output of MSD network. Bottom right: the output of MSD+Deconv network. The visual results shows sharper and higher quality results for MSD+Deconv network.

V. CONCLUSION AND FUTURE WORK

In this article, a deep learning approach has been introduced to remove the ring artefacts from sparse view tomosynthesis reconstruction images. To the best of our knowledge, no other method presented in the literature addresses this problem and the current study shows the effectiveness of Deblurring networks in obtaining sharper and higher quality results using both numerical and visual assessments. Fully convolutional DNNs applied to restoration problems do not induce severe localization errors due to the convolution nature. In our initial visual evaluations (an example is depicted in Fig. 6), the network does not induce localization inconsistencies but our future works include educated evaluations by radiologists to ensure that the network does not alter the diagnosis procedure.

ACKNOWLEDGMENT

This work is financially supported by VLAIO (Flemish Agency for Innovation and Entrepreneurship), through the ANNTOM project HBC.2017.0595 and the Research Foundation Flanders (FWO) through the MetroFlex project S004217N.

REFERENCES

[1] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision." *IEEE Consumer Electronics Magazine*, vol. 6, no. 2, pp. 48–56, 2017.

[2] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[4] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 2, pp. 254–259, 2018.

[5] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in neural information processing systems*, 2014, pp. 1790–1798.

[6] S. Bazrafkan, V. Van Nieuwenhove, J. Soons, J. De Beenhouwer, and J. Sijbers, "Deep neural network assisted iterative reconstruction method for low dose ct," *arXiv preprint arXiv:1906.00650*, 2019.

[7] D. Pelt, K. Batenburg, and J. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks," *Journal of Imaging*, vol. 4, no. 11, p. 128, 2018.

[8] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Optics express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[9] S. Bazrafkan, T. Nedelcu, P. Filipczuk, and P. Corcoran, "Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods," in *Consumer Electronics (ICCE), 2017 IEEE International Conference on*. IEEE, 2017, pp. 217–220.

[10] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv:1512.01274*, 2015.