PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

New reconstruction methodology for chest tomosynthesis based on deep learning

C. F. Del Cerro, A. Galán, J. García Blas, M. Desco , M. Abella

C. F. Del Cerro, A. Galán, J. García Blas, M. Desco , M. Abella , "New reconstruction methodology for chest tomosynthesis based on deep learning," Proc. SPIE 12304, 7th International Conference on Image Formation in X-Ray Computed Tomography, 123042X (17 October 2022); doi: 10.1117/12.2646600

SPIE.

Event: Seventh International Conference on Image Formation in X-Ray Computed Tomography (ICIFXCT 2022), 2022, Baltimore, United States

New reconstruction methodology for thorax tomosynthesis based on deep learning

F Del Cerro. C, Galán. A, García Blas. J, Desco. M, Abella M.

ABSTRACT

Tomosynthesis offers an alternative to planar radiography providing pseudo-tomographic information at a much lower radiation dose than CT. The fact that it cannot convey information about the density poses a major limitation towards the use of tomosynthesis in chest imaging, due to the wide range of pathologies that present an increase in the density of the pulmonary parenchyma. Previous works have attempted to improve image quality through enhanced analytical, iterative algorithms, or including a deep learning-based step in the reconstruction, but the results shown are still far from the quantitative information of a CT. In this work, we propose a reconstruction methodology consisting of a filtered back-projection step followed by post-processing based on Deep Learning to obtain a tomographic image closer to CT. Preliminary results show the potential of the proposed methodology to obtain true tomographic information from tomosynthesis data, which could replace CT scans in applications where the radiation dose is critical.

Keywords: Chest tomosynthesis, Computed Tomography, Deep Learning, FDK-based reconstruction, Transfer Learning.

1. INTRODUCTION

Chest radiography is widely used for the diagnosis of lung diseases. It has a high spatial resolution, but being a projection image, tissues are shown overlapped without depth information and with a masking of low-density structures behind the bones. This limitation has been partially solved by tomosynthesis, a technique mainly used in mammography, but it still cannot convey density information. This is a major limitation for the use of tomosynthesis in chest imaging, due to the wide range of pathologies with increased lung parenchymal density. The only available technique to retrieve a density map is computed tomography (CT), but its use is limited by issues of availability and/or radiation dose control.

Several reconstruction strategies have been proposed to obtain an improved reconstruction of tomosynthesis data. In,[1] the authors proposed a Digital Breast Tomosynthesis (DBT) reconstruction based on the filtered back-projection (FBP) algorithm by a modification of the ramp filter, while[2,3] propose iterative reconstruction (IR) algorithms. Nevertheless, although these approaches increase image quality, the results shown are still far from the quantitative information of a CT.

During the last few years, image reconstruction methods have incorporated Deep Learning strategies as: (1) a post-processing step, after reconstruction,[4] (2) a pre-processing of the projection data,[5] or (3) within an IR algorithm, in the so-called unrolled methods.[6,7] Nevertheless, despite further improving image quality over traditional IR, none of the proposed methods show true tomographic results.

In this work, we propose a reconstruction methodology to obtain tomographic images (close to a CT) from chest tomosynthesis. The methodology consists of an FDK-based reconstruction step followed by post-processing based on Deep Learning.

C. F. Del Cerro, A. Galán, M. Abella and M. Desco are with the Dpto. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid and the Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (e-mail: mabella@ing.uc3m.es, cfernandez@hggm.es, agalan@hggm.es).

J. García Blas is with the Dpto. Arquitectura de Computadores, Comunicaciones y Sistemas, Universidad Carlos III de Madrid, Madrid, España (e-mail: fjblas@inf.uc3m.es).

M. Abella and M. Desco are with the Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España (e-mail: manuel.desco@cnic.es).

M. Desco is with the Centro de investigación en red salud mental (CIBERSAM), Madrid, España. (e-mail: desco@hggm.es).

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 123042X
 · 0277-786X · doi: 10.1117/12.2646600

2. DATABASE GENERATION

The database consists of tomosynthesis reconstructions as inputs and CT scans as a reference. We used 55 chest CT studies, 12 provided by the Hospital General Universitario Gregorio Marañón and 43 extracted from the Medical Imaging and Data Resource Center (MIDRC) portal from three different datasets. All studies were interpolated to a common isotropic voxel size, 1.25×1.25×1.25 mm, which is larger than that of any of the original CT volumes.

To simulate the tomosynthesis projections and perform the reconstructions, we used FuxSim,[8] a software that allows simulating different acquisition geometries based on a density map and system parameters. The simulations were based on a standard chest tomosynthesis protocol: 61 projections with a total source-to-detector distance of 1800 mm and a scan angle of 30 degrees. The resulting projections had a matrix size of 1072×1072 pixels with a pixel size of 0.4×0.4 mm. Since the patient is positioned next to the detector in the acquisition of chest tomosynthesis data, we cropped the volumes to remove the excess of air from the CT field of view and eliminated the bed and other patient supports, as shown in Fig. 1.

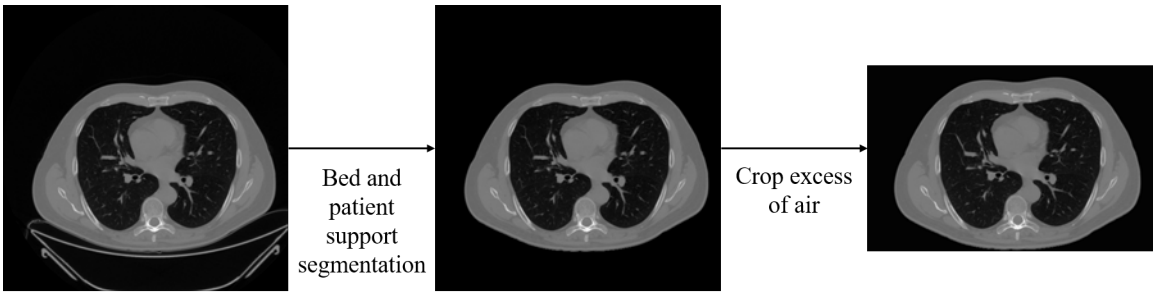

Figure 1. Pre-processing of CT data to simulate patient position in tomosynthesis system: removal of the bed, patient support and excess of air.

To avoid truncation artifacts in the tomosynthesis projections, caused by the spatial limitation of the chest CT scans in the vertical direction, we extended the CT volumes as shown in Fig. 2.

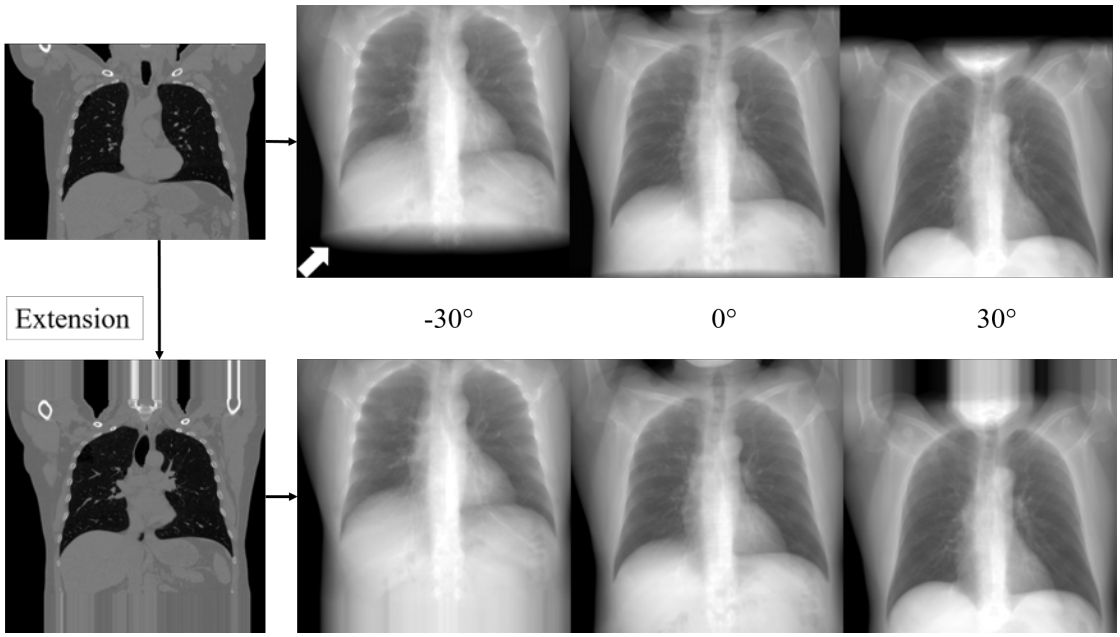

Figure 2. Tomosynthesis projections from CT for three different positions before (top) and after (bottom) extension. White arrow shows truncation artifact.

The simulated tomosynthesis data were reconstructed with an FDK-based algorithm, modifying the minimum value of the ramp filter to avoid the loss of mean value. For each study, we obtained a volume of 275×192×99 corresponding to the central part of the thorax, which is the ROI evaluated for diagnosis, with 1.25×1.25×1.25 mm voxel size. Fig. 3 shows the central coronal slice of the tomosynthesis reconstruction together with that of the reference CT for three of the volumes.

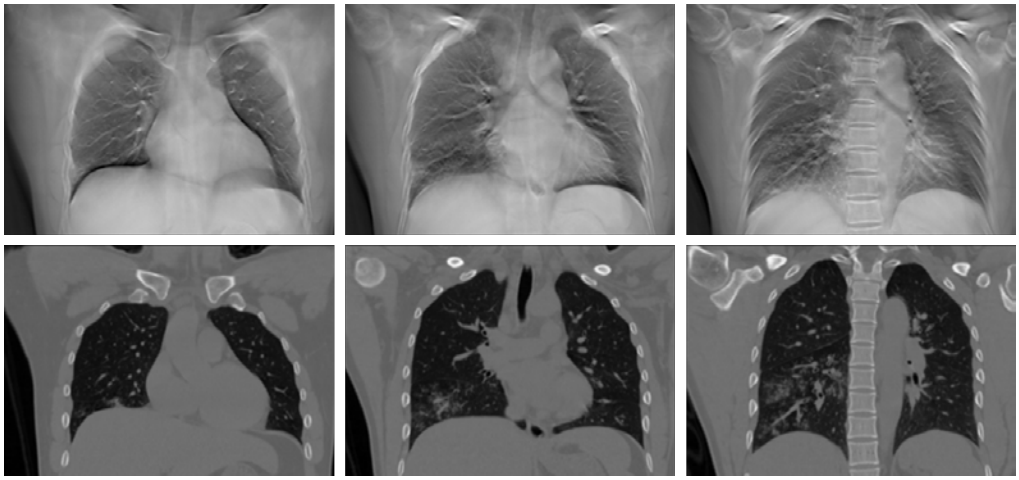

Figure 3. Coronal slices of the result of FDK-based reconstruction from the tomosynthesis data (top) and their corresponding CT (bottom).

The database was divided into the following sets:

- Training set: Composed of 38 volumes, giving a total of 3762 coronal slices and 7296 axial slices.
- Validation set: Composed of 10 volumes, giving a total of 990 coronal slices and 1920 axial slices.
- Test set: Composed of 6 volumes, giving a total of 594 coronal slices and 1152 axial slices.

3. POST-PROCESSING STEP

The post-processing step has as input the FDK-based reconstruction and as output a volume with enhanced tomographic information. The proposed architecture is a modification of the original version of the U-Net architecture[9] to include new strategies recently proposed in the literature. The encoder was replaced by ResNet-34,[10] due to the improved performance offered by the residual blocks over the original version of the U-Net encoder. The decoder consists mainly of oversampling blocks composed of a convolutional layer, followed by an average pooling layer (APL), a batch normalization (BN) layer, a ReLU, and two convolution layers (Conv2D) followed by a ReLU each. At the output of the first convolutional layer, the pixel shuffle operation with ICNR initialization[11] is applied. The complete network is shown in Fig. 4.

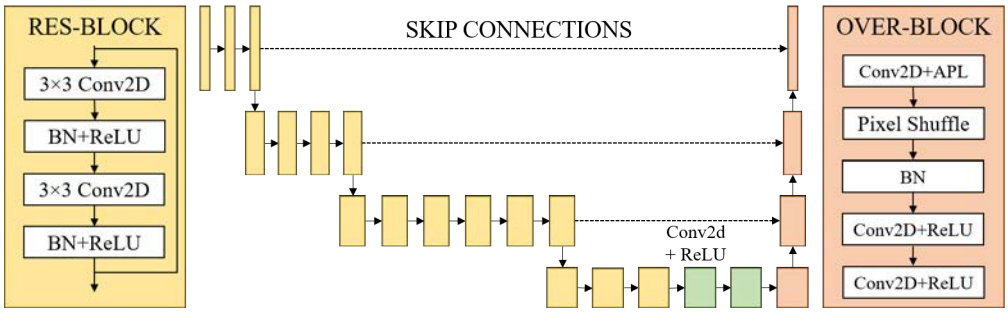

Figure 4. Proposed U-net network architecture.

For the initialization of the encoder, we used the pre-trained version of Resnet-34 for *Imagenet*.[12] The decoder was initialized randomly adapting it to the output of our problem. The training methodology was based on Transfer Learning,[13] first freezing the encoder to train only the decoder and then training the end-to-end network for fine-tuning.

Perceptual Loss (PL)[14] was chosen as the cost function to assure the recovery of fine details. We chose Adam[15] as the optimizer, due to its higher convergence speed, and a weight decay equal to 10^{-3} as regularization strategy. To determine the learning rate, we used the test presented by Leslie N. Smith,[16] which resulted in an optimal learning rate of 10^{-3} for decoder training and 10^{-5} for end-to-end network training.

Although the model processes 2D images, we incorporate 3D information by alternating coronal and axial slices in a multi-stage training strategy. As shown in Fig. 5, in the first stage we trained with the coronal view of the FDK-based reconstruction as inputs and the coronal CT view as target. In the next round, the coronal prediction of the first stage was resliced to obtain the axial view, this being the input at this stage and the target being the CT axial view. The axial view of the new prediction was then resliced to obtain the coronal view again and a new training was performed.

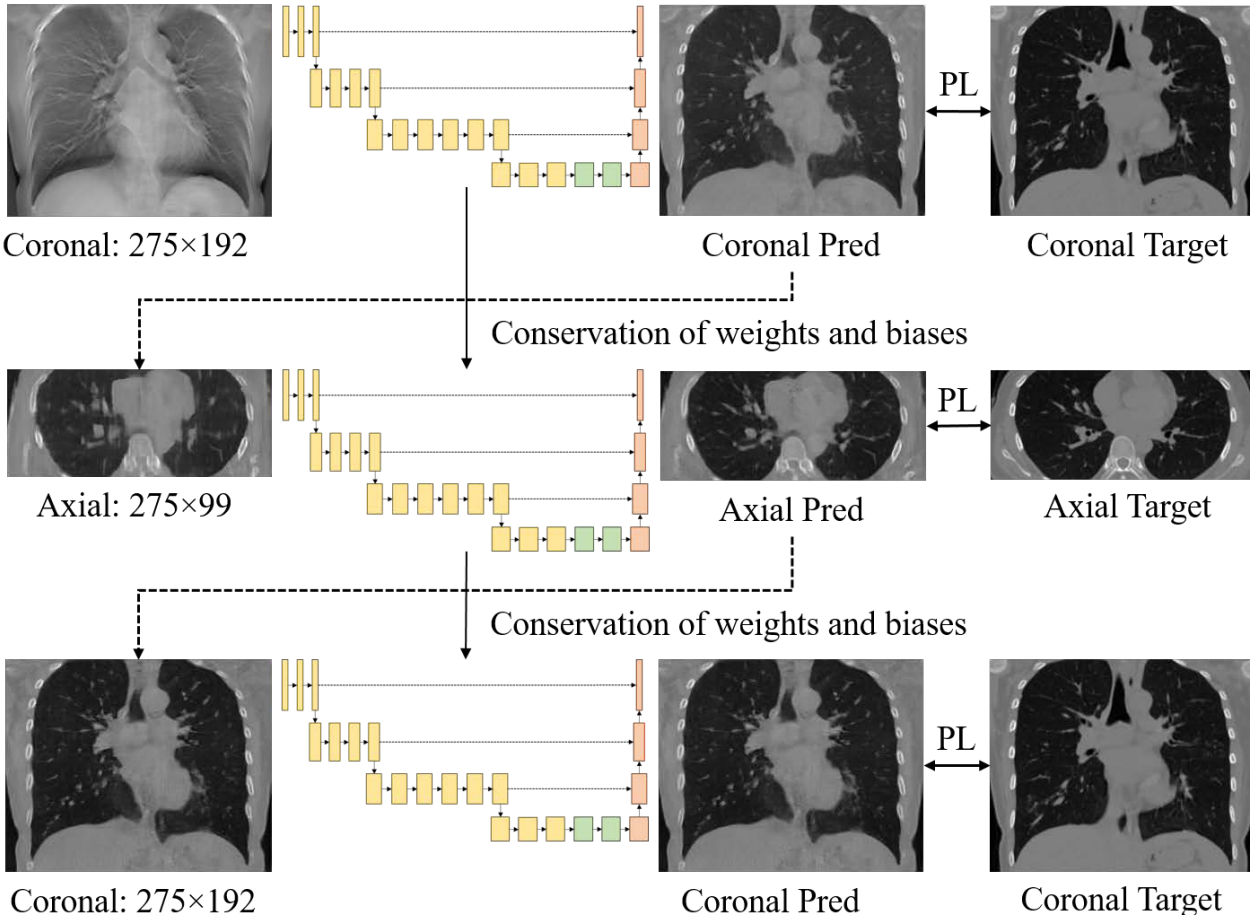

Figure 5. Scheme of the training strategy.

4. RESULTS

Fig. 6 shows a good recovery of general structure and realistic texture with the proposed reconstruction methodology, especially in areas with well-defined in the FDK-based reconstruction, such as the spine.

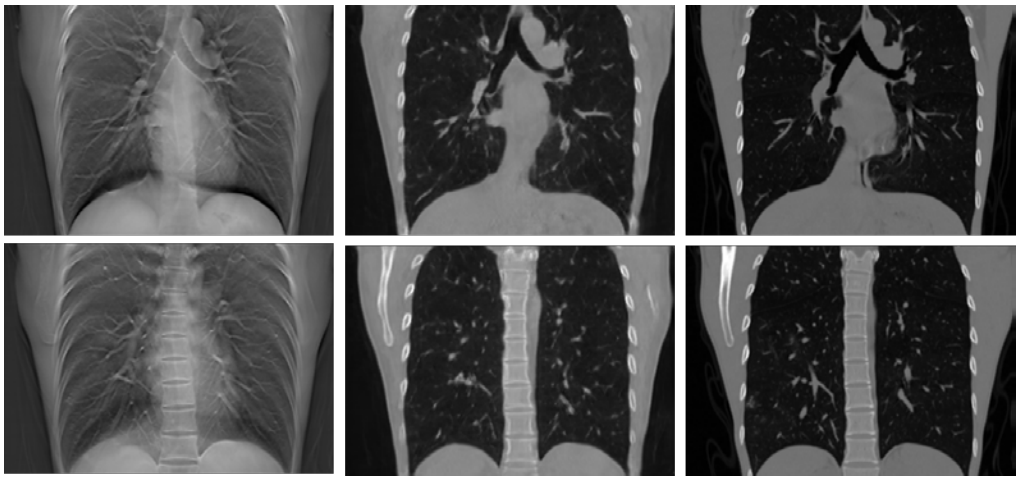

Figure 6. Coronal slices of: FDK-based reconstruction (left), prediction (center) and CT (right).

Fig. 7 shows the mean of the root mean square error (RMSE) with respect to the target CT for the six test volumes. We can see that our proposed methodology reduces the RMSE compared to FDK-based reconstruction in whole volume, but greater in the slices with tissues with higher contrast, such as spine and ribs (see coronal slices shown).

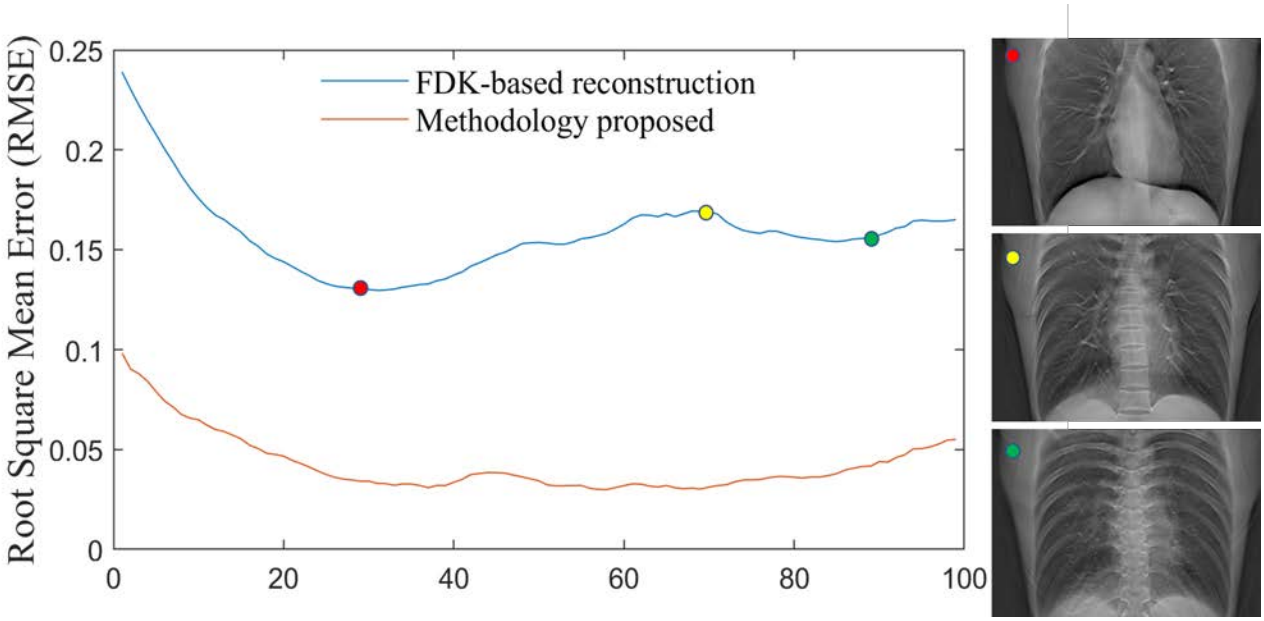

Figure 7. Mean value of RMSE for the six test volumes along the different coronal slices with three FDK-based reconstruction examples.

Fig. 8 show a zoom of coronal slices in Figure 6, where we can see a blurring, and more importantly, structures that do not exactly match those in the original CT (see white arrows).

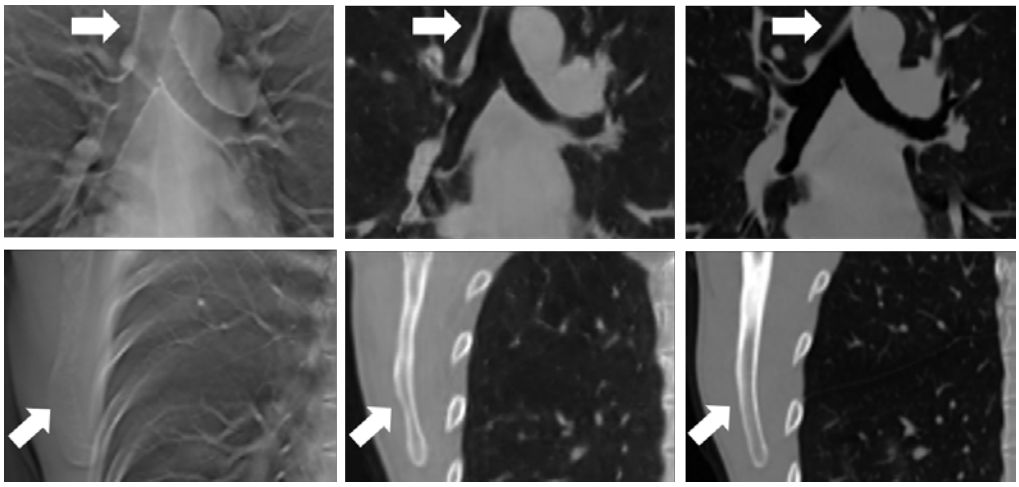

Figure 8. Zoomed ROIs from coronal slices in Fig. 6: FDK-based reconstruction (left), prediction (center) and CT (right). White arrow highlights structures wrongly.

Fig. 9 shows the predicted image after each stage. After the first stage, we can still see artifacts in the axial view from the FDK-based reconstruction, because the prediction is done on the coronal view and the FDK. The second stage, working on the axial view, these artifacts are removed, the contours are recovered, and texture is improved. The last stage further improves texture improves without altering shapes already found in previous stages.

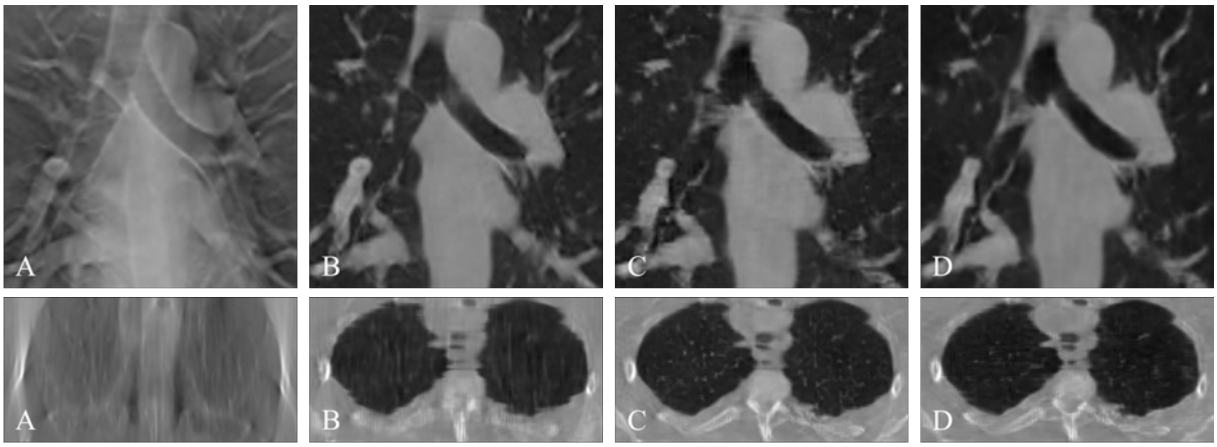

Figure 9. Coronal (top) and axial (bottom) slices of: FDK-based reconstruction (A), the first (B), second (C) and third (D) stages.

5. DISCUSSION

We have presented a reconstruction methodology for chest tomosynthesis based on FDK followed by post-processing based on deep learning. We use the transfer learning workflow, with U-Net architecture having a ResNet34 as encoder. Although our network processes 2D images, we incorporate 3D information by alternating coronal and axial slices in a multi-stage training strategy. The first stage of training uses the coronal view because it is the one with most information in FDK-based reconstruction. The evaluation shows a good reconstruction, especially in areas with high contrast and well-defined structures in the FDK-based reconstruction, such as the spine, and a very realistic tomographic texture. However, in the predicted images, we can find structures that do not match exactly with the original CT (biases). Looking at the result of the different stages, we can see that the biases appear in the first stage and are maintained through subsequent stages. Future work will evaluate

the combination of both prediction of previous stage together with the original FDK-based reconstruction at the input of the second and third stages to minimize the propagation of bias along the stages.

In this work we use a voxel size bigger than that of the original CTs due to computational limitations. This might be together with the subsampling layers of our U-Net might be responsible of the loss of spatial resolution. Future work will try to overcome the memory limitation by distributed training in using multiple GPUs.

ACKNOWLEDGMENTS

This work has been supported by Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación: PID2019-110369RB-I00/AEI/10.13039/501100011033 (RADHOR); PDC2021-121656-I00 (MULTIRAD), funded by MCIN/AEI/10.13039/501100011033 and by the European Union 'NextGenerationEU'/PRTR. Also funded by Comunidad de Madrid: Multiannual Agreement with UC3M in the line of 'Fostering Young Doctors Research' (DEEPCT-CM-UC3M), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation"); S2017/BMD-3867 RENIM-CM, co-funded by European Structural and Investment Fund. And also partially funded by CRUE Universidades, CSIC and Banco Santander (Fondo Supera Covid19), project RADCOV19 and by Instituto de Salud Carlos III through the project " PT20/00044", co-funded by European Regional Development Fund "A way to make Europe". The CNIC is supported by Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and the Pro CNIC Foundation.

The imaging and associated clinical data downloaded from MIDRC (The Medical Imaging Data Resource Center) and used for research in this publication was made possible by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contracts 75N92020C00008 and 75N92020C00021.

REFERENCES

[1] Rose, S. D., Sidky, E. Y., Reiser, I. S., and Pan, X., "Filtered back-projection for digital breast tomosynthesis with 2d filtering," in [*Medical Imaging 2019: Physics of Medical Imaging*], **10948**, 1094851, International Society for Optics and Photonics (2019).

[2] Rose, S. D., Sanchez, A. A., Sidky, E. Y., and Pan, X., "Investigating simulation-based metrics for characterizing linear iterative reconstruction in digital breast tomosynthesis," *Medical physics* **44**(9), e279–e296 (2017).

[3] Rodriguez-Ruiz, A., Teuwen, J., Vreemann, S., Bouwman, R. W., van Engen, R. E., Karssemeijer, N., Mann, R. M., Gubern-Merida, A., and Sechopoulos, I., "New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers," *Acta Radiologica* **59**(9), 1051–1059 (2018).

[4] Lee, D. and Kim, H.-J., "Restoration of full data from sparse data in low-dose chest digital tomosynthesis using deep convolutional neural networks," *Journal of digital imaging* **32**(3), 489–498 (2019).

[5] Gomi, T., Hara, H., Watanabe, Y., and Mizukami, S., "Improved digital chest tomosynthesis image quality by use of a projection-based dual-energy virtual monochromatic convolutional neural network with super resolution," *PLoS One* **15**(12), e0244745 (2020).

[6] Moriakov, N., Michielsen, K., Adler, J., Mann, R., Sechopoulos, I., and Teuwen, J., "Deep learning framework for digital breast tomosynthesis reconstruction," in [*Medical Imaging 2019: Physics of Medical Imaging*], **10948**, 1094804, International Society for Optics and Photonics (2019).

[7] Su, T., Deng, X., Yang, J., Wang, Z., Fang, S., Zheng, H., Liang, D., and Ge, Y., "Dir-dbtnet: Deep iterative reconstruction network for three-dimensional digital breast tomosynthesis imaging," *Medical Physics* (2021).

[8] Abella, M., Serrano, E., Garcia-Blas, J., García, I., De Molina, C., Carretero, J., and Desco, M., "Fux-sim: Implementation of a fast universal simulation/reconstruction framework for x-ray systems," *PloS one* **12**(7), e0180363 (2017).

[9] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[10] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[11] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1874–1883 (2016).

[12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision* **115**(3), 211–252 (2015).

[13] Kornblith, S., Shlens, J., and Le, Q. V., "Do better imagenet models transfer better?," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 2661–2671 (2019).

[14] Johnson, J., Alahi, A., and Fei-Fei, L., "Perceptual losses for real-time style transfer and super-resolution," in [*European conference on computer vision*], 694–711, Springer (2016).

[15] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[16] Smith, L. N., "Cyclical learning rates for training neural networks," in [*2017 IEEE winter conference on applications of computer vision (WACV)*], 464–472, IEEE (2017).