Fast Reconstruction of non-circular CBCT orbits using CNNs

Tom Russ[1], Wenying Wang[2], Alena-Kathrin Golla[1], Dominik F. Bauer[1], Matthew Tivnan[2], Christian Tönnes[1], Yiqun Q. Ma[2], Tess Reynolds[3], Sepideh Hatamikia[4,5], Lothar R. Schad[1], Frank G. Zöllner[1], Grace J. Gang[2], and J. Webster Stayman [2]

[1]Computer Assisted Clinical Medicine, Heidelberg University, Heidelberg, Germany
[2]Department of Biomedical Engineering, Johns-Hopkins University, Baltimore, USA
[3]ACRF Image X Institute, University of Sydney, Australia
[4]Austrian Center for Medical Innovation and Technology, Wiener Neustadt, Austria
[5]Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria

Abstract Non-circular acquisition orbits for cone-beam CT (CBCT) have been investigated for a number of reasons including increased field-of-view, minimal interference within an intraoperative environment, and improved CBCT image quality. Fast reconstruction of the projection data is essential in an interventional imaging setting. While model-based iterative reconstruction can reconstruct data from arbitrary geometries and provide superior noise suppression for a wide variety of non-circular acquisitions, such processing is particularly computationally intensive. In this work, we present a scheme for fast reconstruction of arbitrary non-circular orbits based on Convolutional Neural Networks (CNNs). Specifically, we propose a processing chain that includes a shift-invariant deconvolution of backprojected measurements, followed by CNN processing in a U-Net architecture to address artifacts and deficiencies in the deconvolution process. Synthetic training data is produced using orbital specifications and projections of a large number of procedurally generated objects. Specifically, attenuation volumes are created via randomly placed Delaunay tetrahedrons. We investigated the reconstruction performance for different sets of acquisition orbits including: circular, sinusoidal and randomized parametric trajectories. Our reconstruction scheme yields similar image quality when compared to simultaneous algebraic reconstruction technique (SART) reconstructions, at a small fraction of the computation time. Thus, the proposed work offers a potential way to utilize sophisticated non-circular orbits while maintaining the strict time requirements found in interventional imaging.

1 Introduction

The advent of robotic interventional x-ray systems has opened the door to dramatically increased flexibility in the design of CBCT acquisition trajectories. Such orbits have been used to increase the imaging field-of-view and to minimally interfere with the other equipment in the interventional suite; but also to improve image quality. For example, a large variety of non-circular orbits has been investigated to improve data completeness, metal artifacts, and task-based detectability [1, 2, 3, 4]. Typically, reconstruction algorithms for non-circular data have relied on both analytical and model-based methods. Analytical solutions exist for specific classes of non-circular orbits such as saddle trajectories [5]. Model-based iterative reconstruction (MBIR) implicitly handles arbitrary geometries (providing a “best” estimate based on the available data). These algorithms, however, are computationally expensive, which poses a major limitation particularly for interventional applications. The recent proliferation of data-driven and machine-learning-based reconstruction methods provides opportunities for superior reconstruction speed *and* image quality comparable to MBIR.

In this work, we propose a reconstruction scheme that leverages Convolutional Neural Networks (CNNs). In particular, we develop a processing chain where data backprojection is followed by a shift-invariant deconvolution step followed by CNN processing. The deconvolution is based on the orbital trajectory and the intrinsic system response but is only approximate. The CNN step is trained to mitigate deficiencies in this approximate deconvolution. Each of these steps is computationally efficient and non-iterative leading to a fast processing chain. The following sections detail this processing chain and its application to five different sets of orbit geometries. For comparison, an iterative reconstruction scheme, the simultaneous algebraic reconstruction technique (SART), is also applied and quantitative performance measures (relative to truth) are computed.

2 Materials and Methods

2.1 The Tomographic Reconstruction Problem

Presuming log-transformed projection data, tomographic reconstruction seeks to solve the following inverse problem:

$$y = \mathbf{A}(\Omega)\mu, \tag{1}$$

where y denote the measured line integrals of attenuation (e.g., projections) and μ is the distribution of attenuation values in the object. Here, we identify the dependence of the projection matrix, $\mathbf{A}$, on some parameterization of the acquisition orbit Ω. Classic inversion approaches often seek to find the pseudo-inverse:

$$\mu = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{2}$$

The pseudo-inverse has the advantage that solutions can be found for non-square and rank-deficient $\mathbf{A}$ that are possible for arbitrary trajectories.

We note that $\mathbf{A}^T$ represents a backprojection operation. Thus, the operator $(\mathbf{A}^T\mathbf{A})^{-1}$ represents a kind of generalized filtering operation. In fact, under idealized imaging conditions (parallel beam, sufficient sampling, etc.) and a circular acquisition geometry, $(\mathbf{A}^T\mathbf{A})$ represents the operator that applies the well-known intrinsic response of tomography - a $1/r$ blur function. Thus, in the ideal case, $(\mathbf{A}^T\mathbf{A})^{-1}$ is the inverse filter that removes $1/r$ blur. For non-circular orbits, divergent beams, etc., the blur induced by $(\mathbf{A}^T\mathbf{A})$ is not generally shift-invariant nor of the form $1/r$. However, these observations suggest a potential scheme for fast reconstruction using similar processing stages.

2.2 Proposed Reconstruction Pipeline

Motivated by the above observations, we propose a new reconstruction pipeline using neural networks but leveraging what we already know about the required reconstruction process. Specifically, we maintain the backprojection step and address the operator $(\mathbf{A}^T\mathbf{A})^{-1}$.

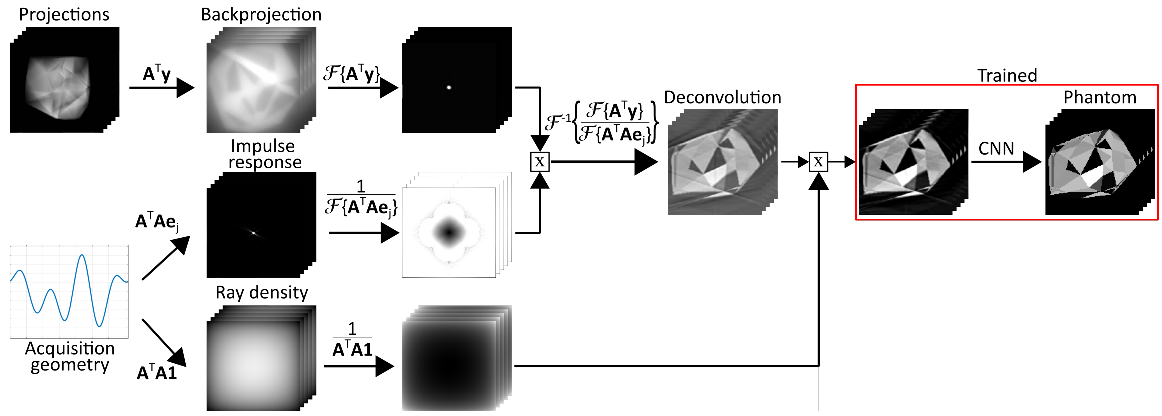

Figure 1: Flow chart illustrating the proposed reconstruction pipeline. We first deconvolve an approximation of the system response then deploy a CNN to remove residual artifacts.

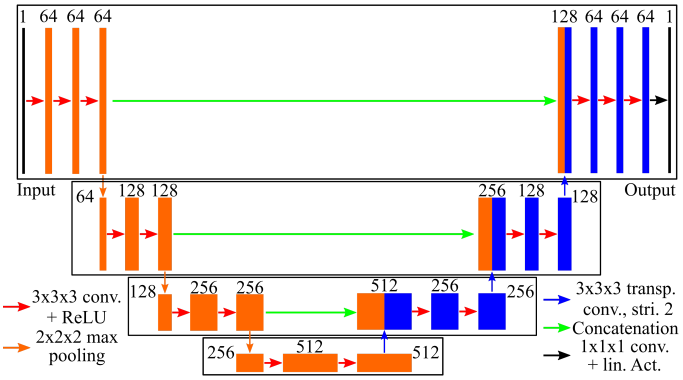

Figure 2: U-Net in the last step of the propose pipeline. Numbers over feature channel blocks indicate the number of channels. Max pooling halves the size of each dimension, whereas transposed convolution with stride two doubles the size of each dimension.

While one could develop a neural network to learn this inverse transform, there is an opportunity to provide a better network input. We will presume that the geometry, and hence $\mathbf{A}$ and $\mathbf{A}^T$ of non-circular acquisitions are known *a priori*. We can therefore devise network inputs to leverage such prior information. Specifically, if we can first deconvolve (in the case of a shift-invariant system) the system response, $\mathbf{A}^T\mathbf{A}$, from the backprojection, $\mathbf{A}^Ty$, we can effectively remove the dependency on geometry in the reconstruction process. Of course, such deconvolution procedure is noise amplifying and prone to artifacts. We therefore deploy a post-deconvolution CNN to remove residual artifacts.
Towards this end, we developed a processing chain illustrated in Fig.1. For initial investigation in this work, we assumed the system response to be approximately shift-invariant (true for small objects and/or long geometries) and approximated the system response as $\mathbf{A}^T\mathbf{A}e_j$ where e_j denotes an impulse at the center of the image. We first deconvolved $\mathbf{A}^T\mathbf{A}e_j$ from $\mathbf{A}^Ty$ via direct Fourier inversion, i.e.:

$$\mathscr{F}^{-1}\left\{\frac{\mathscr{F}\left\{\mathbf{A}^Ty\right\}}{\lfloor\mathscr{F}\left\{\mathbf{A}^T\mathbf{A}e_j\right\}\rfloor}\right\}. \tag{3}$$

We adopted several techniques to mitigate artifacts associated with the deconvolution process. First, a threshold operator was used in the denominator to avoid division by zero. (Specifically, a value of 0.0025 was applied.) Second, the backprojection volume was expanded to approximately four times the reconstruction volume to mitigate spurious frequencies as a result of the fast Fourier transform of signals with discontinuities at the boundaries. Third, to mitigate artifacts in $\mathscr{F}\left\{\mathbf{A}^T\mathbf{A}e_j\right\}$ due to the combined effect of voxel sampling and ray-based projector, we computed $\mathbf{A}^TAe_j$ at eight voxel locations around the central voxel of the image and averaged the responses.
After the deconvolution, we additionally corrected for sampling density by performing an element-wise division of the volume by $\mathbf{A}^T\mathbf{1}$, where $\mathbf{1}$ denotes a volume of 1s. We truncated the image to the same size as the reconstruction image volume to save memory. The resulting image volume was used as input to the CNN. In summary, the input to the CNN is represented mathematically as:

$$\mathbf{x} = \mathscr{F}^{-1}\left\{\frac{\mathscr{F}\left\{\mathbf{A}^Ty\right\}}{\lfloor\mathscr{F}\left\{\mathbf{A}^T\mathbf{A}e_j\right\}\rfloor}\right\}\frac{1}{\mathbf{A}^T\mathbf{A}\mathbf{1}}. \tag{4}$$

For the CNN processing step, we chose a U-Net architecture consisting of seven convolutional blocks illustrated in Figure 2. The U-net architecture was chosen due to its successful application in image deconvolution and CT reconstruction.
The input of the network detailed above consists of 128x128x128 voxel volumes. The network is trained to predict the ground truth phantom images of the same size. For training, the root-mean-square error (RMSE) between the prediction and the ground truth image was chosen as the loss function. Optimization was performed using an Adam optimizer with a learning rate of 0.001 and terminated after 100 training epochs. Among the 1000 phantom images, 800 images were used for training, 100 for validation, and 100 for testing. Details of the training and evaluation data follow.

2.3 Phantom and Data Generation

For imaging phantoms used in training and evaluation, we procedurally generated 1000 random realizations of Delaunay tetrahedrons. We randomly sampled 40 vertex locations in 3D, then created a tetrahedron mesh by connecting these

vertices using the 3D Delaunay triangulation algorithm in MATLAB. Within each tetrahedron, a uniform attenuation coefficient was randomly assigned based on the distribution of voxel values in an abdomen CT scan (*sans* background). The phantoms were then formed by voxelizing the meshes on a 128x128x128 grid with 0.5x0.5x0.5 mm^3 voxel spacing.
Data were simulated using the ASTRA toolbox [6, 7]. The imaging geometry used a source-axis distance and source-detector distance of 1 m and 0.5 m, respectively. The reconstruction volume matched the voxel size and spacing of the ground truth. The projection data were simulated on a 256x256 detector with pixel size 0.75 mm x 0.75 mm. The detector size was large enough to avoid data truncation. Noiseless projection data were simulated.

2.4 Experimental Design

We exercised the proposed reconstruction pipeline on five different sets of acquisition geometries. For each geometry, 512 rotation angles, θ, are evenly distributed between 0° and 360°. The elevation angles, ϕ, are parameterized as sinusoidal functions of θ at varying frequencies. The amplitude has been set to 25° for all cases (except the circular geometry). Four networks were trained on data of only one orbit, while one network was trained on data with two different geometries in a common pool. This was done to investigate if our proposed approach is able to reconstruct data of more than one geometry.
The five acquisition geometries are as follows:

- circular, $\phi = 0$ for all θ
- $\phi = \sin(2\theta)$,
- $\phi = \sin(3\theta)$,
- $\phi = \sin(2\theta)$ and $\phi = \sin(3\theta)$,
- one linear combination of sinusoidal basis functions with randomly generated coefficients.

2.5 Evaluation Metrics

Reconstruction performance was evaluated in terms of the the normalized RMSE (nRMSE), the feature similarity index (FSIM) and the structural similarity index (SSIM) between the network output and the ground truth phantom images. To compare the proposed reconstruction pipeline with state-of-the-art algorithms, we additionally performed reconstructions using 50 iterations of the SART algorithm (using the GPU-based TIGRE toolkit for arbitrary trajectories [8, 9]).

3 Results

Intermediate images and final reconstruction outputs from the proposed reconstruction pipeline are illustrated in Figure 3. Note the residual artifacts in the deconvolved volumes. The calculated evaluation metrics are compared in Table 1. The CNN-based approach consistently outperforms the SART reconstructions in terms of nRMSE and FSIM. This is also the case for SSIM except for the network trained on two sinusoidal geometries. While SART performs comparably for all geometries with only slight deviations, the performance of the CNNs show noticeable differences for the different geometries. Specifically, reconstruction performance decreases with increasing orbital complexity (possibly due to increased shift-variance). This is apparent in the slightly decreasing evaluation metrics as well as in the magnified areas and the difference images in Figure 3. The magnified region in particular contains fine-grain details, which every CNN struggles to reconstruct accurately.
The majority of the computation time for the proposed method is spent on the calculation of the impulse response (5 minutes), the ray density (30 seconds), and ultimately for the deconvolution operation (20 seconds). The CNN prediction takes around 1 second. In comparison, SART reconstructions take approximately 50 minutes for 50 iterations on a workstation with comparable specifications. Aside from the CNN, the mentioned implementations have not been optimized for runtime.

4 Discussion and Conclusion

In this work, we proposed a novel pipeline for fast reconstruction of non-circular geometries. The pipeline consists of an initial deconvolution step to remove an approximation of the system response followed by an artifact removal step using a CNN. We tested the pipeline in five sets of imaging geometries of single and mixed sinusoidal orbits. Our proposed method offers $\sim 90\,\%$ reduction in computation time and is comparable or superior to SART in terms of the nRMSE, FSIM, and SSIM. These results suggest that the pipeline offers a promising approach to reconstruct data acquired with non-circular orbits when time is of the essence.
This work has several limitations that are being addressed in ongoing work. First, the pipeline was only trained and assessed on piecewise-constant phantoms. Extending the reservoir of phantoms to include non-piecewise-constant phantoms will help to improve the generalizability of the proposed approach. Second, the case where two sinusoidal orbits were trained simultaneously illustrates some capacity of the method to accommodate multiple geometries within the same class. We plan to extend the reconstruction capability of the method to arbitrary orbits within classes (e.g., sinusoids).

Acknowledgements

The authors acknowledge support, in part, by the US National Institutes of Health through grant R01EB027127; and by the state of Baden-Wuerttemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG. This research project is part of the Research Campus M^2OLIE and funded by the German Federal Ministry of Education and Research (BMBF) within the Framework 'Forschungscampus - Public-Private Partnership for Innovation' under the funding code 13GW0388A.

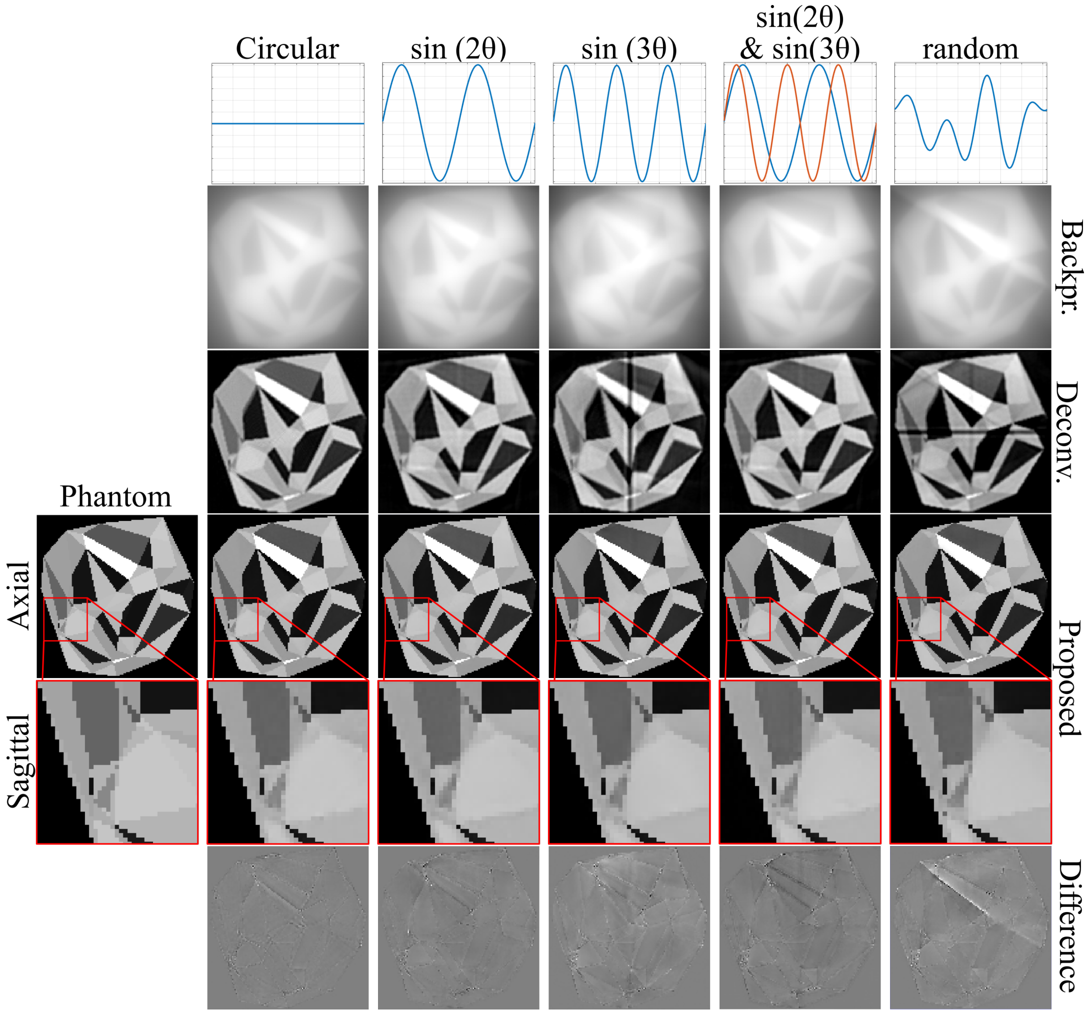

Figure 3: Intermediate image volumes and final reconstruction outputs from the reconstruction pipeline. Each column corresponds to a set of imaging geometries. Rows from top to bottom: elevation angle ϕ as a function of rotation angles θ; backprojected volume; volume after deconvolution; CNN reconstruction (axial slice); zoomed in ROI within the slice; difference image between the reconstructions and ground truth phantom images.

		circular	sin(2θ)	sin(3θ)	sin(2θ)&sin(3θ)	random
Proposed	nRMSE ↓	**0.033 ± 0.005**	**0.048 ± 0.007**	**0.060 ± 0.008**	**0.062 ± 0.007**	**0.061 ± 0.009**
	FSIM ↑	**0.991 ± 0.005**	**0.983 ± 0.010**	**0.979 ± 0.010**	**0.977 ± 0.013**	**0.979 ± 0.013**
	SSIM ↑	**0.994 ± 0.002**	**0.987 ± 0.004**	**0.984 ± 0.003**	0.944 ± 0.013	**0.985 ± 0.007**
SART	nRMSE ↓	0.116 ± 0.013	0.105 ± 0.016	0.109 ± 0.015	0.107 ± 0.016	0.108 ± 0.015
	FSIM ↑	0.937 ± 0.019	0.943 ± 0.015	0.940 ± 0.015	0.942 ± 0.015	0.941 ± 0.015
	SSIM ↑	0.941 ± 0.011	0.963 ± 0.011	0.956 ± 0.011	**0.960 ± 0.011**	0.958 ± 0.010

Table 1: Evaluation metrics for the propose pipeline compared with SART. All metrics were evaluated between the reconstructions and ground truth phantom images. The better of two values is marked **bold**.

References

[1] Tang, X and Ning, R, *Medical Physics*, vol. 28, no. 6, pp. 1042–1055, 2001.

[2] Gang, G. J *et al.*, *International Conference on Image Formation in X-Ray Computed Tomography*, 2020.

[3] Stayman, J. W *et al.*, *Journal of Medical Imaging*, vol. 6, no. 02, p. 1, 2019.

[4] Capostagno, S *et al.*, *Journal of Medical Imaging*, vol. 6, no. 02, p. 1, 2019.

[5] Pack, J. D *et al.*, *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 70–85, 2005.

[6] Palenstijn, W. J *et al.*, *Journal of Structural Biology*, vol. 176, no. 2, pp. 250–253, 2011.

[7] van Aarle, W *et al.*, *Optics Express*, vol. 24, no. 22, p. 25129, 2016.

[8] Biguri, A *et al.*, *Biomedical Physics and Engineering Express*, vol. 2, no. 5, 2016.

[9] Hatamikia, S *et al.*, *Medical Physics*, vol. 47, no. 10, pp. 4786–4799, 2020.