Deep Learning-Based Detector Row Upsampling for Clinical Spiral CT

Jan Magonov[a,b,c], Julien Erath[a,b,c], Joscha Maier[a,c], Eric Fournié[b], Karl Stierstorfer[b], and Marc Kachelrieß[a,c]

[a]German Cancer Research Center (DKFZ), Heidelberg, Germany
[b]Siemens Healthcare GmbH, Forchheim, Germany
[c]Heidelberg University, Heidelberg, Germany

ABSTRACT

Due to longitudinal undersampling multislice spiral computed tomography (MSCT) scans may suffer from windmill artifacts in reconstructed images. To fulfill the sampling condition and achieve double sampling in z-direction, some CT scanners use the z-flying focal spot (zFFS) technique, a hardware-based solution that effectively doubles the number of detector rows. To obtain a software-based solution we developed a convolutional neural network that is trained in a supervised manner with clinical projection raw data that were acquired with zFFS enabled. We presented this approach as the row interpolation with deep learning (RIDL) network. In this work we simplified the network architecture, extended the clinical dataset and generated an experimental synthetic dataset consisting of two-dimensional projection data. We were able to observe a reduction in windmill artifacts for both datasets used for training. Especially the synthetic dataset is very promising as we could observe a superior reduction of artifacts with this dataset.

Keywords: Spiral CT, z-flying focal spot, deep learning, windmill artifact reduction

1. INTRODUCTION

Multislice spiral computed tomography (MSCT), also known as multidetector CT, has become an integral part of modern medical imaging after the theoretical introduction of spiral CT in 1989.[1] The most common application of these systems is spiral scanning, in which the patient is continuously moved through the gantry, resulting in shorter scan times and higher temporal resolution.[2] Nevertheless, artifacts can occur with this modality that degrade quality of reconstructed images. The windmill artifact is an image distortion in the axial plane whose appearance is characterized by bright streak-like patterns emerging from high contrast structures along the longitudinal axis.[3] When scrolling through the reconstructed slices these streaks appear to rotate. The cause of this artifact can be attributed to inadequate data sampling in the z-plane resulting in not satisfying the Nyquist condition and thus leading to aliasing.[3,4]

A hardware-based method to fulfill the sampling condition and reduce windmill artifacts is provided by the z-flying focal spot (zFFS).[2,4] This technique doubles the effective number of detector rows acquired during the scan by periodically deflecting the X-ray focal spot in longitudinal direction. The resulting higher sampling rate in z-direction reduces the occurrence of windmill artifacts. Figure 1 shows a scan acquired without zFFS compared to a corresponding scan with zFFS enabled.

However, this method also has some drawbacks, as it is technically complex and thus prevents the use of zFFS in CT systems that do not meet these requirements. Previous works, such as in Ref. 5, focus on the reduction of windmill artifacts in the image domain. In contrast we try to solve the problem in projection domain. In Ref. 6 we presented the row interpolation with deep learning (RIDL) network, which was similar to the zFFS designed to double the effective number of acquired raw detector rows in projection domain. The network was based on the SRResNet presented in Ref. 7 to compute super-resolution images, i.e. very high-resolution images. In this paper we simplified the network architecture in order to reduce complexity of training process while

Further author information: (Send correspondence to J.M.)
J.M.: E-mail: jan.magonov(at)dkfz.de

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 1230407
© 2022 SPIE · 0277-786X · doi: 10.1117/12.2647101

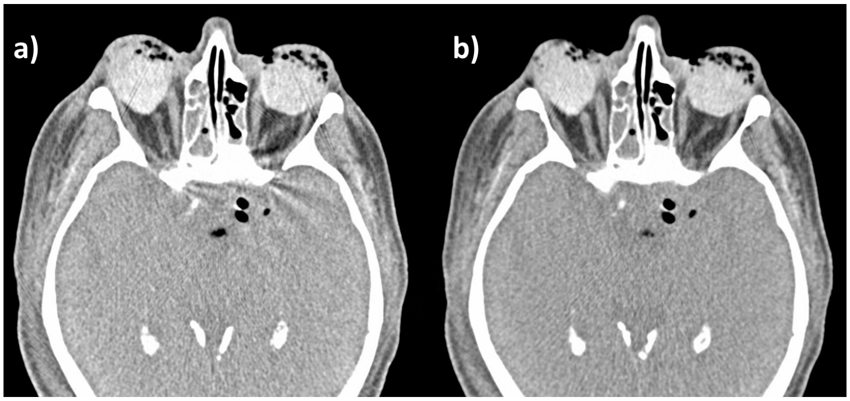

Figure 1. Reduction of windmill artifacts by using zFFS. The left image was taken without zFFS (32×0.6 mm collimation, pitch 1.4) while for the right reconstructed image ($2 \cdot 32 \times 0.6$ mm collimation, pitch 1.4) zFFS was enabled for acquisition ($C = 0$ HU, $W = 200$ HU).

maintaining existing results. Furthermore, the clinical dataset used for network training was extended and an experimental synthetic dataset was generated. Two separate networks were trained with the individual datasets and the network predictions were compared in image domain by reconstructing two clinical spiral CT scans.

2. METHODS AND MATERIAL

2.1 Clinical Data Preparation

For the clinical dataset, we selected raw projection data from a total of 40 clinical CT scans from different patients. The scans covered different body regions such as head, thorax and abdomen and were acquired with Somatom Flash and Somatom Force dual source CT scanners (Siemens Healthineers, Forchheim, Germany) with zFFS enabled. The dataset was split into two disjoint subsets so that 32 of the scans were used as training dataset and 8 scans served as validation dataset. It was ensured that the different body regions and CT systems used in the images were equally distributed in both datasets. Before training, some preprocessing steps were performed, i.e. instead of using the complete projection data of the scans as a whole, randomized image patches were generated to simplify the training process.

2.2 Synthetic Data Preparation

In addition to the clinical dataset, the acquisition of synthetic data for training the RIDL network was investigated. An advantage of using synthetic data would be that any number of data could be generated without requiring a CT scanner with zFFS. For the simulation we used the software package CT_SIM which is based on the deterministic ray propagation simulation software Deterministic Radiological Simulation (DRASIM). These tools allow to simulate the radiation properties in a defined X-ray imaging setup through geometrically defined phantoms.[8] In our first experimental setup, we generated two-dimensional projection data of a water cylinder (length: 10 cm, diameter: 40 cm, density: 1.0 g/cm^3) containing overlapping water spheres with varying densities (0.5 - 3.0 g/cm^3) and diameters (1 - 20 cm). Each water sphere was overlaid with another, smaller water sphere of density 1.0 g/cm^3, resulting in narrow circular edges with a width of 0.3 to 2 mm. These structures are particularly difficult to interpolate. Figure 2 shows an example representation of a projection from the synthetic dataset. A total of 200,000 projections with 80 detector rows and 800 channels containing randomly arranged water spheres were simulated. Comparable to real clinical projection data acquired with zFFS, the detector rows were simulated overlapped. No noise was added to the data. The dataset was split into a training dataset with 160,000 projections and a validation dataset with 40,000 projections. In addition, the range of values of the synthetic projection data was linearly scaled to the value range of the clinical data. Similar to the clinical dataset, random image patches were selected from the projection data for network training, as we will describe in more detail below.

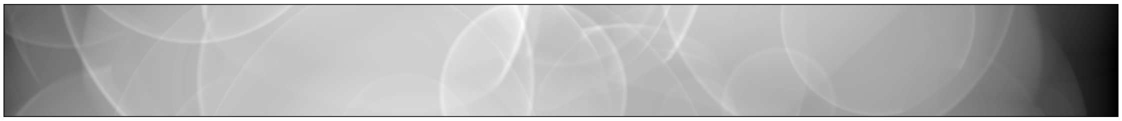

Figure 2. Example projection from the synthetic dataset with different sized water spheres consisting of 80 overlapped detector rows and 800 channels.

2.3 Row Interpolation with RIDL-CNN

In the previous approach of our work, a neural network was trained that received raw projection data and generated a prediction of the input with interpolated rows to effectively double the number of rows. The clinical projection data used to train the network were obtained after the rebinning, which is the rearrangement of the measured fan-beam data to parallel beam geometry. These projections were then divided into alternative rows so that projections containing all rows (acquired with zFFS) were used as the desired output y, and every other row from the corresponding projections was used for the network input x in training. In order to predict an upsampled version of the input data using the network, a so-called subpixel convolutional layer[9] was used, which essentially performs an upsampling of the generated feature maps within the network by a specific type of image reshaping. However, this procedure is time-consuming in network training, as well as in the subsequent use of the trained network for the prediction of rows.

In further experiments, we could observe that a much simpler convolutional neural network without subpixel convolution is able to produce results comparable to the RIDL-SRResNet. In the following, we will refer to this network architecture as RIDL-CNN. Similar to the previous architecture, random patches with the size of 64×32×1 pixels were generated from the underlying projection data to train the network. Also in this case, every other row from these patches serves as network input so that it has a size of 32×32×1. The desired output has the same dimension and is obtained from the intermediate rows in the generated patches. Before network training, all patches were linearly normalized to a value range in the interval from 0 to 1. Slope and offset were set according to the minimum and maximum value of the clinical dataset. After network prediction, the input and output rows have to be interlaced to obtain corresponding interpolated projections. Furthermore, the value range of these projections must be denormalized to the original range of the clinical data. In total, the RIDL-CNN consists of an input layer followed by 12 convolutional layers with 128 filters and 3×3 kernels. The network output is computed by a final convolution. The number of trainable parameters is 1,625,857.

2.4 Implementation and Training

The RIDL-CNN was trained on both the clinical and synthetic dataset, resulting in two separately trained networks. For both datasets, 500,000 examples were selected from the corresponding training dataset and 125,000 from the corresponding validation dataset. For the training we used the Adam optimizer and a combined loss function that is described by:

$$L_{\text{comb}}(y, \hat{y}) = \alpha \cdot L_{\text{MS-SSIM}}(y, \hat{y}) + (1 - \alpha) \cdot L_{\text{MAE}}(y, \hat{y})$$

This loss function was proposed in Ref. 10 and takes into account the pixel-wise computed error between the network output $\hat{y}$ and ground truth y by the mean absolute error (MAE) but also the structural similarity between the two images by the multi-scale structural similarity index (MS-SSIM). The weighting factor was empirically determined as $\alpha = 0.84$. The networks were trained with a batch size of 256 and an initial learning rate set to 1×10^{-5}, which was halved if the error could not be minimized for 25 consecutive epochs.

2.5 Evaluation and Validation

In order to evaluate and validate the results, two scans of a skull phantom with real human bones were acquired with a Somatom Force system. For the first scan, a basic scan mode with a collimation of 96×0.6 mm and activated zFFS was used. For the second scan we used a scan mode of the CT system with a collimation of 48×1.2 mm. In this mode, no zFFS can be enabled and the acquired images show very strong windmill artifacts due to the lower sampling in the z-direction. In both scans the pitch factor was set to 1, since especially scans with pitch values in this range suffer from windmill artifacts.[2,3] A summary for both scan settings can be found in Table 1.

Table 1. Summary of the settings for the two scans used to evaluate and validate the results.

Scan	Collimation	Pitch	zFFS	Reconstructed slices
1	96×0.6 mm	1.0	yes	1.0 mm
2	48×1.2 mm	1.0	no	1.5 mm

As in our previous work, the trained networks were applied to reconstruct these phantom CT scans. For this purpose, a plugin we developed for the Siemens-specific reconstruction software was used to adjust the raw projection data after the rebinning. In the first scan, every second row, i.e. the zFFS-generated rows were replaced by rows predicted by the RIDL networks. In addition, a reconstruction with linear interpolated rows was performed, which should correspond to an acquisition without zFFS. For all reconstructions, the error measures RMSE and SSIM were calculated in relation to the ground truth reconstruction with zFFS enabled. Since no zFFS can be used in the acquisition setting employed in the second scan, there is no ground truth data. The results in these reconstructions can therefore only be evaluated qualitatively. In this case, the number of rows in the raw data was doubled by extending them with predictions from the RIDL networks.

3. METHODS AND MATERIAL

3.1 Scan with 96×0.6 mm Collimation

Figure 3 shows reconstructions of a specific slice with differently modeled projection raw data from the first scan. Difference images are calculated to the ground truth data acquired with zFFS. Comparing the reconstruction without zFFS to the result of the RIDL-CNN trained on the clinical dataset, only a very slight reduction of the windmill artifacts can be seen in the image domain. Furthermore, there are noisy structures noticeable in the difference image in Figure 3c in the area of the bones. However, MSE and SSIM indicate a quantitatively slightly better result compared to omitting the zFFS. Looking at the reconstruction with the network trained with synthetic data (see Figure 3d), we find an improvement in the image quality both in the image domain and in the difference image. Especially the problem with the noisy structures in the bone areas does not occur. With regard to the error measures, this reconstruction also provides the best result quantitatively.

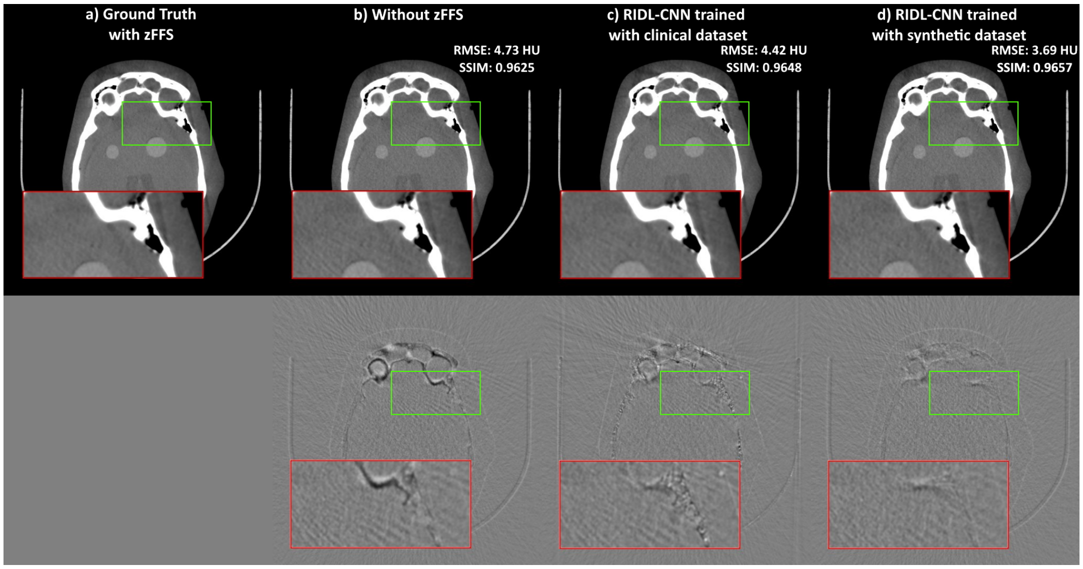
a) Ground Truth with zFFS
b) Without zFFS
RMSE: 4.73 HU
SSIM: 0.9625
c) RIDL-CNN trained with clinical dataset
RMSE: 4.42 HU
SSIM: 0.9648
d) RIDL-CNN trained with synthetic dataset
RMSE: 3.69 HU
SSIM: 0.9657

Figure 3. Qualitative and quantitative comparison of a reconstructed slice (1. scan) without zFFS, with the RIDL-CNN trained with the clinical dataset and the RIDL-CNN trained with the synthetic dataset compared to the ground truth scan with zFFS (C = 60 HU, W = 360 HU). Below the difference images to the ground truth are shown (C = 0 HU, W = 150 HU).

3.2 Scan with 48×1.2 mm Collimation

Figure 4 compares the results for two reconstructed slices from the second scan. In both slices reconstructed with WFBP without zFFS, very dominant windmill artifacts can be observed. Comparing these slices with the results of the network trained with clinical data, a slight reduction of the artifacts can be seen qualitatively. The results obtained with the network trained with the synthetic dataset can most effectively reduce the occurring windmill artifacts and provide superior image quality compared to the network trained with clinical data. The comparison of the reconstructions can only be performed qualitatively due to missing ground truth data, since the applied scan mode does not allow for enabling zFFS.

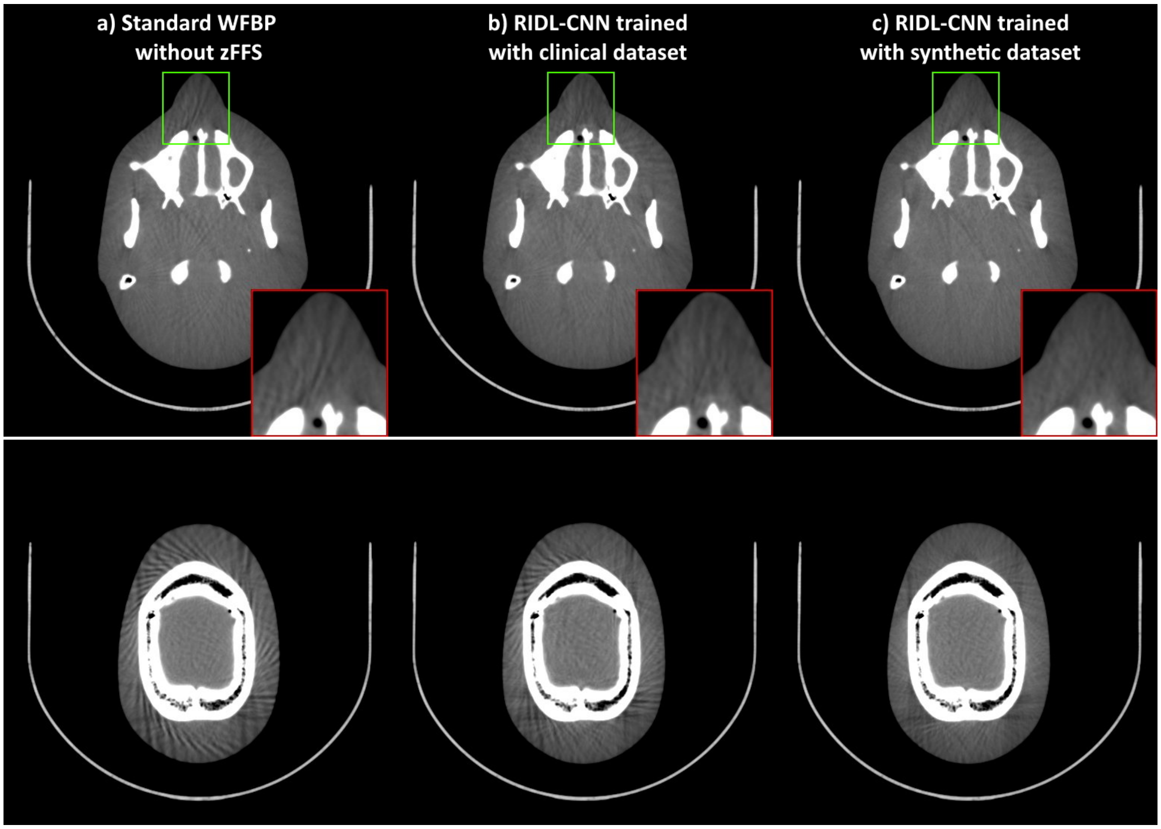
a) Standard WFBP without zFFS
b) RIDL-CNN trained with clinical dataset
c) RIDL-CNN trained with synthetic dataset

Figure 4. Qualitative comparison of two selected slices (2. scan) without zFFS, with the RIDL-CNN trained with the clinical dataset and the RIDL-CNN trained with the synthetic dataset ($C = 60$ HU, $W = 360$ HU).

4. DISCUSSION AND CONCLUSION

In this work, we further adapted our RIDL network and simplified the network architecture. In addition, we extended the clinical dataset and generated an experimental synthetic dataset. This was done by simulating two-dimensional raw data containing different sized overlapping spherical structures. In our experiments presented here, we observed that the results with the synthetic data are very promising. Although no clinical data were included in this dataset, windmill artifacts were reduced more effectively than with the RIDL-CNN trained with the current setup of clinical data. This observation suggests that training with clinical data can still be optimized. One problem could be the noise in clinical projection data. Denoising the clinical data before network training could be considered. However, it is valuable that training with synthetic data can address the problem of windmill artifacts, without having to rely on raw clinical projection data acquired with a CT system that supports zFFS. The next step is to investigate how the synthetic dataset can be adapted more efficiently to our task. In addition, data with a concrete CT system geometry will be simulated. Furthermore, we will optimize the training with clinical data and investigate whether the results can be improved by a combination of synthetic and clinical data.

REFERENCES

[1] Kalender, W. A., Seissler, W., Klotz, E., and Vock, P., "Spiral volumetric CT with single–breath–hold technique, continuous transport, and continuous scanner rotation," *Radiology* **176**, 181–183 (July 1990).

[2] Flohr, T., Stierstorfer, K., Raupach, R., Ulzheimer, S., and Bruder, H., "Performance evaluation of a 64-slice CT system with z-flying focal spot," *RöFo: Fortschritte auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin* **176**, 1803–10 (2005).

[3] Silver, M. D., Taguchi, K., Hein, I. A., Chiang, B., Kazama, M., and Mori, I., "Windmill artifact in multislice helical CT," *SPIE Medical Imaging Proc.* **5032**, 1918 – 1927 (2003).

[4] Kachelrieß, M., Knaup, M., Penßel, C., and Kalender, W. A., "Flying focal spot (FFS) in cone–beam CT," *IEEE Transactions on Nuclear Science* **53**, 1238–1247 (June 2006).

[5] Brown, K. M. and Žabic, S., "Method for reducing windmill artifacts in multislice CT images," *SPIE Medical Imaging Proc.* **7961**, 491 – 495 (2011).

[6] Magonov, J., Kachelrieß, M., Fournié, E., Stierstorfer, K., Buzug, T., and Stille, M., "Row interpolation in spiral CT with deep learning," *16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine* , 376–380 (Oct. 2021).

[7] 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," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , 105–114 (July 2017).

[8] Kappler, S., Niederlohner, D., Wirth, S., and Stierstorfer, K., "A full-system simulation chain for computed tomography scanners," *IEEE Nuclear Science Symp. Conf. Record* , 3433 – 3436 (Dec. 2009).

[9] 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," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2016).

[10] Zhao, H., Gallo, O., Frosio, I., and Kautz, J., "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging* **3**, 47–57 (Dec. 2016).