Published in final edited form as:

Med Phys. 2021 October ; 48(10): 5712–5726. doi:10.1002/mp.15176.

Deep learning enabled ultra-fast-pitch acquisition in clinical X-ray computed tomography

Hao Gong, Liqiang Ren, Scott S. Hsieh, Cynthia H. McCollough, Lifeng Yu

Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

Abstract

Objective: In X-ray computed tomography (CT), many important clinical applications may benefit from a fast acquisition speed. The helical scan is the most widely used acquisition mode in clinical CT, where a fast helical pitch can improve the acquisition speed. However, on a typical single-source helical CT (SSCT) system, the helical pitch p typically cannot exceed 1.5; otherwise, reconstruction artifacts will result from data insufficiency. The purpose of this work is to develop a deep convolutional neural network (CNN) to correct for artifacts caused by an ultra-fast pitch, which can enable faster acquisition speed than what is currently achievable.

Methods: A customized CNN (denoted as ultra-fast-pitch network (UFP-net)) was developed to restore the underlying anatomical structure from the artifact-corrupted post-reconstruction data acquired from SSCT with ultra-fast pitch (i.e., $p \geq 2$). UFP-net employed residual learning to capture the features of image artifacts. UFP-net further deployed in-house-customized functional blocks with spatial-domain local operators and frequency-domain non-local operators, to explore multi-scale feature representation. Images of contrast-enhanced patient exams ($n = 83$) with routine pitch setting (i.e., $p < 1$) were retrospectively collected, which were used as training and testing datasets. This patient cohort involved CT exams over different scan ranges of anatomy (chest, abdomen, and pelvis) and CT systems (Siemens Definition, Definition Flash, Definition AS+, Siemens Healthcare, Inc.), and the corresponding base CT scanning protocols used consistent settings of major scan parameters (e.g., collimation and pitch). Forward projection of the original images was calculated to synthesize helical CT scans with one regular pitch setting ($p = 1$) and two ultra-fast-pitch setting ($p = 2$ and 3). All patient images were reconstructed using the standard filtered-back-projection (FBP) algorithm. A customized multi-stage training scheme was developed to incrementally optimize the parameters of UFP-net, using ultra-fast-pitch images as network inputs and regular pitch images as labels. Visual inspection was conducted to evaluate image quality. Structural similarity index (SSIM) and relative root-mean-square error (rRMSE) were used as quantitative quality metrics.

Results: The UFP-net dramatically improved image quality over standard FBP at both ultra-fast-pitch settings. At $p = 2$, UFP-net yielded higher mean SSIM (> 0.98) with lower mean rRMSE ($< 2.9\%$), compared to FBP (mean SSIM < 0.93 ; mean rRMSE $> 9.1\%$). Quantitative metrics at $p =$

Correspondence: Lifeng Yu, Professor of Medical Physics, Department of Radiology, Mayo Clinic, 200, 1st ST SW, Rochester, MN, USA. Yu.Lifeng@mayo.edu.

CONFLICT OF INTEREST

Dr. McCollough receives industry support from Siemens Healthcare, which is unrelated to this work. No other authors have anything to disclose.

3: UFP-net—mean SSIM [0.86, 0.94] and mean rRMSE [5.0%, 8.2%]; FBP—mean SSIM [0.36, 0.61] and mean rRMSE [36.0%, 58.6%].

Conclusion: The proposed UFP-net has the potential to enable ultra-fast data acquisition in clinical CT without sacrificing image quality. This method has demonstrated reasonable generalizability over different body parts when the corresponding CT exams involved consistent base scan parameters.

Keywords

convolutional neural network (CNN); deep learning; helical CT; helical pitch; image reconstruction

1 | INTRODUCTION

X-ray computed tomography (CT) is one of the leading medical imaging modalities used in routine clinical practice, wherein X-ray projections at successive view angles are acquired to reconstruct a stack of transverse images of internal anatomy. The helical scan is the most commonly used data acquisition mode in clinical CT: the X-ray tube and detector rotate around the patient at a high speed (up to four revolutions per second) as the patient table continuously translates at a constant velocity, which forms a helical trajectory of the X-ray source from the patient frame of reference (Figure 1a). The helical pitch determines the available amount of X-ray projection data for reconstructing a transverse image. The projection data sufficiency condition needs to be met and appropriate weighting factors need to be applied so that standard clinical CT image reconstruction algorithms such as filtered-back-projection (FBP) can provide satisfactory image quality^{1,2} (Figure 1b,c). To reduce reconstruction artifacts, the helical pitch $p < 1.5$ is commonly used in routine practice to make sure accurate reconstruction can be achieved in a desired field-of-view (FOV). Nevertheless, a faster helical pitch is desirable in many clinical CT tasks involving patient and organ motion, for example, pulmonary CT, cardiovascular CT, and pediatric CT. Fast-pitch acquisition (pitch up to 3.4) is currently available on clinical dual-source CT (DSCT) scanners since DSCT uses two X-ray source-detector systems to synergistically fill in the projection data gap for accurate image reconstruction. The corresponding clinical merits have been demonstrated in free-breathing pulmonary CT, coronary CT angiography, and the imaging of traumatic/un-cooperative/pediatric patients.^{3–7} For example, the whole lung can be scanned within 1 s using the ultra-fast-pitch mode, which can be performed without the need of breath holding. For many patients who are unable to cooperate with the breath-holding instruction, such as young pediatric patients, this is an extremely useful acquisition mode. However, such ultra-fast-pitch acquisition cannot be used on single-source helical CT (SSCT) scanners since the single-source system will have non-negligible gaps in the acquired projection data when the helical pitch is greater than ~ 1.5 . Thus far, only one CT vendor (Siemens Healthcare) provides DSCT, while other vendors provide only SSCT. Therefore, most CT scanners available in clinical practice today are SSCT. In this work, we are motivated to develop novel methods to enable ultra-fast-pitch acquisition mode on SSCT such that many more patients can benefit from faster acquisition speed than what can be achieved on current scanners.

Emerging studies involving deep learning-based algorithms have suggested the potential of using deep learning technology to further improve CT image quality under different data conditions. For instance, some image-domain methods have shown promising results in CT image noise reduction^{8–11} and sparse-view artifact reduction.^{12,13} Several projection-domain methods demonstrated the potential of becoming generic reconstruction paradigms in X-ray CT.^{14–16} More pertinently, multiple papers have presented the deep learning-based algorithms dedicated for limited-angle CT reconstruction. These algorithms can also be roughly categorized into image-domain, projection-domain, and hybrid methods. Several examples are briefly summarized as follows. Image-domain methods: Zhang et al. applied standard two-layer convolutional neural networks (CNNs) to separately reduce artifacts at in parallel-beam CT with three angular coverage (130°, 150°, and 170°)¹⁷; Gu and Ye used U-net to suppress artifacts in wavelet transform of fan-beam axial CT images with 120° or 150° coverage.¹⁸ Projection-domain methods: Hammernik et al. used a fan-beam geometry-based neural network for initial limited-angle reconstruction and then applied a variational neural network to reduce remaining limited-angle artifacts¹⁹; the same group later presented a neural network framework dedicated for cone-beam circular CT geometry and evaluated this framework on a half-scan setting (i.e., 180° coverage).²⁰ Hybrid methods: Bubba et al. fused model-based iterative sparse regularization and U-net over shearlet transform of fan-beam axial CT images (with 120° or 150° coverage), solving visible and invisible components of shearlet coefficients, respectively²¹; Huang et al. presented deep learning prior-based limited-angle reconstruction in cone-beam circular CT geometry (120° coverage), using U-net-derived initial reconstruction as prior information in a conventional iterative reconstruction (IR) framework (i.e., simultaneous algebraic reconstruction technique with total variation minimization)²²; Zhou et al. proposed an unsupervised Generative Adversarial Networks (GAN)-based framework to jointly optimize sinogram adaption and image reconstruction and evaluate it in parallel-beam CT geometry (with 120° coverage).²³ However, these methods were not specifically designed for the missing data problem presented in helical CT reconstruction with an ultra-fast pitch. It is unclear if they can be readily applied since the dimensionality of the reconstruction problem is largely increased in helical geometry, compared to two-dimensional (2D) parallel-/fan-beam geometry or cone-beam circular geometry. In older scanners with fewer slices, the three-dimensional (3D) reconstruction problem was often transformed to 2D by rebinning 3D helical projections to axial 2D fan-/parallel-beam projections. One could apply these aforementioned limited-angle reconstruction algorithms on these rebinned 2D datasets, but there are several complications. First, due to helical geometry, the continuous angular shift is imposed on the angular coverage of scanning per slice on the helical path and then results in image artifacts with varying directional characteristics (Figure 1c). Notice that the aforementioned limited-angle algorithms were mostly established and evaluated with CT data associated with stationary angular coverage of scanning. Second, rebinning to axial 2D slices yields artifacts with the 32- or 64-slice CT scanners commonly used today. Newer analytical reconstruction algorithms filter the data along oblique lines or curves and backproject in three dimensions to reduce artifacts.^{24,25} Third, the amount of missing data in the ultra-fast-pitch problem is dependent on distance from the isocenter. In contrast to the 2D problem where a fixed angular range is absent, in ultra-fast pitch, there is less missing data at the isocenter, compared to a location far from the isocenter. Finally, the tremendous

memory requirement could induce challenges to the practical deployment of those deep learning methods that involved direct projection-to-image mapping. For instance, Liang et al. had to reconstruct image data in a coarse grid due to the limited memory of their graphics processing unit (GPU) server.²⁶ In addition, the limited access to patient raw projection data and vendors' proprietary tools may also hinder the practical use of custom algorithms that involve projection data.

In this work, we proposed an ultra-fast-pitch CNN (denoted as ultra-fast-pitch network (UFP-net)) to reconstruct underlying true anatomical structure from artifact-corrupted FBP images acquired from ultra-fast-pitch SSCT. We chose to establish an approach upon post-reconstruction data while deploying deep learning over the image domain and frequency domain in a hybrid manner. In-house-designed functional blocks were developed to account for the directional characteristics of helical artifacts caused by ultra-fast-pitch acquisition. The parameters of UFP-net were optimized with a customized loss function, which consists of Mean square error (MSE), Image-gradient correlation, and Feature-reconstruction loss (denoted as MIF loss), using a customized multi-stage training strategy (see Methods section for further details). The proposed method was benchmarked by a standard FBP algorithm. Visual inspection and quantitative quality metrics were both employed in the evaluation. Further detail of our methods and experimental results is presented in the following sections.

2 | METHODS

2.1 | Architecture of UFP-net

The artifact-corrupted fast-pitch images were used as the inputs to UFP-net (Figure 2), while the corresponding regular-pitch CT images were used as labels in CNN training. A skip connection was added between the inputs and outputs of UFP-net, to enable residual learning, that is, UFP-net mainly learned the feature representation of helical artifact. The architecture of the UFP-net may appear to be similar to the typical U-net like CNNs. Nonetheless, in-house-designed functional blocks were used to form the backbone of UFP-net, that is, two versions of inception-discrete-cosine-transformation (Incept-DCT) blocks (denoted as Incept-DCT v1/v2 blocks). Incept-DCT blocks provided a hybrid of local operators and non-local operators to model the local and non-local components of the helical artifact. Local operators involve standard convolutional operations over the image and spatial latent space, that is, these operators were implemented as parallel convolutional layers with different-sized kernels at different depths of UFP-net to enable multi-scale representation of local image structure. Non-local operators involve convolutional operations over the frequency representation of spatial latent space: These operators were implemented as convolutional layers stacked between paired 2D DCT and inverse 2D DCT, that is, the convolutional operation across the DCT domain to ensure the representation of non-local image structure. DCT was selected to provide frequency representation, considering several practical benefits of DCT, for example, compact representation and fast computation. With the proposed non-local operators, UFP-net was able to extend the effective receptive field to the entire FOV of the input images, and thus explore the global context information for modeling the spatially variant helical artifact. Of note, as suggested by Figure 1b, the generic orientation of artifacts is periodically varying along the helical path due to the

continuous shift of angular coverage of scanning. Thus, we believe that non-local operators could effectively capture the major directional characteristics of helical artifacts, while local operators providing the complementary correction to local artifacts with the data-driven approach. Zero padding was used in all convolutional layers. In addition, a drop-out layer was added to improve the CNN generalizability, using a constant drop-out rate of 0.1. The total number of parameters in UFP-net was roughly 3 000 000.

2.2 | Loss function of UFP-net

The parameters of UFP-net were optimized by minimizing an in-house-designed loss function (denoted as MIF loss, i.e., Equation 1). In Equation (1), the first, second, and third terms refer to MSE, image-gradient-correlation (IGC) regularizer,^{28–30} customized feature reconstruction (FR) loss (modified from reference³¹), respectively. Specifically, the MSE between CNN outputs and labels was used as the fidelity term, and the IGC and the FR loss were used as the regularization terms.

$$L = \frac{1}{M} \sum_{m=1,2,\dots,M} \left(\|f_{CNN,m} - f_{GT,m}\|_2^2 + \frac{1}{\rho(\nabla f_{CNN,m}, \nabla f_{GT,m}) + \epsilon} \right. \\ \left. + \frac{\lambda}{N} \sum_{n=1,2,\dots,N} \|\phi_n(f_{CNN,m}) - \phi_n(f_{GT,m})\|_2^2 \right) \quad (1)$$

$$\nabla f_{i,j,m} = |f_{i+1,j,m} - f_{i,j,m}| + |f_{i,j+1,m} - f_{i,j,m}|, \quad (2)$$

where $f_{CNN,m}$ and $f_{GT,m}$ denote the m^{th} paired network output and label per mini-batch of M samples, $\nabla f_{i,j,m}$ denotes the image gradient of the voxel located at i^{th} row, j^{th} column of the m^{th} sample, $\rho(\cdot)$ denotes the Pearson correlation, ϵ is a small constant fixed at $1.0e-4$, $\phi_n(\cdot)$ denotes the feature maps extracted from the n^{th} layer out of N pre-selected layers from a pre-trained VGG-19 network, and λ is a relaxation parameter. In this work, the first, fourth, seventh, 12th, and 13th layers of VGG-19 were used for feature extraction, and the value of λ was empirically fixed at 0.01. The two regularization terms provided complementary regularization. The IGC loss implemented edge preservation in the image voxel level, while the FR loss enforced the similarity of higher-level image features between network outputs and the corresponding labels.

2.3 | Data preparation

A patient cohort from a prior study³² was re-used in data preparation, which involved 83 retrospectively collected contrast-enhanced CT exams. Half of the cases involved clinically proven liver metastases, a quarter of the cases had proven benign lesions, and the remaining cases had normal livers. CT exams were performed on 128-slice CT systems from the same vendor (Siemens Definition, Definition Flash, or Definition AS+, Siemens Healthcare). For each patient, original CT projection data was acquired using the appropriate clinical protocol determined by the attending radiologist on the day of examination. The major parameters of the original scanning protocols are summarized in Table 1. The retrospective use of these cases was approved by Mayo Clinic Institutional Review Board. Mayo Clinic Institutional Review Board waived informed consent; however, patient authorization for use of images

was required. All methods were performed in accordance with the relevant guidelines and regulations from Mayo Clinic Institutional Review Board. The raw projection data of these exams were acquired from a clinical DSCT system (SOMATOM Flash, Siemens Healthcare, Inc.) with low helical pitch ($p < 1.0$), using scanning/reconstruction protocols at our institution. These cases were randomly split into training subset ($n = 71$), validation subset ($n = 4$), and testing subset ($n = 8$). We simulated virtual SSCT exams with different helical pitches, to generate paired samples for each subset. The simulation was used for two reasons: no commercial SSCT systems could exert helical pitch greater than 1.75; we retrospectively selected CT exams from our clinical data registry, and the helical pitch less than 1.5 is commonly used in clinical routine practice. Briefly, the raw projection data of virtual SSCT was numerically synthesized by calculating the forward projection, using the original patient images as the digital subjects. For each case, a ray-driven method³³ was used to calculate forward projection with the helical pitch across 1.0, 2.0, and 3.0. FBP algorithm with a medium sharpness filter was employed to reconstruct virtual SSCT images at a different helical pitch.^{2,34,35} The specific simulation setup is listed in Table 2. Finally, the full FOV fast-pitch SSCT images (i.e., $p = 2.0$ or 3.0) served as the inputs to UFP-net, while the regular-pitch SSCT images (i.e., $p = 1.0$) were used as the corresponding labels.

2.4 | CNN training and evaluation

The UFP-net was implemented with *Keras* 2.24 and *Tensorflow* 1.14, and the training and inference were carried out using a single GPU (TITAN V, NVIDIA, Inc.). The CNNs were trained separately for the dataset with a helical pitch of 2 and 3. A customized coarse-to-fine training strategy was proposed to incrementally optimize CNN parameters, which used four training stages to gradually transit the focus of the learning task from generic image structure toward finer details. Briefly, the training images were down-sampled to be 64×64 , 128×128 , and 256×256 in the first three stages, whereas the standard-sized images (i.e., 512×512) were used in the last stage. Meanwhile, the size of the mini-batch was decreasing across the same four stages, that is, 1024, 256, 128, and 32, to achieve a reasonable convergence speed while suppressing the risk of converging at sharp local minima. Adam (Adaptive moment estimation) optimizer³⁶ was used to minimize the training loss. The initial learning rate was fixed at 0.001, and the norm of gradient clipped to be 1.0. The number of training epochs was 40, 20, and 20 for the first three stages. Only five epochs were used in the last stage to avoid overfitting. The validation subset was only used to monitor the occurrence of overfitting, and no additional hyper-parameter optimization was used. In this work, we only used a single GPU for training, and the training time per pitch setting was ~38 h. Once the model is trained for a given scanning/reconstruction condition, the deployment for high-pitch image reconstruction is very efficient (e.g., seconds to minutes per case).

The proposed method was evaluated separately over the data with a helical pitch of 2 and 3. Of note, the dataset with a helical pitch of 3 was used for the “pressure test,” that is, to empirically explore the performance limit of the proposed method under severe data insufficiency. Visual inspection was carried out for qualitative evaluation of image quality. Structural similarity index (SSIM) and relative root-mean-square error (rRMSE)¹⁴

were calculated over the entire FOV of each testing image to evaluate the quantitative accuracy.

3 | RESULTS

3.1 | Loss curves of CNN

No obvious overfitting was observed in training (Figure 3). The training loss was a bit higher than validation loss throughout the first three training stages (i.e., the 1st to 79th epoch), whereas the situation was reversed in the final stage. This phenomenon could be possibly attributed to the following two reasons. First, the validation set ($n = 4$) was much smaller than the training set ($n = 71$), and thus training set may have more difficult cases than the validation set. Then, the drop-out regularization was only turned on during training but not validation or testing, which can induce additional “jittering noise” to the CNN model during training. After training, the Incept-DCT block was able to represent local and non-local anatomical structure/artifacts (Figure 4).

3.2 | Image quality comparison at the pitch of 2

The proposed method (i.e., UFP-net with MIF loss) was benchmarked against the standard FBP algorithm. With a helical pitch of 2, FBP yielded prominent image artifacts (e.g., the shading across lung parenchyma and distortion of blood vessels/bonny structure in Figure 5). The comparison of line profiles confirmed this observation (Figure 6a). The corresponding mean global SSIM per case ranged from 0.80 to 0.93, and the mean global rRMSE per case ranged from 9.1% to 15.9% (Figure 6b). The proposed method largely suppressed spatially varying artifacts, improved delineation of finer anatomical structure, and achieved higher quantitative accuracy (mean global SSIM per case range between [0.98 and 0.99]; mean global rRMSE per case between [1.4% and 2.9%]).

3.3 | Pressure test at the pitch of 3

In the pressure test, we observed similar findings as described in Section 3.2. FBP demonstrated significant anatomical distortion and spatially varying CT number bias (Figures 7 and 8): mean global SSIM [0.36, 0.61] and mean global rRMSE [36.0%, 58.6%]. The proposed method still effectively improved the fidelity of major image structure and CT number: mean global SSIM [0.86, 0.94], and mean global rRMSE [5.0%, 8.2%].

3.4 | Additional evaluation on pathologic cases and coronal slices

Several lesion examples are presented in Figure 9. UFP-net obviously reduced the helical artifact across lesions, restored lesion-to-background contrast, and improved lesion delineation. Randomly selected examples of coronal slices from the testing cases are presented in Figure 10. For both settings of helical pitch, UFP-net obviously suppressed helical artifact across the different depths of the image volume. At $p = 2$, the patient anatomy in UFP-net outputs had been very similar to the reference one. At $p = 3$, the residual artifact can be observed, yet a lot of anatomical detail was still restored in both lung and abdomen. It is important to note that: we used $p = 3$ as the “pressure test” in which we aim to explore the performance limits of the presented method under extreme projection data insufficiency.

4 | DISCUSSION

In this study, we proposed a deep learning technique for artifact correction on SSCT with ultra-fast pitch settings (i.e., $p = 2$). Briefly, an in-house ultra-fast-pitch neural network was developed to reconstruct anatomical structure from the artifact-corrupted FBP images. The customized loss function and multi-stage training strategy were developed to optimize CNN parameters. The proposed method largely suppressed artifacts and improved quantitative accuracy, compared to FBP. The proposed method has the potential to enable high-quality ultra-fast-pitch data acquisition on the widely used SSCT platform without major hardware changes. This is especially beneficial for many patients who cannot follow breathing instructions or cannot keep stationary during the CT scan.

By the nature of the standard FBP algorithm, insufficient projection data acquisition yielded severe image reconstruction artifacts in ultra-fast-pitch CT images. For a given FBP image, the appearance of artifacts is mainly affected by two factors: the underlying human anatomy and the angular range of projection data used to reconstruct the same image. The latter factor is functionally determined by X-ray beam helical trajectory (Figure 1b). The Incept-DCT blocks jointly explored global and multi-scale local context at different depths of UFP-net, which could respectively approximate the non-local and local components of the anatomy-/location-dependent artifacts. This modeling mechanism also enabled UFP-net to use fewer trainable parameters while exploring large image context, compared to many standard CNN models (e.g., U-net) that used fixed-sized local operators and max pooling at different depths to explore different-scaled context information. The width and depth of those CNNs could be increased to explore larger image context, and the network performance may be further improved thereafter. However, this strategy would increase the number of trainable parameters, which could possibly raise the request for more training data and/or additional regularization strategies to suppress the occurrence of overfitting. Further, note that the presented UFP-net can be readily extended to be 2.5D or 3D CNN. Yet we did not carry out such extension due to the following two major reasons: First, training the 2.5D or 3D versions of UFP-net would require large GPU memory and much more powerful computational hardware; second, with 2.5D or 3D convolution, UFP-net would likely yield much more trainable parameters and then more training data would be needed to reduce the risk of overfitting. Therefore, we only implemented the 2D UFP-net in this preliminary study. In addition, the presented UFP-net employed residual learning, which made the neural network largely focused on the feature representation of the helical artifact. This is because the helical artifact dominated the difference between the network inputs and the corresponding labels used in training. Compared to patient anatomy, helical artifact tends to have a much simpler image structure and is dependent on the given imaging geometry and helical pitch. Thus, the use of residual learning reduced the difficulty of the learning task of UFP-net and also likely improved the network generalizability over different patients. Nevertheless, we consider it is extremely challenging to fully restore all image content when the un-corrected images involve extreme projection data insufficiency. It is theoretically possible that UFP-net could create or eliminate anatomy that is visually similar to the helical artifacts that it is seeking to correct, although we did not observe this effect in our test set. Note that the condition of $p = 3$ was used as a “pressure test” in which

we mainly aimed to heuristically explore the limits of the presented method. Further, the presented method of UFP-net training and deployment can be readily generalizable over different body parts and even CT systems, if consistent major scan parameters (e.g., target collimation and helical pitch) are used in the corresponding base CT scanning protocols (e.g., as illustrated in Table 1). Further, if the network is applied to CT image data associated with drastically different base protocols, we expect model performance would degrade (e.g., stronger residual artifacts). In such scenario, network re-training or fine-tuning would be needed to achieve optimal image quality. Of note, there is the potential that a universal CT scanning protocol can be generated for examinations that require (ultra) high-pitch scan, which would help maintain the generalizability of the proposed method and also reduce the need of network re-training/fine tuning. With proper computational acceleration technique, the additional image reconstruction time for preparing training data and performing clinical scans could be largely reduced. For instance, on-board reconstruction software can be used for initial CT image reconstruction (for training and/or deployment), which typically yield the minimal reconstruction time (e.g., seconds to minutes per case). Further, the deployment of the trained neural network model is also computationally efficient, even when only single GPU is used for computational acceleration (e.g., seconds to minutes per case).

In practice, it is challenging to repeatedly scan real patients to establish training data with the paired low-/high-pitch images. Nevertheless, as we have already demonstrated, the simulation tool can be readily used to generate a large amount of training data, using the existing low-pitch images from the clinical registry. More realistic simulation tools could be used to involve multiple physical aspects (e.g., finite focal spot, detector collimation, etc.) of the given CT systems. If vendors' proprietary tool is accessible, the simulated projection data can be loaded to on-board or off-line CT reconstruction software to involve the effects of vendors' proprietary post-processing. Further, an alternative approach is to repeatedly scan CT phantoms and cadavers, to experimentally generate the paired low-/high-pitch images for training the neural network model.

We acknowledge several limitations in the present study. First, the raw projection data of ultra-fast-pitch CT scans was numerically simulated using a forward projection method, and thus the real ultra-fast-pitch CT data acquisition process was absent in the evaluation. Of note, no commercial SSCT systems could exert the helical pitch greater than 1.75. Second, our method was not compared with commercial FBP or IR algorithms, which involve vendors' proprietary pre-/post-processing techniques. Such comparison would require access to vendors' proprietary reconstruction pipeline, which is not readily accessible. Third, the two objective quality metrics, that is, SSIM and rRMSE, are not the complete descriptors of diagnostic image quality. In particular, SSIM is inadequate for comparing images with a high dynamic range,³⁷ for example, 16-bit CT images. Mathematical model observer studies could provide reliable and objective quality assessment in clinically relevant CT tasks.^{38,39} Human experts (e.g., radiologists and medical physicists) would also be needed for diagnostic image quality assessment within the context of specific clinical CT tasks (e.g., pulmonary embolism/liver metastases detection), which will be the topics of our follow-up studies. Despite the aforementioned limitations, the preliminary experimental results have demonstrated the potential of using the proposed method to further boost helical pitch and maintain image quality in SSCT scans. It should also be noted that a similar strategy can

be applied to dual-source CT scanners to enable data acquisition with an even higher helical pitch (e.g., $p = 4$ or 5) than what is currently available (e.g., $p = 3$).

5 | CONCLUSION

In this study, a fast acquisition mode with ultra-fast helical pitch was proposed for SSCT. An in-house CNN model (UFP-net) with customized functional blocks was developed to tackle the helical artifacts caused by insufficient sampling of projection data. The network parameters were optimized using the in-house loss function and multi-stage training strategy. UFP-net was able to largely suppress image artifacts and restore the underlying anatomical structure at two exemplar ultra-fast-pitch settings. In summary, the presented method has the potential to enable ultra-fast-pitch data acquisition mode in SSCT, which can be used to dramatically improve scanning speed without major hardware changes.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Kudo H, Rodet T, Noo F, Defrise M. Exact and approximate algorithms for helical cone-beam CT. *Phys Med Biol.* 2004;49(13):2913–2931. [PubMed: 15285256]
2. Stierstorfer K, Rauscher A, Boese J, Bruder H, Schaller S, Flohr T. Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch. *Phys Med Biol.* 2004;49(11):2209–2218. [PubMed: 15248573]
3. Bauer RW, Schell B, Beeres M, et al. High-pitch dual-source computed tomography pulmonary angiography in freely breathing patients. *J Thorac Imaging.* 2012;27(6):376–381. [PubMed: 22487993]
4. Booiij R, Dijkshoorn ML, van Straten M, et al. Cardiovascular imaging in pediatric patients using dual source CT. *J Cardiovasc Comput Tomogr.* 2016;10(1):13–21. [PubMed: 26524989]
5. Flohr TG, Leng S, Yu L, et al. Dual-source spiral CT with pitch up to 3.2 and 75 ms temporal resolution: image reconstruction and assessment of image quality. *Med Phys.* 2009;36(12):5641–5653. [PubMed: 20095277]
6. Ertel D, Lell MM, Harig F, Flohr T, Schmidt B, Kalender WA. Cardiac spiral dual-source CT with high pitch: a feasibility study. *Eur Radiol.* 2009;19(10):2357–2362. [PubMed: 19565245]
7. Achenbach S, Marwan M, Schepis T, et al. High-pitch spiral acquisition: a new scan mode for coronary CT angiography. *J Cardiovasc Comput Tomogr.* 2009;3(2):117–121. [PubMed: 19332343]
8. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys.* 2017;44(10):e360. [PubMed: 29027238]
9. Yang Q, Yan P, Zhang Y, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging.* 2018;37(6):1348–1357. [PubMed: 29870364]
10. Shan H, Padole A, Homayounieh F, et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nat Mach Intell.* 2019;1(6):269–276. [PubMed: 33244514]

11. Nakamura Y, Higaki T, Tatsugami F, et al. Deep learning-based CT image reconstruction: initial evaluation targeting hypovascular hepatic metastases. *Radiology: Artificial Intelligence*. 2019;1(6):e180011. [PubMed: 33937803]
12. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process*. 2017;26(9):4509–4522. [PubMed: 28641250]
13. Han YS, Yoo J, Ye JC. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis. *arXiv preprint arXiv:161106391*. 2016.
14. Li Y, Li K, Zhang C, Montoya J, Chen G. Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. *IEEE Trans Med Imaging*. 2019;38(10):2469–2481. [PubMed: 30990179]
15. Shen L, Zhao W, Xing L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nat Biomed Eng*. 2019;3(11):880–888. [PubMed: 31659306]
16. Würfl T, Ghesu FC, Christlein V, Maier A. Deep Learning Computed Tomography: Medical Image Computing and Computer-Assisted Intervention- MICCAI 2016, Athens, Greece, 17-21 10 2016. Springer; 2016.
17. Zhang H, Li L, Qiao K, et al. Image prediction for limited-angle tomography via deep learning with convolutional neural network. 2016. *arXiv preprint arXiv:160708707*.
18. Gu J, Ye JC. Multi-scale wavelet domain residual learning for limited-angle CT reconstruction. 2017. *arXiv preprint arXiv:170301382*.
19. Hammernik K, Würfl T, Pock T, Maier A. A Deep Learning Architecture for Limited-Angle Computed Tomography Reconstruction: Bildverarbeitung für die Medizin, Berlin, Heidelberg, 12-14 3. Springer; 2017.
20. Würfl T, Hoffmann M, Christlein V, et al. Deep learning computed tomography: learning Projection-domain weights from image domain in limited angle problems. *IEEE Trans Med Imaging*. 2018;37(6):1454–1463. [PubMed: 29870373]
21. Bubba TA, Kutyniok G, Lassas M, et al. Learning the invisible: a hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Probl*. 2019;35(6):064002.
22. Huang Y, Preuhs A, Lauritsch G, Manhart M, Huang X, Maier A. Data Consistent Artifact Reduction for Limited Angle Tomography with Deep Learning Prior: Machine Learning for Medical Image Reconstruction, Shenzhen, China, 17 10 2019. Springer; 2019.
23. Zhou B, Lin X, Eck B. Limited angle tomography reconstruction: synthetic reconstruction via unsupervised sinogram adaptation. In: Chung A, Gee J, Yushkevich P, Bao S. eds. *Information Processing in Medical Imaging*, Springer; 2019:141–152.
24. Katsevich A Theoretically exact filtered backprojection-type inversion algorithm for spiral CT. *SIAM J Appl Math*. 2002;62(6):2012–2026.
25. Tang X, Hsieh J, Nilsen RA, Dutta S, Samsonov D, Hagiwara A. A three-dimensional-weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT-helical scanning. *Phys Med Biol*. 2006;51(4):855–874. [PubMed: 16467583]
26. Liang K, Yang H, Xing Y. Comparison of projection domain, image domain, and comprehensive deep learning for sparse-view X-ray CT image reconstruction. 2018. *arXiv preprint arXiv:180404289*.
27. Gong H, Ren L, McCollough C, Yu L. Ultra-fast-pitch acquisition and reconstruction in helical CT. *SPIE*. 2020; 11312:1131209.
28. Gong H, Leng S, McCollough C, Yu L. A Deep-Learning Based Lower-Dose CT Simulation Technique in Image Domain: 61st AAPM Annual Meeting & Exhibition, San Antonio, TX, 14-18 7 2019. American Association of Physicist in Medicine; 2019.
29. Gong H, Leng S, Yu L, et al. Convolutional Neural Network Based Material Decomposition with a Photon-Counting-Detector Computed Tomography System: 60th AAPM Annual Meeting & Exhibition, Nashville, TN, 29 7-2 August 2018. American Association of Physicist in Medicine; 2018.
30. Gong H, Tao S, Rajendran K, Zhou W, McCollough CH, Leng S. Deep-learning-based direct inversion for material decomposition. *Med Phys*. 2020;47(12):6294–6309. [PubMed: 33020942]

31. Gatys L, Ecker AS, Bethge M. Texture Synthesis Using Convolutional Neural Networks:Advances in Neural Information Processing Systems, Montreal, Canada, 7-12 12 2015. Springer; 2015.
32. Fletcher JG, Fidler JL, Venkatesh SK, et al. Observer performance with varying radiation dose and reconstruction methods for detection of hepatic metastases. *Radiology*. 2018;289(2):455–464. [PubMed: 30204077]
33. Joseph PM. An improved algorithm for reprojecting rays through pixel images. *IEEE Trans Med Imaging*. 1982;1(3):192–196. [PubMed: 18238275]
34. Hoffman J, Young S, Noo F, McNitt-Gray M. Technical note: freeCT_wFBP:a robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT. *Med Phys*. 2016;43(3):1411–1420. [PubMed: 26936725]
35. Flohr TG, Stierstorfer K, Ulzheimer S, Bruder H, Primak AN, McCollough CH. Image reconstruction and image quality evaluation for a 64-slice CT scanner with z-flying focal spot. *Med Phys*. 2005;32(8):2536–2547. [PubMed: 16193784]
36. Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
37. Aydin TO, Mantiuk R, Myszkowski K, Seidel H-P. Dynamic range independent image quality assessment. *ACM Trans Graph*. 2008;27(3):1–10.
38. Gong H, Hu Q, Walther A, et al. Deep-learning-based model observer for a lung nodule detection task in computed tomography. *J Med Imaging*. 2020;7(4):042807.
39. Gong H, Yu L, Leng S, et al. A deep learning- and partial least square regression-based model observer for a low-contrast lesion detection task in CT. *Med Phys*. 2019;46(5):2052–2063. [PubMed: 30889282]

FIGURE 1.

(a) Graphical illustration of projection data acquisition in single-source helical computed tomography (CT) with one exemplar setting of pitch value of 1. The dotted helix indicates the X-ray source trajectory. The small sphere on the beam trajectory represents the X-ray source. The curved grid on the opposite side of X-ray source represents the curved X-ray detector in a clinical CT scanner. (b) Geometric illustration of data condition with respect to pitch = 1 or 3. For illustration purpose, the detector was simplified to have a single column with four detector elements. Each solid line indicates the measured projection data per detector element per helical turn (i.e., a 360° revolution). “View angle (λ)”—the in-plane angular position of X-ray source per 360° revolution. “Pos #1” and “Pos #2”—the detector longitudinal position at the beginning and the end of a single helical turn. Of note, the missing projection data may be partially restored by the complementary projection data acquired at the corresponding opposite view angle, which depends on the specific configuration of pitch setting and reconstruction field-of-view (FOV). (c) Examples of paired chest CT images of the same patient (with 10-mm interval between two consecutive slices) corresponding to two pitch settings of 1 and 3. Standard clinical CT image reconstruction algorithm, that is, filtered-backprojection algorithm, was used to

reconstruct the chest image from the measured projection data. Severe image artifacts can be observed with a pitch of 3 because of the missing data problem

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 2.

(a) Structure of the proposed ultra-fast-pitch network (UFP-net). UFP-net uses fast-pitch images to predict regular-pitch images. This figure was modified from our recent conference proceeding.²⁷ (b) The diagrams for the first block and the two versions of inception-discrete cosine transformation blocks (i.e., Incept-DCT, v1/v2). “Conv”—standard convolutional layer. “Up-conv”—up-convolutional layer. “BN”—batch-normalization. “Leaky ReLU”—leaky rectified linear unit (the corresponding negative slope fixed at 0.1). No max pooling was used in the UFP-net. In the contracting path, the “ 1×1 conv” in Incept-DCT v1 was used with stride 2 to achieve down-sampling. In the up-sampling path, the “ 1×1 up-conv” in Incept-DCT v2 (the up-sampling path) was used with stride 2. The other convolutional layers used stride 1. Dropout layer was added after each Incept-DCT block, and the dropout rate was empirically fixed at 0.1

FIGURE 3.

Loss curves of UFP-net trained for two helical pitch settings at $p = 2$ and 3. “Train loss”—training loss; “Val loss”—validation loss

FIGURE 4.

Examples of an artifact-corrupted CT image from a randomly selected testing case (with a pitch of $p = 3$, i.e., “pressure test”), the corresponding reference image, and the corresponding feature maps from the proposed neural network. Feature maps #1 and #2 were randomly selected from the neural activation corresponding to the local operators in the first Incept-DCT v1 block, while feature map #3 was randomly selected from the neural activation corresponding to non-local operators in the same block. Feature maps #4, #5, and #6 were similarly extracted from a deeper Incept-DCT block. “FBP”—filtered-back-projection image, that is, standard clinical CT image reconstruction algorithm. “Reference”—the FBP image corresponding to the regular helical pitch of $p = 1$. Of note, local operators tend to explore the local anatomical structure and helical artifact, while non-local operators

could represent more generic structure information. Display window for CT images: $[-1350$
 $150]$ Hounsfield Unit (HU). Display window for activation maps: $[-1\ 1]$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 5.

Examples of thoracic and abdominal images (with helical pitch setting at $p = 2$) reconstructed with standard FBP algorithm and the proposed method (i.e., “UFP-net” with in-house MIF loss). The two slices per case were selected with 20-mm interval along the z -direction (i.e., along the depth of image volume). “Reference”—FBP images with regular helical pitch of 1, which served as the reference in evaluation. The zoom insets were ~ 1.6 times of the square region-of-interests (ROIs). The dotted lines (i.e., L1–L4) indicate the line profiles plotted in Figure 6. The local structural similarity index (SSIM), that is, SSIM per ROI, is presented next to the inset. Image display window (W/L): thorax 1500/–600 HU; abdomen 400/40 HU. Of note, these two display window settings were commonly used in routine clinical CT diagnostic tasks

FIGURE 6.

(a) Line profiles extracted from Figure 5. (b) Comparison of mean global SSIM and relative root-mean-square error (rRMSE) per testing case. Both SSIM and rRMSE were calculated across the complete image FOV

FIGURE 7.

Examples of thoracic and abdominal images (with helical pitch setting at $p = 3$, i.e., the “pressure test”) reconstructed with the standard FBP algorithm and the proposed method (i.e., “UFP-net with MIF.” The two slices per case were selected with 20-mm interval along the z-direction (i.e., along the depth of image volume). “Reference”—FBP images with a regular helical pitch of 1, which served as the reference in evaluation. The zoom insets were ~ 1.6 times of the square ROIs. The dotted lines (i.e., L1–L4) indicate the line profiles plotted in Figure 8. The local SSIM, that is, SSIM per ROI, is presented next to the inset. Display window (W/L) of thorax images was fixed at a typical clinical setting for chest CT: 1500/–600 HU. Display window of abdominal examples was fixed at 1000/40 HU for the convenience of illustration

FIGURE 8.

(a) Line profiles extracted from Figure 7. (b) Comparison of mean SSIM and rRMSE per testing case. Both SSIM and rRMSE were calculated across the complete image FOV

FIGURE 9.

Examples of hypo-attenuated liver lesion images from testing cases at different helical pitches ($p = 2$ and 3). Left column: standard FBP before correction. Middle column: the presented method, that is, UFP-net with MIF loss. Right column: the reference images, that is, FBP with regular helical pitch ($p = 1$). The displayed SSIM was calculated at the lesions. Display window (W/L) for $p = 2$ was fixed at the standard setting of clinical abdominal CT: 400/40 HU. Display window (W/L) for $p = 3$ was 1000/40 HU for the convenience of illustration. Arrows indicate lesion location

FIGURE 10.

Examples of coronal slices of thorax and abdomen from four randomly selected testing cases at different helical pitches ($p = 2$ and 3). Left column: standard FBP before correction. Middle column: the presented method, that is, “UFP-net with MIF.” Right column: the reference images—FBP images with regular helical pitch of 1. The display window of thorax examples was fixed at a typical clinical setting (W/L): 1500/−600 HU. Display window of abdomen examples was fixed at 1000/40 HU for the convenience of illustration

TABLE 1

Major parameters of scanning protocols used in original patient cohort *

CT scanner	128-slice computed tomography (CT) systems: Siemens Definition, Definition Flash, or Definition AS+ (Siemens Healthcare, Inc.)
Range of anatomy in scan	Chest-abdomen-pelvis, chest-abdomen, or abdomen-pelvis
Reference tube voltage (kV) **	120
Reference quality-reference-mAs	200
Scanning field-of-view (FOV)	500 mm
Detector collimation †	128 × 0.6
Helical pitch	0.6
Image thickness/increment (mm)	3/3

* The parameters are from routine CT scanning protocols used at our institution.

** CareKV was used.

† Z-flying focal spot: double sampling along the longitudinal direction. Actual beam width 64 × 0.6 mm.

TABLE 2

Simulation setup for virtual projection data acquisition

Source-to-detector distance	1085.6 mm
Source-to-isocenter distance	595 mm
Scanning FOV	500 mm
Detector size	16 × 368
Views per gantry rotation	576
Image reconstruction	Filtered-back-projection with medium sharpness filter
Reconstruction FOV	380 mm
Image thickness/increment	1/1 mm
Image size	512 × 512