Multi-Scale Dilated Convolution Neural Network for Image Artifact Correction of Limited-Angle Tomography

HAICHUAN ZHOU1,2, YINING ZHU1,2, (Member, IEEE), QIAN WANG3, (Student Member, IEEE), JINQIU XU4, GE LI5, DEFENG CHEN1,2, YINGYING DONG6, AND HUITAO ZHANG1,2

1 School of Mathematical Sciences, Capital Normal University, Beijing 100048, China
2 Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China
3 Department of Electrical and Computer Engineering, University of Massachusetts–Lowell, Lowell, MA 01854, USA
4 Paul Scherrer Institut, 5232 Villigen, Switzerland
5 Ping An Technology, Beijing 100000, China
6 Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

ABSTRACT Limited-angle computed tomography (CT) has arisen in some medical and industrial applications. It is also a challenging problem since some scan views are missing and the directly reconstructed images often suffer from severe distortions. For such kind of problems, we analyze the features of limited-angle CT images and propose a multi-scale dilated convolution neural network (MSD-CNN) to correct the artifacts and to restore the image. In this network, the dilated convolution layer and multi-scale pooling layer are combined to form a group and exited in the whole encoder-decoder process. Since the dilated convolutions support an exponential expansion of the receptive field without losing resolution and coverage, the obtained artifact features possess the multi-scale characteristic. Furthermore, to improve the effectiveness and accuracy of the training step, we employ a preprocessing method, which extracts image patches. Numerical experiments verify the out-performance of the proposed method compared with some conventional methods, such as Unet based deep learning, TV- and $L_0$-based optimization methods.

INDEX TERMS Limited-angle tomography, artifact correction, multi-scale, dilated convolution.

I. INTRODUCTION Limited-angle tomography, artifact correction, multi-scale, dilated convolution.
of the gradient $L_0$-norm minimization problem have been studied [13]–[16]. Because the $L_0$-norm is superior to the $L_1$-norm in sparsity expression, the gradient $L_0$-norm based optimization method outperforms to the TV based one, especially for edge maintenance.

In recent years, some researchers have introduced the idea of deep learning to limited-angle CT problems as well as low does problem [17]–[20]. According to the differences in the processing steps, it can be classified into three categories: image post-processing, projection pre-processing and reconstruction processing.

- **Image post-processing**: image post-processing which combines traditional methods with deep learning methods is an end-to-end training to learn full-angle images from limited-angle images. The artifacts of limited-angle CT images have certain directionality and global distribution. In 2016, Hanming Zhang et al. first applied deep learning to the limited-angle CT problem and achieved satisfactory results when the missing-angle range is small [21], [22]. Jawook Gu et al. studied the characteristics of limited-angle image artifacts and proposed a method to restore the image by learning multi-scale wavelet coefficients [23].

- **Projection pre-processing**: the incomplete projection data can be learned to achieve a complete version before the reconstruction process. Anirudh et al. proposed a system consisting of a one- and a two-dimensional convolutional neural network to recover the missing projection data. This method improved the reconstruction quality and was applied to baggage CT systems [24], [25]. Projection pre-processing is to learn full-angle projection data from limited-angle projection data. The disadvantage of projection pre-processing is that small changes in projection domain have a great impact on the reconstructed result.

- **Reconstruction processing**: reconstruction processing is to employ deep learning in the reconstruction process and to reconstruct images from limited-angle projection data. The process is staged and the learning process is more complex. Deep learning can also be employed in the reconstruction process. Hammernik proposed a method that can directly reconstruct images from finite-angle projections [26]. This method first learned the missing weights of the data in the projection field and correspondingly corrected the intensity changes. Then by employing a variational network, the stripe artifacts are effectively suppressed.

Considering the impact from data sets and processing procedure, in this paper, we propose an end-to-end deep neural network for the limited-angle CT problem. The network contains two subnetworks for main structure correction and fine artifact removal respectively. Experimental results show that the proposed network outperforms the conventional methods for the limited-angle CT, even in the case that the scan angle range is restricted to 130°.

II. METHOD

The post-processing method for the limited-angle artifact correction is to establish a mapping relationship between a limited- and a full-angle reconstructed image via a deep neural network. The parameters of the network are obtained by supervised learning for a training dataset. In this work, we design a new end-to-end network structure, where the CT image reconstructed from limited-angle projection data is employed as the network input, and the one from complete projection data is adopted as the network label value. Without loss of generality, we define the input image as $F$, the label image as $G$, and the output image as $P$, with the same size of $N \times N$, i.e., $F, G, P \in R^{N \times N}$. We establish a relationship as follow,

$$P \equiv \Phi(F),$$  \hspace{1cm} (1)

where the mapping function $\Phi : R^{N \times N} \rightarrow R^{N \times N}$ represents the image recovery process. Eq. (1) is then converted to an optimization model,

$$\Phi = \arg \min_{\Phi} \|\Phi(F) - G\|_2^2.$$  \hspace{1cm} (2)

To develop a deep learning based solution method for optimization problem (2), we propose a multi-scale dilated convolution neural network (MSD-CNN). As is illustrated in Fig. 1, the entire network is based on an “encoder-decoder” structure with a refining stage.
FIGURE 2. The schematic diagram of Dilated Convolution. Left image is generated from the original image by a 1-dilated convolution with a receptive field of $3 \times 3$; Middle image is generated from Left image by a 2-dilated convolution with a receptive field of $7 \times 7$; Right image is generated from Middle image by a 4-dilated convolution with a receptive field of $15 \times 15$.

Dilated convolution was originally developed for wavelet decomposition [28]. The main idea of dilated convolution is to insert “holes” (zeros) between pixels in convolutional kernels to increase image resolution, thus enabling dense feature extraction in deep CNNs [29]. These zeros are able to be considered as “gaps” and the different width of the gap corresponds to the different dilation rate. Under such a definition, the dilated convolution on 2D data is illustrated in Fig.2. For a given 2D image, when the dilation rate equals 1, it is a normal convolution, when rate = 2 means skipping one pixel per input, and when rate = 4 means skipping 3 pixels. In the example of Fig. 2, red dots are the inputs to a filter which is $3 \times 3$, and yellow area is the receptive field captured by each of these inputs, where the receptive field implies the area captured on the initial input by each input to the next layer.

Dilated convolution provides a method to exponentially increase receptive views of the network via linearly enlarging the field of convolution kernels. Thus, it is suitable for the applications caring more about integrating knowledge of the wider context with less cost. For example, in the semantic segmentation framework, Yu and Koltun [30] employed serialized layers with increasing rates of dilation to enable context aggregation. More recently, dilated convolution has been applied to a broader range of tasks, such as object detection [31] and audio generation [32].

Multi-scale pooling (MSP) is to use pooling layer (or convolution kernel) with different sizes to extract the features of images at various scales, so as to prevent the loss of network information and to improve the network training. As shown in Fig. 3, assuming the size of input image is $M \times N$ and the channel number is $n$, the channels are added up after 2 × 2, 4 × 4 and 6 × 6 Pool&Conv with stride 2, padding 0, 1, 2 respectively, and the n-channel images are obtained with the size of $M/2 \times N/2$. MSP is adapted according to the spatial pyramid pooling and initially applied in the field of object detection [33]. By extracting features through different-size pooling and generating fixed-size feature vectors, MSP avoids the information loss in candidate region images caused by clipping, deformation and other factors.

In our neural network, we combine a dilated convolution layer and a multi-scale max pooling layer to form a block named as MSD block. These blocks effectively impact the encoding and decoding stages. The left part of MSD-CNN is the encoding stage, which contains 7 convolution layers and 4 MSD blocks. After 4 down-sampling operations, the resolution of the image is reduced to $N/4 \times N/4$. In the decoding stage, 4 up-sampling operations (deconvolution) are performed (including BN and ReLu layers). The result of each up-sampling step is combined with the same resolution image, i.e. the result of the MSD block in the encoding stage. Finally, the $N \times N$ output result is obtained. The refining stage is a shallow convolution network consisting of four $3 \times 3$ convolution layers and one $1 \times 1$ convolution layer. The input is a dual-channel image that combines the same resolution image, i.e. the result of the MSD block in the encoding stage. Finally, the $N \times N$ output result is obtained. The refining stage is a shallow convolution network consisting of four $3 \times 3$ convolution layers and one $1 \times 1$ convolution layer. The input is a dual-channel image that combines the result of the decoding stage with the original input image, and the final output is the corrected limited-angle CT image [34]. Although there are 3 stages in the neural network, the MSD-CNN can be learned as a whole network without pre-training.

III. EXPERIMENTS

A. DATASET GENERATION AND PREPROCESSING

In this paper, the training and testing datasets are from an open source dataset called Liver Tumor Segmentation Challenge
FIGURE 4. An example of the chest CT image data set employed in this work. The scan angle range is \([0^\circ, 150^\circ]\).

(LiTS) (https://competitions.codalab.org/competitions/1559). The LiTS dataset contains 200 three-dimensional (3D) CT scans of the human body, which cover the chest to the pelvic region. In this experiment, we choose the chest (lung) as the target object and generate a limited-angle dataset according to the following steps:

1) randomly select 10 lung-region slices (2D image) from each 3D CT data as label images. The size of each slice is \(512 \times 512\);
2) perform a parallel projection to each slice with a scan angle range of \([0^\circ, 150^\circ] \cup [0^\circ, 130^\circ]\) to obtain the limited-angle projection and the angular sampling rate is \(0.5^\circ\);
3) implement the SART algorithm for the projection data obtained in step 2) to reconstruct the limited-angle CT images.

The limited-angle CT image obtained in step 3) is the input image of the network, and the 2D slice of lung obtained from the first step is the label image. We randomly choose 160 out of the 200 3D CT data as the source of the training set and the rest 40 as the source of the test set. Thus, the data set includes \(160 \times 10 = 1600\) training images and \(40 \times 10 = 400\) testing images. Fig. 4 shows an example of the data set, where the first row is a label image and two zoom-in patches, and the second row is correspondingly from the network input, i.e., the limited-angle CT image.

From the limited-angle CT images, it is obvious that the large structural defects and artifacts usually affect a local region rather than the entire image. Hence, we utilize a strategy that divides the image into small patches for the training usage. As a result, both the training efficiency and accuracy are improved.

As shown in Fig. 5, the sliding window is used to extract the patches from the input and the label images. Then, a \(256 \times 256\) patch pair is obtained per 128 pixels. For each image with the size of \(512 \times 512\), we divide it to nine \(256 \times 256\) small patches for local training usage. Therefore, after the extracting operation, the sample number of the data set is \(14400(1600 \times 9)\). However, in the test experiments, we only use the entire image instead of the patches to avoid the merging operation.

B. ENVIRONMENT AND HYPER-PARAMETER
We use Pytorch as the deep learning framework. The specific software and hardware environments are shown in Table 1. The elaborated network parameter settings are shown in Table 2.

C. EXPERIMENTAL RESULTS
We design three experiments to validate the MSD-CNN structure and the artifact correction performance respectively. Some image quality assessments (IQAs) are employed for quantitative comparison, such as Peak Signal Noise Ratio (PSNR), Structural Similarity Index (SSIM) [35] and Universal Image Quality Index (UIQI). The UIQI is caculated
TABLE 1. The software and hardware environments of the numerical experiments.

|                      |         |
|----------------------|---------|
| **software**         |         |
| Operation System     | Ubuntu 16.04 64-bit |
| Deep Learning Framework | Pytorch 0.3.1 |
| GPU Platform         | Cuda9 0, Cudnn7 |
| **hardware**         |         |
| GPU                  | NVIDIA TITAN X 12G |
| Memory               | 64G     |
| CPU                  | Inter i7-4930k |

TABLE 2. The hyperparameter settings of the network.

| Parameter                              | Value                  |
|----------------------------------------|------------------------|
| optimizer                              | Adam                   |
| convolution layer bias initial parameters | 0                     |
| convolution layer weight initial parameters | Gaussian(0, \(\sqrt{2/n}\)) |
| base learning rate                     | 1e-4                   |
| train batch size                       | 18                     |
| test batch size                        | 3                      |
| training iterations                    | 150 epochs              |
| Training time                          | 33h 19m                |

as follow,

\[Q = \frac{4\sigma_{xy} \bar{x} \bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)}\]

where \(\bar{x}\) and \(\bar{y}\) represent the mean of the reference image and the reconstructed image, \(\sigma_x\) and \(\sigma_y\) the variances of \(\bar{x}\) and \(\bar{y}\), and \(\sigma_{xy}\) the covariance of \(\bar{x}\) and \(\bar{y}\). If the two images are identical, \(Q = 1\). And the closer \(Q\) is to 1, the better image quality is achieved.

First, we compare the MSD-CNN with different module combinations, e.g., without dilated convolution layers (undilated), without MSP layers (MSP-free) and without refining stage (unrefined). For each scan angle range \([0^\circ, 150^\circ]\) \(\cup\) \([0^\circ, 130^\circ]\), the neural networks are trained with the same aforementioned parameters and datasets. The images predicted from these neural networks are shown in Fig.6 and Fig.7. The corresponding IQA values are listed in Table 3.

Loss curves of training and testing datasets with different MSD-CNN modules are shown in Fig. 8.

According to Figs. 6 and 7, it is obvious that the proposed original MSD-CNN superiorly recover local details and fine structures comparing with other modifications. The quantitative evaluations in Table 3 have the consistent performances. The loss curves of training and testing datasets with different MSD-CNN modules also prove the superiority of MSD-CNN network. Thus, the MSD-CNN visually and numerically outperforms in image prediction. The results indicate that the dilated convolution layer, the MSP layer as well as the refining stage are all essential to the MSD-CNN.

Then in the second experiment, we compare the MSD-CNN with a classical network i.e. Unet, and a network without input image in the refining stage (Single-Channel).

The initialization parameters and datasets are fixed the same for all the comparison methods. The images predicted from these neural networks are shown in Figs. 9 and 10. The corresponding IQA values in Tables 4 and 5 have consistent performances. The loss curves of training and testing datasets with different network structures are shown in Fig. 11.

According to Figs. 9 and 10, Unet and Single-Channel networks are blurred in some small structures and appear black shadows, while MSD-CNN performs well. The quantitative evaluations in Tables 4 and 5 have consistent performances. The loss curves of training and testing datasets
with different network structures also prove the superiority of MSD-CNN network. Thus, the MSD-CNN has a visual and numerical outperformance in image prediction. The results indicate that MSD-CNN is superior to the other two network structures.

Finally, in the third experiment, we employ some traditional approaches as comparisons, including the SART, $L_0$- and TV-based regularization methods, to verify the
superiority of the MSD-CNN in artifact correction of the limited-angle CT.

Figs. 12 (example 1) and 13 (example 2) show the results by using different methods. It is noticeable that the $L_0$- and TV-based method can recover most of the main image boundaries, but the restored fine details are not satisfactory. Some “block artifacts” and blurring effect still exist in the $L_0$- and TV-based results. However, the images processed by the proposed neural network dramatically improve in structure restoration and artifact removal.

Figs. 14 (example 1) and 15 (example 2) show the processed CT results by using different methods with the same limited-angle range of $0^\circ - 130^\circ$. It indicates that the MSD-CNN still works well with a narrow scan rang.
with TV and $L_0$ regularizations fail to restore some important image structures and the blurring artifacts are still remained.

From Table 6, it is noticed that the image processed by the MSD-CNN consistently obtains better quality assessments than the conventional methods when the scan angle range varies.

The scan angle of Fig. 15 is 130° which leads more obvious artifacts in reconstructed image than the case of 150°. Hence, the quality (especially in the soft-tissue regions) of the images predicted by MSD-CNN decreases for the case of 130° than 150° (Fig. 13). However, comparing to other conventional methods or deep learning networks (Unet) with the same scan angle setting, the proposed method has the best performance in structure maintainance. It is noticed that even with a narrow scan angel (130°), the image processed by MSD-CNN still has a better performance than the other methods with 150° scan angle (Figs. 15 and 13, and Table 6).

To further demonstrate the effectiveness of the proposed network, we calculate an average IQA values for all the output images, seen Table 7. The corresponding results show that the proposed method achieves the desired performance and effectively overcome the limited-angle-caused blur vision.

### IV. CONCLUSION

In this paper, we propose a multi-scale dilated neural network (MSD-CNN) for the artifact correction of limited-angle CT problems.

The design of the MSD-CNN is based on an “encoder-decoder” structure with several dilated convolution layers
and multi-scale pooling layers. Meanwhile, a shallow convolution layer is composed to refine the limited-angle artifact correction. These different modules make the MSD-CNN have a superior capability to deal with the artifacts caused by limited-angel tomography. In addition, considering the local characteristics of the limited-angle artifacts, we design a method to extract the image patches, which retains the original image information and meanwhile increases the size of the training dataset. Thus, the training efficiency and accuracy are further improved. We verify the MSD-CNN with different module combinations and compare it with some conventional networks and methods, such as U-Net, TV-and $\ell_0$-based optimization methods. The results demonstrate the necessity of each MSD-CNN module and combined superiority. Moreover, it shows the outperformance of the proposed MSD-CNN than the conventional networks and methods for artifact correction of limited-tomography.

REFERENCES
[1] F. Natterer and F. Wübbeling, Mathematical Methods in Image Reconstruction, vol. 5. Philadelphia, PA, USA: SIAM, 2001.
[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress, “Practical cone-beam algorithm,” J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 1, no. 6, pp. 612–619, Jun. 1984.
[3] R. Gordon, R. Bender, and G. T. Herman, “Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography,” J. Theor. Biol., vol. 29, no. 3, pp. 471–481, Dec. 1970.
[4] A. Andersen, “Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm,” Ultrason. Imag., vol. 6, no. 1, pp. 81–94, Jan. 1984.
[5] E. Y. Sidky and X. Pan, “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization,” Phys. Med. Biol., vol. 53, no. 17, pp. 4777–4807, Sep. 2008.
[6] J. Tang, B. E. Nett, and G.-H. Chen, “Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms,” Phys. Med. Biol., vol. 54, no. 19, pp. 5781–5804, Oct. 2009.
[7] J. Bian, J. H. Siewersden, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, “Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT,” Phys. Med. Biol., vol. 55, no. 22, pp. 6575–6599, Nov. 2010.
[8] G.-H. Chen, J. Tang, and S. Leng, “Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets,” Med. Phys., vol. 35, no. 2, pp. 660–663, Jan. 2008.
[9] J. Liu, J. Ma, Y. Zhang, Y. Chen, J. Yang, H. Shu, L. Luo, G. Couteix, W. Yang, Q. Feng, and W. Chen, “Discriminative feature representation to improve projection data inconsistency for low dose CT imaging,” IEEE Trans. Med. Imag., vol. 36, no. 12, pp. 2499–2509, Dec. 2017.
[10] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Couteix, and W. Chen, “Artifact suppressed dictionary learning for low-dose CT image processing,” IEEE Trans. Med. Imag., vol. 33, no. 12, pp. 2271–2292, Dec. 2014.
[11] J. Liu, Y. Hu, J. Yang, Y. Chen, H. Shu, L. Luo, Q. Feng, Z. Gui, and G. Couteix, “3D feature constrained reconstruction for low-dose CT imaging,” IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 5, pp. 1232–1247, May 2018.
[12] H. Yu and G. Wang, “A soft-threshold filtering approach for reconstruction from a limited number of projections,” Phys. Med. Biol., vol. 55, no. 13, pp. 3905–3916, Jul. 2010.
[13] L. Xu, C. Lu, Y. Xu, and J. Jia, “Image smoothing via Lo gradient minimization,” ACM Trans. Graph., vol. 30, no. 6, 2011, Art. no. 174.
[14] L. Xu, Q. Yan, Y. Xie, and J. Jia, “Structure extraction from texture via relative total variation,” Togacm Trans. Graph., vol. 31, no. 6, p. 1, Nov. 2012.
HAICHUAN ZHOU received the B.S. degree in applied mathematics from Henan Normal University, Xinxiang, China, in 2016, and the M.S. degree in computational mathematics from Capital Normal University, Beijing, in 2019, where she is currently pursuing the Ph.D. degree. Her current research interests include image reconstruction, deep learning-based artifact correction for computed tomography, and image processing.

YINING ZHU (Member, IEEE) received the B.S. degree in information and scientific computing from the Beijing Information and Technology Institute, in 2006, and the M.S. and Ph.D. degrees in mathematics and information technology from Capital Normal University, Beijing, China, in 2009 and 2012, respectively. He was a Postdoctoral Researcher with Peking University, Beijing, from 2013 to 2014. He is currently an Associate Professor with the School of Mathematical Science, Capital Normal University. His current research interests include image reconstruction, deep learning-based artifact correction, and image processing.

QIAN WANG received the B.S. degree in mathematics and applied mathematics from Hebei Normal University, Hebei, China, in 2012, and the M.S. degree in mathematics and information technology from Capital Normal University, Beijing, in 2016. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Massachusetts–Lowell, Lowell, MA, USA. Her research interests include image reconstruction, spectral CT, and medical image processing.

GE LI received the B.S. degree in applied mathematics and the M.S. degree in mathematics and information technology from Capital Normal University, Beijing, China, in 2015 and 2018, respectively. She is currently with Ping An Technology. Her current job interest in medical image processing based on deep learning.

DEFENG CHEN received the B.S. degree in computer science and technology from the Beijing Information and Technology Institute, in 2005, and the M.S. and Ph.D. degrees in mathematical and information technology from Capital Normal University, Beijing, China, in 2008 and 2016, respectively. He is currently a Laboratory Technician with the School of Mathematical Science, Capital Normal University. His current research interests include image processing, deep learning, and image reconstruction.

YINGYING DONG received the B.S. degree in information and computation science from Shandong Normal University, Shandong, China, in 2006, the M.S. degree in mathematics of computation from Capital Normal University, Beijing, China, in 2009, and the Ph.D. degree in agricultural remote sensing and information technology from Zhejiang University, Zhejiang, China, in 2013. She was a Postdoctoral Researcher with the University of Lethbridge, Canada, from 2013 to 2015. She is currently an Associate Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. Her research interest includes research and application in vegetation remote sensing.

HUITAO ZHANG received the B.S. degree in mathematics from Hebei Normal University, Anhui, China, in 2012, and the M.S. and Ph.D. degrees in mathematics and information technology from Capital Normal University, Beijing, in 2015 and 2019, respectively. She is currently a Postdoctoral Researcher with the Paul Scherrer Institute, Switzerland. Her research is focused on image reconstruction, X-Ray phase contrast imaging, and image processing.

JINQIU XU received the B.S. degree in applied mathematics from Huaibei Normal University, Anhui, China, in 2002, and the M.S. and Ph.D. degrees in mathematics and information technology from Capital Normal University, Beijing, in 2005 and 2008, respectively. He was a Postdoctoral Researcher with the Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, from 2008 to 2010. He is currently an Associate Professor with the School of Mathematical Science, Capital Normal University. His current research interests include image reconstruction, deep learning-based artifact correction, and X-Ray phase contrast imaging.