PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

CT artifact reduction via U-net CNN

C. Zhang, Y. Xing

C. Zhang, Y. Xing, "CT artifact reduction via U-net CNN," Proc. SPIE 10574, Medical Imaging 2018: Image Processing, 105741R (5 March 2018); doi: 10.1117/12.2293903

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

CT artifact reduction via U-net CNN

C. Zhang[a], Y. Xing[ab]
[a]The Department of Engineering Physics, Tsinghua University, Beijing 100084 China.
[b]Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing 100084 China.
Corresponding author: Yuxiang Xing.
Contact information: Email: xingyx@mail.tsinghua.edu.cn, Telephone: (+8610) 62782510, Fax: (+8610) 62782967

ABSTRACT

Purpose: Our preliminary study showed us the capability of a deep learning neural network (DLNN) based method to eliminate a specific type of artifact in CT images. This work is to comprehensively study the applicability of a U-net CNN architecture in improving the image quality of CT reconstructions by respectively testing its performance in various artifact removal tasks.

Methods: A U-net architecture is trained by a big dataset of contaminated and expected image pairs. The expected images known as reference images are acquired from groundtruths or using superior imaging system. A proper initialization of network parameters, a careful normalization of original data and a residual learning objective are incoorprated into the framework to boost training convergence. Both numerical and real data studies are conducted to validate this method.

Results: In numerical studies, we found that the DLNN-based artifact reduction is powerful and can work well in eliminating nearly all artifacts of various types and recovering detailed structural information in low-quliaty images (e.g. plain FBP reconstructions) if the network is trained with groundtruths provided. In real situations where groundtruth is not available, the proposed method can characterize the discrepancy between contaminated data and higher-quality reference labels genertated by other techniques, mimicking their capability of reducing artifacts. Generalization to disjointed data is also examined using testing data. All results show that the DLNN framework can be applied to various artifact reduction tasks and outperforms conventional atifact reduction methods with shorter runtime.

Conclusion: This work gained promising results of the U-net network architecture successfully characterizing both global and local artifact patterns. By forward propagating contaminated images through the trained network, undesired artifacts can be greatly reduced while structural information maintained for an input of CT image. It should be noted that the proposed deep network should be trained independently for each specific case.

I. INTRODUCTION

Diagnostic CT images can be contaminated by various types of artifacts such as metal artifacts, low-dose artifacts, and few-view artifacts, etc. Tons of reconstruction and post-processing techniques were proposed to eliminate a certain type of artifact[1–4]. However, few methods have been developed to remove multiple artifact patterns. Recent works on artifact reduction techniques based on deep learning neural network (DLNN) demonstrate their power to characterize specific artifact patterns under the "instruction" of big training datasets. Chen et al. constructed a three convolutional step network and a residual encoder-decoder CNN as well to mitigate low-dose noise[5,6]; Zhang et al. proposed a simliar three-layer CNN to eliminate limited-angle artifacts[7]; Gjesteby et al. showed pilot results using six convolutional layers to reduce metal streak artifacts[8]. Zhang et al also proposed a metal artifact reduction method based on CNN[9]. While achieving satisfactory performance, they vary a lot in network design and training implementation and mostly focus on single artifact removal task.

In 2015, a U-Net embracing both global and local features is proposed by Ronneberger for image segmentation[10]. Han et al. leveraged this architecture to solve globally distributed few-view problems[11]. Inspired by this work, we attempt to address multiple artifact reduction tasks in image domain such as metal artifact reduction, low-dose artifact reduction, and few-view reduction tasks within a single U-Net framework.

Medical Imaging 2018: Image Processing, edited by Elsa D. Angelini, Bennett A. Landman, Proc. of SPIE
Vol. 10574, 105741R · © 2018 SPIE · CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293903

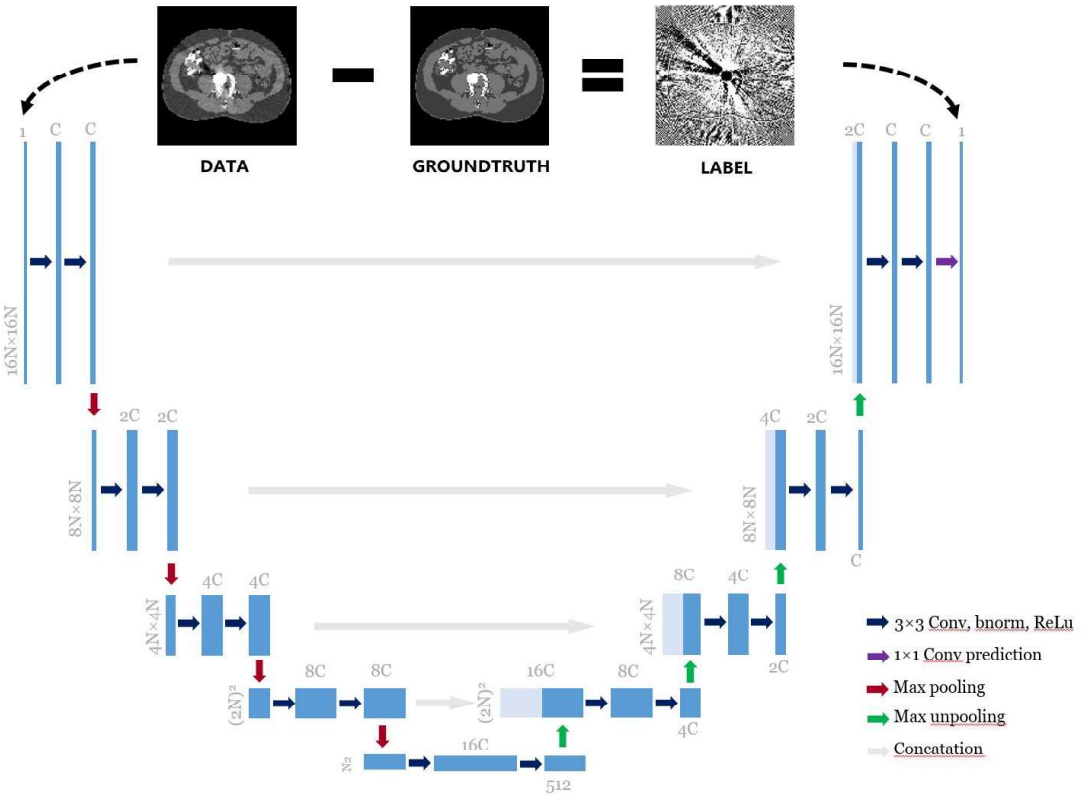

Fig.1 The schematic diagram of the proposed DLNN (example for N×N pixels in the lowest resolution and 16N×16N in the highest resolution). The input, output, layer operations, image size, number of channels and concatation paths are shown with arrows of different colors. Each blue box corresponds to a multi-channel image. Image size is specified on the left of the box. The number of channels are denoted on the top or the bottom of the box. The variable C is the number of channels in the first convolution layer. The light-colored box is the concatation image copied from the previous box. The arrows represent different operations listed in the right-bottom legend.

II. THEORETICAL AND EXPERIMENTAL METHODS

2.1 Deep learning framework

The DLNN based artifact reduction technique can be considered as an implicit recovering operator:

$$\hat{\boldsymbol{\mu}}_{\text{res}} = \mathcal{O}_{\text{res}}\{\hat{\boldsymbol{\mu}}\} \tag{0.1}$$

The operator $\mathcal{O}_{\text{res}}\{\}$ propagates the attenuation map of a contaminated image $\hat{\boldsymbol{\mu}}$ through the corresponding network and acquires the artifact-only image $\hat{\boldsymbol{\mu}}_{\text{res}}$ learned by DLNN. An artifact-free image can be denoted as $\hat{\boldsymbol{\mu}}_{\text{DL}}$. Using a large number of data pairs of artifact contaminated images (denoted by $\hat{\mathbf{M}}$) and reference images (denoted by $\hat{\mathbf{M}}^*$) which we aim to achieve, an optimal operator $\mathcal{O}^*_{\text{res}}$ can be trained via minimizing the following objective function of an ensemble L2 norm:

$$\mathcal{O}^*_{\text{res}} = \arg\min_{\mathcal{O}} \sum_{k\in\mathcal{T}} \left\| \mathcal{O}_{\text{res}}\{\hat{\mathbf{M}}_k\} - \hat{\mathbf{M}}^*_k \right\|_2^2 \tag{0.2}$$

where k indexes a single data pair belonging to the whole dataset $\mathcal{T}$. The objective represents the ensemble error between DLNN learned images $\hat{\mathbf{M}}$ and reference images $\hat{\mathbf{M}}^*$. Trained by the big dataset $\mathcal{T}$, the optimal operator $\mathcal{O}^*_{\text{res}}$ covers all its systematic and statistical information. Thus the operator $\mathcal{O}^*_{\text{res}}$ can be applied to any data of the same format generated by the same scanning system. Note that to avoid numerical problems and stabilize the network during training, the data pairs are firstly normalized as follows:

$$\hat{\mathbf{M}}_{\text{scaled}} = \frac{\hat{\mathbf{M}} - \{\hat{\mathbf{M}}\}_{\min}}{\sigma_{\hat{\mathbf{M}}}} \tag{0.3}$$

$\{\hat{\mathbf{M}}\}_{\min}$ and $\sigma_{\hat{\mathbf{M}}}$ are minimum value and the standard deviation of all pixel values in the training dataset $\hat{\mathbf{M}}$.

We utilize the U-net network architecture[10,11] illustrated in Fig. 1 as the backbone network with respect to the operator $\mathcal{O}$. Specifically, a U-net consists of multiple stages connected by pooling layers in the first half and unpooling layers in the second half. Each stage has two groups and each group comprises of a 3×3 convolution layer (Conv), a batch-normalization layer (Bnorm), and a ReLu layer. The number of channels for each convolution layer is doubled after each pooling layer. A scale by scale concatation strategy is leveraged after each unpooling layer to incorporate the higher resolution structural information from previous stages. The last stage employs a 1×1 Conv layer to predict the output. This multiscale network enlarges the receptive field and potentially achieves better performance in characterizing both local and global artifacts.

The training is implemented using Matconvnet[12] which is a Matlab programming deep learning toolbox. The proposed U-net is trained by stochastic gradient descent (SGD) update with momentum at a constant learning rate and weight decay. The parameters of Conv layers are initialized by Xavier method, while uniform initialization using the following criterion is performed for Bnorm layers:

$$\text{filters} \sim \sigma(\hat{\mathbf{M}}); \text{bias} \sim \text{mean}(\hat{\mathbf{M}}); \text{momentum} \sim 0; \tag{0.4}$$

where $\hat{\mathbf{M}}$ denotes the whole inputs in the dataset. The sysmbol $\sim$ represents the same order of magnitude. The training of 10^4 images generally takes a day on GPU Nvida Tesla-80, CPU Intel Xeon E5-2460 @2.40GHz

2.2 **Datasets preparation**

In simulation studies, we prepared a set of 2D numerical phantoms by segmenting real abdominal CT volumes. A polyenergetic forward projection model[13] is leveraged to generate CT projection data of these 2D phantoms. After projections acquired, CT images are reconstructed using the conventional FBP reconstruction method. The groundtruths are the corresponding monoenergetic attenuation maps of the phantoms at the mean energy of incidented X-ray spectrum. The residual reference labels are thus set as the difference between FBP images and groundtruths. Using this data preparation work flow, we obtained simulated CT images comtaminated by various artifacts, such as metal artifacts, low-dose artifacts, and few-view artifacts. Specifically, metal artifacts are simulated by implanting high attenuation objects, e.g. Titanium of diverse shapes into random positions of numerical phantoms. Circular scanning geometry is employed and configured as 1200mm source-to-detector distance and 600mm source-to-axis distance. A flat 1-D detector is configured to have 600 bins of 0.2774mm size. The projections are acquired using 10^5 incident photon counts with the X-ray spectrum of 120kVp with 2mm Al and 0.2mm filtration. Angular coverage is $360°$ at $1°$ interval. Thus, a reasonable size of dataset for metal artifact reduction ([MAR]) is built. Low-dose([LD]) artifacts are simulated based on [MAR] datasets by lowering the kVp and mAs settings to 80kVp and 10^4 incident photon flux; Few-view([FV]) artifacts are