Analysis of Deep Convolutional Neural Network for Inverse Problems in Imaging

Kutay Ugurlu

Contents

List of Figures

List of Tables

I. Introduction

In this term project report, the study *Analysis of Deep Convolutional Neural Network for Inverse Problems in Imaging* by Jin *et al.* [1] is going to be reviewed, the results are going to reproduced and the experiments are going to be conducted using another dataset which is going to be formed by hand.

The authors of the paper propose a convolutional neural network architecture mainly for the inverse problem of computerized tomography(CT) problems, while they stress that the proposed method is available for all normal-convolutional inverse problems. The study aims reconstructing the image from lower views by direct inversion followed by a convolutional neural network, more specifically a U-Net [2] based architecture. Starting from the observation that the normal operator H^*H, where H^* is the adjoint operator of H that satisfies the relation $\langle f, H^*g \rangle = \langle Hf, g \rangle$, that appears as a forward model in a set of inverse problems, the authors investigate the relationship between the CNN models and the iterative optimization models.

II. Theory

A. Shift-invariant Normal Operator

To define a shift-invariant normal operator, the authors made the following definitions for the continuous domain:

1) *Isometry:* An isometry T is a linear operator such that $T^*Tf(x) = f(x)$
2) *Multiplication:* A multiplication M_m is a linear operator such that $M_m f(x) = m(x)f(x)$ where $m(x)$ is a continuous and bounded function.
3) *Convolution:* A convolution H_h is a linear operator such that $H_h f = \mathcal{F}^* M_{\hat{h}} \mathcal{F} f$ where $\mathcal{F}$ is the Fourier transform and $\hat{h}$ is the Fourier transform of h.
4) *Reversible Change of Variables:* A reversible change of variables $\Phi_\psi : L_2(\Omega_1) \to L_2(\Omega_2)$ is a linear operator such that $\Phi_\phi f = f(\phi(.))$ for some $\phi : \Omega_2 \to \Omega_1$ and such that its inverse exists. That is, it is a linear operator that represents the same function via a reversible domain transformation, such as Cartesian to Polar coordinate transformation.

Theorem 1. *If an operator is in the form $H = TM_m\Phi_\phi^{-1}\mathcal{F}$, then normal operator H^*H represents a convolution operation with $|detJ_\phi|M_{\Phi_\phi|m|^2}$*

where J_ϕ is the Jacobian matrix.

B. Forward Model of X-Ray CT Problem

The main focus of the paper is reconstructing the X-ray CT image from low-view projections. When we consider the forward model as the 2 dimensional X-Ray Transform $R : L_2()\mathbb{R}^2) \to L_2([0, 2\pi) \times \mathbb{R})$, also known as Radon Transform, we can express it using Fourier Slice Theorem as follows:

$$R = T\Phi_\phi^{-1}\mathcal{F} \tag{1}$$

where the coordinate transformation operation is defined as conversion from Cartesian to polar coordinates and T is inverse Fourier transform. The operation expressed by Eqn. 1 takes the Fourier transform of the image, express the resultant signal in polar coordinates and takes the inverse Fourier transform. This process is illustrated in Figure 1. Theorem 1 states that R^*R is a convolution with $\frac{1}{||w||}$ where w is the frequency variable for the 2D Fourier transform.

C. Direct Inversion

One may apply direct inversion methods to solve the inverse problem of tomography, *i.e.*, obtaining the image from measurements. The problem can be formulated as $g = Hf$ where H is a normal convolutional operator. According to [1], the solution can be obtained by direction inversion in two ways as follows:

$$f = W_hH^*g \tag{2}$$
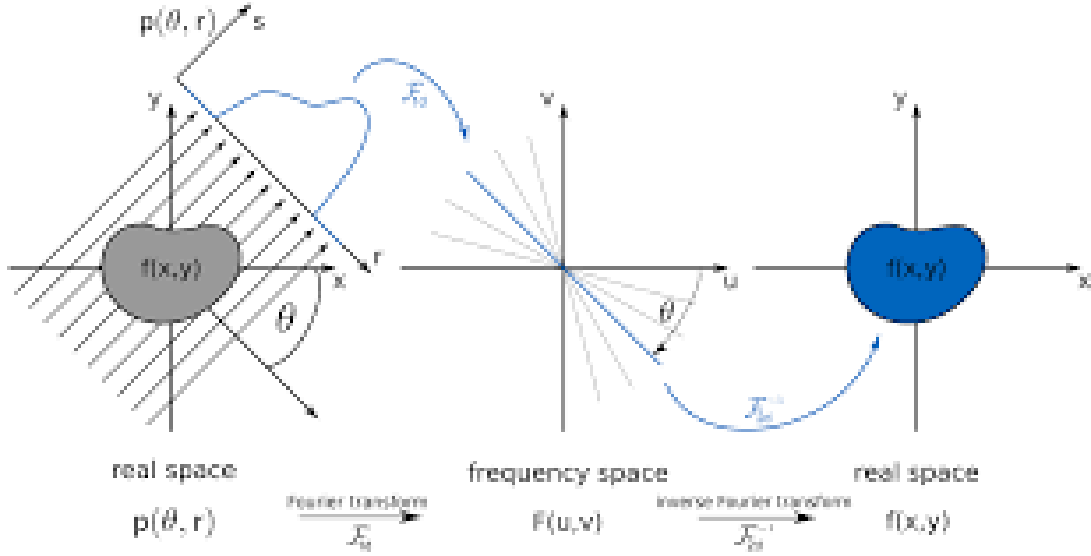
$$f = H^*TM_hT^*g \tag{3}$$

Figure 1: Illustration of reconstruction via Radon Transform [3]

where W_h in Equation (2) is the convolution operator with $\frac{1}{|det J_\phi |\Phi_\phi| m(w)|^2}$ and M_h is $\frac{1}{|det J_\phi||m(w)|^2}$. The physical interpretation behind these equations can be considered as follows: The operation in Equation (2) corresponds to the deconvolution in the reconstruction space, whereas Equation (3) corresponds to a filtering operation in Fourier domain followed by a back- projection, if T is Fourier transform.

D. Iterative Inversion

There also exist iterative approaches to solving inverse problems and they are formulated as follows:

$$\underset{a}{\mathrm{argmin}} \, ||y - HWa||_2^2 + \lambda ||a||_1 \tag{4}$$

where $x = Wa$ and a is a sparse representation of x in a domain where x is transformed to by W.

III. Proposed Method: FBPConvNet

The authors state that the filtering and pointwise linearity operations in the iterative reconstruction methods may suggest that CNNs may be a good fit for the solution of the inverse problems.

The approach simply consists of applying a direct inversion method to the measurements (sinograms in CT case) by Matlab's `iradon` function and feed the resultant projections to the CNN to train the network to learn the mapping between the back projected measurements and suitable ground truth images. The authors explain the reason for using a direct inversion method, instead of following an end-to-end neural net training approach between measurements and ground truth images as follows: By projecting the measurements to image domain, the learning is greatly simplified due to the fact that the network does not have to learn the coordinate transformation between Cartesian and Polar space. Since, filtered back projection encapsulates this operation and the physical information about the forward model.

A. Neural Network Model

The design of the proposed network was based on U-Net. The modified architecture is illustrated in Figure 2. The authors summarized the reasons for basing their architecture on U-Net as follows:

- *Multilevel Decomposition:* The use of pooling kernels with different sizes help the network effectively invert H^*H.
- *Multichannel filtering:* The convolutional kernels have the ability to make arbitrary combinations of channels, analogous to wavelet subbands in ISTA or split variables in ADMM.
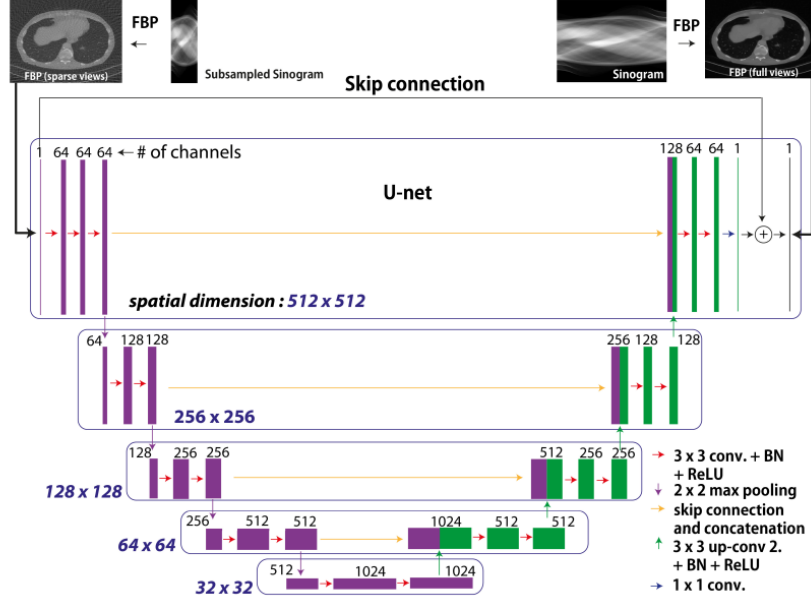
Figure 2: Modified U-Net Architecture

- *Residual Connection:* In addition to the original U-Net work, the authors added a skip connection between the input and output, allowing the network to learn the difference between the input and output. These connections have significant improvement in overcoming the gradient vanishing problem.
- *Last Layer:* The last layer of the network is replaced with a convolution layer, since the original U-Net implementation has 2 channels: foreground and background.

B. Experiments

Jin *et al.* begin the reconstruction via full view sinogram from both real and simulated data. Then the reconstruction is performed on the subsampled sinogram. This type of procedure holds an important place in human imaging because lowering the views lower the radiation dose received by the patient. The full-view filtered back projection data is used as ground truth images, rather than the actual image, since the authors believe that this is a more realistic setting in practice and the oracle information will never be available in CT.

C. Training Procedure

The sample size of the train and test splits of the datasets are given in Table I.

Dataset \ Split	Training Data	Test Data	Total
Synthetic (Ellipsoidal)	475	25	500
Real (Biomedical)	327	25	352

Table I: Samples for Datasets

The CNN part of the network is trained with pairs of low view and full view FBP images. To reduce the overfitting, data augmentation is realized by mirroring the images in both vertical and horizontal directions. To prevent the cost function from diverging, the gradient clipping procedure is applied. The hyperparameters for state-of-the-art TV reconstruction [4] for comparison study is set experimentally by hand.

IV. Results and Discussion

A. Reconstruction Performance Metrics

The authors use SNR as a quantitative metric. The SNR of the reconstruction $\hat{x}$ with the ground truth image x is given by Equation (5).

$$SNR = \max_{a,b \in \mathbb{R}} \frac{||x||_2}{||x - a\hat{x} + b||_2} \tag{5}$$
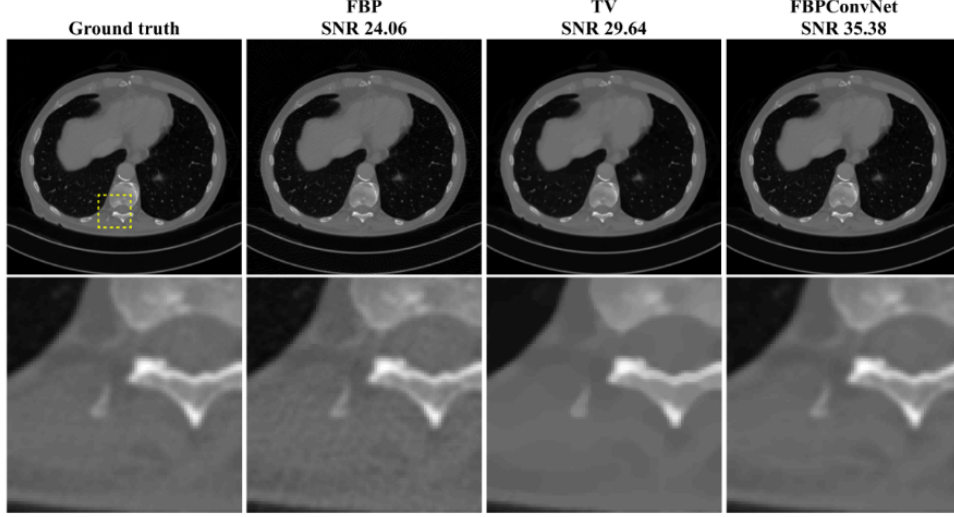
B. Sample Reconstructions

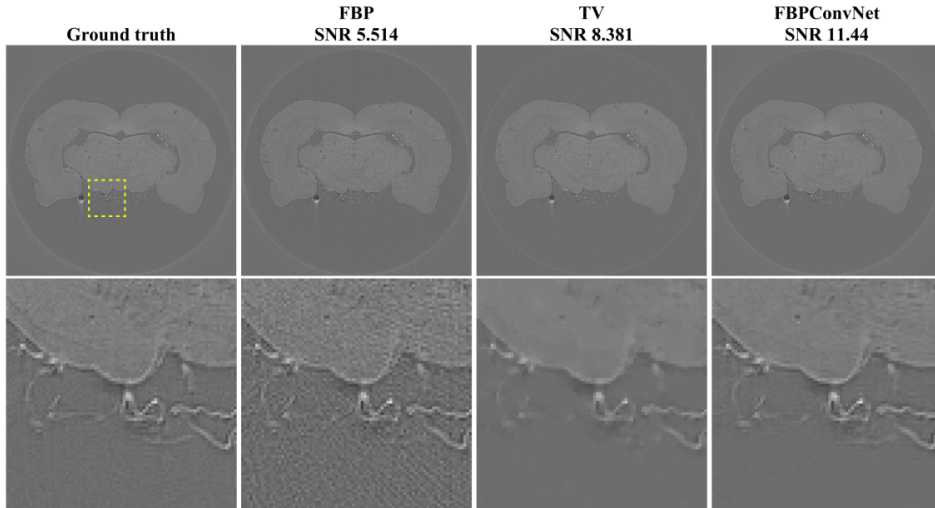Figure 3: Reconstructed images of biomedical dataset from 143 views

Figure 4: Reconstructed images of experimental dataset from 145 views

The results of the experiments showed that the choice of CNN for the set of inverse problems where the forward model is convolution is suitable. In both Figure 3 and Figure 4 the denoising performance of the FBPConvNet against back projection can be observed qualitatively. The proposed method yields compelling results in both real and synthetic data. It outperformed when the state-of-the-art iterative reconstruction methods. Once the training is finished, the computation time for an image is computed to be under a second.

References

[1] Kyong Hwan Jin et al. "Deep Convolutional Neural Network for Inverse Problems in Imaging". In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522. DOI: 10.1109/TIP.2017.2713099.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention.* Springer. 2015, pp. 234–241.

[3] *Instruction for the advanced lab course (FOPRA).* URL: https://www.ph.tum.de/academics/org/labs/fopra/?language=en.

[4] Michael T McCann et al. "Fast 3D reconstruction method for differential phase contrast X-ray CT". In: *Optics express* 24.13 (2016), pp. 14564–14581.