Low Dose CT Enhancement Using Deep Learning Method

by

Boyang Pan

Graduate Program in Medical Physics
Duke Kunshan and Duke University

Date: _______________________

Approved:

___________________________
Lei Ren, Advisor

___________________________
Zhiheng Wang

___________________________
Bowsher James

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Graduate Program of Medical Physics in the Graduate School of Duke Kunshan and Duke University

2021
ABSTRACT

Low Dose CT Enhancement Using Deep Learning Method

by

Boyang Pan

Graduate Program in Medical Physics
Duke Kunshan and Duke University

Date: _______________________
Approved:

___________________________
Lei Ren, Advisor

___________________________
Zhiheng Wang

___________________________
Bowsher James

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Graduate Program of Medical Physics in the Graduate School of Duke Kunshan and Duke University

2021
Abstract

Purpose:

Deep learning has been widely applied in traditional medical imaging tasks like segmentation and registration. Some fundamental convolution neural network (CNN) based deep learning methods have shown great potential in low dose CT (LDCT) enhancement. This study applied U-Net++ model to enhance CT images with low dose and compared the performance of U-net++ and U-net quantitatively and qualitatively.

Method:

30 patient CT images were chosen as the ground truth in the training process. Under-sampled projections were simulated from the ground truth volumes with a uniform distribution. LDCT was then reconstructed from the under-sampled projections using the ASD-POCS TV algorithm with 40 iterations and was treated as the input of the models. The U-net++ model was improved based on U-net model by connecting the decoders, reserving better dense feature along skip connections. Deep supervision (DS) was used to make a combined loss between each upper node and the ground truth to enhance the image feature preserving capacity. U-net was used as standard model for comparison. L1 loss and structure similarity (SSIM) loss were used in different attempts. The generated images were compared quantitatively using SSIM and peak signal-to-noise ratio (PSNR).

Results:

Both models succeeded to improve the quality of the low dose CT images. The U-net++ model trained with MSE loss had best average PSNR of 17.8 on the test dataset and average SSIM of 0.779 in terms of the whole images compared with the
original under-sampled LDCT with SSIM of 0.532 and PSNR of 16.7. U-net model trained using L1 loss had the best average SSIM of 0.756 and the average PSNR of 17.5. Conclusion: Deep learning method showed its potential dealing with the high dose caused by modern CT technique. Different CNN models could influence the quality of the generated image on different evaluation criterions.
2.2 Network Architecture ....................................................................................29
2.3 Network Training .........................................................................................31
2.4 Evaluation Metrics ......................................................................................32
3. Results ...........................................................................................................33
  3.1 Feasibility Study .........................................................................................33
  3.2 Model Evaluation .......................................................................................36
4. Discussion ......................................................................................................45
5. Conclusion .....................................................................................................47
References .........................................................................................................48
List of Tables

Table 1 Typical Organ Radiation Doses from Various Radiologic Studies .....................15
Table 2 SSIM comparison result ...............................................................................43
Table 3 PSNR comparison result .............................................................................44
List of Figures

Figure 1 An example of three-layer neural network ....................................................18
Figure 2 Sigmoid function .........................................................................................20
Figure 3 ReLU(x) = max(0, x) ..............................................................................20
Figure 4 An example of imaging convolution ...........................................................23
Figure 5 A demonstration of padding process...........................................................23
Figure 6 Max pooling layer with size of 2 and stride of 2 .........................................25
Figure 7 Average pooling layer with size of 2 and stride of 2 .................................25
Figure 8 Demonstration of U-net Architecture .......................................................27
Figure 9 Demonstration of U-net++ architecture. Each node in the graph rep- resents a convolution block..................................................30
Figure 10 L1 loss of one input image decreased from 323.3 to 2.9 in the training process ........................................................................................................33
Figure 11 Comparison of LDCT, DL-enhanced LDCT, NDCT. .................................35
Figure 12 Comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT. ..............................................................................37
Figure 13 Bone structure comparison of LDCT, DL-enhanced LDCT, NDCT..............38
Figure 14 Lung structure comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT. .................................................................39
Figure 15 Soft tissue comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT. .................................................................40
Figure 16 Histogram comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT. .................................................................41
Acknowledgements

First, I’d like to express my gratitude to my supervisor, Prof. Lei Ren, and committee members, Prof. James Bowsher, and Prof. Zhiheng Wang. I would also appreciate the warm concern from Prof. David Huang.

2020 and 2021 are tough years among my limited life for the COVID-19 pandemic. It’s a pity we could not go to Durham for graduate study. During this hard period, thank my parents for their meticulous care. Thanks for the company of my classmates and friends. I will miss the happy days we stayed together. Finally, I would appreciate the help from my supervisor, Dr. Nanjie Gong, and the instruction from Dr. Lei Xiang during my internship in Subtle Medical.
1. Introduction

1.1 Computed Tomography

1.1.1 Medical Imaging

Shortly after the discovery of X-ray by Wilhelm Conrad Rontgen in 1895\cite{1}, X-ray was used for body examination and disease diagnosis. That was the beginning of diagnostic radiology and the foundation of medical imaging.

In 1950s, ultrasound and radionuclide were used to imaging the human body leading to the development of ultrasonography and $\gamma$—scintigraphy. Then in 1970s and 1980s, X-ray computed tomography, magnetic resonance imaging and emission computed tomography were invented and applied to the medical examination. Though these imaging technologies base on the different imaging principles, all imaging modalities use a kind of energy source to disturb the human body and receive the response signals that reveal the human tissue properties and form the imaging contrast. Computed tomography (CT) displays perspective images that separate different tissues and bones which could not be obtained by the human eyes. Magnetic resonance imaging (MRI) provides outstanding anatomic images and exhibits different contrast for different tissues according to the imaging sequences. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) show the biologic information of human body.

Nowadays, medical imaging plays a more and more important role in modern healthcare system, such as diagnose, therapy and monitor. Imaging modalities were
deeply integrated to the modern digital system and workflow. Quantitative analysis makes it possible to reveal the deeper relationship between the imaging abnormality and pathology, thus provide a solid foundation for diagnosis.

1.1.2 Development of Computed tomography (CT)

Introduced by Geoffrey Hounsfield in 1972, the first computed tomography scanner was designed to make brain images in 80x80 array of 2.4-mm pixels[2]. Though the image quality looks poor in today’s criterion, the first generation of CT scanner helped physicians see the anatomy of human body that did not require open surgery.

The second generation of CT developed fast with the success of the first generation[3]. Multiple detectors were introduced to the scanner system and a narrow fan beam replaced the original pencil beam geometry. This system could acquire more data in shorter time and generate a 160x160 matrix image at a time.

However, the scan speed remained slow even with the increase of detector number. To speed up the CT examination, a rotating gantry was built. It enabled the rotate-rotate geometry while remain the fixed relative position of the x-ray tube and detector array. In this geometry, if the x-ray source to isocenter distance is L and the swept fan angle of the detector array is α, we can get the diameter of the FOV

\[
\text{FOV} = 2L \sin(\alpha/2)
\]  

In the clinical practice, the potential ring artifact lead by the rotate-rotate geometry was found. To address this issue, the fourth-generation CT scanner was designed, which consisted of a rotating x-ray tube and the entire stationary detector array. Hence, this system had a rotate-stationary geometry.
With the improvement of high-resolution CT images, most structural lesions could be easily detected nowadays, such as tumor, edema and aberration.

1.1.3 Principle of CT Imaging

A series of rays that are emitted from x-ray tube, pass through the patient at the same orientation and measured by detectors to form the signal is called a projection or a view. CT images are reconstructed from projections that cover the complete volume of scanned object using mathematical algorithms.

Traditional straightforward image reconstruction algorithms include simple back-projection, Fourier reconstruction, filtered back projection and convolution back projection. However, simple back projection algorithms are not accurate for image reconstruction while the latter ones have slow computation speed. An advanced reconstruction algorithm is called iteration reconstruction. Iteration reconstruction has a small amount of calculation in each iteration and the image quality gradually approach to the original one by iterations.

1.1.4 Artifacts of CT Images

Image artifact is a structure or pattern within an image which does not exist in original object. CT artifacts could come from system design, scanner components or could be induced from patient or helical and multi-slices.

- Beam-hardening Artifact

If a polychromatic x-ray beam is used as the imaging source and go through an object, the outcome beam would have a higher average energy due to the different
attenuation effect. Different material could lead to different beam hardening effect. And the longer the penetration range, the greater the impact. The common beam hardening appearance includes shading, streaks and cupping.

- Aliasing Artifact

According to Shannon sampling criterion, the raw data must be sampled at a rate of at least twice the highest spatial frequency which contained in the signal to avoid aliasing. If the data is sampled under the required rate, the reconstructed image shows the overlap in structures.

- Partial Volume Artifact

The HU value of voxels represents the attenuation coefficient of the tissue in the volume. If a voxel contains several materials, the detection process will average the beam attenuation effect. However, if a high-density organ is mixed with a low-density organ in one voxel, the attenuation performance varies greatly depending on the beam angle which leads to the partial volume artifact. Partial volume artifact arises mostly from reconstructing low resolution CT images, and substantially reduced with the development of modern CT examinations.

1.1.5 Organ Dose in CT

Dose absorbed in the CT examination is commonly computed using table lookup. However, the normalization process is not the same with radiography. The computed tomography dose index (CTDI) was originally designed as an index[4]. With the enhancements and modifications over the years, CTDI concept have attempted to
make it a more accurate patient dosimetry method. Now, CTDI-based dosimetry is the worldwide standard for estimation of patient dose in CT.

### 1.1.6 Low Dose CT

Organ doses from CT examination are considerably greater than those from other traditional imaging modalities like conventional anterior-posterior x-ray. For example, a conventional lateral chest radiography result in 0.15mGy dose absorbed in lung, while abdominal CT usually gives 10mGy dose to the stomach which is about 70 times greater than the former as shown in Table 1[5].

According to a survey published in 2013, the ownership of CT scanner per million people was 32.2 in the United State and 92.6 in Japan. Also, with the evolution of CT scanner, time spent on the examination shortened greatly during the past thirty years. All these facts lead to the rapid growth of the annual number of CT scans. With the fast development of CT scanner and the worldwide use of CT examination, the dose accumulated in the patient cannot be ignored. It’s estimated that CT scans might be responsible for up to 2% of cancers in the U.S., and CT scans performed in each year would cause 29,000 new cancer cases in the future due to ionizing radiation exposures to patients.

**Table 1 Typical Organ Radiation Doses from Various Radiologic Studies**

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Relevant Organ</th>
<th>Relevant Organ Dose (mGy or mSv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental radiography</td>
<td>Brain</td>
<td>0.005</td>
</tr>
</tbody>
</table>
The use of low-dose CT is a possible solution to reduce the impact of dose induced by CT examination. Since the radiation doses from CT depend on several factors like the number of scans, the tube current, scanning time, the size of patient and the scan pitch, the impact can be controlled by the radiologist. However, it’s always the case that the noise level in CT images will increase as the radiation dose decreases. The tradeoff between the need of low-noise CT images and the desirability of low doses in examination requires more consideration.

To solve the problem of high noise present in low-dose CT, various iterative reconstruction (IR) techniques were developed. But due to the nature of IR algorithm, radiation dose can be reduced by 17-44% which is not sufficient for the increasing demand for CT examination. Furthermore, IR algorithm have a high requirement for computing source to meet the daily clinical needs. As an example, GE Veo takes 30-45min to reconstruct one set of raw data on a specialized massively parallel computer of 112 CPU cores. Recent study has found the potential of deep learning method in

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Anatomy</th>
<th>Dose (mSv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior–anterior chest radiography</td>
<td>Lung</td>
<td>0.01</td>
</tr>
<tr>
<td>Lateral chest radiography</td>
<td>Lung</td>
<td>0.15</td>
</tr>
<tr>
<td>Screening mammography</td>
<td>Breast</td>
<td>3</td>
</tr>
<tr>
<td>Adult abdominal CT</td>
<td>Stomach</td>
<td>10</td>
</tr>
<tr>
<td>Barium enema</td>
<td>Colon</td>
<td>15</td>
</tr>
<tr>
<td>Neonatal abdominal CT</td>
<td>Stomach</td>
<td>20</td>
</tr>
</tbody>
</table>
synthesizing normal dose CT images from low dose CT especially in ultra-low-dose CT recovery.

With the concerns for the high dose retained in a patient by CT examinations, the American Association of Physicists in Medicine (AAPM) and National Institutes of Health (NIH) conducted the 2016 Low Dose CT Grand Challenge[6]. Over 500 groups of researchers from all over the world have requested the access to the abdominal low dose CT dataset. Multiple methods were proposed to reconstruct the image from projection data or reduce the noise in the images. However, the size of the dataset is restricted which consist of projection data and corresponding images from 30 patients. And due to the proprietary information and formatting of manufacturer-specific projection data files, the access to clinical CT projection data has been extremely limited.

### 1.2 Deep Learning

#### 1.2.1 Introduction of Neural Network

Neural network is an old algorithm, which is originally designed to simulate the operation mode of human brain[7]. Brain can finish all kinds of tasks, like calculation, language understanding, object identification et al. If we expect the computer to accomplish the same missions, the easiest way to think about is letting the computer work just like the brain. As a part of artificial intelligence, brain simulation is always a fascinating subject and give enlightenments to the most advanced machine learning algorithms.

Thanks to the development of neurology, now we know the neuron is the most basic elements in human intelligence. Each neuron consists of nucleus, dendrites and
axons. Dendrites receive input signals from outside environment and pass the signal to
the nucleus. nucleus can be treated as processing unit. It deals with the input signals.
Axons output the signal to the next neuron or the effectors. A complete reflecting arc is
composed of multiple neurons. To process different tasks, a lot of neurons form the
neural network. Further study shows that the architecture of neural network is
complicated. To understand the function of each neuron, a concise way is to treat the
different level of neurons as different layers, which illustrate the practical structure of
neural network algorithm.

1.2.2 neural Network: Representation

A three-layer neural network can be designed as Figure 1.

\( x_1, x_2, x_3 \) in layer 1 are the input units, which represent the input data. \( a_1, a_2, a_3 \)
in layer 2 are the middle units. They are responsible for dealing with the input data and
sent the processed data to the next layer. In each calculation unit, different input can be
treated differently with isolated weights. Also, an activation function \( g(x) \) is always used
to restrict the output feature. Layer 3 is the output unit. It gives us the output of neural
network, here, \( h_\Theta(x) \). An extra element would be added to each layer as a bias unit.

Figure 1 An example of three-layer neural network
Sigmoid activation function (shown in Figure 2) and ReLU activation function (shown in Figure 3) are the most widely used activation functions. Take the above model as an example. We use \( a_i^j \) as the \( i \) element in the \( j \) layer, \( \theta_{ik}^j \) as the weight of the \( k \) element in \( j \) layer to form the \( i \) element in \( j+1 \) layer, \( g(x) \) as the activation function. Then we will have

\[
a_1^{(2)} = g\left(\theta_{10}^{(1)}x_0 + \theta_{11}^{(1)}x_1 + \theta_{12}^{(1)}x_2 + \theta_{13}^{(1)}x_3\right)
\]

\[
a_2^{(2)} = g\left(\theta_{20}^{(1)}x_0 + \theta_{21}^{(1)}x_1 + \theta_{22}^{(1)}x_2 + \theta_{23}^{(1)}x_3\right)
\]

\[
a_3^{(2)} = g\left(\theta_{30}^{(1)}x_0 + \theta_{31}^{(1)}x_1 + \theta_{32}^{(1)}x_2 + \theta_{33}^{(1)}x_3\right)
\]

And

\[
h_\theta(x) = g\left(\theta_{10}^{(2)}a_0^{(2)} + \theta_{11}^{(2)}a_1^{(2)} + \theta_{12}^{(2)}a_2^{(2)} + \theta_{13}^{(2)}a_3^{(2)}\right)
\]

The algorithm representing the layer from left to right is called forward propagation. If we use matrix to express \( x, \Theta, a \), like,

\[
X = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}, \Theta = \begin{pmatrix} \theta_{10} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \theta_{33} \end{pmatrix}, a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}
\]

We have

\[
g(\Theta \cdot X) = a
\]

In this neural network, the relationship between input \( X \) and the output \( h \) can be expressed as

\[
h = F(X)
\]

where \( F(X) \) represent the neural network.
Figure 2 Sigmoid function

\[ \sigma(x) = \frac{1}{1 + e^x} \]

Figure 3 ReLU(x) = \text{max}(0, x)
1.2.3 Neural Network: Learning

In the model shown in Figure 1, the output $h_\Theta(x)$ is determined for each input $X$, if the network parameters $\Theta$ is fixed. To make the model implement the function exactly desired, $\Theta$ need to be chosen properly.

Let $\theta_0$ be the parameters we want, and $\theta_1$ be the parameters we have. The cost function is defined as

$$ J = J_X(\theta_0, \theta_1) $$

As an example, the sum of square error can be expressed as

$$ J_X(\theta_0 - \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (F_{\theta_0}(x^i) - F_{\theta_1}(x^i))^2 $$

where $m$ is the input data number. The learning process of the neural network is defined as for a set of input $X$ and corresponding output $h$, minimizing the cost function. One popular way to minimize the cost function is gradient descent algorithm, which can be defined as

$$ \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) $$

For multi-layer neural network, the partial derivative calculating process for $\Theta$ in the first several layers can be complicated. Backpropagation algorithm is designed for such purpose, which is to calculate the derivative in the last layer, and then, one layer by another, till the layer we want.

The greatest challenge in neural network training process is overfitting. It occurs when the number of input data is less than the number of parameters in the neural network. In this case, the network memorizes the input and output pair rather than
learning the feature of the dataset. To solve this problem, a big dataset for training is necessary.

1.2.4 Convolution Neural Network (CNN)

In the advanced neural network, the form of processing unit various, which means $\Theta \cdot X$ in forward propagation could be replaced with $P(X)$. The most famous and widely used structure to deal with imaging input is convolution. The core idea behind the convolution is that instead of processing the whole input layer with a large and dense linear connection from every input to every output, a small linear layer can save the calculating resource. The benefit of small layer is to focus on the feature of small field and search the translation invariance of input information.

With the convolution layer, the number of parameters in one layer reduces from the number of input data to a small convolution kernel. The formula of convolution can be written as

$$P(X) = \int_{-\infty}^{+\infty} f(\tau) g(x - \tau) d\tau$$  \hspace{1cm} (1.12)

where $f(x)$ is the input, $g(x)$ is the filter function. Figure 4 shows an example when the input of the convolution layer is two dimensional, like gray scale image.
In this case, the input is a 5x5 image and the filter is 3x3 large. So, we get the output image of 3x3 large. To keep the size of output image the same as the input one, we need to add padding to the edge of the input image. Figure 5 shows an example of padding one zeros.
For an input image of $m \times n$, a filter of $k \times l$, with filter stride $s$ and padding $p$, the output image size $x \times y$ would be

$$x = (m - k + p)/s$$  \hspace{1cm} (1.13)

$$y = (n - l + p)/s$$  \hspace{1cm} (1.14)

The size of the filter various from 3 to 11 and always take odd numbers depending on the feature of input.

### 1.2.5 Pooling

The convolution layer is very effective in creating the feature map of input images. However, the size of convolution kernel limits its ability of local information acquisition for adjacent pixels in images may contain the similar structural features. To condense the information stored in the adjacent pixels, The pooling layer is a useful non-linear layer which can reduce the size of feature maps without introducing new parameters. There are two commonly used pooling layers, the first is the max pooling (shown in Figure 6), and the second one is the average pooling (shown in Figure 7).

Average pooling calculates the average value for each patch on the feature while max pooling uses the maximum value. The size and the stride of pooling layer can be adapted to change the compression level. The most widely used pooling layer would be the size of two and the stride of two.
The inverse operation of pooling is up-sampling, and the corresponding layer is up-sampling layer. Up-sampling layers translate from coarse salient features to a more dense and detailed output. A simple version of up-sampling could just repeat the rows and columns of the input. More precise model uses bilinear interpolation to predict the value of synthetic voxels.

With the deepening of network architecture, the performance of the network gets better at first but goes down when layer number is too big. The reason behind this phenomenon is the loss of original information as the layer goes deeper. Deep residual
network [8] avoid the degradation problem by utilizing the skip connection or concatenate connection.

The skip connection commonly skips over one or two layers and can be presented as

$$X^l = g(\theta^{l-1} \times X^{l-1} + b^l + \theta^{l-2} \times X^{l-2})$$  \hspace{1cm} (1.15)

The concatenate connection just adds former layer to the extra channels of the latter layer if the size of two layers is equal. After concatenate connection, the information of former layer remains complete, but the calculation burden increases.

1.2.6 Deep Supervision

Proposed by Lee et al. [9], deep supervision was first designed as a solution to deal with the problem of gradient vanishing in CNN training process. This method gives weights to each hidden layer on the final loss function. The size difference between hidden layer and output layer can be solved by introducing a simple convolution layer. With deep supervision, the robust features of early layers learned fast.

1.2.7 U-net

In the early days of deep learning, multiple structures were designed to meet the need of different tasks. An encoder-decoder like structure was found effective in imaging classification and segmentation, which inspired the invention of U-net[10]. U-net is a 'U'-shape architecture based on the encoder-decoder design which is shown in Figure 8. It has four 1/2 pooling layers and symmetrical up-sampling layers. Each pooling layer and corresponding up-sampling layer was followed by a residual block and connected by a skip connection. It can effectively extract high dimensional characteristics of the input image and feedback on the lower-level images.
Since the first day of U-net, its excellent performance on multiple imaging tasks like object detection and structure segmentation while saving the calculation resource had made it popular among researchers. Now, U-net has been the baseline model in image semantic segmentation task especially for medical images. A great number of advanced models have been proposed based on the U-net like Res-Unet, U2-net[11], U-net++, Attention-Unet[12].

1.2.8 Research Aim

Inspired by the successful practice of deep learning method in medical imaging enhancement for different imaging modalities, the aim of this research was to investigate
the feasibility of using CNNs to enhance the quality of low dose CT images. We were also curious about the performance of different architectures like U-net and U-net++ on this target. The influence of hyper-parameters was investigated by setting the different convolution kernels and activation functions during the process of building neural networks[13]. The imaging quality was assessed both qualitatively and quantitatively.
2 Methods and Materials

2.1 Dataset

In this study, we downloaded the public low dose CT data from The Cancer Imaging Archive which includes 299 groups of CT images and corresponding projection data stored in DICOM format[14]. The available dataset was collected from either a GE Discovery CT750i from GE healthcare, SOMATOM Definition AS+ or SOMATOM Definition Flash dual-source CT system from Siemens Healthiness. 100 groups of chest CT cases, 99 groups of head CT cases and 100 groups of abdomen CT cases were included in the dataset. Approximately 1/2 of the cases was diagnosed as negative. Each case consists of CT images, projection data and clinical diagnostic report. Low dose projection data were generated by adding noise into those of full dose to simulate a low dose scan.

50 cases were chosen for this research where 30 were used for training the network, 10 were used as validation and 10 were used for testing. Limited by the research resources, this study investigates the relationship between the low dose CT images and the full dose CT images.

2.2 Network Architecture

In this study, U-net and U-net++ were built to enhance the low dose CT images. The architecture of U-net was shown in Figure 8. Four pooling layers and four up-sampling layers were used to extract the high-level information of input images. Two convolution layers and ReLU layers was adjacent to each pooling layer and up-sampling layer. The size of the convolution kernel was set as 3x3. The padding of each convolution was set as 1 with stride of 1 to keep the size of input images.
The U-net++ architecture was proposed by Zhou et al. [15] and is shown in Figure 9. This architecture adds nodes in varying depths and dense connects each node in the same depth. Each node also acquires the features from the previous lower node. This architecture introduces deep supervision to accelerate the learning process and gains the stable features.

To acquire the best model for low dose CT enhancement, different hyper-parameters were used to train the U-net++ network. The convolution layer was separately set as 3x3 kernel with padding of 1 and stride of 1 or 5x5 kernel with padding of 2 and stride of 1. ReLU function and Leak-ReLU function were used as different activation functions. The influence of different types of polling layer was also assessed.

![Figure 9 Demonstration of U-net++ architecture. Each node in the graph represents a convolution block.](image-url)
2.3 Network Training

Typical loss functions include L1-loss, L2-loss and structural similarity (SSIM) loss. L1-loss is defined as

\[
Loss_{l1} = \frac{1}{N} \sum_{i=1}^{N} |F(x_i) - y_i|
\]  

(2.1)

where \(x_i, y_i\) represent the pixel value in the input and reference CT images with \(N\) pixels. L2-loss is also called MSE loss, and is defined as

\[
Loss_{l2} = \frac{1}{N} \sum_{i=1}^{N} [F(x_i) - y_i]^2
\]  

(2.2)

SSIM is an index describing the similarity of two structures. It’s defined as

\[
SSIM = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)}
\]  

(2.3)

where \(\mu_x\) and \(\mu_y\) are the average pixel value in the input and reference patch, \(\sigma_x\) and \(\sigma_y\) are the covariance of pixel value respectively. \(C1\) and \(C2\) is small numbers to prevent the numerator equals to zero.

SSIM loss is the difference of structural similarity in input and reference images, and is defined as

\[
Loss_{ssim} = \frac{1}{M} \sum_{i=1}^{M} (1 - SSIM_i)
\]  

(2.4)

where \(M\) is the number of input patch.
2.4 Evaluation Metrics

The CT images enhanced by the U-net and U-net++ were evaluated qualitatively and quantitatively with the comparison to input low dose CT images and the ground truth full dose CT images.

For qualitative evaluation, presentation and sharpness of different tissues was assessed. Histogram map was used to evaluate voxel-wise accuracy quantitatively. Overall image quality was assessed by the SSIM and peak signal to noise ratio (PSNR) within the regions of interest. PSNR is defined as

$$PSNR = 10 \times \log_{10}\left(\frac{MaxPixelValue}{RMSE}\right)$$

(2.5)

And RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [F(x_i) - y_i]^2}$$

(2.6)
3. Results

3.1 Feasibility Study

We used a basic U-net to assess the feasibility of the low dose CT enhancement using deep learning method. The kernel size of each convolution layer was set to 3 with padding of 2 and stride of 1 to keep the size of each feature map. The feature channel was set to 8 of the first convolution layer. The L1 loss of the whole image and SGD optimizer with learning rate of 0.1 in the first 1500 epochs and 0.01 in the rest 500 epochs were used to training the network.

![Figure 10 L1 loss of one input image decreased from 323.3 to 2.9 in the training process](image)

The loss changed in the training process was shown in Figure 10. We could find the loss function dropped fast in the previous epochs and turned stable with the training process.

The comparison of low dose image, DL-enhanced image and full dose image was shown in Figure 11. (a, d, g) were low dose CT images. (b, e, h) were DL-enhanced
low dose CT images. (c, f, i) were normal dose CT images. (g, h, i) were the details show of figure (d, e, f)

Two representative slices of axial views from the test dataset were included to evaluate the performance of proposed deep learning method.
Figure 11 Comparison of LDCT, DL-enhanced LDCT, NDCT.
It can be concluded that the deep learning enhanced images had better image quality compared with the original low dose images especially in the contrast of bone structure and the presentation of small soft tissue structure. To be specific, bone structures look blurry in LDCT images, but have clearer boundary in NDCT images and DL-augmented images.

Small soft tissue structures cannot be distinguished from the background tissue in LDCT but exist in NDCT images. Deep learning method gives more contrast to small soft tissue and makes it recognizable.

3.2 Model Evaluation

The performance of deep learning method is greatly influenced by the network structure and the cost function. After the assessment of different hyper-parameters, we set the kernel size of each convolution layer to 3 with padding of 2 and stride of 1 to keep the size of each feature map. The feature channel was set to 8 of the first convolution layer and the max pooling layer was applied.

The result of U-net and U-net++ augmentation with voxel-wised l1 loss or structural-wise ssim loss for low dose CT images in patient study were shown in Figure 12. (a) was low dose CT image. (b) was U-net with l1 loss augmented image. (c) was U-net with ssim loss augmented image. (d) was U-net++ with l1 loss and deep supervision augmented image. (e) was U-net++ with l1 loss but without deep supervision augmented image. (f) was normal dose image.
Figure 12 Comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT.
Figure 13 Bone structure comparison of LDCT, DL-enhanced LDCT, NDCT.
Figure 14 Lung structure comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT.
Figure 15 Soft tissue comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT.
Figure 16 Histogram comparison of LDCT, DL-enhanced LDCT for different architectures and NDCT.
Bone structure comparison of different images was shown in Figure 13. (a) was low dose CT image. (b) was U-net with l1 loss augmented image. (c) was U-net with ssim loss augmented image. (d) was U-net++ with l1 loss and deep supervision augmented image. (e) was U-net++ with l1 loss but without deep supervision augmented image. (f) was normal dose image.

Lung structure comparison of different images was shown in Figure 14. (a) was low dose CT image. (b) was U-net with l1 loss augmented image. (c) was U-net with ssim loss augmented image. (d) was U-net++ with l1 loss and deep supervision augmented image. (e) was U-net++ with l1 loss but without deep supervision augmented image. (f) was normal dose image.

Soft tissue comparison of different images was shown in Figure 15. (a) was low dose CT image. (b) was U-net with l1 loss augmented image. (c) was U-net with ssim loss augmented image. (d) was U-net++ with l1 loss and deep supervision augmented image. (e) was U-net++ with l1 loss but without deep supervision augmented image. (f) was normal dose image.

The DL-enhanced CT images have higher image quality than LDCT images. The general image quality of U-net with l1 loss and U-net++ augmented images were close. U-net with ssim loss augmented images have more clear soft tissue structures but may be different from that of corresponding normal dose CT images.

The histogram of 3-D images was compared in Figure 16. (a) was low dose CT image. (b) was U-net with l1 loss augmented image. (c) was U-net with ssim loss augmented image. (d) was U-net++ with l1 loss and deep supervision augmented image.
(e) was U-net++ with l1 loss but without deep supervision augmented image. (f) was normal dose image.

Little difference was found between LDCT, U-net with l1 loss, U-net++ with deep supervision and the ground truth NDCT in the aspect of image intensity distribution and the position of the peaks. U-net with ssim loss and U-net++ without deep supervision caused great change of Hounsfield unit (HU) numbers in high intensity regions of the input images.

**Table 2 SSIM comparison result**

<table>
<thead>
<tr>
<th>ROI Region</th>
<th>LDCT</th>
<th>U-net(l1)</th>
<th>U-net++(DS)</th>
<th>U-net++(no-DS)</th>
<th>U-net(SSIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole image</td>
<td>0.532</td>
<td>0.756</td>
<td>0.779</td>
<td>0.711</td>
<td>0.529</td>
</tr>
<tr>
<td>tissue</td>
<td>0.791</td>
<td>0.791</td>
<td>0.771</td>
<td>0.784</td>
<td>0.653</td>
</tr>
<tr>
<td>tissue</td>
<td>0.876</td>
<td>0.856</td>
<td>0.846</td>
<td>0.851</td>
<td>0.753</td>
</tr>
<tr>
<td>lung</td>
<td>0.800</td>
<td>0.835</td>
<td>0.842</td>
<td>0.840</td>
<td>0.820</td>
</tr>
<tr>
<td>lung</td>
<td>0.722</td>
<td>0.769</td>
<td>0.773</td>
<td>0.776</td>
<td>0.714</td>
</tr>
<tr>
<td>bone</td>
<td>0.896</td>
<td>0.918</td>
<td>0.920</td>
<td>0.914</td>
<td>0.895</td>
</tr>
<tr>
<td>bone</td>
<td>0.885</td>
<td>0.908</td>
<td>0.911</td>
<td>0.908</td>
<td>0.876</td>
</tr>
<tr>
<td>bone</td>
<td>0.888</td>
<td>0.917</td>
<td>0.918</td>
<td>0.915</td>
<td>0.853</td>
</tr>
</tbody>
</table>

The quantitative comparison result was shown in Table 2 and Table 3. For the whole image, deep learning enhanced images had better SSIM and PSNR performance.
than original LDCT images. U-net++ trained with deep supervision had best SSIM and PSNR result compared with all other method. U-net trained using SSIM loss had worst quantitative image quality, which could be a result from no voxel-wise criterion for the training process. The deep learning methods had better performance in lung and bone area, while had little contribution to the soft tissue enhancement.

<table>
<thead>
<tr>
<th>ROI Region</th>
<th>LDCT</th>
<th>U-net(l1)</th>
<th>U-net++(DS)</th>
<th>U-net++ (no-DS)</th>
<th>U-net(SSIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole image</td>
<td>16.78</td>
<td>17.48</td>
<td>17.80</td>
<td>16.15</td>
<td>11.01</td>
</tr>
<tr>
<td>tissue</td>
<td>10.61</td>
<td>16.28</td>
<td>11.41</td>
<td>15.28</td>
<td>6.485</td>
</tr>
<tr>
<td>tissue</td>
<td>18.30</td>
<td>18.50</td>
<td>15.77</td>
<td>16.44</td>
<td>9.256</td>
</tr>
<tr>
<td>lung</td>
<td>14.57</td>
<td>14.82</td>
<td>15.01</td>
<td>14.40</td>
<td>11.71</td>
</tr>
<tr>
<td>lung</td>
<td>14.02</td>
<td>14.42</td>
<td>14.6</td>
<td>14.30</td>
<td>10.39</td>
</tr>
<tr>
<td>bone</td>
<td>17.12</td>
<td>17.20</td>
<td>17.66</td>
<td>17.82</td>
<td>11.94</td>
</tr>
<tr>
<td>bone</td>
<td>17.30</td>
<td>17.38</td>
<td>17.75</td>
<td>17.93</td>
<td>12.15</td>
</tr>
<tr>
<td>bone</td>
<td>16.57</td>
<td>15.12</td>
<td>16.00</td>
<td>15.80</td>
<td>10.16</td>
</tr>
</tbody>
</table>
4. Discussion

In this research, deep learning method was proposed to improve the quality of LDCT images. The performance of different types of network architecture and cost function were evaluated both qualitatively and quantitatively. The result proved the capacity of deep learning method in generating more precise CT images with low radiation dose within the examination process.

As shown in Table 2 and Table 3, U-net++ architecture achieved the best solution in this task. Compared with the training plan without deep supervision, U-net++ made a great progress in predicting the voxel value for bone structures. This ability was acquired at the price of approximately 1.5 times the training time and corresponding calculating resource. Considering the computational and time cost, U-net could still be the most economically available solution to the task.

We also found the capacity of deep learning method in analyzing small soft tissue structure is insufficient. The improvement in visual effect is not obvious compared with other tissues. This shortage may be caused by the low quality of training samples even on which the soft tissue contrast is not legible. We also notice the small structures on images synthesized by U-net trained using SSIM loss. These structures show the texture within bulk organizations but do not exist on full dose images. So, it’s hard to evaluate whether these textures exist. However, since the original full dose CT is the exact result, we want and these components are uncontrollable in the system, we would abandon these small structures to form a more robust function for clinical use.

Though the quantitative analysis of U-net trained using SSIM loss was unsatisfactory, the distinguishable soft tissue in CT images processed by SSIM loss U-
net could give further study a slight. A combined loss of SSIM and L1 may generate CT images more balanced on voxel-wise error and the clarity of tiny structures.

The possible future direction on this study could be applying deep learning method on clinic data, improving the network architecture for a better performance on soft tissue and use other organ images like head LDCT.
5. Conclusion

In this study, deep learning method demonstrated the capability of synthesizing full dose CT from low dose CT images. Networks performed well on bone structure enhancement but behaved mediocrely on soft tissue augmentation. U-net++ trained using L1 loss with deep supervision technique reached the best performance among different quantitative and qualitative evaluation metrics.
References


