Kill Two Birds With One Stone: Solving Deblurring And Super Resolution with One Neural Network

Your Exam Number

word count

Master of Science

Speech and Language Processing

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2022

Abstract

Computed tomography (CT) is a three-dimensional imaging modality that uses X-ray beams to produce cross-sectional information of an object or a patient. CT suffers from multiple inherent limitations of image quality. The purpose of this study is to propose three deep learning models to address two of the common limitations, namely image blurring and limited spatial resolution. Two of the models, residual dense network (RDN) and Sharp U-Net are first seen their application in these medical imaging tasks. Moreover, we incorporate the advantages of these two methods to design a novel neural network architecture named Sharp RDN. The performance of these methods was benchmarked on two data sets through both qualitative and quantitative comparisons. The data sets were AAPM simulated data and TCIA clinical data. The metrics for the quantitative measure were PSNR, SSIM, and RMSE. Experimental results showed that both RDN and Sharp RDN drastically improved the corrupted image quality by giving a 65% improvement in the deblurring task, and 32% improvement in the SR task. The performance of Sharp U-Net was poorer than RDN and Sharp RDN, indicating that U-Net-type architectures are not suitable for these image tasks.

Contents

Chapter 1

Introduction

1.1 Medical CT Imaging

Computed tomography (CT) was developed by British electrical engineer Godfrey N. Hounsfield and South African-American physicist Allan M. Cormack at EMI Laboratories in 1972. CT is a diagnostic imaging procedure that combines image formation using the interactions between X-rays and matter, data capture/conversion with high-end detector systems, and sophisticated algorithms to produce tomographic (cross-sectional) images. It has not only become one of the most popular ways for internal inspection of a patient's body, but also become one of the most popular medical imaging methods for screening, diagnosis, and image-guided intervention. More than 70 million CT scans are performed every year in the US alone, and 300 million CT scans worldwide[Hsieh, 2003].

The diagnostic reliability of CT images hinges on the degree to which the images retain fidelity to the features being scanned. High-quality images provide better visibility, anatomical details, and fidelity of radiomic features. However, CT image quality can be degraded by various hardware and software limitations. They create artifacts that can complicate or even preclude the use of the images[Ketcham and Hanna, 2014, Herman, 2009]. Common CT non-idealities fall into two categories - artifacts and detector limitations. Artifacts in radiology refer to abnormalities seen on an image that is not present in reality. Common artifacts include beam hardening, quantum, and electronic noise, scattering, focal spot, detector blurring, etc. The size and number of detector elements limit the spatial resolution of a CT image, which further determines the quality of an image and describes how detailed an object can be represented by the image[Ghekiere et al., 2017].

1.2 Non-idealities in CT

In this work, we focus on addressing two CT non-idealities - image blurring and limited spatial resolution - with deep learning methods.

1.2.1 Image blurring

Image blurring is mainly caused by the finite size of an X-ray focal spot and detector pixel. Ideally, the X-ray focal spot and detector pixel size should be infinitely small. However, under typical diagnostic CT conditions, the focal spot size is 0.5–2.0 mm, and the detector pixel size is 1.2 mm[Tilley et al., 2015]. At the edge of objects, a blurred margin, which is called a penumbra, will be produced because of the finite size of the focal spot. The result of x-rays will arrive from slightly different locations in the focal spot[Bivard et al., 2013]. The resultant loss of sharpness is called focal spot blur or geometric blur and results in limited spatial resolution[Hu et al., 2022].

1.2.2 Limitation of image spatial resolution

For CT, the detector pixel size will limit the spatial resolution, that is because it can not shows an object whose size is less than one pixel. The Physiological and pathological units in human bodies are on scales of 10 μm. However, the resolution of clinical CT systems is on a scale of 1mm. For example, CT angiography is inferior as regards spatial resolution for the detection of small vessels at the neck of an aneurysm, which is important for planning treatment[Dammert et al., 2004].

1.3 Image restoration methods

For better diagnostic value, image restoration is often needed in CT. Essentially, there are two strategies that work for image restoration: one hardware approach and another software approach. A hardware approach might evolve the replacement of the x-ray tube with a finer focal spot or a better detector element of smaller pitch resulting in more precise CT scanning. However, these hardware approaches are usually more expensive and require more time-consuming protocols and radiation doses. Software approaches are more cost-effective and easier to be integrated into the existing hardware. Therefore, we will focus on software approaches in the following sections.

1.3.1 Conventional methods

Image deblurring aims to recover a clean image from a degraded blurry image. For image deblurring, conventional methods are either non-blind deconvolution or blind deconvolution approaches. Non-blind deconvolution strategies use the point spread function (PSF), which is used to describes the responses of an imaging system to a nearly infinitely small object, of the optical system. By using a calibration piece of a shape that is known, we can measure the PSF of an imaging system. However, a calibration piece is not always available. Therefore, blind deconvolution methods were used to derive an approximated PSF straight from the resulting images, either iterative or non-iterative. The former iteratively estimate the PSF from a set of parameters, and the latter calculates the PSF immediately by using arbitrarily determined parameters[Makarkin and Bratashov, 2021]. Blind deconvolution also has limitations. It is an ill-conditioned problem that can have a large number of solutions. An insufficiently accurate PSF estimate can further induce artifacts to the image. Therefore, deep learning-based methods algorithms are useful because they are specifically tailored for a task as such extracting information from data.

Super Resolution (SR) deals with the problem of reconstructing the high-resolution image from a low-resolution image. For the SR, the conventional method was to use interpolation: taking the average of the surrounding pixels to produce a smooth transition between the known and unknown pixels. There are different interpolation methods, such as bilinear interpolation and bicubic Interpolation. However, all of these methods will cause a big chunk of information loss.[Siu and Hung, 2012]

1.3.2 Deep learning-based methods

The past decade has seen an unprecedented increase in deep learning in computer vision tasks. The deep learning-based methods also have been used for both deblurring and SR.

A convolutional neural network(CNN) is the one of most commonly used neural network architectures, consisting of convolutional kernels and each kernel learns certain features of an image. A CNN can therefore learn to approximate the input to the target by using many convolutional kernels. The progress would be that the data set is collected and then fed to the neural network and the neural network would train itself using this vast domain-specific data. After the training phase has ended, the desired result can be achieved by using a feed-forward propagation.

U-net is one widely used CNN that has been tested on deblurring. Then [Zunair and Hamza, 2021] proposed a new U-net by combining Sharpen and U-net. The resulting network did, in fact, perform better than the original U-net. Another would be the Residual Dense Network(RDN)[Zhang et al., 2020]. The RDN would keep the information contiguous and add a residual to improve the matching process.

For the image SR, the Sharp U-net can also be used. Although the Sharp U-net[Zunair and Hamza, 2021] is not tested for SR, others have already used the u-net structure for SR. [Hu et al., 2019] Although the input size and the output size of Sharp U-net also have to be the same, as a normal U-net, so it may seem like it can't be used for SR, but it can. By first up-scaling the input image—so that the input shape matches the output shape—the Sharp U-net would then train the input data to match as close to the ground truth as near as possible. Although strictly speaking, the Sharp U-net is not doing the SR, it did, however, give a way for the Sharp U-net to be used for SR. The RDN can also be used for image SR. It would extract abundant local features through the dense connected convolutional layers by connecting all of the layers together and contains both Local feature fusion and Global feature fusion.

Another way to do both of the operations is Cycle-GAN. Cycle-GAN is an Image-to-image translation network that can be used for both SR and deblurring. The input and the output would be paired. In fact, there is already studies exploring Cycle-GAN on SR[You et al., 2019][Tien et al., 2021]. In their study, they produced some desirable results, comparable to other SOA methods. Furthermore,

One problem with applying these deep learning methods to medical image deblurring and SR is the gap between the pace of development of deep learning-based CV and its adoption in the medical domain. While there are many effective methods for deblurring and SR, most of them are designed for natural image datasets such as the ImageNet[Deng et al., 2009]. Due to considerable differences between natural and medical images (content, texture, color vs. gray value, etc), these deep learning methods may not have transferability to the medical imaging domain. It is therefore important to evaluate the feasibility of these methods in medical imaging tasks.

1.4 Proposed method

In this work, the authors delve into some very popular deep learning techniques used for deblurring and SR in the medical domain. We propose to use three deep learning methods to address these two limitations. These include two dense connection-based

architectures and one U-Net-based architecture. The dense connection-based architectures are RDN and a novel sharp RDN. While dense connection-based architectures are shown to improve the information flow between layers, these networks are computationally power-consuming. Therefore, we also include one U-Net-based method to examine if this memory-efficient architecture is capable of the two tasks.

Chapter 2

Methodology

2.1 Datasets and preparation

Training of a supervised CNN requires paired input and target, i.e. paired blurry input and sharp target for deblurring, and low spatial resolution images as input and the corresponding high-resolution images as the target for image SR. Since it is difficult to obtain paired images, we used simulation strategies for the training input. For image SR, we used the bicubic downsampling model to simulate LR images with scaling factors $\times 2$, $\times 3$, $\times 4$, and $\times 8$. For image deblurring, we used a 25×25 Gaussian blur kernel of standard deviation 1.6 to blur the images; afterward, an additive Gaussian noise of $\sigma = 2$ will be added to input images. All the image preprocessing procedures were conducted with Skimage Python toolbox[Van der Walt et al., 2014].

We used two public data sets - one simulation and one clinical - for our study. The simulation data set in this study was obtained from the AAPM DL-spectral CT challenge[Sidky and Pan, 2022]. The 512×512 pixel training images are from a breast phantom simulation. Since data was designated for dual-energy CT material decomposition, 50 and 80 kVp CT data and the corresponding adipose, fibroglandular, and calcification spatial maps were simulated. We only used the adipose images for our experiments as the other data have similar geometric shapes, so there is no added value to train on all data. The clinical data set was obtained from the TCIA Pancreas-CT data[Roth et al., 2016]. The dataset includes 80 abdominal contrast-enhanced 3D CT scans from 53 male and 27 female subjects. The image size is 512×512. In our study, we randomly selected 63 patient scans for training, 6 for validation, and 11 for testing. In total of 1230 images divided into three groups with 968 for training, 103 for validation, and 159 for testing.

Deblurring (all model)		Input	Target
		64×64 patch+blur+Gaussian noise	64×64 patch
SR	RDN	$\times 2$ downsampled 32×32 patch	64×64 patch
	Sharp U-Net	$\times 2$ downsampled+$\times 2$ upsampled 64×64 patch	64×64 patch

Table 2.1: Training studies and the corresponding data configurations. Note that *RDN* uses downsampled images as the input for SR, whereas the input and target of *Sharp U-Net* are of the same size, so downsampling followed by upsampling is used for the input preparation.

The training was applied to patch-based data to reduce memory requirements. From both input and target images, we extracted 64×64 patches with a stride of 32. A stride refers to the sampling interval in pixels between the neighboring patches. The inference was performed with the full-size images. The training input and target for each task can be seen in Table 2.1.

2.2 Proposed method

2.2.1 Network architectures

In this paper, we proposed to use three neural network architectures to realize medical CT image deblurring and SR. They are RDN, SRDN, and Sharp U-Net.

RDN is a neural network architecture originally proposed for image SR[Zhang et al., 2018]. It further demonstrated the state-of-the-art (SOTA) performance in other image restoration tasks[Zhang et al., 2020]. The key module of RDN is the Residual Dense Block (RDB). As its name suggests, RDB combines both residual and dense learning. Each of the RDBs would encompass a local feature fusion (LFF) and local residual learning, an integrated residual jump connection. Through connecting all of the layers in the channel dimension, it would achieve feature reuse. Two convolutional layers named shallow feature extraction (SFENet) are used to extract shallow features at the end of the network. There are mainly two traits of RDN that contribute to its SOTA performance. First, the preceding layers have direct access to all the subsequent layers. The local feature information is therefore fully used. Second, unlike many common CNN architectures, pooling layers are also removed in RDN. This design helps to retain pixel-level information. The RDN and RDB architectures are shown in Fig. 2.3.

Sharp U-Net[Zunair and Hamza, 2021] is a modification from the classic U-Net architecture[Ronneberger et al., 2015]. U-Net consists of a contracting path and an ex-

pansive path. The contracting path follows the typical architecture of a convolutional network and downsamples the feature in each step. The expansive path consists of an upsampling in each step. This way, different levels of image features can be learned efficiently. In order to alleviate the fusion of dissimilar features between the contracting path and expansive path, the architecture of the shape U-Net consists of a depthwise convolution to perform sharpening on the prior encoder features and then fuses them with the later decoder features. Through the use of sharpening, not only the information gap between the layer, which came from the high-level process from the decoder sub-network, will reduce, but the details in the feature maps will also be sharpened, which means low-frequency noise propagated from the untrained parameters in the feature space will be reduced. The sharpening operation does not introduce any additional learnable parameters, hence is an inexpensive add-on module for any backbone architecture.

Inspired by the Convolution Network, we adopted the local dense connections into our proposed network. Including Sharpening, the Sharp RDN might be better trained because the low-frequency information is ignored and only the high-frequency information is focused upon, resulting in a better trained neural network. The RDN was been used because it is considered as the SOTA for both deblurring and SR, therefore, the Sharpen is added to see if it can be even furthermore improved. The RDN has been tested and measured in a previous study for both deblurring and SR and proved to be SOTA. [Zhang et al., 2020]

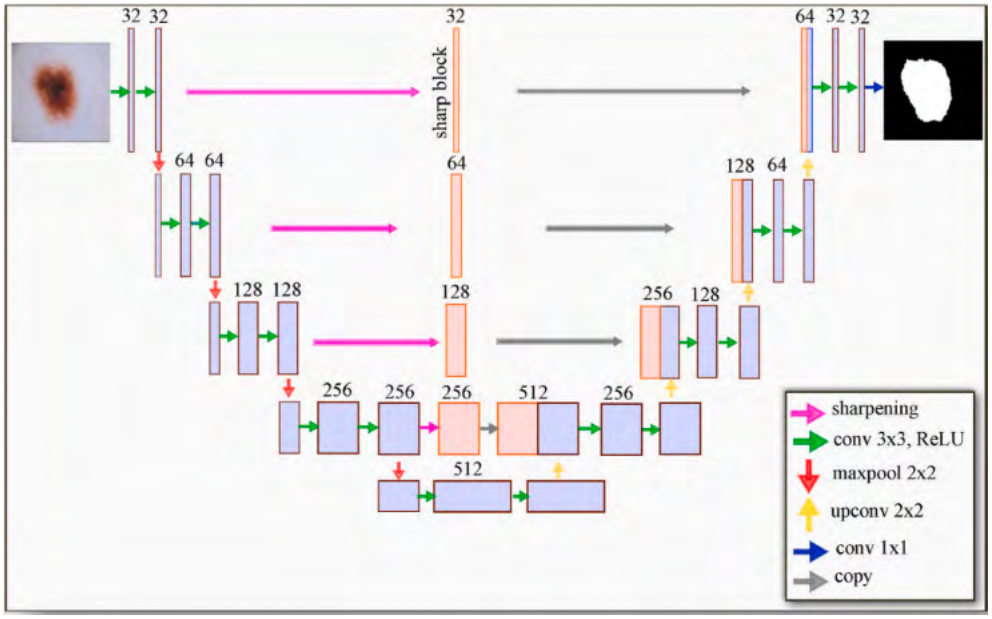

Figure 2.1: The sharpen U-Net architecture. Sharpen U-Net is a modification of the classic U-Net. Each output of the contracting path is sharpened by a predefined kernel and concatenated with the corresponding output of the expansion path[Zunair and Hamza, 2021].

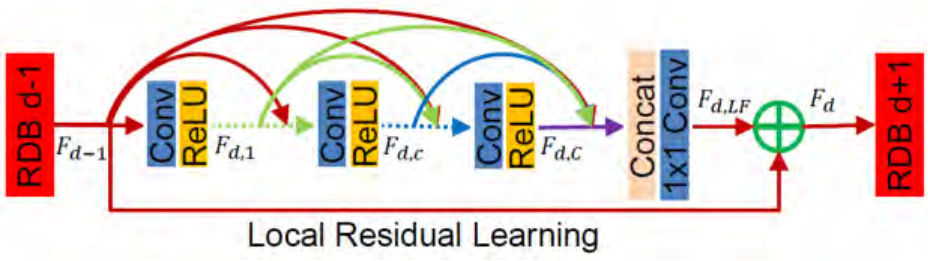

Figure 2.2: Residual dense block (RDB) is composed of three convolutional layers and each is followed by a ReLU activation. The output of each convolutional layer is concatenated with the output of the next layer, leading to a contiguous memory (CM) mechanism[Zhang et al., 2018].

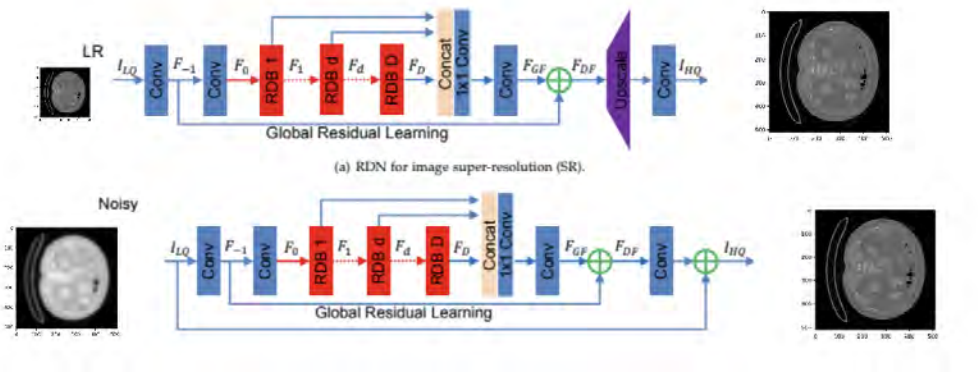

Figure 2.3: RDN mainly consists of four parts: shallow feature extraction net (SFENet), residual dense blocks (RDBs), dense feature fusion (DFF), and finally the up-sampling net (UPNet)[Zhang et al., 2018].

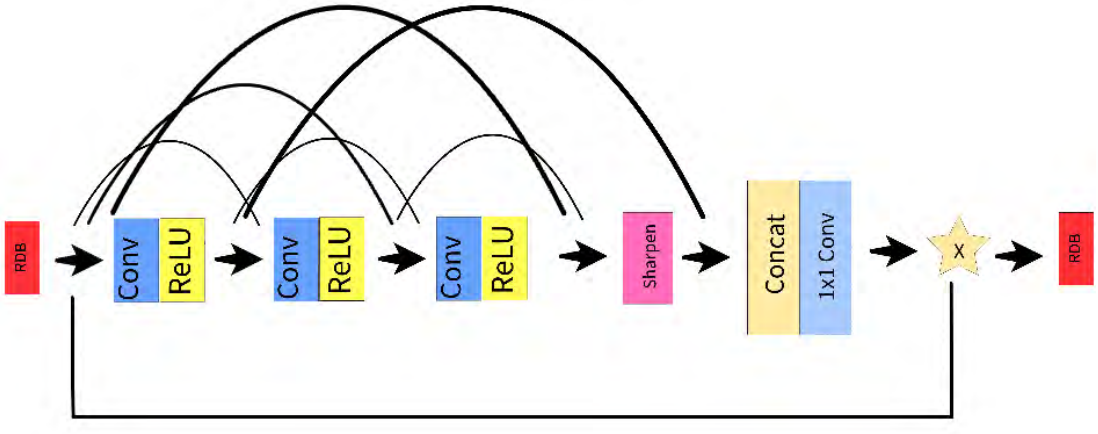

Figure 2.4: The network is composed of the original Conv and ReLu and an extra Sharpen.

2.2.2 Training environment

Experimented were performed on the Ubuntu operating system on the NVIDIA 2080Ti GPU hardware environment. The models were implemented using the TensorFlow 2.8.2 and Keras 2.8.0 frameworks, which are in Python 3.6.

In RDN architectures, we set 3×3 as the size of all convolutional layers except that in local and global feature fusion, whose kernel size is 1×1. Shallow feature extraction layers and local and global feature fusion layers have 64 filters. For the layers in each RDB, each has 64 filters and is followed by a ReLU activation. For image SR, we use ESPCNN[Shi et al., 2016] to upscale the coarse resolution features to fine ones. Five RDBs were used for the RDN architectures.

With a starting learning rate of 1×10^{-3}, the networks were trained through the use of the Adam optimizer over 200 epochs with a batch size of 32. Every time when there is no improvement for 10 epochs, each of whose duration is about 20 minutes, the learning rate will be reduced by a factor of 0.2. 1×10^{-7} is set to be the lower bound for the learning rate.

The squared L2 norm loss, or the mean squared error (MSE) loss, is chosen to be the loss function. The MSE is the mean of all of the squared distances between the model's target values and predicted values. By using the MSE loss function, the network would seek to minimize the value of MSE:

$$MSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{2.1}$$

Where N is the number of samples that are tested, i is the index of each value, $\hat{y}$ means the target value, and y means the predicted value.

2.3 Performance evaluation

The metrics we used to evaluate the model's performance were Root-Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). [Chai and Draxler, 2014][Hore and Ziou, 2010]

RMSE (Equation 2.1) represents the average of all of the root-squared differences between the actual values and the predicted.

$$RMSE = \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2.2}$$

A higher value of RSME means a low quality of the reconstructed image. PSNR, which measures in units of decibels (dB), represents the ratio between the maximum possible value, or power, of a signal and the value of the noise. The equation for PSNR is below, where MAX_I represents the maximum possible pixel value of the image. MSE in the equation represents the MSE between the two images.

$$PSNR = 10 \times \log_{10}(\frac{MAX_I^2}{MSE}) \tag{2.3}$$

The high value of PSNR means a high quality of the reconstructed image. In most cases, an image with a PSNR value higher than 35dB would be considered a good image.

SSIM measures the similarity between two images by using a weighted comparison of the luminance, contrast, and structure of the images. The maximum possible value of the SSIM is 1 and the larger it is, the better it is. So in other words, it is best when the SSIM is close to 1.

All metric calculations used the original image as the ground truth reference. However, for SR the metrics cannot be calculated for the original input as the input and target are of different sizes. To quantify the SR improvement from the original input, all input images were upsampled to the original image size before calculation.

Chapter 3

Results

3.1 Image deblurring

3.1.1 AAPM simulated data

The AAPM data set is simulated data. Since CT is subject to various artifacts that degrade the image quality (e.g. noise, beam hardening, scattering, etc.), by using simulation we can exclude these artifacts and focus on intuitively measuring the magnitude of changes caused by image blurring and the improvement of each deblurring deep learning method.

These quantitative results were calculated slice by slice in the test images. The results were averaged and presented in Tab 3.4. The Sharp U-Net performed poorly. It only showed 10% improvement in PSNR, 20% improvement in SSIM, and 38% in RMSE compared to the original blurred image; The RDN and the Sharp RDN performed similarly and both are significantly better than Sharp U-net, with the RMSE being only one-tenth of Sharp U-Net, and with PSNR and SSIM distinctly greater than the Sharp U-Net. However, the sharp layer added to the original RDN did not improve

Table 3.1: Quantitative comparison of image quality metrics evaluated against the corresponding ground truth images. This table is the result of the Deblur

	Blurred	RDN	Sharp RDN	Sharp U-Net
PSNR	27.34	48.86	48.45	30.12
SSIM	0.61	0.99	0.99	0.73
RMSE	0.0431	0.0036	0.0038	0.0312

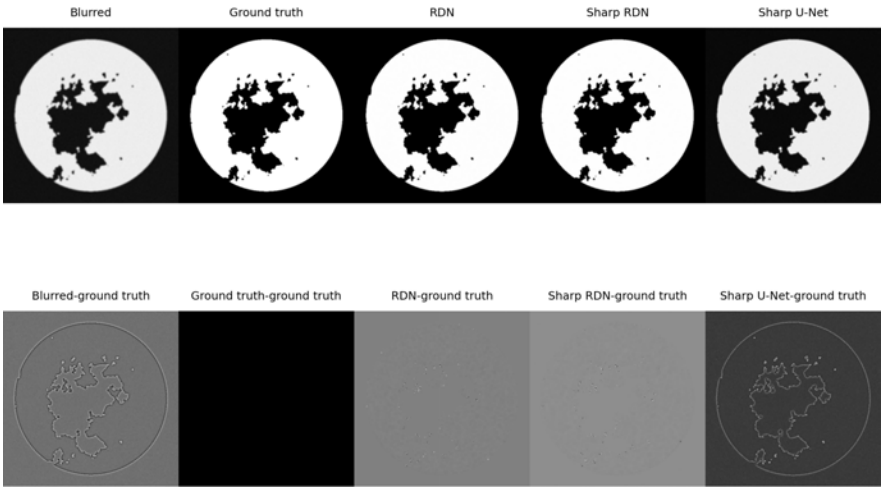

Figure 3.1: A visual comparison of the AAPM data by using different methods.

Table 3.2: Quantitative comparison of image quality metrics evaluated against the corresponding ground truth images.

	Blurred	RDN	Sharp RDN	Sharp U-Net
PSNR	13.14	37.34	36.10	21.56
SSIM	0.92	0.99	0.99	0.96
RMSE	50.48	10.85	11.78	33.11

the performance.

3.1.2 TCIA clinical data

Contrary to the AAPM data, TCIA data set is collected from the real patient so the performance of the TCIA data set would reflect more practical situations.

Just like the AAPM, when the Sharp U-net is used, it performs poorly compared to RDN and Sharp RDN. On the other hand, the RDN and Sharp RDN did significantly better than the Sharp U-Net, as the PSNR of RDN and Sharp RDN has a score of 37.34—184%—and 36.1—274%—whereas the Sharp U-Net only has a score of 21.56——only 64 percent. However, it can be seen that adding a Sharpen layer did not help the network to improve.

The validation loss with the number of the epochs train graph:

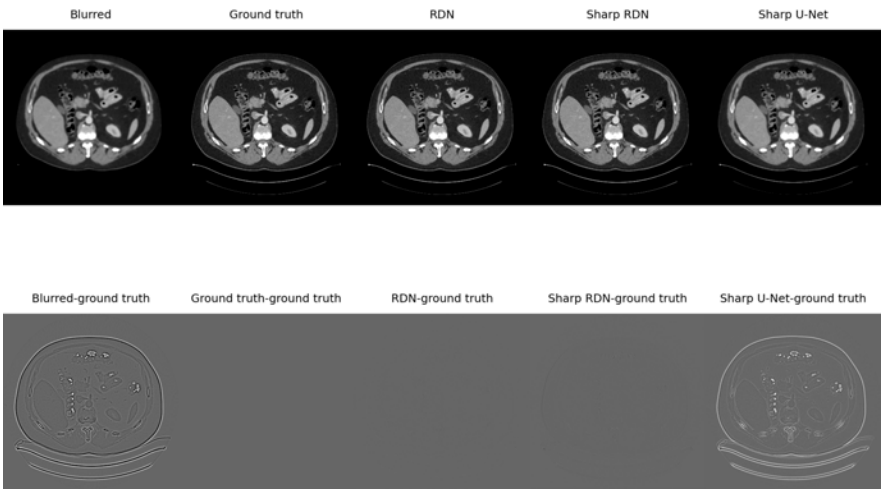

Figure 3.2: A visual comparison of the TCIA data by using different methods.

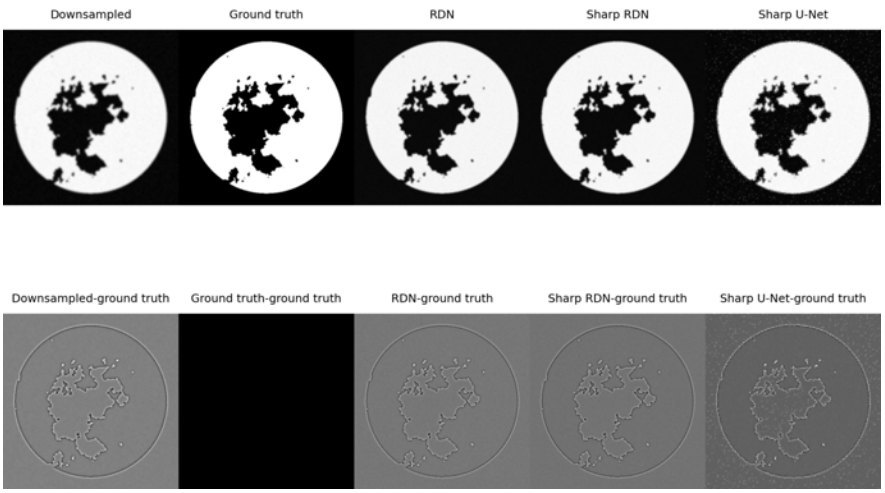

Figure 3.3: A visual comparison of the AAPM data by using different methods.

3.2 SR

3.2.1 AAPM simulated data

Although the Sharpen RDN did not perform well on the deblurring part, it did perform better than the RDN on the SR part for the AAPM data set. For all three——PSNR, SSIM, and RMSE——the sharp RDN did better than the normal RDN, which is somewhat surprising for the other tasks and data set, in which the sharp RDN did not perform better than the normal RDN. Also can be seen is that the image produced by the Sharp U-net contains some sparkle

From the table, it can be seen that the Sharp U-net performs the worst again, as all three metrics of the image produced by the Sharp U-Net were worst than the down

Table 3.3: Quantitative comparison of image quality metrics evaluated against the corresponding ground truth images. This table is the result of the SR.

	down resolution	RDN	Sharp RDN	Sharp U-Net
PSNR	23.65	26.49	26.35	22.84
SSIM	0.79	0.80	0.81	0.76
RMSE	0.0609	0.0382	0.0385	0.0420

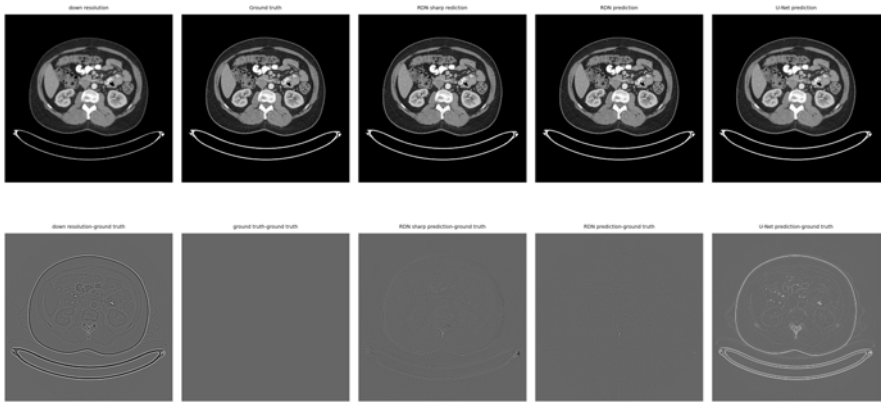

Figure 3.4: A visual comparison of the TCIA data by using different methods.

resolution itself——only 96% as the original. The Sharp RDN also performed better than the normal RDN, although their measures were somewhat close——about 12%.

3.2.2 TCIA clinical data

For the SR part, the RDN also performed better than the Sharp U-Net. As can be seen from the graph, adding a sharpened layer did not help to improve the proximity to the original image. The Sharp U-Net performed worst of all methods.

As also can be seen from the table, RDN and sharp RDN did better than Sharp

Table 3.4: Quantitative comparison of image quality metrics evaluated against the corresponding ground truth images. This table is the result of the SR

	Down resolutioned	RDN	Sharp RDN	Sharp U-Net
PSNR	37.219	48.667	42.791	40.329
SSIM	0.44	0.538	0.407	0.586
RMSE	42.927	11.837	23.7	30.126

U-Net in both PSNR and RMSE; however, the Sharp U-Net did perform better than those two in SSIM—33%—showing that although the image produce by the RDN—an improvement of 30%—and sharp RDN—an improvement of 15%—is closer to the real image than Sharp U-Net, the structure of the image produced by the Sharp U-Net is closer to the real image.

Chapter 4

General Discussion

4.1 Discussions

The contributions of our work are twofold. First, our work has provided multiple deep learning-based solutions to two commonly encountered problems in medical CT images, which are image blurring and limited image resolution. One SOTA method in image restoration, RND, and one SOTA method in medical imaging was chosen for our study. As there are many modifications of the U-Net and Sharp U-Net is designed for restoring edge artifacts, we chose Sharp U-Net for our study. Correspondingly, a sharpen layer was also added to RDN to evaluate the effect of this operation. Moreover, our work is the first to evaluate the efficacy of RDN on medical images. Although there are many deblurring and SR models proposed in the medical imaging domain, most of them used GAN architectures which are hard to train and require a large data set that is not always available in medical imaging. Therefore, they lack efficiency and practical usage. In comparison, RDN requires much less time and data size to train and optimize. Our work has demonstrated its great potential in medical imaging.

There are mainly three design features that contribute to the superior performance of RDN compared to U-Net. First, there are no pooling layers. Pooling operations may discard pixel-level information, which is important for image restoration. Second, dense connections are used. The dense connections enhance signal propagation and stimulate feature reuse. The dense connections would substantially curtail the model size through the use of a relatively small growth rate, such as a low number of channels in dense blocks and squeezed channels after connecting all of the input feature maps. Last but not least, local residual learning in RDB improves the information flow. These features together make RDN a model suitable for deblurring and SR.

4.2 Limitations and future work

However, there are some limitations to be noted; one would be that all of the data used here was simulated, so it may not reflect the real circumstances. Another would be that an arbitrary number of RDBs was used, the number was picked merely for the convenience of training. In the future, we will conduct an experiment to determine the optimal number of RDBs for each of the tasks for CT medical images. Furthermore, inspired by the SSIM value of the Sharp U-Net, we will also conduct an experiment of SR by combining Sharp U-net and RDN——that is, let the Sharp U-Net first up resolution then train the RDN to match the corresponding images.

Chapter 5

Conclusions

All in all, in this paper, we tried to use a pre-existing CNN(the RDN[Zhang et al., 2020]) to perform deblurring and SR on CT images. We trained the networks by using data sets from AAPM and TCIA. The networks we train include RDN, Sharp RDN, and Sharp U-net. The measurements used to compare the reconstructed images are PSNR, SSIM, and RMSE. Based on the PSNR, SSIM, and RMSE, the regular RDN performed the best on image deblurring; when doing SR, based solely on PSNR and RMSE, the RDN did better than the other two, yet when based on SSIM, the result showed that the Sharp U-Net did the best.

Bibliography

[Bivard et al., 2013] Bivard, A., Levi, C., Spratt, N., and Parsons, M. (2013). Perfusion ct in acute stroke: a comprehensive analysis of infarct and penumbra. *Radiology*, 267(2):543–550.

[Chai and Draxler, 2014] Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250.

[Dammert et al., 2004] Dammert, S., Krings, T., Moller-Hartmann, W., Ueffing, E., Hans, F., Willmes, K., Mull, M., and Thron, A. (2004). Detection of intracranial aneurysms with multislice ct: comparison with conventional angiography. *Neuroradiology*, 46(6):427–434.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[Ghekiere et al., 2017] Ghekiere, O., Salgado, R., Buls, N., Leiner, T., Mancini, I., Vanhoenacker, P., Dendale, P., and Nchimi, A. (2017). Image quality in coronary ct angiography: challenges and technical solutions. *The British journal of radiology*, 90(1072):20160567.

[Herman, 2009] Herman, G. T. (2009). *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media.

[Hore and Ziou, 2010] Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.

[Hsieh, 2003] Hsieh, J. (2003). Computed tomography: principles, design, artifacts, and recent advances.

[Hu et al., 2019] Hu, X., Naiel, M. A., Wong, A., Lamm, M., and Fieguth, P. (2019). Runet: A robust unet architecture for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0.

[Hu et al., 2022] Hu, X., Zhong, Y., Huang, Y., Shen, C., and Jia, X. (2022). Improving small animal cone beam ct resolution by mitigating x-ray focal spot induced blurring via deconvolution. *Physics in Medicine & Biology*.

[Ketcham and Hanna, 2014] Ketcham, R. A. and Hanna, R. D. (2014). Beam hardening correction for x-ray computed tomography of heterogeneous natural materials. *Computers & geosciences*, 67:49–61.

[Makarkin and Bratashov, 2021] Makarkin, M. and Bratashov, D. (2021). State-of-the-art approaches for image deconvolution problems, including modern deep learning architectures. *Micromachines*, 12(12):1558.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[Roth et al., 2016] Roth, H. R., Farag, A., Turkbey, E. B., Lu, L., Liu, J., and Summers, R. M. (2016). Data from pancreas-ct. *The cancer imaging archive*, 32.

[Shi et al., 2016] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883.

[Sidky and Pan, 2022] Sidky, E. Y. and Pan, X. (2022). Report on the aapm deep-learning sparse-view ct grand challenge. *Medical Physics*.

[Siu and Hung, 2012] Siu, W.-C. and Hung, K.-W. (2012). Review of image interpolation and super-resolution. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–10. IEEE.

[Tien et al., 2021] Tien, H.-J., Yang, H.-C., Shueng, P.-W., and Chen, J.-C. (2021). Cone-beam ct image quality improvement using cycle-deblur consistent adversarial networks (cycle-deblur gan) for chest ct imaging in breast cancer patients. *Scientific Reports*, 11(1):1–12.

[Tilley et al., 2015] Tilley, S., Siewerdsen, J. H., and Stayman, J. W. (2015). Model-based iterative reconstruction for flat-panel cone-beam ct with focal spot blur, detector blur, and correlated noise. *Physics in Medicine & Biology*, 61(1):296.

[Van der Walt et al., 2014] Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, 2:e453.

[You et al., 2019] You, C., Li, G., Zhang, Y., Zhang, X., Shan, H., Li, M., Ju, S., Zhao, Z., Zhang, Z., Cong, W., et al. (2019). Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE transactions on medical imaging*, 39(1):188–203.

[Zhang et al., 2018] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481.

[Zhang et al., 2020] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2020). Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495.

[Zunair and Hamza, 2021] Zunair, H. and Hamza, A. B. (2021). Sharp u-net: depthwise convolutional network for biomedical image segmentation. *Computers in Biology and Medicine*, 136:104699.