

Super resolution of Covid-19 CT-Scan Images

by

VAIDIK GAUTAM PATEL
202011038

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY



July, 2022

Declaration

I hereby declare that

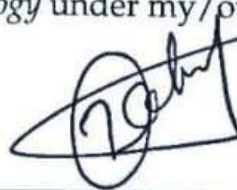
- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the text to all the reference material used.



Vaidik Gautam Patel

Certificate

This is to certify that the thesis work entitled INSERT YOUR THESIS TITLE HERE has been carried out by INSERT YOUR NAME HERE for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.



Bakul Gohel
Thesis Supervisor

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Bakul Gohel for seeing me fit for this thesis research. Words can't express how thankful I am for his constant guidance, patience and invaluable feedback. His vision, continuous critique over my work has vastly enhanced my knowledge. Without his continuous support this thesis would not have concluded.

I would also like to thank my friends and peers who have supported me and helped me in times of need. Their emotional support has motivated me the a lot. Lastly, I would be remiss in not mentioning my family, whose belief in me has always kept my spirits high.

Contents

Abstract	v
List of Principal Symbols and Acronyms	v
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Background	1
1.2 Outline	3
2 Fundamentals	4
2.1 Super Resolution	4
2.2 Lung CT image	5
3 Related Work	6
3.1 SR using conventional methodologies	6
3.2 SR using Deep Learning	7
3.3 SR of lung CT	8
4 Methodolgy	10
4.1 Introduction to the chapter	10
4.2 Network Architecture	10
4.3 Pre-processing	13
4.3.1 Z-Score	14
4.3.2 Scaling	14
4.4 Evaluation Metrics	14
4.4.1 PSNR	15
4.4.2 SSIM	15

5 Experiments	17
5.1 Dataset	17
5.1.1 Medseg Part	17
5.1.2 Radiopaedia Part	18
5.2 Implementation Details	18
5.2.1 Training and Testing Split	18
5.2.2 Pre-processing	19
6 Results	21
6.1 Image Super resolution Results	21
7 Discussion and Conclusion	27
7.1 Introduction to the chapter	27
7.2 Discussion	27
7.3 Conclusion	28
References	29

Abstract

Acquisition of high quality CT images is difficult, because it requires exposing patients to high doses of radiation. Super resolution algorithms can help in overcoming this problem and obtain higher spatial resolution in CT images. Much deep learning based architecture have been proposed in the literature to overcome this problem. We perform the task of super resolution on a U-Net and study the effects of 2 pre-processing methods which are scaling and z-score. The evaluation strategy for the super resolution of CT images in the literature uses the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM), however the results are published for the entire image. This is not a good practice for the evaluation of SR, we propose a novel region based similarity measurement practice and a lung specific or region of interest based similarity measurement. We further bifurcate the SSIM metric into it's 3 component, i.e. luminance, contrast and structure, and study the impact of super resolution on each of these components.[5]

List of Tables

6.1	The results of PSNR computed locally for different tissues for down-sampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic	22
6.2	The results of SSIM and score for each of it's components i.e. L(luminance), C(contrast) and S(structure) for a downsampling factor of 4 on the Medseg dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic.	22
6.3	The results of SSIM and score for each of it's components i.e. L(luminance), C(contrast) and S(structure) computed locally for different tissues for downsampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic.	22
6.4	The results of PSNR computed locally for different tissues for down-sampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic	23
6.5	The results of SSIM and PSNR computed for the full image for downsampling factor of 4 on the Medseg dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic	23
6.6	The results of SSIM and PSNR computed for the full image for downsampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic	23

List of Figures

4.1	UNET architecture[7]	12
5.1	Raw high resolution CT images from Medseg data-set	18
5.2	Image (a) is a high resolution CT image from the Radiopaedia data-set. Image (b), (c) and (d) are the Ground Glass Opacity mask, the consolidations mask and the healthy lungs mask respectively.	20
6.1	HR,LR,SR and Bicubic CT images	24
6.2	Region specific comparison of SSIM for SR and Bicubic with HR image patch.	25
6.3	Region specific compasrison of SSIM for SR and Bicubic with HR image patch.	26

CHAPTER 1

Introduction

1.1 Background

Super-resolution refers to the difficult process of predicting a high-resolution (HR) image from its low-resolution (LR) equivalent (SR). High-resolution (HR) images are in high demand for a variety of real-world applications, including medical imaging for clinical duties, geographic information systems, security video monitoring, finger-print image enhancement and others. The imaging hardware, on the other hand, usually limits the image resolution and quality. With the dawn of super resolution algorithms and techniques obtaining HR images from their LR counterparts became possible without upgrading the hardware, because of these reasons SR garnered a lot of attention within the research community. The SR being an ill posed problem, typically the SR reconstructed image lacks textural details. With the advent of AI and Deep Learning, the performance of SR algorithms have improved significantly for natural images. However medical images are different from natural images. In contrast to natural images, medical data does not typically include orientation, and objects in medical photos have no canonical orientation. Tissue slices in digital pathology photographs, for example, are placed on the glass without any predetermined alignment. Background tissues in a medical image can have an orientation-dependent structure, the anomalies themselves do not. For example in a lung CT image, the lungs have a certain shape that the DL framework can learn, however the anomalies themselves are irregular in shape These pre-trained models based on real-world images may generate unreal patterns, affecting clinical interpretation and diagnosis[14].

In the medical field the acquisition of HR images is highly desirable as it would provide crucial details of the physiological, functional, anatomical and metabolic information of patients. The aforementioned hardware limitations exist in the medical field as well. Apart from these hardware limitations in the medical field, health limitations exist as well. MRI, CT and ultrasound are the widely used imag-

ing modalities. The technical quality of the device as well as the conditions of the scan influence the acquisition of a medical image. In MRI, high spatial resolution usually means a longer scan time, less spatial coverage, and a worse signal to noise ratio (SNR). Another most used medical imaging technology for screening, diagnosis, and image-guided intervention is X-RAY computed tomography (CT). CT scanners use ionized radiations, high doses of such radiations are harmful for the human body. Therefore in clinical practice CT scans are obtained at low doses of radiations.

There are two general approaches to increasing CT image resolution: (1) hardware-oriented and (2) computational. For starters, more complex hardware components, such as an x-ray tube with a narrow focal spot size, small-pitch detector elements, and improved mechanical precision for CT scanning, can be used. These hardware-based approaches are often costly, increase CT system costs and radiation exposure, which can induce genetic damages and be a cause of cancerous diseases. SR can help overcome the hardware limitations by increasing the spatial resolution of the LR medical images. In this study we perform the task of SR on a lung CT dataset.

SR can be applied in the frequency domain as well as the spatial domain. Maximum research for SR has been done in the spatial domain. The simplest of these approaches convert LR images to HR images in the spatial domain is using the various interpolation methods. The interpolation techniques uses cubic spline or other polynomial techniques to enlarge or sharpen a 2D grid of pixels. The most commonly used interpolation technique is the bicubic interpolation. In the spatial domain there are two approaches : (i) Multiple Image SR (ii) Single Image SR. The multiple image based algorithm uses multiple LR images and a single target HR image. The multiple LR observations have some geometric or photometric similarity as that of the HR image. Single Image SR approaches learns the mapping from a single LR image to the HR image and based on these mapping constructs the SR image. The multiple image based SR algorithms generally construct SR images that fail to retain the high frequency information and smoothens the images. In the medical field retaining this high frequency information and textural details is important as lack of it can lead to misdiagnosis. In the SISR, learning based approaches have shown boosted performance with the advent of deep learning. This is in part because of increase in computational power as well as availability of big data.

In this study we use the U-Net architecture for the task of super resolution. U-Net has achieved state of the art performance for the segmentations based tasks

in the medical domain. We have also performed experiments using two different pre-processing methods to analyse which pre-processing method gives good results across datasets. In the literature no study has used both the pre-processing methodologies, and as a result no proof is available as to which one performs better across datasets. For evaluation, PSNR and SSIM are the most widely used performance metrics for the task of super resolution. These are full – reference metrics, which means that these metrics require a full reference image for evaluation. The metric score for a full image makes sense for natural images, as the information content is spread throughout the image. In medical images the background information is redundant and is low frequency information, or the pixels in the background of medical images has a constant pixel intensity. Retaining this information is an easy task for any DL based method. Subject to this, calculating the metric score for the entire image will yield a high score. This higher score however will not give us an accurate evaluation of our SR algorithm or model. To mitigate this we calculate the ROI based or region specific or tissue specific PSNR and SSIM scores. We also test our model on a different dataset, using the two pre-processing methods and performing region specific similarity measurements.

1.2 Outline

The entire thesis is organised in 7 chapters. Chapter 2 contains fundamentals needed for SR. Chapter 3 contains literature survey. Chapter 4 presents the methodology proposed. Chapter 5 contains experiments. Chapter 6 elaborates on the results. Chapter 7 concludes the thesis.

CHAPTER 2

Fundamentals

2.1 Super Resolution

Super resolution is the task of enhancement of low resolution image (LR) to a high resolution image (HR). Let's consider a LR image y and HR image x and define a degradation function F as given in Equation 2.1.

$$\mathbf{y} = F(\mathbf{x}; \theta_\eta) \quad (2.1)$$

In the above equation θ_η refers to the degradation parameters. In theory, these parameters can be noise and other artifacts but practically such a degradation function is not possible. It is our assumption that the degradation parameters degrade the quality of the HR image x and as a result give us a low resolution LR image y . In reality, we have a LR image y and we need to devise a method, Super-Resolution, to enhance its quality and get the HR image x . The Equation 2.2 represents the method F^{-1} super-resolution method.

$$\mathbf{x}' = F^{-1}(\mathbf{y}, \theta_\zeta) \quad (2.2)$$

The θ_ζ refers to the parameters of the function F^{-1} . The resultant image is x' , an approximation of x . Getting x is an ideal scenario and not pragmatic approach. The equations given above explain the overall scenario of the super-resolution task i.e. to identify an approximation method to get HR image. However, the degradation process is very complex and meekly known. It involves several parameters like blurring, scaling noise etc. and so many research work prefer the degradation method as given in Equation 2.3.

$$\mathbf{y} = (\mathbf{x} \otimes \mathbf{k}) \downarrow_{sf} + \mathbf{m} \quad (2.3)$$

The k is the blurring kernel and \otimes represents the convolutional operation be-

tween the HR image x and k . The \downarrow represents the downsampling with scaling factor of sf and m denotes the additive white Gaussian noise (AWGN).

In image super-resolution the aim is to minimize the data fidelity term with model $y = x \otimes k + m$ with a regulariser called Image prior Ψ with a balancing factor. The entire super-resolution method is denoted in the Equation 2.4. The α is the balancing factor.

$$J(x', \theta_c, k) = \underbrace{\|x \otimes k - y\|}_{\text{data fidelity term}} + \alpha \underbrace{\Psi(x, \theta_c)}_{\text{regularizer}} \quad (2.4)$$

2.2 Lung CT image

In lung CT images, there are 2 types of regions, namely Ground Glass opacity and consolidations. Ground Glass Opacity is a region of lung that is hazy. There is an increase in the opacity of lung in that region but vessels and bronchial structures are still visible. At high-resolution computed tomography, pulmonary ground-glass opacity (GGO) is characterised as a hazy opacity that does not cover underlying bronchial structures or pulmonary arteries (HRCT). Therefore for the differential diagnosis it is important that the SR algorithms, or the deep learning architectures deployed for the task of super resolution successfully retain high frequency information in the GGO regions of CT images[3]. The second type of abnormality found in the CT scans is known as consolidations. The small airways present in the lungs are generally filled with air, consolidations in the lung occur when these airways are filled with something else. Depending on the cause the accumulation of inflammatory cellular exudate in the alveoli and surrounding ducts causes consolidation.. Induration (swelling or hardening of typically soft tissue) of a normally aerated lung characterises consolidations. To diagnose pneumonia, consolidation must be present: the indications of lobar pneumonia are distinct and clinically referred to as consolidation.

CHAPTER 3

Related Work

3.1 SR using conventional methodologies

There is an abundance of SR algorithms in the literature, frequency and spatial domain. SR is the process of constructing a high resolution image from a low resolution. The hardware-based solution is to increase the amount of pixels per unit area. This can be achieved by either of two ways: (i) decreasing the pixel size or (ii) increasing size of the sensor. But such hardware based approaches are expensive. Therefore, software based approaches are usually preferred. Classification of SR algorithms in the literature is based on the domain employed, number of LR images involved and type of reconstruction methods. On the basis of the LR images involved, SR algorithms can be classified into two types single image and multiple image based algorithms. In the frequency domain, images are first converted into the frequency domain and the HR image is estimated in this domain, after which it is transformed back into the spatial domain. There are two groups of methods in the frequency domain ; wavelet based methods and fourier transform based methods. In the spatial domain, algorithms are classified into two groups as mentioned earlier. The multiple image based algorithms includes iterative back propagation (IBP), iterative adaptive filtering, direct methods and maximum likelihood estimates. The iterative back propagation method first creates an initial estimate of the LR image on to the HR image space and then the target image is generated by refinement . The iterative adaptive filtering algorithms are generally used to generate super resolved videos from an LR video. It models the relationship between the current and the previous HR frame. The maximum a posteriori methods find an estimate of the HR image from one or more LR images, using the Bayes's rule. The single image based algorithms can be classified into two categories: (i) Learning based algorithms and (ii) Reconstruction based algorithms. One of the first learning based approach used a neural network. Algorithms in this category use a training step to learn the relationship between

the HR example and its LR counterparts. This newly acquired knowledge is subsequently added into the reconstruction's a priori term. The Feature pyramids algorithm creates a pyramid of LR images by applying Gaussian blurs and latter after training retrieves the best LR image from this pyramid of blurred LR images. Both the LR image and its HR counterparts are separated into patches by a belief network such as a Markov network. A so-called observation function is then used to link the corresponding patches in the two images. After training the model (parameters), it uses a belief propagation approach to infer the missing HR details of LR input images, resulting in a MAP super-resolved image. The other SISR algorithms are the reconstruction based algorithms, which address the aliasing artefacts in the input LR image.

3.2 SR using Deep Learning

The increase in the computational power and big data has made training deep learning models efficient. Deep learning has given state of the art results for many image processing tasks like object detection, semantic segmentation, object classification etc. This is because of the tremendous learning capabilities of deep learning models. Researchers have developed deep learning models for the task of super resolution as conventional models do not give optimal results for SR. Although the existing deep learning models for SR vary widely, they can be classified into four model frameworks: (i)Pre-Upsampling (ii)Post-Upsampling (iii)Progressive Upsampling (iv) Iterative Up-and-down sampling[10]. Because learning the mapping from low-dimensional to high-dimensional space is complex, using typical upsampling procedures to produce higher-resolution images and then improving them using deep neural networks is a simple solution, this type of model architectures are called pre-upsampling architectures. As a result, Dong et al. propose SRCNN [11] to learn an end-to-end mapping from interpolated LR images to HR images. In the post-upsampling framework researchers propose performing most computation in low-dimensional space by replacing predetermined upsampling with end-to-end learnable layers merged at the end of the models to improve computational efficiency and make full use of deep learning technologies to raise resolution automatically. In such models LR images are passed as input and layers that can learn end-to-end upsampling are added at the end of the networks. The computation and spatial complexity are greatly decreased because the feature extraction process, which has a high computational cost, occurs only in low-dimensional space and the resolution rises

only at the conclusion. Although the SR framework with post-upsampling has greatly decreased the computational cost, it still has certain flaws. These models only conduct one step of upsampling, which dramatically increases the learning complexity for big scaling factors. Another issue with these models is that each scaling factor necessitates the training of a separate SR model, which cannot cope with the need for multi-scale SR. To overcome these limitations, the progressive upsampling framework employs a cascade of CNNs to rebuild higher-resolution images gradually[4].

3.3 SR of lung CT

Recent years have seen rapid rise in utilisation of methods based on deep learning for image super-resolution. Traditional interpolation methods need prior information about the images in order to produce a mapping between LR and SR images. However, deep learning methods directly extract features from images and learn a mapping function between LR and SR images. Deep learning methods have powerful learning ability and therefore produce superior results.

The success of Convolution Neural Networks (CNNs) in computer vision tasks has motivated a plethora of researches to use them for Super-resolution task. The main reason for its success is the high level feature extraction capability and representation capability [13]. More recently, residual learning has also been used extensively. Coupled with CNN networks residual networks prevent vanishing gradient issues and helps the model to yield compelling results.

Chen et al. [1] proposed a dual path residual network that leverages the convolutional networks. The dual path is for capturing low and high frequency features respectively. These feature maps are upsampled to generate HR images.

Yu et al. [13] have also employed convolutional neural networks integrated for the task of super resolution in single as well multiple CT slices. For single slice, they have integrated a residual model that helps to fetch high-frequency details that reconstruct the HR outputs with higher quality. For multiple slices, coherency between neighboring slices is used in for better reconstruction of SR images.

You et al. [12] proposed a semi-supervised, robust model (GAN-CIRCLE) to produce denoised, deblurred high resolution (HR) CT images. Its a novel architecture

with residual CNN-based network in CycleGAN framework. With the generative adversarial network (GAN) as the foundation model, they have taken care of cycle consistency in order to get non-linear mapping from the LR to HR images to maintain across-domain consistency. The input images are preprocessed using BMD3 for noise removal and A+ for super solving. Using the windowed sync interpolation method they have reconstructed the dataset as the images were not isotropic.

Farias et al. [2] have proposed a modification in GAN-CIRCLE model. They have added spatial pyramid pooling (spp) in the discriminator part in order to incorporate multiple size inputs. Along with that they have cropped the input images to get better image qualities. These modifications have, as claimed, improved the perceptual quality of the HR images.

CHAPTER 4

Methodolgy

4.1 Introduction to the chapter

In this chapter we discuss the deep learning architecture, the pre-processing methods used and the evaluation metrics used in this study. In the first section the structure of the model is explained layer by layer. In the next section the two pre-processing methods used for comparative analysis have been discussed along with their formulas. The final section discusses the evaluation metrics used for the purpose of evaluation. Peak signal to noise ratio and Structural Similarity index measurement are the two metrics used for evaluation of Super Resolution methods, we have used the same in this study.

4.2 Network Architecture

For the task of single image super-resolution many state of the art deep learning architectures have been proposed, these include SRCNN, VDSR, ESPCN, SRGAN to name a few. We have hereby used U-NET[7] for the task of super resolution of biomedical CT images. U-NET has been widely regarded as the best deep learning network architecture for medical imaging tasks. The U-Net architecture has two paths, a contracting path and an expansive path. Typically CNNs are used for the classification task, however in biomedical imaging tasks output should contain information about the localization. Each pixel is expected to be allocated a class label. The U-net is a fully convolutional network (FCN). FCN is a network that lacks "Dense" layers (as in typical CNNs) and instead uses 1x1 convolutions to serve the function of fully connected layers (Dense layers). U-net is designed in such a way that it can work with very few training images and still yield precise segmentations. There are two paths in the U-net architecture, the contracting path and the expansive path. The contracting path is made up of convolutional blocks, these convolutional blocks are repeated throughout the contracting path.

It consists of two 3×3 convolutions which are unpadded, these are followed by a ReLU layer, which is the rectified linear unit layer and finally a 2×2 max pooling layer with a stride of 2 for the down-sampling operation. The number of feature channels are doubled after each down-sampling operation. The U-net architecture has a large number of feature channels in the upsampling section, allowing the network to pass context information to higher resolution layers. As a result, the expansive path is roughly symmetrical to the contracting path, resulting in a U-shaped structure. The expansive path consists of an up - convolution layer with size of 2×2 . The number of feature channels are halved after the up-convolution layer. The feature map generated after this up-sampling layer is then concatenated with the feature map from the corresponding block of the contracting path. As a result the number of feature channels are doubled. The concatenation of high resolution feature map is done so as to localize. This is then followed by two 3×3 convolution layers and then a ReLU. Because every convolution loses border pixels, cropping is required. The final layer is a 1×1 convolutional layer. The network consists of 23 convolutional layers, 4 max pooling layers and 4 up-sampling layers.

Figure 3.1 depicts the U-Net architecture used. Here the input size of the image is 512×512 . A pre-processing layer is added before the input layer to up-sample the LR images to the spatial resolution of the HR images. Bi-linear interpolation is used to up-sample the LR images from 128×128 to 512×512 . At the first convolution block 16 feature channels are extracted from our original gray scale image with 1 channel. In the subsequent blocks 32, 64, 128 and 256 feature channels are extracted before the expansive path. The horizontal gray arrows represent the concatenation process, the feature map from the contracting path is concatenated to the feature map of the corresponding convolution block in the expansive path. The final layer in the expansive path is 1×1 convolution layer which returns a SR grayscale image. The U-Net architecture used can be categorised as a pre-upsampling architecture as the LR image is first up-sampled into the HR spatial space and then the U-Net learns the mapping between them in the HR spatial space.

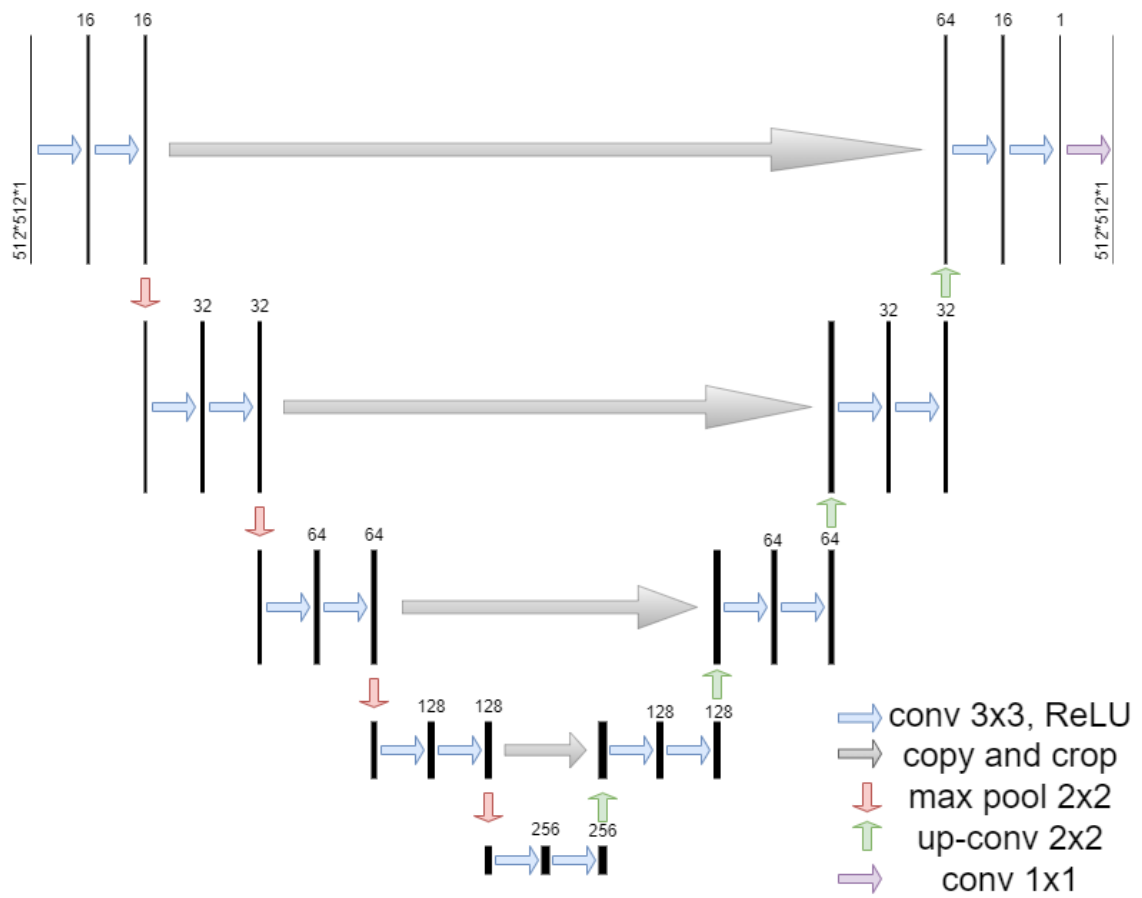


Figure 4.1: UNET architecture[7]

4.3 Pre-processing

Natural and radiology images are acquired in different ways to one another. The technological parameters employed to generate radiography images in each modality [i.e., computed tomography (CT) and magnetic resonance imaging (MRI)] are directly related to the technical parameters of the scanners used to obtain of these images. There are frequently modest variations in picture quality, field of view, and resolution when a big dataset is produced from data received from different patients, scanners, or numerous institutions, which should be taken into account through a few pre-processing steps before training deep learning models. In transfer learning, a model is fine-tuned using radiological images after it has been pre-trained on natural images. Natural images with a globally meaningful range of intensity levels are obtained, saved, and loaded into memory. Clinical pictures are often obtained in a variety of methods depending on the situation, with varying interpretations for their intensity levels. Thus, multiple pre-processing steps are beneficial to enabling a robust transfer learning from natural images into a dataset of medical images. Different pre-processing procedures can be seen as a quality check for radiology images that goes beyond the scanner's integrated image quality filters. Getting a specified level of image quality in the training dataset can help with deep learning quantification. The nature of the pre-processing approach is highly dependent on the following processing algorithm's specific goal as well as the image type. The Hounsfield unit is a quantifiable assessment of tissue density relative to water that results from CT acquisition (HU). Following the conventional temperature and pressure standards, voxel HU values in CT scans are mainly regarded reproducible, with minor variations across different scanners and patients. Pre-processing can be applied at a certain level based on the task of deep networks. For example, data normalisation can be done at the slice level (each individual slice inside an image), , patient level, image level, scanner level, institution level, or overall training data level utilising image statistics[6]. Most of the studies apply pre-processing at the patient-level, but that can change based on the application. CT images are collected in 12- or 16-bit DICOM format to display tissue density measurements in HU, which are integer integers between (1024 Hu and 4000 Hu). After possible de-noising, interpolation, and registration, these values may transform to real numbers. Direct display of these numbers to a radiologist as a picture has its limitations. The human eye can only identify 100 shades of grey at a time, yet it can distinguish thousands of colours. Image windowing can be used to boost contrast over a specific area of interest. In lit-

erature, the latter is widely employed, with CT image values trimmed around a specific band of HU values for each organ, based on unique texture-based dictionary requirements. After CT windowing comes the normalisation step which is performed to stretch or squeeze the CT data intensity values. This is to fit the CT data intensity values in the data range for provided image type. The data values are mapped to 0-255 if it is a 8 bit pixel or to 0-65,535 if it is a 16 bit image[6]. If the normalisation is performed at the data-set level then the minimum and maximum pixel values among all the patients or all the CT images in the data-set. We have used two normalisation techniques : (i) Z-score or zero mean unit variance and (ii) Scaling.

4.3.1 Z-Score

Z-score is a statistical measurement which describes the relation of a value to the mean of a group of values. To calculate the z-score for a group of values one has to first compute the mean and the standard deviation for the group of values. A z-score of 0 indicates that the value of the data-point is identical to 0, and a z-score of 1 would indicate that the value is 1 standard deviation from its mean. The z-score would have a positive value if the data-point lies above the mean and a negative value if the data-point lies below the mean. Formula to calculate z score:

$$Z = (x - \mu) / \text{std} \quad (4.1)$$

Here μ is the mean of the dataset , and std is the standard deviation of the dataset. In our case the data-points would be the pixel intensity values of the image. We have applied z-score at the slice-level i.e. mean and standard deviation are calculated for each slice and then z-score is calculated using this mean and standard deviation.

4.3.2 Scaling

The second method used in the pre-processing step is scaling. In this method we divide all the data-points with a scalar value, in our case 512.00.

4.4 Evaluation Metrics

The metrics that are widely used in the super resolution literature are MSE, PSNR and SSIM [?]. All three metrics are full-reference metrics, which means that they

require a complete reference image for comparison [?].

4.4.1 PSNR

The MSE is a quantitative similarity measure and it depends on the pixel intensity range. It compares the pixel wise intensity between two images and does not match the human perception. This is one of the reasons that degraded images which are clearly different can have similar MSE scores Formula for MSE is given as :

$$e_{MSE} = \frac{1}{KL} \sum_{l=1}^K \sum_{k=1}^L [\hat{f}(l,k) - f(l,k)]^2 \quad (4.2)$$

Here, \hat{f} is the original or the reference image and f is the degraded image. Peak signal to Noise Ratio (PSNR) is another highly used metric used in the literature. PSNR overcomes the issues in MSE by scaling it according to the image range.

$$PSNR = -10 \log_{10} \frac{e_{MSE}}{S^2} \quad (4.3)$$

As MSE approaches 0, the value of PSNR reaches infinity which means that higher values of PSNR means better results. Although PSNR mitigates the problem of image pixel intensity that MSE poses [5], it is still different than the ways human perceive images.

4.4.2 SSIM

Structural similarity index is a highly used quality metric which was developed by Wang et al. [9] . SSIM is a perceptual quality metric that loosely mimics the human visual system (HVS). SSIM models image degradation by combining three components – luminance, contrast and structure.

$$S(\mathbf{x}, \mathbf{y}) = f(l(\mathbf{x}, \mathbf{y}), c(\mathbf{x}, \mathbf{y}), s(\mathbf{x}, \mathbf{y})) \quad (4.4)$$

As shown in equation (3) the structural similarity is a function of $l(\mathbf{x}, \mathbf{y})$, which is the luminance comparison function, $c(\mathbf{x}, \mathbf{y})$, which is the contrast comparison function and $s(\mathbf{x}, \mathbf{y})$ which is the structure comparison function. The luminance of an image is the mean of the pixel intensities.

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (4.5)$$

The luminance comparison function is then calculated using the μ_x and μ_y of the two images under comparison. The standard deviation of the image is estimated as the contrast for an image.

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2}. \quad (4.6)$$

Using c_x and c_y , the contrast comparison function is calculated. For the structure comparison the image is normalised using it's own standard deviation, $(x - \mu_x)/\sigma_x$ the image then has a unit standard deviation. The structure comparison function is calculated using these normalised image signals. The equation for the luminance comparison function is defined in eq. 4.7.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4.7)$$

Here C_1 is a constant to avoid instability when denominator reaches 0. The contrast comparison function is defined in eq. 4.8.

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (4.8)$$

Here C_2 is a constant to avoid instability when denominator reaches 0. The structure comparison function is defined in eq. 4.9.

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (4.9)$$

Here C_3 is a constant and is defined as $C_3 = C_2/2$.

CHAPTER 5

Experiments

The environment for the experimentation consists of a workstation (Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz) with GPU NVIDIA GeForce RTX 2080.

5.1 Dataset

We have used a CT lung dataset which is divided into 2 parts: (i) Medseg Part and (ii) Radiopedia part.

5.1.1 Medseg Part

The medseg part [8] is a dataset with 100 axial CT images from more than 40 different patients. The size of the images is 512×512 , this dataset also contains mask with the same number of slices and same size as that of its images. There are 4 channels in the mask array, where channel 0 represents the ground glass opacity, channel 1 represents consolidations, channel 2 represents 'lung others', which is mask for the remaining parts of lungs apart from GGO (ground glass opacity) and consolidations, the last channel, channel 3 represents the background, which can be inverted to obtain the complete lung mask.



Figure 5.1: Raw high resolution CT images from Medseg data-set

5.1.2 Radiopaedia Part

The second part of this data-set is called Radiopaedia part. It contains 9 segmented axial volumetric CTs from Radiopaedia, this data-set contains both positive and negative slices, there are a total of 829 slices out of which 373 are as Covid-19 positive by a radiologist and are segmented. The masks for this data-set also have 4 channels and follow the same distribution as the previous data-set. The importance of this data-set lies in the fact that we have segmentation of ground glass opacity (GGO) and consolidations available which can be used to observe the result and effects of super resolution on these regions separately.

5.2 Implementation Details

5.2.1 Training and Testing Split

We have trained our U-net network for the up-scale factor of d ($d = 4$), which is the most common up scaling factor in the medical super resolution literature. For the training part 50 CT scans from the Medseg data-set have been used, another 20 CT scans from the same data-set were used for the validation part and remaining 30 CT images were used for the testing part. Testing has also been done on the Radiopaedia data-set. the Radiopaedia data-set contains both negative as well as positive CT slices. In this study we focus on the SR results on the lungs and the various abnormalities present in the lungs and therefore, only the positive CT slices were used for the testing purpose.

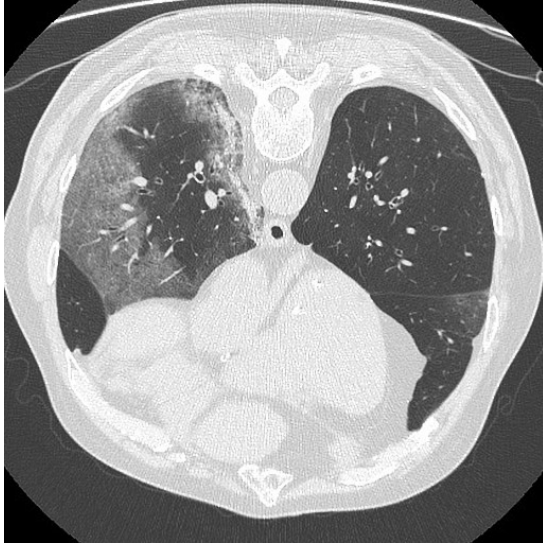
5.2.2 Pre-processing

Scaling

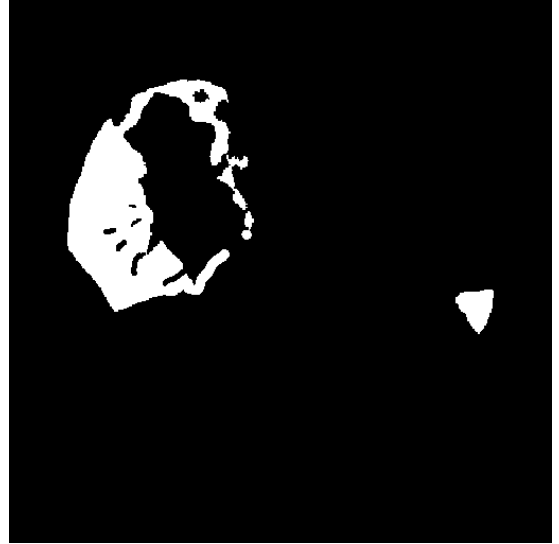
Two methods were used in the pre-processing step. The first method was a simple scaling of the pixel intensities, whereby the pixel intensities were divided by 512.00, after which lung mask is applied on the CT image to remove the background information. Final CT images that were used in the training consisted of only the lung segmented part.

Z-score

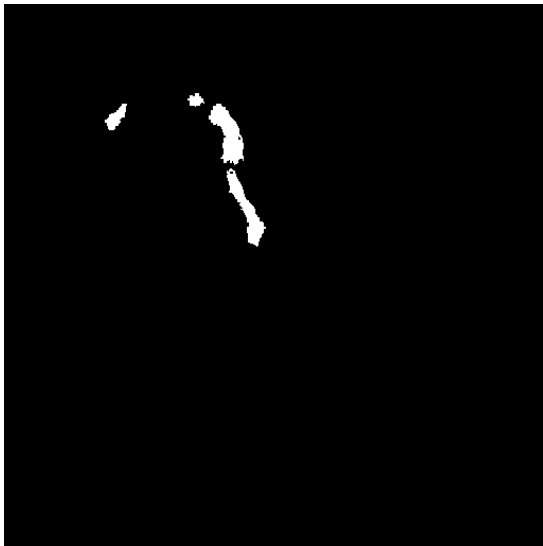
The second method used for the pre-processing part is the z-score. In this method the mean and standard deviation of the pixel intensities are calculated for the lung part. The CT images are then normalised using these mean and standard deviation values. Finally, lung segmentation mask is applied on the CT image to remove the background information and CT images which contain only the lung region are obtained.



(a)



(b)



(c)



(d)

Figure 5.2: Image (a) is a high resolution CT image from the Radiopaedia data-set. Image (b), (c) and (d) are the Ground Glass Opacity mask, the consolidations mask and the healthy lungs mask respectively.

CHAPTER 6

Results

6.1 Image Super resolution Results

In the super resolution literature, the metrics used for similarity measurements are SSIM and PSNR. These metrics are calculated for full-reference metrics, meaning they require a prior full reference image. And these metrics are computed for the entire images, so a manual localisation approach is needed. We compute local SSIM and PSNR for GGO and consolidations. For local ssim computation, we first compute the ssim map, the size of this ssim map will be 512×512 , same as that of the image size. Local masks are then applied on this ssim map, and values for that region are extracted into a 1-D array. The mean of this final 1-D array will give us the local ssim value for the region of interest.

In the case of SSIM we go on further and calculate the values of its three components namely luminance, contrast and structure. Similar to the ssim map, we obtain a map for each of the 3 components. Local masks are applied to each of these maps, and a 1-D array containing the values for that component will be returned, the mean of this 1-D array will give us the local component specific value for that region.

In table 6.1 we have reported the SSIM scores for super resolution using our U-NET model for an upscale factor of 4. Table 2 contains score for super resolution using the bicubic interpolation for an upscale factor of 4. It is clearly visible that the U-Net model proposed by us performs better than the bicubic interpolation. Table 3 and Table 4 represent super resolution performed by our U-Net model and bicubic interpolation for an upscale factor of 2 respectively. We report the SSIM score for the entire image as well as region specific SSIM score. It is reported that our U-Net model performs better than bicubic for all the local regions. We further bifurcate the SSIM metrics into its components. The luminance score and the contrast is high for both U-Net and bicubic interpolation for both the pre-processing methods. It is observed that the structure component across all the

table is highly impacted.

Table 6.1: The results of PSNR computed locally for different tissues for down - sampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic

Medseg	U-Net(Z-score)	U-Net(Scaling)	Bicubic
	PSNR	PSNR	PSNR
Lungs	21.25 dB	21.68dB	18.84 dB
Healthy Lungs	21.00 dB	21.51dB	19.12 dB
GGO	21.76 dB	22.11 dB	19.98 dB
Consolidations	19.87 dB	20.71 dB	18.00 dB

Table 6.2: The results of SSIM and score for each of it's components i.e. L(luminance), C(contrast) and S(structure) for a downsampling factor of 4 on the Medseg dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic.

Medseg	U-Net(Z-Score)				U-Net (Scaling)				Bicubic			
	SSIM	L	C	S	SSIM	L	C	S	SSIM	L	C	S
Lungs	0.66	0.99	0.97	0.68	0.65	0.99	0.97	0.67	0.53	0.98	0.95	0.56
Healthy	0.66	0.99	0.97	0.70	0.67	0.99	0.97	0.67	0.53	0.98	0.94	0.57
GGO	0.68	0.99	0.97	0.69	0.69	0.99	0.97	0.76	0.56	0.99	0.96	0.58
Conso	0.66	0.94	0.97	0.71	0.66	0.93	0.97	0.60	0.55	0.92	0.97	0.59

Table 6.3: The results of SSIM and score for each of it's components i.e. L(luminance), C(contrast) and S(structure) computed locally for different tissues for downsampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic.

Radiopedia	Unet(Z-score)				Unet(Scaling)				Bicubic			
	SSIM	L	C	S	SSIM	L	C	S	SSIM	L	C	S
Lungs	0.58	0.99	0.97	0.60	0.58	0.99	0.97	0.60	0.46	0.99	0.96	0.48
Healthy	0.62	0.99	0.97	0.64	0.61	0.99	0.96	0.63	0.49	0.99	0.95	0.52
GGO	0.54	0.99	0.97	0.55	0.54	0.99	0.97	0.56	0.43	0.99	0.96	0.44
Conso	0.60	0.99	0.97	0.56	0.58	0.97	0.97	0.60	0.46	0.95	0.96	0.48

Table 6.4: The results of PSNR computed locally for different tissues for down-sampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic

Radiopedia	U-Net(Z-score)	U-Net(Scaling)	Bicubic
	PSNR	PSNR	PSNR
Lungs	18.96 dB	18.31 dB	18.00 dB
Healthy Lungs	18.86 dB	18.24 dB	17.83 dB
GGO	18.74 dB	18.35 dB	19.56 dB
Consolidations	16.81 dB	17.06 dB	18.00 dB

Table 6.5: The results of SSIM and PSNR computed for the full image for down-sampling factor of 4 on the Medseg dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic

Medseg	Unet(scaling)	Unet(z-score)	Bicubic
SSIM	0.89	0.91	0.86
PSNR	27.97 dB	27.96 dB	25.83 dB

Table 6.6: The results of SSIM and PSNR computed for the full image for down-sampling factor of 4 on the Radiopedia dataset between U-Net(using z-score as normalization), U-Net(using scaling as normalization) and Bicubic

Radiopedia	Unet(scaling)	Unet(z-score)	Bicubic
SSIM	0.90	0.92	0.88
PSNR	26.87 dB	30.84 dB	29.87 dB



(a)



(b)



(c)



(d)

Figure 6.1: Here (a) is the HR CT image which has is of size 512x512, (b) is the LR of (a) down-sampled by a factor of 4,size 128x128, (c) is the up-scaled image using the bicubic interpolation and (d) is the SR image using our U-Net model.

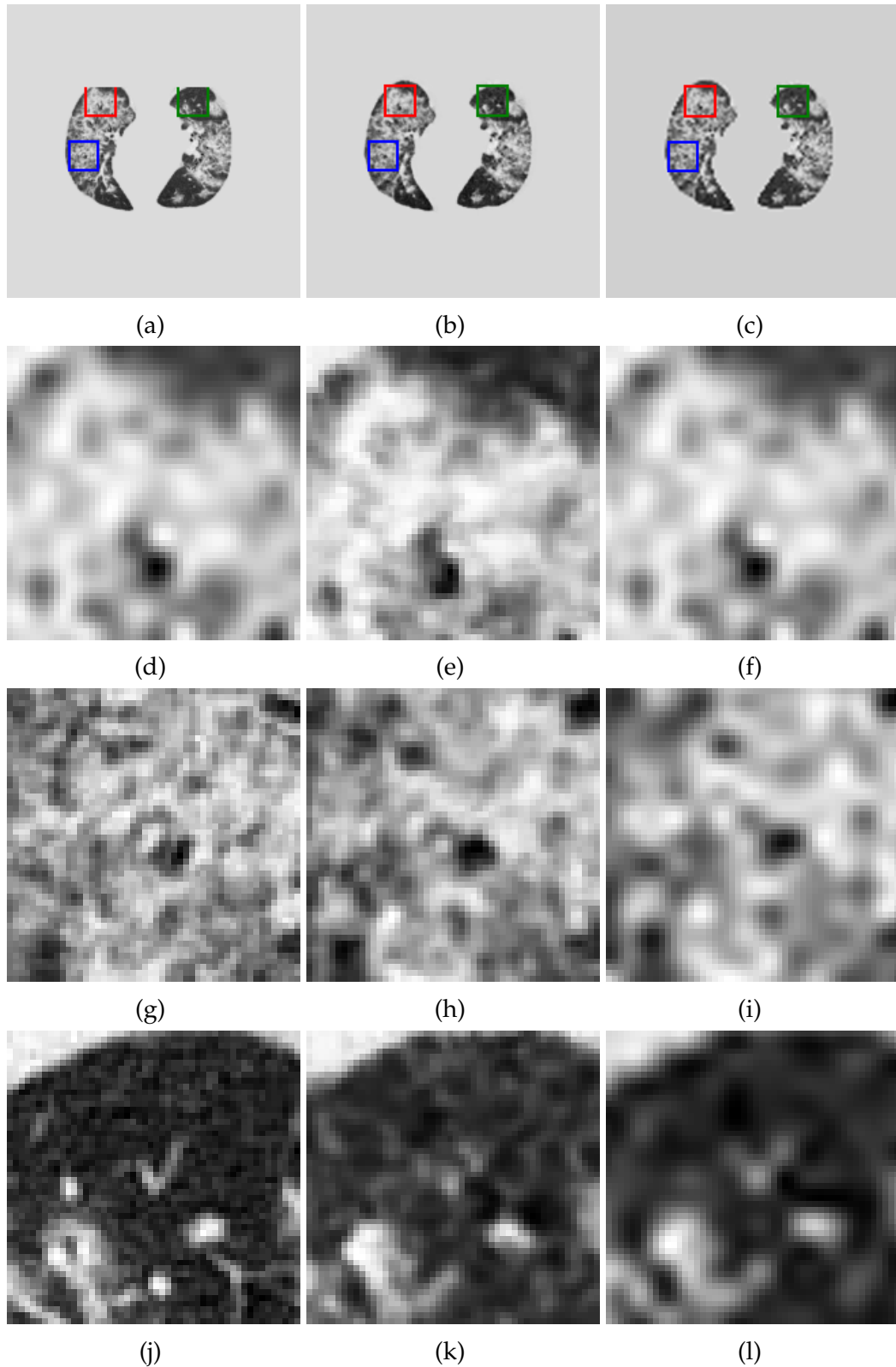


Figure 6.2: Image (a) is an HR image, size 512×512 , (b) is the SR obtained by U-Net and (c) is the image obtained using bicubic interpolation from LR of (a). (d),(e)(SSIM = 0.64) and (f)(SSIM = 0.55) are healthy lung tissue from (a), (b) and (c) respectively. (g),(h)(SSIM = 0.53) and (i)(SSIM = 0.44) are consolidation tissue from (a), (b) and (c) respectively. (j),(k)(SSIM = 0.56) and (l)(SSIM = 0.50) are GGO tissue from (a), (b) and (c) respectively.

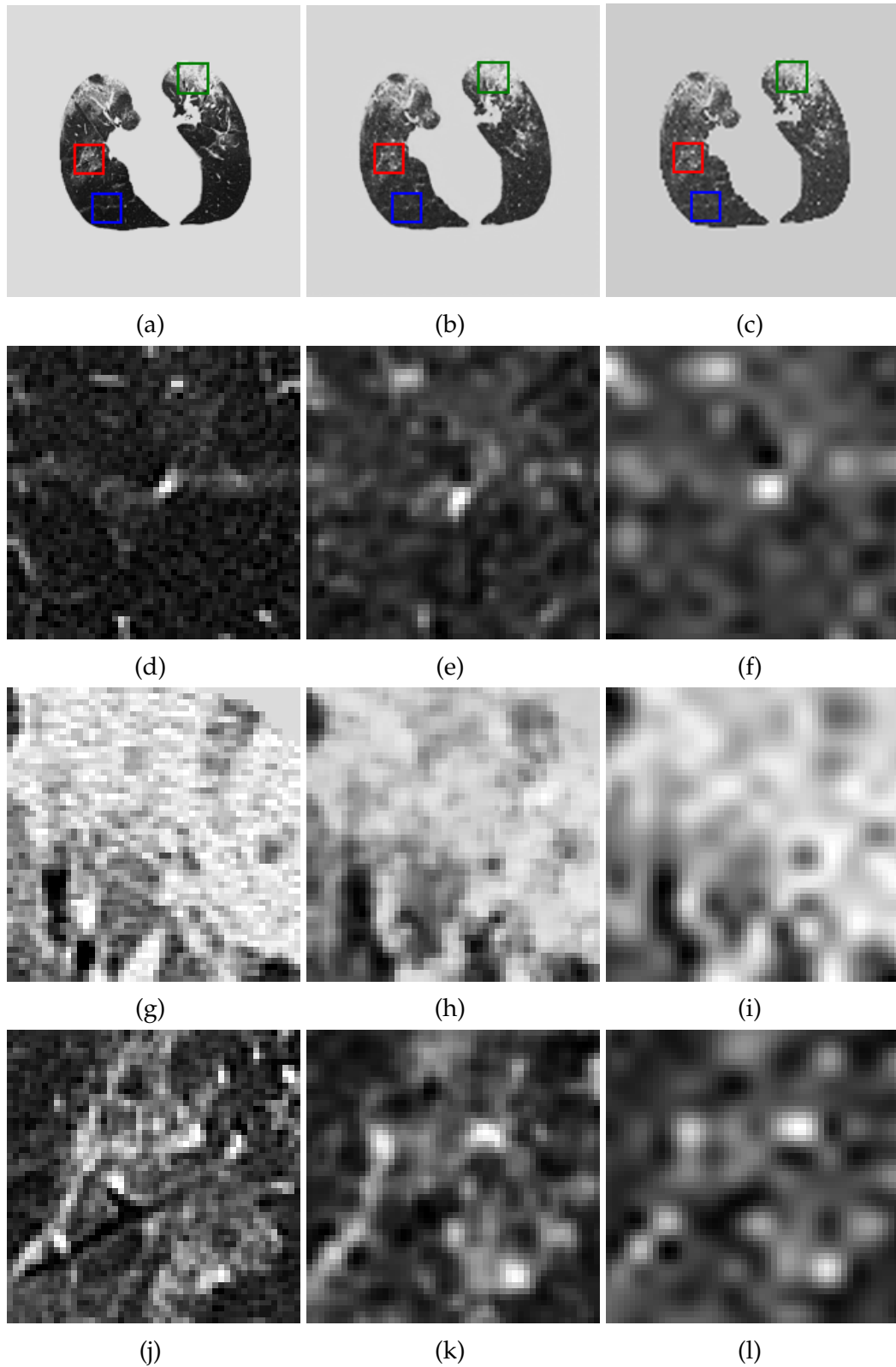


Figure 6.3: Image (a) is an HR image, size 512×512 , (b) is the SR obtained by U-Net and (c) is the image obtained using bicubic interpolation from LR of (a). (d),(e)(SSIM = 0.35) and (f)(SSIM = 0.33) are healthy lung tissue from (a), (b) and (c) respectively. (g),(h)(SSIM = 0.44) and (i)(SSIM = 0.34) are consolidation tissue from (a), (b) and (c) respectively. (j),(k)(SSIM = 0.47) and (l)(SSIM = 0.36) are GGO tissue from (a), (b) and (c) respectively.

CHAPTER 7

Discussion and Conclusion

7.1 Introduction to the chapter

This chapter discusses the results, interpret the meaning of the results and discuss it's relevance.

7.2 Discussion

As mentioned in the earlier sections we have evaluated similarity metrics for different tissues. There are 4 different types of tissues: (i)Lungs (this comprises of entire lung, i.e. the healthy tissue along with the abnormalities, (ii) Healthy Lungs (this is the region apart from the abnormalities), (iii)Ground Glass Opacity(GGO) and (iv) Consolidations. Table 6.1 contains tissue specific PSNR scores of medseg test dataset for U-Net with Scaling as normalization technique, U-Net with Z-score as the normalization technique and Bicubic interpolation. We obtain equivalent PSNR scores for all tissues for the 3 methods. We also see an improvement for 2-3 dB between Bicubic and both U-Net. Table 6.2 contains the tissue specific SSIM, Luminance, Contrast and Structure scores for the medseg dataset. We can observe an improvement of 10 % between the U-Net and bicubic. We further observe that among the 3 components of SSIM only Structure is affected while the other two components are preserved. We observe similar results in table 6.3 for radiopedia dataset. In table 6.2 no difference is observed in the scores among tissues, however in table 6.3 we observe that healthy tissue has the best SSIM score, while GGO has the least SSIM score. Tables 6.5 and 6.6 contain SSIM and PSNR scores for full images for medseg dataset and radiopedia dataset respectively. We observe that the SSIM score for medseg for full image is 0.91 in case of U-Net while the SSIM score for the tissues is between 0.66 to 0.68. This is because the CT images contain a lot of low frequency information in the background while the high frequency information is only present in the lung tissue region. It is a

difficult task for SR algorithms to retain this high frequency information and so although the tissue specific scores are low, they provide an accurate evaluation of our model compared to scores for entire image.

7.3 Conclusion

- From the results we conclude that calculating the similarity scores for full images yields a high score in comparison to the tissue specific scores.
- This is because most of the high frequency information in a chest CT is present in the lung tissue.
- Secondly, our hypothesis that using z-score as the normalization technique can mitigate the inter subject variability present in the medical field fails. Our results are equivalent for both the normalization techniques.
- An improvement of 1-2% is observed for z-score.
- In the medical field test subjects are bound to change, therefore it is imperative to perform cross data analysis.
- Upon cross data testing we see that we get equivalent results for the full image metrics calculations.
- However, the results deteriorate significantly for the tissue specific metrics measurements

References

- [1] Y. Chen, Q. Zheng, and J. Chen. Double paths network with residual information distillation for improving lung ct image super resolution. *Biomedical Signal Processing and Control*, 73:103412, 2022.
- [2] E. C. de Farias, C. di Noia, C. Han, E. Sala, M. Castelli, and L. Rundo. Impact of gan-based lesion-focused medical image super-resolution on the robustness of radiomic features. *Scientific Reports*, 11(1):21361, Nov 2021.
- [3] J.-W. Gao, S. Rizzo, L.-H. Ma, X.-Y. Qiu, A. Warth, N. Seki, M. Hasegawa, J.-W. Zou, Q. Li, M. Femia, et al. Pulmonary ground-glass opacity: computed tomography features, histopathology and molecular pathology. *Translational Lung Cancer Research*, 6(1):68, 2017.
- [4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [5] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.
- [6] S. Masoudi, S. A. A. Harmon, S. Mehralivand, S. M. Walker, H. Raviprakash, U. Bagci, P. L. Choyke, and B. Turkbey. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. *Journal of Medical Imaging*, 8(1):1 – 14, 2021.
- [7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

- [8] J. S. Suri, S. Agarwal, A. Carriero, A. Paschè, P. S. Danna, M. Columbu, L. Saba, K. Viskovic, A. Mehmedović, S. Agarwal, et al. Covlias 1.0 vs. med-seg: Artificial intelligence-based comparative study for automated covid-19 computed tomography lung segmentation in italian and croatian cohorts. *Diagnostics*, 11(12):2367, 2021.
- [9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [10] Z. Wang, J. Chen, and S. C. H. Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3365–3387, 2021.
- [11] C. M. Ward, J. Harguess, B. Crabb, and S. Parameswaran. Image quality assessment for determining efficacy and limitations of super-resolution convolutional neural network (srcnn). In *Applications of Digital Image Processing XL*, volume 10396, page 1039605. International Society for Optics and Photonics, 2017.
- [12] C. You, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, and Z. Zhang. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). 39(1):188–203, jan 2020.
- [13] H. Yu, D. Liu, H. Shi, H. Yu, Z. Wang, X. Wang, B. Cross, M. Bramler, and T. S. Huang. Computed tomography super-resolution using convolutional neural networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3944–3948, 2017.
- [14] J. Zhu, G. Yang, and P. Lio. How can we make gan perform better in single medical image super-resolution? a lesion focused multi-scale approach. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1669–1673, 2019.