Tissue-Specific Analysis of Super Resolution Methods for Medical Images

by

NISARG DOSHI
202111033

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

May, 2023

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,
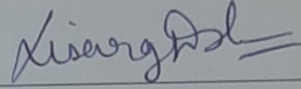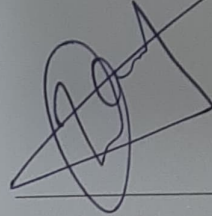
ii) due acknowledgment has been made in the text to all the reference material used.

_____

Nisarg Doshi

Certificate

This is to certify that the thesis work entitled "Tissue-Specific Analysis of Super Resolution Methods for Medical Images" has been carried out by Nisarg Doshi for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my supervision.

_____

Prof. Bakul Gohel
Thesis Supervisor

Acknowledgments

I want to express my sincere gratitude to my thesis supervisor, Professor Bakul Gohel. His invaluable guidance, expertise, dedication, and enthusiasm have shaped my research. He has been an excellent mentor, always available to help me with any questions or concerns I had. His subject knowledge and insightful feedback have helped me improve my work and become a better researcher.

I would also like to thank Vaidik Patel for his work whose work provided a solid foundation for my research and inspired me to pursue new ideas.

I thank all the faculty members at Dhirubhai Ambani Institute of Information and Communication Technology for providing the required courses related to Machine Learning. Their knowledge and expertise have been invaluable in shaping my understanding of the field. I also want to express my appreciation to the staff members at the college for their assistance throughout my research.

I am deeply grateful to my family for their unwavering support, understanding, and encouragement throughout this challenging process. Finally, I would also like to thank my friends whose engaging discussions and joyful distractions helped me maintain a healthy work-life balance.

Contents

Abstract

Image super-resolution (SR) techniques are widely used in various domains to enhance the resolution of low-resolution images, producing visually appealing high-resolution versions. However, regarding medical images, SR methods must produce precise results. Therefore, a thorough evaluation of the performance of different SR methods on various tissues is essential to determine their suitability.

In particular, evaluating SR methods on region-specific organs, such as the lung, liver, and kidney in CT scans and brain in MRI scans, is essential. When these organs are individually enhanced using Bi-cubic interpolation and Modified-ESPCN methods, along with standard evaluation metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), it is observed that SR methods exhibit inferior performance on most individual regions of interest compared to the entire image. This difference in performance can lead to misleadingly high results when evaluated over the entire image, which includes irrelevant non-tissue regions.

We propose using a tissue-specific model incorporating a region-based loss function to overcome this limitation. This approach allows for a more accurate and informative evaluation of SR methods in the context of tissue-specific performance analysis for CT images.

Keywords: *image super-resolution, SSIM, ESPCN model, region-of-interest, CT Scan, MRI, tissue-specific*

List of Principal Symbols and Acronyms

AI   Artificial Intelligence

BraTS   Brain Tumor Segmentation Challenge

CLIS   COVID-19 CT Lung and Infection Segmentation

CNN   Convolutional Neural Network

CPU   Central Processing Unit

CT   Computed Tomography

dB   Decibel

DL   Deep Learning

ESPCN   Efficient Sub Pixel Convolutional Network

GAN   Generative Adversarial Network

GHz   Gigahertz

GPU   Graphics Processing Unit

HR   High-Resolution

LR   Low-Resolution

MISR   Multiple Image Super-Resolution

MRI   Magnetic Resonance Imaging

MSE   Mean Squared Error

OSIC   Open Source Imaging Consortium

PSNR   Peak Signal-to-Noise Ratio

ROI   Region of Interest

SISR Single Image Super-Resolution

SNR Signal-to-Noise Ratio

SRCNN Super-Resolution Convolutional Neural Network

SRGAN Super-Resolution Generative Adversarial Network

SR Super-Resolution

SSIM Structural Similarity Index

VDSR Very Deep Super-Resolution

List of Tables

List of Figures

CHAPTER 1

Introduction

Super-resolution is a critical process in which low-resolution (LR) images are enhanced to obtain high-resolution (HR) versions. The demand for high-resolution images in various fields, including medical imaging, has spurred extensive research. However, medical images pose unique challenges that require specialized approaches. Unlike natural images, medical images lack a canonical orientation, and pre-trained models based on real-world images may generate unrealistic patterns that affect clinical interpretation and diagnosis. This chapter introduces the research topic, aiming to address the limitations of existing super-resolution methods and improve the quality of super-resolved medical images for accurate diagnosis. The chapter presents the background of super-resolution, highlights the importance of high-resolution medical images, and discusses the challenges specific to the medical imaging domain. Additionally, it outlines the motivation for the research, emphasizing the need for specialized techniques that preserve high-frequency information and textural details in the region of interest (ROI) while excluding irrelevant regions. The organization of the thesis is also outlined, providing a roadmap for the subsequent chapters.

1.1 Background

Super-resolution is a complex process that aims to predict a high-resolution (HR) image from its low-resolution (LR) equivalent (SR). Various real-world applications require high-resolution images, such as medical imaging for clinical duties, geographic information systems, security video monitoring, and fingerprint image enhancement. However, the quality of the imaging hardware often limits the resolution of the images. Therefore, super-resolution algorithms and techniques have garnered much attention within the research community. These methods allow us to obtain HR images from their LR counterparts without needing hardware upgrades. However, super-resolution is an ill-posed problem, and the re-

constructed SR image often needs more textural details. Deep Learning and Artificial Intelligence have significantly improved the performance of SR algorithms for natural images.
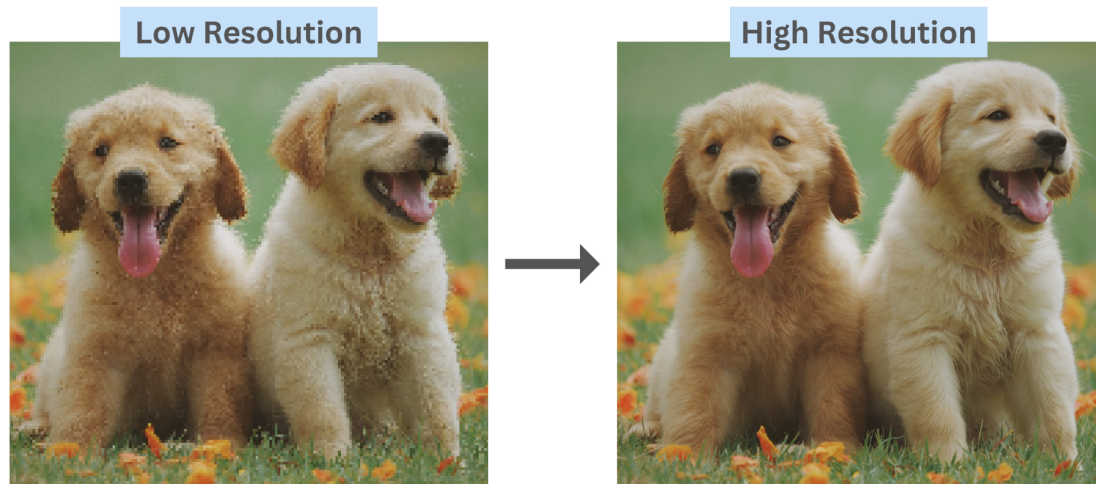

Figure 1.1: Example of SR aiming to obtain HR image when given an LR image

Though, medical images pose unique challenges that require different approaches. Unlike natural images, medical images do not typically include orientation, and objects in medical photos have no canonical orientation. For example, tissue slices in digital pathology photographs are placed on the glass without any predetermined alignment. Additionally, background tissues in a medical image can have an orientation-dependent structure, while the anomalies themselves do not. Therefore, pre-trained models based on real-world images may generate unrealistic patterns that can affect the clinical interpretation and diagnosis. [1]

In medical imaging, obtaining high-resolution (HR) images is paramount as they provide vital physiological, functional, anatomical, and metabolic information about patients. However, as with other imaging domains, hardware limitations also exist in the medical field, resulting in the acquisition of low-resolution (LR) images. Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Ultrasound are the widely used imaging modalities in medical imaging. The technical quality of the device and the conditions of the scan play an essential role in acquiring a medical image. For instance, in MRI, higher spatial resolution leads to longer scan time, less spatial coverage, and a worse signal-to-noise ratio (SNR). One of the most commonly used imaging technologies for screening, diagnosis, and image-guided intervention in the medical field is X-Ray computed tomography (CT). However, as CT scanners use ionizing radiation, high radiation doses harm the human body. Therefore, in clinical practice, CT scans are obtained at

low doses of radiation.

One way to increase the resolution of CT images is by using more advanced hardware components, which include a narrow focal spot size X-Ray tube, small-pitch detector elements, and improved mechanical precision for CT scanning. However, these hardware-based approaches are often expensive and can increase radiation exposure, leading to genetic damage and potential cancer. Alternatively, computational approaches can increase image resolution without upgrading the hardware. Super-resolution techniques have shown promising results in overcoming hardware limitations and increasing the spatial resolution of medical images.

Super-resolution (SR) has been a prominent research topic in image processing. While SR can be applied in frequency and spatial domains, most research has focused on the latter. Spatial domain approaches for converting LR images to HR images include using interpolation techniques, such as bicubic interpolation. Two common approaches for SR in the spatial domain are Multiple Image SR and Single Image SR. While the former uses multiple LR images and a single target HR image, the latter learns the mapping from a single LR image to the HR image to construct the SR image. However, multiple image-based SR algorithms tend to smooth out the images, losing important high-frequency information and textural details crucial for medical imaging. Deep learning-based approaches have recently shown significant improvement in Single Image SR due to increased computational power and availability of big data. This is particularly important for medical imaging, where retaining high-frequency information and textural details is crucial to avoid misdiagnosis.

1.2 Motivation

Medical images are unique and different from real-world images due to the specific region of interest (ROI) that they typically have. Other organs/tissues and a constant pixel background accompany the ROI. Super-resolution techniques applied to medical images should be evaluated to exclude non-important regions and backgrounds, as the enhancement effects can vary across different parts of the image. The inclusion of irrelevant regions in the evaluation of super-resolution metrics can lead to inaccurate results. These differences in medical images emphasize the need for specialized super-resolution techniques that can enhance the ROI and preserve the high-frequency information and textural details crucial in medical diagnosis.

Our current work primarily evaluates the performance of Super Resolution (SR) methods, including Bicubic interpolation and Modified Efficient Sub-Pixel Convolution Network (Modified ESPCN), in specific tissue regions. We are using standard loss functions for our analysis: Mean Squared Error (MSE) Loss and Structural Similarity Index (SSIM) Loss with the Modified ESPCN model. In order to assess the efficiency of these techniques, we are utilizing popular metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). These metrics rely on a complete reference image for evaluation and are commonly used for natural images but may not be suitable for medical images. Medical images often have a constant pixel intensity in the background, which contains low-frequency information and may be redundant. Therefore, retaining this information is relatively easy for deep learning (DL) methods. So, calculating the metric score over the entire image will yield a higher score which is spurious. Therefore, this will not give an accurate evaluation of SR methods. We compared the localized results with the complete image results to prove this hypothesis. We have also proposed a custom loss-based Deep Learning (DL) approach to address the limitations of the current SR methods and improve the performance in localized regions, aiming to improve the super-resolved image quality and assist in accurate diagnosis.

1.3   Organization of thesis

The thesis is organized as follows:

Chapter 2 provides an overview of the fundamental aspects of image super-resolution. It includes a detailed explanation of image super-resolution methods, types, processes, and metrics used for evaluation.

Chapter 3 presents a literature survey of related works in the field of image super-resolution, with a specific focus on medical image super-resolution.

Chapter 4 explores the methodology used for conducting experiments. This includes the description of the datasets used, preprocessing steps, model architecture, and the usage of evaluation metrics.

Chapter 5 discusses the experimental setup and implementation details, accompanied by supporting results.

Chapter 6 analyzes the observations derived from the results and provides a comprehensive discussion of their implications. The chapter concludes the work undertaken and outlines future directions for research.

CHAPTER 2

Fundamentals

This chapter provides an overview of image super-resolution, which is the task of enhancing the spatial resolution of low-resolution images. It explains the goal of recovering high-frequency components and fine textures lost during image acquisition or downscaling. The chapter introduces the concept of degradation functions and discusses the challenges in accurately modeling them. It covers different scenarios, focusing on single-image super-resolution and presents various techniques including interpolation-based, sparse representation-based, and deep learning-based methods. The importance of evaluation metrics such as PSNR and SSIM for assessing the performance of super-resolution algorithms is also emphasized.

2.1   Image Super Resolution

Image super-resolution (SR) is a critical task in the field of computer vision that aims to enhance the spatial resolution of a given low-resolution (LR) image to generate a high-resolution (HR) image with finer details. The goal is to recover the high-frequency components and fine textures lost during the image acquisition or downscaling. Image super-resolution has widespread applications in various domains, including medical imaging, surveillance systems, remote sensing, and digital photography. This chapter provides an overview of the fundamental concepts and techniques used in image super-resolution.

Let us consider an LR image denoted as $y \in \mathbb{R}^{m \times n}$ and its corresponding HR counterpart denoted as $x \in \mathbb{R}^{M \times N}$, where $M > m$ and $N > n$. Image super-resolution aims to estimate the HR image, x', from the given LR image, y, using appropriate algorithms and techniques. The process involves reconstructing the missing high-frequency details to produce a visually appealing and perceptually realistic HR image.

The LR image is obtained by a downscaling operation from the HR image.

To describe the relationship between these images, we can define a degradation function F, as expressed in Equation 2.1.

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

The above equation introduces the parameter θ_η as the degradation parameter. While, in theory, this parameter can represent various factors such as noise, it is not practically feasible to accurately model such a degradation function with a single parameter. We assume that this degradation parameter contributes to the reduction in quality of the high-resolution image x, resulting in the generation of a low-resolution image y. However, in reality, we are presented with a low-resolution image y, and our goal is to develop a Super-Resolution method that effectively enhances its quality and reconstructs the high-resolution image x.

Single-image super-resolution (SISR) is the most common scenario where only a single LR image is available for the super-resolution task. SISR aims to estimate an HR image visually similar to the ground truth HR image. The problem can be formulated as follows:

$$x' = F^{-1}(y; \theta_\zeta) \tag{2.2}$$

The parameters θ_ζ in the function F^{-1} represent the parameters of the super-resolution method. The resulting image, denoted as x', approximates the high-resolution image x. While obtaining the exact image x is an ideal scenario, it is not a practical approach. The equations presented above outline the overall concept of the super-resolution task, which aims to develop an approximation method to obtain a high-resolution image. However, the degradation process is highly intricate and not fully understood. It involves several parameters, such as blurring, scaling, noise, and other factors. The mapping function F^{-1} can be realized using various techniques, including interpolation-based methods, sparse representation-based methods, and deep learning-based methods.

Interpolation-based methods aim to increase the resolution of the LR image by introducing additional pixels between the existing pixels. These techniques utilize simple mathematical operations to estimate the missing high-frequency details. Bilinear interpolation, bicubic interpolation, and Lanczos interpolation are commonly used interpolation methods. However, these methods need more fine details due to the limited information in the LR image.

Sparse representation-based methods leverage the assumption that a linear combination of a dictionary of HR patches can sparsely represent HR patches. These methods seek a sparse representation of LR patches in the HR dictionary

and use this representation to estimate the HR image. One popular sparse representation-based method is the sparse coding-based super-resolution algorithm.

Deep learning-based methods have significantly succeeded in various computer vision tasks, including image super-resolution. These methods utilize deep neural networks to learn the mapping function between LR and HR images directly from a large set of training pairs. The most popular architecture in deep learning-based super-resolution is the convolutional neural network (CNN). The CNN architecture consists of multiple convolutional layers for feature extraction and reconstruction. One example is the Super-Resolution Convolutional Neural Network (SRCNN), one of the simplest CNN for SR and consists of three main layers: the patch extraction layer, the non-linear mapping layer, and the reconstruction layer. The objective is to learn the optimal parameters of the network to minimize the reconstruction error between the estimated HR image and the ground truth HR image.

Several evaluation metrics are commonly used to evaluate the performance of super-resolution algorithms. These metrics compare the estimated HR image, $\hat{HR}$, with the ground truth HR image, HR, and provide objective measures of the quality of the super-resolved image.

PSNR measures the quality of the estimated HR image by computing the ratio of the maximum possible pixel intensity to the mean squared error (MSE) between the $\hat{HR}$ and HR images.

SSIM evaluates the structural similarity between the estimated and ground truth HR images. It considers luminance, contrast, and structural information. The SSIM index ranges from 0 to 1, with a value of 1 indicating an exact match. SSIM consists of three components – luminance, contrast, and structure.

2.2 Efficient Sub-Pixel Convolutional Network

There are certain limitations associated with convolutional neural network (CNN) approaches, such as SRCNN, FSRCNN, and VDSR. Firstly, these CNN-based methods require the utilization of interpolation techniques like bicubic interpolation to upscale the low-resolution (LR) image. Secondly, these approaches perform the upscaling operation either before or at the initial layer of the network. In essence, the CNN-based approach directly applies the convolutional neural network to the upsampled LR image, leading to increased computational complexity and increased memory requirements.

ESPCN tackles these issues by incorporating an efficient sub-pixel convolu-

tional layer into the CNN network architecture. Unlike previous methods, ESPCN performs the upscaling operation at the very end of the network. Consequently, the LR image of smaller dimensions is directly inputted into the network, eliminating the need for interpolation techniques. This approach allows the network to learn a more accurate LR to HR mapping compared to the conventional method of upscaling with interpolation filters prior to network input. ESPCN can utilize smaller filter sizes to extract relevant features by working with the reduced input image size. As a result, the computational complexity and memory requirements are significantly reduced, enhancing overall efficiency.
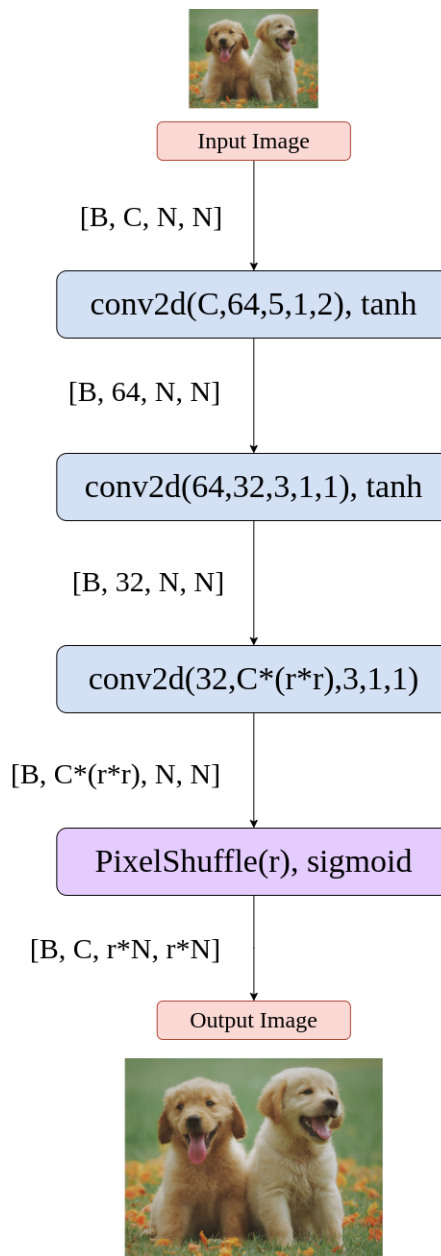

Figure 2.1: ESPCN Model

Figure 2.1 illustrates the network architecture of ESPCN where B represents the batch size, C represents the channel size, N represents the image size, and r represents the upscale factor. The conv2d refers to the convolutional layer with parameters (cin, cout, kernel size, stride, padding). Additionally, tanh denotes the tanh activation function, while sigmoid represents the sigmoid activation function. Finally, PixelShuffle corresponds to the Sub-Pixel Convolution Layer. Assuming that the network consists of L layers, the initial L-1 layers are convolutional layers responsible for extracting feature maps from the input LR images. The final layer, known as the efficient sub-pixel convolutional layer, is specifically designed to restore the output image size based on a predetermined upscale factor.

The sub-pixel convolution technique can be utilized in super-resolution (SR) models to generate high-resolution images. Unlike traditional deconvolution methods that involve padding with zeros, sub-pixel convolution avoids the need for padding. Instead, it employs pixel shuffle at the final layer of the network to reconstruct the LR image. This process combines each pixel from multiple-channel feature maps into a square area in the output image, resulting in each pixel on the feature maps representing a sub-pixel on the final generated output image. Sub-pixel convolution comprises two key steps: a regular convolution operation and pixel rearrangement. The network's last layer must have an output channel of $C \times r \times r$ where C is the channel size and r is the upscale factor to ensure the total number of pixels matches the desired high-resolution image. In the ESPCN network, the convolutional layers implicitly learn the interpolation process, eliminating the need for explicit interpolation methods. Furthermore, since the convolution operations are performed on smaller LR images, the network achieves greater efficiency. The decision to choose ESPCN from all available CNN models and modify it for our analysis due to advantages like reduced computational complexity and implicit learning of interpolation.

Chapter Summary

The chapter provided an overview of image super-resolution, which aimed to enhance the spatial resolution of low-resolution images. It discussed the primary objective of recovering high-frequency components and fine textures that may have been lost during image acquisition or downscaling. The chapter explored fundamental concepts, including degradation functions and the challenges associated with accurately modeling them. It presented different scenarios such

as single-image super-resolution and delved into various techniques, including interpolation-based, sparse representation-based, and deep learning-based approaches. The chapter also highlighted the importance of evaluation metrics such as PSNR and SSIM for assessing the performance of super-resolution algorithms.

CHAPTER 3

Literature Review

The Literature Review chapter comprehensively overviews conventional and deep learning-based image super-resolution techniques. Conventional methods, such as Bayesian frameworks, sparse representation, and multi-frame fusion, have significantly contributed to the field. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have gained prominence in both Single Image Super-Resolution (SISR) and Multiple Image Super-Resolution (MISR). Pioneering models like SRCNN, VDSR, RDN, and SRGAN have demonstrated improved performance in enhancing image resolution. The chapter further focuses on applying these techniques in medical imaging, particularly in MRI and CT, highlighting the challenges and advancements in synthesizing high-resolution medical images.

3.1 Conventional Image Super Resolution

Various conventional methods have been extensively studied and have significantly contributed to the image super-resolution field. One unique approach introduced by Huang et al. [2] employed a Bayesian framework with a statistical prior based on maximum a posteriori estimation, demonstrating promising results. Another effective method by Yang et al. [3] utilized sparse representation, employing an over-complete dictionary to represent low-resolution images and recovering high-resolution details through sparse coding.

In the domain of image denoising and super-resolution, Buades et al. [4] proposed a non-local means filter, which was further extended by Darbon et al. [5] to enhance high-frequency details during the upsampling process by leveraging self-similarity in images. Multi-frame image super-resolution techniques have also garnered attention, such as the work by Farsiu et al. [6], where multiple low-resolution images of the same scene were registered and fused to generate a high-resolution output.

Example-based learning approaches have demonstrated their effectiveness in image super-resolution. Freeman et al. [7] introduced a statistical method that learned a mapping between low-resolution and high-resolution image patches using an extensive training set, facilitating the reconstruction of high-frequency details. Additionally, methods based on total variation (TV) regularization has been proposed, such as the work by Elad et al. [8], which incorporated TV constraints to encourage piecewise smoothness in the reconstructed high-resolution images.

3.2 Deep Learning based Image Super Resolution

Image super-resolution techniques can be categorized into two main classes: Single Image Super-Resolution (SISR) and Multiple Image Super-Resolution (MISR). Chen et al. [9] extensively review real-world SISR, covering essential datasets, assessment metrics, and effective methods in this domain. The authors emphasize the significance of super-resolution in improving image quality and highlight the advancements facilitated by deep learning approaches.

In the context of SISR, Convolutional Neural Networks (CNNs) have gained prominence. Dong et al. [10] introduce the Super-Resolution Convolutional Neural Network (SRCNN), among the pioneering methods to utilize deep learning for SISR. The SRCNN successfully enhances single images, leveraging its three-layer architecture. Building upon this foundation, researchers have proposed more effective and deeper architectures. Kim et al. [11] extend the network depth with the Very Deep SR (VDSR) model, which incorporates additional convolutional layers and residual learning to enhance the performance of SISR further.

Zhang et al. [12] propose the Residual Dense Network (RDN) to extract hierarchical features and improve the overall outcome of SISR. By incorporating dense residual connections after residual learning, the RDN achieves superior results. Ledig et al. [13] introduce the SRGAN approach, employing Generative Adversarial Networks (GANs) to generate perceptually pleasing super-resolution images. However, it is worth noting that GAN-based models with excessively deep architectures may produce artefacts or textures that lack authenticity.

In the case of MISR, Kawulok et al. [14] highlight the benefits of information fusion, which enables higher reconstruction accuracy. They introduce a novel method that combines multiple-image fusion with deep network-based learning, allowing for mapping low-resolution images to high-resolution counterparts. These techniques have found applications in general image processing and have

proven helpful in the medical domain for enhancing medical images.

3.2.1 Medical Image Super Resolution

Image super-resolution (SR) in medical imaging is a well-studied specialized field, particularly in Magnetic Resonance Imaging (MRI) and Computed Tomography (CT). A technical review by Prince et al. [15] provides comprehensive insights into medical image super-resolution, covering various techniques and approaches used in the field. The review discusses the challenges and advancements in synthesizing high-resolution medical images. Yamashita and Markov [16] conducted a review focusing on medical image enhancement using super-resolution methods. The study explores different techniques and algorithms used to improve the resolution of medical images, highlighting their applications and effectiveness. Isaac and Kulkarni [17] presented a study on super-resolution techniques for medical image processing. The authors discussed various methods used to enhance the resolution of medical images, including interpolation-based techniques, patch-based methods, and deep-learning approaches. They provide insights into the strengths and limitations of different approaches and highlight the challenges specific to medical image super-resolution. Yu et al. [18] proposed a CNN-based approach for super-resolution of single and multiple slice CT images. Their method utilizes residual learning, leveraging the power of deep neural networks to reconstruct high-resolution CT scans. Rajeshwari and Shyamala [19] focused on super-resolution of chest CT images using deep-learning network models. They applied deep learning techniques, specifically convolutional neural networks (CNNs), to enhance the resolution of chest CT scans. Umehara et al. [20] investigated the application of super-resolution convolutional neural networks for enhancing image resolution in chest CT. Their study focused on developing a network architecture tailored to chest CT images. Kolarik et al. [21] proposed an unbalanced 3D Dense-U-Net network for super-resolution of MRI brain images. Their study focused on developing a specialized network architecture that generates high-resolution MRI brain images. Recently, convolutional neural network (CNN)-based approaches have been increasingly used for medical image SR applications in different imaging modalities, e.g. MRI, CT, X-ray, fundoscopy and histopathological images and showed impressive performance. [22]

Most of the literature on medical radio imaging has evaluated SR methods by measuring the similarity metrics over entire images [22], [20] - [23]. In contrast, in the present work, we aimed to look at tissue or region-specific analysis of the performance measure, which are PSNR and SSIM for SR approaches like Bicubic

interpolation and Modified-ESPCN deep learning model.

Chapter Summary

The Literature Review chapter provided an overview of studies on image super-resolution methods, with a specific focus on their application in medical imaging. It extensively discussed the advancements in the field using Convolutional Neural Networks (CNNs) for both Single Image Super-Resolution (SISR) and Multiple Image Super-Resolution (MISR). The chapter highlighted the significant progress facilitated by deep learning approaches, including pioneering methods such as SRCNN, VDSR, RDN, and SRGAN. It emphasized the application of these methods in medical imaging, addressing the challenges encountered and showcasing their effectiveness in enhancing resolution in domains such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT).

CHAPTER 4
Methodology

In the methodology chapter, we address the challenges related to dataset availability, variations in imaging protocols, and the scarcity of expert-annotated masks in medical imaging. We discuss the three publicly available datasets used in our research: CT-ORG, COVID-19 CT Lung and Infection Segmentation and OSIC Pulmonary Fibrosis Progression. Preprocessing steps were applied to standardise the data, including down-sampling, scaling, and normalization. We present the modified ESPCN model, incorporating additional convolutional layers and skip connections, as our chosen architecture for image super-resolution. Finally, we describe the performance metrics used, namely Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), to evaluate the quality and accuracy of the reconstructed images.

4.1   Dataset and Pre-processing

During thesis work, we encountered several challenges related to datasets. One of the primary issues was the scarcity and limited availability of high-quality medical imaging data. A considerable hurdle was obtaining a diverse and comprehensive dataset for training and evaluation purposes. Due to privacy concerns and institutional policies, access to large-scale, annotated medical image datasets was restricted.

Furthermore, the dataset suffered from significant variations in imaging protocols and acquisition parameters. The medical images were acquired from different scanning devices, resulting in variations in image resolution, noise levels, and contrast. This heterogeneity posed a challenge regarding data standardization and normalization, as the images needed to be processed and appropriately aligned to ensure consistent and reliable training. In addition, the presence of noise and artefacts in the dataset, caused by patient motion or equipment limitations, further complicated the training process.

Another issue was that we needed the datasets with corresponding masks for the particular organ or tissue for our analysis of tissue-specific regions. Mask annotations in such medical datasets are usually done by expert radiologists separately. Such datasets with high-quality mask annotations by expert radiologists for particular organs of our interest were very scarce.

After a great deal of research, we were able to procure three datasets across the imaging modalities CT scan, for our purpose of analysis, which were:

- CT-ORG Dataset [24]

- COVID-19 CT Lung and Infection Segmentation Dataset [25]

- OSIC Pulmonary Fibrosis Progression Dataset [26]

These datasets we utilized were publicly available and consisted of three-dimensional volumes, that is, CT scans of patients. This three-dimensional nature allowed us to slice the volumes as needed for our analysis. Moreover, these datasets included tissue masks labelled alongside the images or separately provided by expert radiologists. These tissue masks were crucial for our research as they helped identify and isolate specific organs or tissues for further analysis.

Most studies evaluated the SR methods for down-sample factors x2, x3 or x4 [27] [15] [16] [17]. We applied down-sampling to the slices for all the datasets using scaling factors 2, 3, and 4. This down-sampling process involved reducing the resolution of high-resolution (HR) slices to obtain corresponding low-resolution (LR) slices. By down-sampling the HR slices by x2, x3, and x4 factors, we created LR slices that served as input for various super-resolution methods. This approach allowed us to simulate different levels of image degradation and evaluate the super-resolution techniques' performance in enhancing the image's resolution and quality.

We also performed preprocessing on the slices of all the datasets, which usually involved appropriate scaling and normalization techniques. This step aimed to standardize the pixel values and ensure consistency across the dataset. Scaling was applied to adjust the pixel intensities to a desired range, typically between 0 and 1. Normalization refined the data by centring the pixel values around zero and scaling them to a standard deviation of 1. The slices were also centre-cropped to maintain uniformity before applying super-resolution methods for all the datasets. By applying these preprocessing techniques, we aimed to mitigate variations in pixel intensities that could arise from different acquisition settings and scanner sensitivities. This helps to enhance the comparability and generalizability of the data, facilitating more effective analysis and training of deep learn-

ing models for our super-resolution tasks. More details about each of the datasets are mentioned below.

4.1.1   CT-ORG

CT-ORG dataset comprises imaging data obtained from publicly available data and data collected from Stanford Healthcare. The dataset also consists of labelled masks annotated by experts. The dataset consists of 140 CT scans with each slice of shape 512x512 pixels in the scans and having annotations for five organs in three-dimensional space: lung, bones, liver, kidneys, and bladder. While a subset of the scans also includes brain annotations, we excluded them in our analysis because of fewer annotations.

Different CTs had different spatial resolutions. Most CTs have a spatial resolution higher than one mm in an axial plane. Therefore, each CT was resized using Bicubic interpolation to make one mm spatial resolution in the axial plane. Using this dataset, we evaluated the SR methods for the down-sample factor of x4. To evaluate various super-resolution methods, we later down-sample the CT slices by factors of 2, 3 and 4, followed by preprocessing steps.

4.1.2   COVID-19 CT Lung and Infection Segmentation (CLIS)

COVID-19 CT Lung and Infection Segmentation (CLIS) dataset consist of 20 CT scans, with each slice having a height and width of 512. They are labelled for lung segmentation and infection detection. The labelling process involved two radiologists assigning labels to the left lung, right lung, and areas of infection within the scans, which an experienced radiologist further verified. This dataset does not present any challenges related to different spatial resolutions. Therefore, the axial slices can be directly utilized for our work without additional adjustments.

Similar to the CT-ORG dataset, we applied down-sampling to the CT slices for our analysis. The down-sampling was performed by a factor of 2, 3, and 4 on the CT slices from the dataset. These down-sampled slices, after preprocessing, were then utilized as inputs for various super-resolution methods.

4.1.3   OSIC Pulmonary Fibrosis Progression (OSIC)

The Pulmonary Fibrosis Progression Dataset was provided by the Open Source Imaging Consortium (OSIC) as part of a Kaggle-administered the competition AI competition aimed at predicting lung function decline in individuals with pul-

monary fibrosis. This dataset comprises chest CT scans and associated clinical information for a group of patients. Precisely, the dataset consists of a total of 200 CT scans in DICOM format [28].

Dr. Konya [29] has manually annotated lung masks for 110 scans from the dataset, which does not originally include organ masks. These annotations were done separately for slices with a height of 512 and width of 512. This dataset was also downsampled by a factor of 2, 3 and 4 to evaluate SR approaches followed by scaling and normalization of the data.

4.2 Model Architecture

Several CNN-based approaches have been proposed for image super-resolution tasks in the reviewed literature. The most commonly employed deep learning architectures include Super-Resolution Convolutional Neural Network (SRCNN), Efficient Sub-Pixel Convolutional Network (ESPCN), Very Deep Super Resolution Model (VDSR), and Super Resolution Generative Adversarial Network (SRGAN). Given that our research focuses on analyzing the tissue-specific performance of super-resolution techniques and aiming to enhance their capabilities, we selected and modified the ESPCN model for implementation as it is a lightweight and resource-efficient CNN architecture [30].

To further improve the classical ESPCN model, we modified it by incorporating two additional convolutional layers and integrated skip connections within our modified implementation as illustrated by Figure 4.1. The additional convolutional layers can capture more complex and high-level features from the input data. This deeper representation allows the model to learn and exploit more intricate patterns and details. Adding more convolutional layers allows the modified model to have a larger receptive field, enabling it to consider a broader context when making predictions. Including skip connections allow for direct connections between layers at different depths. This0 facilitates the flow of information from earlier layers to later layers, enabling the model to access both low-level and high-level details simultaneously. By preserving and reusing information from shallower layers, the modified model can retain important image characteristics and prevent information loss during the up-sampling process.
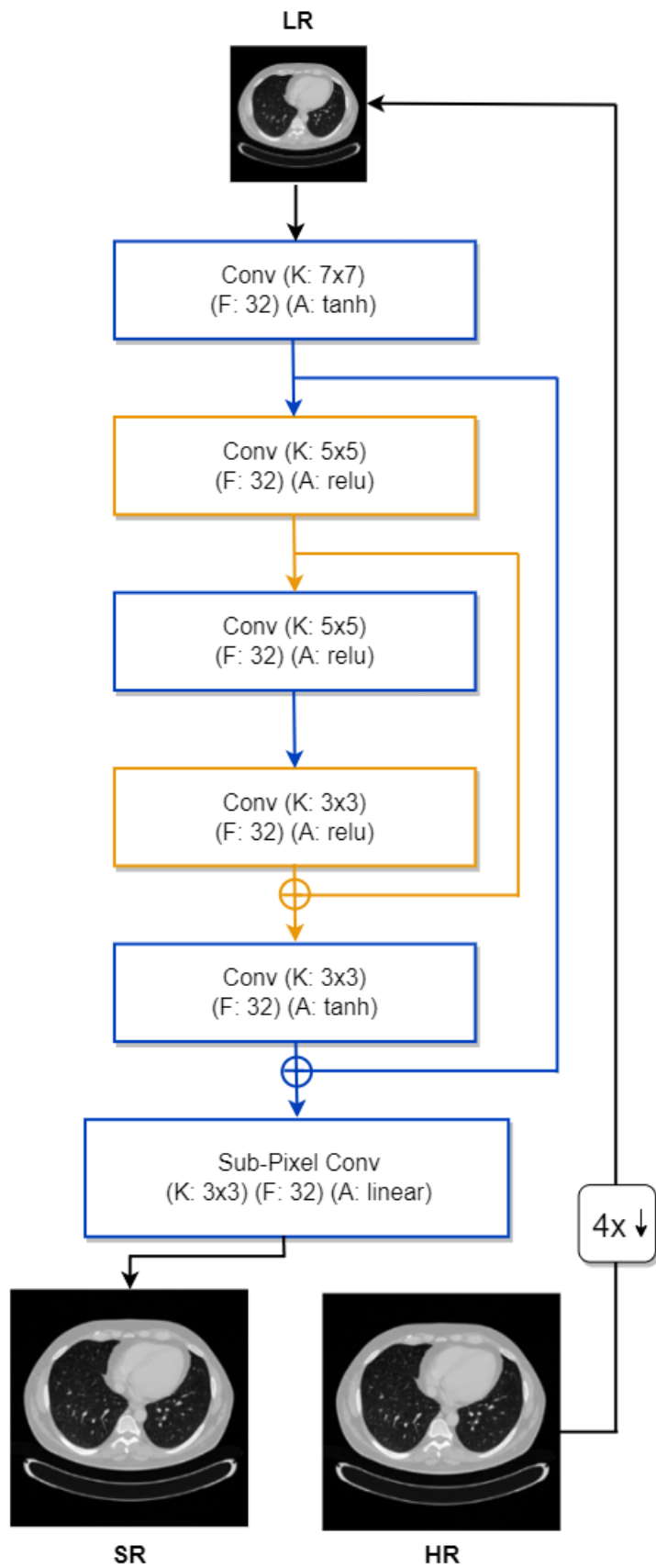
Figure 4.1: Modified ESPCN Architecture

4.3 Performance Metrics

Performance metrics play a crucial role in evaluating the effectiveness of image super-resolution techniques. These metrics quantify the quality of the generated high-resolution images by comparing them with the ground truth or reference images. Two widely used metrics for evaluating super-resolution performance are Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

In the field of super-resolution tasks, it is common practice to utilize both the PSNR (Peak Signal-to-Noise Ratio) and the SSIM (Structural Similarity Index) metrics in conjunction to evaluate the performance of various algorithms. These metrics provide complementary insights into different aspects of the reconstructed images. The combination of PSNR and SSIM allows us to evaluate the performance of super-resolution methods from multiple perspectives, considering both quantitative accuracy and perceptual quality.

4.3.1 Peak Signal-to-Noise Ratio (PSNR)

The Peak Signal-to-Noise Ratio or PSNR metric is commonly used to assess the quality and fidelity of the reconstructed high-resolution images. PSNR quantifies the level of distortion or error between the reconstructed and the ground truth image, providing a numerical measure of the reconstruction accuracy.

The calculation of PSNR relies on the Mean Squared Error (MSE), which quantifies the average squared error between the pixel values of the reconstructed image and the corresponding pixel values of the ground truth image. The MSE is mathematically defined as:

$$MSE = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i,j) - T(i,j))^2 \tag{4.1}$$

where $I(i,j)$ and $T(i,j)$ represent the pixel values of the reconstructed image and the ground truth image at the (i,j) location, respectively, and m and n are the dimensions of the images.

PSNR is then calculated as follows:

$$PSNR = 10 \cdot \log_{10} \left(\frac{\max^2}{MSE} \right) \tag{4.2}$$

where max is the maximum possible pixel value of the images. For example, for 8-bit images, the maximum pixel value is 255.

The PSNR value is expressed in decibels (dB). A higher PSNR value indicates a

smaller distortion between the reconstructed and ground truth images, implying better reconstruction quality. Conversely, a lower PSNR value indicates a higher level of distortion or error in the reconstructed image.

It is important to note that PSNR has some limitations. It primarily focuses on pixel-wise errors and does not capture the perceptual quality of the images. Therefore, two images with similar PSNR values may appear different to the human eye. Additionally, PSNR is sensitive to changes in image resolution and scaling. Hence, it is advisable to use PSNR with other metrics, such as Structural Similarity Index (SSIM), to comprehensively evaluate image super-resolution performance.

4.3.2   Structural Similarity Index (SSIM)

The Structural Similarity Index or SSIM metric is widely used to evaluate the quality and similarity between the reconstructed high-resolution images and the ground truth images. SSIM considers structural information and perceptual similarities, offering a more comprehensive assessment of the reconstruction accuracy.

SSIM measures three components of image similarity: luminance (l), contrast (c) and structure (s).

- **Luminance Component**: The luminance component evaluates the similarity in terms of brightness and intensity between the reconstructed and ground truth images. It is computed using the mean values of the image pixels.

- **Contrast Component**: The contrast component measures the similarity in terms of the image contrast or the difference in intensity levels. It is calculated based on the standard deviations of the pixel values in the images.

- **Structure Component**: The structure component assesses the similarity of the image structure or the arrangement of edges and textures. It is determined using the covariance of the pixel values in the reconstructed and ground truth images.

The SSIM index is calculated on various windows of an image. The measure between two windows $\mathbf{x}$ and $\mathbf{y}$ of common size $N \times N$ based on the individual component functions are:

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_\mathbf{x}\mu_\mathbf{y} + c_1}{\mu_\mathbf{x}^2 + \mu_\mathbf{y}^2 + c_1} \tag{4.3}$$

$$\mathbf{c}(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_\mathbf{x}\sigma_\mathbf{y} + c_2}{\sigma_\mathbf{x}^2 + \sigma_\mathbf{y}^2 + c_2} \tag{4.4}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_\mathbf{xy} + c_3}{\sigma_\mathbf{x}\sigma_\mathbf{y} + c_3} \tag{4.5}$$

SSIM is a weighted combination of :

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = l(\mathbf{x}, \mathbf{y})^\alpha \cdot c(\mathbf{x}, \mathbf{y})^\beta \cdot s(\mathbf{x}, \mathbf{y})^\gamma \tag{4.6}$$

Setting the weights α, β, γ to 1 and $c_3 = \frac{c_2}{2}$ the formula can be reduced to:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_\mathbf{x}\mu_\mathbf{y} + c_1)(2\sigma_\mathbf{xy} + c_2)}{(\mu_\mathbf{x}^2 + \mu_\mathbf{y}^2 + c_1)(\sigma_\mathbf{x}^2 + \sigma_\mathbf{y}^2 + c_2)} \tag{4.7}$$

where:

$\mu_\mathbf{x}$ is the pixel sample mean of $\mathbf{x}$;

$\mu_\mathbf{y}$ is the pixel sample mean of $\mathbf{y}$;

$\sigma_\mathbf{x}^2$ is the variance of $\mathbf{x}$;

$\sigma_\mathbf{y}^2$ is the variance of $\mathbf{y}$;

$\sigma_\mathbf{xy}$ is the covariance of $\mathbf{x}$ and $\mathbf{y}$;

$c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are two variables to stabilize the division;

L is the dynamic range of the pixel-values (typically $2^{\#bits\ per\ pixel} - 1$);

$k_1 = 0.01$ and $k_2 = 0.03$ by default.

Equation 4.7 is the final mathematical formula for calculating the similarity between two images. The SSIM value ranges between 0 and 1, where 1 indicates a perfect image similarity. A higher SSIM value implies a greater structural similarity between the reconstructed and ground truth images, indicating better reconstruction quality. [31]

Unlike the Peak Signal-to-Noise Ratio (PSNR) metric, which focuses primarily on pixel-wise errors, SSIM considers the structural and perceptual aspects of the images. This makes SSIM a valuable metric for assessing visual quality and image similarity.

4.4   Usage of metrics

To compute the Structural Similarity Index (SSIM), we utilized the SSIM function provided by the scikit-image Python library. We made customizations to the function to access the individual components of SSIM, including luminance (l), contrast (c) and structure (s) values. We employed a window size 11x11 for computing the local pixel-wise SSIM map between the reconstructed super-resolution (SR) images and the ground truth high-resolution (HR) images.

For our evaluation, we calculated the Peak Signal-to-Noise Ratio (PSNR) and SSIM scores for the entire CT slice. We also determined the PSNR and SSIM scores for specific tissues within the CT slices. To compute tissue-specific PSNR and SSIM scores, we averaged these metrics over the pixels belonging to a particular tissue. This was achieved by utilizing tissue-specific masks, allowing us to focus the evaluation on the performance of super-resolution methods for individual anatomical structures.

By employing these customized calculations and tissue-specific evaluations, we obtained a comprehensive understanding of the performance of the super-resolution techniques not only for the entire CT slice but also for specific tissue regions of interest. This approach provided valuable insights into the effectiveness of the methods in enhancing image quality and details for different anatomical structures within medical images.

Chapter Summary

In this chapter, we delved into the datasets utilized for our research on super-resolution analysis. The scarcity and heterogeneity of high-quality datasets posed obstacles, and expert-annotated masks for tissue-specific regions were limited. We procured four publicly available datasets, consisting of CT and MRI scans with corresponding tissue masks. Preprocessing steps were employed, including down-sampling, scaling, and normalization, to standardize the data. We selected the modified ESPCN model for implementation, which incorporated additional convolutional layers and skip connections. Performance evaluation utilized PSNR and SSIM metrics for the entire slices and tissue-specific regions, providing insights into the effectiveness of the super-resolution techniques.

CHAPTER 5

Experiments and Results

In this chapter, we present the experiments conducted and the corresponding results obtained. We rescaled the image intensity of the dataset and trained two models using different loss functions: mean square error (MSE) and structural similarity index metric (SSIM). Evaluations were performed using four super-resolution approaches, including bicubic interpolation and three deep learning-based methods. The results demonstrate that the Modified ESPCN approach with SSIM loss outperformed other methods, and our custom mean squared error (Custom MSE) loss function improved resolution for tissue regions. Comprehensive assessments using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) scores were conducted for all scale factors, revealing the most effective approach for enhancing image resolution.

5.1 Experiments

Before training and testing, the image intensity of the dataset was re-scaled to a range between zero and one. The training process involved training the model for a total of 100 epochs, with each epoch consisting of five steps. Each training step utilized a batch size of 16 CT slices. Within each step, 16 CT slices were randomly selected from a CT volume randomly chosen from the training dataset.

Adam optimizer with a learning rate 0.001 was used. The training was performed on a workstation equipped with an Intel(R) Xeon(R) Silver 4214R CPU operating at 2.40GHz and an NVIDIA GeForce RTX 2080ti GPU featuring 11GB of RAM. This computational setup provided the necessary resources for the training process.

Two distinct models were trained, each employing a different loss function. The first model utilized the mean square error (MSE) loss function, while the second model incorporated the structural similarity index metric (SSIM) loss [18]. By training multiple models with different loss functions, we aimed to explore and

compare their respective performances in the context of image super-resolution.

In our experiments, we conducted evaluations using four super-resolution approaches across all the datasets. As a baseline, we employed the conventional bicubic interpolation method. Additionally, we utilized three deep learning-based methods, all trained on a modified ESPCN architecture with varying loss functions. The first deep learning method employed the mean squared error (MSE) loss function, while the second incorporated the structural similarity index (SSIM) loss. These two methods served as established approaches commonly used in the literature.

In addition, we proposed a custom mean squared error (Custom MSE) loss function to address the limitations observed in existing super-resolution techniques, particularly their poor performance in enhancing tissue regions compared to complete images. Our custom loss function was explicitly designed to improve the super-resolution results for tissue regions, often the regions of interest in many medical imaging applications. By incorporating this custom loss function, we aimed to achieve better outcomes and enhance the resolution of specific tissue regions, ultimately improving the overall performance of our proposed method.

For all four methods, including bicubic interpolation, MSE, SSIM, and Custom MSE, we reported the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) scores. These metrics were computed for all scale factors, including 2x, 3x, and 4x, providing a comprehensive assessment of each method's performance across different image magnification levels. By utilizing these four distinct methods and evaluating them using PSNR and SSIM scores, we could compare their effectiveness and determine the most suitable approach for enhancing the resolution of the images across various datasets.

5.2   Results

Initially, we present the findings of our experiment conducted on the CTORG dataset, which substantiate our hypothesis that distinct tissues possess unique structural information, leading to varied performance of super-resolution methods.

Table 5.1 shows that the application of the Modified-ESPCN approach with SSIM loss exhibited a significantly higher PSNR and SSIM score for all regions compared to MSE loss for both Bicubic interpolation and the Modified-ESPCN approach. Notably, the PSNR and SSIM scores were significantly higher for the entire image than for the tissue-only region. Additionally, there were significant

differences in PSNR and SSIM scores among various tissues, such as the liver, bladder, lung, kidney, and bone. Specifically, the lung and bone tissues displayed significantly lower PSNR and SSIM scores than the liver, kidney, and bladder. Furthermore, the SSIM score for the structural component (S) was significantly lower than the luminance (L) and contrast (C) components across regions for all super-resolution methods, as summarized in Table 5.2.

Table 5.1: Region or tissue-specific comparison of performance metrics for CT images.

Performance Metrics →	PSNR			SSIM		
Regions ↓	Bicubic	Modified ESPCN (MSE Loss)	Modified ESPCN (SSIM Loss)	Bicubic	Modified ESPCN (MSE Loss)	Modified ESPCN (SSIM Loss)
Full Image	49.1	46.63	50.8	0.8	0.84	0.88
Tissue only	40.23	40.77	43.2	0.67	0.75	0.79
Liver	48.56	47.21	49.71	0.84	0.86	0.88
Bladder	48.12	46.36	49.51	0.83	0.85	0.87
Lung	39.82	40.3	43.39	0.63	0.72	0.77
Kidney	44.81	45.92	47.32	0.8	0.86	0.88
Bone	34.83	36.65	38.35	0.61	0.72	0.76

Table 5.2: Region or tissue-specific comparison of luminance(L), contrast(C) and structure(S) of SSIM metric for CT images.

SR Methods	Bicubic			Modified ESPCN (MSE loss)			Modified ESPCN (SSIM loss)		
Performance Metrics → Regions ↓	SSIM (L)	SSIM (C)	SSIM (S)	SSIM (L)	SSIM (C)	SSIM (S)	SSIM (L)	SSIM (C)	SSIM (S)
Full Image	0.98	0.96	0.84	0.99	0.97	0.87	0.99	0.97	0.90
Tissue only	0.99	0.96	0.71	1.00	0.96	0.78	1.00	0.96	0.82
Liver	1.00	0.98	0.85	1.00	0.99	0.87	1.00	0.99	0.89
Bladder	1.00	0.99	0.85	1.00	0.99	0.86	1.00	0.99	0.89
Lung	0.98	0.93	0.68	0.99	0.96	0.75	1.00	0.94	0.81
Kidney	1.00	0.98	0.81	1.00	0.99	0.87	1.00	0.98	0.89
Bone	1.00	0.96	0.63	1.00	0.95	0.76	1.00	0.96	0.79

After effectively substantiating our hypothesis through the outcomes presented in Table 5.1 and Table 5.2, we proceed to demonstrate the outcomes obtained from our extensive datasets utilizing the newly devised custom loss function based on local mean squared error (MSE) for various tissue types. The results have been presented comprehensively, encompassing all relevant parameters. These parameters encompass various super-resolution methods at different scale factors. Notably, the scores have been computed for the entire image and different tissue

regions. This methodology has been consistently applied throughout all the reported results.

Based on the analysis of Table 5.3 and Table 5.4 for the CLIS dataset, it can be observed that the Modified ESPCN model with Local MSE loss consistently outperforms the Modified ESPCN model with MSE loss in terms of both PSNR and SSIM metrics. This indicates the superior performance and effectiveness of the Local MSE loss in improving the model's results for the CLIS dataset.

Table 5.3: PSNR results for CLIS dataset

Scale Factor	Regions	SR Methods			
		Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
2x	Full Image	36.82	36.16	34.26	33.53
	Lungs Only	31.99	31.99	31.27	26.79
3x	Full Image	34.9	32.29	33.87	33.08
	Lungs Only	29.86	27.99	28.69	26.37
4x	Full Image	32.12	30.31	31.31	25.73
	Lungs Only	26.07	25.81	26.33	24.99

Table 5.4: SSIM results for CLIS dataset

Scale Factor	Regions	SR Methods			
		Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
2x	Full Image	0.935	0.944	0.944	0.937
	Lungs Only	0.918	0.922	0.942	0.917
3x	Full Image	0.838	0.875	0.873	0.877
	Lungs Only	0.777	0.781	0.783	0.791
4x	Full Image	0.781	0.831	0.834	0.832
	Lungs Only	0.683	0.696	0.698	0.701

Table 5.5 and Table 5.6 derived by experimentation on the OSIC dataset provide further support for our anticipated results. Specifically, the Modified ESPCN with Local MSE loss exhibits slightly superior performance compared to the model using MSE loss. Notably, we observe significant improvements in the lung region's PSNR for the 4x scale factor when employing the Local MSE model, highlighting its superiority over the standard MSE model.

Table 5.7 and Table 5.8 derived by experimentation on the CTORG dataset provide further support for our anticipated results for lung as well as liver region. We can note that for any of the scale factor of 2x, 3x or 4x for lungs as well as liver the

Table 5.5: PSNR results for OSIC dataset

Scale Factor	Regions	SR Methods			
		Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
2x	Full Image	28.82	27.31	26.85	17.81
	Lungs Only	28.26	27.36	27.91	20.45
3x	Full Image	28.35	22.78	27.22	21.92
	Lungs Only	27.26	24.48	25.97	20.88
4x	Full Image	25.82	18.34	23.10	24.98
	Lungs Only	23.99	19.31	22.82	22.33

Table 5.6: SSIM results for OSIC dataset

Scale Factor	Regions	SR Methods			
		Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
2x	Full Image	0.788	0.807	0.802	0.783
	Lungs Only	0.859	0.867	0.862	0.827
3x	Full Image	0.645	0.684	0.702	0.695
	Lungs Only	0.712	0.709	0.728	0.732
4x	Full Image	0.623	0.628	0.649	0.673
	Lungs Only	0.649	0.648	0.654	0.675

Table 5.7: PSNR results for CTORG dataset

Scale Factor	Regions	SR Methods			
		Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
2x	Full Image	44.24	45.09	45.64	44.28
	Lungs Only	40.71	40.39	40.44	41.33
	Liver Only	43.83	43.17	43.27	35.62
3x	Full Image	41.75	42.96	42.27	41.36
	Lungs Only	36.59	36.14	36.91	35.88
	Liver Only	40.53	39.95	41.76	39.68
4x	Full Image	38.69	40.44	40.56	39.84
	Lungs Only	32.17	32.65	32.87	32.01
	Liver Only	38.59	39.04	39.15	36.83

Modified ESPCN has performed better in both the metrics, especially more so in PSNR. These results align with our intended outcomes.

Figure 5.1, 5.2 and 5.3 displays a single original slice image from CLIS dataset, OSIC dataset and CTORG dataset respectively. The figure contains ground truth images which are cropped from original slice for better visualization of details

Table 5.8: SSIM results for CTORG dataset

Scale Factor	Regions	Bicubic	Modified ESCPN (MSE Loss)	Modified ESCPN (Local MSE Loss)	Modified ESCPN (SSIM Loss)
			SR Methods		
2x	Full Image	0.978	0.980	0.980	0.981
	Lungs Only	0.975	0.974	0.976	0.976
	Liver Only	0.960	0.958	0.960	0.960
3x	Full Image	0.934	0.946	0.945	0.948
	Lungs Only	0.902	0.892	0.893	0.905
	Liver Only	0.899	0.909	0.910	0.911
4x	Full Image	0.886	0.913	0.915	0.922
	Lungs Only	0.815	0.809	0.815	0.821
	Liver Only	0.872	0.885	0.887	0.889

for various resolution of 2x, 3x and 4x agianst all the super resolution methods including Bicubic, Modified ESPCN with MSE loss, Modified ESPCN with local loss and Modified ESPCN with SSIM loss.
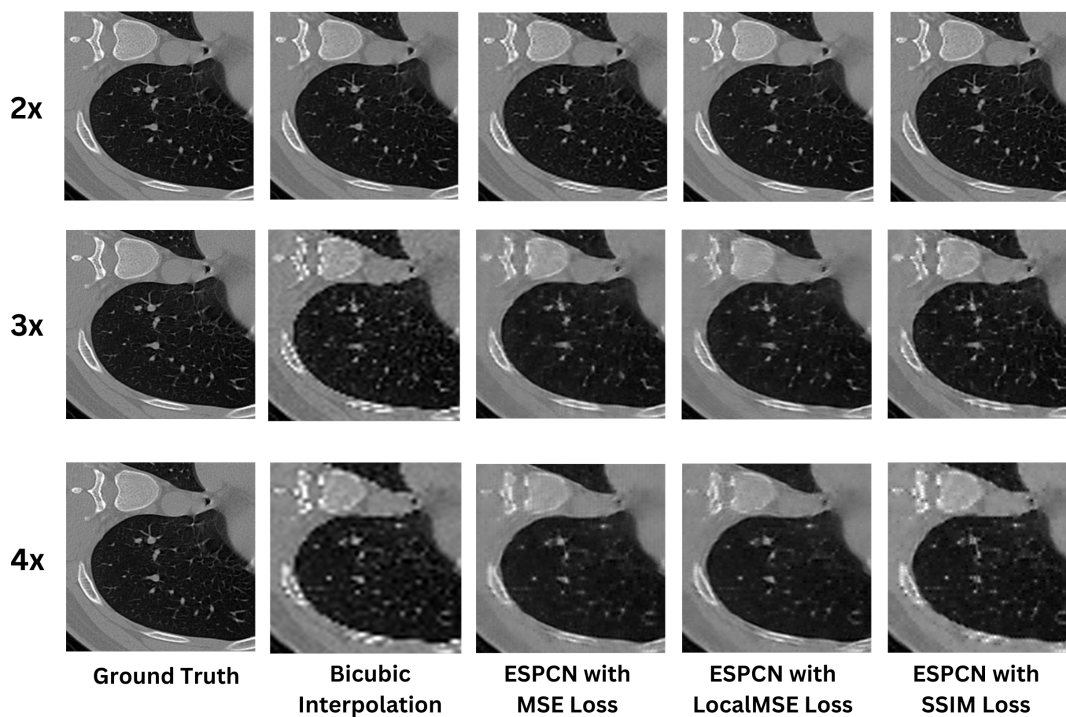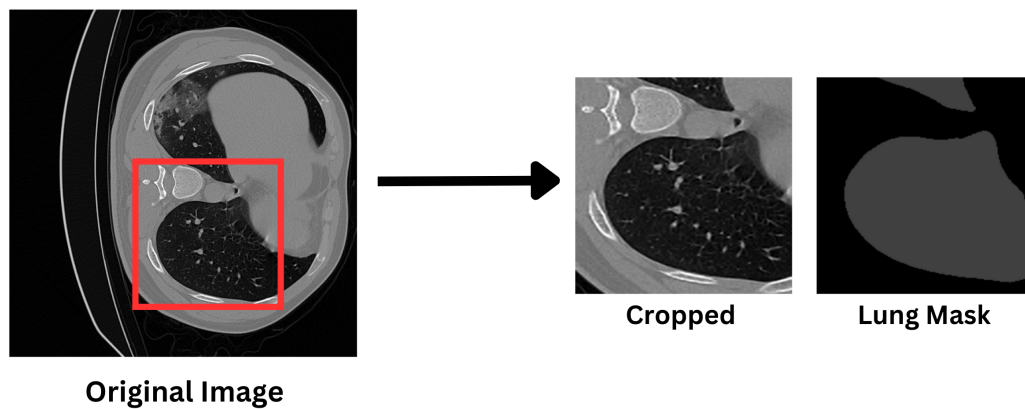
Figure 5.1: Original full size slice from CLIS Dataset is displayed along with cropped version and its mask. Below it Bicubic Interpolation, Modified ESPCN with MSE loss, Modified ESPCN with LocalMSE loss and Modified ESPCN with SSIM loss across 2x, 3x and 4x scale factors is shown.
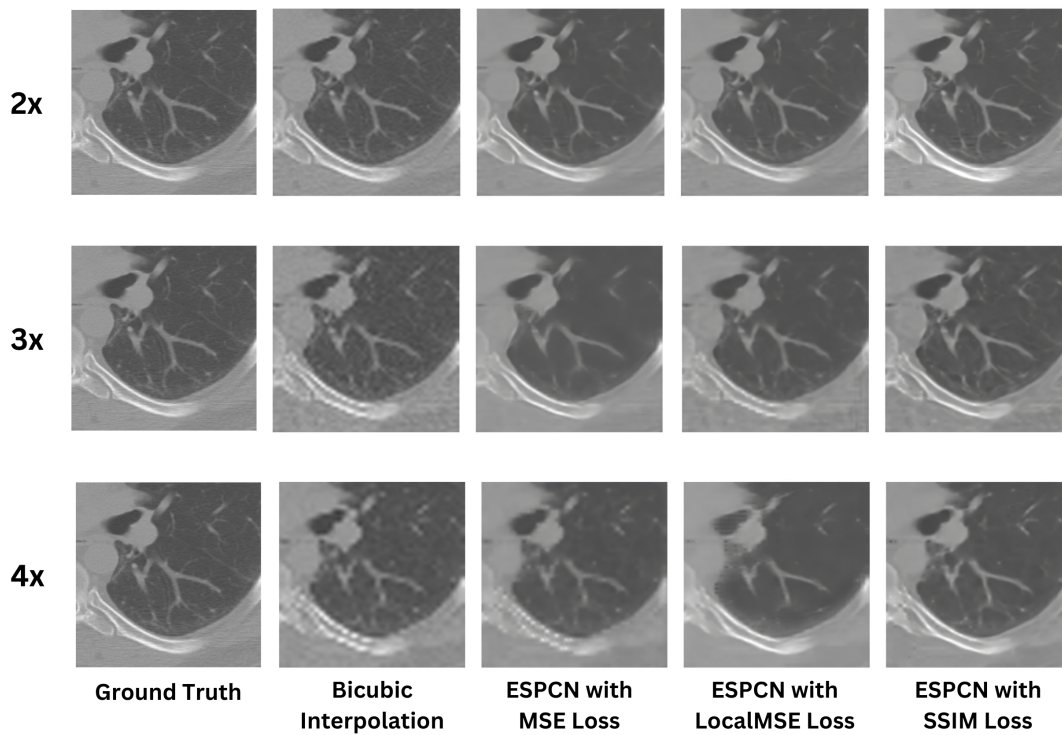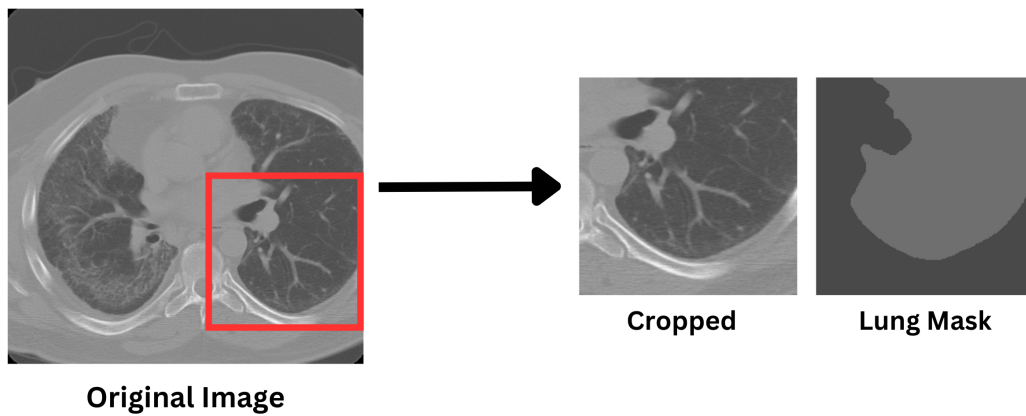
Figure 5.2: Original full size slice from OSIC Dataset is displayed along with cropped version and its mask. Below it Bicubic Interpolation, Modified ESPCN with MSE loss, Modified ESPCN with LocalMSE loss and Modified ESPCN with SSIM loss across 2x, 3x and 4x scale factors is shown.

Original Image

Cropped   **Lung Mask**

2x

3x

4x

Ground Truth   **Bicubic Interpolation**   **ESPCN with MSE Loss**   **ESPCN with LocalMSE Loss**   **ESPCN with SSIM Loss**
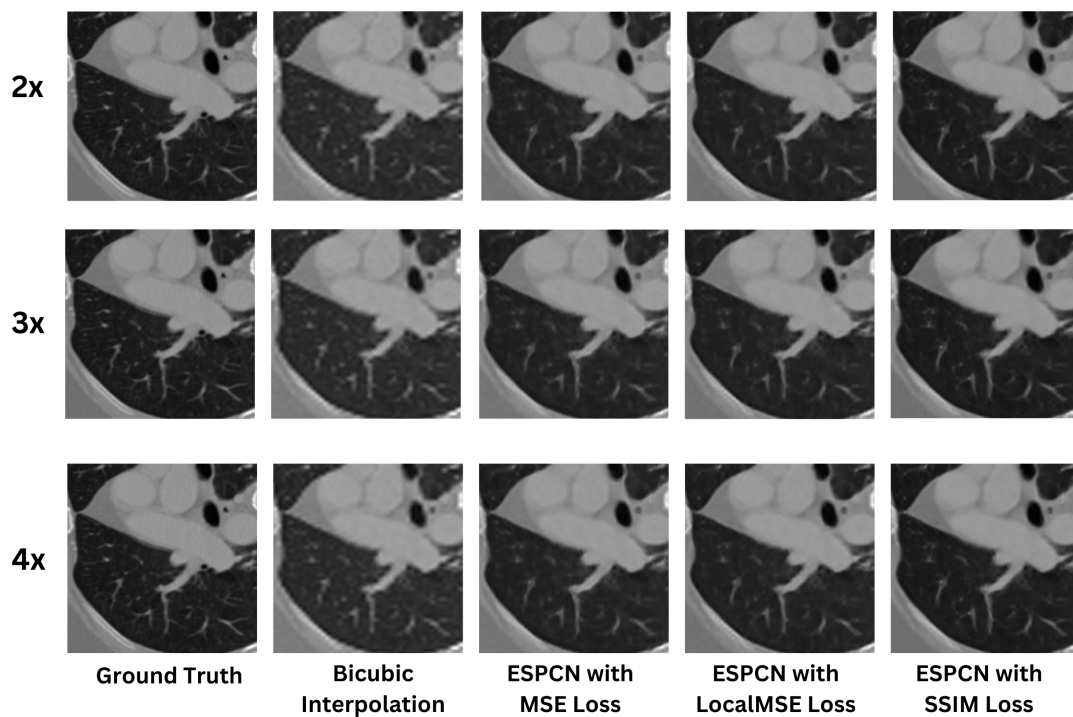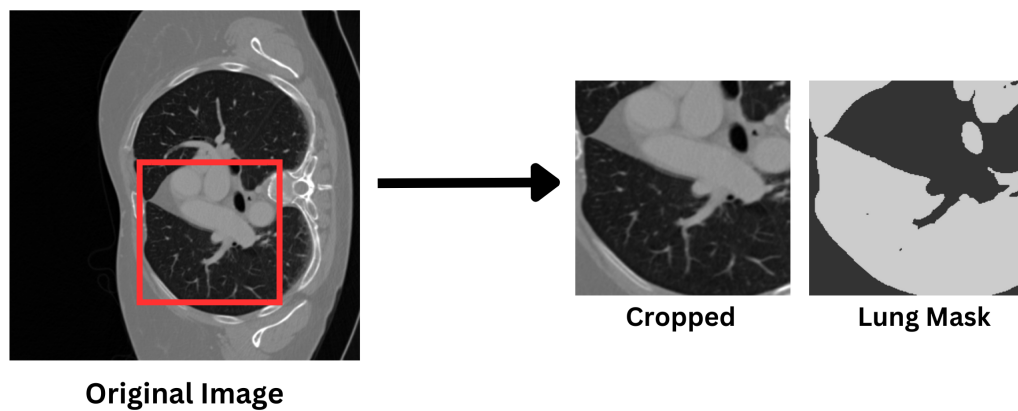
Figure 5.3: Original full size slice from CTORG Dataset is displayed along with cropped version and its mask. Below it Bicubic Interpolation, Modified ESPCN with MSE loss, Modified ESPCN with LocalMSE loss and Modified ESPCN with SSIM loss across 2x, 3x and 4x scale factors is shown.

Chapter Summary

Three models were trained using different loss functions: mean square error (MSE), structural similarity index (SSIM), and a custom MSE loss function designed to enhance tissue regions. Evaluations were conducted using four super-resolution approaches: bicubic interpolation, ESPCN model with MSE loss, SSIM loss, and the proposed custom MSE loss. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) scores were reported for all methods and scale factors (2x, 3x, and 4x), enabling a comprehensive performance assessment. All the results are reported in this chapter.

CHAPTER 6

Discussion and Conclusions

Our first objective was to assess the performance of CT image super-resolution methods on individual tissues. Interestingly, we observed significantly lower SSIM scores when the metrics were computed locally, focusing on specific tissues or regions, compared to when computed over the entire image (Table 5.1). This finding suggests that background regions and non-body tissue areas, artificially inflate the performance ratings. Therefore, excluding the background region when calculating performance metrics is crucial to accurately represent the outcomes.

Previous studies on medical radio-imaging super-resolution have primarily assessed the similarity metrics over entire images without considering the variations in different body tissues. [22], [20] - [23] However, different tissues in CT images exhibit distinct image characteristics and patterns. SR methods are sensitive to the information content, such as higher spatial frequencies or variance. Our study revealed significantly different performances of SR methods for different tissues in CT images (Tables 1 and 2), with lung and bone tissues exhibiting notably lower performance. This finding is particularly important considering lung tissue is of great interest to radiologists and physicians, especially in cases involving respiratory symptoms and COVID-19 infection. Therefore, evaluating SR quality specifically for the region of interest, such as lung tissue, is imperative. This aligns with a recent study that reported lower SSIM scores for lung tissue patches in SR CT images produced by various SR methods [22].

Furthermore, our analysis indicates that SR methods tend to incur higher penalties on the structural component (S) of the SSIM measure compared to the contrast (C) and luminance (L) components (Table 5.2). This outcome is expected since the structural component represents high spatial frequency or variance information, which is challenging to approximate accurately. While our analysis focused on CT scans of different organs, the implications of this study can be generalized to other medical imaging modalities with diverse tissue types. In conclusion, ex-

cluding non-tissue regions when computing performance measures for SR methods is essential. Specifically, it is crucial to present performance measures specific to the tissue or region of interest for medical image enhancement applications using super-resolution approaches.

The next objective was to develop a region based custom loss function to improve performance of SR methods on specific area of interest like lungs. Based on the results presented in Table 5.3 and Table 5.4 for the CLIS dataset and Table 5.5 and Table 5.6 for OSIC dataset and table 5.7 and 5.8 , we can state that the Modified ESPCN model with Local MSE loss generally demonstrates superior performance compared to the model using MSE loss as we are getting higher PSNR and SSIM scores for the Local MSE model across various tissue types and scale factors. These findings suggest that considering the local mean squared error within specific tissue regions improves super-resolution results for the CLIS dataset.

Additionally, there are several noteworthy findings that can be inferred from the obtained results. The LocalMSE loss corresponds to the Mean Squared Error (MSE) loss specifically applied to a particular region of interest, while the MSE loss is predicated on Peak Signal-to-Noise Ratio (PSNR). Consequently, the enhancement in PSNR scores outweighs the improvements in SSIM scores, as SSIM primarily emphasizes image structure. Another critical observation pertains to the perceptual image quality evident in the results. It is readily apparent that despite achieving comparable performance ratings, the Bicubic interpolation method yields images with significant noise and lacks meaningful information. Furthermore, the model employing the SSIM loss exhibits the highest level of perceptual quality compared to all other approaches.

These observations and results highlight the importance of considering tissue-specific characteristics and local information when evaluating and developing super-resolution methods for medical imaging. By incorporating custom loss functions based on local mean squared error, we can achieve improved performance and quality in the super-resolution of CT images, particularly for specific tissue regions of interest. These findings have implications for enhancing the accuracy and reliability of medical image enhancement techniques using super-resolution approaches across different modalities and tissue types.

References

[1] 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)*, 2019, pp. 1669–1673.

[2] T. Huang, G. Yang, and G. Tang, "A maximum a posteriori estimation approach for image super-resolution," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 984–998, 2015.

[3] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[4] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2. IEEE, 2005, pp. 60–65.

[5] J. Darbon, J. Fadili, and G. Peyré, "Image decomposition into a bounded variation component and an oscillating component," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 263–276, 2007.

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[8] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[9] S. Chen, H. Li, X. Yuan, Q. Tan, and W. Zuo, "Real-world single image super-resolution: A survey," *arXiv preprint arXiv:2203.05932*, 2022.

[10] C. Dong, C. C. Loy, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *arXiv preprint arXiv:1511.04587*, 2016.

[12] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," *arXiv preprint arXiv:1802.08797*, 2018.

[13] C. Ledig, L. Theis, F. Husz'ar, 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," *arXiv preprint arXiv:1609.04802*, 2017.

[14] M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa, and J. Nalepa, "Deep learning for multiple-image super-resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1062–1066, 2020.

[15] J. L. Prince, A. Carass, C. Zhao, B. E. Dewey, S. Roy, and D. L. Pham, "Image synthesis and superresolution in medical imaging," in *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2020, pp. 1–24.

[16] K. Yamashita and K. Markov, "Medical image enhancement using super resolution methods," in *International Conference on Computational Science*. Springer, 2020, pp. 496–508.

[17] J. S. Isaac and R. Kulkarni, "Super resolution techniques for medical image processing," in *2015 International Conference on Technologies for Sustainable Development (ICTSD)*. IEEE, 2015, pp. 1–6.

[18] H. Yu, J. Park, Y.-G. Yoo, and D. Hwang, "Computed tomography super-resolution using convolutional neural networks," *Computers in biology and medicine*, vol. 82, pp. 59–68, 2017.

[19] P. Rajeshwari and K. Shyamala, "Chest ct image super resolution using deep learning network models," in *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. IEEE, 2021, pp. 1–5.

[20] K. Umehara, D. Ueda, T. Takenaga, H. Yoshida, M. Yamada, and M. Jinzaki, "Application of super-resolution convolutional neural networks to chest ct," *Journal of Medical Imaging*, vol. 5, no. 1, p. 013502, 2018.

[21] M. Kolarik, E. Belyaev, and F. Šroubek, "Superresolution of mri brain images using unbalanced 3d dense-u-net," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 544–552.

[22] E. C. d. Farias, C. d. 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*, vol. 11, no. 1, 2021.

[23] C. Jiang, Q. Zhang, R. Fan, and Z. Hu, "Super-resolution ct image reconstruction based on dictionary learning and sparse representation," *Scientific Reports*, 2018.

[24] B. Rister, D. Yi, K. Shivakumar, T. Nobashi, and D. L. Rubin, "Ct-org, a New Dataset for Multiple Organ Segmentation in Computed Tomography," *Scientific Data*, vol. 7, no. 1, pp. 1–9, 2020.

[25] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, and et al., "Covid-19 CT Lung and Infection Segmentation Dataset," Apr 2020.

[26] D. Ourselin, M. Tustison, A. L. Ardon, B. M. van Ginneken, and C. R. Vermont, "OSIC Pulmonary Fibrosis Progression Dataset," Dec 2021.

[27] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A comprehensive review of deep learning-based single image super-resolution," *PeerJ Computer Science*, vol. 7, p. e621, 2021.

[28] J. Bidgood, W. D., S. C. Horii, F. W. Prior, and D. E. Van Syckle, "Understanding and using DICOM, the data interchange standard for biomedical imaging," *Journal of the American Medical Informatics Association : JAMIA*, vol. 4, no. 3, pp. 199–212, 1997.

[29] S. Konya, "CT Lung Heart Trachea Segmentation Dataset," https://www.kaggle.com/datasets/sandorkonya/ct-lung-heart-trachea-segmentation, 2021.

[30] W. Shi, J. Caballero, F. Huszár, J. Totz, A. Aitken, R. Bishop, and D. Rueckert, "Sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[31] Wikipedia contributors, "Structural similarity," https://en.wikipedia.org/wiki/Structural_similarity, accessed on May 22, 2023.