

Spatially Constrained Compound Magnification Framework for Histopathology Whole Slide Images

by
Mehdi Afshari

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2022

© Mehdi Afshari 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Joel Saltz
Professor, Dept. of Biomedical Informatics, Stony Brook University

Supervisors: H.R. Tizhoosh
Professor, Mayo Clinic, Rochester, MN, USA

Shahryar Rahnamayan
Professor, Dept. of Engineering, Brock University

Internal Member: Nima Maftoon
Assistant Professor, Dept. of Systems Design Engineering,
University of Waterloo

Internal-External Member: Mark Crowley
Associate Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Oleg Michailovich
Associate Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Histopathology whole slide images (WSIs) contain a considerable amount of valuable information required for diagnosis, prognosis, and treatment planning. These scans are the product of a contemporary technology of high-precision digitizers that seemed to aid in the clinical procedure. The availability of internet tools might improve treatment process in a variety of ways, like remotely communicating opinions amongst pathologists. Due to the significance of details in glass tissue slides, these images can be as big as $100,000 \times 200,000$ pixels and are often scanned in often three like red, green, and blue or luma component, blue-difference, and red-difference. Furthermore, WSIs are often kept in pyramidal magnification structures that include images of varying resolutions as well as clinical information. At higher magnification levels of the pyramid, a much greater quantity of information is stored. Although lower level magnification does not incorporate all information held at higher levels, it may considerably aid the clinical evaluation by producing high-density pixel information from lower resolution. The purpose of this research is to create a mechanism capable of recovering information at high-resolution such as 20X magnification, from lower-level magnifications such as 5X. This strategy reduces the amount of storage space and bandwidth needed for image transfer. Having models to reproduce the 20X magnification slides out of 5X magnification with high accuracy can considerably facilitate the adoption of digital pathology. one of the barriers to going digital is the high cost of storing many gigapixel WSIs. Transferring histopathology slides to a remote location can be facilitated by this method, while smaller size slides are easier for data sharing. Finally, in digitizing glass slides, scanners may not always correctly focus on the tissue, resulting in blurry portions negatively affecting diagnosis in certain cases. The same techniques may overcome the optical hurdles of obtaining crisp images at high magnifications. The suggested solutions would allow for the storing of digital slides in much smaller-scale versions while maintaining the reproducibility of high-resolution images. On the other hand, storing merely a lower magnification rather than the actual high magnification image and depending only on the software may not result in an adequate results. To improve the capabilities of the suggested approach even further, it is advisable to save some portions of the original high quality picture to guarantee a more accurate outcome. The second challenge that this research is focusing on is identifying the areas that may need further attention in order to store a part of them. The anomalous region of a slide is the piece that substantially proposes the diagnosis when defining the concern of an essential disease. A strategy for detecting this region is also explained here. To detect this area, a technique that can do it unsupervised is appropriate since it reduces pathologists' effort by not needing them to provide accurate annotations for a certain kind of cancer before using the approach. This thesis further relies on the unsupervised localization of an abnormal WSI region.

Acknowledgements

My sincere gratitude goes out to my supervisor, Professor Hamid R. Tizhoosh, for his unwavering advice and guidance. He offered to be my supervisor just as I was having my most difficult time. I want to thank Professor Shahryar Rahnamayan, my co-supervisors, for helping me with the most challenging questions.

Thanks to my thesis committee members for their constitutive comments and guidelines for the current and also future work.

My loving parents and brother, who were always there for me whenever I was in a difficult position, have my heartfelt thanks.

I want to thank all of friends in Kimia Lab who helped me in the ways they could to complete this task.

Last but not least, I want to express my admiration to Tahereh Karami Shoar for supporting me throughout my academic career in both an emotional and intellectual sense.

Table of Contents

List of Tables	x
List of Figures	xi
List of Abbreviations	xv
1 Introduction	1
1.1 Problem Statement	3
1.2 Motivation	4
1.3 Organization of the Thesis	6
2 Background and Literature Review	8
2.1 Histopathology Image Analysis	9
2.2 Image-to-Image Translation	11
2.2.1 Single Image Super-Resolution (SISR)	12
2.2.2 Problem formulation	13
2.2.3 Interpolation	14

2.2.4	SR Deep Networks	16
2.2.5	Architectures	16
2.2.6	Loss Functions	18
2.2.7	Reconstruction Quality Measurement	20
2.3	Image-to-Image Distance	20
2.3.1	Peak signal-to-noise ratio (PSNR)	21
2.3.2	Structure Similarity Index Method (SSIM)	21
2.3.3	Feature Similarity (FSIM)	22
2.3.4	Gradient Magnitude Similarity Deviation (GMSD)	23
2.3.5	Visual Saliency-based Index (VSI)	23
2.3.6	Mean Deviation Similarity Index (MDSI)	24
2.3.7	Learned Perceptual Image Patch Similarity (LPIPS)	24
2.3.8	Haar Perceptual Similarity Index	24
3	Compound Magnification Framework	26
3.1	Problem Definition	26
3.2	Spatially Constrained Compression	29
3.2.1	Preliminary Approaches	31
3.3	Summary	33

4	Super-Resolution Reconstruction	35
4.1	Background	35
4.1.1	Super-resolution in Histopathology	35
4.2	Dataset Preparation	36
4.2.1	Publicly available SR data-sets	37
4.2.2	Publicly available Histopathology data	37
4.2.3	Creation of a data-set	38
4.2.4	Data-set for histopathology SR	39
4.3	Super-resolution reconstruction	39
4.3.1	Networks	39
4.3.2	Training	41
4.4	Human Assessment	41
4.4.1	Diagnostic dataset	42
4.4.2	Questionnaires	42
4.5	Analysis of results	43
4.6	Summary	44
5	Unsupervised WSI Patch Selection	55
5.1	Objective	55
5.2	Introduction	56
5.3	Background	58
5.3.1	Unsupervised Detection of Visual Anomalies	58

5.3.2	Image-to-Image Translation	58
5.3.3	Unsupervised Learning in Medical Imaging	59
5.4	Method	59
5.4.1	I2I Translation	60
5.4.2	Measurement System	61
5.5	Materials	62
5.5.1	Histopathology Dataset	62
5.5.2	Processing Units	63
5.5.3	Frameworks	63
5.6	Results and Discussions	63
5.6.1	Magnification Selection	64
5.6.2	Gradient Calculation Application	66
5.6.3	Distance Distributions	66
5.6.4	Annotation Results	72
5.7	Summary and Conclusion	73
6	Summary and Remarks	86
6.1	Summary	86
6.2	Remarks	87
6.3	Future research	89
	References	90

List of Tables

4.1	Super-Resolution Famous Datasets and Data Overview	37
4.2	The TCGA codes (in alphabetical order) of all 32 primary diagnoses and corresponding number of evidently diagnosed patients in the dataset	45
4.3	Accuracy of image reconstruction using six networks trained on TCGA dataset; the PSNR/SSIM numbers are shown where the bold blue numbers are best reconstructions across networks, bold red and green are the worst and best bicubic reconstruction among data categories, respectively.	46
4.4	Human assessment of SR vs. HR pathology patches; Type defines the Benign or Malignancy; Subtype is the eight categories of BreakHis dataset used for diagnosis purpose	54
5.1	Patch-Selection, AUROC of Metastasis region detection results	73

List of Figures

1.1	An illustration of a WSI in SVS format that includes: Left, images stored in various levels; Right, scanner properties, image file specifications, and etc.	5
2.1	A TCGA brain patch of the size 256×256 pixels and the interpolation of its down-sampled patch of the size 64×64 pixels; (a) original image and the down-sampled image in top right, (b) nearest neighbor interpolation, (c) bilinear interpolation and (d) bicubic interpolation	15
2.2	Four general frameworks of deep neural networks used for image super resolution including; (a) Pre-upsampling SR (b) Post-upsampling SR (c) Progressive upsampling SR (d) Iterative up-and-down Sampling SR; Grey boxes, SR blocks; Green boxes, upsampling units; Orange boxes, downsampling units	17
3.1	A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches). The 5x image can be upsampled to 20X when needed for visual inspection. The 20X patches depict relevant diagnostic histology, hence preserved in original resolution.	27
3.2	A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches)	28
3.3	A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches)	30

3.4	High resolution patch selection using WSI color histogram clustering in low resolution; the lower resolution slide is first tiled, then color histograms are computed and centroids of clustering are selected	32
3.5	Clustering by deep features; (a) Sample images from CRC dataset used for training, (b) Confusion matrix of clustering based on the deep features of images, (c) The DCEC network architecture which has been trained to extract features (green area is the input, blue area is the decoder output used for training, yellow area shows the deep features used for clustering)	34
4.1	Patches of the created SR dataset in nine TCGA cancer subtypes	40
4.2	The image reconstruction accuracy (PSNR and SSIM) distribution of the RCAN network of the 26 TCGA subtypes categorized in the nine biopsy sites	47
4.3	The image reconstruction accuracy (PSNR and SSIM) distribution of the ESRGAN network of the 26 TCGA subtypes categorized in the nine biopsy sites	48
4.4	Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box	49
4.5	Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box	50
4.6	Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box	51
4.7	Google-form questionnaire for diagnostic tests; left: overview of the form page, right: options available for selection on the form	52
4.8	Two pages of the Google-form questionnaire for quality comparisons tests	53
5.1	General overview of the patch selection method	57

5.2	Initial distribution experiments results on 4X, and 16X magnification with PSNR, and SSIM	65
5.3	HR, LR, and Super Resolution (SR) Patches with magnification of 16X . .	66
5.4	Visualization of differences distribution assessment for 4X magnification . .	67
5.5	Visualization of differences distribution assessment for 8X magnification . .	68
5.6	Visualization of differences distribution assessment for 16X magnification .	69
5.7	Gradient calculation visual inspection for normal vs. metastasis patch . . .	70
5.8	Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the HR patch to the SR patch	74
5.9	Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the HR patch to the SR patch histogram matched to the HR	75
5.10	Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the gradient of HR patch to the gradient of SR patch	76
5.11	Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the gradient of HR patch to the gradient of SR patch histogram matched to the HR patch	77
5.12	Visualization of WSI, Otsu Thresholding of the tissue region, and the metastasis region shown in blue	78
5.13	Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the SR generated patch	79
5.14	Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the SR generated patch	80
5.15	Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the histogram matching SR generated patch	81

5.16 Visualization of MS-SSIM, and SSIM measure on the WSI comparing the original patch to the histogram matching SR generated patch	82
5.17 Visualization of differences distribution assessment for 16X magnification .	83
5.18 Visualization of differences distribution assessment for 16X magnification .	84
5.19 Proposed method for patch selection of a WSI	85

List of Abbreviations

AUROC Area Under the Receiver Operating Characteristic curve

CAD Computer-Assisted Diagnosis

CMF Compound Magnification Framework

CNN Convolutional Neural Network

DL Deep Learning

FSIM Feature Similarity Index Method

GAN Generative Adversarial Network

GMSD Gradient Magnitude Similarity Deviation

HaarPSI Haar Perceptual Similarity Index

HR High Resolution

I2I Image-to-Image

IQA Image Quality Assessment

LPIPS Learned Perceptual Image Patch Similarity

LR Low Resolution

ML Machine Learning

MS-GMSD Multi-Scale Gradient Magnitude Similarity Deviation

MS-SSIM Multi-Scale Structural Similarity Index Method

PSNR Peak Signal-to-Noise Ratio

SR Super Resolution

SSIM Structure Similarity Index Method

VSI Visual Saliency-based Index

WSI Whole-Slide Image

Chapter 1

Introduction

The field of pathology is experiencing a paradigm change due to the widespread use of Whole-Slide Image (WSI) in lieu of traditional light microscopy. Digital pathology is critical in modern clinical practice and is fast gaining acceptance as an essential tool in the laboratory setting [1] [2]. The capacity to digitally visualize microscopic pathology features opens up new avenues for pathology study. Processor improvements, quicker networks, and more affordable storage have allowed pathologists to handle digital slide images more quickly and adaptable than ever before. Additionally, remote primary diagnostic work, teleconsultation, workload efficiency and balance, collaborations, central clinical trial evaluation, image analysis, virtual training, and innovative research are facilitated by access to digital slides [3]. Over the past two decades, the invention and growth of whole-slide imaging in digital pathology imaging have permitted the imaging and permanent storage of whole slides at high resolution. The computer vision and artificial intelligence communities have expanded the potential for computational pathology via WSI technology, including deep learning methods [4]. Recent years have seen incredible advancements in computer-aided approaches, owing primarily to deep learning. Numerous deep architectures have been trained on large-scale datasets to uncover significant biological informatics insights and perform outstanding image processing, and analysis tasks [5].

Scientists and pathologists benefit from the digital scans of tissue slides in considerably high resolutions because of the details of the histologic features of biopsy samples. For example, as very detailed information, the cells are a critically examined reference in histopathology since their shape and distribution include vital information that may be used to differentiate between conditions. Breast cancer, for example, is the most prevalent

malignant tumor affecting females. Diagnosis of cancer is critical for effective treatments. Classification of cell nuclei is critical in histopathology image analysis because many issues, particularly malignancies, may be defined by their expression in cell nuclei. [6] For breast cancer, historically, pathologists were required to manually extract cytological descriptions of cells at high magnification (e.g., cell size, shape, distribution, and cellular spaces) from histopathology images and then identify the cell nuclei using these human-specified criteria [7]. Examining a large number of biopsy slides manually requires considerable effort and has a poor degree of repeatability [8]. Additionally, since the attention lesion's extent is often tiny compared to the WSI, details become much more important. Certain crucial areas of attention in a WSI may account for less than 1% of the image. Localization of lesions is also a time-consuming process, even for experienced pathologists. While computer-assisted approaches for quick processing have received much interest, the need for high-resolution images that depict slides at high magnifications persists. [9].

During the past few years, microscopic images have been the focus of many researchers in deep learning [10][11]. Overall, proposed approaches mainly concentrate on classifying patches resulting from sliding a window in WSI [12]. Primarily, manual annotations are used to create bounding boxes in WSIs and train a network to achieve the desired purpose. In most cases, a classifier applies to a set of "patches" (selected sub-images) for training. However, these perspectives come with limitations that require attention. To achieve this, pathologists must create annotations via bounding boxes around regions of interest in each WSI of the training set images. Pathology knowledge and expertise are vital for aggregating patch classifications to develop a heuristic. In the proposed methods, a patch is usually selected independent of neighbors or other parts of a WSI, which means possible correlations may have been neglected.

For many diagnostic or prognostic applications, a high-resolution image is necessary to resolve any of the abovementioned concerns. The ultimately sensitive and accurate processing of high-quality scans comes at costs, increasing the treatment and medical exposures, extra time, and staff and expert requirements. The proposition of digital solutions helps with these issues and accelerates patient access to proper treatment.

Additionally, similar problems may exist in other disciplines. Theoretically, any gigantic image may need a similar procedure to this study while keeping the properties of the proposed technique. For instance, astronomical and civil engineering images may approach the same size as histology pictures. In this thesis, however, histopathology slides were evaluated to establish the applicability of this method's dependability.

1.1 Problem Statement

This doctorate dissertation aims to design two methods for transforming histopathology images, as well as a novel image storage framework. WSIs are often used at high magnifications in histopathology, such as 20X and 40X. While high-resolution images are necessary for accurate clinical evaluation, the goal is to present an alternate framework that provides comparable image quality where it is required while also delivering storage advantages. To the best of the authors' knowledge, this research introduces the first *Compound Magnification Framework (CMF)* for tissue slides. A CMF is made up of two primary components. The original slide at a lower magnification (and resolution), accompanied by selected patches at the highest magnification necessary. A CMF is an alternative to conventional WSI built from a WSI and may be used to construct a comparable WSI.

The first method addressed in this research is converting a CMF to a WSI, which is the most accurate way of obtaining information. As a consequence of this transformation, a synthetic WSI is created that successfully delivers information comparable to the original WSI. The term "information" refers to the image's perceptual data, represented by pixel values. The correctness of the information is determined by several criteria that, when combined, result in a proper diagnostic material. A reliable diagnostic material may provide critical information to a medical professional and a computer-assisted analytical approach. The problem is designing and implementing a technique for reconstructing a detailed medical image from a CMF.

The second method is used to convert a WSI to a CMF. From a memory standpoint, a CMF is substantially smaller than a WSI. The WSI is maintained at a decreased magnification, along with a selection of patches, to form a CMF. Essentially, the transformation denotes extracting WSI information from the perceptual data included in an image. In contrast, some researchers contributed significantly to compression regardless of the clinical relevance of information in a snap [13]. Perceptual information includes the shapes, relative positions, textures, colors, and orientations of features (e.g., cells) in a WSI. Here, we would choose a collection of patches that convey the most valuable information for future use. The goal implies that this part's primary objective is to handle the challenge of selecting a subset of patches for storage at high magnification.

1.2 Motivation

Histopathology is the science of investigation of tissue specimens that traditionally employ light microscopy. High-resolution scanning and storing slides enable researchers to transition to digital pathology due to technical improvements. Large pathology databases facilitated digital pathology for various applications, including research and clinical usefulness. Additionally, digital pathology allows for the examination and comparison of tissue samples with annotated digital images. Annotations are available for various data types, ranging from segmentation, classification, and diagnosis to clinical data such as pathology reports and molecular data. In many case scenarios, the diagnosis may improve by correlating comparable patients and retrieving similar images through “image search”. However, several issues are slowing or even halting this progress.

Histopathology diagnosis is based on the information contained in tissue samples mounted and fixed on glass slides and evaluated at various magnification levels. During a medical process, it is typical for pathologists to repeatedly zoom in and out on different slide areas to find certain diagnostic features. As a result, the digitized slides are often scanned at an exceptionally high resolution that correlates to high magnification to achieve equivalent performance in digital pathology. Computational pathology may then be performed on this high-resolution imaging. The gigapixel scanned images are huge files, often in the SVS format, necessitating more storage and transfer precautions. An SVS file is a Tiled TIFF image that contains a high magnification image, an overview image, multiple smaller, scaled copies of the scanned slide, and properties such as the slide name and scanner parameters. A sample WSI in SVS format is shown in Figure 1.1.

Each scan generates a gigabyte-sized image; typically, multiple slides are accessible for each patient. Every day, tens of thousands of people undergo biopsies. These variables contribute to the development of massive archives that are difficult to access and evaluate for research and clinical reasons. Hospitals and clinics may delete previous scans to make room for new ones to address the storage issue. Instead of deleting older slides, another approach is to maintain an alternative version and a mechanism that could reconstruct a similar image comparable to the original version upon access without compromising data accuracy.

Digital Scanners are a new generation of devices that automatically exploit the principles of a light microscope inside a box; the microscope is no longer visible to the user. Among the capabilities of digital scanners, resolution, and magnification are the factors that cause inconsistencies in automated scans. For instance, various scan magnifications

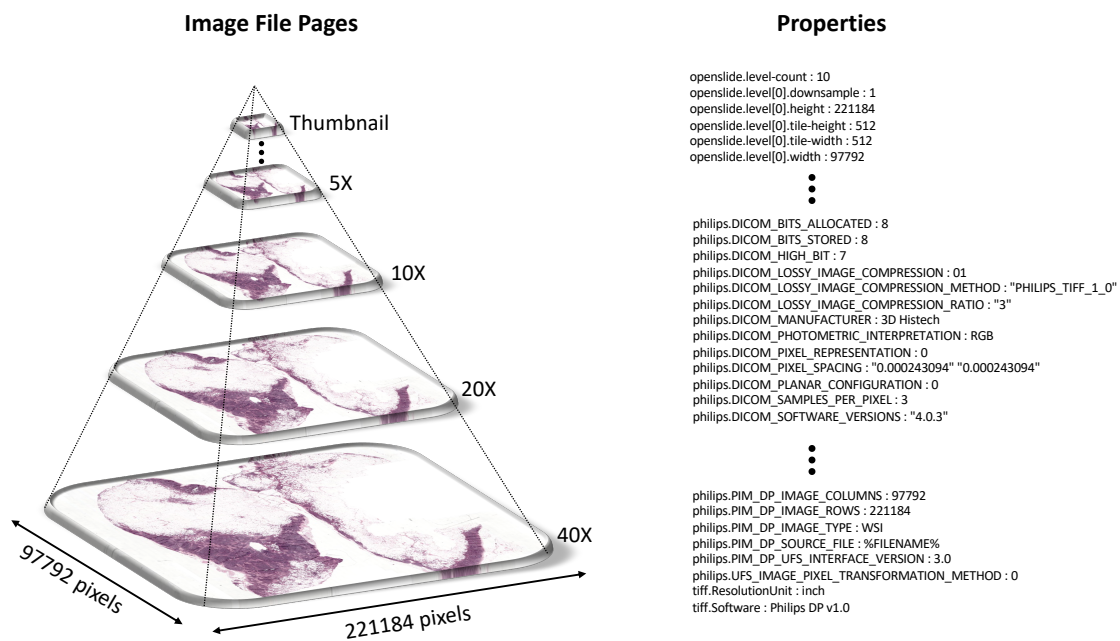


Figure 1.1: An illustration of a WSI in SVS format that includes: Left, images stored in various levels; Right, scanner properties, image file specifications, and etc.

may be inspected even in a large dataset. In a sensitive context like digital pathology, to design an automatic method, data consistency is pivotal. While it would be ideal if all data collection centers used similar digital devices with matching specifications, it is recognized that dataset collection variances are unavoidable. To maintain the magnification consistency across all diagnostic cases, a technique capable of magnifying scans from lower resolution would be the desired solution.

On the other hand, digital scanners are expensive and may not be accessible to all pathologists worldwide. One of the primary reasons for the high cost of digitization technologies is their high magnification scanning capacity. Although the availability of lower-cost equipment may be possible, the resulting reduction in scanning resolution and magnification would be detrimental. A practical tool to enhance the resolution at any magnification of scanned slides by affordable devices facilitates the adoption of digital pathology, hence the advancement of computational pathology.

Finally, the scans and scanners themselves are far from ideal. It has been noticed that scanned slides are not extraordinarily precise at providing a high-quality image. This issue is not rare since the optical devices of the scanner need to focus on the biopsy. By measuring the depth of the tissue surface, digital scanners maintain correct focussing. Focusing is error-prone due to issues like specks of dirt or stains on a slide and physical limitations of optics. A digital tool to help this problem is benefiting from methods capable of image enhancement.

All in all, some notions of compression and decompression of the scanned images may help solve the abovementioned problems. This study mainly aims at designing a method capable of down- and up-sampling ideas. The overall approach would help with storing smaller images while capable of restoring high-resolution versions. The lower-resolution images are magnifiable by the decompression method. Besides, blurry scans can also get enhanced in a decompression manner based on other informative areas of the WSI.

1.3 Organization of the Thesis

This thesis is split into six chapters, one of which is an introduction. The second chapter reviews the primary areas of literature relevant to the study done for this doctoral dissertation. This would begin with examining the techniques used to up-sample pictures in deep learning, including the problem formulation, the most frequently used network architectures, performance measures for the network, and training loss functions. Following

that, a description of the most commonly used image distance metrics is included, which will be established in Chapter five. Chapter three discusses the proposed framework, the tools required to implement it, and the resultant compression. Chapter four discusses the technique of image transformation, which upsamples the images. Chapter five discusses the issue of unsupervised localization, which is a necessary component of the second transformation. Finally, Chapter six summarises the study's findings, conclusions, and possible future work.

Chapter 2

Background and Literature Review

Digital pathology plays a critical part in contemporary clinical practice and is becoming an increasingly necessary technology prerequisite in the scientific laboratory setting. The examination of digitized slides for diagnostic reasons is referred to as digital pathology. The digitized slides are scans of the physical slides that allow computer-based evaluation. The section 2.1 provides an overview of the WSI analysis process.

Images are not always of the best standard, and their resolution may be improved. The image processing method known as image SR is an important family of image processing techniques used in computer vision to increase the resolution of pictures. Image SR is a technical term that refers to recovering High Resolution (HR) images from Low Resolution (LR) ones. The application of deep learning algorithms has enabled remarkable breakthroughs in image super-resolution in recent years [14]. The purpose of this method is to use perceptual information to recover pixel-level details about an image. This method, however, is not limited to images. Numerous researchers have used SR to increase the resolution of video for computer vision applications [15]. By and large, studies of image SR seek to propose a solution from three perspectives. SR approaches such as supervised, unsupervised, and domain-specific are all conceivable. The section 2.2.1 examines SR as a component of this study. The SR examination should provide the groundwork for the CMF to WSI transition process that we will undertake.

The study of SR is not restricted to images taken from nature or human beings on a daily basis. Additionally, the researchers have adapted it to medical imaging. Researchers may use publicly available datasets to train networks for SR tasks. However, the currently

available SR datasets are seldom focused on histopathology. Due to a lack of substantial SR datasets in histopathology, only a few investigations have been conducted. The section 4.1 summarises these studies and their findings about the use of SR methods in medical imaging.

2.1 Histopathology Image Analysis

Digital pathology dates all the way back to the twentieth century, when specialized equipment was employed to record information from a microscope [16]. Telepathology is a notion that has existed for about 50 years. It involves sending microscope pictures between far-away sites. However, it is only in the last decade that pathology has started practically digitalizing, shifting from an analog to an electronic setting [17]. The term "digital imaging" refers to digitally capturing, storing, and presenting scenes. Compared to standard digital imaging methods that process static pictures using cameras, whole slide imaging is a more sophisticated and regularly used tool in modern pathology [18]. This approach consists of two stages. To start, digitizing a whole glass histology or cytopathology slide needs the use of a specialized scanner. Second, a virtual slide viewer displays the acquired digital content. Although the scanning operation might be time-consuming with the present technology, this approach can quickly visualize whole slides in high resolution. A full slide image often contains billions of pixels and occupies between 200 MB to several GB [19]. Since a WSI captures data on the whole tissue sample, the amount of a patient's data collected from a biopsy may reach TB during the course of therapy. The WSI makes a substantial contribution to systematic image analysis by directing it toward microscopic morphology [20].

The purpose of digital pathology analysis is to aid in determining the extent of cancerization in tissue and in forecasting its prognosis. One of the critical endeavors in digital pathology is exploring alternative ways to current procedures for histopathology image analysis. The methods examine a surgical tissue or biopsy. Computer-Assisted Diagnosis (CAD) of histopathology images is practical due to repetitive patterns that end in the study of texture and automated recognition [21]. To achieve CAD, two phases of image scan and process are combined with traditional histopathology. As a relatively new technology, whole-slide imaging allows assessing various techniques to analyze high-resolution images. However, problems such as storage, certifying the image, labeling (annotations), and scan quality are still in practice [22].

Previously, scientists utilized conventional Machine Learning (ML) algorithms to evaluate tissue slides in digital pathology. These approaches include the following steps: The first step is to acquire an image. This stage allows for the acquisition of a full glass slide in bright field or fluorescent settings, at a magnification equivalent to that of a microscope, using high-throughput, automated digital pathology scanners. The second is to use pre-processing methods, which include several fundamental tasks such as color normalization (staining) and illumination. Thirdly, morphological properties of structures such as nuclei and glands are detected and segmented. Fourth, feature extraction was tackled by considering features at the object and relationship level in addition to global and window-based multi-resolution features. Fifth is feature selection for improved generalization and run-time efficiency, often paired with dimensionality reduction approaches. Finally, the characteristics of an image allow for proper categorization [23]. However, deep learning has shown state-of-the-art performance in various applications, including digital pathology analysis. While traditional machine learning approaches often need explicit domain-specific features to learn representations, deep learning techniques may do so without them. Deep learning methods need less effort than typical machine learning techniques to extract features [20].

With recent research advancements, Deep Learning (DL) methods as an alternative could surpass traditional machine learning approaches. In the DL techniques, nonlinear transformations are composited over the image pixels, which leads to a suitable abstract representation for the learning process. Internal network parameters are randomly established at the start and updated through error-backpropagation. In massive data sets, deep learning may identify complicated structures. High amounts of training data typically result in better performance. As a concern, magnification level plays a significant role because of patterns; therefore, the selection of the region of interest and image or patch size gain attention. The first step of applying DL to the WSI is data acquisition; the input data (i.e., the patch of WSIs) is then augmented by rotations, scaling, flipping, and other augmentation methods. A variety of DL methods are exercised at this stage. The nature of the desired problem finally derives the solution, such as classification, segmentation, or detection [24].

One of the substantial issues in both ML and DL methods is *supervision*. Researchers extended the ideas to unsupervised learning methods in which samples may not be fully labeled. Unsupervised learning includes the process of reconstructing an input based on the original image. Afterward, it is used as the feed of a supervised method. Besides, convolutional neural networks as a supervised learning method enhance attributes like minimum pre-processing time, robustness to distortions, and variability of the pattern.

Generally, the challenging tasks in CAD could be categorized into three groups. The first group is detecting the region of interest in which the area can be very small. Classification of images is the second group in which the architecture of patterns in tissue samples is the basis. The third group refers to either localization or segmentation of particular tissue structures. For instance, delineating a boundary around glands belongs to the third group.

In summary, these findings provide a potential mechanism for workload reduction of pathologists and sensible coordination. Complementary to the mentioned benefits, the most satisfactory outcomes are delivered by DL methods and compelling viewpoints for the future. The developments are very much the critical component in future attempts to overcome at least three types of imperfections:

- The sensibility of models to the fine-tuning for particular sets
- The impact of observer variability on parameter adjustment of methods
- The lack of robustness of the methods which utilize images

This research is an attempt to delimitate the capability of improved interaction between variant scanned images and pathologists. Also, creating methods that enable training from weak or unlabelled data might provide a vital area for future research.

2.2 Image-to-Image Translation

Numerous problems in computer vision may be framed as image-to-image translation challenges, which imply converting an image from one domain to another when sufficient training data is available. The computer science field has already demonstrated effectiveness in this direction, with Convolutional Neural Network (CNN) establishing themselves as the de facto standard for various visual prediction tasks. While CNNs learn to minimize a loss function that assesses the quality of results, a considerable amount of human labor is still crucial to obtain effective losses [25]. Multiple topics employ the "translation" problem, including super-resolution [26], colorization [27], inpainting [28], style transfer [29], etc [30]. For example, image SR may be thought of as the problem of mapping a low-resolution image to a similar high-resolution image, and colorization can be thought of as the problem

of mapping a gray-scale image to a similar color image. The topic may be examined in supervised and unsupervised learning situations. Supplied images from multiple sites are matched in the supervised context. When the image translation problem is seen through the perspective of probabilistic modeling, the major challenge is to uncover a shared distribution of images across domains. In the unsupervised setting, the two sets include images from two marginal distributions in two separate domains to infer the joint distribution from the images. According to coupling theory, [31], an infinite number of joint distributions may be generated from the specified marginal distributions. Consequently, deriving the joint distribution from the marginal distributions is a highly ill-posed problem [26].

2.2.1 Single Image Super-Resolution (SISR)

Image super-resolution is a class of image-processing (“translation”) techniques that aim to retrieve a high-resolution image based on a lower-resolution image. Image SR has a history of advancement. During the subject’s early proposals, SR was one of the early ways to increase digital consumer images, improve printer resolution, and convert PAL or NTSC resolution video to HDTV quality video. As was the case with general resampling, SR was a common technique carried out by technology such as digital cameras and computers [32]. Over the last decade, deep learning has been used to tackle this image enhancement challenge in computer vision. Image SR manifested itself in a variety of ways depending on the application. Among the most important types, lightweight versions were produced for usage on mobile devices and for rapid processing; [33] [34] for video improvement, multi-frame approaches were provided; [35] and for scenes reference-based super-resolution was presented [36]. However, for this study, single-image SR is the most relevant strategy since the WSI are images, and the additional images required for the other approaches may not always be available. Additionally, when quick techniques are used, accuracy is often compromised to provide a more compact solution that can be implemented on smaller devices, which is irrelevant for this thesis application [37]. The single image super-resolution is a subcategory of image SR in which a network uses a single low-resolution image to reconstruct its’ high-resolution mapping. The essence of the problem has been proven by multiple real-world applications ,including, but not limited to, Medical Imaging [38] [39] [40] [41] , and security [42] [43] [44]. In addition, image super-resolution improves the performance of methods for other tasks [45] [46] [47]. SISR is a considerably challenging issue which also is an ill-posed problem due to its problem definition. A single low-resolution image can act as the respective lower-resolution image of multiple high-resolution images. Conventional techniques to solve this started to emerge a few decades ago [48]. Among the

classical methods, researchers introduced methods based on predictions [49], edges [50], statistics [51], patches [52], and sparse representation [53].

Deep learning has primarily affected image SR. In this area, state-of-art results have been achieved during the past five years. Specific benchmarks for this task are also designed, which are briefly named in the section dataset. A researcher could attain leading performances assessed by a vast range of metrics on the various SR benchmarks. Deep networks that have been used to address this problem are from the very early CNN such as SRCNN [26] to more recent ideas like Generative Adversarial Network (GAN), which have been employed in SRGAN [54].

Deep learning SR networks have differences in aspects like the architecture of the network [55] [56] [57], loss function [58] [59], and learning strategies [60] [61] that are discussed in the following sections.

2.2.2 Problem formulation

The main goal of image SR is to find the most relevant HR image \mathbf{I}_{hr} that corresponds to a LR image \mathbf{I}_{lr} . We assume that the LR image is an image of $l_h \times l_v$ pixels (where l_h and l_v are the number of pixels in horizontal and vertical axis, respectively); therefore, LR image consists of $s_{lr} = l_h \times l_v$ pixels and the $s_{hr} = m \times s_{lr}$ is the number of pixels of the HR image \mathbf{I}_{hr} . The integer parameter m is the factor that shows the increase in the image size. Now, the degradation is defined as equation 2.1.

$$\mathbf{I}_{lr} = \mathcal{D}(\mathbf{I}_{hr}; \delta), \quad (2.1)$$

Here \mathcal{D} is the degradation function and δ corresponds to the related parameters (e.g. kernel size, noise, scaling factor, etc.). The degradation function is assumed to be unknown in many problems called blind SR; however, we can consider it known if the degradation is digitally applied. The approximate higher resolution image \mathbf{I}_{sr} , which is called the SR image of the low-resolution image, is then constructed equation 2.2.

$$\mathbf{I}_{sr} = \mathcal{F}(\mathbf{I}_{lr}; \theta), \quad (2.2)$$

Here \mathcal{F} and θ correspond to the SR function and the approximation parameters, respectively. The degradation model could be modeled by a down-sampling function applied to the high-resolution image as equation 2.3.

$$\mathcal{D}(\mathbf{I}_{\text{hr}}; \delta) = (\mathbf{I}_{\text{hr}}) \downarrow_{\mathbf{d}}, \{\mathbf{d}\} \subset \delta, \quad (2.3)$$

where $\downarrow_{\mathbf{d}}$ denotes the downsampling operation and $\{\mathbf{d}\}$ are the parameters of it. Many downsampling methods are introduced in the literature [62] [63]. Among these methods, bicubic downsampling is commonly used in image SR applications, although other methods are used to operate the downsampling. The bicubic upsampling method has been covered in section 2.2.3. Generally, it is possible to model the downsampling as equation 2.4.

$$\mathcal{D}(\mathbf{I}_{\text{hr}}; \delta) = (\mathbf{I}_{\text{hr}} \otimes k) \downarrow_{\mathbf{d}} + n_{\zeta}, \{k, d, \zeta\} \subset \delta, \quad (2.4)$$

Here $\mathbf{I}_{\text{hr}} \otimes k$ denotes a kernel of k convolution with the high-resolution image to apply a filter (e.g., blur) on the image. n_{ζ} is the gaussian noise added to the model with an average of zero and the standard deviation of ζ . The model described by equation 2.4 has been proven to have more relativity with real-world problems [64].

Finally, the image SR objective is formulated as equation 2.5.

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(\mathbf{I}_{\text{sr}}, \mathbf{I}_{\text{hr}}) + \lambda \Phi(\theta), \quad (2.5)$$

Here, $\mathcal{L}(\mathbf{I}_{\text{sr}}, \mathbf{I}_{\text{hr}})$ is the loss function which is measuring the difference between the generated SR image and the ground truth high-resolution image. The regularization term is formulated by $\Phi(\theta)$, and λ is the trade-off parameter. The losses are usually a combination of multiple functions.

2.2.3 Interpolation

Interpolation of the data is a part of either upsampling or downsampling. Here, the upsampling method has been briefly explored, while the downsampling can be analogously described. In image resampling, the main aim is to predict pixel values based on the available data. However, this task has been conventionally investigated to introduce relatively fast and easy methods. Some of these methods are the nearest neighbor, bilinear and bicubic interpolations. These are aimed at constructing a smoother image. Images of these methods are shown in Figure 2.1 [65].

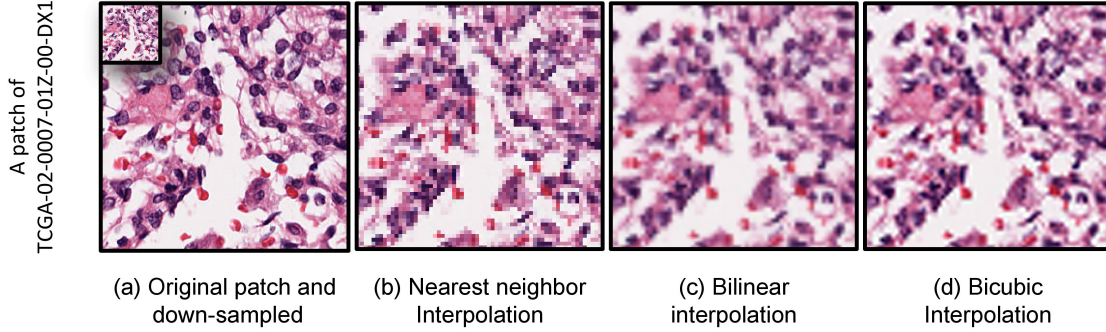


Figure 2.1: A TCGA brain patch of the size 256×256 pixels and the interpolation of its down-sampled patch of the size 64×64 pixels; (a) original image and the down-sampled image in top right, (b) nearest neighbor interpolation, (c) bilinear interpolation and (d) bicubic interpolation

Bicubic Interpolation

The bicubic method takes 16-pixel values into account to compute the smoothest result. On the other hand, a more basic method of bilinear would benefit from 4-pixel values. In the bicubic method, a polynomial technique is used to enlarge the image. The surface of the image function f is then formulated as equation 2.6.

$$f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 c_{ij} x^i y^j, \quad (2.6)$$

where x and y are the pixel-wise location of the image points. In order to solve this surface equation for additional points, the coefficient of c_{ij} are required; Besides, the derivatives f_x, f_y, f_{xy} are computed as equation 2.7 2.7 2.9.

$$f_x(x, y) = \sum_{i=1}^3 \sum_{j=0}^3 i * c_{ij} x^{i-1} y^j, \quad (2.7)$$

$$f_y(x, y) = \sum_{i=0}^3 \sum_{j=1}^3 j * c_{ij} x^i y^{j-1}, \quad (2.8)$$

$$f_{xy}(x, y) = \sum_{i=1}^3 \sum_{j=1}^3 i * j * c_{ij} x^{i-1} y^{j-1}. \quad (2.9)$$

Now, having the corner points (0,0), (0,1), (1,0), and (1,1) in the above equations, the problem is solved. The interpolation can be then simplified as the equation 2.10.

$$f(x, y) = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix} \begin{bmatrix} c_{00} & c_{01} & c_{02} & c_{03} \\ c_{10} & c_{11} & c_{12} & c_{13} \\ c_{20} & c_{21} & c_{22} & c_{23} \\ c_{30} & c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix}. \quad (2.10)$$

2.2.4 SR Deep Networks

During the past few years, SISR, like other areas, has significantly improved. These improvements were not possible without deep neural networks. Overall, the deep neural networks for SR are categorized into four main types. The categories with samples of the successful networks are discussed in section 2.2.5. Then the loss function and evaluation metrics are given in sections 2.2.6 and 2.2.7, respectively.

2.2.5 Architectures

Researchers have proposed a range of network architectures. Although these networks exhibit considerable distinctions, Wang et al. classified them into four categories [14]. The general categorization of frameworks is given in Figure 2.2.

One of the first networks for image super-resolution is entitled SRCNN, which Dong. et al. introduced in 2014. The network they introduced enhanced the classical mapping methods between the LR and HR images. Their success gained attention, and many other networks were proposed ideas based on their initial proposal. Their network design

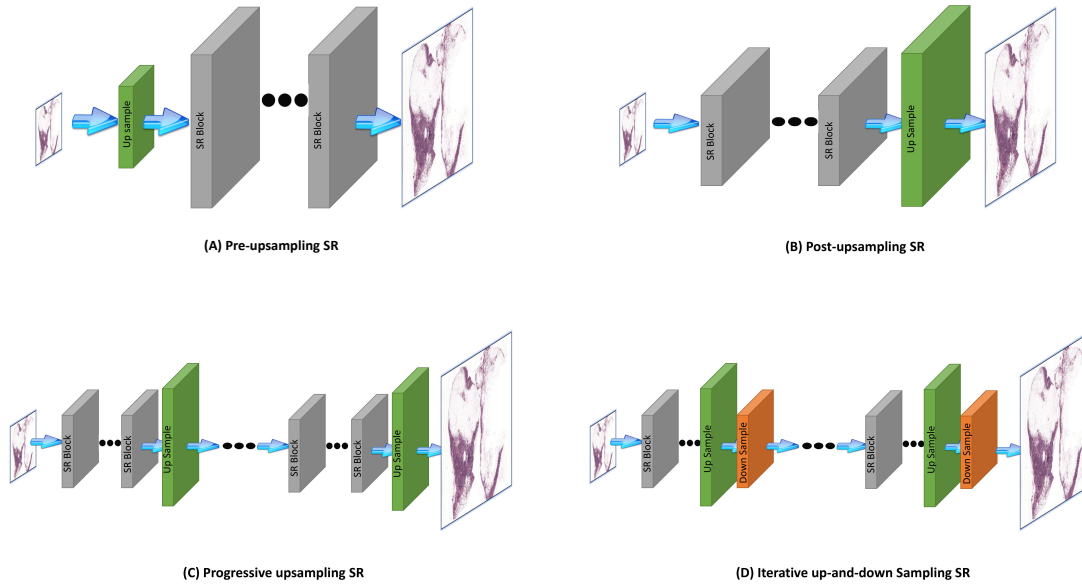


Figure 2.2: Four general frameworks of deep neural networks used for image super resolution including; (a) Pre-upsampling SR (b) Post-upsampling SR (c) Progressive upsampling SR (d) Iterative up-and-down Sampling SR; Grey boxes, SR blocks; Green boxes, upsampling units; Orange boxes, downsampling units

could best fit the category (a) of Figure 2.2 where a CNN followed a pre-upsampling to enhance the initial upsampling. Despite their success, these structures faced issues like noise amplification. Later, researchers showed these networks' considerably lengthy operation time due to the higher dimensional computation framework comparing other networks [26].

Post-upsampling networks then solve the computational expense problem. Researchers benefitted from similar ideas; however, faster training and testing time. To name a few, RCAN and SRGAN were among the networks that benefitted from this framework. These networks were successful enough to consider state-of-the-art results. Although RCAN presented higher accuracy values in many studies, SRGAN presented more realistic images by hallucinating the textures [66] [67].

The benefits of the framework (b) were outstanding; however, there were shortcomings. The single-step up-sampling module made the training an arduous task when a higher

magnification, like 4 or 8, was required. To help this issue, researchers that experienced the high-quality results of progressive networks such as StyleGAN implemented similar approaches in the areas of SISR [68]. One of these networks that were upsampling iteratively is ProSR which achieved relatively high-performance [61].

Finally, the last framework of iterative up-and-down sampling super-resolution was investigated since even the framework (c) encountered some problems. For instance, the learning strategies required to train them were complicated. In contrast, framework (d) applies back-projection refinement. To name a successful example, SRFBN fits this category which benefits from feedback block and offers better representations [69].

2.2.6 Loss Functions

In deep learning, loss functions play a crucial role in guiding the model to optimize with desired weights. In image super-resolution, the optimal outcome is a network that enables a quality reconstruction; therefore, the loss functions are designed to help the training in leading the network to reconstruct an image close to the HR image. Here we discuss three of the loss functions which are used in many studies.

Reconstruction loss evaluates the correctness of picture production based on the difference in pixel values between the intended and created images. This loss would instruct the network to create pictures with identical pixels. For example, produced picture colors and histograms tend to be most comparable for images with more blurriness. This is because an edge that arrives just one pixel after where it was expected to occur results in a significant loss value. As a result, the pixel value numbers are often averaged to achieve lower loss values. On the contrary, perceptual loss influences the content of pictures. This suggests that this loss function prioritizes the perception of similar shapes and structures instead of pixel value equality. It implies that a pixel change in the appearance of an edge, for example, may not significantly impact the loss value; nevertheless, if a shape of perception system interest is absent, the loss value increases. Ultimately, the adversarial loss causes two networks to compete against one another for superior outcomes. The generator network would be steered toward more realistic-appearing pictures, but the discriminator network would seek distinctions between genuine and created images. This loss should result in the development of pictures that resemble the originals independent of pixel or content comparison.

Reconstruction Loss

Reconstruction loss or pixel loss indicates the similarity of the generated SR image and the desired HR image. This is computed by L1-norm or L2-norm of the images differences as equations 2.11 and 2.12.

$$\mathcal{L}_{rec.}(I_{HR}, I_{SR}) = \frac{1}{v} \|I_{HR} - I_{SR}\|_1, \quad (2.11)$$

$$\mathcal{L}_{rec.}(I_{HR}, I_{SR}) = \frac{1}{v} \|I_{HR} - I_{SR}\|_2, \quad (2.12)$$

where $v = h \times w \times c$ is the volume which is the multiplication of height and weight and number of channels. L1-norm showed better performance from sharpness in addition to easier convergence. The sharper images emerge since the L2-norm penalizes the significant deviation exceedingly while minor errors slightly.

Perceptual Loss

A *perceptual loss*, named content loss, evaluates the perceptual similarity of the generated image with the HR image. This is done by comparing the semantic content of the images using a pre-trained network (e.g., VGG16). Researchers computed the distance of the l -th layer of the network like equation 2.13.

$$\mathcal{L}_{percept.}(I_{HR}, I_{SR}) = \frac{1}{v} \|\phi^{(l)}(I_{HR}) - \phi^{(l)}(I_{SR})\|_2, \quad (2.13)$$

Where $\phi^{(l)}(.)$ denotes the output of the l -th layer of the pre-trained network, in other words, the perceptual loss guides the network based on the hierarchical image features of a network that has been trained for a task (e.g., ImageNet classification) [70].

Adversial Loss

The application of GANs has reached the SR as well. GAN-based training generally consists of two main networks: the generator and the discriminator. These two networks compete

against each other for training purposes. Consequently, the adversarial loss is defined as equations 2.14 and 2.15 for the generator and discriminator networks, respectively.

$$\mathcal{L}_{GAN_G}(I_{HR}; D) = -\log D(I_{SR}), \quad (2.14)$$

$$\mathcal{L}_{GAN_D}(I_{HR}, I_{SR}; D) = -\log D(I_{HR}) - \log(1 - D(I_{SR})), \quad (2.15)$$

where D denotes the discriminator for a binary decision whether the image is real.

2.2.7 Reconstruction Quality Measurement

Deep networks can produce higher-quality images that seem identical to the original image in image SR. However, since this mapping is the inverse of down-sampling, which results in the loss of significant amounts of information, the resulting picture may not have identical pixel values to the original. In pathology, one of the primary goals of these images is to aid in the clinical evaluation of human experts. As a result, the most reliable method of assessing the image quality is by human inspection [71]. Although human review ensures dependability, most approaches are computational as the size of an image dataset increases. These distinctions represent the apex of a method’s capacity. Among the most often used measures, two metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index Method (SSIM) in this field, are described in the following section for use in the analysis of image SR.

2.3 Image-to-Image Distance

One of the primary objectives of image translation is to determine the method’s effectiveness by comparing the produced image’s deviation to the original. This subject is often referred to as Image Quality Assessment (IQA) in the literature [72]. This is not, however, the only circumstance in which a distance may be favorable. The difference in image generation is a critical factor in determining the location of an object while doing unsupervised detection. The following paragraph will cover several of the most often utilized approaches for determining generational differences, which may help achieve the thesis’s objective.

2.3.1 Peak signal-to-noise ratio (PSNR)

PSNR is used to calculate the ratio of the maximum possible signal intensity to the power of the distorting noise that reduces the representational quality of the signal. The signal represents the original image, while the noise denotes image generation errors experienced by comparing the original image to the generated image. It is a frequently used metric that compares the rate of maximum pixel values to the image’s mean squared error (MSE). The PSNR ratio of two pictures is given in decibels. Because signals have a reasonably high dynamic range, PSNR is often calculated as the decibel scale’s logarithm term. This dynamic range is defined as the difference between the highest and least possible values, which vary depending on the problem [73]. PSNR of the two images I_{SR} and I_{HR} is then computed by equation 2.16.

$$PSNR = 10 \log_{10} \left(\frac{L^2}{\|I_{HR} - I_{SR}\|_2^2} \right), \quad (2.16)$$

where L is the maximum pixel-value equal to 255. Despite the common use of this metric, it does not reflect the human perception. [74]

2.3.2 Structure Similarity Index Method (SSIM)

The primary function of the human visual system is to identify visible structures. The Universal Quality Index (UQI), or Wang–Bovik Index, was established in 2001 by Zhou Wang and Alan Bovik [75]. This grew into the present version of SSIM. The Structural Similarity Index Method is a technique that is based on perception. In this approach, image degradation is seen as a change in the perception of structural information. This metric comprises three main components: luminance, contrast, and structure. Luminance refers to reducing the visibility of an image’s distortion section around the image’s edges. Contrast, on the other hand, is a term that relates to the act of decreasing the effect of the appearance of textural imperfections in a picture. The term structural information refers to pixels that are strongly reliant on one another or that are spatially constrained. These highly interconnected pixels offer crucial information to the image domain about visual objects [76] [73]. SSIM metric determines the perceived quality of images and videos. It makes a comparison between the original and reconstructed images. It is formulated as

$$\text{SSIM} = \frac{2\mu_{SR}\mu_{HR} + k_1}{\mu_{SR}^2 + \mu_{HR}^2 + k_1} \times \frac{\sigma_{SR,HR} + k_2}{\sigma_{SR}^2 + \sigma_{HR}^2 + k_2}, \quad (2.17)$$

where μ_{SR} and σ_{SR} are the mean and variance of the SR image, μ_{HR} and σ_{HR} are the mean and variance of the HR image, and $\sigma_{SR,HR}$ is the covariance between SR and HR image. Finally, k_1 and k_2 are the relaxation constants.

SSIM is commonly used in different fields, like medical imaging. Researchers commonly use this metric, despite its limitations [77]. It is practically used in various medical imaging domains like Histopathology, CT scans, and MRI, and ultrasound [78] [79] [80].

Multi-Scale Structural Similarity Index Method (MS-SSIM)

The structural similarity image quality paradigm assumes that the human visual system is highly specialized for acquiring structural information from scenes to approximate perceived picture quality. Multi-Scale Structural Similarity Index Method (MS-SSIM) is a multiscale structural similarity methodology that includes more viewing scenarios than the previous single-scale method. An image synthesis technique is developed to calibrate the relative importance of scales. The Multi-Scale Structural Similarity Index Method (MS-SSIM) analyses a large number of structurally similar images at various image sizes. MS-SSIM compares two images of the same size and resolution. When comparing two images, changes in luminance, contrast, and structure are analyzed similarly to SSIM. It may outperform SSIM on subjective image and video datasets, depending on the task [81].

2.3.3 Feature Similarity (FSIM)

Index of Feature Similarity The method maps the characteristics and compares the similarity of two images. To further understand Feature Similarity Index Method (FSIM), we must first define two conditions. They are as follows: Phase Congruency and Gradient Magnitude. At the time, phase congruency was a novel approach for recognizing visual characteristics. One of the essential properties of phase congruency is that it is insensitive to picture light change. Furthermore, it is capable of detecting other interesting traits. It focuses on the image's characteristics in the frequency domain. Contrast does not affect phase congruency [82].

2.3.4 Gradient Magnitude Similarity Deviation (GMSD)

This metric was designed not just to provide high-quality prediction accuracy but also to be computationally efficient. Because of the rising proliferation of high-volume visual data on high-speed networks, metric efficiency is becoming more crucial. They provide Gradient Magnitude Similarity Deviation (GMSD), a novel, effective, and efficient IQA model. Picture gradients are affected by image distortions, although various local structures of a distorted image suffer varying degrees of deterioration. This prompted the researchers to investigate the usage of a gradient-based local quality map with global variance for overall picture quality prediction. They discovered that paired with a unique pooling approach, the pixel-wise gradient magnitude similarity between the reference and distorted pictures may reliably predict perceptual image quality. The resultant GMSD algorithm outperformed most earlier approaches in terms of prediction accuracy [83].

Multi-Scale Gradient Magnitude Similarity Deviation (MS-GMSD)

This measure builds on the work of GMSD by proposing a more efficient metric. First, a unique similarity index was presented, allowing the masking value to be tuned to mimic the human visual system more closely. The authors suggested a multi-scale GMSD technique incorporating luminance distortion ratings at multiple scales. Furthermore, based on the measure, they presented a method for quantifying chromatic aberrations in YIQ color space. Multi-Scale Gradient Magnitude Similarity Deviation (MS-GMSD) is calculated by combining luminance and chrominance scores [84].

2.3.5 Visual Saliency-based Index (VSI)

During the past decade, psychologists, neurobiologists, and computer scientists have extensively researched visual saliency to determine which picture portions will draw the most attention from the human optical system. Intuitively, VS is similar to IQA in that suprathreshold aberrations may significantly impact VS maps of pictures. With this in mind, a solution was given that was a simple but very efficient complete reference IQA approach using VS. VS has a dual purpose in the proposed IQA paradigm. First, VS is employed as a feature for calculating the distorted image's local quality map. Second, VS is used as a weighting mechanism when pooling the quality score to indicate the significance of a particular area. The suggested IQA index is referred to as a Visual Saliency-based Index (VSI) [85].

2.3.6 Mean Deviation Similarity Index (MDSI)

Researchers suggested a trustworthy whole reference IQA model based on gradient similarity, chromaticity similarity, and deviation pooling. Considering the drawbacks of the frequently used GS for modeling the human visual system (HVS), a new GS was developed using a fusion approach that was more likely to follow HVS. The suggested process is a quick and easy way to compute the joint similarity map of two chromatic channels to assess color changes. The recommended CS map was proven more efficient and produces equivalent or higher quality predictions than a regularly used formulation in the literature. Motivated by a previous article that employs standard deviation pooling, their study presented a universal formulation of the DP to calculate a final score from the suggested GS and CS maps. This suggested DP formulation takes use of the Minkowski pooling as well as a proposed power pooling [86].

2.3.7 Learned Perceptual Image Patch Similarity (LPIPS)

While assessing the visual closeness of two pictures is almost straightforward for humans, the underlying mechanisms are regarded to be highly complicated. Despite this, the most extensively used perceptual measures today, such as PSNR and SSIM, are simplistic, hollow functions that ignore a large number of subtleties in human perception. The deep learning community recently discovered that characteristics of the VGG network trained on ImageNet classification were beneficial as a training loss for visual synthesis. Learned Perceptual Image Patch Similarity (LPIPS) metric may not be well-suited for use as a loss function when training a network. This is despite the fact that it may be a practical quantitative assessor closely correlating with the human impression of picture quality.

Researchers demonstrated that deep features trained on supervised, self-supervised, and unsupervised goals explain low-level perceptual similarity remarkably effectively, exceeding prior, widely-used measures [87].

2.3.8 Haar Perceptual Similarity Index

In most real cases, images and videos cannot be compressed or transported without introducing distortions that a human viewer would eventually notice. Most image and video

restoration techniques, such as inpainting and denoising, aim to enhance the viewing experience for humans. Correctly anticipating a picture’s similarity to an undistorted reference image, as viewed subjectively by a human viewer, may result in significant benefits in any transmission, compression, or restoration approach.

In the following respects, Haar Perceptual Similarity Index (HaarPSI) outperforms the previously discussed measures: It provides better correlations with human opinion evaluations on massive benchmark datasets in almost every case. It is significantly more effective and fast to compute than most other measures [88].

Chapter 3

Compound Magnification Framework

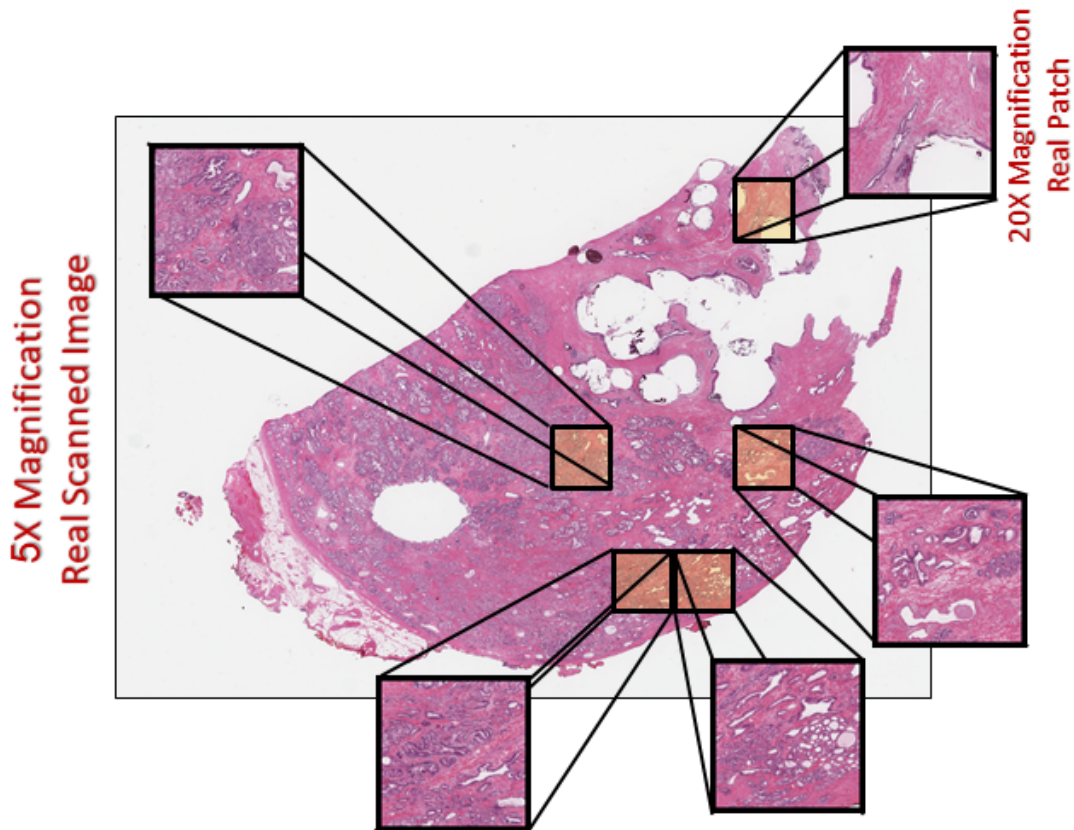
The gigapixel images that are presented as WSIs are generally stored in SVS format. These images contain valuable information in deep magnifications like 20X and 40X. Consequently, this has resulted in scans with sizes even more enormous than $100,000 \times 200,000$ pixels. Here we propose a Compound Magnification Framework (CMF) as an alternative to help with smaller size scans. Figure 3.1 shows a TCGA slide in the proposed CMF.

As shown in Figure 3.1, a CMF contains two crucial sets of information. One is the low-resolution image with a 5X magnification here corresponding to the whole area of the slide. This low-resolution image is a 4X down-sampling of the original WSI at 20X. Besides, two are a set of high-resolution patches presented as 20X magnification patches. The patches are the same size tiles of the original slide at high resolution. CMF is generated based on a WSI, and synthetic WSI generation is based on the CMF. The detailed problem definition of transformations between CMF and WSI is given in the next section.

3.1 Problem Definition

The spatially constrained CMF refers to a low-magnification whole slide image in addition to several patches. The generation of this framework is done through a procedure that we call "Spatially constrained compression." Based on this procedure, a WSI is turned into a CMF. The CMFs are meant to represent a smaller size of their WSI while capable of reconstruction. Figure 3.2 shows an overview of the compression method.

Compound Magnification Framework (CMF)



Slide: TCGA-XJ-A83F-01Z-00-DX1

Figure 3.1: A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches). The 5x image can be upsampled to 20X when needed for visual inspection. The 20X patches depict relevant diagnostic histology, hence preserved in original resolution.

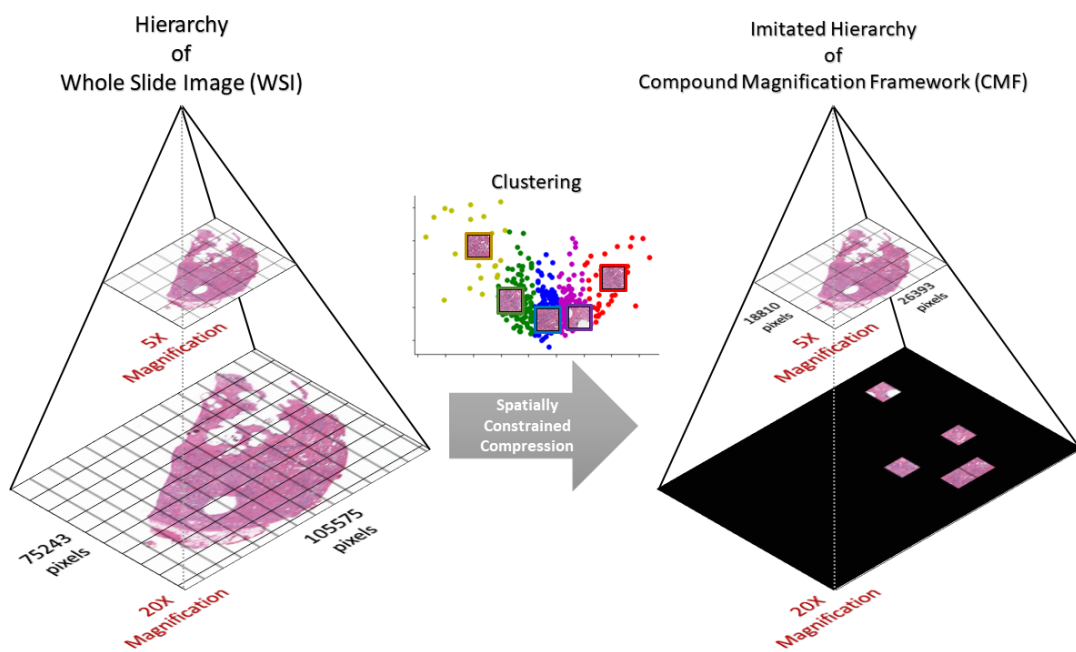


Figure 3.2: A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches)

As shown in Figure 3.2, the compression route transforms a WSI into a CMF. In this procedure, two parts are considered. First, the WSI is presented in the hierarchy of magnifications. In this hierarchy, the 5X magnification is fully available in the CMF. However, the high-resolution level of the slide is mainly dropped other than a few patches. The patches are selected based on the result of a clustering method. The clustering method applies to the patches extracted from the original slide at a low magnification level. Lower magnification patches are equal to smaller size patches that result in faster processing for patch selection. The selection of patches at low magnification is important due to the extremely high number of patches in a slide.

On the other hand, the opposite procedure is named “Super-resolution decompression”. This task refers to the problem of reconstruction of a synthetic WSI based on the CMF. Figure 3.3 gives an overview of the process.

As shown in Figure 3.3, the super-resolution decompression is responsible for creating a synthetic WSI based on a CMF. This procedure is done with the help of an SR deep network where the available information in the CMF is utilized to generate the synthetic WSI. The goal is to create a synthetic WSI as much as possible close to the original WSI.

3.2 Spatially Constrained Compression

Construction of a CMF is done through a procedure that we named “spatially constrained compression”. The high-resolution slide is downsampled to the lower resolution. In addition, several patches are extracted at high resolution. The patches, in addition to the low-resolution slide, are called the CMF. This structure is beneficial because of two main reasons. While the reconstruction of the synthetic WSI out of a CMF could be done based on single-image super-resolution, having the original high-resolution patches would help ensure the availability of accurate information as a reference. The references can be helpful for assessments, including medical doctor diagnosis, computer-assisted diagnosis, and computer assessment of methods. Second, the benefit of having original high-resolution patches includes the possibility of practicing more complex techniques. These complex methods include but are not limited to reference-based super-resolution. In a reference-based super-resolution method, the network input is at least two images of the lower-resolution and high-resolution. Besides, it is worth noting that the more enhanced process may benefit from multiple references to retrieve the best high-resolution image.

Compound Magnification Framework (CMF)

Synthetic Whole Slide Image

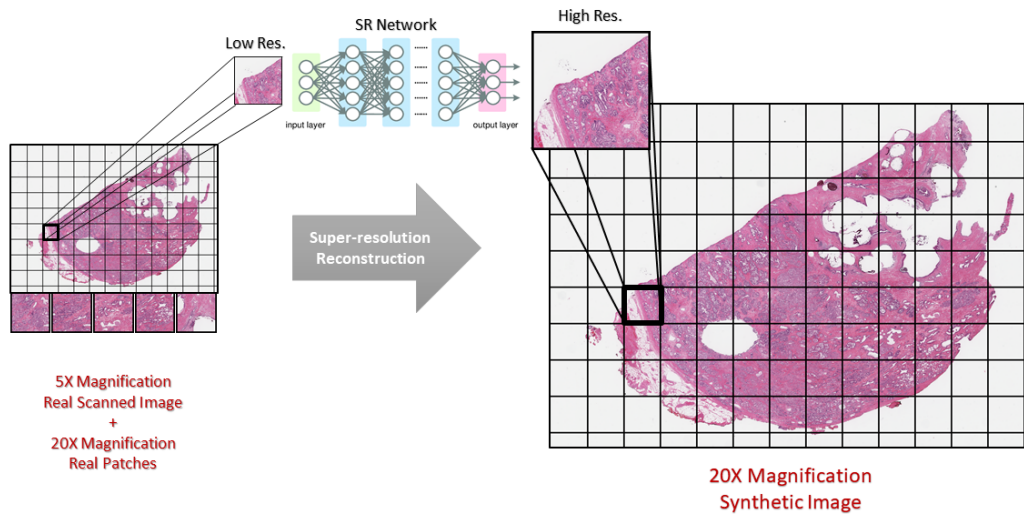


Figure 3.3: A TCGA slide in Compound Magnification Framework (CMF) representing the slide in 5X magnification (low-resolution image) and five 20X magnification patches (high-resolution patches)

As previously discussed, the WSIs are gigapixel images. This means that a large number of patches have resulted from tiling. However, a selection of these patches should be stored. A proper selection should contain images that can be used most in the following step of reconstruction. As discussed, this optimal selection is then dependent on the problem goals. We assume that a good reconstruction assures the other problem needs, such as diagnosis accuracy. Consequently, the stored patches should represent a comprehensive sampling of tissue texture used to help the super-resolution network. Among the methods to select relatively similar texture patches, color-based and feature-based patching have been experimented.

3.2.1 Preliminary Approaches

Although clustering is an unsupervised technique, the assessment of the outcomes could be compared to the labeled data. The TCGA dataset offers labels at the WSI level, while the labels at the patch level are required. In this part, we benefit from the Human Colorectal Cancer dataset to experiment at patch level [89]. The following section includes the dataset description, followed by two clustering method sections based on color and deep features.

Colorectal Cancer (CRC) Dataset

This is a collection of 107,000 non-overlapping picture patches extracted from hematoxylin and eosin (H&E) stained histological images of human colorectal cancer (CRC) and normal tissue. All images are 224x224 pixels (px) @ 0.5 microns per pixel (MPP). Using Maconko’s approach, each picture is color-normalized. Adipose (ADI), backgrounds (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM) are the tissue classes presented in this dataset. Figure 3.5 (a) depicts several examples of the images.

These images were taken manually from N=86 H&E stained human cancer tissue slides from formalin-fixed paraffin-embedded (FFPE) samples obtained from the NCT Biobank (National Center for Tumor Diseases, Heidelberg, Germany) and the UMM pathology archive (University Medical Center Mannheim, Mannheim, Germany). Tissue samples comprised CRC original tumor slides and CRC liver metastasis tumor tissue; normal tissue classes were supplemented with non-tumorous sections from gastrectomy specimens to

improve variety. A portion of the dataset consists of one hundred thousand images that were used to train the networks. For the validation, a collection of 7180 image patches from $N=50$ patients with colorectal cancer is also employed.

Color histogram

Color-based patch selection is an intuitive method that has been previously experimented with in a method; namely, mosaic creation [90]. It has been evident that this method is among the fastest ways to extract features of images. Figure 3.4 image shows a sample of how this procedure is done in our experiment.

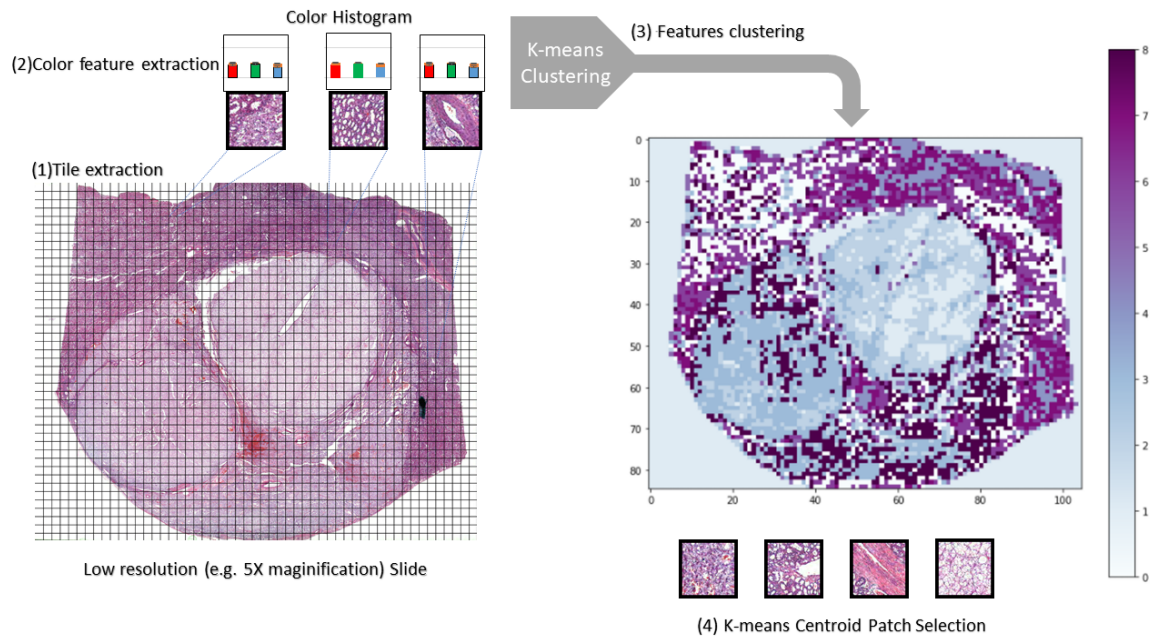


Figure 3.4: High resolution patch selection using WSI color histogram clustering in low resolution; the lower resolution slide is first tiled, then color histograms are computed and centroids of clustering are selected

In this Figure, a WSI is first tiled into same-size small images, then the color features representing the image's histogram are extracted. These features are then clustered based on k-means clustering. It is worth noting that the same randomness value should be used in other parts of the problem to avoid differences in results. Finally, the patches respective

of the cluster centroids are used. The least distant patches to the centroids are used in this problem. However, if more than one patch has a similar minimum distance, randomly, one is chosen.

Deep Clustering with Convolutional Autoencoders (DCEC)

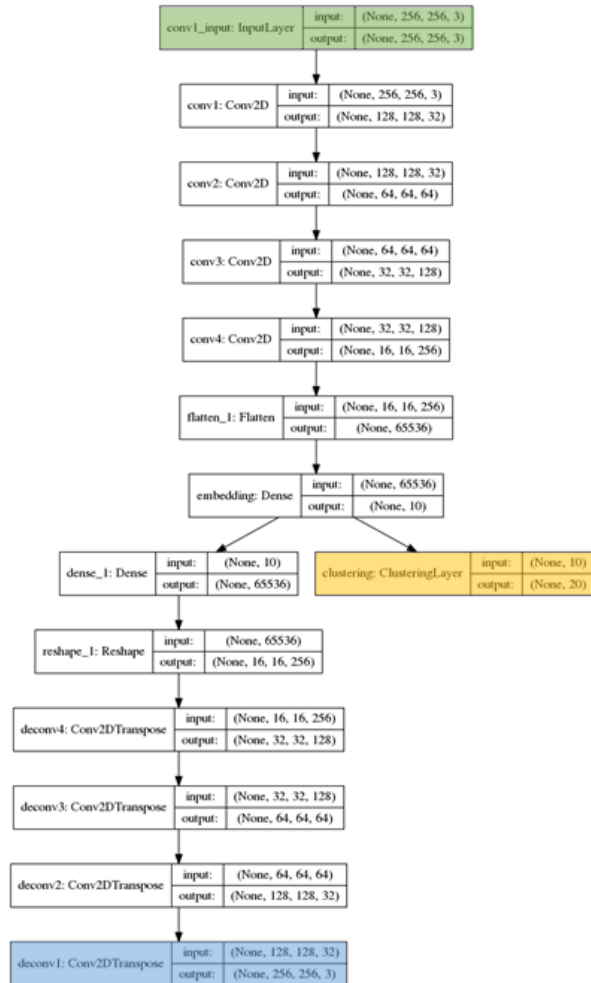
The implemented clustering technique comes from a DCEC network [91]. This is helpful since it is reasonably shallow to operate relatively fast and offers excellent accuracy. The network architecture is modified to Figure 3.5 (c), where the input and output of the autoencoder are shown in green and blue boxes, respectively. The yellow box does the clustering, which uses a feature vector of 10 elements. The feature vector is produced by dense extraction of the flattening of the four convolutional layers.

The network is trained by the CRC training dataset, which has been split into two parts of training and validation data. The validation part takes 20% of the data. The final network is selected based on the evaluation of the validation loss. Where the autoencoder loss is borrowed from DCEC and the clustering loss is borrowed from the study of Unsupervised deep embedding for clustering analysis[91] [92].

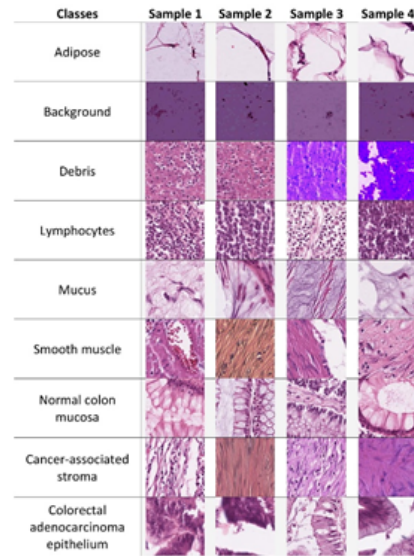
The confusion matrix of the results of the trained network is shown in 3.5. (b). The results are based on the test dataset, which is identified as validation in the CRC dataset.

3.3 Summary

The compound magnification framework is made up of three fundamental elements. The first is the framework that was developed and presented in this study. In addition, a super-resolution network for creating high-resolution pictures has been trained, and the results are given in Chapter 4. The patch selection approach is the third. Two techniques based on color and deep characteristics were presented for ideally rapid selection. However, these approaches need an expert's time and only give basic information about a slide. In order to better handle this issue, an unsupervised unhealthy area detector is included in Chapter 5.



(c) DCEC network architecture



(a) CRC dataset sample images

True Class	1	2	3	4	5	6	7	8	9
1	393			64		1	4	8	
2		407		223					
3			116	65	1	67	1		1
4	4			164	166	92	9	10	9
5	52			2	362	325	10	9	11
6				210	1	207	3	18	5
7	2			53	106	212	64	54	50
8	1			124	68	108	6	4	2
9	4	1		366	38	173	85	124	98
	1	2	3	4	5	6	7	8	9

(b) Confusion matrix of CRC test-set by DCEC Clustering

Figure 3.5: Clustering by deep features; (a) Sample images from CRC dataset used for training, (b) Confusion matrix of clustering based on the deep features of images, (c) The DCEC network architecture which has been trained to extract features (green area is the input, blue area is the decoder output used for training, yellow area shows the deep features used for clustering)

Chapter 4

Super-Resolution Reconstruction

This study suggests using an SR-Deep network in order to rebuild a (synthetic) WSI from a CMF. This chapter describes the use of image SR to histopathology pictures from numerous vantage points, including training and testing. Initially, the history of SR and medical imaging is presented. The dataset used for this investigation is then described. The networks, training, and other outcomes are then offered.

4.1 Background

Medical imaging contains a vast number of image modalities for diagnostic and treatment-panning purposes. To name a few, MRI, fMRI, and PET are among the important ones. Researchers initially investigated classical SR methods on these modalities and contributed to the advancement of the field. Radiology images are usually small (mainly megapixel and not gigapixel sizes) and hence easier to process [93]. However, with the emergence of deep learning, in addition to better results, more complex structures and data like pathology images are being processed.

4.1.1 Super-resolution in Histopathology

Among the earliest super-resolution in histopathology studies was a paper published by L. Mukherjee et al. in 2018. Their study introduced a deep framework for reconstructing

HR images in the pathology field. This was important since the data in pathology does not contain high-level features in comparison to natural images. Their results showed promising outcomes, which they suggested as a comparative counterpart to the expensive scanners [94]. Later, they investigated a recurrent network to enhance the quality by a multi-resolution approach. Lately, Bin Li et al. proposed a framework that benefits from the hierarchical structure of the WSIs and achieved the best results among others. Their practice showed that downsampling could act as a training data solution for these deep networks. The studies were examined against Tissue microarray (TMA) datasets and a two-site whole tissue section dataset [95].

4.2 Dataset Preparation

In deep learning, the dataset is the essential material for training a network, and super-resolution is no exception. Super-resolution deep networks are trained in various methods, from supervised to unsupervised training. The supervised methods require the LR and HR paired images as the training material [60]. These images are usually synthetic pairs based on standard down-sampling methods. During the past few years, researchers have introduced and investigated many SR datasets. Some of the most frequently used SR datasets are presented in section 4.2.1. These datasets are helpful since a considerable number of successful networks are trained based on these images. In addition, transfer learning as a general method of training enhancement of deep networks requires understanding the initial data. Furthermore, looking at the considerations about previously created datasets is essential to create a convenient dataset of super-resolution in pathology.

Histopathology images are presented in WSIs. These images are much larger than images in other areas like face recognition and autonomous driving. WSIs, as gigapixel images, contain valuable information in high resolution. However, due to computational limitations, a complete WSI cannot be processed by conventional graphical processing units. These images are usually split into patches to enable efficient processing. In order to create a dataset, we have first investigated the largest publicly available histopathology dataset in section 4.2.2. Having a relatively complete source of WSIs, the creation of an SR dataset for histopathology is described in section 4.2.3.

The generated dataset enables the investigation of two sides of this problem. First, the compression path, which includes the clustering experiments, is described. Second, the image super-resolution steps for reconstruction are given. Finally, section 4.6 concludes the chapter.

4.2.1 Publicly available SR data-sets

During the past few years, improvements in deep learning required more data, and as a result variety of datasets have been introduced. Datasets offer images with different properties like source domain, quality, and resolution in the image SR domain. SR datasets contain HR images, while some include the paired LR images too. The LR pairs are, in most cases, extracted by a bicubic interpolation with anti-aliasing. Some of the most famous datasets are in Table 4.1.

Table 4.1: Super-Resolution Famous Datasets and Data Overview

Dataset	#images	Avg. Resolution	Img. Format	Img. Category
BSDS300 [96]	300	(435, 367)	JPG	animal, people, plant, building/landscape, food, etc.
BSDS500 [96]	500	(432, 370)	JPG	animal, people, plant, building/landscape, food, etc.
DIV2K [97]	1000	(1972, 1437)	PNG	environment, flora, fauna, handmade object, people, scenery, etc.
General100 [98]	100	(435; 381)	BMP	animal, people, plant, daily necessity, food, texture, etc.
L20 [99]	20	(3843, 2870)	PNG	animal, building, landscape, people, plant, etc.
Manga109 [100]	109	(826, 1169)	PNG	manga volume
PIRM [101]	200	(617, 482)	PNG	environments, flora, natural scenery, objects, people, etc.
Set5 [102]	5	(313, 336)	PNG	baby, bird, butterfly, head, woman
Set14 [103]	14	(492, 446)	PNG	humans, animals, insects, flowers, vegetables, comic, slides, etc.
Urban100 [104]	100	(984, 797)	PNG	architecture, city, structure, urban, etc.

The image datasets cover a range of image sizes, while the number of images provided does not exceed 1000 images. In some studies, to have a more comprehensive range of images and better training results, in addition to image augmentation as a general enhancement method, these datasets are combined for training purposes. For instance, the datasets Set5 and Set14, which are primarily used for test/validation, are usually combined.

Although the datasets in Table 4.1 are widely used, as described, the images provided by these datasets do not cover the medical imaging and histopathology area.

4.2.2 Publicly available Histopathology data

WSIs are the scanned histopathology slides. The Cancer Genome Atlas (TCGA) is the most enormous publicly available dataset of scanned slides and reports. This dataset (available at <https://www.cancer.gov/tcga>) allows researchers to not only experiment and propose new methods but also compare their results efficiently. The TCGA contains over 20,000 primary cancer in addition to the normal samples for 33 cancer types and started

this mission in 2006. The generated data by TCGA is now over 2.5 petabytes and spans genomic, epigenomic, transcriptomic, and proteomic data [105].

TCGA repository (i.e., Genomic Data Commons, GDC) contains 11,007 cases and 30,072 SVS files of the slides. The WSIs of this repository span 26 organs (primary sites) with 32 cancer subtypes. The subtypes are abbreviated by a few letters in the repository, and the same structure is used in this document. Complete subtype names and the distribution of the number of patients in each category are explained in table 4.2. The demographic information attached to each scan includes "Morphology", "Primary Diagnosis", "Tissue or Organ of origin", "Patient Age at the time of Diagnosis", "Tumor Stage", "Age", "Gender", "Race" and "Ethnicity" and some other information like the patients current status (e.g., dead or alive).

4.2.3 Creation of a data-set

In this section, the creation of a dataset of histopathological slides is described. The dataset includes patches, and the labels of the patches are the information attached to the WSIs. This dataset is created to help research in the area of digital pathology. The scanned slides of the TCGA repository include frozen section biopsy WSIs. A frozen section (cryosection) is a laboratory technique that helps to get microscopic analysis of a specimen rapidly. The fast diagnosis is beneficial for the management of the patient during an operation. However, frozen sections are not usually high-quality due to their procedure. These lower-quality slides were dropped to avoid confusion during deep network training sessions.

To assign labels to the images, the information provided by the repository is used. The beneficial information is decided to be the primary diagnosis and the section site (/tissue/organ of origin). These two labels are presented under the "Diagnoses / Treatments" of the slides. Although other information is also available for further diagnosis, we have decided to use the before-mentioned labels. Finally, some slides did not include the required information to create the image labels; these also were dropped. In addition, the classes with a number of WSIs less than 20 were also eliminated. This has been done due to the fact that this low number of images may not result in a comprehensive assessment.

4.2.4 Data-set for histopathology SR

To create a dataset for histopathology SR, some considerations to eliminate the appearance of undesired data. We have first removed the background patches to minimize the artifact since these do not contain valuable information for training. This task has been done and evaluated previously in another study. Second, the slides are cropped into the same size tiles. The tiles are at the size of 640×640 , 320×320 , 160×160 , 80×80 , and 16×16 pixels in three channels of RGB at 40X, 20X, 10X, 5X, and 1X magnification levels. Figure 4.1 provides one sample per cancer subtype.

4.3 Super-resolution reconstruction

4.3.1 Networks

In the reconstruction path, the lower resolution patches are used, and a deep network that is trained for single image super-resolution generates the patches for synthetic WSI. In order to better find a network that can adequately enhance the images, six networks are trained. First, we benefit from the Deep Back-Projection Networks (DBPN). The network benefits from iterative up- and down-sampling layers, which are called stages. The stages include an error feedback mechanism for projection errors. The network can learn multiple content information by its up-and-down-sampling layers [106]. Residual Dense Networks (RDN) are then trained, which exploit the hierarchical features from all the convolutional layers [107]. The third is a very deep residual channel attention network, RCAN, that benefits from residual in residual blocks in a post-upsampling structure [66]. Fourth is a Feedback Network for Image Super-Resolution (SRFBN) that focuses more on feedback mechanism because the human vision system follows a similar method [69]. Fifth is Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR), which is proposed in addition to the multi-scale deep super-resolution system [60]. Finally, a network that we use to do the super-resolution is among the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). This network architecture has achieved some state-of-the-art results in addition to providing a high perceptual index [108].

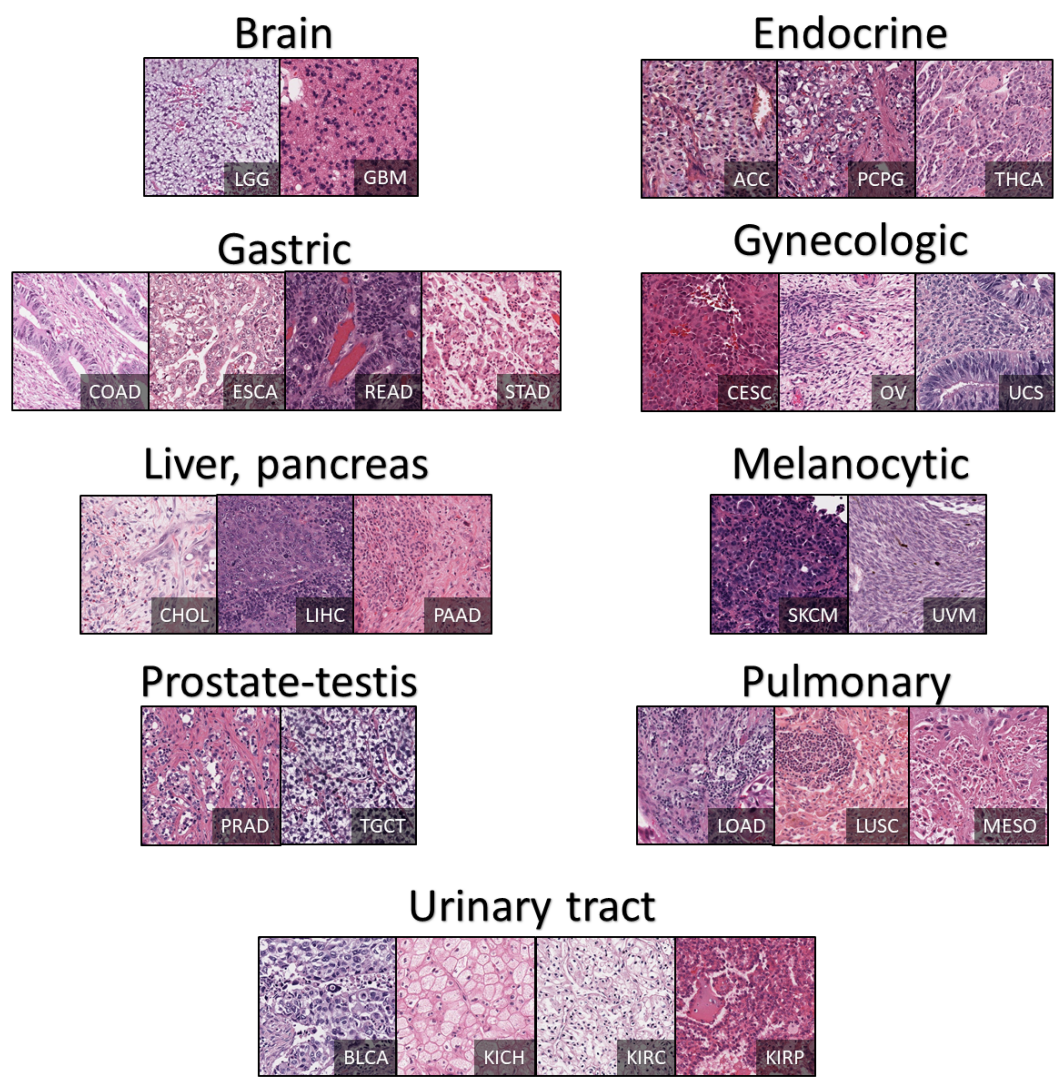


Figure 4.1: Patches of the created SR dataset in nine TCGA cancer subtypes

4.3.2 Training

In order to train the network, the training patches of the extracted SR dataset are used. To achieve the best results, the relativistic discriminator is used for adversarial training. The discriminator classifies the images as not only fake or real but also relatively compares the extent of fake and real images, which enhances the learning procedure. The networks are all trained for at least 100 epochs. The optimal weights selection is based on the validation accuracy and loss. The training was done on a Tesla V100 32 GB GPU.

4.4 Human Assessment

One of the most practical ways to assess synthetic images' usefulness is to evaluate the results of the human rates. This type of assessment has been previously experimented with in similar fields in addition to retinal images [71] [109]. The goal is to assess the classification accuracy of field experts (pathologists) to ensure the images' suitability further, though images may show promising aspects from computational assessment metrics. Additionally, the practice of assessing the quality of the images generated is also exercised in literature [110]. Although asking the question of quality preference may seem like an option, it suffers from observer bias which may adversely affect the results. In other words, asking a participant to assess if they can identify the SR image vs. the HR image does not reveal meaningful information. This issue is due to the differences in the quality of the images, though the SR images may still be a viable alternative. Therefore, rather than a straightforward question of identifying the HR image, the questionnaire is designed to assess the practicality of images based on the comparison.

Breast cancer-focused human assessment studies have been conducted to evaluate the applicability of SR in histopathology. For the pathologist inspection and categorization of synthetic pictures, I developed two Google Forms questionnaires. A comprehensive evaluation by board-certified pathologists was done to determine the quality of these images. Three pathologists specializing in breast cancer have filled out the forms for this evaluation. The standard approach for breast cancer identification entails accurate WSIs assessment; however, in this work, patches were offered to make the task implementable. As a result, it was not anticipated that the accuracy of the evaluations would be comparable to that of WSI assessments, given the vastly different amount of information accessible to specialists. Even though SR pictures are obtained from a lower-quality source, it was predicted that

the degree of assessment accuracy would be comparable between SR and HR images. The ESRGAN network with 4X magnification generates the SR images in this part.

4.4.1 Diagnostic dataset

The Breast Cancer Histopathological Image Classification (BreakHis) dataset was utilized to evaluate the diagnostic accuracy of the human evaluation [111]. The dataset consists of 9,109 microscopic pictures of breast tumor tissue obtained from 82 individuals at four magnification levels. The primary categories of the BreakHis dataset are benign and malignant tumors. There would be a total of 1,820 pictures at 400X magnification, of which 588 are malignant, and 1,232 are benign. The images are 700x460 pixels in size, 3-channel RGB with 8-bit depth per channel, and PNG format.

4.4.2 Questionnaires

First, a diagnostic test was conducted where pathologists were asked to classify images into eight categories. Figure 4.7 shows screenshots of two pages of this test. The categories (types) included were Breast tumors, benign and malignant, where each had four subtypes. The questions included four distinct histological categories of benign breast tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA); and four malignant breast tumors: carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). Each participant evaluated 64 images, of which half were SR and the other half were HR. There was an equal share of classes among the images.

Second, we assessed the image quality preference of field specialists by presenting them with 50 pairs of images, each consisting of the original image and the SR image. Figure 4.8 shows screenshots of two pages of this test. In order to reduce the possibility of bias, the poll developers did not review the images. The arrangement of the images was undisclosed to the experts. To prevent participant bias toward any of the picture categories, the images were arranged in random order (e.g., HR and SR). The outcomes are shown in Table 4.4.

The diagnostic findings reveal that there is no significant difference between the usage of SR and HR images. This is owing to the greater accuracy or kappa score of the evaluation diagnosis in 6 instances with HR pictures vs. 6 cases involving SR images. As shown in

Table 4.4, assessing SR pictures, participant three received higher diagnostic score values. Participant two obtained higher diagnostic scores for HR pictures. However, participant one had slightly better findings (e.g., 3.1% higher malignancy detection accuracy) for type and subtype identification in SR and HR visuals, respectively.

Although this may be a question of personal opinion, the image quality preference findings show that specialists generally prefer SR images. According to the data, just one pathologist (e.g., two) favored the HR images, and even in this instance, the expert prioritized or was unable to identify substantial differences in 46% of the images. One participant regarded 64% of the imagegraphs to be reasonably comparable, although, in 34% of the instances, they preferred the SR images. Ultimately, the second pathologist deemed 68% of SR images to be preferable, while the rest of the images were without discernible distinctions.

4.5 Analysis of results

The results are assessed from multiple perspectives. First, we provide the quantitative comparison of trained network results for image generation based on PSNR and SSIM in Table 4.3. This table compares different sites and subtypes against each other in this table. The comparison shows the details and complexity of the images in each category. The minimum and maximum reported values of the 4X bicubic upsampling of the low-resolution images are shown in red and green, respectively. It is evident that Brain/GBM has the lowest SSIM, which could be interpreted as the highest structural loss. In contrast, the Endocrine/THCA offers the highest SSIM value. On the other hand, the lowest and highest PSNR values were observed in Testis/TGCT and Liver CHOL. These values show the extent of details lost in the interpolation.

Another comparison is also provided in Table 4.3. In this table, the highest values with respect to the subtype are reported in blue. Two networks, DBPN and RCAN, mostly achieved the best results. The DBPN network is superior to others in 15 subtypes, while RCAN is also superior in 15 subtypes. If one network outperforms the other for one measure (SSIM or PSNR), both are mentioned in blue.

The qualitative comparison of the generated images shown in Figures 4.4, 4.5, and 4.4. A randomly selected patch per site is shown in which another randomly selected window has been taken into focus. The window has been compared from eight approaches, including the high-resolution, low-resolution, and six networks generation. The low-resolution

image shows the bicubic interpolation of the 4X downsampled image. Network inputs were the low-resolution image, while the ideal outcome was a high-resolution image generation. The PSNR and SSIM of the images are also written below them, in which the reference for measurement is the high-resolution image. Taking a look at the images, the ESRGAN-generated images look most like the high-resolution image. It seems that adversarial training helps more acceptable reconstruction based on previously visited patterns. Finally, to take a more comprehensive look at the results generated by ESRGAN as the best perceptual generation and RCAN as the best PSNR-based generation, the distribution of the results is presented in Figures 4.3, and 4.2, respectively. The distribution of the results is very similar, while the values are better in RCAN.

As shown in the figures, the network is capable of mapping the LR images to a reasonable HR image which seems to include many details. This has been achieved due to the usage of generative adversarial networks.

4.6 Summary

Ultimately, the architecture for compound magnification includes three essential phases. This chapter discusses the path of regenerating the synthetic WSI, which contains several HR patches and a majority of SR patches. Multiple SR-network architectures have been quantitatively and qualitatively evaluated and reported. The images created by deep SR were evaluated based on professional human evaluation. In the last test, the quality of the pictures from the standpoint of pathologists specializing in the same field of WSIs was evaluated, revealing a remarkable correlation with HR imagegraphs. In order to guarantee that the use of SR pictures does not impact a diagnostic procedure, another human evaluation test comparing the diagnostic accuracy of SR and HR images has been done. All of the findings indicate that SR pictures obtained at four times magnification may effectively replace HR images.

Table 4.2: The TCGA codes (in alphabetical order) of all 32 primary diagnoses and corresponding number of evidently diagnosed patients in the dataset

Code	Primary Diagnosis	# Patients
ACC	Adrenocortical Carcinoma	86
BLCA	Bladder Urothelial Carcinoma	410
BRCA	Breast Invasive Carcinoma	1097
CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenoc.	304
CHOL	Cholangiocarcinoma	51
COAD	Colon Adenocarcinoma	459
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	48
ESCA	Esophageal Carcinoma	185
GBM	Glioblastoma Multiforme	604
HNSC	Head and Neck Squamous Cell Carcinoma	473
KICH	Kidney Chromophobe	112
KIRC	Kidney Renal Clear Cell Carcinoma	537
KIRP	Kidney Renal Papillary Cell Carcinoma	290
LGG	Brain Lower Grade Glioma	513
LIHC	Liver Hepatocellular Carcinoma	376
LUAD	Lung Adenocarcinoma	522
LUSC	Lung Squamous Cell Carcinoma	504
MESO	Mesothelioma	86
OV	Ovarian Serous Cystadenocarcinoma	590
PAAD	Pancreatic Adenocarcinoma	185
PCPG	Pheochromocytoma and Paraganglioma	179
PRAD	Prostate Adenocarcinoma	499
READ	Rectum Adenocarcinoma	170
SARC	Sarcoma	261
SKCM	Skin Cutaneous Melanoma	469
STAD	Stomach Adenocarcinoma	442
TGCT	Testicular Germ Cell Tumors	150
THCA	Thyroid Carcinoma	507
THYM	Thymoma	124
UCEC	Uterine Corpus Endometrial Carcinoma	558
UCS	Uterine Carcinosarcoma	57
UVM	Uveal Melanoma	80

Table 4.3: Accuracy of image reconstruction using six networks trained on TCGA dataset; the PSNR/SSIM numbers are shown where the bold blue numbers are best reconstructions across networks, bold red and green are the worst and best bicubic reconstruction among data categories, respectively.

Data categorization		Accuracy (PSNR/SSIM)						
Site	Subtype	Bicubic	DBPN	RDN	RCAN	SRFBN	EDSR	ESRGAN
Brain	GBM	(23.46/0.57)	(24.64/0.66)	(24.44/0.66)	(24.44/0.66)	(23.68/0.62)	(23.99/0.63)	(22.43/0.57)
	LGG	(23.50/0.58)	(24.68/0.67)	(24.50/0.67)	(24.50/0.67)	(23.69/0.63)	(24.02/0.64)	(22.35/0.58)
Endocrine	ACC	(22.30/0.59)	(23.52/0.68)	(18.68/0.43)	(23.55/0.69)	(22.64/0.65)	(22.81/0.64)	(21.85/0.63)
	PCPG	(22.17/0.59)	(23.42/0.68)	(18.39/0.41)	(23.52/0.70)	(22.47/0.65)	(22.65/0.64)	(21.73/0.63)
	THCA	(24.30/0.67)	(25.78/0.75)	(19.02/0.39)	(25.96/0.76)	(25.12/0.73)	(24.75/0.70)	(24.21/0.72)
Gastro.	COAD	(23.56/0.62)	(24.92/0.71)	(18.80/0.40)	(24.90/0.72)	(24.14/0.69)	(24.00/0.67)	(23.26/0.67)
	ESCA	(23.52/0.65)	(25.17/0.74)	(18.83/0.42)	(25.17/0.75)	(24.14/0.71)	(24.13/0.70)	(23.46/0.70)
	READ	(23.38/0.59)	(24.61/0.68)	(19.13/0.38)	(24.66/0.69)	(23.97/0.66)	(23.82/0.64)	(23.00/0.63)
	STAD	(23.07/0.64)	(24.56/0.73)	(18.22/0.41)	(24.56/0.74)	(23.68/0.70)	(23.61/0.69)	(22.98/0.68)
Gynaeco.	CESC	(22.75/0.62)	(24.26/0.71)	(18.13/0.40)	(24.31/0.72)	(23.45/0.69)	(23.22/0.66)	(22.56/0.66)
	OV	(23.25/0.59)	(24.45/0.67)	(19.22/0.40)	(24.53/0.69)	(23.90/0.66)	(23.75/0.64)	(23.11/0.64)
	UCS	(23.02/0.62)	(24.48/0.71)	(18.43/0.41)	(24.50/0.72)	(23.68/0.69)	(23.44/0.66)	(22.63/0.67)
Liver, panc.	CHOL	(24.70/0.66)	(26.30/0.74)	(19.97/0.41)	(26.19/0.74)	(25.45/0.71)	(25.30/0.70)	(24.52/0.69)
	LIHC	(22.86/0.60)	(24.26/0.69)	(18.22/0.38)	(24.25/0.70)	(23.45/0.66)	(23.29/0.64)	(22.54/0.63)
	PAAD	(24.03/0.61)	(25.47/0.70)	(19.36/0.39)	(25.34/0.70)	(23.45/0.66)	(24.60/0.66)	(23.54/0.65)
Melono.	SKCM	(22.67/0.66)	(24.56/0.76)	(17.78/0.42)	(24.68/0.77)	(23.67/0.73)	(23.21/0.70)	(22.83/0.72)
	UVM	(24.32/0.58)	(25.39/0.66)	(19.97/0.38)	(25.36/0.67)	(24.68/0.64)	(24.75/0.63)	(23.77/0.62)
Prostate, testis	PRAD	(23.90/0.59)	(25.15/0.68)	(19.42/0.38)	(25.08/0.69)	(24.41/0.66)	(24.34/0.64)	(23.42/0.63)
	TGCT	(21.46/0.61)	(23.02/0.72)	(17.64/0.44)	(23.04/0.73)	(22.00/0.68)	(22.00/0.66)	(21.35/0.67)
Pulmonary	LUAD	(23.25/0.63)	(24.78/0.72)	(18.37/0.39)	(23.04/0.73)	(23.93/0.69)	(23.77/0.67)	(22.99/0.67)
	LUSC	(23.12/0.62)	(24.57/0.71)	(18.42/0.40)	(24.48/0.71)	(23.67/0.68)	(23.62/0.67)	(22.81/0.66)
	MESO	(22.49/0.61)	(24.01/0.71)	(18.59/0.43)	(23.97/0.72)	(23.05/0.68)	(23.09/0.67)	(22.31/0.66)
Urinary tact	BLCA	(22.58/0.60)	(24.02/0.70)	(18.42/0.40)	(23.98/0.71)	(23.11/0.67)	(23.10/0.65)	(22.32/0.65)
	KIRC	(24.45/0.65)	(25.96/0.73)	(19.29/0.39)	(25.93/0.74)	(25.09/0.70)	(25.02/0.69)	(24.24/0.69)
	KIRP	(22.56/0.59)	(23.99/0.69)	(18.03/0.37)	(24.09/0.71)	(23.14/0.66)	(23.04/0.64)	(22.43/0.65)
	KICH	(23.79/0.64)	(25.18/0.72)	(18.37/0.38)	(25.22/0.73)	(24.47/0.70)	(24.12/0.67)	(23.54/0.69)

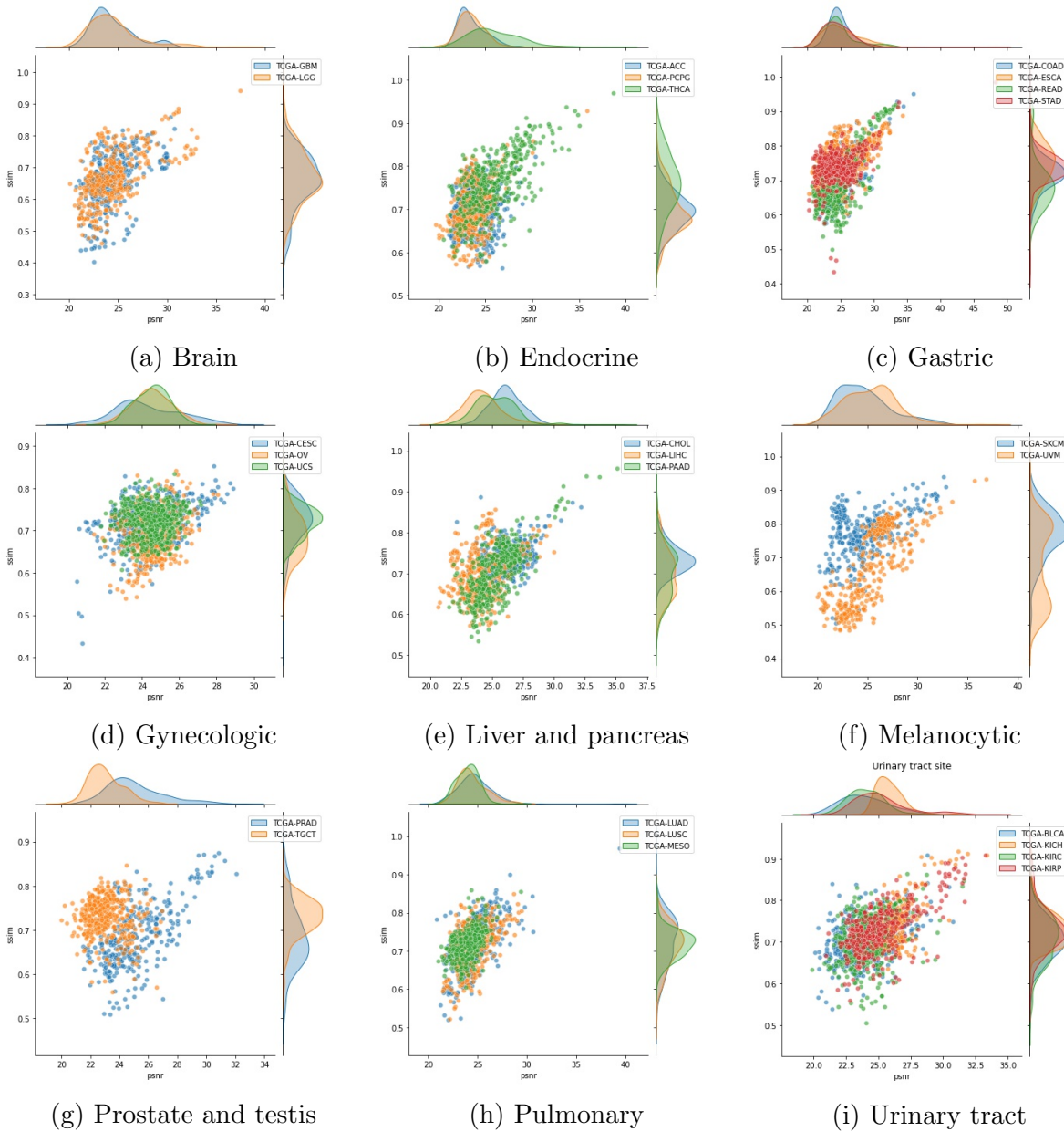


Figure 4.2: The image reconstruction accuracy (PSNR and SSIM) distribution of the RCAN network of the 26 TCGA subtypes categorized in the nine biopsy sites

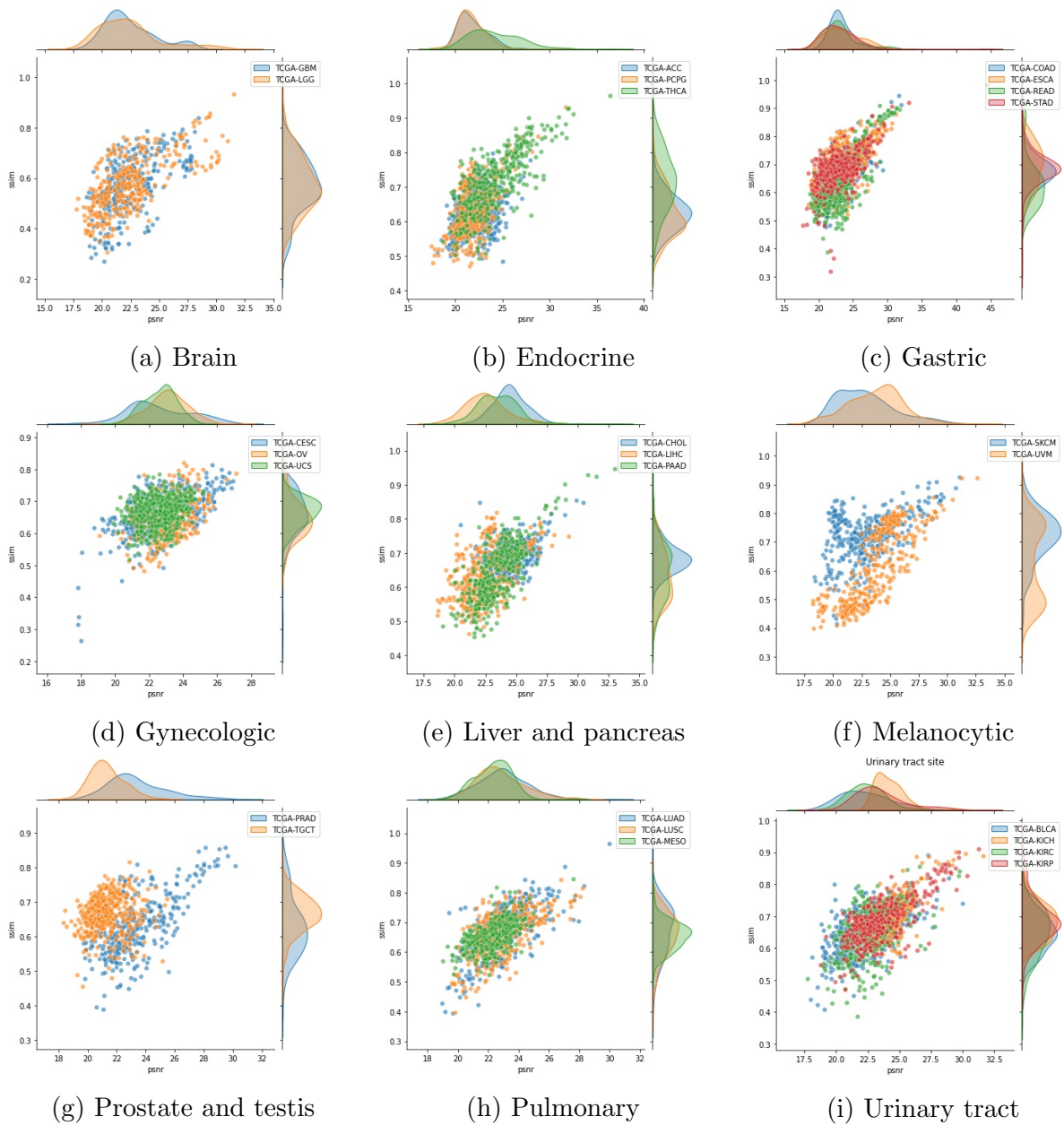


Figure 4.3: The image reconstruction accuracy (PSNR and SSIM) distribution of the ESRGAN network of the 26 TCGA subtypes categorized in the nine biopsy sites

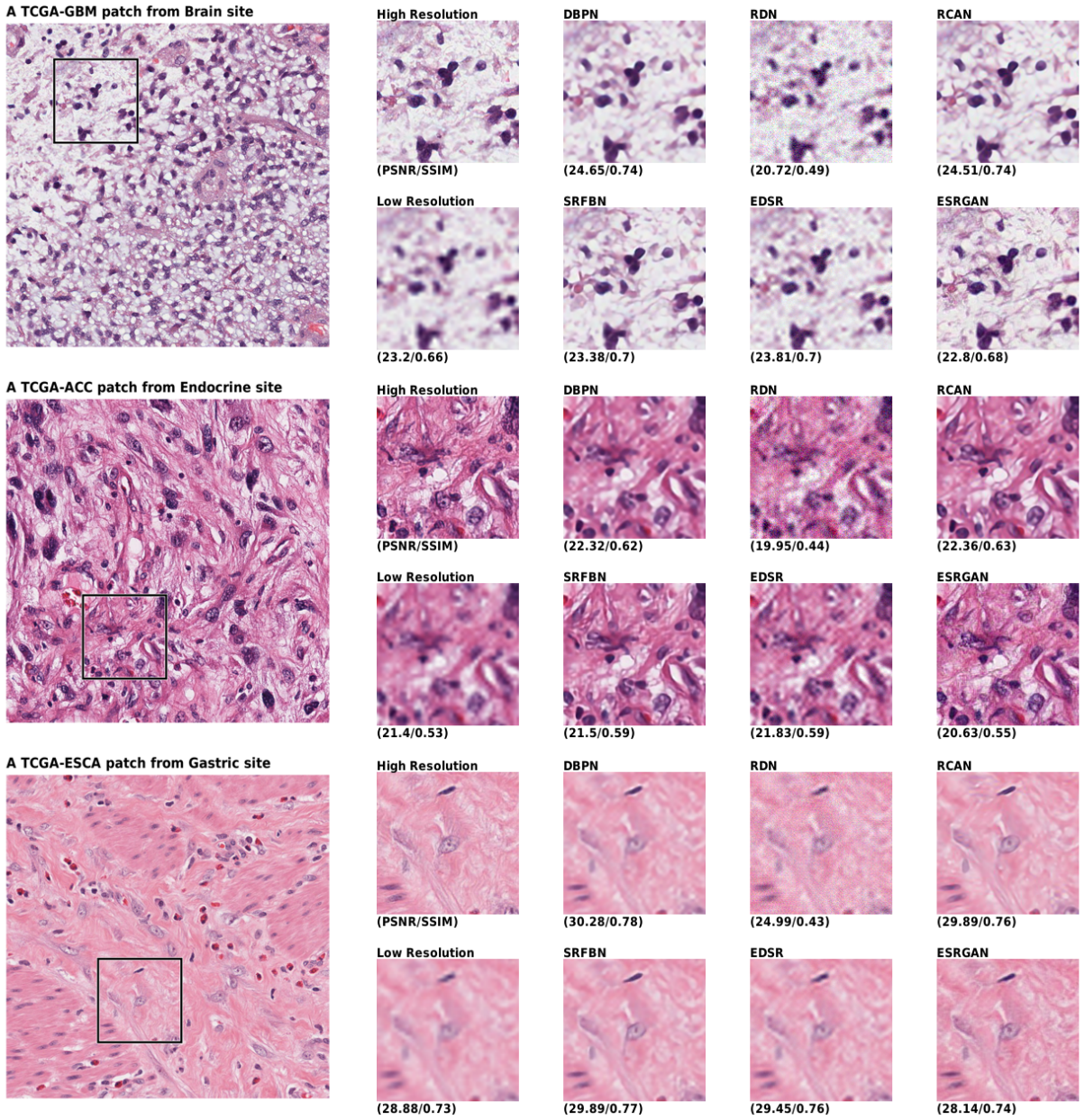


Figure 4.4: Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box

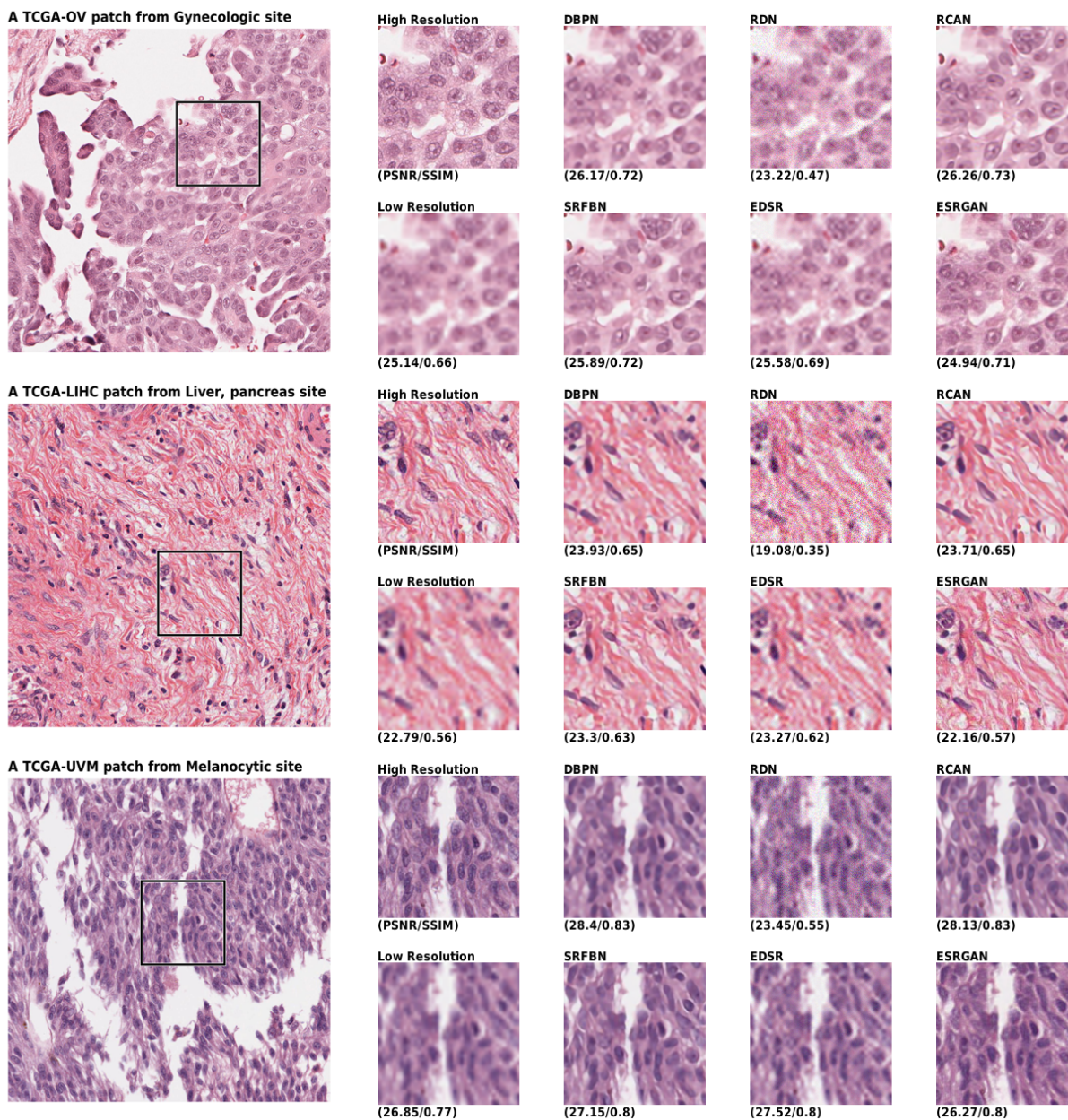


Figure 4.5: Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box

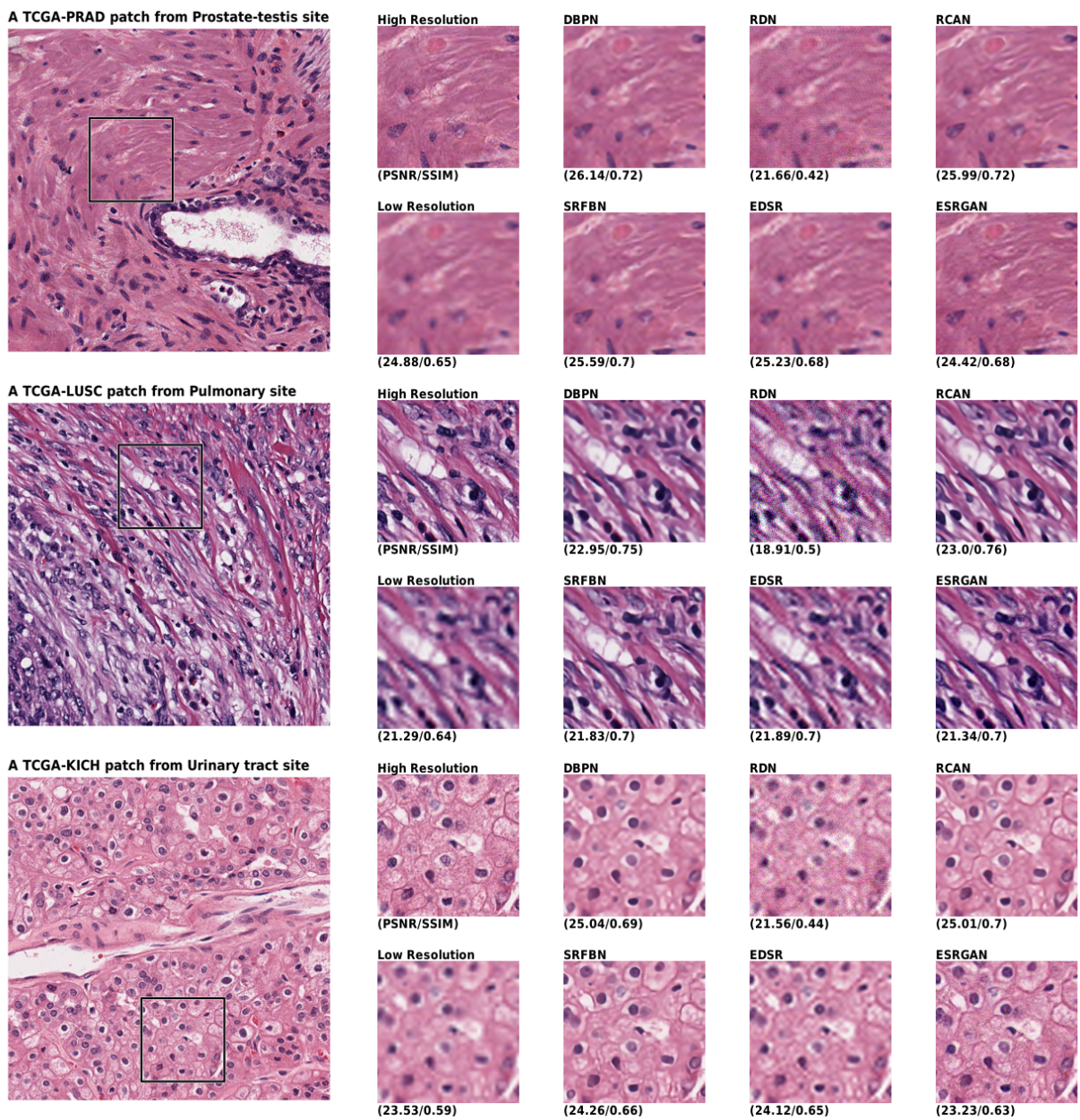


Figure 4.6: Qualitative results comparison of the six networks and the high- and low-resolution images; The low resolution is fed to network to generate images and The big image on left shows the span box

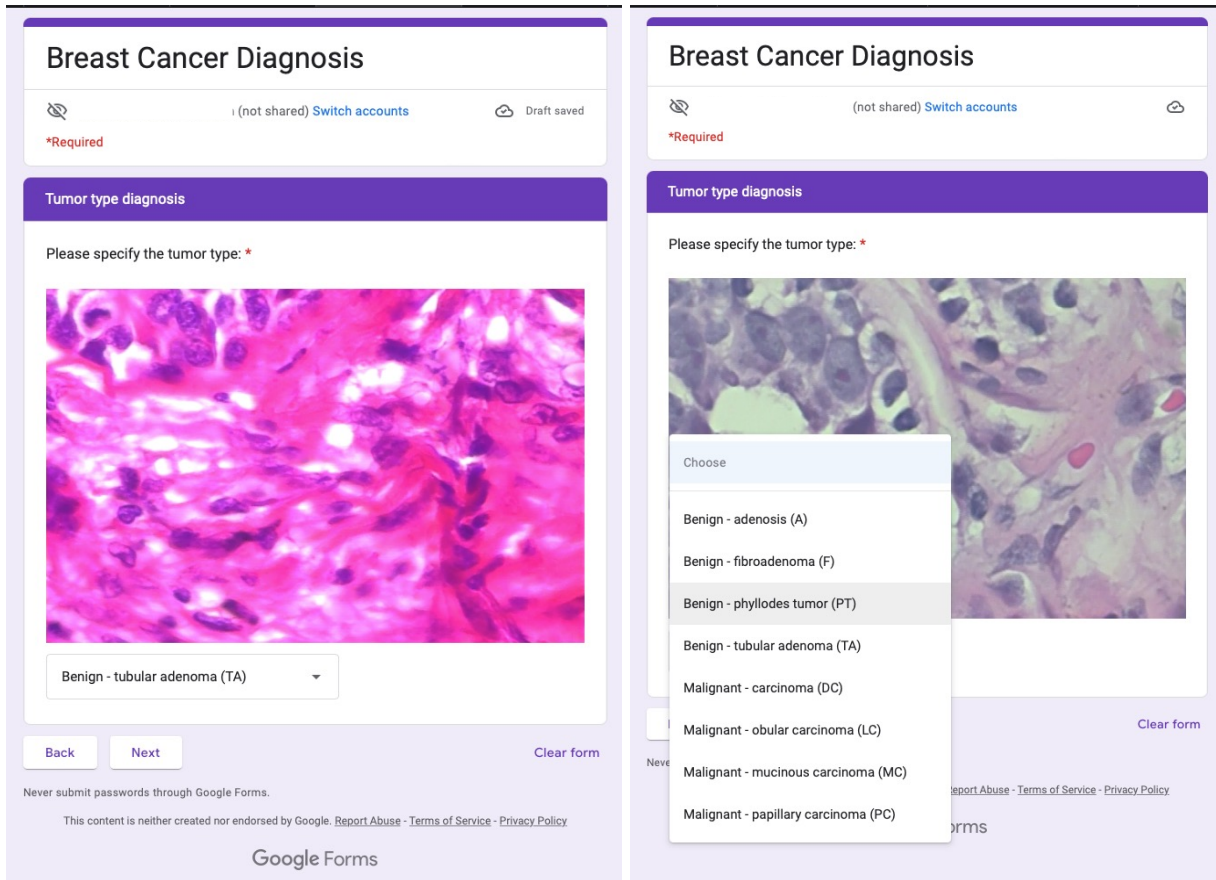


Figure 4.7: Google-form questionnaire for diagnostic tests; left: overview of the form page, right: options available for selection on the form

Qualitative Image Assessment

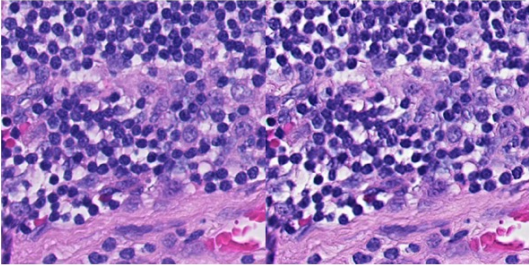
(not shared) [Switch accounts](#)

***Required**

Comparing Histopathology Images

Please choose the patch (left or right) that you believe is of higher quality in comparison to the other for clinical evaluation purposes. *

The order of the images (options) conveys no information.



- Strongly left
- Slightly left
- No discernible distinction
- Slightly right
- Strongly right

Back Next Page 2 of 52 Clear form

Never submit passwords through Google Forms. This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)

Qualitative Image Assessment

(not shared) [Switch accounts](#)

***Required**

Comparing Histopathology Images

Please choose the patch (left or right) that you believe is of higher quality in comparison to the other for clinical evaluation purposes. *

The order of the images (options) conveys no information.



- Strongly left
- Slightly left
- No discernible distinction
- Slightly right
- Strongly right

Back Next Page 9 of 52 Clear form

Never submit passwords through Google Forms. This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)

Figure 4.8: Two pages of the Google-form questionnaire for quality comparisons tests

Diagnostic classification assessment results					
	SR Patch		HR Patch		
	Accuracy	Kappa	Accuracy	Kappa	
<i>Pathologist 1</i>					
Type	78.1	0.56	75.0	0.50	
Subtype	15.6	0.04	21.9	0.11	
<i>Pathologist 2</i>					
Type	78.1	0.56	71.9	0.44	
Subtype	31.2	0.21	18.7	0.07	
<i>Pathologist 3</i>					
Type	71.9	0.44	81.3	0.62	
Subtype	25.0	0.14	34.4	0.25	
<i>Average</i>					
Type	76.0	0.52	76.0	0.52	
Subtype	24.0	0.13	25.0	0.14	
Quality preference results distribution(%)					
	Strong SR	Slight SR	No Distinc.	Slight HR	Strong HR
<i>Pathologist 1</i>	4.0	30.0	64.0	6.0	0.0
<i>Pathologist 2</i>	6.0	26.0	14.0	48.0	6.0
<i>Pathologist 3</i>	4.0	64.0	32.0	0.0	0.0
<i>Average</i>					
	4.7	40.0	36.7	18.0	2.0

Table 4.4: Human assessment of SR vs. HR pathology patches; Type defines the Benign or Malignancy; Subtype is the eight categories of BreakHis dataset used for diagnosis purpose

Chapter 5

Unsupervised WSI Patch Selection

5.1 Objective

Deep learning techniques have been used to evaluate gigapixel histopathology images. One issue is that images can be millions of pixels in size, but the region that acquires the most attention is frequently a rather small portion of the image. The region of importance is the section that is primarily concerned with diagnosing the patient. For instance, detecting a cancerous Whole Slide Image (WSI) may be based primarily on identifying a relatively tiny metastasis area. Frequently, annotations for these regions are unavailable. The purpose of this study is to use an unsupervised deep learning approach to identify these regions. To learn the repetitive pattern, a generative model was trained on a set of healthy WSIs. Following that, similar patterns are discovered by comparing the trained network-generated image to the original image. Figure 5.1 depicts the flowchart of the suggested method in general. Assuming we are dealing with a database of WSIs, each WSI undergoes Otsu thresholding to create the tissue region area. It would be carried out at a low magnification to facilitate this procedure since the specifics have no immense effect. Based on the tissue area, the WSI is then divided into tiles termed patches. Each patch is subsequently subjected to a calculating distribution procedure. The patch is first downsampled and then regenerated to the same size by the SR network. The SR and HR patches may then undergo post-processing. Lastly, the similarity or distance between these images is determined. After computing the distances between each patch, a distribution of the distance between each tile's SR and HR pictures is created. Based on the resulting

distribution, a number (determined by the framework use case) of the most distant (or least similar) patches would be chosen.

This research focuses on a patch-selection system that stores a set of high-resolution patches for future usage. Although the area corresponding to primary diagnosis increases significance in this issue, additional patches may also be beneficial. The restoration procedure may benefit from the high-quality patches that are kept in the framework. Therefore, another class of patches worth keeping are those that are technically difficult to regenerate.

Lastly, while developing a system for processing all the spots on a slide, the processing time may cause challenges. This is crucial because, in order to process a WSI with 100,000 x 100,000 pixels and 256 x 256-pixel patches, the approach may need to process up to 160,000 patches per WSI. Therefore, recommending a procedure that can be completed in a fair amount of time gets relevant.

5.2 Introduction

Digital pathology is a relatively new and rapidly growing area of medical imaging. In digital pathology, whole-slide scanners are used to capture high-resolution images of glass slides containing tissue specimens. The rising digitization of WSIs enables machine learning-based advancements in detecting, visualising, and segmenting lesions, which can ultimately reduce medical experts' workloads [90] [112] [113]. Nevertheless, supervised machine learning requires precisely labeled datasets, which is commonly difficult or impossible to gather, or at the very least time-consuming and thus expensive. As a result, methods that require only partially labeled data (semi-supervised methods) or none at all (unsupervised methods) are now used more frequently [114] [115].

Our research will focus on the detection of micro- and macrometastases in digitized lymph node images. This is a highly relevant subject since these are related to breast cancer diagnosis; metastases to lymph nodes occur in the majority of cancer types (e.g., breast, colon, prostate). Small glands called lymph glands or nodes filter lymph, the fluid that circulates throughout the lymphatic system. Breast cancer spreads most frequently via the lymph nodes in the underarm. The involvement of lymph nodes in metastatic breast cancer is a significant prognostic factor. The prognosis is worse when cancer has spread to the lymph nodes. Meanwhile, the diagnostic procedure used by pathologists is laborious, time-consuming, and prone to misinterpretation [116].

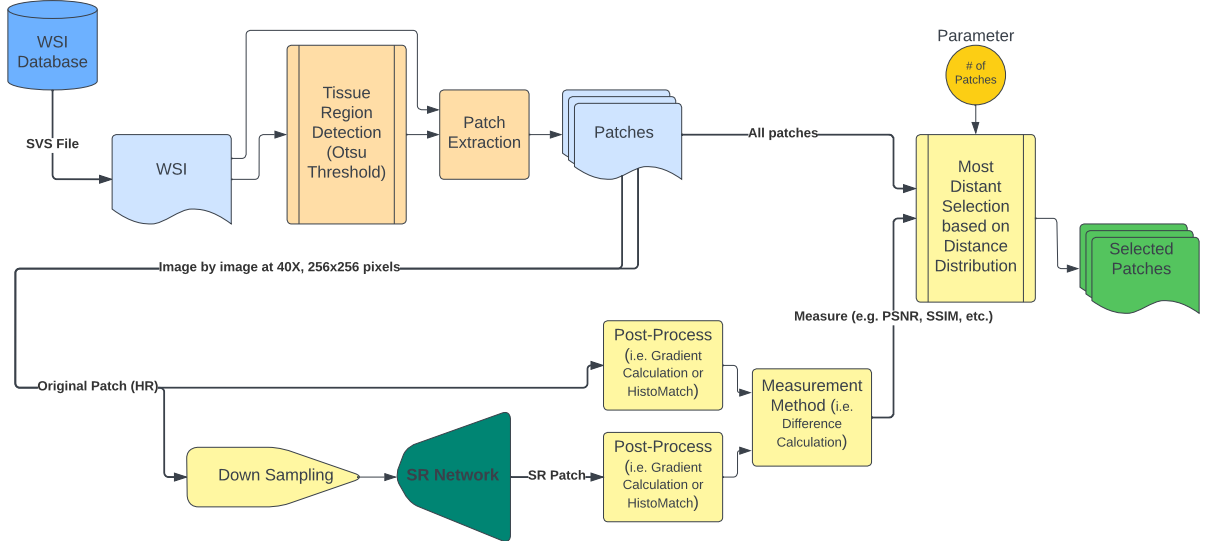


Figure 5.1: General overview of the patch selection method

Automated lymph node metastasis detection has the potential to significantly assist pathologists and reduce their workload. The field has made advances in recent years toward ambitious ideals with significant potential diagnostic impact: Automated assessment of WSIs for the purpose of detecting and grading cancer, assessing prognosis, and identifying metastases [117].

For numerous decades, anomaly detection has been a thriving but active study subject in a variety of research groups. There are still certain complicated and unique problems that demand advanced approaches. Deep learning-enabled anomaly detection, or deep anomaly detection, has emerged as a major trend in recent years. The efficacy of such approaches is particularly visible in areas where the anomalies are well-characterized (and maybe limited to a small set), and there is an abundance of labeled data [118]. In addition to detecting visual abnormalities, we often need to identify the image’s actual anomalous region. Obtaining such comprehensive labels for the purpose of learning supervised models is a time-consuming and, in many circumstances, impractical procedure. While there is a wealth of data accessible in the biomedical area, it is often significantly more complicated and diverse. Typically, such data may be annotated only by qualified biomedical professionals, precluding large-scale crowdsourcing operations [119].

In this study, we present a novel high-resolution image-to-image translation-based

method for unsupervised detection of cancerous regions in histology images that outperforms existing GAN-based unsupervised approaches in that domain for the purpose of patch-selection.

5.3 Background

5.3.1 Unsupervised Detection of Visual Anomalies

Point anomaly identification is the task of finding a single outlier in a large dataset. Most anomaly detection methods can identify point abnormalities. Anomaly detection comes in three varieties: supervised, semi-supervised, and unsupervised. Detecting anomalies with supervision is like classifying a skewed dataset. Semi-supervised attempts to train a model on a single class of data, usually healthy, and then apply it to both healthy and diseased data. Without labels, unsupervised anomaly detection uses both normal and anomalous data and just the dataset’s intrinsic properties (via distances or densities) [3]. Semi-supervised and unsupervised anomaly detection (UAD) are often used interchangeably in anomaly detection.

5.3.2 Image-to-Image Translation

Image-to-Image (I2I) translation aims to discover a mapping between an input image and an output image that spans multiple domains. In the supervised setting, paired correlating images from different domains (e.g., low-resolution version of the same image) are available for learning the mapping, and conditional GANs [16] are used to understand the mapping. Without supervision, only two independent sets of images from the source and target domains are available, with no paired examples (e.g., grayscale-color of different images). Cycle consistency loss, as presented by CycleGANs [120], is a constraint that is frequently used for such inherently ill-posed problems. CycleGANs [120] compel original image reconstruction when mapping the source image to the target domain and back, thereby capturing the target domain’s unique characteristics and determining how to transfer them to the source domain while preserving the source domain’s image characteristics.

Because an I2I job is tasked with learning the mappings between distinct picture domains, how these mappings are represented in order to get the desired outcomes is directly

tied to the generative models. The generative model [121] [122] [123] assumes that data is generated by a particular distribution defined by two parameters (i.e., a Gaussian distribution) or non-parametric variants (each instance contributes to the distribution independently), and it uses specific algorithms to approximate that underlying distribution. This strategy allows the generative model to produce data rather than only differentiate between existing data (classification). For example, deep generative models have shown significant performance gains when making predictions, estimating missing data, compressing datasets, and producing invisible data. In an I2I task, a generative model may simulate the target domain’s distribution by providing convincing ”fake” data, typically translated pictures that look to be taken from the target domain’s distribution.

5.3.3 Unsupervised Learning in Medical Imaging

The diagnosis of tumor metastases is critical for the treatment of breast cancer patients. Numerous CNN-based algorithms achieve good performance in object detection/segmentation. However, for two reasons, metastasis identification in hematoxylin and eosin (H&E) stained whole-slide images (WSI) remains hard. (1) The image’s resolution is very high. (2) a deficiency in labeled training data. Most of the time, full-slide images are kept in a multi-resolution structure with several downsampled tiles. Without compression, it is impossible to store the whole picture in memory. Additionally, categorizing images for pathologists is time-consuming and costly. This writing aims to examine the difficulty of recognizing breast cancer metastases at the patch level in pathological images. To be more exact, we offer an unsupervised patch classifier that operates on slide labels rather than precise annotations.

5.4 Method

To address the challenge of unsupervised unhealthy area detection, as recommended by the literature, it is possible to train an I2I that produces images that distinguish between healthy (patterns/structures observed) and unhealthy (not observed) areas. This signifies that the recommended strategy consists of two essential components. First, there is an I2I that is responsible for altering images, and second, there is a measuring system that recognizes the changes that have become apparent. The following sections discuss the viable options, reasons to choose each component, and the structure of the final method.

5.4.1 I2I Translation

The goal of I2I is to translate pictures from one domain to another while maintaining their content representations. I2I has gained considerable attention and advancement in recent years due to its broad applicability to various computer vision and image processing challenges, including image synthesis, segmentation, style transfer, restoration, and posture estimation.

Although the techniques of many of the previously-mentioned forms might be modified to match the application that is the subject of this thesis, I have selected image SR as a viable solution. In theory, a trained SR network that can upsample low-resolution images to high-resolution can create the differences. Technically, if a network is trained on a collection of images, it is expected that its performance may decline when the domain changes, despite the fact that it may perform well on images from the same domain. If we suppose that an SR network is trained to upsample pictures (patches) of healthy WSIs, then we may argue that if this network is subjected to unhealthy patches, the resulting images may have anomalies. The presence or absence of abnormalities may therefore determine whether a patch (or a tiny area) is healthy or unhealthy (like metastasis in this study).

The usage of image SR signifies that the approach enables a network that is adept at estimating the details of a healthy WSI to attempt to upsample all pictures. Therefore, if it sees cells at a modest magnification, for instance, it believes they are healthy and generates an image with detailed healthy cells. This argument seems logical since this procedure may be viewed as a pathologist’s standard practice. In a clinical evaluation, the pathologist continuously zooms in and out of a slide to discover abnormalities. This may be interpreted as a technique in which a clinician has a general understanding of how a healthy area should look but detects a problem upon seeing unusual structures and forms in high magnifications.

In other words, we assume the existence of healthy patches in one domain and unhealthy patches in another. The network is trained on images from the first domain before being supplied with images from any domain. Finally, the generated images should vary from their respective source images. There are several factors to consider while developing a meaningful solution. These factors include the patch parameters (e.g., size, level), network topology, and SR magnification, which will be explored in further detail below. Evaluating differences of image generation are detailed in section [5.4.2](#).

To implement the I2I, here we use the Structure-Preserving Super Resolution with Gradient Guidance (SPSR) network architecture. The network benefits from two branches:

SR and Gradient. The SR branch helps with a general upsampling problem which uses ESRGAN architecture, while the gradient branch uses RRDB blocks and focuses on structures in reconstruction, which is helpful since the images are going to be compared, and the appearance of similar patterns is the best leading factor. The benefits of gradient guidance further help with the reconstruction of the edges. The network is trained using perceptual, pixel, and adversarial loss for SR image in addition to pixel and adversarial loss for SR gradient and pixel loss of gradient image generated by the gradient branch. The discriminator is based on a VGG-like network and training is done using ADAM optimizer [124].

5.4.2 Measurement System

According to the suggested technique, each tile would result in two pictures. The first is the original picture extracted from the WSI, while the second is the SR, upsampled version of the downsampled tile. In other words, the first picture is the original, and the second is the trained SR network's impression of the original image if it only had access to a lower degree of magnification. According to the logic, the created picture should ideally resemble the original image in which we have a healthy patch. In contrast, the images may display differences if the patch comes from a metastatic location. To determine whether or not there is a difference between these images, we need to utilize some measures to evaluate their similarity. The pertinent measures and the selection rationale are detailed below.

The PSNR value embodies the principle of image comparison in its most basic form. Technically, if we subtract the pixel values of the two images (i.e., SR image error), the resultant numbers should represent the pictures' fundamental resemblance. However, this statistic does not take picture content into account. In order to better compare the substance of images, measures that are more conceptual may be beneficial.

A perception-based model, SSIM, is employed as an additional measure for content evaluation. The human vision system (HVS) operates conceptually to comprehend a picture based mostly on its low-level features. This has been implemented as an additional statistic of FSIM that could be useful here. Regarding the idea of machine vision, there are measures depending on the machine's comprehension of the picture. This can also work by using LPIPS. Additionally, wavelet transformation has been used to develop measures based on human perception. This is referred to as HaarPSI, which is used here. The images of histopathology include information such as cell edges. Therefore, gradient-based similarity may disclose more information and possible distinctions. GMSD has been used

for this objective. An additional measure that incorporates gradients, chromaticity similarity (CS), and deviation pooling (DP) might be advantageous. Thus, MDSI has been implemented. VSI, as a Visual Saliency-Induced Index, is also used. Finally, in the field of histopathology, images, magnification, and level are crucial factors. The Multi-Scale version of SSIM and GMSD are also examined to determine their influence on the suggested technique.

5.5 Materials

In order to apply the concept offered in this study, the materials indicated in this part are used. These resources consist of the datasets, processing units, and computer frameworks addressed in sections 5.5.1, 5.5.2, and 5.5.2 , respectively.

5.5.1 Histopathology Dataset

We utilize the publicly available CAMELYON16 dataset, which was prepared from data gathered at Netherlands medical centers. The WSI files are provided in TIFF format. At the lesion level, annotations are provided as XML files. Each annotation was meticulously prepared under the supervision of pathologists. The data set contains 400 sentinel lymph node WSIs from two distinct datasets at Radboud University Medical Center in Nijmegen, Netherlands, and University Medical Center Utrecht in Utrecht, Netherlands [125].

The original dataset is divided into two parts: the first contains 170 WSIs of lymph nodes (100 normal and 70 with metastases), and the second includes 100 WSIs (including 60 normal slides and 40 slides containing metastases). In this work, I utilize just the first 100 WSIs for training and the remainder of the dataset (60 normal slides and 110 slides with metastases) for evaluation and testing.

To enable visualization of the images, WSIs are often saved in a multi-resolution pyramid format. Within image files, multiple downsampled copies of the original image are stored. Each image in the pyramid is kept as a series of tiles in order to facilitate the retrieval of image subregions. Reading these images using conventional image tools or libraries is a challenge since these tools are typically designed for images that can be uncompressed and stored in RAM. This is accomplished using OpenSlide, a C library that

provides a simple interface for reading and writing WSI files in a variety of different formats (code available at <https://openslide.org>). Moreover, ASAP is used to enable the use of annotations. It is a free and open source platform for presenting, annotating, and automating the analysis of whole-slide histopathology image annotations (source code available at <https://github.com/computationalpathologygroup/ASAP>).

5.5.2 Processing Units

Multiple devices and setups were utilized to conduct experiments. Two Nvidia Tesla V100 GPUs with 32 GB of memory each were employed during the training phase, where the CPU and memory were 32 cores and around 80 GBs, respectively. However, testing and validation benefitted from a more common processing configuration consisting of an Nvidia-2080 Super GPU and a ten-core Intel 9900X CPU paired with 32 GBs of RAM.

5.5.3 Frameworks

OpenSlide is the WSI research framework. To parse the annotations, a method was developed from scratch that utilizes xmlTree to read XML files and PIL image to display them. PyTorch is the deep learning framework used for the training and execution of SR networks in this study. This was picked because of its versatility for evaluating a variety of techniques. The measuring system is based on the PIQA framework, which is based on PyTorch since it offers all the necessary tools for this study, and the Tensor datatype may save processing time. There have been numerous more frameworks to implement the remaining project details, which the reader may request access to.

5.6 Results and Discussions

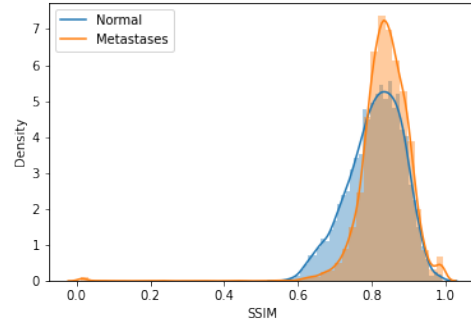
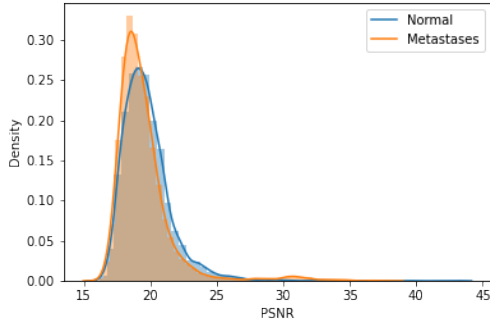
This part details the outcomes of applying the recommended strategy to the mentioned dataset. This thesis argues that enlarging images exposes their normal or abnormal qualities. However, the first trials undertaken did not provide the intended results. A sample distribution result of implementing an ESRGAN network with 4X and 16X upsampling and comparing PSNR and SSIM of images is shown in Figure 5.2. Figures 5.2, subfigures a and b demonstrate that while the network was trained on upsampling the healthy patches, in

practice, the metastasis patches are also upsampled with a comparable degree of precision, resulting in significant overlap of the distributions. This has occurred owing to the fact that, despite knowing that healthy patches and unhealthy patches are not identical, the information provided by four times upsampling may be correlated. Similar difficulties are observed when upsampling 16 times in subfigures c and d; nonetheless, the PSNR distributions displayed a small degree of disparity. In addition, the SSIM comparison demonstrates a moderate degree of disentanglement. Two peaks unrelated to the desired diagnostic emphasis are observed in both the positive and negative distributions. However, the trials have shown that findings may improve as magnification rises. Experiments on a subset of deliberately chosen, strongly opposed pictures have been done to confirm that the proposed solution is indeed a viable one. This is accomplished by evaluating the concept of magnification using a bicubic technique and monitoring the changes using standard statistics such as the mean, median, and standard deviation.

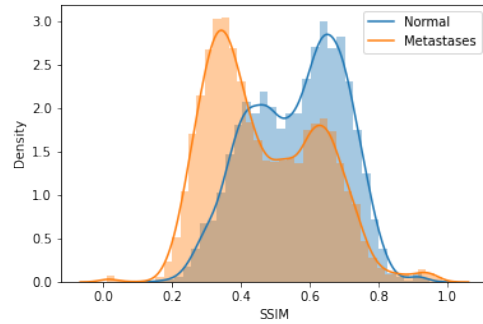
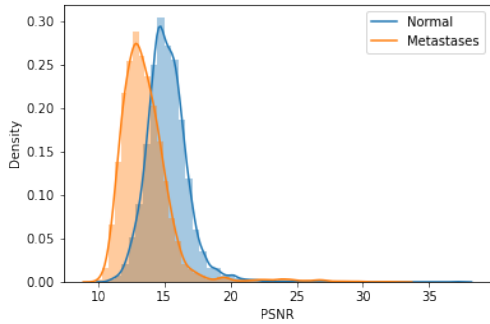
According to the first section of the results, the distributions of pictures with positive or negative labels were more dispersed as the degree of magnification increased. This would indicate that boosting the magnification to its maximum setting may provide better results. However, given that the picture patch size is assumed to be 256 by 256 pixels, the input image to the network should get smaller as the magnification level rises. For example, at 4X, 8X, and 16X, the width and height of the input picture will be 64, 32, and 16 pixels, respectively. In technical terms, this indicates that a rise in magnification corresponds to a network that receives substantially less information at the input, finally preventing it from creating any meaningful information at the output. After careful evaluation of the images, overtraining or incapability of network was also viewed at 16X, rendering it useless. The images in Figure 5.3 depicts instances of 16X magnification. Instead of upsampling patches and generating meaningful information, the network discovered the appearance of tissue patches and used a coloring-like technique on them. This issue could be interpreted as the incapability of the network in upsampling 16 times which is not reasonably expected considering the number of unknown pixels in the process. This has not been doable since the network is supposed to find 256-pixel values instead of each pixel, which is extremely hard.

5.6.1 Magnification Selection

The correlation between greater magnification and improved disentanglement has been examined in order to quantify the effect of increasing magnification more precisely. The patches are extracted at the most significant attainable magnification (i.e., 40X) since



(a) PSNR comparison of 4X magnification (b) SSIM comparison of 4X magnification



(c) PSNR comparison of 16X magnification (d) SSIM comparison of 16X magnification

Figure 5.2: Initial distribution experiments results on 4X, and 16X magnification with PSNR, and SSIM

literature suggests this is the most informative. Since we are using an SR technique, the input picture should be a downsampled replica of the original patch. Multiple levels of downsampling and, subsequently, upsampling may be performed by the method. For instance, the patch may be downsampled/upsampled 4, 8, or 16 times. As a preliminary evaluation, the pictures must depict distinctions between the approach’s most fundamental version, the bicubic method. Here the distance differences and instances of patches in the normal vs. metastasis are shown.

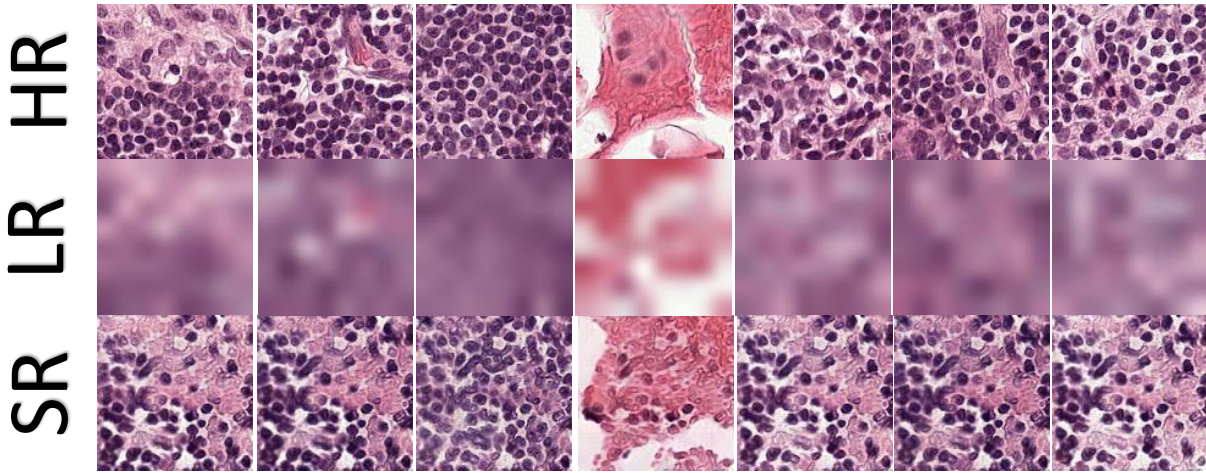


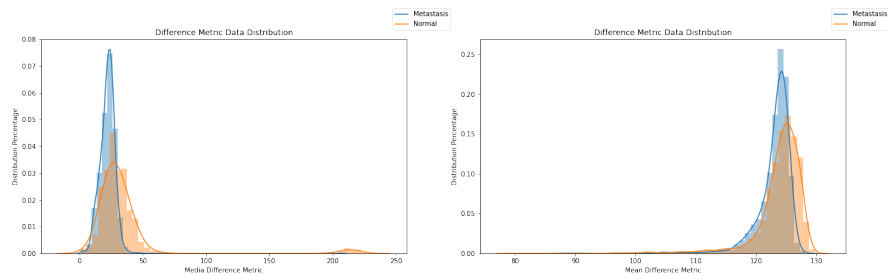
Figure 5.3: HR, LR, and SR Patches with magnification of 16X

5.6.2 Gradient Calculation Application

One of the characteristics that might aid in discovering an abnormal region is the patch's edges. In order to detect edges, gradients of images are used to both guide the super-resolution network in which the Structure-Preserving Super Resolution with Gradient Guidance (SPSR) network has been employed, and additionally, instead of computing the image distances, the distances of images gradients are calculated when comparing images. Figure 5.7 has been provided to evaluate the use of gradient computation. As demonstrated in the images, the cells that seem to be associated with metastases are more eradicated by SR reconstruction, but the healthy patch exhibits less of the same impact. In addition, gradient computation, which computes the differences between two X and Y directions, has been supplied to amplify the SR network's impact. This impact has the potential to enhance the method's capacity to select a more specific and sensitive location.

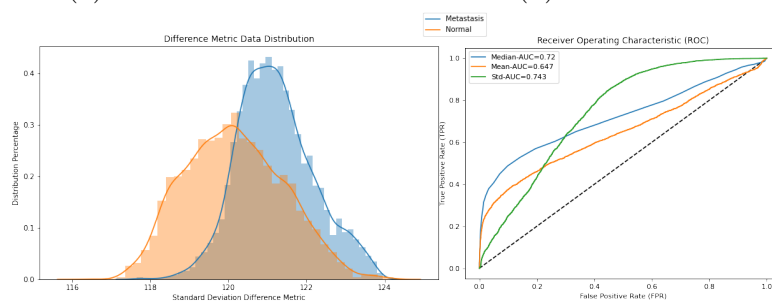
5.6.3 Distance Distributions

On the basis of the data obtained so far, it is clear that a greater magnification might better separate the distribution. In addition, gradients may improve networks' outputs in terms of cell borders and other possible edges. However, the concept of gradient comparison may need more evaluation. In addition, unlike the supervised technique, the images utilized in this method are not initially normalized. Image stain normalization is well recognized



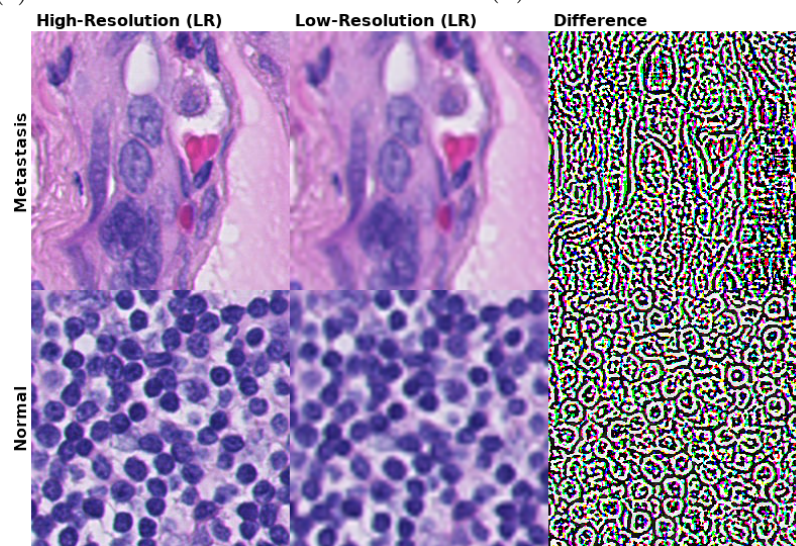
(a) Median of Dist.

(b) Mean of Dist.



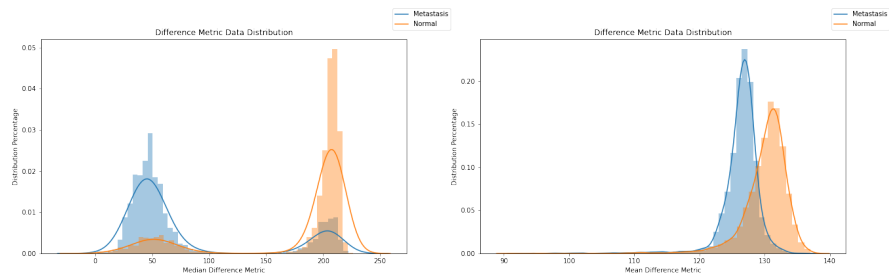
(c) Standard deviation of Dist.

(d) ROC Curve at 4X



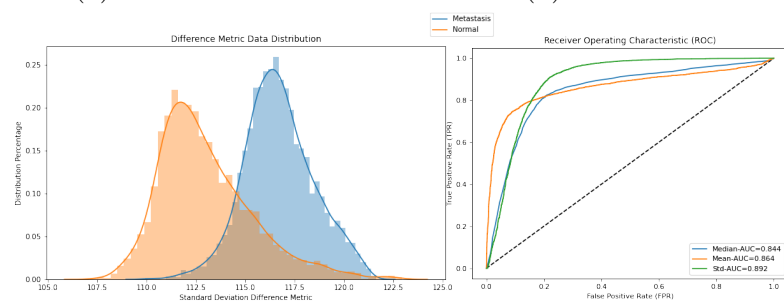
(e) Visualization of differences at 4X

Figure 5.4: Visualization of differences distribution assessment for 4X magnification



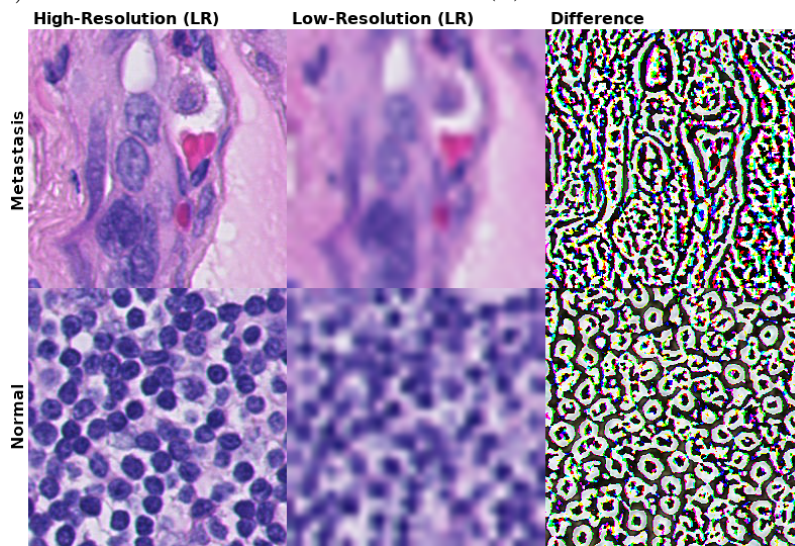
(a) Median of Dist.

(b) Mean of Dist.



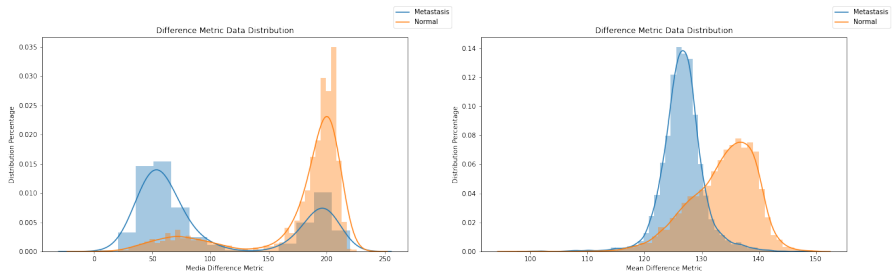
(c) Standard deviation of Dist.

(d) ROC Curve at 8X



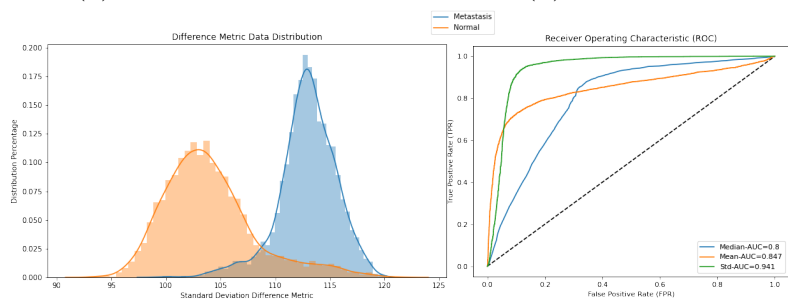
(e) Visualization of differences at 8X

Figure 5.5: Visualization of differences distribution assessment for 8X magnification



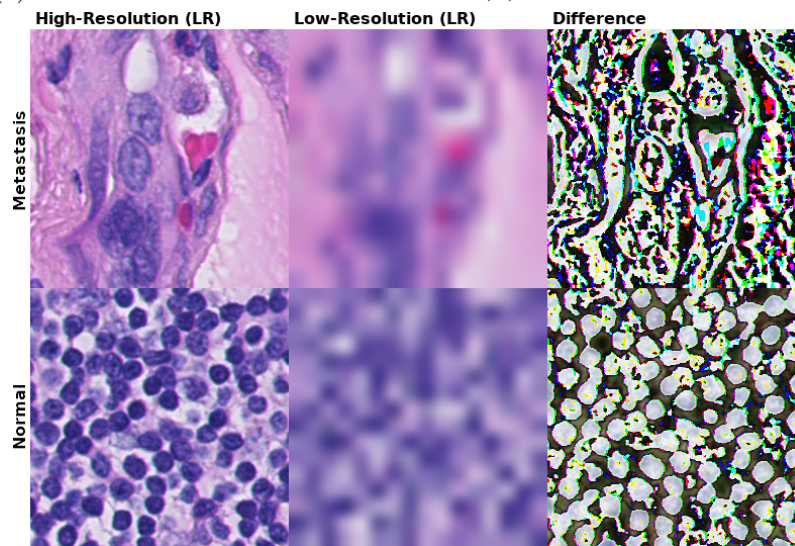
(a) Median of Dist.

(b) Mean of Dist.



(c) Standard deviation of Dist.

(d) ROC Curve at 16X



(e) Visualization of differences at 16X

Figure 5.6: Visualization of differences distribution assessment for 16X magnification

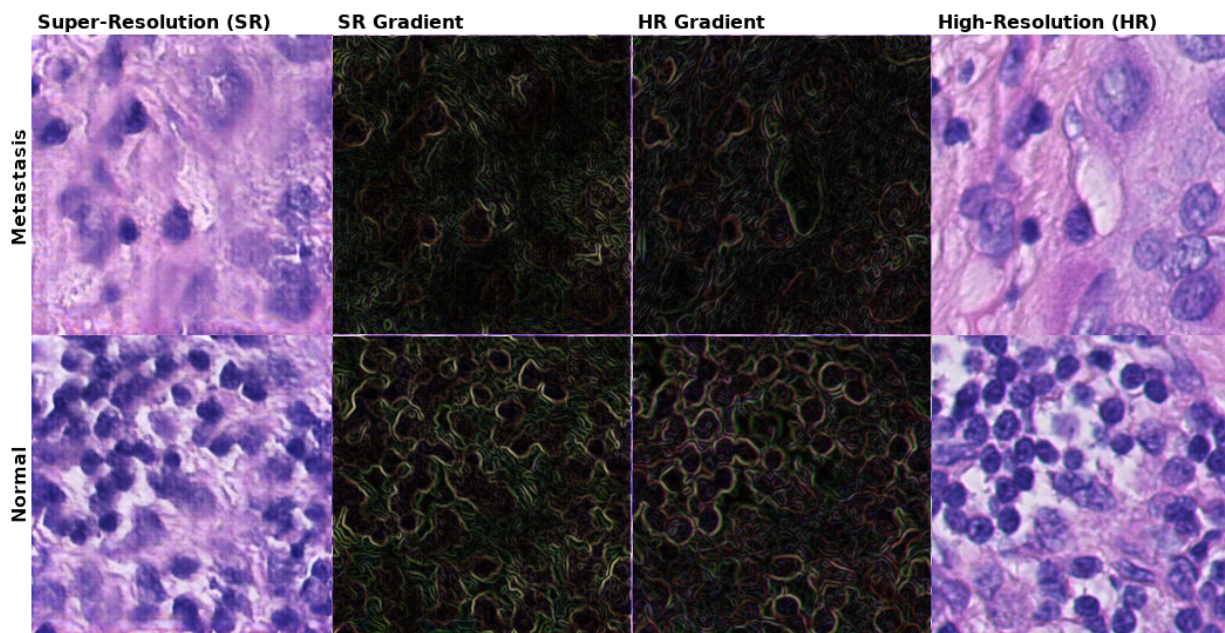


Figure 5.7: Gradient calculation visual inspection for normal vs. metastasis patch

as one of the uses of generative adversarial training models. During training, the SR-network learns a mapping between the pictures it observes. This indicates that if the tested images are not in a highly comparable domain, they may be mapped to the known zone that the network observed in the training phase. This might occur on the hues of the images or, in other words, staining. To further guarantee that this problem does not reduce the procedure’s efficiency, the resultant images can be simply matched using the histogram. This is allowed since the approach provides access to the original patches, and a histogram match would not compromise the method’s generalizability. Based on the scenarios discussed, the system has been evaluated under four distinct settings. The specifications derive from the option’s feasibility for gradient and/or histogram matching.

Having trained the image SR network for eight times magnification, the next step is to test the similarity/distance measurement’s ability to detect any differentiating distributions. At this point, distance distributions must be computed. Each picture is being downsampled, and the SR version is created for this purpose. The images are compared under the four cases mentioned earlier. Comparison is also implemented using the ten similarity/distance measures given. Having obtained all patch comparison data, the distance distribution is next displayed to evaluate the method’s applicability. Their graphs are shown in the Figures 5.8, 5.10, 5.9, and 5.11. Each figure depicts the distance distribution

under one of the four scenarios.

The approach requires that the distribution exhibit two local maxima corresponding to the positive and negative labels. Ideally, the peaks should be separated, and the two distributions of normal versus metastatic should be correctly divided. Nonetheless, this is not often the case in all the experiments. Here, not all of the figures exhibit two locally distinct maximums. This is used as a hint for limiting the number of measures in this approach.

Figure 5.8 demonstrates the ten measures used to calculate the distance or similarity between the original HR patch (image) and the SR-generated version. The majority of the figures have only one distinct maximum. For instance, even though the SSIM distribution seems to be the sum of two distributions, it is unclear where each distribution lies. Although the distributions may exhibit a Gaussian distribution, this theory has not been implemented since there is no supporting literature or research. In addition, a minor peak is noticed in the PSNR distribution, which is shown in subfigure a, which may indicate a correlation with the unhealthy zone. However, this cannot be the case since the little peak appears at greater PSNR levels. Higher PSNR values suggest a greater signal-to-noise ratio, or, in other words, that the associated patches have been upsampled more effectively. This peak cannot thus represent the unhealthy area. Only four of the ten distributions have two distinct maximums. FSIM, MS-SSIM, VSI, and LPIPS are these. Finally, these four have been deployed on a WSI for visual observation of the outcomes in Figure 5.13. The sample WSI and its Otsu threshold defining the tissue region are shown in Figure 5.12.

Figure 5.10 demonstrates the ten measures used to calculate the distance or similarity between the gradient of the original patch (image) and the gradient of the SR-generated version. The same argument as the previous case applies here. Only two of the ten distributions have two distinct maximums. SSIM and MS-SSIM are these. In addition, these two have been deployed on a WSI for visual observation of the outcomes in Figure 5.13.

Figure 5.9 demonstrates the ten measures used to calculate the distance or similarity between the original patch (image) and the histogram-matched SR-generated version. The same argument as the previous case applies here. Only five of the ten distributions have two distinct maximums. FSIM, LPIPS, MDSI, MS-SSIM, and VSI are these. In addition, these two have been deployed on a WSI for visual observation of the outcomes in Figure 5.13.

Figure 5.11 demonstrates the ten measures used to calculate the distance or similarity between the gradient of the original patch (image) and the gradient of the histogram matched SR-generated version. The same argument as the previous case applies here. Only two of the ten distributions have two distinct maximums. MS-SSIM and SSIM are these. In addition, these two have been deployed on a WSI for visual observation of the outcomes in Figure 5.16.

5.6.4 Annotation Results

In the end, adopting the provided technique reveals the area that may need the most attention. In our situations, this zone is preferably the Metastases region. The outcomes of the tests are shown in the Figures 5.17 and 5.18.

The method’s precision must be evaluated to assess the effect and applicability of the recommended idea. The approach generates a distribution that, ideally, reflects two distinct distributions representing the normal and anomalous portions of a WSI. Due to the fact that tests have shown that these two distributions are not fully separate, it is likely that some of the patches will be misclassified depending on the chosen threshold. In order to leave the door open for any threshold selection, or selection of the number of patches, this research evaluates the method’s overall performance. A receiver operating characteristic curve, or ROC curve, is recommended for this purpose because it is a graphical representation of the diagnostic capability of a binary classifier system when its discrimination threshold is altered. The ROC curve is derived by graphing the true positive rate vs. the false positive rate at different threshold levels. The true-positive ratio is known as sensitivity, recall, and detection probability. The false-positive rate, commonly known as the likelihood of false alarm, is equal to $(1 - \text{specificity})$. The Area Under the Receiver Operating Characteristic curve (AUROC) is then determined in order to evaluate the method’s general utility. The bigger the area, the more beneficial the test; AUROC is used to compare the utility of tests. Final results are provided in the Table 5.1. The use of the metric LPIPS was avoided because the distributions indicated that no significant difference was observed. Moreover, because this perceptual metric is based on the features extracted from a deep classifier network, the processing time tends to increase, which is contrary to the purpose of this study. As demonstrated by the findings, the optimal choice is 8X upsampling and downsampling using the MS-SSIM as the distance measurement. As we get to 16X, the findings for MS-SSIM indicate a significant decline in value. This is because, as previously shown, the network acquired a coloring capacity rather than a significant upsampling. In order to guarantee that the inability or lack of a decent reconstruction was

not a result of the network’s insufficient capacity, 16X magnification has been implemented by two 4X successive networks and is labeled as 16X ext. Nonetheless, findings do not demonstrate a successful area identification, which may be construed as an inability to magnify 16X. SSIM and MS-SSIM had the highest performance at 4X and 8X magnification, as was previously established based on distribution observations. In addition, while comparing the MS-SSIM data for 4X and 8X, the fact that a greater magnification level might disclose superior findings was evident.

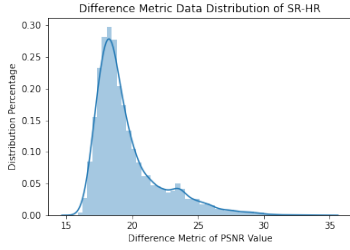
Table 5.1: Patch-Selection, AUROC of Metastasis region detection results

M	PSNR	SSIM	MS-SSIM	FSIM	GMSD	MS-GMSD	HaarPSI	MDSI	VSI
4X	0.522	0.751	0.787	0.666	0.428	0.436	0.683	0.484	0.721
8X	0.604	0.766	0.808	0.622	0.649	0.636	0.577	0.601	0.702
16X	0.684	0.268	0.358	0.373	0.587	0.601	0.583	0.662	0.669
16X ext.	0.673	0.379	0.384	0.276	0.728	0.729	0.287	0.677	0.349

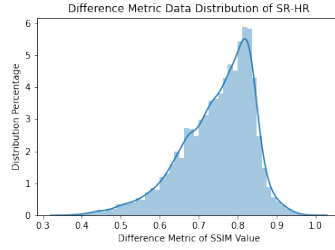
5.7 Summary and Conclusion

This section focused on recommending a process for selecting patches. As explained, this approach has restrictions that make its design difficult. The conditions stem from the lack of available data annotations. Given that not all cancer types and subtypes are annotated, access to annotation should not be regarded as a baseline. Nevertheless, it is well established that if a slide is malignant, the area that shows this acquires heightened significance. In addition, while annotations are rare and difficult to obtain, generic slide labels are widely accessible. This culminated in the strategy recommended in this chapter. Figure 5.19 illustrates the summary of the proposed method depicted applying on a WSI.

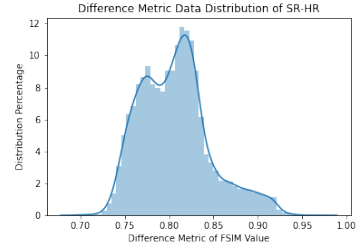
The technique shown here uses the publicly accessible labels of images in the datasets to offer a meaningful annotation for the patch selection that may eventually approach the accuracy of supervised approaches.



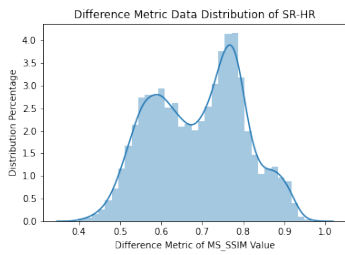
(a) PSNR distribution



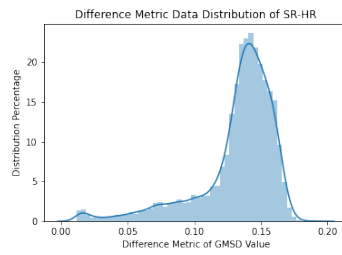
(b) SSIM distribution



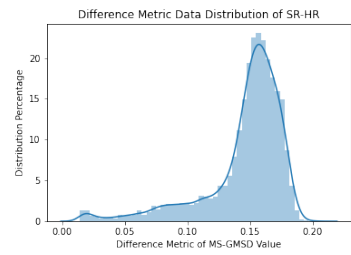
(c) FSIM distribution



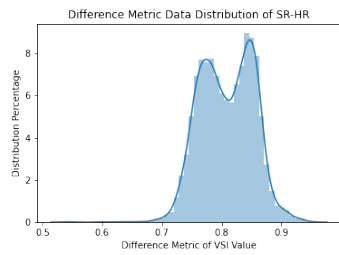
(d) MS-SSIM distribution



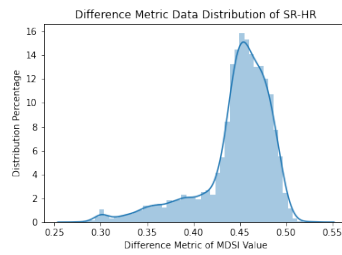
(e) GMSD distribution



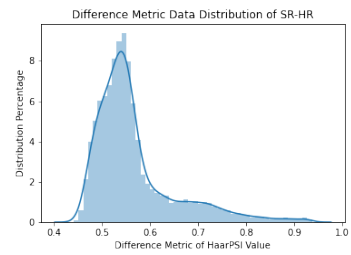
(f) MS-GMSD distribution



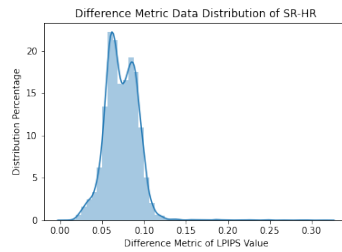
(g) VSI distribution



(h) MDSI distribution

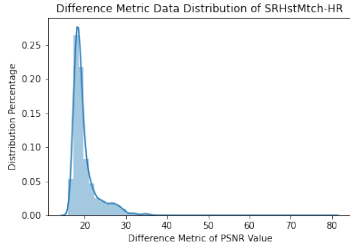


(i) HaarPSI distribution

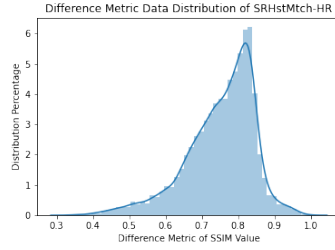


(j) LPIPS distribution

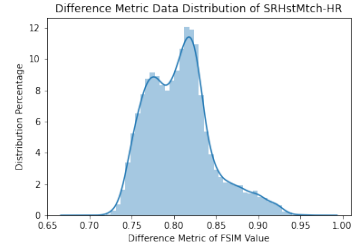
Figure 5.8: Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the HR patch to the SR patch



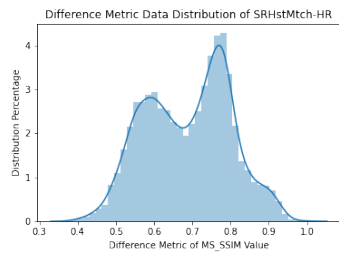
(a) PSNR distribution



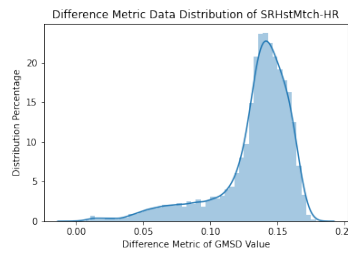
(b) SSIM distribution



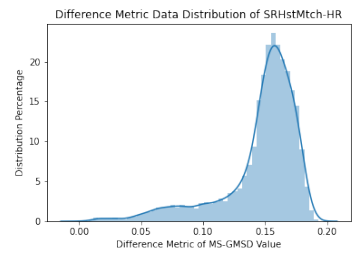
(c) FSIM distribution



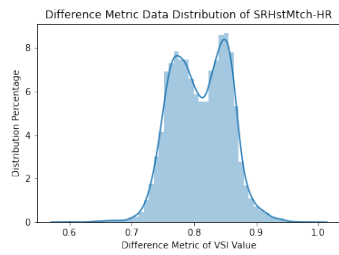
(d) MS-SSIM distribution



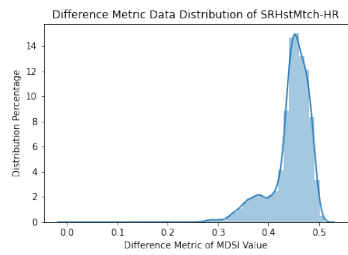
(e) GMSD distribution



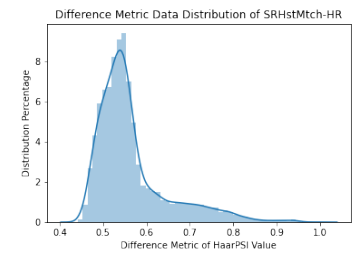
(f) MS-GMSD distribution



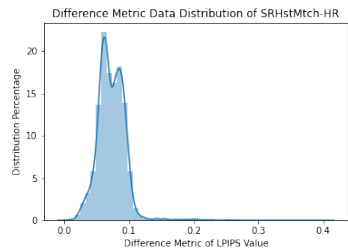
(g) VSI distribution



(h) MDSI distribution

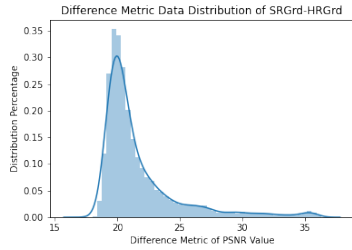


(i) HaarPSI distribution

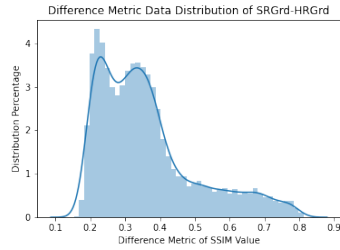


(j) LPIPS distribution

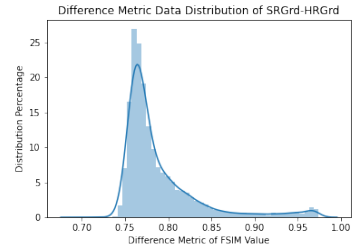
Figure 5.9: Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the HR patch to the SR patch histogram matched to the HR



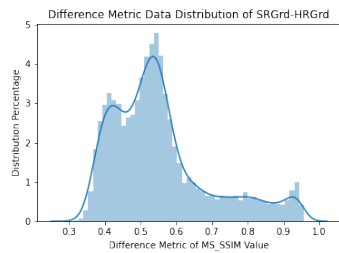
(a) PSNR distribution



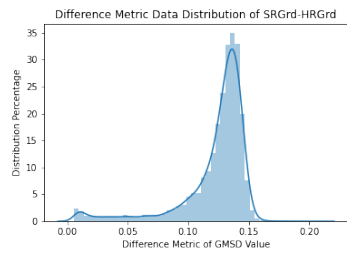
(b) SSIM distribution



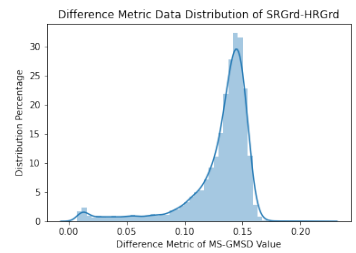
(c) FSIM distribution



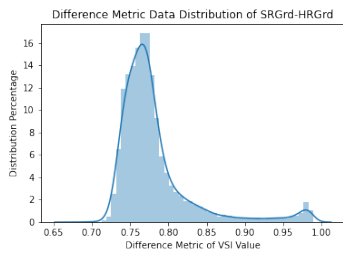
(d) MS-SSIM distribution



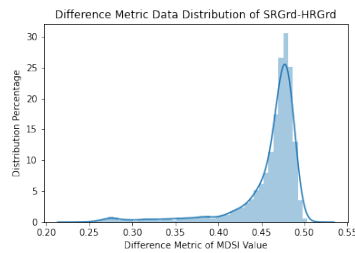
(e) GMSD distribution



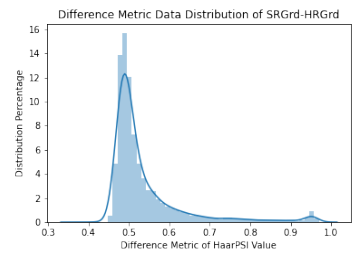
(f) MS-GMSD distribution



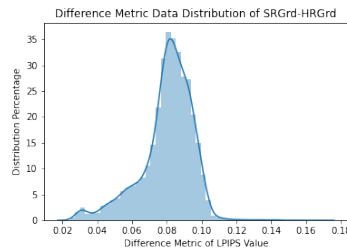
(g) VSI distribution



(h) MDSI distribution

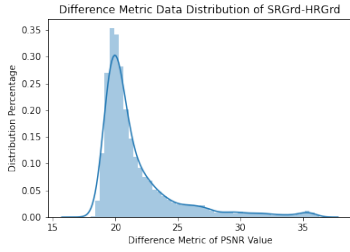


(i) HaarPSI distribution

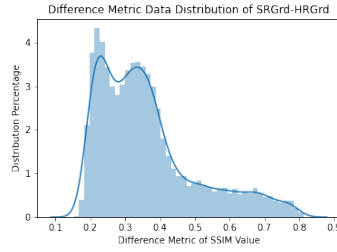


(j) LPIPS distribution

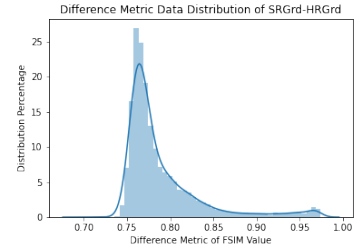
Figure 5.10: Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the gradient of HR patch to the gradient of SR patch



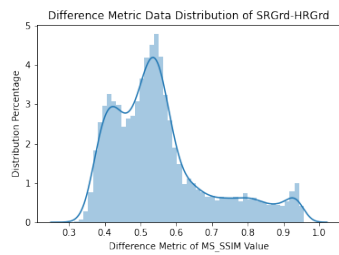
(a) PSNR distribution



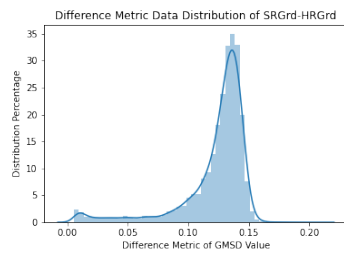
(b) SSIM distribution



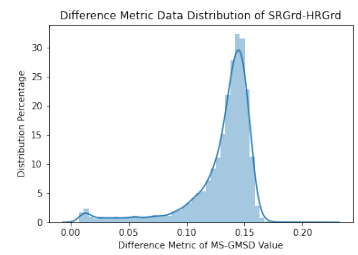
(c) FSIM distribution



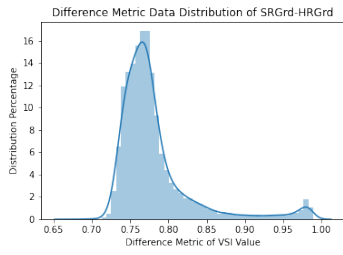
(d) MS-SSIM distribution



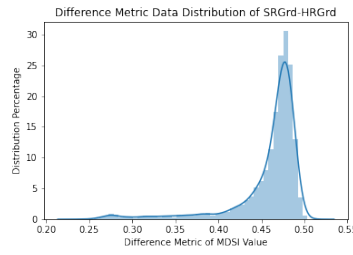
(e) GMSD distribution



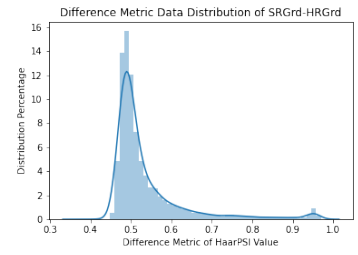
(f) MS-GMSD distribution



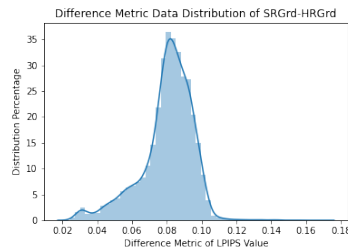
(g) VSI distribution



(h) MDSI distribution



(i) HaarPSI distribution



(j) LPIPS distribution

Figure 5.11: Data distribution visualization of the ten difference/similarity measures for 8X magnification comparing the gradient of HR patch to the gradient of SR patch histogram matched to the HR patch

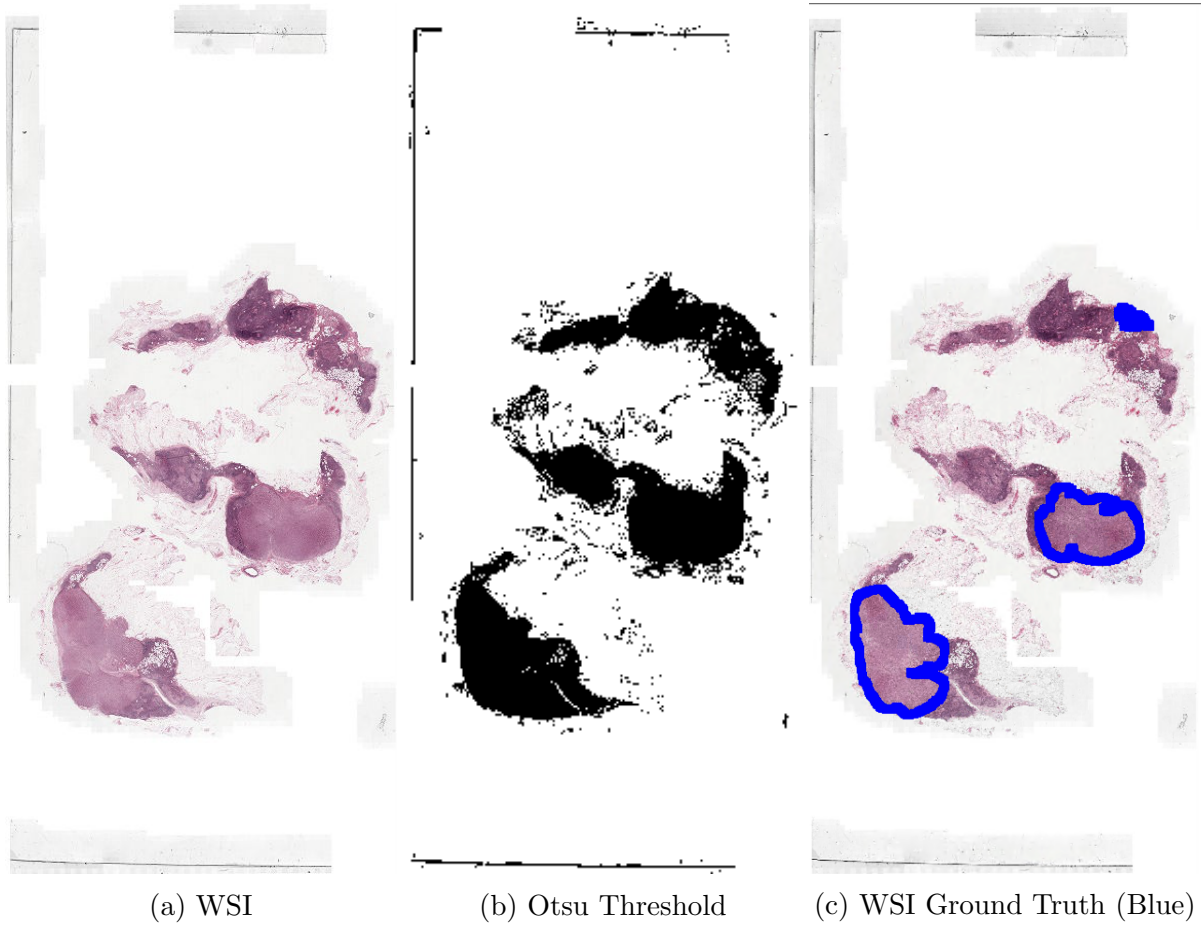
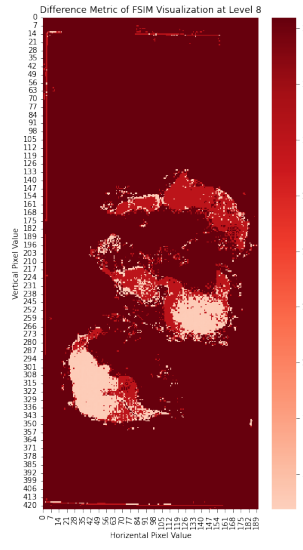
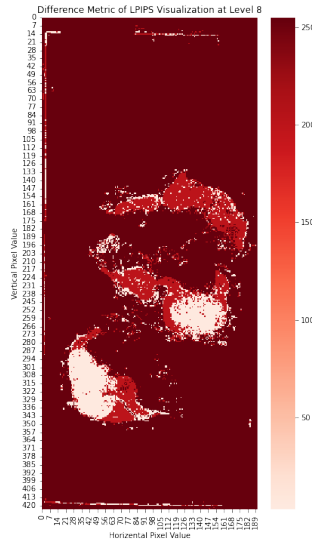


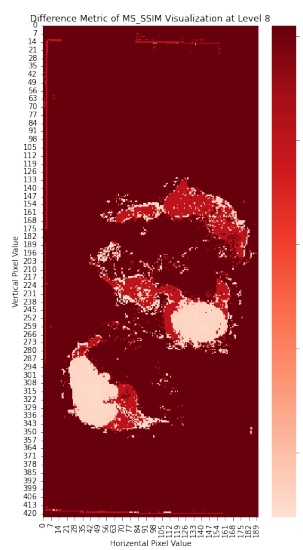
Figure 5.12: Visualization of WSI, Otsu Thresholding of the tissue region, and the metastasis region shown in blue



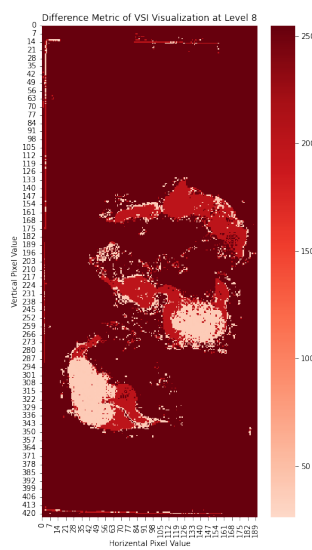
(a) FSIM



(b) LPIPS

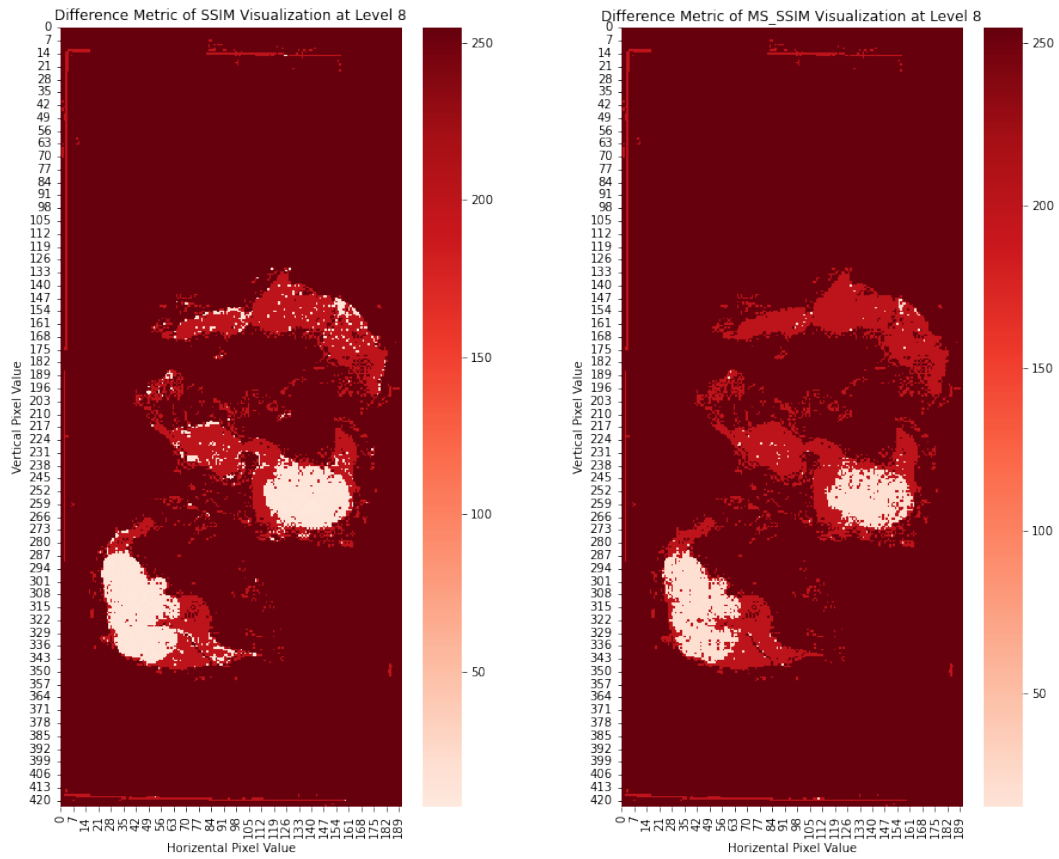


(c) MS-SSIM



(d) VSI

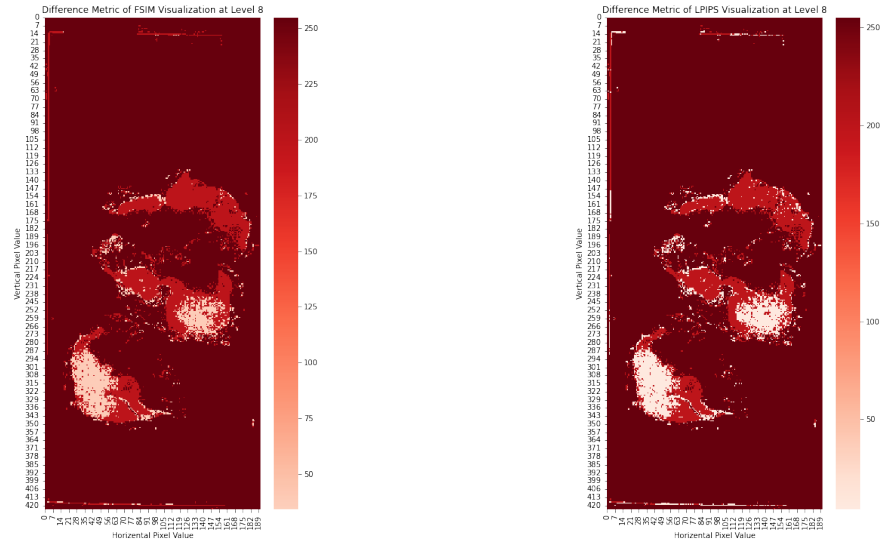
Figure 5.13: Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the SR generated patch



(a) SSIM

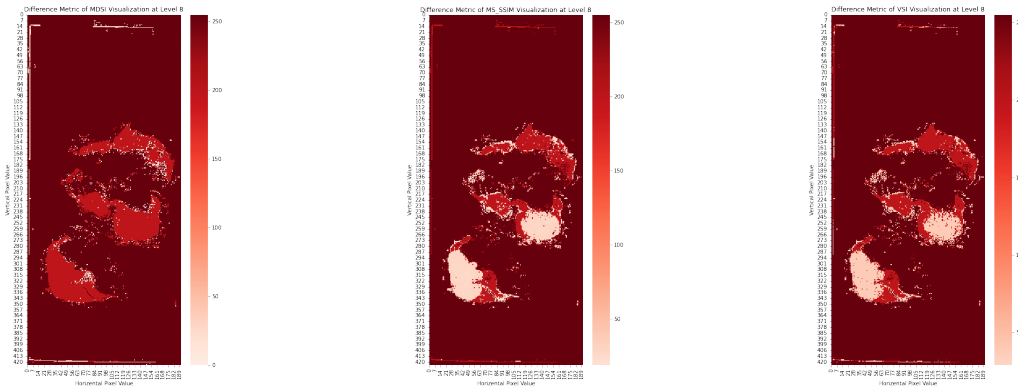
(b) MS-SSIM

Figure 5.14: Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the SR generated patch



(a) FSIM

(b) LPIPS

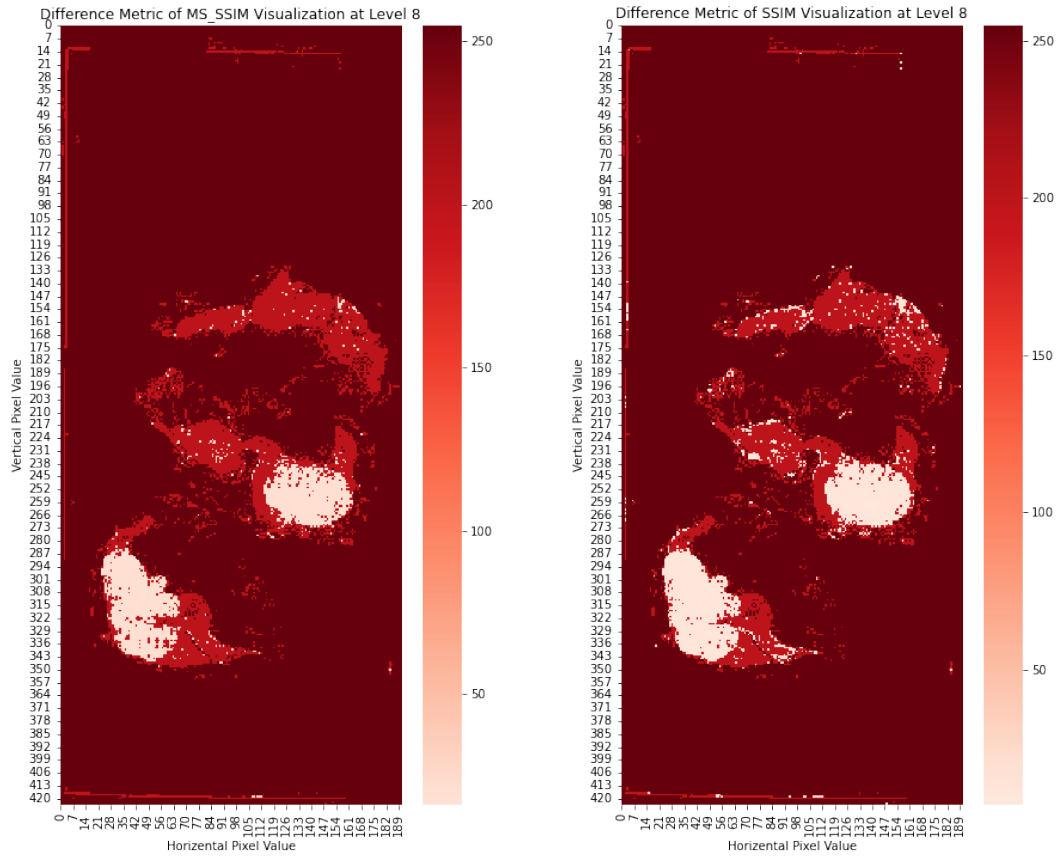


(c) MDSI

(d) MS-SSIM

(e) VSI

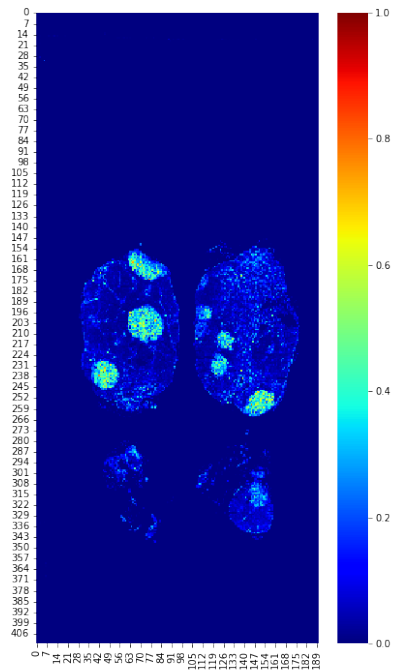
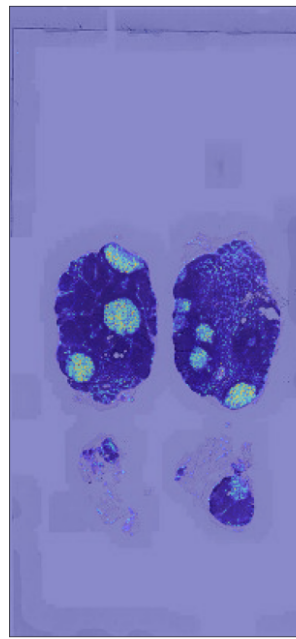
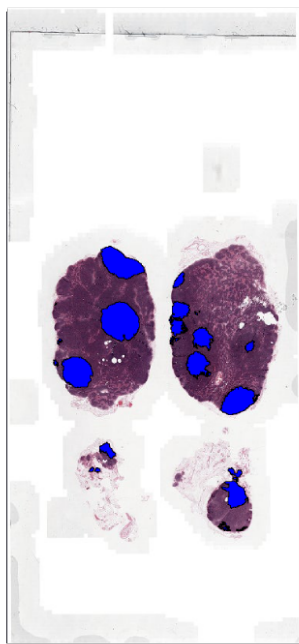
Figure 5.15: Visualization of FSIM, LPIPS, MS-SSIM, and VSI measure on the WSI comparing the original patch to the histogram matching SR generated patch



(a) MS-SSIM

(b) SSIM

Figure 5.16: Visualization of MS-SSIM, and SSIM measure on the WSI comparing the original patch to the histogram matching SR generated patch

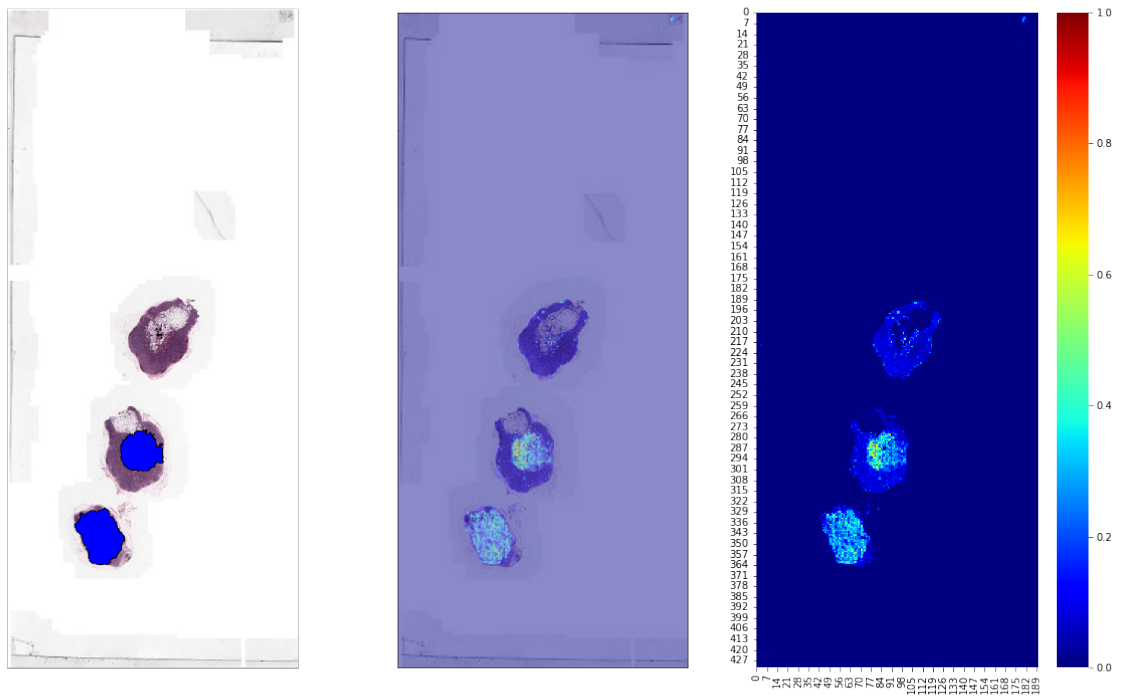


(a) GT overlay on WSI

(b) Pred. overlay on WSI

(c) Prediction

Figure 5.17: Visualization of differences distribution assessment for 16X magnification



(a) GT overlay on WSI

(b) Pred. overlay on WSI

(c) Prediction

Figure 5.18: Visualization of differences distribution assessment for 16X magnification

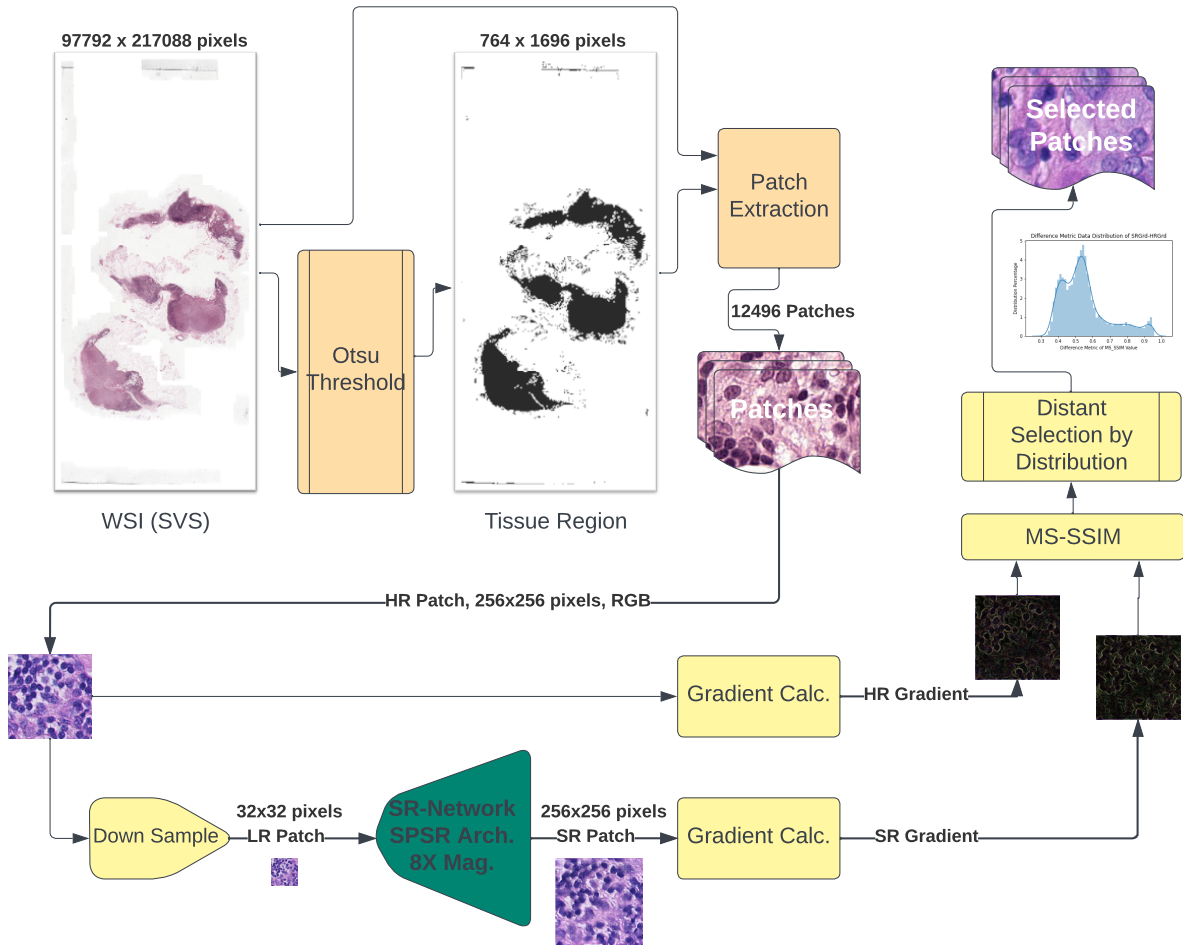


Figure 5.19: Proposed method for patch selection of a WSI

Chapter 6

Summary and Remarks

6.1 Summary

The field of digital histopathology relies heavily on the invaluable past experiences that specialists in the field have accumulated while treating patients. These experiences, i.e., evidently diagnosed cases, may help specialists develop more effective treatment strategies and analyses. The glass slides are the gold standard for diagnosis and prognosis in the field of pathology. However, due to the enormous number of patients and numerous slides for each patient, this information is expanding at an almost exponential rate.

Currently, digital pathology aims to improve the treatment process by using cutting-edge computer approaches for faster and more accurate diagnosis. Many of the approaches, in fact, depend on digital slides. This has resulted in extensive WSI archives, a process that is accepted in the post-pandemic world. During the past few years, research on deep networks in the area of super-resolution has gained considerable attention. Among the most critical applications of these deep networks is medical imaging. A few researchers have recently started working on similar problems. This thesis gives a framework to aid in the efficient storage of these images. Technically, this framework saves a drastically reduced version of the original WSI. In addition, it offers a way to restore a comparable version of the original WSI from the smaller version.

In summary, this thesis contains three significant assertions. A reduced version format has been offered initially. This version includes the downsampled version of the original

WSI as well as a set of patches at the original quality and resolution. Second, it provides Super-Resolution to restore the original WSI from the suggested reduced version. In order to guarantee a meaningful implementation, it has also been proposed that by storing the original patch set, the final rebuilt picture may provide the expert with more confidence. Ultimately, the set of chosen patches is decided, which is offered as a patch selection method.

6.2 Remarks

To summarise the findings of this study, they may be divided into three categories. First, this research is aimed at designing a reliable compound magnification framework (CMF) that creates a new file format for the enhancement of digital pathology. CMF is a file format that combines a low-resolution slide with selected high-resolution patches. Within this framework, large WSI files are stored in much smaller sizes, enabling the storage of more slides and, consequently, fewer expenses for hospitals and more research data for physicians. CMF provides high-resolution images that could be further used as references to enhance the LR-HR mapping. The structure of the tissues has similarities that can play a significant part in this area. Besides, two methods for transforming CMF to WSI and wise-verse were advised. Second, The applications of generative models in this area can enhance the accessibility of affordable medical diagnoses to everyone in the world due to the adoption of digital pathology. Experiments with the super-resolution in the reconstruction path have established a baseline to generate synthetic WSIs. The created synthetic WSIs are trustworthy from numerous angles. These include computer measurement from PSNR and SSIM viewpoints, qualitative evaluation, human assessment by field specialists for the quality of the synthetic pictures, and, most crucially, the synthetic images' diagnostic capacity. Overall, the findings indicate that the synthetic pictures generated using SR deep neural networks are reliable replacements for the original image. In addition, the evaluation revealed disparities in the particulars of the various tissue organ slides. This information may then be used to make choices about storage, patch selection, etc. Third, the potential patch selection methods were also assessed. The current techniques are aimed to perform as fast as possible. The reason behind this selection is that the hospitals, as the primary collector of WSIs, may not always be equipped with high-performance storage devices and fast computers (e.g., small clinics and community hospitals). On the basis of the proposed framework, there is a possibility for potential future adjustment. The fact that a collection of patches are kept at a high resolution may facilitate several future research projects relevant to pictures stored using the framework. Under the assumption

that the suggested framework is implemented, the patch selection approach provides the following advantages:

- **Unsupervised selection:** The patch selection should be applicable to a variety of tissue types. However, not all tissue types are extensively annotated, and even if they are for a given dataset, it is uncertain if a domain change would result in relevant consequences. Changes in the scanner or image source (i.e., hospital) may suffice to signal a domain shift.
- **Content aware selection:** The selection of patches is a crucial component of this strategy. The procedure that picks patches should base its decision on the patch's content. This has been accomplished by directing the selection procedure toward anomalous or abnormal regions since they are the most diagnostically valuable locations.
- **Edges importance:** The edges amplify crucial information, such as the cell boundaries. The suggested technique would benefit from gradients to identify these edges and choose patches depending on their area. In addition, gradients have guided the SR generation procedure for patch selection to enhance the existing edge presentations further.
- **Benefiting from the inaccuracies:** Moving from supervised to unsupervised techniques, it is evident that, due to the quick shift in the amount of information utilized for training, the output accuracy levels would often not be comparable. In the suggested method, however, we benefit from the lower accuracy in the sense that if a patch is picked from beyond the abnormality zone, it is due to a lack of precision during the upsampling, which is a salient signal of the significance of keeping that patch.
- **Speed:** The evaluation of WSIs, in terms of assessing all patches on a slide, is a highly costly method that needs either extremely fast processors (which are expensive) or lengthy waiting periods until the findings are available. Nonetheless, given the framework offered by various computer optimizations, the approach is capable of analyzing large WSIs at the highest magnification (i.e., 40X) within a realistic period of minutes, dependant on the size of the tissue area.

On the other hand, another set of methods that are proposed in the literature focuses on lossy compression (and decompression) of images. Although these methods may offer

more storage savings comparing the provided solution, there still is a problem where the images that are stored lose their original specifications. In the solution provided here, the stored data is the original form, which could offer outstanding reliability for the future use-case. In other words, the method provided here only removes the sets of pieces of information that could be reliably restored.

6.3 Future research

The author feels that this discovery may pave a new thriving venue in digital pathology. To create a production-level framework, however, there are suggestions that might further improve the process:

- **Super-Resolution:** The proposed upsampling technique depends mostly on a single patch to obtain the higher resolution version. However, WSI contains a vast quantity of information, such as the position, surrounding patch information, and diagnostic reports, that might be utilized to enhance the approach.
- **Patch Selection:** The suggested patch selection method examines WSI abnormalities with more specificity in an effort to emulate the regeneration process. Although this is very significant, the author argues that further study is necessary to determine which patches are really essential. If an expert is asked to identify a subset of patches, his rationale and response may still be ambiguous.

Another viable approach is to store a summary or a feature vector of dropped images to be used in a reconstruction process. This can also be experimented with using reference-based SR approaches where the reference may be a vector created during the compression procedure. Potentially this solution could be combined with a lossy compression solution where the constraints are the usage of the downsampled image and the summary, which can act as a leading key for reconstruction. Finally, the author experimented with discriminator networks for the patch selection method, though the results did not provide helpful information. It is reasoned that despite its failure, it still seems to be a better solution than the provided solution. Potential changes for discriminator training can result in better accuracy. For instance, the gradient calculation could be entered into the discriminator network input.

References

- [1] Liron Pantanowitz, Paul N Valenstein, Andrew J Evans, Keith J Kaplan, John D Pfeifer, David C Wilbur, Laura C Collins, and Terence J Colgan. Review of the current state of whole slide imaging in pathology. *Journal of pathology informatics*, 2(1):36, 2011.
- [2] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of pathology informatics*, 1, 2010.
- [3] Ewout W Steyerberg et al. *Clinical prediction models*. Springer, 2019.
- [4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [5] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.
- [6] Yangqin Feng, Lei Zhang, and Zhang Yi. Breast cancer cell nuclei classification in histopathology images using deep neural networks. *International journal of computer assisted radiology and surgery*, 13(2):179–191, 2018.
- [7] Freddie Bray, Jian-Song Ren, Eric Masuyer, and Jacques Ferlay. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International journal of cancer*, 132(5):1133–1145, 2013.
- [8] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

- [9] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: challenges and opportunities. *Journal of pathology informatics*, 9, 2018.
- [10] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, 65(20):20TR01, 2020.
- [11] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps*, pages 323–350, 2018.
- [12] Chetan L Srinidhi, Ozan Ciga, and Anne L Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, page 101813, 2020.
- [13] Xin Yuan and Raziel Haimi-Cohen. Image compression based on compressive sensing: End-to-end comparison with jpeg. *IEEE Transactions on Multimedia*, 22(11):2889–2904, 2020.
- [14] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020.
- [15] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4472–4480, 2017.
- [16] Sidney L Shaw, ED Salmon, and RS Quatrano. Digital photography for the light microscope: results with a gated, video-rate ccd camera and nih-image software. *Biotechniques*, 19(6):946–955, 1995.
- [17] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. Openslide: A vendor-neutral software foundation for digital pathology. *Journal of pathology informatics*, 4, 2013.
- [18] Ronald S Weinstein, Anna R Graham, Lynne C Richter, Gail P Barker, Elizabeth A Krupinski, Ana Maria Lopez, Kristine A Erps, Achyut K Bhattacharyya, Yukako Yagi, and John R Gilbertson. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human pathology*, 40(8):1057–1069, 2009.

- [19] Xiaoyou Ying and Thomas M Monticello. Modern imaging technologies in toxicologic pathology: An overview. *Toxicologic pathology*, 34(7):815–826, 2006.
- [20] Shujian Deng, Xin Zhang, Wen Yan, Eric I Chang, Yubo Fan, Maode Lai, Yan Xu, et al. Deep learning in digital pathology image analysis: a survey. *Frontiers of medicine*, 14(4):470–487, 2020.
- [21] Leila H Eadie, Paul Taylor, and Adam P Gibson. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European journal of radiology*, 81(1):e70–e76, 2012.
- [22] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The lancet oncology*, 20(5):e253–e261, 2019.
- [23] Tonghe Wang, Yang Lei, Yabo Fu, Jacob F Wynne, Walter J Curran, Tian Liu, and Xiaofeng Yang. A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of Applied Clinical Medical Physics*, 22(1):11–36, 2021.
- [24] Oscar Jimenez-del Toro, Sebastian Otálora, Mats Andersson, Kristian Eurén, Martin Hedlund, Mikael Rousson, Henning Müller, and Manfredo Atzori. Analysis of histopathology images: From traditional machine learning to deep learning. In *Biomedical Texture Analysis*, pages 281–314. Elsevier, 2017.
- [25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [26] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- [27] Aditya Deshpande, Jiajun Lu, Mao-Chuang Yeh, Min Jin Chong, and David Forsyth. Learning diverse image colorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6837–6845, 2017.
- [28] Omar Elharrouss, Noor Almaadeed, Somaya Al-Maadeed, and Younes Akbari. Image inpainting: A review. *Neural Processing Letters*, 51(2):2007–2028, 2020.
- [29] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

- [30] Aziz Alotaibi. Deep generative adversarial networks for image-to-image translation: A review. *Symmetry*, 12(10):1705, 2020.
- [31] Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.
- [32] JD Van Ouwerkerk. Image super-resolution survey. *Image and vision Computing*, 24(10):1039–1052, 2006.
- [33] Chunwei Tian, Ruibin Zhuge, Zhihao Wu, Yong Xu, Wangmeng Zuo, Chen Chen, and Chia-Wen Lin. Lightweight image super-resolution with enhanced cnn. *Knowledge-Based Systems*, 205:106235, 2020.
- [34] Andrey Ignatov, Radu Timofte, Maurizio Denna, and Abdel Younes. Real-time quantized image super-resolution on mobile npus, mobile ai 2021 challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2525–2534, 2021.
- [35] Mehdi SM Sajjadi, Raviteja Vemulapalli, and Matthew Brown. Frame-recurrent video super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6626–6634, 2018.
- [36] Haitian Zheng, Mengqi Ji, Haoqian Wang, Yebin Liu, and Lu Fang. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In *Proceedings of the European conference on computer vision (ECCV)*, pages 88–104, 2018.
- [37] Rushi Lan, Long Sun, Zhenbing Liu, Huimin Lu, Cheng Pang, and Xiaonan Luo. Madnet: a fast and lightweight network for single-image super resolution. *IEEE transactions on cybernetics*, 51(3):1443–1453, 2020.
- [38] Kirsten Christensen-Jeffries, Olivier Couture, Paul A Dayton, Yonina C Eldar, Kullervo Hynynen, Fabian Kiessling, Meaghan O’Reilly, Gianmarco F Pinton, Georg Schmitz, Meng-Xing Tang, et al. Super-resolution ultrasound imaging. *Ultrasound in Medicine & Biology*, 46(4):865–891, 2020.
- [39] Hayit Greenspan. Super-resolution in medical imaging. *The computer journal*, 52(1):43–63, 2009.
- [40] François Rousseau, Kio Kim, Colin Studholme, Meriam Koob, and J-L Dietemann. On super-resolution for fetal brain mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 355–362. Springer, 2010.

- [41] Eric Van Reeth, Ivan WK Tham, Cher Heng Tan, and Chueh Loo Poh. Super-resolution in magnetic resonance imaging: a review. *Concepts in Magnetic Resonance Part A*, 40(6):306–325, 2012.
- [42] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859, 2010.
- [43] Hilario Seibel, Siome Goldenstein, and Anderson Rocha. Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos. *IEEE access*, 5:20020–20035, 2017.
- [44] Pourya Shamsolmoali, Masoumeh Zareapoor, Deepak Kumar Jain, Vinay Kumar Jain, and Jie Yang. Deep convolution network for surveillance records super-resolution. *Multimedia Tools and Applications*, 78(17):23815–23829, 2019.
- [45] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. Is image super-resolution helpful for other vision tasks? In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.
- [46] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Task-driven super resolution: Object detection in low-resolution images. *arXiv preprint arXiv:1803.11316*, 2018.
- [47] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4491–4500, 2017.
- [48] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.
- [49] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8):1016–1022, 1979.
- [50] Jian Sun and Heung-Yeung Shum. Image super-resolution using gradient profile prior, June 23 2015. US Patent 9,064,476.
- [51] Kwang In Kim and Younghee Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE transactions on pattern analysis and machine intelligence*, 32(6):1127–1133, 2010.

- [52] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, 22(2):56–65, 2002.
- [53] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.
- [54] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan 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.
- [55] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.
- [56] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.
- [57] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018.
- [58] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [59] Adrian Bulat and Georgios Tzimiropoulos. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2018.
- [60] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

- [61] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 864–873, 2018.
- [62] Weisi Lin and Li Dong. Adaptive downsampling to improve image compression at low bit rates. *IEEE Transactions on Image Processing*, 15(9):2513–2521, 2006.
- [63] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239, 1991.
- [64] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018.
- [65] Ralph E Carlson and Frederick N Fritsch. Monotone piecewise bicubic interpolation. *SIAM journal on numerical analysis*, 22(2):386–400, 1985.
- [66] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks, 2018.
- [67] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2017.
- [68] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2019.
- [69] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution, 2019.
- [70] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [71] Adrian B Levine, Jason Peng, David Farnell, Mitchell Nursey, Yiping Wang, Julia R Naso, Hezhen Ren, Hossein Farahani, Colin Chen, Derek Chiu, et al. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *The Journal of pathology*, 252(2):178–188, 2020.

- [72] Shahrukh Athar and Zhou Wang. A comprehensive performance evaluation of image quality assessment algorithms. *Ieee Access*, 7:140030–140070, 2019.
- [73] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.
- [74] Quan Huynh-Thu and Mohammed Ghanbari. The accuracy of psnr in predicting video quality for different video scenes and frame rates. *Telecommunication Systems*, 49(1):35–48, 2012.
- [75] Zhou Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.
- [76] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [77] Jean-François Pambrun and Rita Noumeir. Limitations of the ssim quality metric in the context of diagnostic imaging. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 2960–2963. IEEE, 2015.
- [78] C Karthikeyan and B Ramadoss. Comparative analysis of similarity measure performance for multimodality image fusion using dtcwt and sofm with various medical image fusion techniques. *Indian Journal of Science and Technology*, 9(22):1–6, 2016.
- [79] Basant Kumar, Shishu Bind Kumar, and Chandrahas Kumar. Development of improved ssim quality index for compressed medical images. In *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, pages 251–255, 2013.
- [80] Santanu Roy, Alok kumar Jain, Shyam Lal, and Jyoti Kini. A study about color normalization methods for histopathology images. *Micron*, 114:42–61, 2018.
- [81] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.
- [82] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.

- [83] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2):684–695, 2013.
- [84] Bo Zhang, Pedro V. Sander, and Amine Bermak. Gradient magnitude similarity deviation on multiple scales for color image quality assessment. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1253–1257, 2017.
- [85] Lin Zhang, Ying Shen, and Hongyu Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014.
- [86] Hossein Ziaei Nafchi, Atena Shahkolaei, Rachid Hedjam, and Mohamed Cheriet. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access*, 4:5579–5590, 2016.
- [87] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
- [88] Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018.
- [89] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, April 2018.
- [90] Abtin Riasatian, Morteza Babaie, Danial Maleki, Shivam Kalra, Mojtaba Valipour, Sobhan Hemati, Mani Zaveri, Amir Safarpour, Sobhan Shafiei, Mehdi Afshari, et al. Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Medical Image Analysis*, 70:102032, 2021.
- [91] Xifeng Guo, Xinwang Liu, En Zhu, and Jianping Yin. Deep clustering with convolutional autoencoders. In *International conference on neural information processing*, pages 373–382. Springer, 2017.
- [92] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487. PMLR, 2016.

- [93] Qing Lyu, Chenyu You, Hongming Shan, and Ge Wang. Super-resolution mri through deep learning. *arXiv preprint arXiv:1810.06776*, 2018.
- [94] Lopamudra Mukherjee, Adib Keikhosravi, Dat Bui, and Kevin W Eliceiri. Convolutional neural networks for whole slide image superresolution. *Biomedical optics express*, 9(11):5368–5386, 2018.
- [95] Lopamudra Mukherjee, Huu Dat Bui, Adib Keikhosravi, Agnes Loeffler, and Kevin W Eliceiri. Super-resolution recurrent convolutional neural networks for learning with multi-resolution whole slide images. *Journal of biomedical optics*, 24(12):126003, 2019.
- [96] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int’l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [97] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017.
- [98] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.
- [99] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1865–1873, 2016.
- [100] Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. Interactive manga retargeting. In *ACM SIGGRAPH 2011 Posters*, SIGGRAPH ’11, New York, NY, USA, 2011. Association for Computing Machinery.
- [101] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.
- [102] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, pages 135.1–135.10. BMVA Press, 2012.

- [103] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.
- [104] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015.
- [105] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, Cancer Genome Atlas Research Network, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113, 2013.
- [106] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution, 2018.
- [107] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution, 2018.
- [108] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks, 2018.
- [109] Philippe M Burlina, Neil Joshi, Katia D Pacheco, TY Alvin Liu, and Neil M Bressler. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA ophthalmology*, 137(3):258–264, 2019.
- [110] Jongyoo Kim and Sanghoon Lee. Deep learning of human visual sensitivity in image quality assessment framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1676–1684, 2017.
- [111] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.
- [112] Danial Maleki, Mehdi Afshari, Morteza Babaie, and Hamid R Tizhoosh. Ink marker segmentation in histopathology images using deep learning. In *International Symposium on Visual Computing*, pages 359–368. Springer, 2020.

- [113] Mehdi Afshari and HR Tizhoosh. A similarity measure of histopathology images by deep embeddings. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3447–3450. IEEE, 2021.
- [114] Jian Ren, Ilker Hacihaliloglu, Eric A Singer, David J Foran, and Xin Qi. Unsupervised domain adaptation for classification of histopathology whole-slide images. *Frontiers in bioengineering and biotechnology*, 7:102, 2019.
- [115] Meriem Sebai, Xinggang Wang, and Tianjiang Wang. Maskmitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. *Medical & Biological Engineering & Computing*, 58(7):1603–1623, 2020.
- [116] Bashar Emon, Jessica Bauer, Yasna Jain, Barbara Jung, and Taher Saif. Biophysics of tumor microenvironment and cancer metastasis—a mini review. *Computational and Structural Biotechnology Journal*, 16:279–287, 2018.
- [117] David F Steiner, Robert MacDonald, Yun Liu, Peter Truszkowski, Jason D Hipp, Christopher Gammage, Florence Thng, Lily Peng, and Martin C Stumpe. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *The American journal of surgical pathology*, 42(12):1636, 2018.
- [118] Kevin Faust, Quin Xie, Dominick Han, Kartikay Goyle, Zoya Volynskaya, Ugljesa Djuric, and Phedias Diamandis. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC bioinformatics*, 19(1):1–15, 2018.
- [119] Milda Pocevičiūtė, Gabriel Eilertsen, and Claes Lundström. Unsupervised anomaly detection in digital pathology using gans. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1878–1882. IEEE, 2021.
- [120] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [121] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. In *International conference on machine learning*, pages 1462–1471. PMLR, 2015.
- [122] Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*, 2016.

- [123] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [124] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020.
- [125] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermesen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.