Automating Interpretation of Images and Visual Inspections in Modern Manufacturing and Medical Settings

by

Behnam Rasoolian

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 1, 2021

Keywords: Image Processing, Machine Learning, Neural Networks

Approved by

Daniel Silva, Chair, Assistant Professor of Industrial and Systems Engineering
Jeffrey Smith, Joe W. Forehand Jr. Professor of Industrial and Systems Engineering
Richard Sesek, Tim Cook Associate Professor of Industrial and Systems Engineering
Nima Shamsaei, Philpott-WestPoint Stevens Distinguished Associate Professor of Mechanical Engineering

Abstract

Automating visual inspection is rapidly finding its way into non-invasive industrial and medical processes. In this work, we address detection automation opportunities in the task of interpreting both medical and industrial images, we combine domain knowledge with state-of-the-art, domain-independent methods such as Convolutional Neural Networks (CNN) to address them. Specifically, we address two problems in additive manufacturing and one in medical imaging as described below. We provide a method to improve the resolution of and automate the extraction of local and global porosity features from XCT images of additively manufactured Ti-6Al-4V. We also address the automation issues in characterizing local roughness features, particularly the curvature of microscopic surface pits observed in microscopy images of AM Ti-6Al-4V. Finally, we developed a deep learning model to automate the task of detection and segmentation of different muscles in images obtained from MRI of the lower back.

Acknowledgments

To my wife, for her patience and unbeleivable support.

To my parents, for all these years we were apart.

To my advisor, Dr. Silva, for his guidances.

To Dr. Saeed Maghsoodloo, who was like a grandfather to me, and helped me improve at all stages.

Table of Contents

Abstract ii

Acknowledgments iii

1 Introduction 1

2 Technologies and Methods 4

2.1 Visual inspection and imaging technologies 4

2.1.1 Non-destructive evaluation in manufacturing 6

2.1.2 Non-destructive imaging techniques applied to additive manufacturing parts 7

2.1.3 Non-invasive measurements in medicine and ergonomics 7

2.2 Deep Learning, Convolutional and Generative Adversarial Neural Networks 9

2.3 Generative Adversarial Neural Networks 11

2.3.1 GAN-based Super Resolution 12

3 Porosity analysis and resolution improvement of XCT images of additively manufactured metallic parts 14

3.1 Introduction 15

3.2 Materials and Methods: 19

3.2.1 Data 19

3.2.2 Super-resolution Model 20

3.2.3 Pore Detection and Graphical Representation 21

3.3 Results . . . 25
3.3.1 Super-resolution . . . 25
3.3.2 Pair-wise pore feature comparison: . . . 29
3.4 Conclusion . . . 31

4 An Automated Method for Geometrical Surface Characterization for Fatigue Analysis of Additive Manufactured Parts . . . 32
4.1 Introduction . . . 32
4.1.1 Surface roughness parameters . . . 34
4.1.2 Radii of Curvature . . . 36
4.2 Material and methods . . . 37
4.2.1 Data pre-processing . . . 37
4.2.2 Radius of curvature of valleys . . . 39
4.3 Experimental Results . . . 44
4.3.1 Statistical Analysis of SR parameters . . . 44
4.3.2 Comparing different resolutions . . . 45
4.3.3 The effect of sub-areas . . . 46
4.3.4 Radii of curvature . . . 47
4.4 Conclusions . . . 50

5 Muscle Recognition in lower-back MRI Using CNN . . . 52
5.1 Introduction . . . 52
5.2 Recent applications of Deep CNNs in medical images . . . 55
5.3 Data Collection and Preparation . . . 57
5.3.1 Experimental procedures . . . 57
5.3.2 Data . . . 58

5.3.3 Data Preprocessing 59
5.4 Model Design 61
5.4.1 Hyperparameters 63
5.4.2 Accounting for data imbalance 63
5.4.3 Training 65
5.4.4 Post-processing 65
5.5 Results & discussion 66
5.5.1 Performance Metrics 66
5.5.2 Lever arms, area, and centroids 69
5.5.3 Speed-up 69
5.6 Conclusions and future work 70

Bibliography 71

List of Figures

2.1 (left) A traditional neural network (right) A convolutional layer [41] 10

2.2 A conceptual GAN design 12

3.1 VGG loss 21

3.2 Relationship between low- and high-resolution images. The reconstructed 3d version of the low-resolution images is depicted on the top left corner while the high-resolution image is shown in the top right-corner. The bottom image shows how the high-resolution image only covers a small sub-volume of the low-resolution data. 22

3.3 A comparison of pores detected from the same section in the three images (low-resolution, improved resolution, and high-resolution images) 26

3.4 3D formation of pores detected 27

3.5 Visual comparison of a single matched pore. 30

4.1 A surface roughness image(top) with a single profile selected via a red line. The roughness profile corresponding to the heigh of the surface on the red line is plotted in the bottom image 35

4.2 The descriptions of before and after the misalignment correction 38

4.3 (a) 3D demonstration of tilt correction and (b) 2D profiled line before and after tilt/alignment correction with numerical height values. 40

4.4 Valley local minima (valleys) that fall below one standard deviation of average roughness in a roughness profile. 41

4.5 Process of choosing a neighborhood for a typical valley. In order to maintain symmetry, the closest inflection point to the local minimum is chosen (in this case the inflection point that is to the right) and mirrored on the other side of the minimum. . 42

4.6 Valley radii for a roughness profile. The golden horizontal line separates the valleys that are one standard deviation below $\bar{y}$. 44

4.7 The bar plots of several SR parameters based on the height data measured by 500X and 1000X magnifications. 46

4.8 (a) Schematically described sub-areas with various colors. (b) SR parameters of each sub-areas with corresponded colors. 48

4.9 A comparison of fatigue life computations from Pegues et al. and the automated method proposed herein. 50

5.1 (left) A typical MR image. (right) The same image with the muscles of interest traced. 59

5.2 Muscle class extraction from annotated images. First, the manual traces are converted to binary masks representing muscle segmentation (top left). Second, the segments are assigned to a muscle based on their relative position (top right). Finally, muscle features such as shape and lever arms are computed from classified segments (bottom). 60

5.3 Muscle class Numbers overlayed on muscle segments. Generally, the data set consists or two different formation of the same muscles. The first one (left) consists of 10 muscles and the spine while the second one (right) consists of 8 muscles and the spine (left and right Quadratus are not present) 61

5.4 CNN Model Architecture 62

5.5 Adding noise to the background to avoid unbalanced classes. 64

5.6 Original segmented image (left) together with raw (middle) and post-processed (right) segmentations. 65

5.7 A comparison of 90^{o} lever arm distances from original traced (left) and post-processed segmentations (right) 66

List of Tables

3.1 Definitions of true positive, true negative, false positive and false negative in the context of pore pixel prediction 27

3.2 Pixel-wise detection metrics 28

3.3 Comparing images in terms of number of pores predicted 29

3.4 Matched pore feature comparisons between the low-resolution image and the improved-resolution image. 30

4.1 : Surface Roughness Parameters List 34

4.2 P-values for different metrics based on t-tests comparing 1000x and 500x resolution images 45

4.3 : t-test p-values for difference in SR parameters between different surface subsample images and the whole surface 47

4.4 A quantitative comparison of the manual and automatic methods of radii extraction using minimum squared errors 50

5.1 Muscle Names and their corresponding class numbers used in prediction 60

5.2 Hyperparameters used to train the CNN model. (Conv and Up are short forms of Convolutional Layer and Up-sampling layers respectively) 63

5.3 Accuracy, Precision, and Recall: 68

5.4 Lever arm, area, and centroid statistics. All values are median values across all predicted muscles. 70

Chapter 1

Introduction

Automatic visual inspection is rapidly finding its way into non-invasive industrial and medical processes. This development especially benefits from increasingly powerful computers and reasonably priced imaging technologies. Visual inspection means inspecting by looking. On the one hand, visual inspection can ensure specific quality characteristics, and on the other hand, it can be used to detect and analyze variations, anomalies and patterns. Nearly all industrially produced goods are visually inspected during manufacturing, assembly and/or transportation [8]. In the health care industry, radiologists and pathologist inspect a large number of medical images in order to detect patterns and anomalies. Traditionally, industrial and medical inspection tasks are either carried out by specialized personnel, appropriately trained for this, or by regular employees in addition to their regular duties.

Although the human visual system, consisting of the eyes and brain, has numerous astonishing capabilities that surpass machines in many ways, visual inspection carried out by humans comes with some unavoidable disadvantages. Tediousness, fatigue, lack of objectivity and reproducibility, and costliness in terms of both time and money are among the most important disadvantages of manual inspection [8]. For example, in the health care industry, the abundance of data and the progress in imaging has changed how radiologists interpret images. For a single patient with multiple trauma, a radiologist typically views 4000 CT scan images, leading to fatigue [36].

As a result, it is desirable to automate visual inspection where saving time and human effort is possible. A good automated system works without fatigue, provides objective and reproducible

results, and allows a complete and detailed documentation. In the last few decades, computer vision has been used to detect large defects in industrial parts or to characterize surface texture. In health care, Watson, an IBM prototype for artificial intelligence, can identify pulmonary embolism on CT images and detect abnormal wall motion on echocardiography [36].

In this dissertation, we address detection automation opportunities in the task of interpreting both medical and industrial images. To do this, we combine domain knowledge with state-of-the-art, domain-independent methods such as Convolutional Neural Networks (CNN) and deep Generative Adversarial Neural Networks (GAN) to address those tasks. In Chapter 2 we provide some background on both the imaging technologies and image processing methods relevant to this dissertation. Then, in the remaining chapters, we use these technologies and methods to address automatic image interpretation in two problems in additive manufacturing and one in medical imaging, as described below.

In Chapter 3, we automate extraction of local and global porosity features from X-ray CT (XCT) images of additively manufactured (AM) parts of titanium alloy Ti-6Al-4V. We developed a methodology for utilizing the data obtained from XCT in order to visualize and extract features of porosity from two sets of low- and high- resolution scans. We first developed a deep GAN model to improve the resolution of the low-resolution images. We then developed a method that extracts 3-dimensional contours as a representation of pores. Finally, features of position, size, shape and orientation of the pores are extracted. In order to characterize orientation of the pore, we propose an ellipsoid estimation to calculate how much a pore extends in three dimensions. The features can be used to classify pores and investigate their effect on mechanical properties such as fatigue failure and crack initiation. The main contributions of this work are:

- Developing a deep GAN model to improve the resolution of low-resolution XCT images of AM parts, using high-resolution XCT images for training.
- Developing a framework to characterize pores using numerical features that can be used in classification algorithm, such as size, shape and spatial orientation of individual pores.

In Chapter 4, we address automation of characterizing surface roughness features, particularly curvature of microscopic surface pits, in microscopy images of AM parts made of Ti-6Al-4V. Here, we extract geometric features of individual valleys in order to investigate their relationship with fatigue failure and crack initiation points. We developed a method to extract both traditional, and newly proposed global roughness parameters of the whole surface from images collected using microscopy. Additionally, local roughness parameters are also extracted and used to estimate the elastic stress concentration factor. The contributions of this chapter are:

- Automating the currently laborious process of extracting valley radii in roughness data, resulting in significant time reduction while maintaining accuracy and objectivity.
- Using valley radii together with other automatically extracted features of the valleys to predict fatigue life.
- Data-intensive statistical analysis of the effects of different microscopic imaging scenarios on surface roughness parameters.

Finally, in Chapter 5, we move from additive manufacturing to medical images. In this study, we automate the task of detection and segmentation of different muscles in images obtained from MRI of lower back. This automated method facilitates, among other tasks, the analysis of mechanical interactions between muscles for a considerably larger number of individuals. The contributions of this chapter are:

- Development of a deep CNN model to segment muscles from a stack of MRI images.
- Using the model to automate the process of calculating lever arms and other biomechanical measurements.

In all three of the applications above, we developed tools that can handle large image data from different parts/individuals. The three models developed automate the measurement and interpretation of the data, leading to significant time efficiency and equal or higher accuracy of measurements compared to traditional manual approaches.

Chapter 2

Technologies and Methods

This chapter is intended to give the reader the necessary background information necessary to more easily follow subsequent chapters. It is divided into two sections. The first section discussed the technologies involved to obtain the images analyzed. The second section provides an overview of Convolutional Neural Networks and associated methods used throughout the dissertation.

2.1 Visual inspection and imaging technologies

We start this section with an overview of non-invasive technologies. Next, we discuss applications of imaging, especially X-ray Computed Tomography (XCT) in additive manufacturing (AM). We then deal with imaging techniques in surface roughness analysis and roughness parameters used to characterize roughness. Finally, we provide a history of imaging in medical settings and provide some recent works on applications of AI in the field of image analysis and segmentation for human vertebrae.

At the most fundamental form, non-destructive testing (NDT) involves visual inspection – or simply looking at the specimen using a system or set of criteria for evaluation to find defects. Some non-destructive techniques were initially developed for clinical applications. An example of a non-destructive testing technique that has been transferred over from a medical context to an industrial context is digital radiography. This method provides immediate feedback to health care practitioners, particularly dentists, by sending scans to a computer in real-time. Other technological

improvements in non-destructive methods that have been transfered over from medical equipment advances to manufacturing settings include phased array ultrasonic testing, and endoscopy.

In a similar manner, NDT tools the were developed to address industrial applications have been repurposed to be used in medical contexts. For example, medical ultrasound was first used in industry before it was used for medical imaging. Anohter is example is provided in [51] which discusses how an industial NDT method that is used for electronic components can be transfered over for medical imaging with high accuracy.

Seen in this light, non-destructive testing can be considered an interdisciplinary field that involves researches that have different interests and aims.

In both medical and industrial settings, the purpose of performing non-invasive tests is similar. Early and accurate diagnosis in medicine leads to a more effective treatment, in some cases even preventing a health issue. In industry, detecting possible failures by non-destructive techniques helps avoid accidents and plan repairs at convenient times, which minimizes down time.

The difference between the two disciplines is that in industry, non-destructive testing provides a tool for designers to improve the product by experimenting with different designs. In medicine, one cannot change the design but might be able to improve the patient's life. Although medical and industrial imaging occur in different settings and address different issues, they both fall in the same category of non-invasive evaluation in terms of the mechanisms involved. According to [23], such methods have five essential features:

- Providing energy from an external source to the test object.
- Possibly, as a result of defects in the part, the energy distribution in the test object changes
- Detecting the change in energy properties by a sensitive detector.
- Measuring and recording energy readings from the detector
- Interpreting the results and deciding whether the part is defective or not

It is the automation of that last step that this dissertation is concerned with.

2.1.1 Non-destructive evaluation in manufacturing

Imaging in industrial settings falls into the category of non-destructive evaluations (NDE). NDE can be defined as a set of methods used to examine an object without reducing its usefulness in the future. Historically, the term NDT was used to refer to a pass/fail mechanism for parts and components. It was usually used for finding large defects which lead to the rejection of the part. Breakthroughs in NDT technology resulted in the ability to detect smaller defects and it could be used to evaluate the probability of failure of the part given features of the defects leading to the broader term NDE. Later, a basic change in philosophy required designers to incorporate the probability of failure in the design of the product to produce a part that is robust to component failure. This led to the emergence of the field of fracture mechanics. Now flaw detection was not enough and flaw features needed to be gathered to serve as an input to fracture mechanics models to predict the remaining life of the components. This led to the emergence of terms of Quantitative NDE (QNDE) and Non-destructive Characterization (NDC). One categorization of NDE methods provided in [47] is as follows:

1. EM radiation, including • X-ray and Gamma ray • Optical testing with visible photons • Thermal or Infrared imaging • Gigahertz-Terahertz Waves

2. Acoustic Energy, including • Ultrasonic testing • Acoustic Emission

3. Tracer, including • Dye penetrant • Magnetic Particle inspection • Leak testing • Nuclear Tracers (Medicine)

4. Eddy Current

Our work in Chapters 3 and 4 can be categorized as Non-destructive Characterization (NDC). The goals of NDC and QNDE include determining fundamental micro-structural characteristics such as grain size, porosity and texture, or the mechanical properties related to them such as fatigue, creep, and fracture toughness. In such applications, industrial imaging is generally used for two kinds of features. The first involves capturing surface features such as roughness, which we explore in

Chapter 4. The methods used include various microscopy techniques as well as laser scanning. The second kind of feature involves characterizing the interior parts that are impossible or extremely difficult to access non-invasively, which is addressed in Chapter 3. These use imaging methods including Magnetic resonance imaging (MRI), Computed Tomography (CT) and its variants.

2.1.2 Non-destructive imaging techniques applied to additive manufacturing parts

As will be discussed in Chapter 3, it is desirable to find a quick non-destructive means of measurement to investigate porosity of additively manufactured metallic parts that is fast and cost-efficient. Such non-destructive methods can also be incorporated into AM quality control. X-ray computed tomography (XCT) has emerged as a novel tool that can capture images of AM parts with high sensitivity to internal defects. Using such images, it is possible to extract pore features and to use these features for indirectly predicting fatigue failure resistance of AM parts [70].

In Chapter 3, we use XCT images of internal structure of AM parts to characterize their fatigue failure in terms of the distribution of pores inside them. Our work is novel in the sense that it uses a resolution-improvement technique and introduces new characterization features such as convex hulls and minimum-volume bounding ellipsoids to characterize the size and *orientation* of pores, which will then help pick out *individual* pores that can potentially serve as crack initiation sites.

2.1.3 Non-invasive measurements in medicine and ergonomics

According to [57], any measurement system that does not physically breach the skin or enter the body deeply through an external orifice is considered non-invasive. Throughout the history of medicine, up to the end of the nineteenth century, most medical diagnoses had to be invasive. In the past 50 years, however, non-invasive medical imaging systems have gained tremendous sophistication and effectiveness, largely due to the evolution of 2-D and 3-D digital signal-processing algorithms implemented on modern computers.

The main imaging techniques used on humans include X-rays images, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound imaging, Positron Emission Tomography,

Single-photon Emission Tomography, and Optical Coherence tomography. The first medical image was published by Wilhelm Roentgen in 1895, which was an X-ray (although not named X-ray at the time) radiograph from his wife's hand. A few months later, the first clinical use of X-rays occurred in 1896 ([60, 47]). For a long time, this type of radiography, called projection radiography, which mapped the 3-dimensional structure of the body to a 2-dimensional plane was used [47].

Finally, in 1972, this method was extended to XCT by Godfrey Hounsfield. He used mathematical methods for image reconstruction, which had been developed by Allan Cormack. Projection tomography and CT both use the transmission imaging modality, since they transmit electromagnetic waves through the body, which are then attenuated in intensity to form an image on an sensitive surface, capturing the internal structure of the body. Projection radiography mainly uses an X-ray tube, to generate an X-ray pulse that passes through the body. CT, unlike projection radiography, uses multiple projections of the same tissue that are collected from different angles by moving the X-ray source around the body. These projections are fed to a computer which produces an image of a single plane through the object with other planes eliminated ([60, 47]). This XCT technology was later adopted in industry for NDT and NDC purposes that are discussed in Section 2.1.2.

Magnetic Resonance Imaging (MRI) is and imaging method that is based on the phenomenon of nuclear magnetic resonance, first described by Felix Bloch and Edward Purcell. Application of MRI in medical imaging was first suggested in 1971 by Raymond Damadian. Magnetic resonance scanners work by exposing the body to a strong magnetic field. The vast number of hydrogen atom nuclei in the body are then aligned with the field. By partially exciting regions of the body, these alignments are disturbed, leading to the generation of electromagnetic signatures that can be detected by an antenna [47].

Our work in Chapter 5 falls in the category of non-invasive measurements of lower-back spine and muscles. We used stacks of images captured using MRI and applied a deep learning model to segment the various muscles in the images.

2.2 Deep Learning, Convolutional and Generative Adversarial Neural Networks

In this section we give an overview of the machine learning techniques used in this dissertation. Traditional Neural Networks (NNs) can be considered functions that transform an input vector into an output vector via several (hidden) layers. In each layer, a linear transformation is followed by a non-linear transformation called an activation function. Each cell in the hidden layers is called a neuron. The layers in a traditional neural network are also called fully connected layers. In a fully connected layer, each neuron in a layer is connected to all the neurons in the next layer. In other words, a single neuron in a fully connected layer is a linear combination of all the neurons in the previous layer that goes through a non-linear activation function. Figure 2.1-left demonstrates the architecture of traditional Neural Network models. In recent years, Convolutional Neural Networks (CNN) have gained wide acceptance especially in image processing applications. CNNs were originally inspired by the work of Hubel & Wiesel [34, 33] on cats' visual cortex. They discovered that many neurons react only to stimuli that come from a certain area of visual field and a particular geometric feature (vertical, horizontal, corner, etc.). LeCun et al. [39] had one of the first applications of CNNs, where they developed the now well-know LeNet-5 architecture to recognize handwritten digits. The main difference of CNNs from traditional Neural Networks is that they also include a *convolutional* layer in their architecture. The input to a convolutional layer is (in the context of image processing) usually a stack of 2D matrices (as opposed to a 1d vector in the case of traditional NNs). A neuron in a convolutional layer is connected to only a small, usually rectangular, area in the previous layer, called the receptive field of the neuron. A neuron in the output of a convolutional layer is a convolution (element-wise linear combination) of neurons in the input layer that are within its receptive field and a filter. The size of the filter is the same as the size of the receptive field and is a hyper-parameter of the layer, which is usually decided via experimentation. Figure 2.1-right illustrates the mechanism of a convolutional layer.

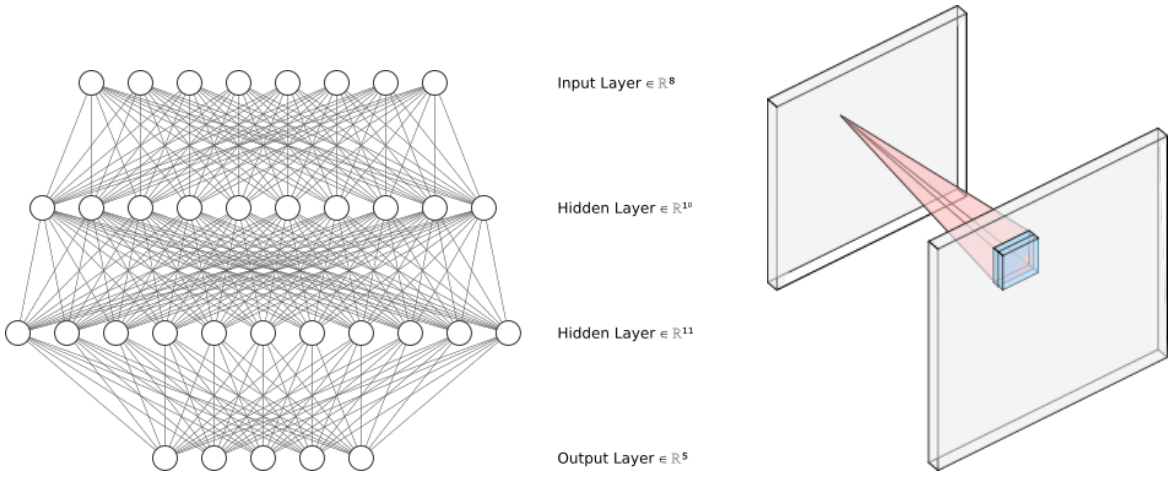


Figure 2.1: (left) A traditional neural network (right) A convolutional layer [41]

Neural Networks comprise one or more (hidden) layers between input and output variables. Deep Neural Networks (DNNs) are neural networks that have more that one hidden layer between the input and output layers [7]. At the time of this writing, the deepest model recorded was GoogleNet [74] with 22 hidden layers. DNNs, specially deep CNNs have had ample use in imaging processing. The volume of works has been so large that several survey papers have been published over the past few years on their architectures [46, 6, 3] and applications [46, 50]. In the rest of this section, we provide a brief introduction to convolutional neural networks. For further information regarding CNNs and deep neural networks please refer to [24] from which the following introduction is also adapted.

For the task of multi-class classification, the final layer usually goes through a SoftMax activation function which assigns a probability to each class. Rectified Linear Unit (ReLU) and its variations (leaky ReLU, ELU, etc.) are popular activation functions that are usually used for hidden layers. The idea behind these activation functions is to keep the neuron intact if its value is non-negative; otherwise, attenuate its negativity.

A neural net is trained by minimizing a loss function. A loss function is a differentiable function that measures the difference between the model output and observed data. For the task of classification, cross-entropy loss function is a common function to use. A common technique

used for minimizing loss functions is gradient descent, which iteratively modifies parameters of a function so that it is reduced at each iteration. An iteration is sometimes called an epoch.

One of the benefits of introducing convolutional layers in neural networks is that it results in a significantly lower number of parameters required to train the model as compared with a fully connected model with the same number of layers, since each neuron is only connected to the neurons in its receptive field in the previous layer. The output of a convolutional layer can be composed of a stack of 2D matrices, called feature maps, that are each the result of the inputs convolved with a different filter. In this manner, several same-scale image features (vertical line, horizontal lines, corners, etc.) can be learnt at the same time. As the feature maps through the layers of a CNN, more complicated image features can be captured via the filters.

2.3 Generative Adversarial Neural Networks

In Chapter 3, we use Generative Adversarial Networks (GANs) to improve the resolution of XCT images. GANs were proposed by [25] in 2014. Conceptually, GANs consist of two different NNs, a *generator* and a *discriminator*. The generator, as the name implies, generates new artifacts such as images, videos, sound, etc. The discriminator is usually a binary classifier whose task is to detect whether the artifact was generated (label = 1) by the generator or is a real artifact (label = 0). These two networks compete with each other in a way that improves both. Figure 2.2 demonstrates how a typical GAN is designed. Training GANs is usually done in iterations in which:

- first the generator produces some images,
- *only* the discriminator is trained using the generated images and (preferably) with label 1 and equal number of real images with label 0,
- then the generator is trained such that the discriminator yields 1 for all generated images. In this stage, *only the generator* is trained and the discriminator is only used to obtain labels.

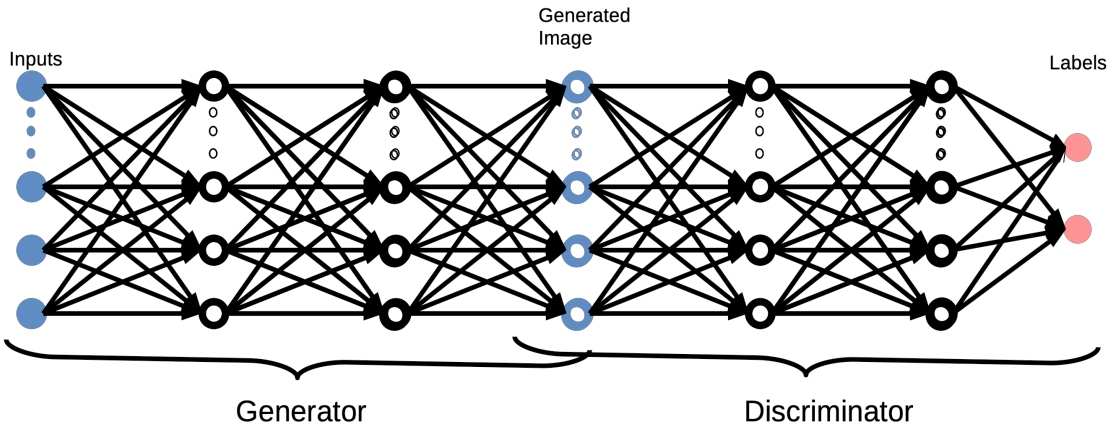


Figure 2.2: A conceptual GAN design

Note that at no stage are the generator and discriminator trained simultaneously. In fact, the generator never directly sees any real image. The only information available to the generator is the loss that is back-propagated through the discriminator.

If trained properly, the generator will be able to generate new data that looks like real data. Training GANs, however, has proved not to be easy. There has been, and still is, works to understand the dynamics of GANs [61, 37, 38]. For example, [61] proposed Deep Convolutional GANs (DCGANs) and provided guidelines on the best practices to train a more stable GAN.

2.3.1 GAN-based Super Resolution

As is discussed in Section 3.2.2, we'll use a GAN model proposed in [68], called SRGAN to improve the resolution of XCT scans taken from AM specimens. This model is a *Super Resolution* model. Super-resolution (SR) is a machine learning technique that refers to the task of estimating a high-resolution version of an image from a lower-resolution version [40]. SR has a wide variety of applications including improving the resolution of satellite, medical, facial, and text images [56]. In recent years there have been advancements in using Generative Adversarial Networks (GANs) for SR. For example, [15], designed a GAN-based deep architecture to improve the resolution of MR images, while [49] developed a GAN model, named GAN-CIRCLE (Constrained

by the Identical, Residual, Cycle Learning Ensemble), to improve the quality of both CT and MR images. A model was developed by [86] to improve the resolution of satellite images. The reader is encouraged to refer to the several survey papers have been published that deal with the topic of GAN-based SR[85, 4, 44] for a much larger list of applications.

Chapter 3

Porosity analysis and resolution improvement of XCT images of additively manufactured metallic parts

Abstract

A major challenge for the continued adoption of metallic parts produced via additive manufacturing (AM) is the uncertainty in their micro-structural and mechanical properties. Fatigue failure resistance is one of these properties that is highly dependent on the microstructure of the parts, specifically defects such as pores and un-melted regions. Size, shape, and orientation of pores in a part play a role in their fatigue life. It is important, therefore, to characterize these features. Information regarding porosity of specimens can be obtained via XCT. However, obtaining high-resolution XCT images is costly and time-consuming. In this chapter, we use a deep Generative Adversarial Network (GAN) to improve the resolution of XCT images. We then present a framework for automatically extracting porosity information including size and shape from XCT images. We also propose using minimum-volume bounding ellipsoid to capture the spatial orientation of pores. A data set, consisting scans with the resolution of $2.9\mu m$, was obtained via XCT. For a small sub-volume of the part, a scan with the resolution of $0.7\mu m$ was also obtained. We then used the $0.7\mu m$ images to improve the resolution of $2.9\mu m$ image Using our porosity framework, we showed that improving resolution does not help much in finding undetected pores, but among the pores that were detected, no significant difference was found between pore features of the images with improved resolution and those of the high-resolution images. On the other hand, these features were significantly different in low-resolution images.

3.1 Introduction

Metallic parts produced via laser/powder-based additive manufacturing (AM) methods, such as laser-powder bed fusion (L-PBF), have appeal to industries such as aerospace, biomedical, nuclear, and more – as they can be customized for specialized applications (low production volumes). However, a major challenge against the continued adoption of AM by such industries is the surrounding uncertainty in micro-structural and mechanical properties of AM parts [43, 79]. Fatigue life is of major concern in such industries, where it is crucial for parts to have high resistance under dynamic loading for prolonged periods. Like the mechanical properties of any metallic part, fatigue failure depends on microstructure [65, 72]. The microstructure in AM parts is primarily driven by thermal history experienced during AM (and subsequent heat treatments, if any), i.e. melt pool temperature, thermal gradient, cooling rate, cyclic reheating.

Thermal history, in turn, depends on powder characteristics, and process parameters, such as laser power, laser speed, layer thickness, and more [9, 66]. These process-property relationships inherent to AM parts, however, are difficult to discern due to the rapidness, small-scale and machine-variation inherent to AM. Even under fixed, optimized process parameters, any change in build/design parameters, such as part size and build orientation, may cause variation in microstructural, and consequently, mechanical characteristics. The first step in finding the structure-property relationships between microstructure and mechanical properties of AM parts is to develop an understanding of the microstructure itself [16].

Fatigue failure starts from defects that serve as crack initiation sites. These defects are typically either in the form of pores [87] which are quasi-spherical in shape and occur due to entrapped gas from the AM process, or as un-melted regions, which are irregularly shaped, have low volume and appear due to lack of fusion between neighboring layers. Optimizing AM process parameters in order to minimize fatigue failure is an important goal for metallic parts that are to be used in critical applications common to the aerospace and biomedical industries. Unfortunately, direct measurement of porosity is complicated for two primary reasons. First, not all pores have the same

impact on fatigue resistance. The size, shape and orientation of each individual pore can contribute differently to crack initiation and thus to fatigue failure. Therefore, simple measures such as overall density do not serve as good predictors of fatigue resistance. It is necessary to develop more detailed pore characterization methods to accurately predict a part's resistance to fatigue. The second challenge in measuring porosity is that it is difficult to accurately take physical measurements of the pores in an AM part. Direct destructive experiments on representative samples can be used to achieve this goal; however, it is desirable to find a quicker means of measurement that is not destructive and more cost-efficient. Such non-destructive methods can also be incorporated into AM quality control.

The most commonly used destructive method to gather information regarding the microstructural formation of AM parts is microscopy, in which a micrograph of a cross-section of the part is analyzed. X-ray Computed Tomography (XCT) is one of the most common non-destructive methods. In this method, the part is placed between an X-ray source on one side and a screen sensitive to X-rays (e.g. silver halide) on the other side. As the X-ray travels through the part, it is attenuated based on its structure and casts a 2D image on the screen. By rotating the part in small angular increments, the 2D image captures more details of the part structure. Finally, the image is processed by algorithms that provide a 3D reconstruction of the part's internal structure. Generally, an X-ray-produced 3D reconstruction is presented as stacks of 2D gray-scale images where a more attenuated pixel is assigned a lower value than a less attenuated one. More attenuated pixels represent voids.

Spierings et al. [71] have made many insightful comparisons between the aforementioned methods. Microscopy provides highly detailed information about pore size, shape and distribution in a particular cross-section of the part [12, 88]. However, it fails to provide information about other sections of the part. Pores are not uniformly distributed in an AM part [80], therefore, results of microscopy for a particular cross-section cannot be generalized to other sections. On the other hand, XCT provides useful information about the distribution of the pores and can provide data descriptive of pore position and shape. However, XCT does not provide high-resolution images

to capture smaller pores that can be observed via microscopy. XCT images also suffer from the presence of artifacts such as Poisson noise and rings around a center of rotation [10].

Analysis of the effect of process parameters on the microstructure of AM parts using the aforementioned porosity measurement methods is becoming highly popular in the literature [65]. Ziolkowski et al. [90] used XCT to characterize porosity in 316L stainless steel produced by Selected Laser Melting (SLM). Parts were intentionally built to have large pores that can be captured via XCT. They reported aggregate pore characteristics in terms of their size (diameter and area) and shape (sphericity). Total porosity was calculated based on the ratio of the area of pores detected and the total area of the material. They also provided a comparison between XCT and microscopy data made from the same cross-section.

Wang et al. [84] used both XCT and the Archimedes Method to analyze AISI 304L stainless steel parts produced using laser-based directed energy deposition (DED). They compared the results from the two methods and provided discussion on the validity of each of them. The subjectivity of thresholding in the XCT method was discussed; meaning that each pixel (with values between 0 and 65,535) cannot be directly used to determine whether a pixel can indicate a pore or not. A threshold value can be chosen and all the pixels having intensities below such a value can be considered black and all the pixels higher than this value considered white. Using this method, all black pixels in the resulting thresholded image can be labeled as pores. Without further information, these threshold values are determined arbitrarily; however, Slotwinski et al. [70] used a microscopic cross-section as a basis of comparison and tried to manually change the threshold so that an acceptable thresholded image could be created and the same threshold was used for all other images. In one of the most recent efforts toward characterizing porosity of AM parts via X-ray CT data, Wang et al. [84] tried to characterize porosity by examining sphericity of the pores. They only managed to characterize pores that had been intentionally inserted during fabrication and thus were uncommonly large, while in practice, pores of such size rarely occur.

Thompson et al. [78] provided a thorough literature review on applications of XCT in AM, including porosity analysis. They mentioned the work of Leuders et al. [42] on relating porosity of

fatigue performance in SLM-processed Ti-6Al-4v by examining as-build and heat- treated (HIP) samples. They showed poor accuracy in failure prediction which they believed was caused by high defect density in SLM parts. Thompson et al. [78] also mentioned the work of Tammas-Williams et al. [75] in investigating the use of XCT to measure porosity distribution on fatigue performance in Ti–6Al–4VEBM. They concluded that although XCT cannot capture small pores, it is capable of capturing of larger pores that are more likely to contribute to fatigue failure. Their study also showed that XCT porosity measurement can be accurate down to a pore size of $\sim 5\mu m$.

Obtaining high-resolution XCT images is time-consuming and costly. The time a typical XCT session takes depends on the desired resolution and the size of the specimen. A high-resolution XCT will provide more detailed porosity information but getting a high-resolution XCT from a large part might take hours. On the other hand, one of the main objectives of AM is to be able to build ready-to-use as-built parts without the need for long post-processing or inspections. It is therefore desirable to strike a compromise between resolution and time.

In this chapter, we use XCT images of the internal structure of AM parts and provide a detailed characterization of the pores inside them. We discussed how generative adversarial networks can be used to improve resolution of images in 2.3. In this work, we first used a super-resolution (SR) model, namely SRGAN [40] to improve the resolution of CT images. Then, we use a novel method using minimum-volume bounding ellipsoid to capture size, location, and the 3D orientation of pores.

SRGAN uses GANs made up of deep Convolutional Neural Networks (CNNs) to train generators and discriminators. SR can significantly reduce the cost of XCT in terms of time and money since lower resolution images can be used to achieve the same porosity measurements as higher resolution ones. These two contributions together will help in identifying individual pores that have the potential to serve as crack initiation points.

Our results showed slight improvement in the number of pores detected as compared to the benchmark high-resolution images. More specifically, among the 38 pores detected in the high-resolution image, only 14 were detected in the low-resolution image. This number increased to

18 when the resolution was improved. However, the main contribution of SR was that the pores detected in images with improved resolution were much closer in features to the high-resolution pores compared to low-resolution images. More specifically, features capturing pore area, volume, elongation, and sphericity were significantly different between the same pores in the low- and high-resolution images while they were not significantly different (with α=0.05) when the resolution was improved.

3.2 Materials and Methods:

3.2.1 Data

In this study we use a dataset of XCT data obtained from a Laser Power Bed Fusion (L-PBF) AM part to develop a novel means for measuring internal porosity. In L-PBF, thin layers of powder are applied to a build surface and melted and merged by a laser beam. The dataset was scanned by X-ray CT at the resolution of 2.9 μm with partial scanning of 0.7 μm for a small region of interest inside the specimen using the powerful Zeiss® Xradia 520 Versa imaging system. We refer to the $2.9\mu m$ resolution image stack as sample D1 and the $0.7\mu m$ image stack as sample D2.

A framework for extracting porosity information from XCT images and characterizing each individual pore using several measurements is proposed. The framework can be applied to either the raw XCT image data, or after the Super-Resolution algorithm has been applied to the XCT data. A three-dimensional (3D) reconstruction of the specimen is rendered in order to visually characterize part porosity. Geometric features of individual pores including size, sphericity and spatial orientation, are extracted in order to quantify the likelihood of them serving as crack-initiation sites using existing fatigue theory. The methodology is designed to be general enough to work with a wide array of data sources. As long as the input data is a stack of 2D images, each representing a thin layer of the part, regardless of what source it comes from, the framework can be applied to it.

The material selected for analysis was Ti-6Al-4V produced via L-PBF. The L-PBF system used herein was the EOS M290 and default process parameters of 280 W laser power, 1200 mm/s

scan speed and 140μm hatch distance were employed. Virgin powder (LPW) with a size distribution of 15-45 μm was used. A *Dog-bone* specimen was built in the vertical orientation consisting of a gage section. The gage section of the specimen is 2mm × 2mm × 2mm.

3.2.2 Super-resolution Model

In this chapter, we are using the SRGAN architecture presented in [40]. This network consists of a generator network (G) and a discriminator network (D). The inputs to the generative network are the lower-resolution images (I^{LR}) and the outputs are generated high-resolution images ($G\left(I^{LR}\right)$). The inputs to the discriminator are a set of generated high-resolution images with label 0 mixed with original high-resolution images with label 1. The task of the discriminator, as the name denotes, is to determine whether the inputs are generated or original. The task of the generator network on the other hand, is to deceive the discriminator network to mistake generated images for original ones. This is the general design of generative adversarial networks proposed in Goodfellow et al [25]]. The discriminator uses a simple binary cross-entropy loss function to detect generated or original images. This is called adversarial loss. The loss function of the generative network is more involved. Traditionally, a simple Mean Squared Error (MSE) loss function would have been used to measure the difference between $G(I^{LR})$ and the high-resolution image (I^{HR}). The authors of [40] improved this loss by introducing VGG loss. VGG is a deep network with an architecture proposed in [69] that is already trained to classify images. VGG loss is defined as the MSE between feature representations of $G\left(I^{LR}\right)$ and I^{HR} via a pre-trained VGG network as presented in [69]. The feature representations are the values of the hidden layers as an image is fed forward in a network. This is illustrated in Figure 3.1. In a traditional MSE loss, only the Euclidean distance between inputs (marked in the orange box) is used as a loss function. In a VGG loss, the Euclidean distance between (some of) the feature maps are also added (marked with the red box).

We used the SRGAN architectures within the default hyper-parameters for both the generative and adversarial models.

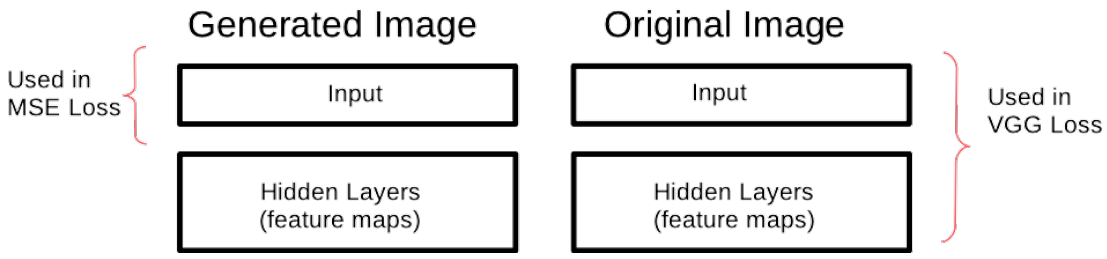


Figure 3.1: VGG loss

Our data consists of a stack of 422 pairs of low- and high-resolution images (respectively corresponding to D1 with the resolution of $2.9\mu m$ and D2 with resolution of $0.7\mu m$ mentioned above). Of those images, 253 (60%) were picked randomly for training and 42 images were used for validation. The remaining 127 images were used for testing. High- vs low-res image pairs were made by putting together the high-res images and the part of low-res images that correspond to the same sub-volume. Figure 3.2 illustrates the relationship between low- and high-resolution images. These pairs were then used to train the model (for pairs in the training set), to validate the model (pairs in the validation set), and to evaluate its performance (pairs in the testing set).

The model was developed using TensorFlow [2]. Training was performed on a Tesla K80 GPU with 32GB memory. The training process took approximately 10 hours before the early stopping criteria was activated. The early stopping criteria activated when the loss function didn't improve on the validation data set for 30 consecutive epochs.

3.2.3 Pore Detection and Graphical Representation

For the image analysis, it was assumed that the voxels adjacent to each other belong to the same pore in the part. Hence, an important step was extracting sets of adjacent voxels. Two voxels $v_1 = [x_1, y_1, z_1]$ and $v_2 = [x_2, y_2, z_2]$ are adjacent if and only if $|x_1 - x_2| \leq 1$,$|(y_1 - y_2)| \leq 1$, $|(z_1 - z_2)| \leq 1$ and $(x_1 - x_2)(y_1 - y_2)(z_1 - z_2) = 0$. Also, let P_1 and P_2 denote two pores, if v_1 and v_2 are adjacent and $v_1 \in P_1$ and $v_2 \in P_2$, then $P_1 = P_2$.

The extracted pore locations provide a means to quantify how pores are scattered throughout a specimen. However, in order to investigate the effects of porosity on properties such as fatigue

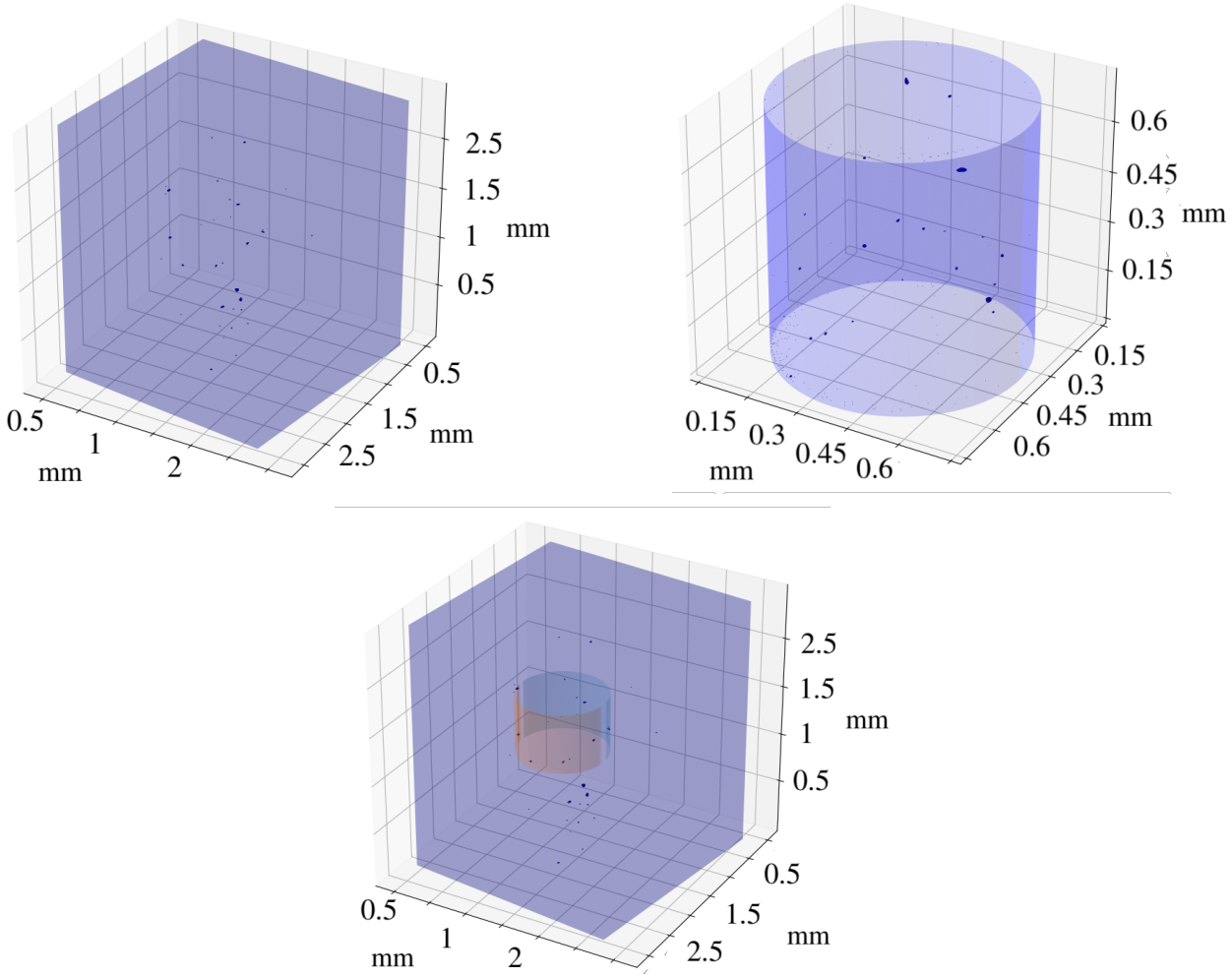


Figure 3.2: Relationship between low- and high-resolution images. The reconstructed 3d version of the low-resolution images is depicted on the top left corner while the high-resolution image is shown in the top right-corner. The bottom image shows how the high-resolution image only covers a small sub-volume of the low-resolution data.

failure resistance, more quantitative measures are required. Features such as porosity ratio, which is the ratio of the total volume of all pores to total volume of the specimen can be very helpful. In general, parts with higher porosity tend to have lower fatigue failure resistance ([81, 87]). Although porosity ratio is an important measure, it may not be a sufficient predictor of fatigue failure resistance since fatigue failures can start from a single pore and develop from there. Global information about porosity does not provide any insight into local features of particular pores. These features include position, size, shape and orientation of the pores. In order to obtain this valuable information, a means to mathematically characterize such features is needed. The rest of this section presents some features that can be extracted from detected pores and describes our procedures to extract them. The size of the pore in voxels is one of these features. This feature can be directly calculated by counting the voxels in a pore. The position of the pore can also be calculated by finding the geometric centroid of the pore.

Shape: Estimating the geometric shape of the pore can help with feature extraction such as area and perimeter. The minimum surrounding convex hull can act as a good geometric representation of the pore. The convex hull or convex envelope of a set of points X in a Euclidean space is the smallest convex set that contains X. The convex hull provides some insight into the shape of the pore; however, the convex hull representation will result in losing some information about pore shapes that are not convex. It is assumed that this loss of information is negligible with respect to the complexity that is posed by trying to serve non-convex representations of the pores. In addition, following the logic of the Murakami's approach [54], an irregular pore — if its size is critical — would quickly develop into a crack with convex shapes as the stress intensity factor along a crack edge tends to be constant

Convexity: Knowing whether a pore is convex is important in itself. It can be a measure of how accurate an estimation based on using a convex hull is for the pore. The convexity measure can be obtained by dividing the volume of the pore in voxels by the volume of the convex hull surrounding the pore.

$$\alpha = \frac{V_p}{V_h} \tag{3.1}$$

where V_p and V_h are volume as number of voxels and volume of the convex hull respectively. If a pore is actually convex then its volume should be almost equal to the volume of its convex hull, thus having a convexity close to 1. On the other hand, if the pore is non-convex, then the volume should be less than that of its convex hull, leading to a convexity smaller than 1.

Sphericity: Pores that are spherical are less likely to initiate cracks leading to failure due to fatigue rather than flat or long pores [66, 87]. Hence measures of sphericity and elongation are needed. The sphericity (ψ), a popular measure defined by Wadell [83], of a pore is the ratio of the surface area of a sphere (with the same volume as the given pore) to the surface area of the pore:

$$\psi = \frac{\pi^{\frac{1}{3}}(6V_p)^{\frac{2}{3}}}{A_p} \tag{3.2}$$

where V_p is volume of the pore and A_p is the surface area of the pore. The sphericity of a sphere is unity by definition and, a particle which is not a sphere has a sphericity less than 1.

Elongation: A simple measure of elongation is:

$$E = \frac{A_p^3}{V_p^2} \tag{3.3}$$

where V_p is volume of the pore and A_p is the surface area of the pore. A perfect sphere has elongation value of 36π. A very long shape with low width and height also has a very large elongation value since volume falls at a much greater rate than surface area.

Orientation and Surrounding Ellipsoid: The orientation of the pores is also of interest. Flat or long pores that are perpendicular to the direction of dynamic loading are more likely to cause fatigue failure than the pores that are perpendicular to the load. Extracting orientation of pores in 3D space can be achieved by fitting an ellipsoid on pore voxels. An ellipsoid can capture the orientation of the pore in 3D space by providing 3 perpendicular vectors that indicate extension of the ellipsoid in the space.

The general equation of an ellipsoid is:

$$(X - C)^T A (X - C) \tag{3.4}$$

where C is a vector of size 3 and denotes the center of the ellipsoid and A is a 3×3 matrix. The eigenvalues of A, (r_1, r_2, r_3) represent radii of the ellipsoid while the eigenvectors (v_1, v_2, v_3) of A represent the directions of the three radii in Cartesian coordinates. The radii values determine elongation of the fitted ellipsoid in the direction of the three axes in Cartesian coordinates. We used the algorithm discussed in Moshtagh and others [53] in order to extract minimum-volume enclosing ellipsoid of pores.

3.3 Results

3.3.1 Super-resolution

Figure 3.3-top demonstrates an example of model output for a layer in the test set. The left-most image is the original lower-resolution image ($2.9\mu m$ resolution). The middle image is the output of the model given the lower-resolution image. Finally, the right-most image is the original high-res image ($0.7\mu m$). Note that this particular high-res image was not used in training the model and was part of the test set. You can see how improving resolution reduces noise and increases the contrast of the pore, so it becomes detectable. This is corroborated by the bottom images which depict the detected pores from each image.

Porosity

The main goal of the task of super-resolution in this context was to increase the contrast of the pores so that automatic pore detection will detect the pores more easily and accurately. As a result, in order to evaluate the performance of the model, we compared the porosity detection between the three cases: low-resolution, improved resolution, and high-resolution. Several different metrics

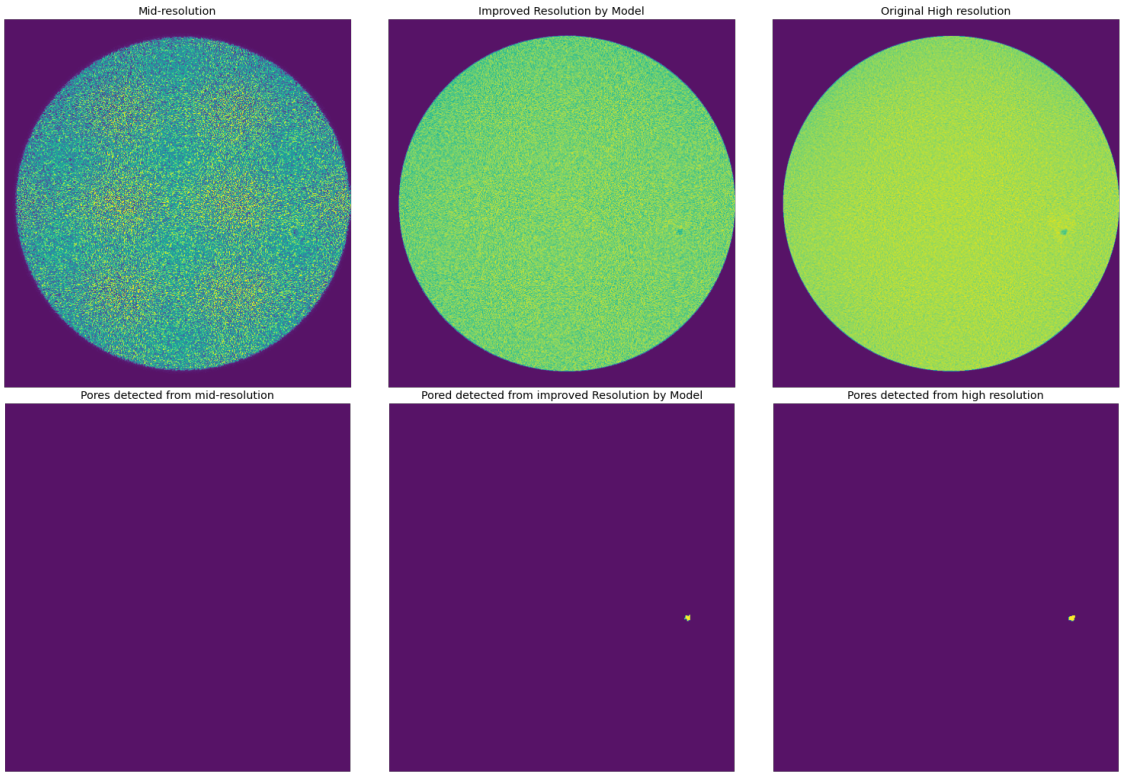


Figure 3.3: A comparison of pores detected from the same section in the three images (low-resolution, improved resolution, and high-resolution images)

were used in order to compare detected pores. In the rest of this subsection, we list these metrics and reports the performance of the model based on these metrics.

Visual comparison

Figure 3.3 illustrates the pores detected from the three images from the same section. Note how the pore is detected in the image with improved resolution (second row) while it remains undetected in the original low-resolution image. A closer look at the low-resolution image reveals that the pore is present there as well, but due to higher noise in surrounding areas, it was also dismissed as noise by the automatic method as well. In Figure 3.4, we have reconstructed 3D formation of pores for low-, improved- and high-resolution images. A visual study of the figure shows that the shapes and relative positions of pores in the improved resolution image are closer to the high-resolution image than the pores in the low-resolution image. This observation is quantitatively investigated in the following subsections.

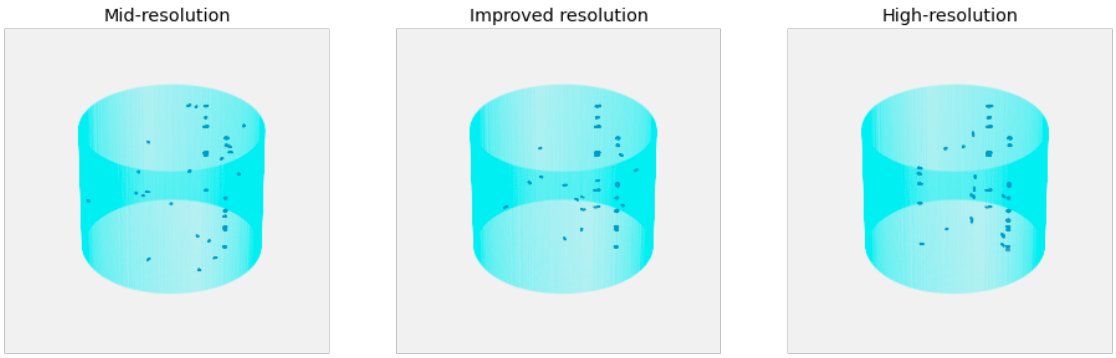


Figure 3.4: 3D formation of pores detected

	Description
True Positive	Pixel correctly classified as pore
True Negative	Pixel correctly classified as not pore
False Positive	Pixel incorrectly classified as pore
False Negative	Pixel incorrectly classified as not Pore

Table 3.1: Definitions of true positive, true negative, false positive and false negative in the context of pore pixel prediction

Pixel-wise pore detection

For a pixel, we denote true positive, true negative, false positive and false negative with TP , TN, FP, and TN respectively and classify them as listed in Table 3.1.

Note that we consider the high-resolution image as ground truth that provides exact information about pixels. Given the aforementioned terms, we then compute accuracy (A), precision (P), and recall (R) respectively as follows:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

$$P = \frac{TP}{TP + FP} \tag{3.6}$$

$$R = \frac{TP}{TP + FN} \tag{3.7}$$

	Accuracy	Precision	Recall
Low Resolution	$\sim$1	0.56	0.31
Improved Resolution	$\sim$1	0.77	0.53

Table 3.2: Pixel-wise detection metrics

In informal terms, accuracy is a measure of how many pixels were correctly predicted. Precision measures how much trust can we have that when the model says a pixel is a pore, it actually is a pore while recall means when the model says a pixel isn't a pore, how likely it is that this is the case. Table 3.2 lists performance metrics of the model in terms of pixels considered to be pores. We compared low-resolution and improved-resolution images in terms of accuracy, precision, and recall for evaluating the performance of our model.

The results of accuracy, precision, and recall are listed in Table 3.2. As can be seen, accuracy is very high. This is because pore pixels are much fewer than non-pore pixels and even if a model predicts all pixels as non-pore, it still acquires a very large accuracy. Therefore, accuracy cannot serve as a good metric for this particular task. Precision and recall, on the other hand, provide much better performance metrics. We can see that both measures improved as a result of resolution improvement procedure. Precision improved from 56% to 77% and recall improved from 31% to 53%. However, the relatively low values of precision and recall even for the improved-resolution image (77% and 53% respectively) show that resolution improvement did not result in an image that is exactly the same as the high-resolution image in terms of pore detection.

Number of pores detected

As is shown in Table 3.3, among the 38 pores detected in the high-resolution image (considered the ground truth), 14 were detected by the low-resolution image and 18 were detected by the improved-resolution image (after super-resolution was applied). This means that, although there is an improvement as a result of the resolution improvement procedure, still more than half of pores go undetected. However, this result is expected to some extent. If the pore is not at all present in the low-resolution image, there is little hope that it can be revealed due to resolution improvement. It

	Number of Pores Detected
High Resolution	38
Low Resolution	14
Improved Resolution	18

Table 3.3: Comparing images in terms of number of pores predicted

should be mentioned that not all pores are the same in terms of their contribution to fatigue failure. Small pores are less likely to serve as crack initiation points than large ones. On the other hand, fatigue failure is usually associated with larger pores. As a result, the points that are detected in the mid-resolution and improved-resolution images are much more likely to serves as crack initiation points than the ones that are missed. In other words, although improving resolution doesn't capture all the pores, it does capture the most important ones.

3.3.2 Pair-wise pore feature comparison:

We matched the pores detected from the low-resolution and improved-res images with the pores detected from the high-res image. These pairs of pores (14 pairs for the low-resolution image and 18 pairs for the high-res image) were then compared in terms of the pore features discussed in the previous section, treating the latter as the ground truth. Figure 3.5 illustrates an example how the improved resolution helps bring the pore features closer to the high-resolution image. In this figure, pore voxels are plotted along with the minimum-volume bounding ellipsoid. It can be seen that by improving resolution the pore shape was improved. In addition, the pore's spatial orientation became more aligned with the real (high-resolution image) pore as can be viewed from the axes of the bounding ellipsoid.

To measure the performance of the model, we take the high-resolution image as the ground truth and then compare the lower-resolution image with improved-resolution image (the output of the super-resolution procedure) in terms of how close they are to the ground truth. Following the same criteria, for each of lower- and improved-resolution images, we extracted pore features of those pores that were matched with the high-resolution images and conducted t-tests on the difference of the average of each performance metric between the lower- or improved resolution

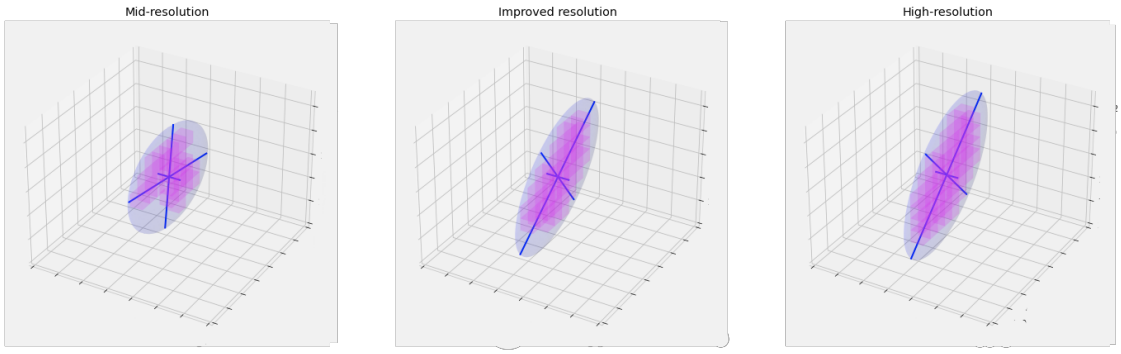


Figure 3.5: Visual comparison of a single matched pore.

Feature	Low-resolution		Improved resolution	
	mean	p-value	mean	p-value
Convex Hull Area	32.91	0.01	8.58	0.06
Convex Hull Volume	14.86	0.01	6.38	0.13
Ellipsoid Area	37.26	0.04	8.47	0.17
Ellipsoid Elongation	1.02	0.04	-0.05	0.82
Ellipsoid Volume	21.08	0.04	10.46	0.23
Pore Elongation	1.36	0.01	0.03	0.89
Pore Sphericity	-0.02	0.01	0	0.59
Pore Volume	38.14	0	12.72	0.03

Table 3.4: Matched pore feature comparisons between the low-resolution image and the improved-resolution image.

images versus the ground truth (high resolution images). The results of the tests are listed in Table 3.4. As can be viewed, p-values for the low-resolution pore features are all smaller than $\alpha = 0.05$, leading to a rejection of the null hypothesis that the features drawn from lower resolution images are statistically similar to the ground truth high resolution images. On the other hand, for the image with improved resolution, except for pore volume with p-value of 0.03, all other features have p-values larger that $\alpha = 0.05$, which means we can't reject the null hypothesis that they are similar. This means that if a pore is detected in an image with improved resolution, its features are much more similar to ground truth than in the original low-resolution image. In fact, the pores detected and characterized in the improved resolution images (with the super resolution algorithm applied) are statistically equivalent to the pores identified in the high-resolution images, while the pores detected and characterized in the lower resolution image are not.

3.4 Conclusion

In this chapter, we provide a method to improve the resolution of XCT porosity images captured from additively manufactured parts using deep generative adversarial networks. we also discussed several important metrics that capture 3D porosity in an AM part. In particular, we proposed using minimum area bounding ellipsoid to capture size, shape, and the 3D alignment of pores with respect to the specimen. We then used these metrics to compare the results for resolution improvement to the high-resolution baseline. We observed some improvement in terms of the number of pores detected. However, we noticed that among the pores that were detected, the images with improved resolution were much more similar in terms of porosity metrics to the high-resolution pores. The resolution-improvement model can be used to while dramatically reduce the time it takes to analyze specimens via XCT. The proposed metrics provide a tool to identify most important pores in terms of fatigue failure. The two methods together, provide a strong framework for non-destructive tests for metallic parts in additive manufacturing.

Chapter 4

An Automated Method for Geometrical Surface Characterization for Fatigue Analysis of Additive Manufactured Parts

Abstract

Recent interest to implement additive manufactured parts into structural applications has created a critical need to better understand the fatigue behavior of these parts. Alloys such as Ti-6Al-4V are popular in the aerospace and biomedical industries due to their superior strength to weight ratio and biocompatibility. Previous works have associated fatigue behavior with surface roughness parameters, especially radius of curvature of notches. It is therefore important to develop a fast, reliable and consistent methodology for extracting such parameters. In this paper, we provide an automated method for quickly extracting surface roughness parameters of the whole surface, including a novel method to extract radius of curvature of valleys using image processing techniques and optimization. We then use the extracted parameters in data-driven analysis. We investigate the effect of microscopy resolution and surface area sampling on the measured parameters. Finally, results on fatigue life prediction indicates that the automated method of calculating radii of curvature yields similar results as that of manually extracted radii while taking a significantly shorter time.

4.1 Introduction

Additive manufacturing methodologies have seen significant progress in recent years in production of net shapes for aerospace and biomedical applications. In particular, additive manufacturing has

gained a significant amount of interest in aerospace and biomedical sectors with Ti-6Al-4V alloy being prominently used in both these industries. Studies have shown that surface roughness can have significant adverse effects on fatigue life of laser powder bed fused (L-PBF) Ti-6Al-4V . Addressing undesirable surface roughness, however, remains a challenging problem in metallic additive manufacturing. Although laser-based powder bed fusion produces parts with relatively better surface roughness due to the small size of particles and thinner layer thickness [18, 27, 22, 31], as-built parts manufactured via this method still demonstrate earlier fatigue failure compared to machined parts. Fatigue resistance of Ti-6Al- 4V parts was improved almost to their fatigue endurance limits as high as 450 MPa [26] when they were highly machined and polished.

Therefore, it is necessary to garner a more comprehensive understanding of the surface roughness in order to characterize the fatigue life of as-built additively manufactured parts. In this work, we provide a comprehensive framework to extract surface roughness parameters obtained from microscopy. The framework includes preprocessing of surface height information obtained from microscopy, and efficient, automatic extraction of surface roughness parameters. The framework covers all traditional roughness profile parameters such as R_a/S_a, R_v/S_v, R_q/S_q, and R_p/S_p, as well as hybrid parameters consisting of multiple parameters such skewness (R_{sk}/S_{sk}), and the kurtosis (S_{ku}/R_{ku}) of the profiled lines. The main contribution of this framework in terms of surface roughness parameters is providing an automatic method to compute 2d and 3d radii of curvature for local minima (also called valleys or pits) as well as introduction of roughness mode as a method to correct for maximum profile valley depth (R_v). We used the framework to characterize the surface roughness of laser-based powder bed fusion Ti-6Al-4V specimens using images taken from a 3D optical microscope. We used statistical techniques to investigate the effect of microscopy resolution and surface area sampling on surface parameter measurements. We then used the radii of curvature to predict fatigue life for parts and compared them to a previous study in which radii were computed manually and showed that the results are statistically not different.

Description	1D	2D
Arithmetical mean height	R_a	S_a
Root mean square deviation	R_q	S_q
Maximum profile valley depth	R_v	S_v
Maximum profile peak height	R_p	S_p
Distance between the mean and the mode	R_{mode}	S_{mode}
Skewness	R_{sk}	S_{sk}
Kurtosis	R_{ku}	S_{ku}
Radius of Curvature of the valleys	ρ_p	ρ_s

Table 4.1: : Surface Roughness Parameters List

4.1.1 Surface roughness parameters

As mentioned earlier, in this work, we compute roughness parameters from an image taken via microscopy. These parameters are listed in Table 4.1. Each parameter listed in the table is computed via in both 1D and 2D dimensions. For 1D computations, each W×H 2D surface image is considered to consist of W linear roughness profiles of size H. Figure 4.1 illustrates the relationship between roughness image and roughness profile. Roughness parameters are then computed for each profile and the results reported either as a vector or described via a statistics such as mean or median. For example, 1D maximum valley depth,R_v, is computed by subtracting global minima from mean roughness of each linear profile. In such a way, we'll get a different R_v for each of the W linear profiles. As to 2D computations, The measurements are taken from the surface as a whole. For example, 2D maximum valley depth, S_v, is computed by subtracting global minima of the surface from mean roughness of the whole surface. Here, we'll have a single value for the whole surface.

The framework provided here computes all the parameters listed in Table 4.1 in an automated fashion. One of the main contributions of this work is that we use a novel method to compute radius of curvature for *all* 1D and 2d valleys in the profile. The following subsection discusses the importance of radii of curvature and discusses models that use it for fatigue life prediction.

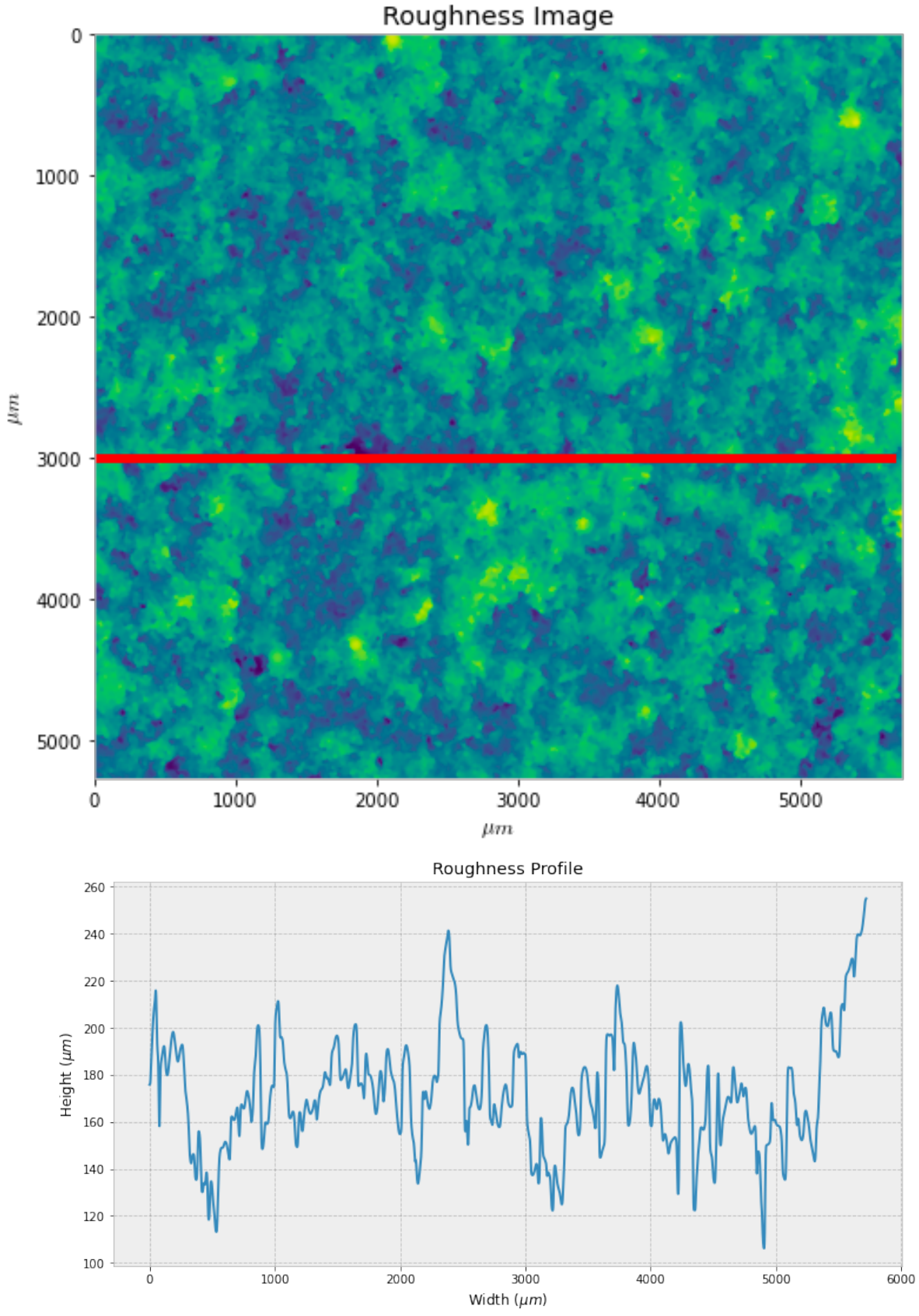


Figure 4.1: A surface roughness image(top) with a single profile selected via a red line. The roughness profile corresponding to the heigh of the surface on the red line is plotted in the bottom image

4.1.2 Radii of Curvature

Pegues. et. al. [59] proposed an approach for estimating the effect of surface roughness on the fatigue strength of additive manufactured Ti-6Al-4V using an approximation of the fatigue notch factor, which was previously used to incorporate roughness in fatigue life in [5]. The fatigue notch factor is expressed by the following equation:

$$\bar{K_f} = 1 + \frac{n\left(\frac{R_a}{\bar{\rho}_{10}}\right)\left(\frac{R_t}{R_{\text{zISO}}}\right)}{1 + \frac{\gamma}{\bar{\rho}_{10}}} \tag{4.1}$$

where n accounts for changes in stress state ($n = 1$ for shear and $n = 2$ for tension), γ is a material characteristic length [87] and R_a, R_t, R_{zISO}, and $\bar{\rho}_{10}$ are mean roughness, max peak-to-valley roughness, 10-point height roughness, and 10-point valley radii respectively. For the results in [58] $n = 2$ was used for the stress state and g was considered as the average acicular grain width of 1.5 μm. The surface roughness parameters are calculated using Equations 4.2-4.5:

$$R_a = \frac{1}{l}\int_0^l |y| \tag{4.2}$$

$$R_t = |y_{max} - y_{min}| \tag{4.3}$$

$$R_{\text{zIso}} = \frac{1}{5}\left(\sum_{i=1}^{5} y_{\text{i-max}} + \sum_{j=1}^{5} y_{\text{j-min}}\right) \tag{4.4}$$

$$\bar{\rho}_{10} = \frac{1}{5}\left(\sum_{j=1}^{5} \rho_{\text{j-min}}\right) \tag{4.5}$$

In Equation 4.5, $\rho_{\text{j-min}}$ are the radii of the five deepest valleys. In order to completely analyze the effect of roughness on fatigue, these radii should be calculated for all roughness profiles. A typical surface roughness image can consist of hundreds to thousands of such profiles, making the measurements of these radii extremely cumbersome, if not impossible. An automated method is

therefore desired not only to improve the time it takes to extract radii, but also to provide objective, reliable and consistent results that don't depend on individual judgements.

In this paper, we propose an objective, consistent, automated technique for extracting each radius of curvature of valleys as described in the proceeding section. The method significantly reduces the time it takes to find the radii, while providing results that are comparable to manually computed radii. The proposed technique is aimed at reducing the time to manually measure the surface characteristics while improving the repeatability of the calculations to obtain accurate and consistent roughness measurements. These results can then be used as inputs for models that predict mechanical properties. Using an objective, repeatable method to calculate the inputs allows researchers to compare different models fairly.

4.2 Material and methods

4.2.1 Data pre-processing

The images were converted from CSV format to a binary format. A pixel in a 500x image represents approximately a $0.429 \times 0.429 \mu m^2$ surface area. The images were appropriately downsampled so each pixel represents $1 \times 1 \mu m^2$. Further smoothing was done by applying a 2D Gaussian kernel of size 5×5 and a standard deviation of 1 to the images. Slight smoothing is necessary in order to remove raggedness that might be caused by noise. Sometimes the specimens may not be aligned exactly parallel to the image frame. Such misalignments were programmatically detected and fixed since it can significantly influence the SR parameters. Figure 4.2 shows misalignments correction for the image of an specimen. The alignment was performed by fitting a line to the upper edge of the specimen in the image. The upper edge is detected via adaptive thresholding which separates the specimen from the background. A rotation angle can be extracted from the slope of the line. Finally, the image is rotated by this angle in the opposite direction, resulting in an upper edge with angle 0.

In addition, as part of microscopy artifacts, the surface of the parts may not be flat. For example, it might be slightly tilted from one side to the other as illustrated in Figure 4.2. These tilts

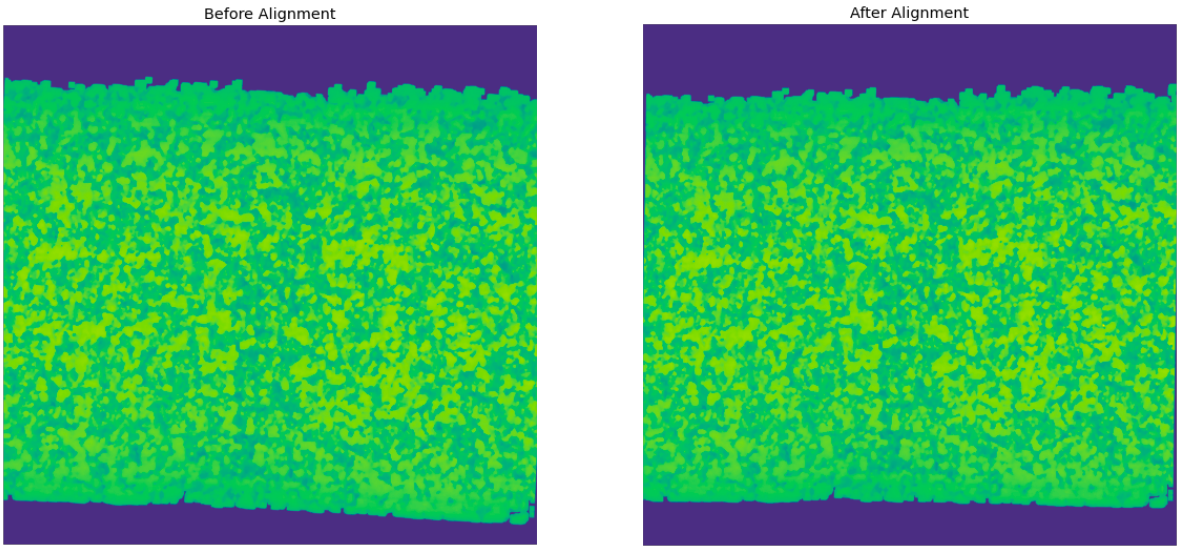


Figure 4.2: The descriptions of before and after the misalignment correction

were also programmatically detected and corrected. The procedure of tilt adjustment is performed by regressing a 2D plane to the surface of the image in the form of

$$aX^2 + bY^2 + cX + dY = Z \tag{4.6}$$

where X and Y are pixel coordinates and Z is the intensity of the pixel in the image. a, b, c, and d are parameters that are estimated by regression. Finally, a correction is made on the intensities subtracting estimated intensity from the actual intensity of each pixel:

$$Z^{'} = Z - \left(aX^2 + bY^2 + cX + dY\right) \tag{4.7}$$

Figure 4.3 contains 3D and 2D demonstrations of how tilt adjustment works. The top and bottom images of Figure 4.3(a) illustrate the same location with original and tilted views. Note how the coloring of minima and maxima changes before and after tilt. In order to verify the alternation of height numerically, the 2D profile of Figure 4.3(a) was generated as shown in Figure 4.3 (b). This figure illustrates the average pixel intensities over the x-axis in the image before and after tilting is performed. Note how the intensity profile is flattened after the process. In this study,

all height data acquired from images had passed through the above steps to ensure consistency in SR analysis.

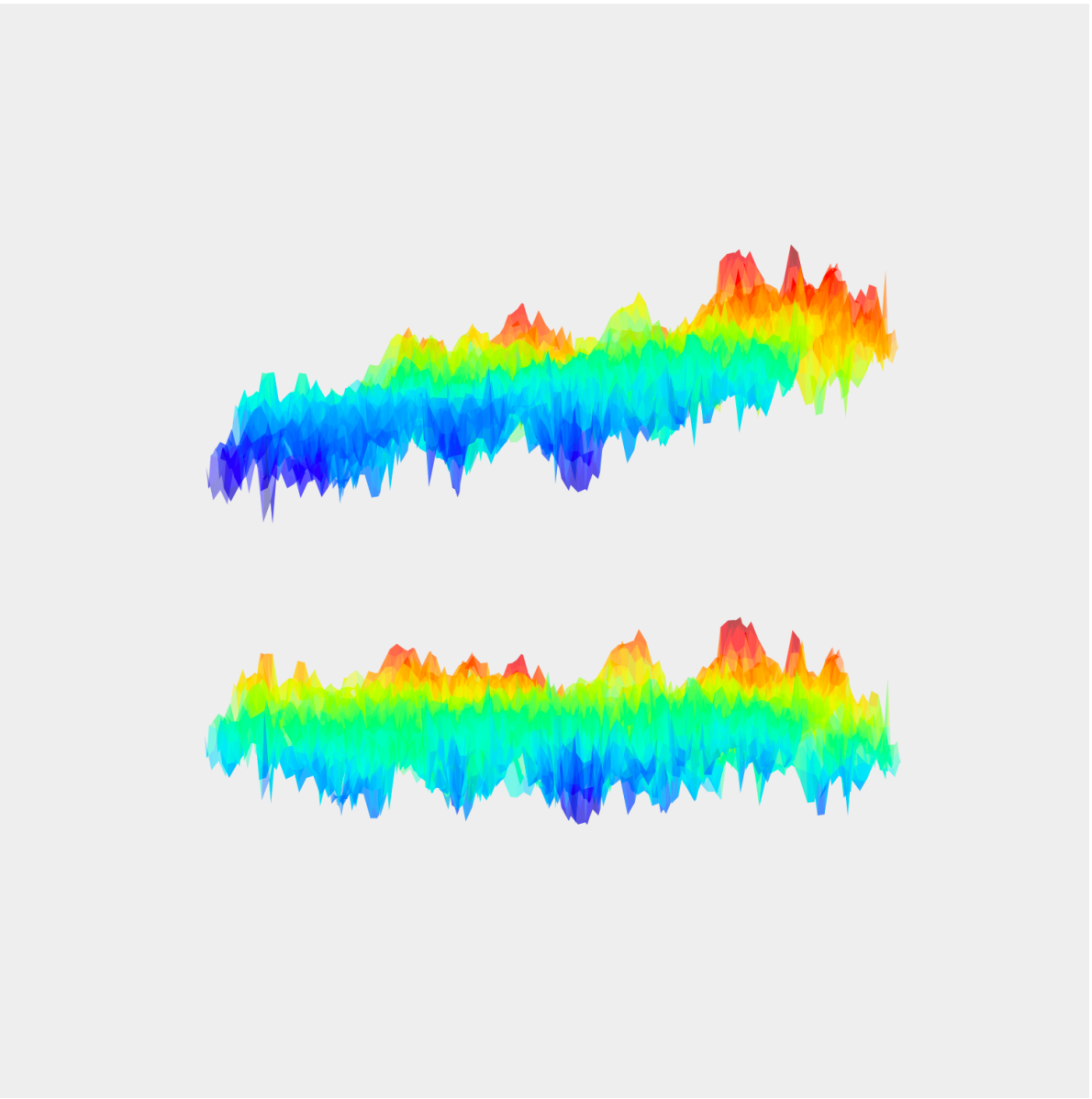
4.2.2 Radius of curvature of valleys

Given a surface roughness profile containing point height data, obtained e.g. through microscopy, our method extracts radii of curvature for valleys. This is done by approximating the radius of curvature at the base of a valley which can be considered a local minimum in the roughness profile. Radius of curvature (R) of a curve at point x can be analytically computed using the following equation:

$$R = \left| \frac{\left(1 + f^{'}(x)^2\right)^{\frac{3}{2}}}{f''(x)} \right| \tag{4.8}$$

Where $f^{'}(x)$ and $f''(x)$ are first and second derivatives of the curve at point x. Note that if x is a local minimum (or maximum), $f^{'}(x) = 0$ and the above formula turns into $R = \left|\frac{1}{f"(x)}\right|$. In a noisy roughness profile, computation of radius of curvature using f' and f'' values computed numerically is highly unstable and prone to numerical errors since a roughness profile is provide as a discrete set of data points and values of f' and f" can change significantly from one data point to and adjacent point. Therefore, there is a need for a more robust method that can be adapted to various roughness profiles.

The computation of valley radii is done in three steps: In the first step, profiles are smoothed and the coordinates of local minima (x_i, y_i) are extracted as valleys. A Gaussian kernel is used for smoothing. The size of the kernel is determined as a fraction (one fifth) of the average valley widths. The valley widths are determined as the distance between the beginning and the end of the valley that is below one standard deviation of the mean roughness. These local minima are then filtered such that only the lowest ones are chosen. The criterion for filtering is that a point is dropped if $y_i > R_a - \sigma_y$ where $\sigma_y = \sqrt{\frac{1}{l}\int_0^l (y - R_a)^2}$ is the standard deviation of the profile. We denote M as the set of all valleys that are selected in this step. This process is illustrated in Figure



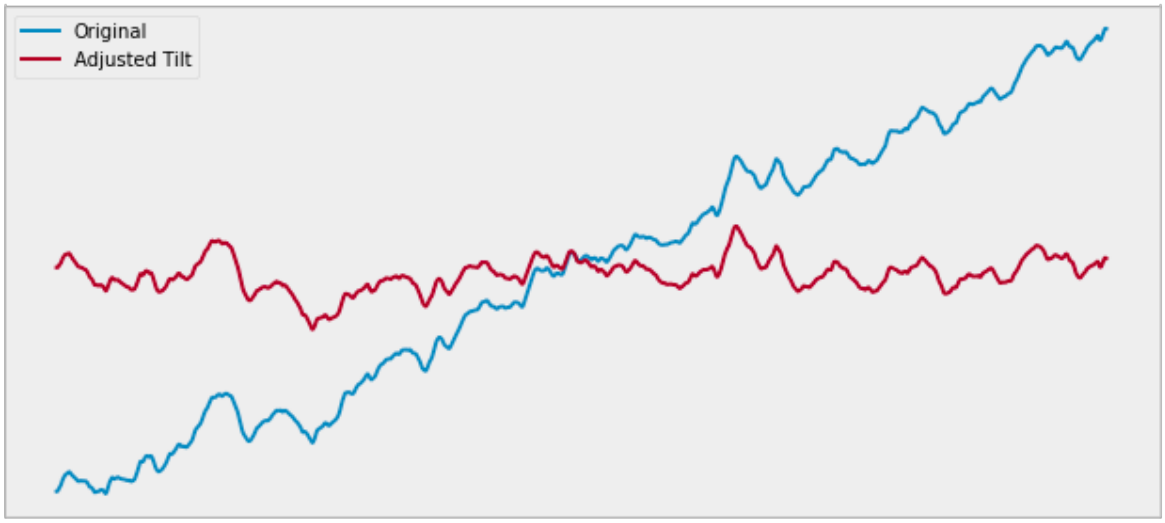


Figure 4.3: (a) 3D demonstration of tilt correction and (b) 2D profiled line before and after tilt/alignment correction with numerical height values.

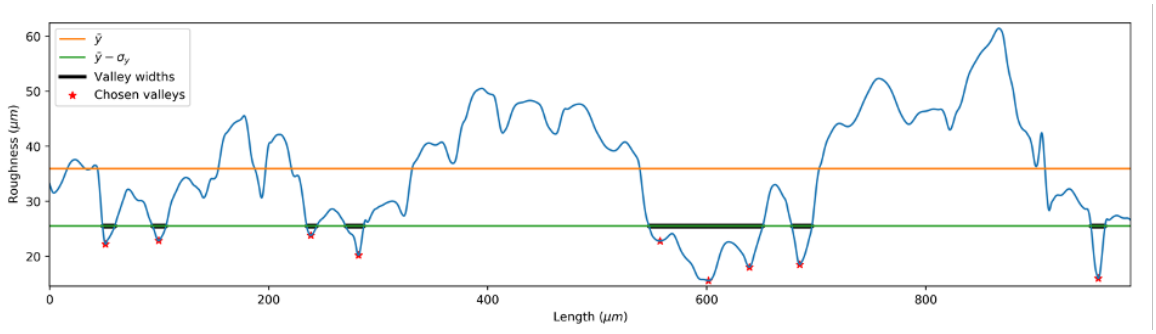


Figure 4.4: Valley local minima (valleys) that fall below one standard deviation of average roughness in a roughness profile.

4.4. In this figure, the valley widths are depicted using black lines. The Gaussian smoothing kernel and the neighborhood from which the local minima are extracted both use kernels sizes that are one fifth of average valley widths. We denote this kernel size with l.

In the second step, for each valley in M a neighborhood is chosen for fitting a circle. This is done by finding the nearest inflection points, that is, the point that the second derivative of the profile changes sign, on the two sides of the valley. This neighborhood ensures that a concave set of points are used to fit a circle. In order to guarantee that the neighborhood is symmetric relative to the local minima, the shorter distance from the two inflection points is chosen and mirrored to the other side of the valley (local minima). The neighborhood size normally does not exceed the size of the kernel used to find the local minima (l). In rare cases where this happens the neighborhood is shrunk to have size l ($\frac{l}{2}$ on either side of the local minimum). Figure 4.5 illustrates the process of choosing the neighborhood. We let N_i denote the points that comprise a neighborhood of the i-th valley in M.

Finally, for each valley i, a circle is fit to the points in N_i. In order to account for different scales and keep numerical stability, the points in N_i are first normalized into N_i'. A circle is then fit to the normalized points and the resulting circle is transformed back to an ellipse in the scale of data points. Finally, the radius of curvature of the data points is computed by finding the radius of curvature at the base of the transformed ellipse. The rest of this section provides more detail on this process and provides illustrated examples of results obtained by following this methodology.

Normalization of points in N_i is done using the following formulae:

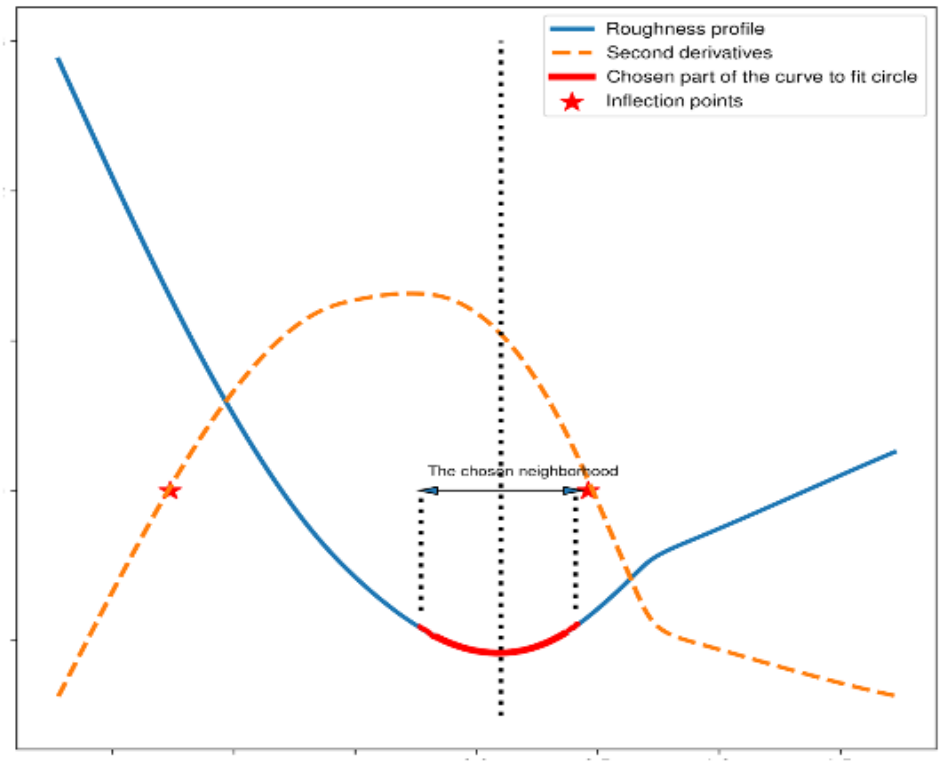


Figure 4.5: Process of choosing a neighborhood for a typical valley. In order to maintain symmetry, the closest inflection point to the local minimum is chosen (in this case the inflection point that is to the right) and mirrored on the other side of the minimum.

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_{x_i}}, y'_{ij} = \frac{y_{ij} - \bar{y}_i}{\sigma_{y_i}} \tag{4.9}$$

Where $\bar{x}_i$ and $\bar{y}_i$ are means of the coordinates of the points in N_i and σ_{x_i} and σ_{y_i} are standard deviations of the points. The circles are then found by minimizing an objective function $f(c_{xi}, c_{yi}, r_i)$ where (c_{xi}, c_{yi}) is the center of the circle and r_i is its radius. The function is expressed by the following equation:

$$f\left(c'_{xi}, c'_{yi}, r'_i\right) = \frac{1}{|N'_i|} \sum_{\left(x'_{ij}, y'_{Ij}\right) \in N_i} \sqrt{r'^2_{ij} - r'^2_i} + \lambda t_{ij} \tag{4.10}$$

where $r'^2_{ij} = \left(x'_{ij} - c'_{x_i}\right) + \left(y'_{ij} - c'_{y_i}\right)$ and λt_{ij} is a penalty term which forces the circle perimeter to align with curve points as opposed to having points inside the circle. t_{ij} is 0 if $r'_i < r'_{ij}$ and 1 otherwise and λ is a large (>500) penalty coefficient. Non-linear optimizers, such as gradient descent can be used in order to find optimal values of c'_{x_i},c'_{y_i}, and r'_i.

c'_{x_i},c'_{y_i}, and r'_i are fit to the scaled transformations of the points in N_i and need to be transformed back to the same scale as the points in N_i. This is done by performing the following transformation:

$$c_{x_i} = c'_{x_i} \times \sigma_{x_i} + \bar{x}_i \tag{4.11}$$

$$c_{y_i} = c'_{y_i} \times \sigma_{y_i} + \bar{y}_i \tag{4.12}$$

$$r_i = r'_i \times \frac{\sigma^2_{x_i}}{\sigma^2_{y_i}} \tag{4.13}$$

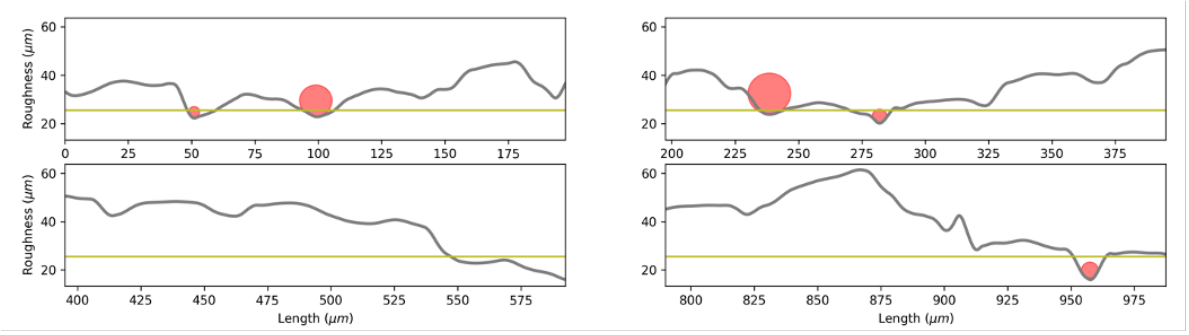


Figure 4.6: Valley radii for a roughness profile. The golden horizontal line separates the valleys that are one standard deviation below $\bar{y}$.

Note that the result of transforming the circle $\left(c'_{x_i}, c'_{y_i}, r'_i\right)$ back to scale is not a circle but an ellipse whose center is at (c_{x_i}, c_{y_i}) and its radii are $(r'_i\sigma_{x_i}, r'_i\sigma_{y_i})$. The radius of curvature at the base of an ellipse with radii (r_1, r_2) is $\frac{r_1^2}{r_2}$. Plugging in $r'_i\sigma_{x_i}$ and $r'_i\sigma_{y_i}$ results in equation 4.2.2.

Results of the automated extraction method for a single profile are illustrated in Figure 4.6, where four sections of a roughness profile are plotted along with the circles fit on the deepest valleys. It can be observed that the extraction method successfully locates the deepest valleys from each profile.

4.3 Experimental Results

4.3.1 Statistical Analysis of SR parameters

We conducted statistical analysis on roughness images taken from 17 specimens. Traditional parameters as well as mode, skewness, kurtosis and radius of curvature of valleys where computed. Both 2d and 3d scenarios were studied. In 2d measurements, the parameters were computed for each profile in the roughness image (~ 6000 profiles each side) of specimens. In the 3d case, the parameters were extracted from the whole surface. Using the extracted parameters, we conducted statistical analysis to test several hypotheses that are discussed in the rest of this section. In subsection 3.1, we discuss the effect of microscopy resolution on SR parameters. In subsection 3.2, we study the effect of surface subsampling. Finally, in subsection 3.4, we validate the results of our

Variation	$\rho^{10}_{\text{p-iso}}$	R_a	R_q	R_v	R_p	$\rho^{10}_{\text{s-iso}}$	s_a	s_q	s_v	s_p
p-value	0.00	0.02	0.02	0.00	0.00	0.0	0.02	0.02	0.00	0.00

Table 4.2: P-values for different metrics based on t-tests comparing 1000x and 500x resolution images

automatic method for calculating valley radius of curvature by comparing fatigue life predictions with a previous work.

4.3.2 Comparing different resolutions

The first question we wanted to answer was what microscopy resolution would be the best resolution to work at. Ideally, we would want to have images with highest resolution; however, increasing resolution, also increases (probably more than linearly) the time it takes to perform microscopy for a part. In the case of our analysis, we compared parameters from a 1000x vs a 500x resolutions for a single part.

Our results show that all parameters except for radii of curvature are relatively the same in both 1000x and 500x. This does not mean that the difference between parameters is not significant. In fact t tests for all parameters, as listed in Table 4.2, show significant differences in the parameters. This significant difference is because we considered a very large pool of profiles (~ 5000) in this study. However, this significance doesn't mean that the SR parameters are not close between the two resolutions. In fact, a close look at Figure 4.7 shows that R_a, R_z, and R_v values are less that $1\mu m$ different from each other. This difference for R_z is slightly higher and close to $3\mu m$. These differences are relatively small compared to the value of these SR parameters. What we mean by "relatively the same" is that the relative error between measurements is low. As to radii of curvature, the radii extracted from 1000x are lower ($5\mu m$ on average) that the 500x images. This can be because the valleys in the 500x image tend to be smoother due to the lower resolution, that is, 1000x captures some extra sharp valleys that the 500x image doesn't.

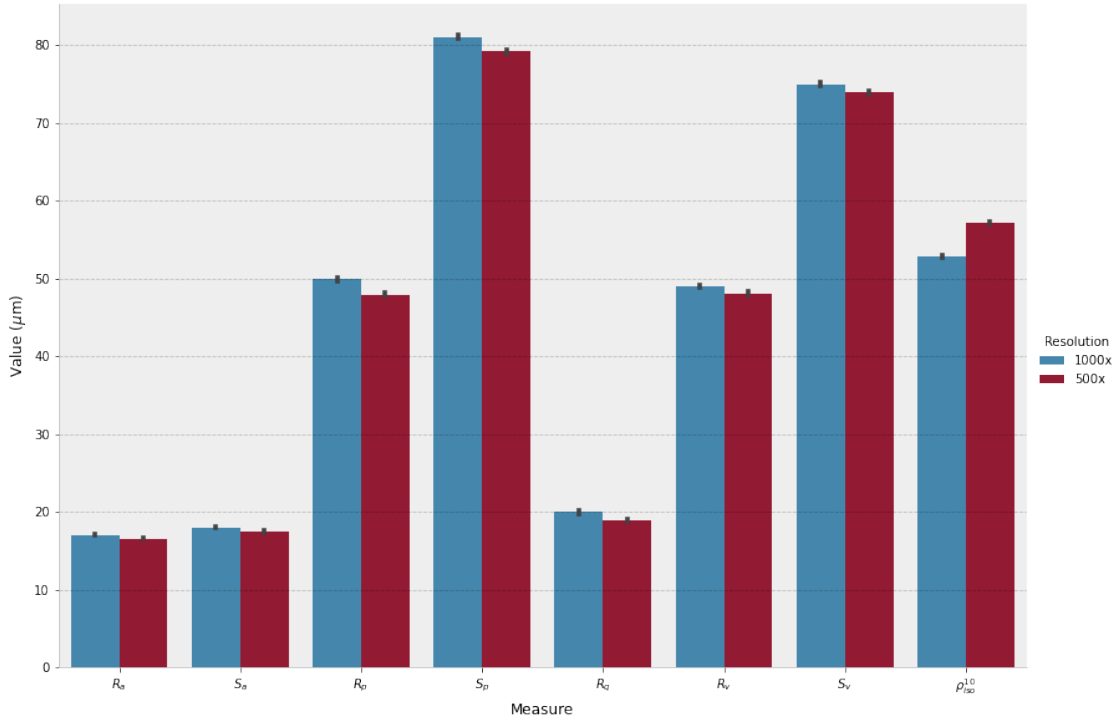


Figure 4.7: The bar plots of several SR parameters based on the height data measured by 500X and 1000X magnifications.

4.3.3 The effect of sub-areas

Sometimes, the entire area of the gage section of the specimen cannot be scanned due to the limitations the microscopy technology, time limitations, etc. Therefore, we decided to study the effect of scanning only a subsection of the specimen and generalizing the extracted SR parameters to the whole part. In order to examine the differences, statistical analysis regarding the effect of sub-area on resulting SR measurements was conducted. The original microscopy image was divided into multiple sub-areas as schematically described with different colors in Figure 4.8(a). SR parameters of each sub-area were investigated and plotted in corresponding colors as shown in Figure 4.8(b). In this image, the length of the colored bars are average measurements over all the parts. The black error bars on top of the color bars are 95% confidence intervals associated with each measurement.

As can be viewed in Figure 4.8 and is listed in Table 4.3, half-length measurements yield similar values of R_a, R_q, but provide lower values of R_v and R_p. This is because information

Variation/Metric	$\rho^{10}_{\text{p-iso}}$	R_a	R_q	R_v	R_p	$\rho^{10}_{\text{s-iso}}$	s_a	s_q	s_v	s_p
25% area vs Whole Surface	0.02	0.22	0.18	0.01	0.02	0.03	0.16	0.13	0.03	0.02
25% length vs Whole Surface	0.00	0.25	0.20	0.01	0.06	0.02	0.18	0.17	0.01	0.05
25% width vs Whole Surface	0.12	0.26	0.24	0.17	0.58	0.14	0.19	0.23	0.21	0.56
50% length vs Whole Surface	0.17	0.18	0.16	0.16	0.00	0.15	0.21	0.30	0.11	0.01
50% width vs Whole Surface	0.28	0.95	0.99	0.71	0.49	0.27	0.97	1.04	0.69	0.50
50% area vs Whole Surface	0.01	0.01	0.30	0.00	0.06	0.01	0.33	0.44	0.01	0.02

Table 4.3: : t-test p-values for difference in SR parameters between different surface subsample images and the whole surface

about the whole profile length is lost when the length is divided. As a result the left and right sides end up having different peak values. Overall, however, this effect is relatively consistent between specimens; that is, R_v and R_p remain in the same order as the original full-scale images. Therefore, it was expected that measured R_v and R_p values for the two halves to be slightly smaller that what they actually are on the whole surface. This will not affect relative SR comparisons since, if these values are smaller in part A than in part B, then actual values are also smaller in A than B. The conclusion here is that although cutting the length of the specimen in half and measuring SR parameters separately for the two halves affects the measured parameters, they remain relatively the same between parts so a comparison of SR parameters still can be made using half-lenths microscopic images.

4.3.4 Radii of curvature

We used two methods to validate the automatic method of computing radius of curvature. We used both synthetic data and actual roughness data for the purpose. The synthetic method is useful since the radii of curvature are known beforehand, so the the results can be easily validated. To test the results using actuall roughness data, we can only compare the effect they have on the fatigue life predictions that are computed via analytical models.

In order to test against synthetic data, the proposed method of fitting circles was tested on local minima of randomly generated polynomial curves of degrees 2 and 3. For each polynomial degree, 1000 expressions were generated using coefficients that were randomly generated from

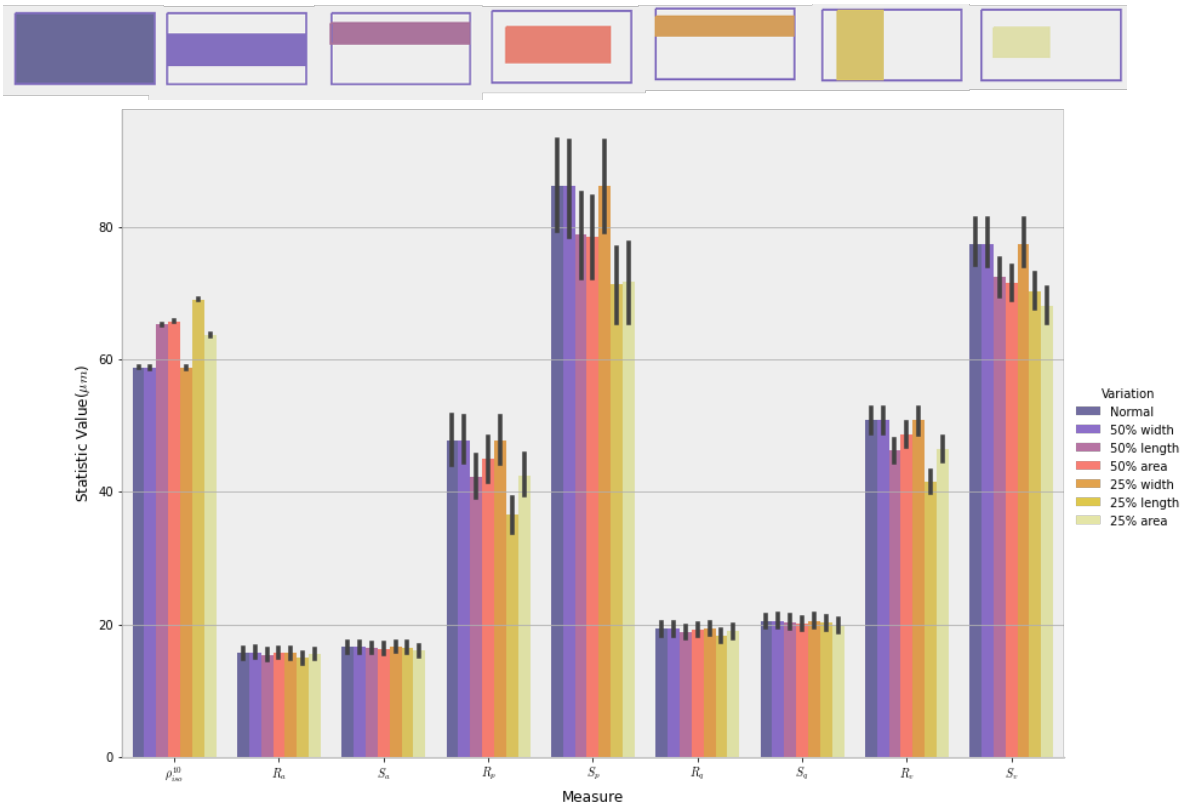


Figure 4.8: (a) Schematically described sub-areas with various colors. (b) SR parameters of each sub-areas with corresponded colors.

values between 0 and 100. For example, to generate a single random polynomial of the third degree, 4 coefficients a, b, c, and d were randomly generated and the expression $ax^3 + bx^2 + cx + d$ evaluated. The radius of circles extracted was then compared to analytically computed radius of curvature at the local minima that were computed using the formula $\frac{1}{f"^2(x)}$ where x is the location of the minima. The relative error was then measured. For polynomials of the third degree, the mean relative error was 0.0836 with standard deviation of 0.0893 and for polynomials of the second degree, mean relative error was 0.0166 with standard deviation of 0.0016. These results show that the fit circles are highly reliable measures of radius of curvature.

In order to validate using real roughness data, the fatigue life estimations by Pegues et al. [59] were reproduced using automatically extracted valley radii. In that paper, the authors provided fatigue life predictions of the endurance limit, using $\bar{K}_f$ (as defined in Equation 4.1), for two sets of roughness geometries, denoted AB1 and AB2. Using basic machine design principles, the authors then estimated the fatigue strength at 1000 cycles which has been traditionally used to signify the transition from low cycle to high cycle fatigue. A straight line fit in log-log coordinates is then constructed to estimate the stress-life curve. The important thing to note here is that only the smooth surface finish endurance limit, ultimate tensile strength, grain size, and roughness parameters are used to estimate the stress-life curves and that no fatigue data from the as-built surfaces are used in the predictions. Comparing the prediction curves with results from the destructive tests, it can be observed that the curve is slightly non-conservative. However, the predictions give reasonable results which, coupled with appropriate safety factors, may be useful in design given no available fatigue results are obtainable.

A comparison of the predictions from the automated method and predictions of [59], manual method are demonstrated in Figure 4.9. Graphs of the same color are predictions of fatigue life for parts with the same roughness geometries, with the dashed line corresponding to the manual method and solid lines corresponding to the automated method. Table 4.4 provides quantitative comparisons of the two methods in terms of mean squared errors (MSE) of the predicted fatigue life from the actual observations. As can be seen, the automated method can provide an accurate

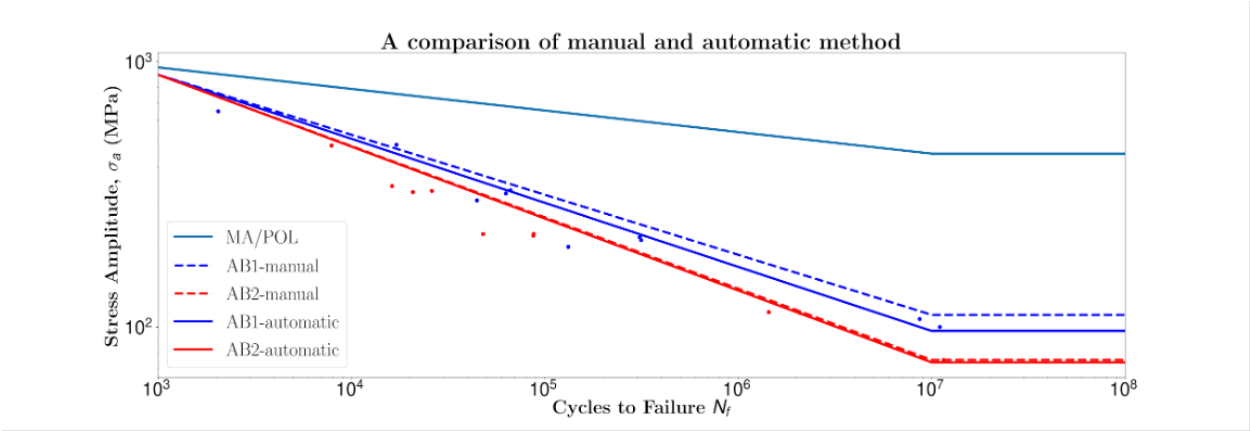


Figure 4.9: A comparison of fatigue life computations from Pegues et al. and the automated method proposed herein.

	Automatic Method	**Manual Method**
AB1 MSE	1883	3042
AB2 MSE	2547	2774

Table 4.4: A quantitative comparison of the manual and automatic methods of radii extraction using minimum squared errors

and reliable alternative for manually extracted radii in prediction of fatigue life. For both surface conditions, the automated method closely matches the results of the manual method. Furthermore, the automatic approach moves the predictive line closer to the fatigue data. The higher accuracy can be due to the fact that the automated method used a larger number of profiles to obtain valley radii. However, additional experiments are required to determine whether one method is consistently more accurate than the other. We can conclude that this approach can significantly reduce the time to obtain the radii measurement while also reducing the ambiguity inherent to manually measuring each radius.

4.4 Conclusions

In this study, a framework is developed to automatically compute surface roughness parameters from microscopy surface images. In particular, an automated approach is developed for predicting the effect of surface roughness on fatigue failure by analyzing surface roughness in metallic parts. Surface data is collected using digital focal variation microscopy and after adjustments for tilt

and misalignment, the roughness parameters are automatically extracted. The proposed extraction method can not only reduce the time costs required for manual measurements but, more importantly, also remove judgement element from the measurement, ensuring repeatability in the results despite the operator performing the experiments. The results of the framework are used in several data-intensive statistical analyses. In particular, the effect of microscopy resolution and surface sampling on SR parameters were studied. Finally, the effective stress concentration and fatigue notch factors for the AM surfaces were estimated using the model proposed in [58] and valley radii computed via the framework. Results show that fatigue life predictions using the automated method are comparable to results from manually extracted radii.

Chapter 5

Muscle Recognition in lower-back MRI Using CNN

Abstract

The purpose of this research is to automate the process of identifying lower back muscles in 2-D Magnetic Resonance Images (MRIs) and subsequently calculating size, location and effective lever arms using Convolutional Neural Networks (CNNs). Currently, lever arm calculations typically rely on human-delineated manual traces of the various low back muscles. Some of these muscles cannot be easily distinguished using MRIs and require trained analysts to reliably trace them. It is therefore possible for analysts to trace images differently. In this study, we use deep learning to trace these muscles. We expect this work will facilitate the study of lower back pain by leading to an automated tool that reduces tracing times significantly, thereby providing more objectivity and robustness to biomechanical calculations. Our model predicts muscles with an overall accuracy of over 99%, average precision of 80% and overall recall of 89%. This results in predicting lever-arms with an average absolute error of less than 8% (0.5 cm). It also reduces the time required to trace images by 98%.

5.1 Introduction

Low back pain (LBP) is one of the most prevalent musculoskeletal disorders (MSDs) in the U.S. and the rest of the world [32]. In fact, 31% of the overall incidence rate of non-fatal occupational injury and illnesses for all workers (public and private sector) is related to MSDs such as sprains

and strains [1]. According to Hales and Bernard [24], more than 80% of the occupational workers experience issues related to LBP during their lives. According to global burden disease study [82], low back pain is the leading contributing factor for overall years lived with disability (YLDs). The cost of low backpain treatment is significant. Dieleman et al. [20] estimated that total low back pain costs, including direct and indirect costs, was around $87.6 billion U.S. dollars from 1996 through 2013. According to the 2018 Liberty Mutual Workplace Safety Index, overexertion related injuries costs $13.67 billion dollars in direct costs alone [35].

LBP risk factors are multifactorial, and result from a combination of physical (biomechanical), psychosocial, and personal factors [17]. Deros et al [19] associated MSDs with manual material handling (MMH) activities. These types of occupational factors increase stresses on low back structures [17]. It is important to understand how a particular body part responds to external forces. If applied forces and moments exceed a structure's capacity, then pain or injury is more likely to occur [29].

A first step to addressing the LBP problem is to characterize the process leading to it. Biomechanical models have been developed to analyze internal reaction of the body against external forces during work [77]. Biomechanical models are often used in ergonomics to estimate forces and moments in various body parts during MMH tasks. Most often, the low back is the focus of studies because it bears significant loads and is frequently injured. Unfortunately, many biomechanical assessment tools share a fundamental model assumption, a simplified description of the lower lumbar musculoskeletal structures, failing to accommodate the large variations in these structures, such as the spinal geometry [76] and the paraspinal muscle mass and lever arm length [29].

McGill and Norman [52] had raised concerns about the accuracy of anatomical features used as a basis in biomechanical modeling. Hansen et al. [30] reviewed historical publications to identify anatomy of the back muscles how they are used in biomechanical modeling. The authors emphasized the reliability of collecting muscle lever (moment) arms and their respective cross-sectional areas and their importance for calculating forces and moments generated by muscles and borne by skeletal structures. Some studies used mean values for muscle lever arms, however; the

adequacy of such crudely estimated muscle lever arms may not be sufficient [30, 55]. Measuring lever arms of each muscle group involved in MMH tasks for each individual with high accuracy is the best way to accurately estimate the loading of intervertebral discs for those specific individuals. However, measuring all workers is not logistically practical. However, broadly sampling subjects to build regressions to estimate biomechanically relevant structures using gross anthropometric features (e.g., height, weight, sex, etc.) is feasible. However, comprehensively analyzing large numbers of subjects is time consuming. Automating the analysis of the most tedious aspects of these measures will save significant time. If this can be done without sacrificing too much accuracy, larger data sets with more varied subjects can be included in the dataset used for developing equations to estimate low back body parameters.

Medical imaging techniques can be used to accurately measure the lever arms that carry biomechanical forces for each individual. Imaging techniques including Magnetic Resonance Imaging (MRI), computed tomography (CT), and X-ray scans have been commonly used in previous studies to obtain more precise data about muscle geometry [28]. However, these images do not directly provide the lever arms of interest. The lever arms are typically calculated by measuring the distance of the muscles' centers of mass with respect to the spine. But, due to close proximity of the muscles and partial overlapping, it is not easy to differentiate some of the muscles in MRIs from each other and/or from neighboring tissues. Hence, an expert is required to trace the outlines of muscles (and the spine) on the MRI, to facilitate locating the centers of mass needed for estimating the lever arms. This is a very monotonous and time-consuming process which can be subject to human errors as well as subjective judgements by the tracer, particularly as the number of images to be traced increases.

In this study, we develop a Machine Learning approach that uses Convolutional Neural Networks (CNN) to segment and trace these muscles (and the spine) automatically on raw MRI images. The CNNs were used to detect muscles and compute mechanical attributes of the muscle automatically. This technique significantly reduces the time it takes to perform, musculoskeletal biomechanical measurements while improving the objectivity of the task.

Our method showed promising predictions for center of mass and lever arms. In our tests, for both cases, 90% of the predictions were within 15% average absolute error (1 cm) as compared to the manually traced muscles. A limitation of the CNN model is that segmented muscle areas tend to be larger than the traced version. This problem is corrected via post-processing. 90% of the postprocessed predicted muscle areas were within 20% of the traced muscle areas.

The remainder of this paper is organized as follows. Section 2 provides a brief introduction to CNNs and discusses previous work on biomedical image analysis. In section 3, we provide information on the methodology. We discuss how we preprocessed the raw MR images. CNN model architecture, hyperparameters, and training are then explored. Finally, the postprocessing technique is briefly discussed. In Section 4, we provide the results on the performance of our developed methodology based on various metrics. Finally, in Section 5, we provide a conclusion and paths for future improvements that can be applied to the method.

5.2 Recent applications of Deep CNNs in medical images

In the field of medical imaging, it is often desirable to automate the process of detection and segmentation of regions of interest in images obtained by various techniques such as XCT or MRI. Automation has the potential to reduce to cumbersome task of manually picking regions of interest in an image and reduce the analysis time dramatically. According to Litjens et al. [45], such automations started from simple rule-based expert systems from the 1970s to 1990s. Towards the end of the 1990s, the progresses in the field of supervised learning made it possible to conduct statistical classification based on some feature vectors. These feature vectors were picked manually depending on domain knowledge. More recently, deep learning approaches, such as Convolutional Neural Networks (CNNs) have taken this a step further and delegate the task of finding optimal features based on the data to the computer [45]. Litjens et al. [45] provides an extensive survey on applications of deep learning in the field of medical imaging. It categorizes papers based on type of images used (MRI, CT etc.), tasks (segmentation, detection, classification etc.) and application area (pathology, brain, cardiac, etc.). It also listed papers dealing with musculoskeletal images.

In a more recent work, Chea et. al. [14] have also conducted a survey on applications of deep learning in musculoskeletal radiology. The mainly categorized papers based on tasks (lesion detection, classification, segmentation, and non-interpretive tasks). Some of applications listed under segmentation task include knee bone and cartilage segmentation, segmentation of the knee for age assessment, and tissue segmentation.

There have been two general approaches to the segmentation problem in deep learning. The first approach uses a patch of surrounding pixels to classify a pixel. This approach can be used on scarce data since it divides each annotated image into hundreds or thousands of smaller patches that can be used in training [45]. The second approach uses an end-to-end architecture to output the segmented output directly. This approach, which was started by U-net [62], uses a much deeper network which includes a conventional convolutional layer followed by mirrored up-convolution (up-sampling) layers. Although this approach can be more accurate, it is much more difficult to train. A major challenge to the application of deep learning in medical imaging is that deep learning algorithms require large number of (mostly annotated) data points in order to be trained properly, while medical images are very cumbersome to annotate. To solve this issue, some works have used Transform Learning, in which some layers are borrowed from a model that is pre-trained on other images in other fields. As an example, Cai et al. [11] used Transformed Deep Convolution Networks to locate vertebrae. They trained their model on both MR and CT images with different resolutions to improve the model's discriminability. Their model's output rotated bounding boxes in order to locate each vertebra. Another work that tried to address limited training data was Zhang et al. [89], where two CNN models were developed. The first model was trained in an unsupervised manner on image patches in order to find inherent relationships and anatomical regions of interest. The second model adds extra layers to the first (already trained) one to predict the coordinates of anatomical regions of interest. Other works, however, claimed that models can be trained on limited data sets. For example, Forsberg et al. [21] used clinically annotated lumbar and cervical images in sagittal view to train a CNN that, in addition to labeling vertebra, locates centroids of vertebra and inter-vertebral disks as wells as their outer edges. Although the training images came

from archives of single institute, the model's sensitivity, precision, and accuracy were all above 99%.

There have been works that used CNNs to increase the resolution of musculoskeletal MRI [13]. Most research on medical images related to musculoskeletal disorders, however, tries to address the problem of locating and labeling vertebrae. For example, one of the early works that used CNNs to distinguish and locate vertebrae is Shen et al. [67]. They used two layers of convolution followed by a max pooling and two fully connected layers. On the other hand, in Suzani et al. [73], deep learning was used to localize and identify vertebrae at the same time. Their approach shows 100% vertebra identification and a mean surface error of below 2.8 mm for 3D segmentation and at the same time runs in a very short time. Lu et. al. [48] used a U-net based CNN for vertebral segmentation and disc-level localization and stenosis grading. However, none of these works tried to identify and measure all 11 muscles in a lower-back MRI.

In this paper, we take a different approach, from the aforementioned works. We use a stack of images in sagittal view of lower back to identify and classify different muscle groups. To the best of our knowledge, there has not been any model yet that achieved that goal. In addition to muscle segmentation, we used the results to automate the task for lever arm calculation. The main contribution of this work will be automating biomechanical measurements in general, leading to quicker, more robust and accurate analysis.

5.3 Data Collection and Preparation

5.3.1 Experimental procedures

This was a retrospective study using previously acquired images. As such, the study was ruled exempt by the local institutional review board and informed consent was waived. Furthermore, the consent given by the participants in the original study allowed further use of the image data by those investigators, who are among the authors of this paper.

The initial study from which we obtained the images had twenty-eight subjects (14 male and 14 female) who participated in the research. The individuals were restricted to only Auburn

University students. The inclusion criteria were that each individual met at the time were that each participant must:

- be between 19-30 years of age
- have no previous lower-back injury
- no current significant lower-back pain symptoms
- have a BMI of no more than 30
- met the MRI safety requirements protocol of Auburn University MRI Research Center (i.e., no metallic objects in their body)

In that original study, the principal investigator (PI) was experienced in operating MRI machines including a Level-3 MRI Safety Certification from the Auburn University MRI Research Center (AUMRIC). MRI scans were performed using a 70cm Open Bore 3T scanner (MAGNETOM Verio, Siemens AG, Erlangen, Germany).

5.3.2 Data

Figure 5.1 shows an example of MR images used in this paper. The data consists of a stack of 850 images from different cross-sections of the lower back from all the study participants. As in Figure 1, the data comes in pairs: the image on the left is the original raw MRI taken and the image on the right is the same image with the muscles of concern traced, that is, their edges have been delineated by an expert analyst using the Osirix© software (Rosset et al. [63]). It takes approximately 5 to 15 minutes for a trained analyst to trace a single MRI image on Osirix©.

Each pair of raw original and traced images can be denoted as (I_o, I_a), where I_o denotes the original image and I_a denotes annotated (traced) image. Each traced image in the training set was manually annotated by an expert. The raw image I_o (with some preprocessing) will be the input to the model and the traced image I_a (again, with some processing) will be the output. After an expert

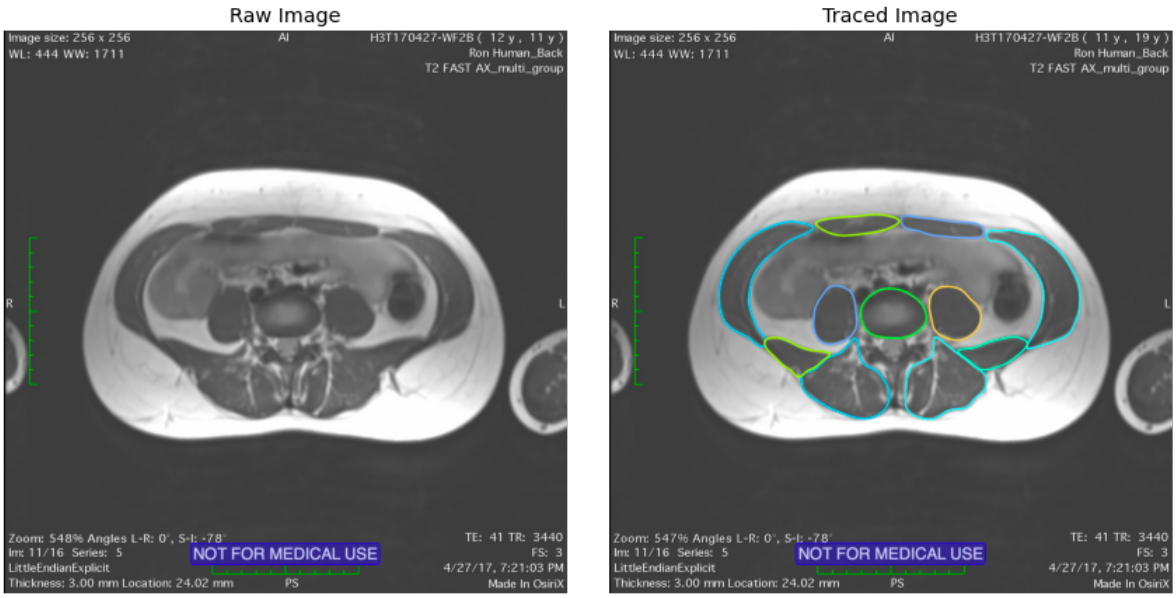


Figure 5.1: (left) A typical MR image. (right) The same image with the muscles of interest traced.

has annotated the image, they used the software to calculate the lever arms. Figure 5.2 illustrates the three main steps in the lever-arm calculation procedure.

The CNN model is designed to substitute the manual tracing process and then directly calculate the lever arms. From the stack of images, 269 images were used to train the model. This training set was divided into 215 (80%) images that were used for the actual training and 58 (20%) images that were used for validation. The remaining images were used as a test set.

5.3.3 Data Preprocessing

As a first step in preprocessing the images, the traced images I_a in each pair (I_o, I_a) must be further modified. Their internal pixels need to be annotated (colored) with their corresponding muscle class numbers. In Figure 5.2, muscle classes and the spine are annotated, and their lever arm distances computed. These muscles are listed in table 5.1.

Note that class numbers are chosen arbitrarily, and classes can be picked in any order. The images (both I_o and I_a) need to undergo further preprocessing before they are ready to be fed to the CNN models. The original images came in different size with height varying between 600 to 1200 pixels and width varying between 800 to 1400 pixels. As a result, in the first step, the images were

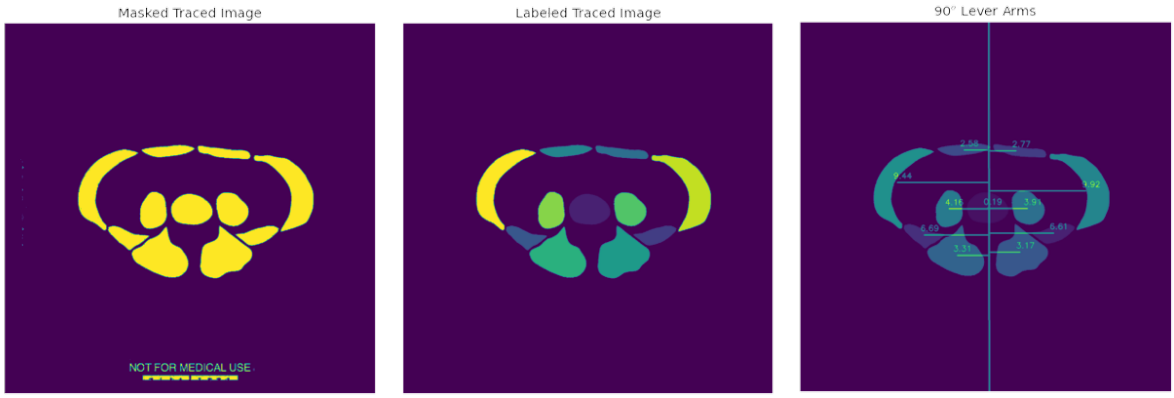


Figure 5.2: Muscle class extraction from annotated images. First, the manual traces are converted to binary masks representing muscle segmentation (top left). Second, the segments are assigned to a muscle based on their relative position (top right). Finally, muscle features such as shape and lever arms are computed from classified segments (bottom).

Muscle Name	**Class #**
Background	0
Spine (disc)	1
Right Quadratus	2
Left Quadratus	3
Right Abdominis	4
Left Abdominis	5
Left Erector Spinae Mass Muscles (ESMM)	6
Right Erector Spinae Mass Muscles (ESMM)	7
Left Psoas	8
Right Psoas	9
Left Oblique	10
Right Oblique	11

Table 5.1: Muscle Names and their corresponding class numbers used in prediction

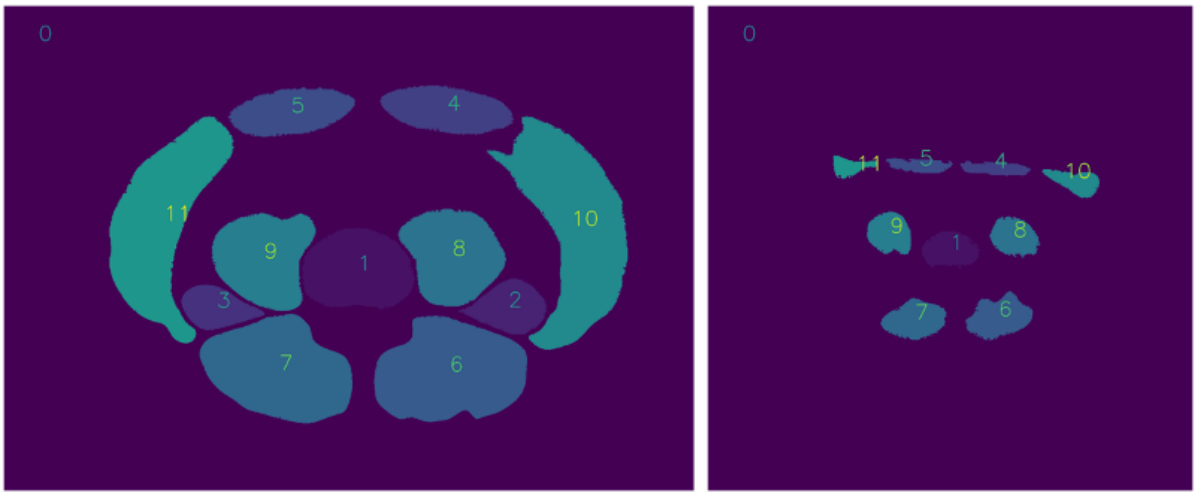


Figure 5.3: Muscle class Numbers overlayed on muscle segments. Generally, the data set consists or two different formation of the same muscles. The first one (left) consists of 10 muscles and the spine while the second one (right) consists of 8 muscles and the spine (left and right Quadratus are not present)

resized to a 256×256 pixel image. We chose this dimension since it strikes a balance between keeping the details of the image and training time. Doubling height and width of the images will increase the size of the training data by 4 times, which significantly increases training time.

As can be viewed in Figure 5.3, the data set contains two types of images taken from different spinal cross-sections. Note that the shared muscle numbers in the two images mean that they refer to the same muscle (for example, muscle number 11 corresponds to right oblique in both images). It is apparent how the sizes of muscles vary between the two images. In fact, there are ten muscles visible on the image on the left, while only eight are present on the right. This contributes to the complexity of muscle detection. Part of the value of our model is the fact that given enough training images from different cross-sections of the lower back, it can perform segmentation without knowing which cross-section the image comes from.

5.4 Model Design

In this section, we describe our proposed model architecture, choice of hyperparameters and training conditions. Finally, we discuss how we deals with data imbalance inherent in the problem.

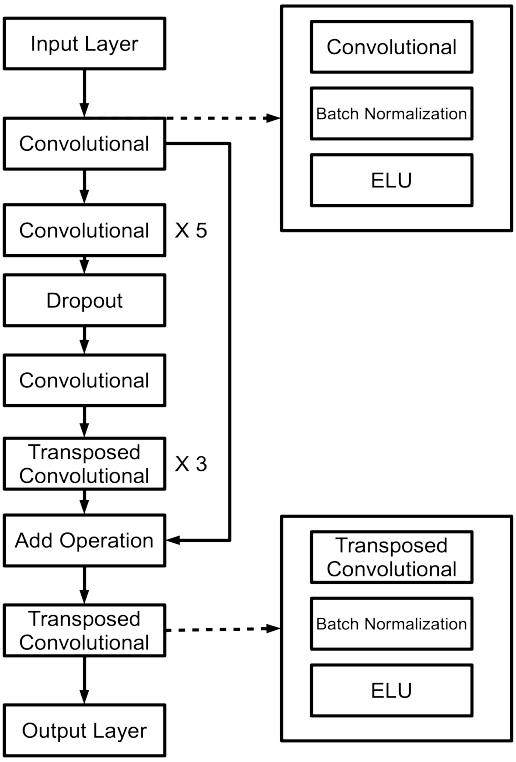


Figure 5.4: CNN Model Architecture

We developed a semantic segmentation model, that is, it assigns a class number to each pixel present in an input image. Figure 5.4 illustrates the architecture of the proposed model.

The model consists of seven convolutional layers (as in Figure 5.4, right) followed by four transposed convolutional (up-sampling) layers. Each of the aforementioned layers except the output layer are followed by batch normalization and an Exponential Linear Unit (ELU) activation function. Since this is a pixel classification task, the output layer has a SoftMax activation function. There is a dropout regularization after the sixth convolutional layer. All convolutional layers have a step of two, so the size of the image is halved as there is a skip layer where the result of the first convolutional layer is added to the result of the third up-sampling layer. This is done to provide finer-level information from the original image to the last up-sampling layer.

A dropout layer following each convolutional layer controls overfitting. A dropout rate of 25% was used in the fourth layer which means that in each iteration, 25% of the neurons are deactivated

Layer	Conv. 1	Conv. 2	Conv. 3	Conv. 4	Conv 5	Conv. 6	Conv. 7	Up. 1	Up. 2	Up. 3	Up. 4
# filters	16	32	64	128	128	256	12	16	16	16	12
Filter size	(3, 3)	(3, 3)	(3, 3)	(4, 4)	(4, 4)	(1, 1)	(1, 1)	(4, 4)	(4, 4)	(4, 4)	(4, 4)

Table 5.2: Hyperparameters used to train the CNN model. (Conv and Up are short forms of Convolutional Layer and Up-sampling layers respectively)

randomly, no matter what values they assume. This is a powerful technique to avoid overfitting since it prevents the model from relying too much on any particular neuron and distributes activations among all neurons. Before each convolutional layer, the inputs are zero-padded such that the size of its outputs remain the same as the size of its input. Without zero-padding, the input will shrink as it moves forward in the network. As mentioned above, the input image Io and the annotated image Ia are both 256×256 pixels.

5.4.1 Hyperparameters

The hyperparameters for the convolutional and up-sampling layers are listed in Table 5.2. Note that filters of size 1×1 are used for convolutional layers 6 and 7. In fact, these two layers could have been replaced by fully connected layers with 256 and 12 neurons respectively. Using convolutional layers will allow the model to process input images of any size. If only fully connected layer had been used, only inputs of size 256×256 could be used as inputs. This architecture is called a fully convolutional network, that is, it only contains convolutional layers. The final up-sampling layer has 12 filters, which equals the number of muscles plus the background. The background could be called class zero or no-muscle class.

5.4.2 Accounting for data imbalance

A particularly important challenge that had to be addressed for the task of muscle segmentation was that the majority of the pixels ($\sim 90\%$) were not associated with any muscles (class zero). This leads to a highly unbalanced dataset. Training a model on this data without addressing this issue will result in misclassifications where accuracy is high, but precision and recall are low (Refer to Section 4 for formal definitions of these terms). Two common approaches to this problem are

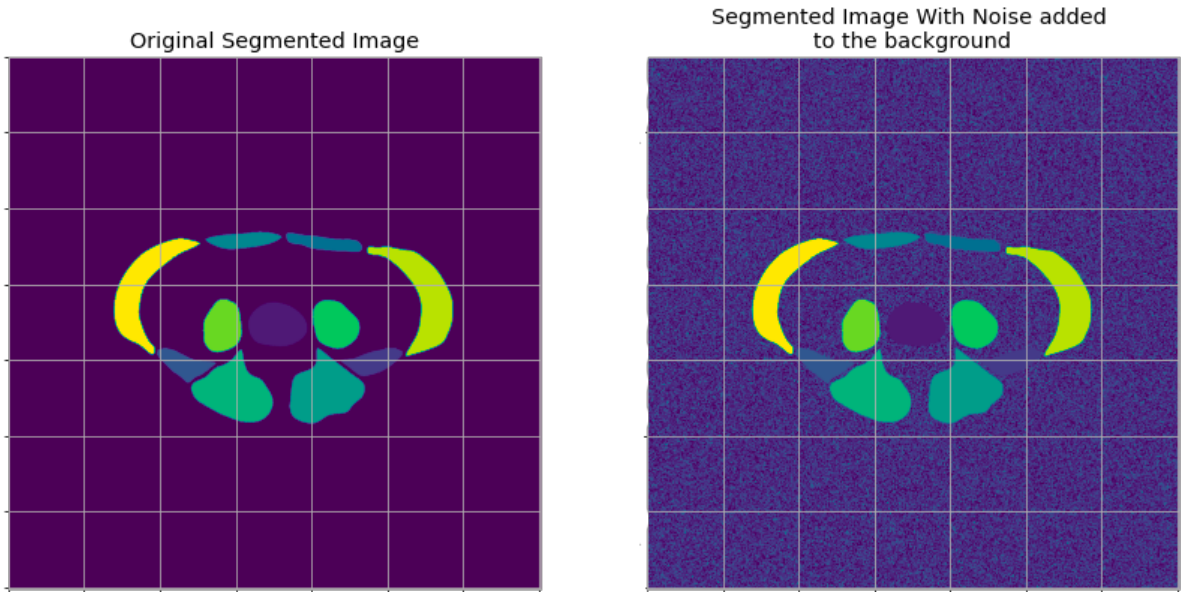


Figure 5.5: Adding noise to the background to avoid unbalanced classes.

subsampling and class weights. Subsampling means taking samples from the data in a biased way, such that class imbalance is reduced. This means picking fewer observations from the class with the majority of instances. This approach is not suitable when it comes to images since an image is a whole entity. Subsampling pixels in an image is simply impossible without changing the nature of an image. Alternatively, class weights could be used to adjust for this issue by providing less weight to the majority class. The loss function will then become a weighted sum of actual losses multiplied by class weights. This approach, although possible, is computationally expensive and requires developing a custom loss function.

The approach we used in this paper is simply replacing a portion of the no-muscle class with white noise. In particular, we replaced 80% of the background pixels with random values from 1 to 11 (corresponding to the 11 muscles). This will reduce the no-muscle class frequency from $\sim 90\%$ to $\sim 16\%$. The justification used here is that this noise does not provide any semantic information to the model. The model cannot learn it and simply ignores it and focuses more on the part of the image that actually demonstrates a pattern. This strategy is depicted in Figure 5.5.

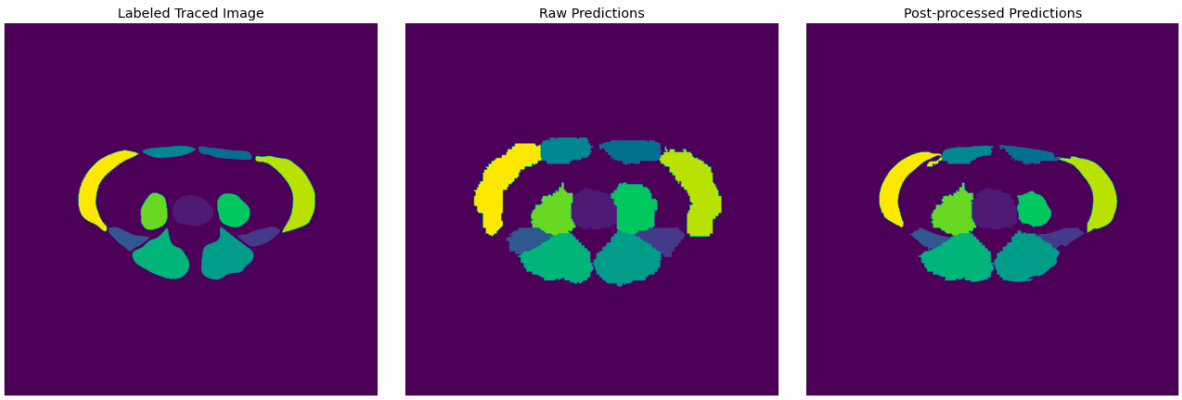


Figure 5.6: Original segmented image (left) together with raw (middle) and post-processed (right) segmentations.

5.4.3 Training

The model was developed using TensorFlow [2]. Training was performed on a Nvidia Tesla K80 GPU with 32GB memory. The training process took approximately 12 hours before the early stopping criteria was activated. The early stopping criteria activated when the loss function didn't improve on the validation data set for 30 consecutive epochs.

5.4.4 Post-processing

Preliminary experiments revealed that the model was susceptible to predicting muscles "fatter" than the actual muscles, that is the predicted muscle would cover, and exceed the boundaries of the ground-truth image. For this reason, we implemented a post-processing strategy that performs more fine-grained segmentation. The idea is that for each muscle, the post-processing procedure performs another segmentation on the segmented area which eliminates background pixels from the foreground. This is achieved via a direct implementation of the GrabCut foreground detection technique, from the image processing literature [64]. A sample comparison of the ground truth, raw prediction and post-processed prediction is given in Figure 5.6.

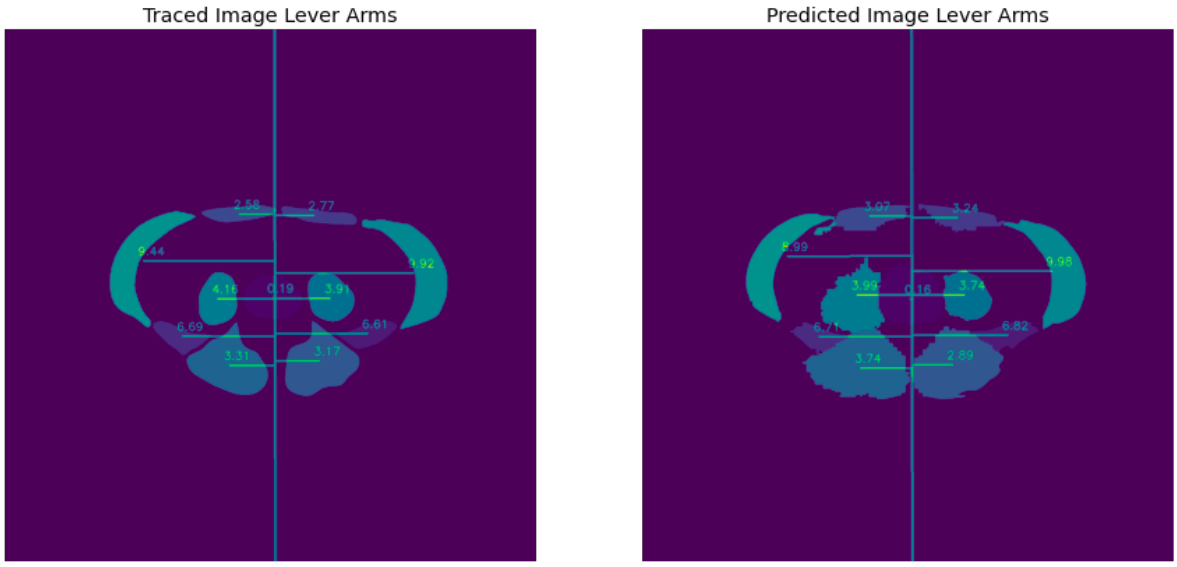


Figure 5.7: A comparison of 90° lever arm distances from original traced (left) and post-processed segmentations (right)

5.5 Results & discussion

We compared the model output with the original images in both the muscle prediction, as well as the prediction of lever arms. Figure 5.7 presents an example. We are interested in the error of the model both in terms of pixel-to-pixel prediction, as well as in the accuracy of the predicted outputs (muscle centroids and lever arms).

5.5.1 Performance Metrics

We used accuracy, precision, and recall to evaluate the performance of our model. For a muscle class i, we denote true positive, true negative, false positive and false negative with TP_i , TN_i, FP_i, and TN_i respectively and compute them as follows:

$$TP_i = n_{y_p=y_o=i} \tag{5.1}$$

$$TN_i = n_{y_p \neq i} \tag{5.2}$$

$$FP_i = n_{y_p=i} \tag{5.3}$$

$$FN_i = n_{y_p \neq i} \tag{5.4}$$

Where y_p and y_o denote predicted and observed (traced) classes and $n_{\text{condition}}$ is the number of pixels that satisfy a certain condition. Given the above definitions accuracy, precision, and recall for pixel i, denoted with A_i, P_i, and R_i respectively are computed as follows

$$A_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{5.5}$$

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{5.6}$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{5.7}$$

Our model achieved an overall accuracy on pixel classification of 95.33% on the training set, 93.1% on the validation set, and 92.5% on the test set. Accuracy is a good metric to measure the overall performance of a model. Its downside, however, is that it works best in balanced datasets. However, in unbalanced datasets, it becomes a poorer metric. Precision and recall can be used in order to provide better understanding of the model and what it captures or misses. Precision is a measure for the correctness of a positive prediction. Recall - or the true positive rate - is the measure for how many true positives get predicted out of all the positives in the dataset. In lay terminology, accuracy is a measure of how many pixels were correctly predicted. Precision measures the level of trust that when the model says a pixel belongs to class i, it actually belongs to that class. Recall means if a pixel is actually part of muscle i, how likely it is to be detected by the model.

muscle	1	2	3	4	5	6	7	8	9	10	11
Accuracy	99.75	99.86	99.87	99.68	99.71	99.62	99.56	99.83	99.80	99.45	99.67
Precision	81.88	71.94	70.73	62.00	65.29	89.27	89.76	87.41	85.01	88.36	84.38
Recall	98.00	89.15	88.23	85.10	88.40	89.17	86.12	93.19	92.73	84.99	93.06

Table 5.3: Accuracy, Precision, and Recall:

The performance metrics for each muscle in the model results are listed in Table 5.3. The model exhibits high values in these three metrics for most of the muscles. Accuracy for individual muscles is very high and close to 1, with a minimum of 99.45%. However, this can be attributed to the highly unbalanced nature of the classification problem. "Unbalanced" here means that most of the pixels in a typical image are background images with class 0 while all the other muscles groups together account for a small part of the image. Accuracy, hence, cannot be considered as a reliable metric in this case. Recall values are also high, with all classes having 85% or higher recall. The lowest recall values belong to classes 4, 5, 8, and 9 (Abdominis and Psoas), which are in the 85% range. However, considering the variety of shapes and sizes of these muscles in these images (as seen in Figure 4), these values are actually relatively high. As to precision, for muscle classes 1, and 6-11 it is relatively high (above 80%) while it is lower for classes 2-5 (Left and right Quadratus and Abdominis). This can be because these muscles are relatively smaller and harder to detect. Especially in the case of muscles 2 and 3 (Quadratus), they are positioned in between larger muscles and are also not present in all the images, which makes the task of prediction more difficult. However, if we censor the data to just the images where these muscles are clearly present the precision measure for them also increases. Muscles 4 and 5 (abdominals) are also the most subject to breathing artifacts (movement from breathing while subjects are in MRI machine).

These results indicate that the model is producing many true positives and few false negatives, but also a relatively large number of false positives. While these metrics are not as high as studies that have focused on identifying a single structure (such as vertebrae), they are relatively high when comparing against works that identify such diverse structures. Furthermore, the objective of the study is not to identify each pixel exactly, but rather size, location, and lever arm of each muscle.

5.5.2 Lever arms, area, and centroids

One of the main goals for this task of muscle segmentations is computing geometric muscle metrics such as centroids, areas, and lever arms. High true and false positives and low false negative means that the predicted muscles will cover the original annotated muscles but be somewhat 'fatter', covering some of the empty area around the original, leading to larger areas. However, this does not necessarily affect centroids or lever arms in a specific direction. In order to measure the model's accuracy in predicting these metrics, we computed absolute difference between the measured and predicted values of 90^{o} lever arms, centroids, and areas. The absolute average differences of these values are listed, for each muscle, in Table 5.4.

Table 5.4 shows that predicted centroid and lever arms are very close to the true measurements. Differences in lever arms are less than 1 cm for all the muscles and close to 0.5 cm overall. You can see that the relative difference between predicted and traced lever arms is below 10% (note that the high relative lever arms for muscle group 1 is because the lever arm for this muscle group is very small an close to 0). Similarly, the distance from the centroid of each muscle to the center of the spine was predicted to within 1 cm on average, for each muscle. Areas, on the other hand, are more different. Table 5.4 also shows both the average absolute difference (in cm^2) as well as the average relative difference of predicted post-processed area over true area. On average predicted muscles are 15% to 22% larger than the manually measured (traced) muscles for muscle groups 1-5. This ratio improves a lot for muscles 5-11 which are below 10%. To summarize the results in Table 5.4, the model results in lever arm distances that are within 10% of manually traced muscles. As to area, for area of muscle groups 6-11. the areas are within 10% of each other but the model tends to predict larger areas for muscle groups 1-5.

5.5.3 Speed-up

As mentioned earlier, it can take between 5 to 15 minutes for a person to trace a single MRI image on Osirix©. Once the model is trained, it takes between 10s to 60s to segment the muscles in the

Metric	1	2	3	4	5	6	7	8	9	10	11
Area absolute difference (cm^2)	1.13	1.09	2.43	2.34	2.46	2.62	1.24	1.39	2.83	3.30	1.80
Area relative difference (%)	15.07	15.60	23.79	22.24	19.98	6.92	9.21	10.45	5.53	9.06	7.84
90° lever arm absolute difference (cm)	0.13	0.33	0.29	0.41	0.43	0.39	0.38	0.23	0.25	0.27	0.28
90° lever arm relative difference (%)	7.03	5.33	3.78	8.06	7.84	7.24	3.90	6.05	6.38	3.50	2.21
Centroid distance (cm)	0.41	0.44	0.75	0.68	0.57	0.71	0.39	0.49	0.64	0.66	0.48

Table 5.4: Lever arm, area, and centroid statistics. All values are median values across all predicted muscles.

image. So, the speedup could be somewhere between 5x to 60x. This is a significant efficiency improvement in terms of specialist time which can be used on analysis rather than preprocessing.

5.6 Conclusions and future work

In this paper, we have developed a CNN model to identify muscles in lower back MRIs. A set of traced (annotated) muscles were used as ground truth for training. A seven-layer network was used for the task. Using the located muscle pixels, muscle features such as area, centroid, and lever arm distances were computed automatically. The model showed median predictions are within 8% (0.5 cm) for lever arms and centroids. Predicted muscle areas were slightly larger compared to the traced muscles (6%-24% larger on average depending on the muscle group). The results of the model are dependent on the quality of input images. The model provides much more accurate results for high-contrast MR images compared to blurry ones. Further work in this area includes improving accuracy of the model by providing location information to the model. Once the muscles (and spine) are detected, further features, such as spine and muscle health, obesity measures, etc. can be extracted automatically.

Bibliography

[1] Nonfatal occupational injuries and illnesses requiring days away from work, 2016. BLS.

[2] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard. Tensorflow: A system for large-scale machine learning. In *12th ${$USENIX$}$ symposium on operating systems design and implementation (${$OSDI$}$ 16)*, pages 265–283, 2016.

[3] Felix Altenberger and Claus Lenz. A non-technical survey on deep convolutional neural network architectures. *arXiv preprint arXiv:1803.02129*, 2018.

[4] Saeed Anwar, Salman Khan, and Nick Barnes. A deep journey into super-resolution: A survey. *arXiv preprint arXiv:1904.07523*, 2019.

[5] D. Arola and M. Ramulu. An examination of the effects from surface texture on the strength of fiber reinforced plastics. *Journal of Composite Materials*, 33(2):102–123, 1999.

[6] Abul Bashar. Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence*, 1(02):73–82, 2019.

[7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[8] Jürgen Beyerer, Fernando Puente León, and Christian Frese. *Machine vision: Automated visual inspection: Theory, practice and applications*. Springer, 2015.

[9] Linkan Bian, Scott M. Thompson, and Nima Shamsaei. Mechanical properties and microstructural features of direct laser-deposited Ti-6Al-4V. *Jom*, 67(3):629–638, 2015.

[10] F Edward Boas and Dominik Fleischmann. CT artifacts: causes and reduction techniques. *Imaging Med*, 4(2):229–240, 2012.

[11] Yunliang Cai, Mark Landis, David T. Laidley, Anat Kornecki, Andrea Lum, and Shuo Li. Multi-modal vertebrae recognition using transformed deep convolution network. *Computerized medical imaging and graphics*, 51:11–19, 2016. Publisher: Elsevier.

[12] Patricio E. Carrion, Arash Soltani-Tehrani, Nam Phan, and Nima Shamsaei. Powder Recycling Effects on the Tensile and Fatigue Behavior of Additively Manufactured Ti-6Al-4V Parts. *JOM*, pages 1–11, 2018.

[13] Akshay S. Chaudhari, Zhongnan Fang, Feliks Kogan, Jeff Wood, Kathryn J. Stevens, Eric K. Gibbons, Jin Hyung Lee, Garry E. Gold, and Brian A. Hargreaves. Super-resolution musculoskeletal MRI using deep learning. *Magnetic resonance in medicine*, 80(5):2139–2154, 2018. Publisher: Wiley Online Library.

[14] Pauley Chea and Jacob C. Mandell. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal radiology*, pages 1–15, 2020. Publisher: Springer.

[15] Yuhua Chen, Feng Shi, Anthony G. Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-Lopez, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*, Lecture Notes in Computer Science, pages 91–99, Cham, 2018. Springer International Publishing.

[16] S. R. Daniewicz and Nima Shamsaei. An introduction to the fatigue and fracture behavior of additive manufactured parts. *International Journal of Fatigue*, (94):167, 2017.

[17] Kermit G. Davis and Michael J. Jorgensen. Biomechanical modeling for understanding of low back injuries: a systematic review. *Occupational Ergonomics*, 5(1):57–76, 2005. Publisher: IOS Press.

[18] Tarasankar DebRoy, HL Wei, JS Zuback, T Mukherjee, JW Elmer, JO Milewski, Allison Michelle Beese, A de Wilson-Heid, A De, and W Zhang. Additive manufacturing of metallic components–process, structure and properties. *Progress in Materials Science*, 92:112–224, 2018.

[19] Baba M. Deros, Dian DI Daruis, Ahmad R. Ismail, and Ahmad RA Rahim. Work posture and back pain evaluation in a Malaysian food manufacturing company. *American Journal of Applied Sciences*, 7(4):473, 2010.

[20] Joseph L Dieleman, Ellen Squires, Anthony L Bui, Madeline Campbell, Abigail Chapin, Hannah Hamavid, Cody Horst, Zhiyin Li, Taylor Matyasz, Alex Reynolds, et al. Factors associated with increases in us health care spending, 1996-2013. *Jama*, 318(17):1668–1678, 2017.

[21] Daniel Forsberg, Erik SjÃ¶blom, and Jeffrey L. Sunshine. Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data. *Journal of digital imaging*, 30(4):406–412, 2017. Publisher: Springer.

[22] William E Frazier. Metal additive manufacturing: a review. *Journal of Materials Engineering and performance*, 23(6):1917–1928, 2014.

[23] ES Gadelmawla, MM Koura, TMA Maksoud, IM Elewa, and HH Soliman. Roughness parameters. *Journal of materials processing Technology*, 123(1):133–145, 2002.

[24] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[25] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014. arXiv: 1406.2661.

[26] Stephane Gorsse, Christopher Hutchinson, Mohamed Goune, and Rajarshi Banerjee. Additive manufacturing of metals: a brief review of the characteristic microstructures and properties of steels, Ti-6Al-4V and high-entropy alloys. *Science and Technology of advanced MaTerialS*, 18(1):584–610, 2017. Publisher: Taylor & Francis.

[27] DD Gu and YC Hagedorn. w. meiners, k. wissenbach, r. poprawe. *Int. Mater. Rev*, 57(3):133, 2012.

[28] Celal Gungor. *Prediction of the erector spinae muscle lever arm distance for biomechanical models*. PhD Thesis, 2013.

[29] Celal Gungor, Ruoliang Tang, Richard F. Sesek, Sean Gallagher, and Gerard A. Davis. Morphological investigation of low back erector spinae muscle: Historical data populations. *International Journal of Industrial Ergonomics*, 49:108–115, 2015. Publisher: Elsevier.

[30] Lone Hansen, Mark De Zee, John Rasmussen, Thomas B. Andersen, Christian Wong, and Erik B. Simonsen. Anatomy and biomechanics of the back muscles in the lumbar spine with reference to biomechanical modeling. *Spine*, 31(17):1888–1899, 2006. Publisher: LWW.

[31] Dirk Herzog, Vanessa Seyda, Eric Wycisk, and Claus Emmelmann. Additive manufacturing of metals. *Acta Materialia*, 117:371 – 392, 2016.

[32] Damian Hoy, Christopher Bain, Gail Williams, Lyn March, Peter Brooks, Fiona Blyth, Anthony Woolf, Theo Vos, and Rachelle Buchbinder. A systematic review of the global prevalence of low back pain. *Arthritis & Rheumatism*, 64(6):2028–2037, 2012. Publisher: Wiley Online Library.

[33] D. H. Hubel. Single unit activity in striate cortex of unrestrained cats. *The Journal of Physiology*, 147(2):226–238.2, September 1959.

[34] David H. Hubel and Torsten N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959. Publisher: Wiley-Blackwell.

[35] Workplace Safety Index. Liberty mutual research institute for safety. *Hopkinton, MA*, 2012.

[36] Saurabh Jha and Eric J. Topol. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information SpecialistsArtificial Intelligence: Radiologists and Pathologists as Information SpecialistsArtificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA*, 316(22):2353–2354, 12 2016.

[37] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[38] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[39] Yann LeCun, LÃ©on Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. Publisher: Ieee.

[40] 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. pages 4681–4690, 2017.

[41] Alexander Lenail. http://alexlenail.me/NN-SVG/index.html. accessed on 2020-11-21.

[42] Stefan Leuders, Malte Vollmer, Florian Brenne, Thomas TrÃ¶ster, and Thomas Niendorf. Fatigue strength prediction for titanium alloy TiAl6V4 manufactured by selective laser melting. *Metallurgical and materials transactions A*, 46(9):3816–3823, 2015.

[43] P. Li, D. H. Warner, A. Fatemi, and N. Phan. Critical assessment of the fatigue performance of additively manufactured Ti-6Al-4V and perspective for future research. *International Journal of Fatigue*, 85:130–143, 2016.

[44] Xiaofang Li, Yirui Wu, Wen Zhang, Ruichao Wang, and Feng Hou. Deep learning methods in real-time image super-resolution: a survey. *Journal of Real-Time Image Processing*, pages 1–25, 2019.

[45] 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. SÃ¡nchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017. Publisher: Elsevier.

[46] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017. Publisher: Elsevier.

[47] Clint M Logan, Daniel J Schneberk, Peter J Shull, and Harry E Martz. *X-ray Imaging: fundamentals, industrial techniques and applications*. CRC Press, 2016.

[48] Jen-Tang Lu, Stefano Pedemonte, Bernardo Bizzo, Sean Doyle, Katherine P. Andriole, Mark H. Michalski, R. Gilberto Gonzalez, and Stuart R. Pomerantz. Deepspine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. *arXiv preprint arXiv:1807.10215*, 2018.

[49] Qing Lyu, Chenyu You, Hongming Shan, Yi Zhang, and Ge Wang. Super-resolution mri and ct through gan-circle. In *Developments in X-Ray Tomography XII*, volume 11113, page 111130X. International Society for Optics and Photonics, 2019.

[50] D. T. Mane and Uday V. Kulkarni. A survey on supervised convolutional neural network and its major applications. In *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*, pages 1058–1071. IGI Global, 2020.

[51] N Martini, V Koukou, G Fountos, I Valais, A Bakas, K Ninos, I Kandarakis, G Panayiotakis, and Ch Michail. Towards the enhancement of medical imaging with non-destructive testing (ndt) cmos sensors. evaluation following iec 62220-1-1: 2015 international standard. *Procedia Structural Integrity*, 10:326–332, 2018.

[52] S. M. McGill and R. W. Norman. 1986 Volvo award in biomechanics: partitioning of the L4-L5 dynamic moment into disc, ligamentous, and muscular components during lifting. *Spine*, 11(7):666–678, 1986. Publisher: LWW.

[53] Nima Moshtagh and others. Minimum volume enclosing ellipsoid. *Convex Optimization*, 111:112, 2005.

[54] Yukitaka Murakami. *Metal fatigue: effects of small defects and nonmetallic inclusions*. Elsevier, 2002.

[55] Wendy M. Murray, Scott L. Delp, and Thomas S. Buchanan. Variation of muscle moment arms with elbow and forearm position. *Journal of biomechanics*, 28(5):513–525, 1995. Publisher: Elsevier.

[56] Kamal Nasrollahi and Thomas B. Moeslund. Super-resolution: a comprehensive survey. *Machine Vision and Applications*, 25(6):1423–1468, August 2014.

[57] Robert B Northrop. *Non-invasive instrumentation and measurement in medical diagnosis*. CRC press, 2017.

[58] Jonathan Pegues, Michael Roach, R. Scott Williamson, and Nima Shamsaei. Surface roughness effects on the fatigue strength of additively manufactured Ti-6Al-4V. *International Journal of Fatigue*, 116:543 – 552, 2018.

[59] Jonathan W. Pegues, Nima Shamsaei, Michael D. Roach, and Randall S. Williamson. Fatigue life estimation of additive manufactured parts in the as-built surface condition. *Material Design & Processing Communications*, 1(3):e36, 2019. Publisher: Wiley Online Library.

[60] Jerry L Prince and Jonathan M Links. *Medical imaging signals and systems*. Pearson Prentice Hall Upper Saddle River, NJ, 2006.

[61] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[62] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[63] Antoine Rosset, Luca Spadola, and Osman Ratib. OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of digital imaging*, 17(3):205–216, 2004. Publisher: Springer.

[64] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. " GrabCut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3):309–314, 2004. Publisher: ACM New York, NY, USA.

[65] Mohsen Seifi, Ayman Salem, Daniel Satko, Joshua Shaffer, and John J. Lewandowski. Defect distribution and microstructure heterogeneity effects on fracture resistance and fatigue behavior of EBM Ti-6Al-4V. *International Journal of Fatigue*, 94:263–287, 2017.

[66] Nima Shamsaei, Aref Yadollahi, Linkan Bian, and Scott M. Thompson. An overview of Direct Laser Deposition for additive manufacturing; Part II: Mechanical behavior, process parameter optimization and control. *Additive Manufacturing*, 8:12–35, 2015.

[67] Wei Shen, Feng Yang, Wei Mu, Caiyun Yang, Xin Yang, and Jie Tian. Automatic localization of vertebrae based on convolutional neural networks. In *Medical Imaging 2015: Image Processing*, volume 9413, page 94132E. International Society for Optics and Photonics, 2015.

[68] Wenzhe Shi, Christian Ledig, Zehan Wang, Lucas Theis, and Ferenc Huszar. Super resolution using a generative adversarial network, March 2018.

[69] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[70] John A Slotwinski, Edward J Garboczi, and Keith M Hebenstreit. Porosity measurements and analysis for metal additive manufacturing process control. *Journal of research of the National Institute of Standards and Technology*, 119:494, 2014.

[71] Adriaan B Spierings, M Schneider, and R Eggenberger. Comparison of density measurement techniques for additive manufactured metallic parts. *Rapid Prototyping Journal*, 17(5):380–386, 2011.

[72] Ralph I. Stephens, Ali Fatemi, Robert R. Stephens, and Henry O. Fuchs. *Metal fatigue in engineering*. John Wiley & Sons, 2000.

[73] Amin Suzani, Abtin Rasoulian, Alexander Seitel, Sidney Fels, Robert N. Rohling, and Purang Abolmaesumi. Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 9415, page 941514. International Society for Optics and Photonics, 2015.

[74] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[75] S Tammas-Williams, H Zhao, Fabien LÃ©onard, F Derguti, I Todd, and PB Prangnell. XCT analysis of the influence of melt strategies on defect population in Ti-6Al-4V components manufactured by Selective Electron Beam Melting. *Materials Characterization*, 102:47–61, 2015.

[76] Ruoliang Tang, Celal Gungor, Richard F. Sesek, Kenneth Bo Foreman, Sean Gallagher, and Gerard A. Davis. Morphometry of the lower lumbar intervertebral discs and endplates: comparative analyses of new MRI data with previous findings. *European Spine Journal*, 25(12):4116–4131, 2016. Publisher: Springer.

[77] Ruoliang Tang, Celal Gungor, Richard F Sesek, Sean Gallagher, Gerard A Davis, and Kenneth Bo Foreman. Prediction models for the cross-sectional areas of lower lumbar intervertebral discs and vertebral endplates. *International Journal of Industrial Ergonomics*, 72:12–34, 2019.

[78] A Thompson, I Maskery, and R K Leach. X-ray computed tomography for additive manufacturing: a review. *Measurement Science and Technology*, 27(7):072001, July 2016.

[79] Scott M. Thompson, Zachary S. Aspin, Nima Shamsaei, Alaa Elwany, and Linkan Bian. Additive manufacturing of heat exchangers: A case study on a multi-layered Ti-6Al-4V oscillating heat pipe. *Additive Manufacturing*, 8:163–174, 2015.

[80] Scott M Thompson, Linkan Bian, Nima Shamsaei, and Aref Yadollahi. An overview of Direct Laser Deposition for additive manufacturing; Part I: Transport phenomena, modeling and diagnostics. *Additive Manufacturing*, 8:36–62, 2015.

[81] Brian Torries, Aidin Imandoust, Stefano Beretta, Shuai Shao, and Nima Shamsaei. Overview on Microstructure- and Defect-Sensitive Fatigue Modeling of Additively Manufactured Materials. *JOM*, 70(9):1853–1862, September 2018.

[82] Theo Vos, Abraham D. Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, Majid Ezzati, Kenji Shibuya, Joshua A. Salomon, Safa Abdalla, and Victor Aboyans. Years lived

with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, 380(9859):2163–2196, 2012. Publisher: Elsevier.

[83] Hakon Wadell. Volume, shape, and roundness of quartz particles. *The Journal of Geology*, 43(3):250–280, 1935.

[84] Zhentian Wang, Nik Hauser, Gad Singer, Mafalda Trippel, Rahel A Kubik-Huch, Christof W Schneider, and Marco Stampanoni. Non-invasive classification of microcalcifications with phase-contrast X-ray mammography. *Nature communications*, 5:3797, 2014.

[85] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[86] Zhongyuan Wang, Kui Jiang, Peng Yi, Zhen Han, and Zheng He. Ultra-dense gan for satellite imagery super-resolution. *Neurocomputing*, 398:328–337, 2020.

[87] Aref Yadollahi and Nima Shamsaei. Additive manufacturing of fatigue resistant materials: Challenges and opportunities. *International Journal of Fatigue*, 98:14–31, 2017.

[88] Bin Zhang, Wen Jin Meng, Shuai Shao, Nam Phan, and Nima Shamsaei. Effect of Heat Treatments on Pore Morphology and Microstructure of Laser Additive Manufactured Parts. *Material Design & Processing Communications*, page e29, 2019.

[89] Jun Zhang, Mingxia Liu, and Dinggang Shen. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. *IEEE Transactions on Image Processing*, 26(10):4753–4764, 2017. Publisher: IEEE.

[90] G Ziolkowski, E Chlebus, P Szymczyk, and J Kurzac. Application of X-ray CT method for discontinuity and porosity detection in 316L stainless steel parts produced with SLM technology. *Archives of Civil and Mechanical Engineering*, 14(4):608–614, 2014.