COMPARISON OF POROSITY DETECTION METHODS IN INFRARED IMAGING FOR ELECTRON BEAM POWDER BED FUSION

A Dissertation
Presented to
The Academic Faculty

by

Brian Aidan Johnstone

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
George W. Woodruff School of Mechanical Engineering

Georgia Institute of Technology
May 2023

COMPARISON OF POROSITY DETECTION METHODS IN INFRARED IMAGING FOR ELECTRON BEAM POWDER BED FUSION

Approved by:

Dr. Christopher Saldana, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Katherine Fu, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Thomas Kurfess
School of Mechanical Engneering
Georgia Institute of Technology

Date Approved: April 14, 2023

[To the students of the Georgia Institute of Technology]

ACKNOWLEDGEMENTS

First, I would like to thank my funding sources for this work: the United State Department of Energy under the Enhanced Preparation for Intelligent Cybermanufacturing Systems grant for providing monetary support and Eaton Aerospace company for providing the parts and images that were analyzed in this work. Additionally, I would like to thank my advisors, Dr. Christopher Saldana and Dr. Katherine Fu, for their constant advice and support while working through these problems, and my lab mates, especially Elliott Jost, Jaime Berez, and Caroline Massey, for their support as well. Finally, I would like to thank my friends and family, both near and far, for their constant prayers, love, and support throughout this entire process. I could not have made it through this process without each and every one of them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AM	Additive Manufacturing
DOF	Degrees of Freedom
EB-PBF	Electron Beam Powder Bed Fusion
FDM	Fused Deposition Modeling
FO	Focus Offset
IR	Infrared
ISO	Isometric Selection
L-PBF	Laser Powder Bed Fusion
ML	Machine Learning
NDE	Non-Destructive Evaluation
PBF	Powder Bed Fusion
RGB	Red-Green-Blue
ROI	Region of Interest
SD	Standard Deviation
SLA	Stereolithography
WAAM	Wire-Arc Additive Manufacturing
XCT	X-Ray Computed Tomography

SUMMARY

Additive Manufacturing has been seeing widespread adoption in a variety of industries over the years. However, due to the inherent uncertainties with these manufacturing methods, process monitoring and qualification is critical. In situ process monitoring has been seeing significant interest over the years, and, especially in metal powder bed fusion process such as electron beam powder bed fusion, imaging is commonly used to detect defects. This work looks at comparing three basic segmentation methods (manual thresholding, statistical thresholding, and gradient) used for detecting porosity in in situ infrared imaging in electron beam powder bed fusion. Samples were manufactured with a variety of focus offsets to induce porosity, and the segmented infrared images were compared to ex situ X-ray computed tomography scans, which served as a ground-truth reference for objective evaluation. Through this analysis framework, influential parameters for each of the three methods were analyzed for effect on performance in properly detecting porosity, then the three methods were compared to each other.

CHAPTER 1. INTRODUCTION

1.1 Problem Statement

Additive manufacturing (AM), which was first developed and shown in the 1980's by Kodama [1], has seen significant push in the last two decades [2]. The processes allow for the manufacturing of components unable to be manufactured by traditional processes. This includes components with features such as unique geometries, increased customizability, and weight optimization. These processes create components in a layer-by-layer manner, building up material incrementally from a build plate using a variety of methods. Some methods, such as fused deposition modeling (FDM) and wire-arc AM (WAAM), use a nozzle to extrude material in specific locations. Other methods, such as stereolithography (SLA), expose liquid material to a specific wavelength of light to solidify the material. Still other processes, such as powder-bed fusion (PBF), lay layers of powdered material and melt specific sections of the powder, which then solidifies into a part. For PBF processes, there are two primary types: laser PBF (L-PBF) and electron beam PBF (EB-PBF). As stated in their names, L-PBF uses a laser to melt the powder while EB-PBF uses a beam of accelerated electrons. Additionally, EB-PBF processes include additional steps of heating the powder bed before (called preheating) and after (called post-heating) melting the powder to keep the powder at a more stable temperature by minimizing cyclical thermal loading. However, like with all AM processes, guaranteeing component quality is difficult and the focus of much current research [3]. Without proper process control, PBF components are prone to defects, such as spatter, surface roughness, and porosity. Porosity is when sections inside the solid component are not fused together with the rest of the part, leading to voids in the component. A common method of detecting the presence of these pores is in situ infrared imaging since the pores have greater thermal emissivity than the

solid material. From these images, pores can be detected using various image processing techniques to segment pores from solid material.

This thesis aims to assist the current research by analyzing some low-level segmentation methods seen in the literature for porosity detection in infrared imaging (manual thresholding, statistical thresholding, and gradient). For each method, adjustable parameters will be analyzed for their effect on their respective segmentation method's detection accuracy using x-ray computed tomography (XCT) as a ground-truth reference. Finally, the accuracy of each method will be compared to each other with respect to the XCT scan data.

1.2 Thesis Organization

This thesis is organized into six chapters: *Introduction*, *Background*, *Methods*, *Results*, *Discussion*, and *Conclusions*. Chapter 2 contains context for in situ infrared imaging for porosity detection within the current literature. Chapter 3 contains the methodology for evaluating the different methods used, as well as generating and preparing images for evaluation. Chapter 4 contains the results of the evaluation. Chapter 5 contains a discussion of the implications from the results. Chapter 6 contains a summary of the work and discussion, as well as limitations in the scope of the work and future work that can be done from the work presented.

CHAPTER 2. BACKGROUND

2.1 Additive Manufacturing and Porosity

Many industries have been implementing AM in the last couple decades [2], utilizing it to manufacturing parts unable to be manufactured by traditional processes. This includes components with features such as unique geometries, increased customizability, and weight optimization. However, guaranteeing component quality in AM processes is difficult and the focus of much current research [3]. Without proper process control, PBF parts are prone to defects such as spatter, surface roughness, and porosity. Porosity is when sections inside the solid component are not fused together with the rest of the part, leading to voids in the component. Previous researchers have shown porosity can lead to a decrease in material performance. Edwards et al. [4] showed porosity can lead to premature fatigue, resulting in a lower component performance, while Bisht et al. [5] showed porosity can lead to a decrease in ultimate tensile strength. Since the porosity spots, called pores, are not solidified to the rest of the component, the material does not aid in performance (structural, vibrational, thermal, etc.). Because of this, a major area of research is focused on reducing the likelihood of these defects from appearing in AM components.

2.2 Ex Situ Evaluation

To guarantee components are built with sufficient quality, a variety of evaluation methods have been developed. Standard evaluation methods (such as tensile and hardness testing) have been used as ex situ quality control. However, due to their destructive nature, the parts evaluated via these methods cannot be used after evaluation. Therefore, these methods are typically used as ground-truth comparison for other non-destructive evaluation (NDE) methods. One such NDE method is x-ray computed tomography (XCT). Previous researchers, such as du Plessis et al. [6], have researched and categorized how defects

detected by XCT impact component quality. This allows for each component to be evaluated for quality while still being usable after evaluation.

2.3 In Situ Evaluation

As accurate as XCT and other ex situ NDE methods are, they require performing additional evaluations and obtaining additional equipment, which is expensive in both time and money, especially in a production setting. This has led to an increase in interest in utilizing in situ evaluation methods [7]. These methods work by incorporating a variety of sensors to detect anomalies in the build process to predict the presence of detects in the components. Some of these sensors are used to detect anomalies in machine functionality (such as gas flow [8] or recoater blade vibrations), while others are used to detect anomalies in the actual parts (such as porosity or surface roughness). One of the more common sensor types is imaging. Both optical (as seen in Aminzadeh et al. [9]) and infrared (IR) cameras are used in industry and research to detect anomalies in PBF processes and have been established to effectively detect certain quality defects. IR cameras are especially effective at detecting porosity because pores have greater thermal emissivity than the solid material and therefore appear as bright spots in IR images.

2.4 Infrared Imaging and Segmentation

Many researchers have used IR imaging to monitor meltpools and detect pores. Williams et al. [10] used it to measure the effect of interlayer cooling time on porosity generation. Rodriguez et al. [11] showed that IR imaging can be effective for detecting defects in EB-PBF processes. Cordero et al. [12] used IR imaging to look at how chimney pores develop throughout parts in EB-PBF. To detect these pores, a variety of image processing techniques have been used to segment pores from the rest of the image.

The most used method is an image binarization with a manually set threshold (called a manual threshold in the literature). This method is the easiest to implement, as any pixel value above the threshold will be designated as part of a pore, while any pixel value below will not. Ridler et al. [13] showed this method could be iterated in a general case (called an iterative selection (ISO) method), and this has become common in industries such as food processing. Lough et al. [14] used this method to determine areas with temperatures above a desired value during a build process, and Mahmoudi et al. [15] used it to determine regions of interest (ROI) for further analysis. Another method is a statistical thresholding method. This is like the manual thresholding method but calculates a threshold for each layer using the mean and standard deviation (SD) of the pixel values. These statistical methods work like the Otsu threshold selection method [16], which uses statistical evaluation of gray values to find distinguished peaks in gray value clusters to denote classes and find a threshold value between them, as both are more unsupervised compared to the manual methods and are therefore as less subjective. Bartlett et al. [17] used a method like this, setting the threshold to one standard deviation above the mean pixel value, to evaluate the effectiveness of using IR imaging for defect detection in PBF processes. A third common method is the gradient segmentation. Since pores have higher grayscale values than the solid material around it, it leads to having a large change in grayscale values with the nearby pixels. Gradient segmentation finds pores by using these grayscale gradients to find the outlines of pores, then uses an imaging filling technique to fill in the rest of the pore. Yoder et al. [18] used this method to identify defects in topology-optimized components.

Finally, many researchers have developed machine learning (ML) methods to detect pores in IR images. However, these methods are typically also detecting a variety of other part defects (such as surface roughness) or are trying to predict the likely cause of the pores (such as lack of fusion or keyholing, as seen in Kozjek et al. [19]). These ML methods will

use some type of segmentation approach like the previously explained image processing techniques in identifying pores. However, they tend to use more feature-based methods (such as identifying certain shapes) in addition to the more basic segmentation methods, as seen in Scime et al. [20], who used gradients as part of a defect prediction ML model. This difference makes them different enough to make direct method comparisons difficult. Additionally, they require significantly more computing power and may even require layer history to accurately identify pores.

2.5 Research Gaps

For these more low-level approaches, little research has been done to optimize the segmentation methods. Manual thresholds are typically used for analyzing individual layers, as the threshold is easy to adjust, but the optimal threshold value is usually determined qualitatively. Statistical methods have only used an n value of 1 SD above the mean pixel value for their threshold and have not looked at how changing this parameter can affect detection accuracy. Gradient methods pick one type of gradient that works for the specific case, once again determining the optimal type qualitatively. Additionally, little research has gone into comparing the effectiveness of these methods, and while frameworks for evaluating these methodologies have been proposed using ex situ evaluation as ground truth, such as in Winton et al. [21], such evaluations have still only been qualitative in nature or use single-layer comparisons to evaluate the defect detection accuracy for whole components.

Finally, some researchers have used XCT scanning to evaluate the validity of IR in situ imaging. Mohr et al. [22] qualitatively compared IR images to XCT scans to observe how areas of increased thermal emissivity led to formation of an increased number of defects. Croset et al. [23] used XCT scans to qualitatively verify the presence of pores and other defects that were detected via IR images. Coeck et al. [24] overlayed point clouds of pores

found in IR images over a point cloud of pores derived via XCT scans to train a prediction model to detect lack of fusions events with a voxel size of 30 μm. Finally, Ren et al. [25] used in situ x-rays as the comparison for in situ IR imaging instead of ex situ to train an ML model for real-time porosity detection. However, these comparisons have either only been done qualitatively or using too large a voxel size to detect smaller defects. Additionally, while in situ detection can be helpful for real-time detection, the evolution of porosity from manufacturing to finished product is not yet fully understood.

This work aims to contribute to the field by addressing the following questions:

1) How are manual segmentation methods affected by their threshold parameter?
2) How are statistical segmentation methods affected by their N value (number of standard deviations above the mean pixel value) parameter?
3) How are gradient segmentation methods affected by their gradient type?
4) How do factors external to the analysis methods (such as porosity levels and overall image brightness) affect porosity detection in these methods?

This work uses XCT scans with a sufficiently small voxel size as a ground-truth comparison to objectively evaluate the effectiveness of each method for both optimization and cross-method comparison. These comparisons will give future researchers greater understanding of which segmentation approaches to use for their relevant applications, and the optimizations will allow them to use the methods with their most accurate parameters. Additionally, the framework for evaluating the methods can be applied to other segmentation methods to determine their detection accuracy.

CHAPTER 3. METHODS

3.1 Sample Design and Registration

The specimens used for this experiment were manufactured out of Ti-6Al-4V powder on an ARCAM Spectra L EB-PBF machine. In total, 27 samples were manufactured with a layer height of 70 μm (see Table 3.1). Porosity was created at varying levels by adjusting the focus offset (FO), as Cordero et al. [12] showed that adjusting focus offset has a strong relationship to porosity formation. Of these samples, 9 samples had an FO of 55 mA, 9 had an FO of 65 mA, and 9 had an FO of 75 mA, and they were placed in the build volume in sets of 9 (three of each FO). The samples were designed as cylinders with a flat vertical face (see Figure 3-1). This was to make registration easier, with the cylindrical section restricting 4 degrees of freedom (DOF) (translational X, translational Y, rotational X, and rotational Y), the vertical face restricting another DOF (rotational Z), and the top restricting the final DOF (translational Z) (see Figure 3-2). It should be noted that both the top and bottom faces were intended to be used to limit translational Z, but due to the poor surface finish on the bottom face, only the top face was used for registration. The samples were 18 mm tall with a diameter of 8 mm and were grouped in the build volume in sets of 9 (3 samples of each FO). The in-situ images were taken with the Spectrum L's built-in IR camera, using a focal length of 350 mm, a bit depth of 8 bits, an image size of 5120 x 5120 pixels, and a pixel size of 70 μm. Once manufactured, the samples were scanned in a Metrotom CT machine. To ensure consistency among the samples, a fixture was made to give the same orientation during scanning. All samples were scanned using a voltage of 130 kV, current of 61 μA, and a voxel size of 12 μm.

Table 3.1: Sample Manufacturing Parameters

Sample Numbers	Focus Offset (mA)	Layer Height (µm)	Sample Height (mm)	Sample Diameter (mm)
1, 2, 3, 10, 11, 12, 19, 20, 21	55	70	18	8
4, 5, 6, 13, 14, 15, 22, 23, 24	65			
7, 8, 9, 16, 17, 18, 25, 26, 27	75			

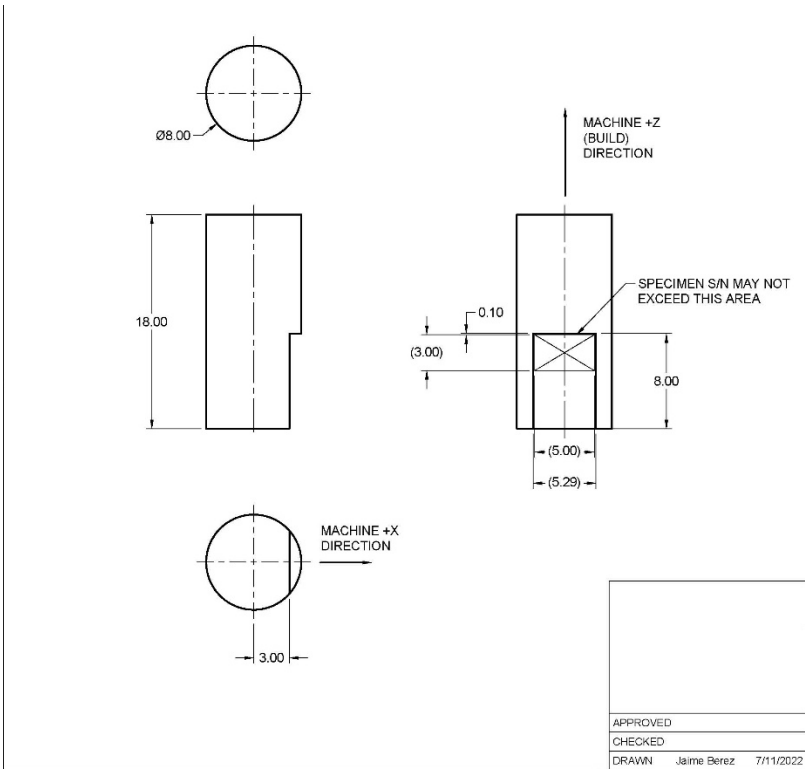

Figure 3-1: Sample Design

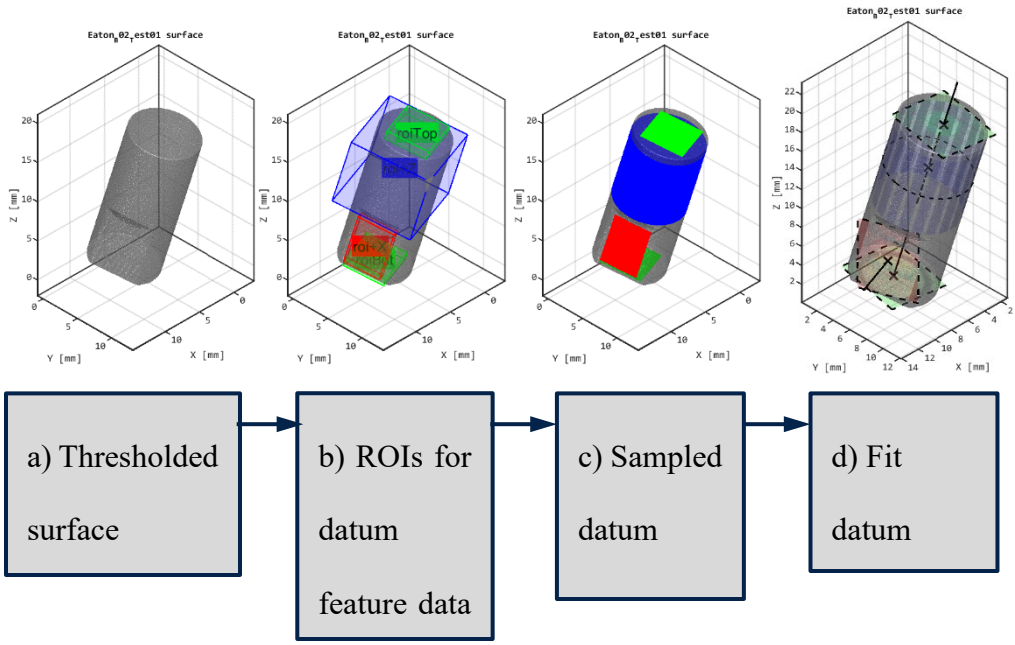

Figure 3-2: Sample Registration Process

3.2 IR Image Processing and Segmentation Methods

The IR images were distortion-corrected in MATLAB 2022b using the program's pinhole distortion correction method toolbox. Once corrected, each image was cropped to create a single image stack for each sample. Then, the images were filtered using an anisotropic diffusion filter, which smooths images while maintaining gradients to remove noisy pixels, before performing each segmentation method on each image stack. For the manual cutoff, the images had their gray values unitized to be between 0 and 1; then, the images were binarized using MATLAB imbinarize function, with thresholds varying from 0.5 to 0.95. For the statistical cutoff, the mean and standard deviation (SD) of the pixel grayscale values for each image were calculated then binarized using the MATLAB imbinarize function, with the threshold being determined by being N * SD above the mean and n ranging from 0 to 6. For the gradient segmentation, the grayscale gradient was determined for each image

using the MATLAB imgradient function, using all the available gradient types. These ranges are listed in Table 3.2. Once the segmented image stacks were generated, they were cropped to be within the contours of the part. Finally, they were resampled to match the smaller CT pixels size and allow for direct pixel-to-pixel comparisons between the CT and IR images.

Table 3.2: Segmentation Methods, Parameters, and Ranges

Method	Parameter	Tested Values/Types	
Manual	Threshold	0.5 0.55 0.6 0.65 0.7	0.75 0.8 0.85 0.9 0.95
Statistical	Number of Standard Deviations Above Mean (N)	0 0.5 1.0 1.5 2.0 2.5 3.0	3.5 4.0 4.5 5.0 5.5 6.0
Gradient	Gradient Type	Sobel Prewitt Central	Intermediate Roberts

3.3 CT Volume Processing

The CT scan volumes were registered using MATLAB code developed by the research team to the IR images. First, volumes were converted from volumes to surfaces by thresholding the volume grayscale values. This allowed for easier detection of regions of interest (ROI) for determining datum features without interference from internal porosity. The first datum feature was the cylindrical section, followed by the vertical face and finally the top face. For each of these features, surface locations were sampled within the respective ROI, then the geometric datum features were best fit to the actual features. A

cylinder feature was fit to the cylindrical sections, while planes were fit to the flat faces. This placed the volumes into the correct orientation to be directly compared to the IR images. Then, the volumes were sliced into image stacks in MATLAB. These image stacks were then processed using Weka, a machine learning image segmentation tool in the image processing program Fiji, to detect porosity. A reference pixel was set that would be aligned to a similar reference pixel in the IR images. Finally, the CT images were cropped to be within the part boundaries to eliminate edge segmentation errors.

3.4 CT to IR Comparison Analysis

Once all these image stacks were processed, the segmented IR image stacks for each sample were compared directly to their CT counterpart on a pixel-by-pixel level. Based on their accuracy, each pixel was given a new red-green-blue (RGB) color value based on the following detection criteria (also listed in Table 3.3): true positive (green, IR detected pore where CT detected pore), true negative (grey, IR did not detect pore where CT did not detect pore), false positive (blue, IR detected pore where CT did not detect pore), and false negative (red, IR did not detect pore where CT did detect pore). The black pixels represent pixels outside the component boundary. For each layer of each sample, the number of pixels for each condition were counted and saved for further analysis.

Table 3.3: Evaluation Conditions and Color Designations

Condition	Detected in IR?	Detected in CT?	RBG Color
True Positive	X	X	Green
True Negative			Grey
False Positive	X		Blue
False Negative		X	Red

CHAPTER 4. RESULTS

4.1 Infrared and X-Ray Computed Tomography Imaging Results

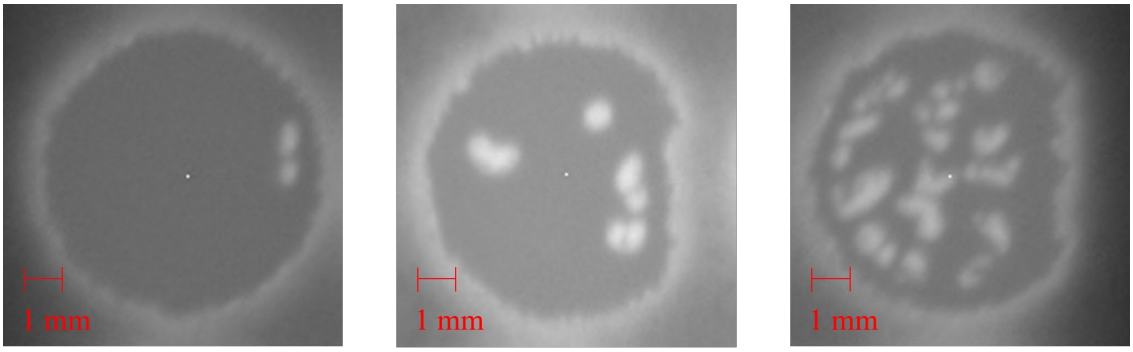

Figure 4-1: Raw IR Images for Sample 11 (Left), Sample 5 (Middle), and Sample 17 (Right)

Figure 4-1 shows 3 examples of raw IR images before performing any segmentation analysis representing the 3 different FO's (Sample 11 on the left with FO = 55mA, Sample 5 in the middle with FO = 65 mA, and Sample 17 with FO = 75 mA). As predicted, the increase in FO resulted in an increase in porosity. The bright dot in the center of each image is the reference pixel, which is used to properly register and align the IR and CT images. Additionally, another aspect to note is that the overall brightness varies between the samples, which may influence the effectiveness on the methods.

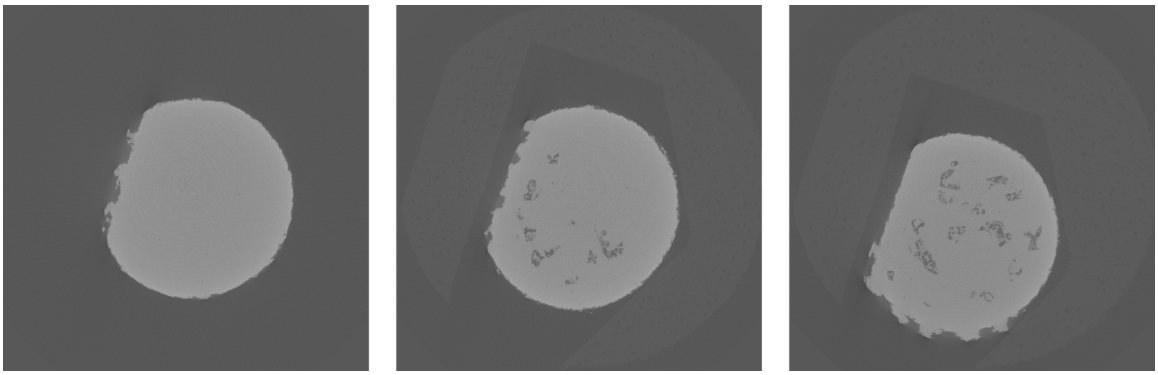

Figure 4-2: Raw CT Images for Sample 11 (Left), Sample 5 (Middle), and Sample 17 (Right)

Figure 4-2 shows the results of CT scanning Samples 11, 5, and 17. The additional object in the image from the scans for Samples 5 and 17 is the holder used for scanning. The changes in FO also resulted in varying levels of detectable porosity, which will allow the CT scans to be used as ground-truth reference for the IR images. As with the IR images, the increase in FO resulted in an increase in porosity, allowing for analysis at different porosity levels.

Figure 4-3shows a side-by-side comparison of a rescaled IR image and registered CT image for Sample 5. It is important to note that the pores have different appearance in each image stack, so evaluating analysis methods need to account for this difference.

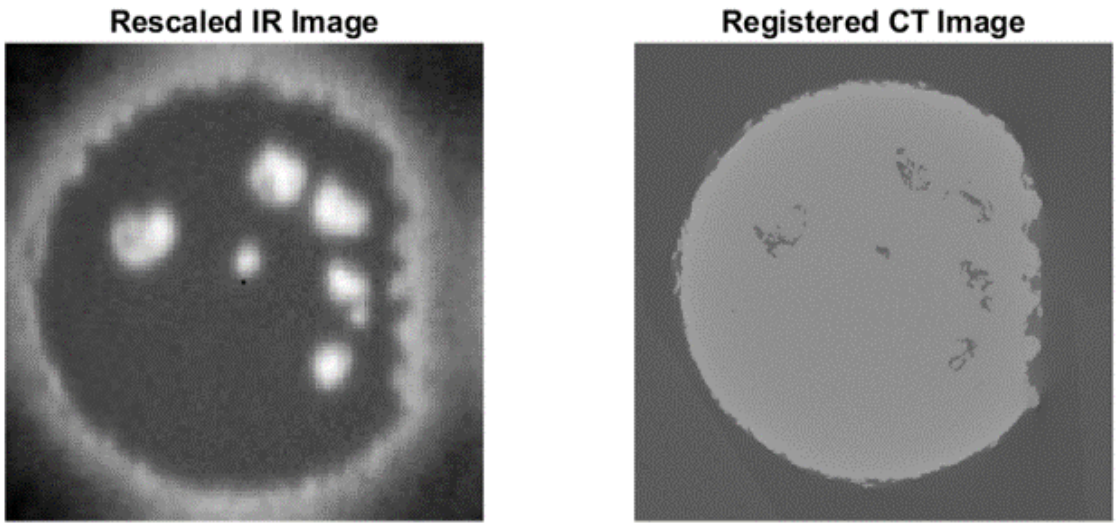

Figure 4-3: Side-by-Side Comparison of Rescaled IR Image and Registered CT Image for Sample 5

4.2 Manual Thresholding Segmentation

Figure 4-4 shows the results of performing the manual thresholding segmentation analysis on Samples 5, 11, and 17 (raw images shown in Figure 4-1). In these resulting images, pixels that are white are classified as pore, while pixels that are black are classified as not a pore. The pixel in the center of each image is the reference pixel, which is used to align the CT and IR image stacks. As can be seen, the threshold value can influence the results of the segmentation. When the threshold is set high enough, the resulting segmented will classify everything as not a pore (called blacking-out). This would mean pores that exist in the part will be missed. When the threshold is set low enough, the resulting segmentation will classify everything as a pore (called whiting-out). This would mean features would be misclassified as a pore when they are not. The most common of these features would be the contours. However, in this work, contours were filtered out and were not considered when determining the accuracy of each method. Since all images with threshold greater than 0.85 were blacked-out, analysis on images with thresholds of 0.90 and 0.95 were not analyzed further.

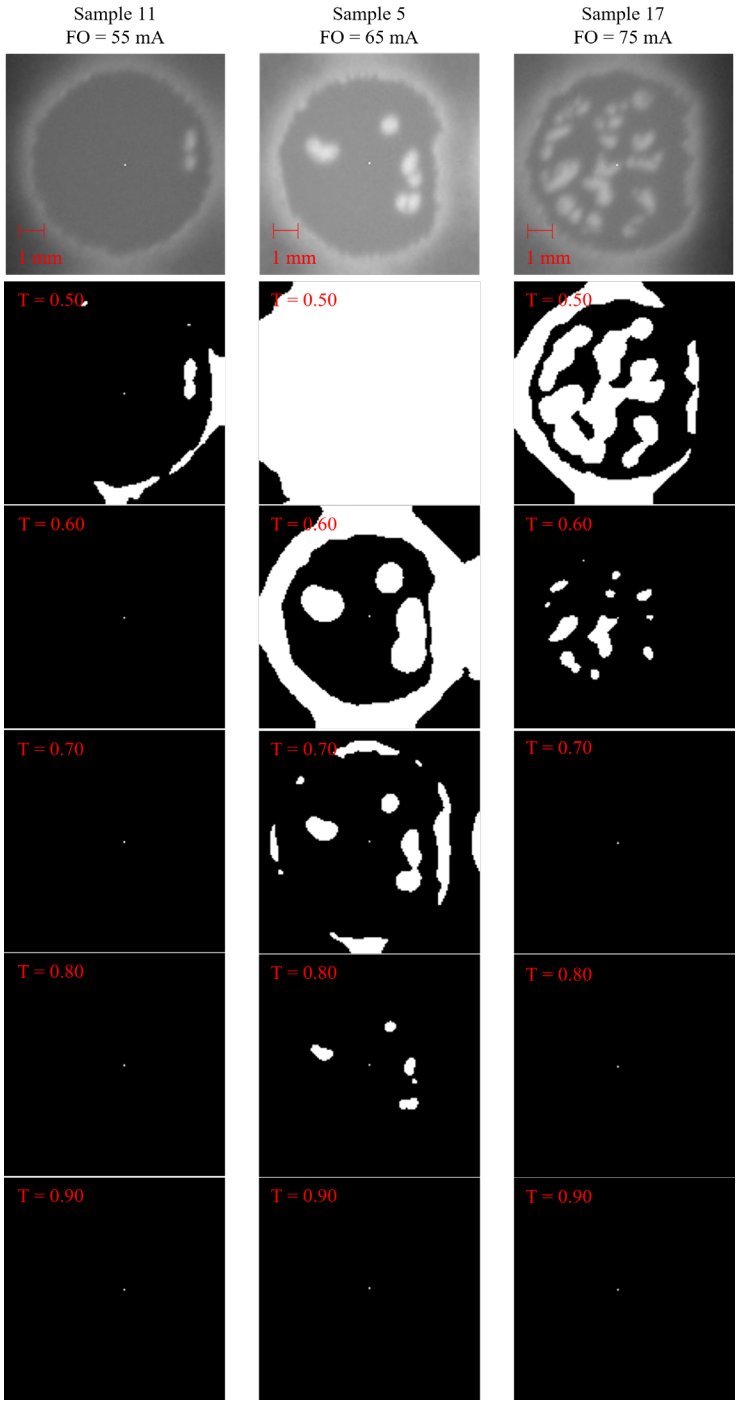

Figure 4-4: Manual Segmentation Results for (Left) Layer 118 (Height = 8.26 mm) of Sample 11 (FO = 55 mA), (Middle) Layer 100 (Height = 7.0 mm) of Sample 5 (FO = 65 mA), and (Right) Layer 40 (Height = 2.8 mm) of Sample 17 (FO = 75 mA) Compared to Their Raw IR Images

Figure 4-5 shows the F1 score of the manual segmentation method as a function of threshold. F1 score is a segmentation evaluation metric that is calculated using Equations 1-3. For this calculation, the number of pixels for each of the four evaluation conditions (true positive, true negative, false positive, false negative) are counted for each image slice of each sample, and the totals for each parameter are summed for the plot. These scores range from 0 to 1, with 1 meaning perfect detection.

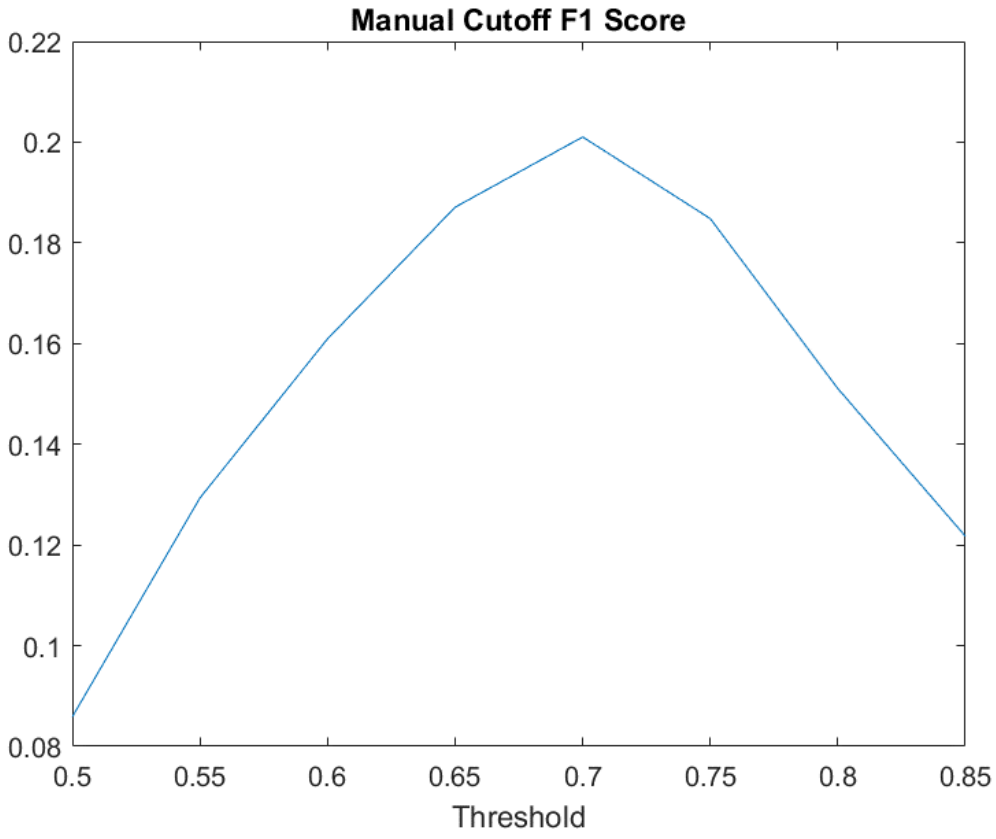

Figure 4-5: F1 Score for Manual Segmentation Method

$$Precision = \frac{True\ Positive\ Pixels}{True\ Positive\ Pixels + False\ Positive\ Pixels} \quad (1)$$

$$Recall = \frac{True\ Positive\ Pixels}{True\ Positive\ Pixels + False\ Negative\ Pixels} \quad (2)$$

$$F1\ Score = \frac{2\ \times Precision\ \times Recall}{Precision + Recall} \quad (3)$$

Figure 4-6 shows a plot of the average area of each analysis condition for each parameter value. This was determined by counting the pixels for each condition, then multiplying that count by the area of each pixel (12 μm x 12 μm). One observation to note is the true negative area is at least an order of magnitude higher than true positive. This is likely due to how most of the part in the image is material, meaning there will be more area associated with not being a pore and result in imbalanced classes. This is why F1 score was chosen to measure classification performance over other metrics (such as classification accuracy) to measure classification performance, as the quantity of TN pixels would skew the results if other performance metrics were used.

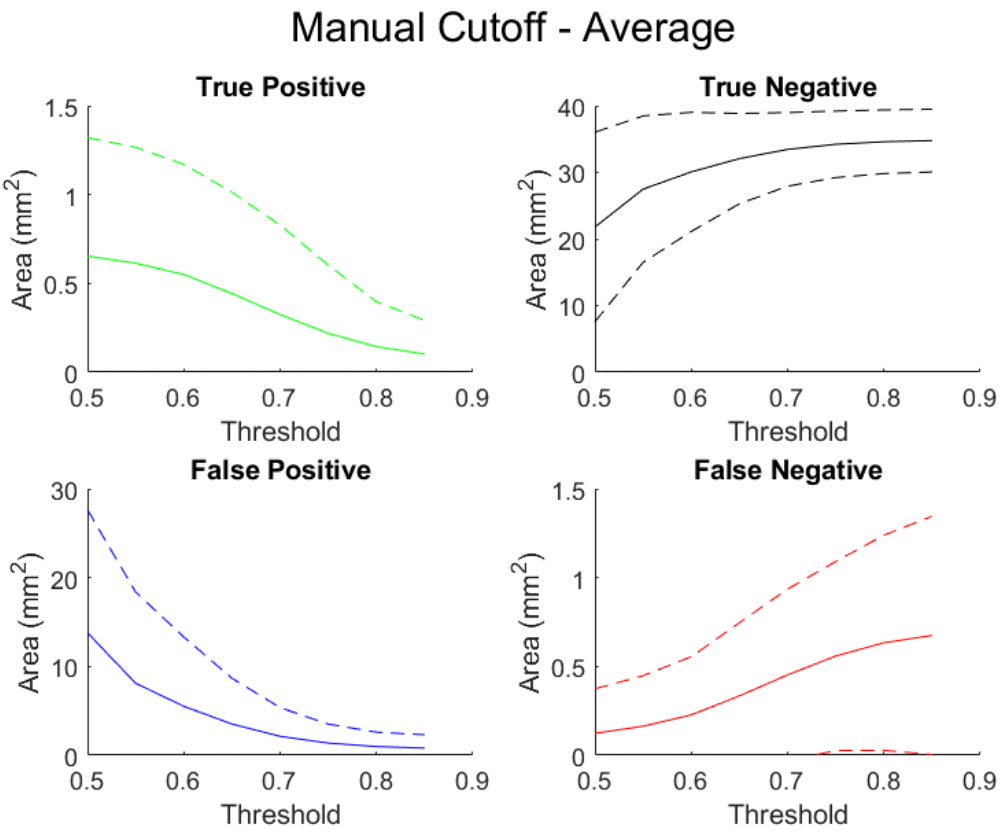

Figure 4-6: Average Condition Count (Solid) +/- 1 SD (Dashed) for Manual Segmentation Methods

Figure 4-7 shows the shows an example of colorized comparison for results comparing manual IR and CT image stacks on Sample 22. As can be seen, the lower thresholds have a fair amount of oversizing the pores, while the higher thresholds have the pores undersized and even disappearing. Therefore, while increasing the threshold can result in increased performance, increasing too much can result in a decrease in performance.

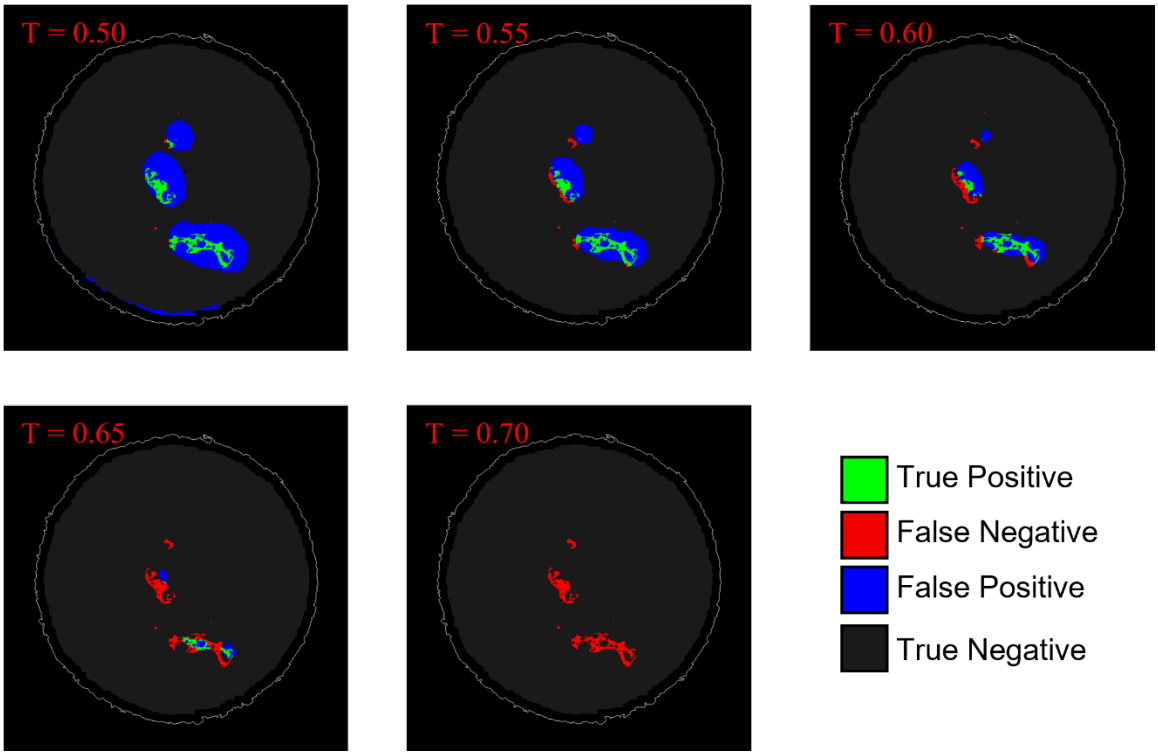

Figure 4-7: Colorized Comparison of Manual Segmentations Applied at Height = 10.92 mm on Sample 22 with Various Threshold Values

4.3 Statistical Thresholding Segmentation

Figure 4-8 shows the results of performing the statistical segmentation on Samples 5, 11, and 17. Like with the manual thresholding segmentation method, the tested parameter (in this case, the number of SD above the mean pixel gray scale value N) influences the detection of porosity. If the value of N is too low (such as with N = 0), the segmentation will over-identify pixels in the image as pores that are not, and if the value of N is too high (such as with N = 6), the segmentation will under-identify pixels in the image as not pore when they are. While there are noticeable differences between the different values of N, the differences are less obvious compared to the other methods. Therefore, qualitative method optimization is more difficult and quantitative method optimization is required.

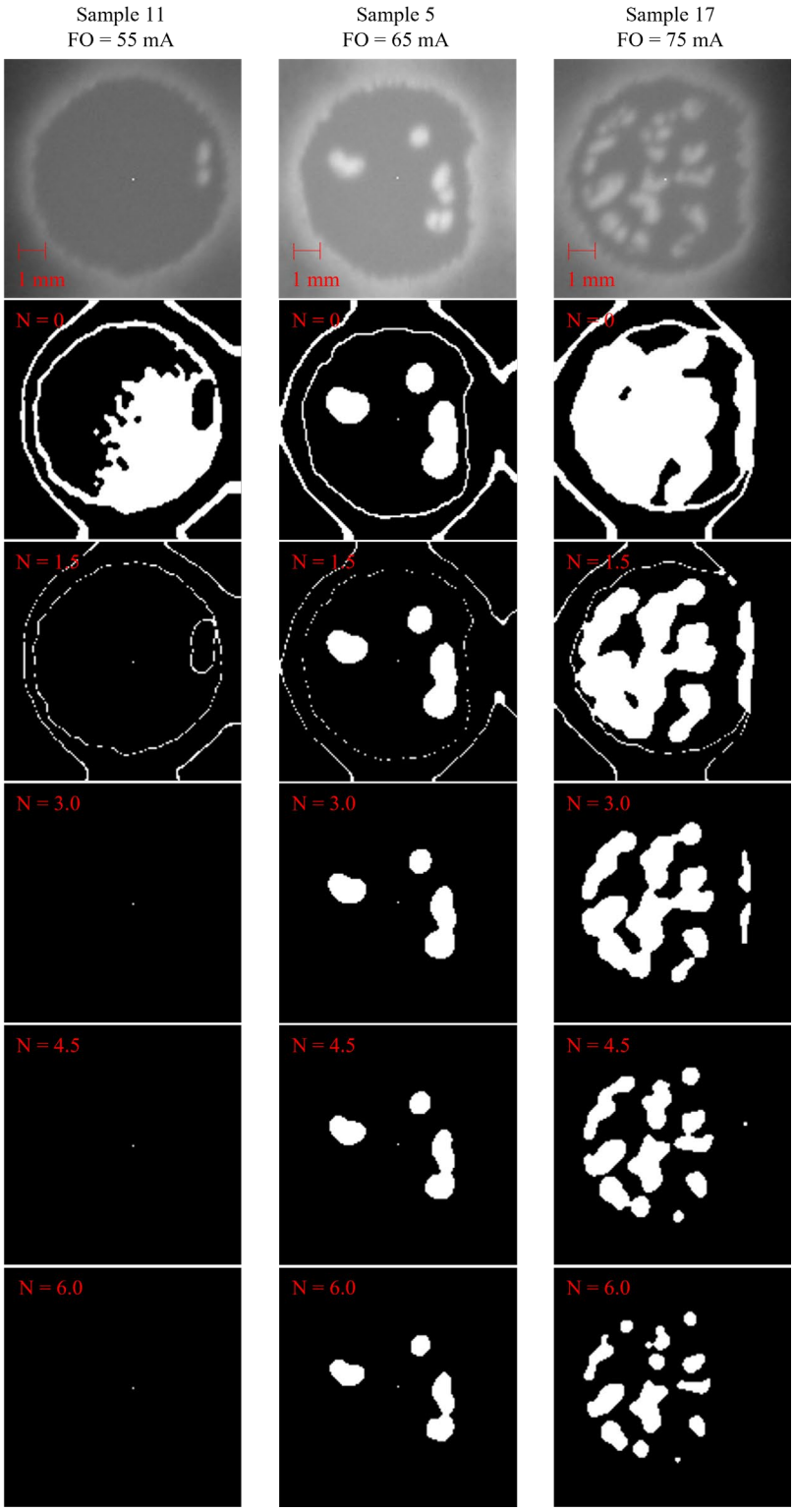
Sample 11
FO = 55 mA
Sample 5
FO = 65 mA
Sample 17
FO = 75 mA
1 mm
1 mm
1 mm
N = 0
N = 0
N = 0
N = 1.5
N = 1.5
N = 1.5
N = 3.0
N = 3.0
N = 3.0
N = 4.5
N = 4.5
N = 4.5
N = 6.0
N = 6.0
N = 6.0

Figure 4-8: Statistical Segmentation Results for (Left) Layer 118 (Height = 8.26 mm) of Sample 11 (FO = 55 mA), Layer 100 (Height = 7.0 mm) of Sample 5 (FO = 65 mA), and Layer 40 (Height = 2.8 mm) of Sample 17 (FO = 75 mA) Compared to Their Raw IR Images

Figure 4-12 shows an example of colorized comparison results for comparing statistical IR and CT image stacks. A legend of the four detection conditions and their respective color classifications is shown in the bottom right corner. As can be seen, the N parameter influences the detection capabilities of the methods.

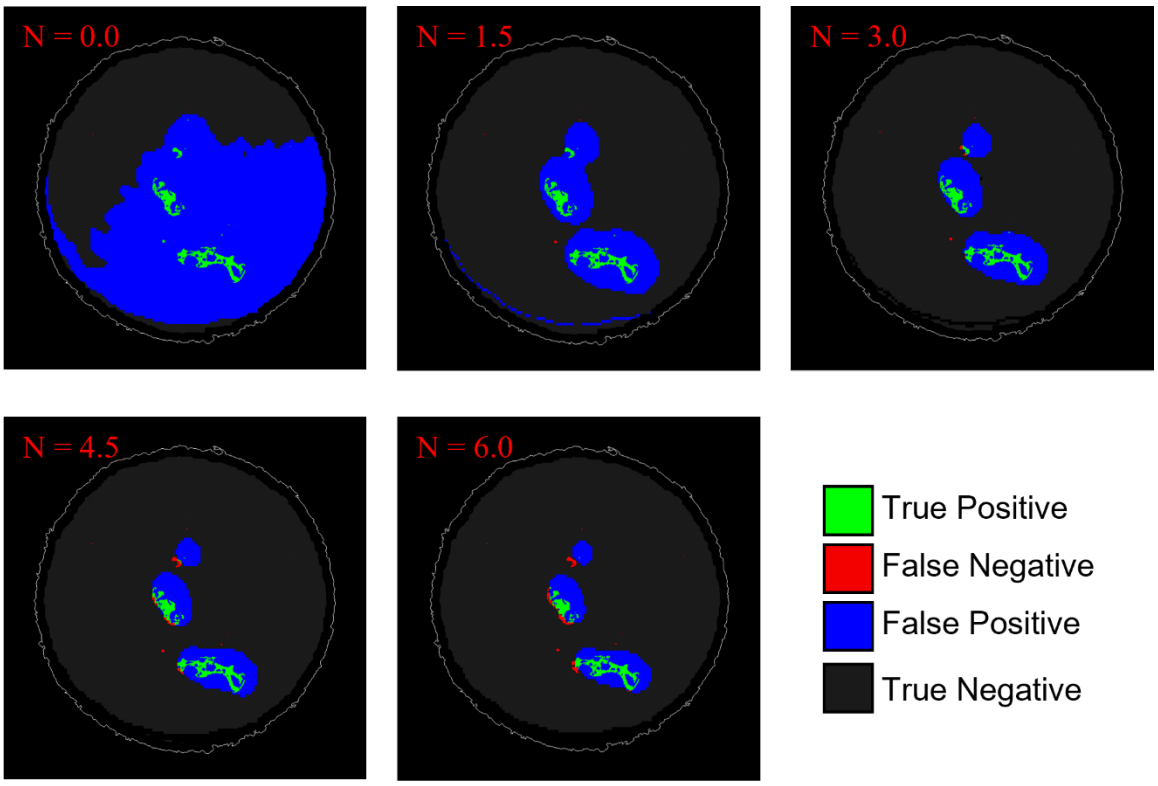

Figure 4-9: Colorized Comparison of Statistical Segmentations Applied at Height = 10.92 mm on Sample 22 with Various N-Values

Figure 4-10 contains a plot of the F1 score for the statistical segmentation as a function of N. The graph shows a positive correlation between N and accuracy, as accuracy increases with an increase in N.

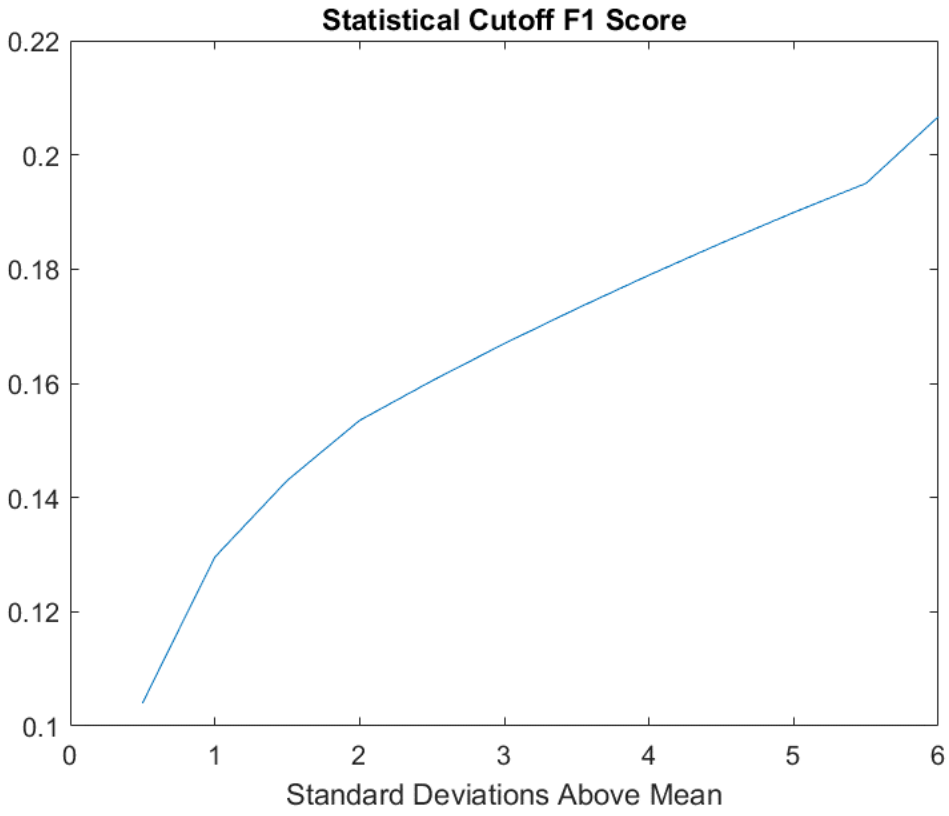

Figure 4-10: F1 Score of Statistical Segmentation Method

Figure 4-11 shows the average condition area (+/- 1 standard deviation) for the statistical segmentation methods. As with the manual segmentation, the true negative area is significantly higher than any of the other conditions. However, unlike the manual, the true positive area always remains above the false positive count.

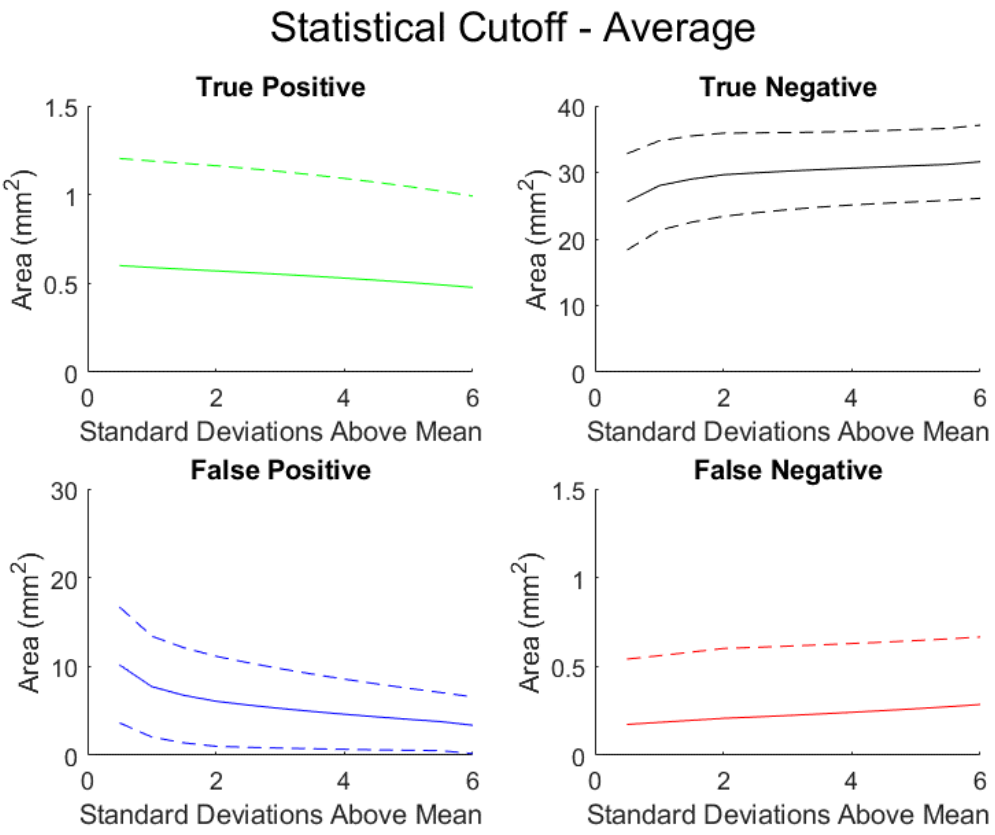

Figure 4-11: Average Condition Count (Solid) +/- 1 SD (Dashed) for Statistical Segmentation Methods

4.4 Gradient Segmentation

Figure 4-12 contains the results of gradient segmentation on Samples 11, 2, and 17. As can be seen from the figures, only the Sobel and Prewitt methods detected gradients large enough to analyze, so the other three methods (Central, Intermediate, and Roberts) were not analyzed further. While this was also the case for many of the results for the Prewitt gradient as well, there were some samples where gradients were still detected in some samples, so Prewitt was still processed further.

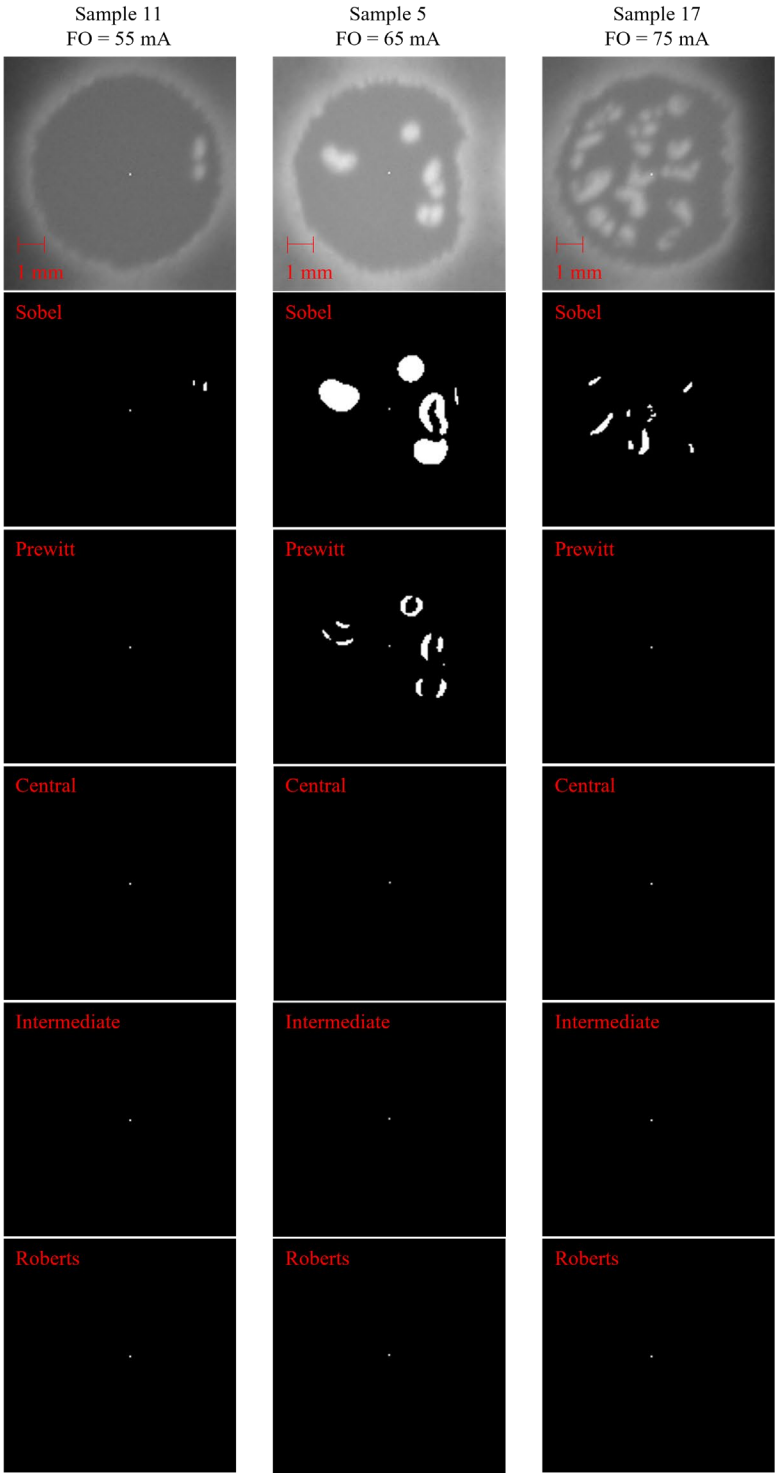

Figure 4-12: Gradient Segmentation Results for (Left) Layer 118 (Height = 8.26 mm) of Sample 11 (FO = 55 mA), Layer 100 (Height = 7.0 mm) of Sample 5 (FO = 65 mA), and Layer 40 (Height = 2.8 mm) of Sample 17 (FO = 75 mA) Compared to Their Raw IR Images

Figure 4-13 shows a colorized comparison example result of comparing the Sobel IR and CT image stacks for Sample 22. The true positive pixels appear clustered around the edges of the pore where the pores are wider. But where the pore gets narrower, there are more false negative pixels without nearby true or false positive pixels.

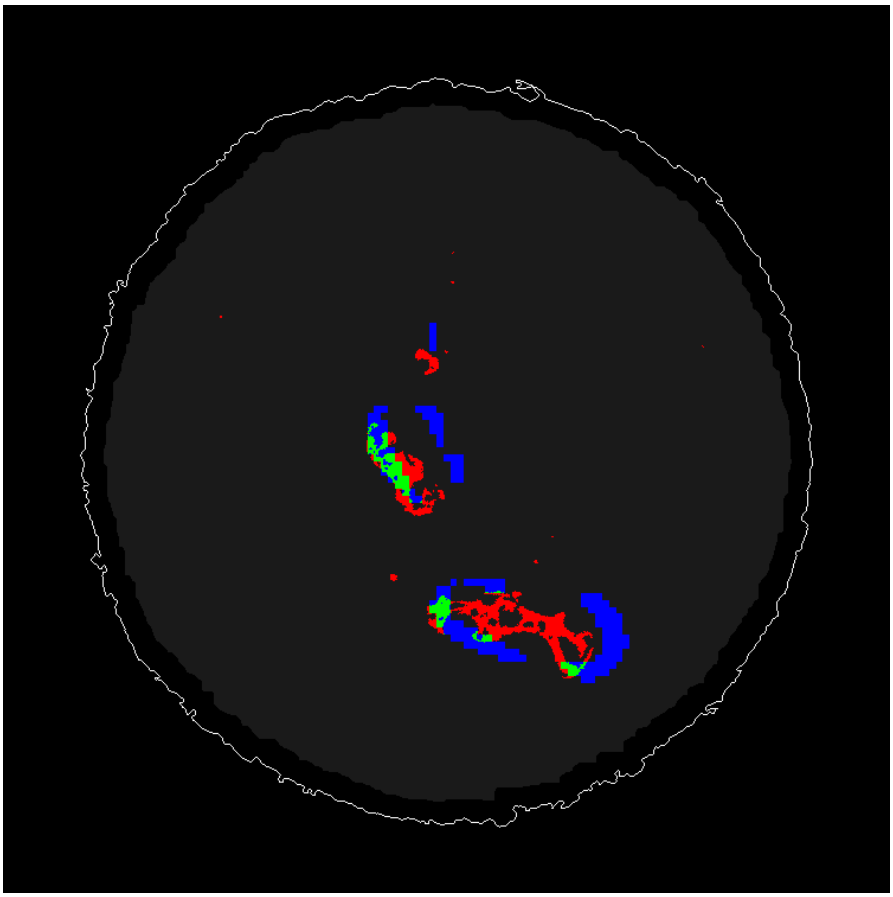

Figure 4-13: Colorized Comparison of Sobel Gradient Performed at Height = 10.92 mm of Sample 22

Table 4.1 contains the classification accuracy of the two analyzed gradient segmentation methods. Like with the other two segmentation methods, these values are based off averages across the analyzed samples.

Table 4.1: F1 Score for Gradient Segmentation Analyses

Gradient Type	F1 Score
Sobel	0.0883
Prewitt	0.0251

Table 4.2 and Table 4.3 show the confusion matrices for the Sobel and Prewitt gradient types, respectively. Between the two methods, it can be noted the true and false positive values are higher in the Sobel method, while the true and false negative values are higher in the Prewitt method. This is likely due to Sobel detecting more porosity, while the Prewitt more often resulted in blacked-out images.

Table 4.2: Confusion Matrix for Average Area in mm (+/- 1 SD) for Prewitt Gradient Analysis

		Sobel	
		Actual	
		Positive	Negative
Predicted	Positive	0.215 (0.481/0)	3.084 (6.077/0.090)
	Negative	0.563 (1.085/0.041)	32.645 (37.690/27.600)

Table 4.3: Confusion Matrix for Average Area in mm (+/- 1 SD) for Prewitt Gradient Analysis

		Prewitt	
		Actual	
		Positive	Negative
Predicted	Positive	0.018 (0.055/0)	0.417 (1.190/0)
	Negative	0.760 (1.399/0.123)	35.310 (39.761/30.860)

4.5 Segmentation Method Comparison

Figure 4-14 shows an example of colorized comparisons for each of the 3 main types of segmentation approaches performed on Sample 22.

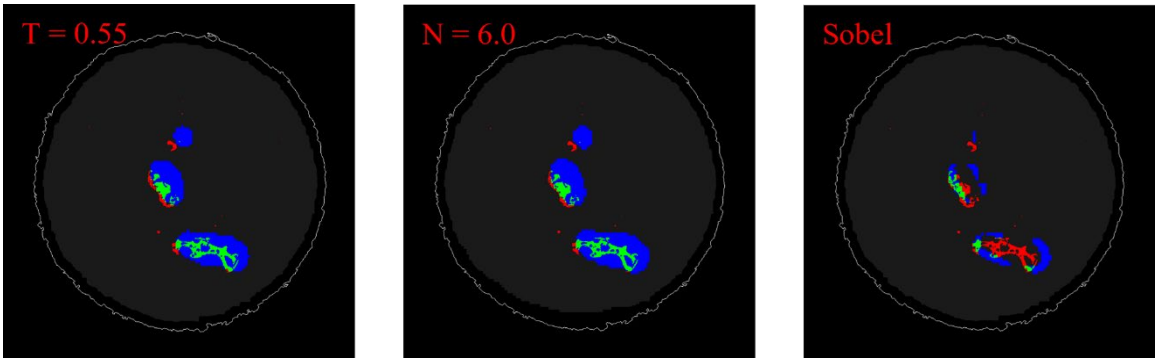

Figure 4-14: Side-by-Side Comparison Between Colorized Comparison Results of Manual (Left), Statistical (Center), and Gradient (Right) Segmentation

CHAPTER 5. DISCUSSION

5.1 Manual Thresholding Segmentation

When observing the raw IR images, there is significant difference between results based on both threshold and porosity levels. As the threshold increases, the porosity representation changes greatly until dropping off completely, resulting in a blacked-out image. This black-out level changes between images, with brighter images having a higher threshold than dimmer images. For example, in Figure 4-4, the image for Sample 11 has a black-out level between T = 0.50 and 0.60, Sample 17 has a black-out level between 0.60 and 0.70, and Sample 5 has a black-out level between 0.80 and 0.90. This shows that image brightness will have a significant effect on the detection of porosity. Additionally, on images with higher overall brightness, the differences in results between threshold values is significantly more. Sample 5 goes from nearly completely whited-out at T = 0.50 to completely blacked out at T = 0.90, while neither of the other show images have that stark difference between their base case and worse case. Therefore, if this method is used, an image filter or normalization operation should be used, as seen in Croset et al. [43].

When analyzing the colorized comparisons, the threshold value appears to have a significant effect on detection accuracy. When the threshold value is too low (as seen when set to 0.50), the cutoff is low enough to catch some of the emissivity of the fused part, leading to the segmented image being whited-out. This led to a significantly higher false positive rate. When the threshold value is too high (as seen when set to 0.80 or higher), the segmented image is blacked-out, with some pores not being detected and others that are detected being represented as significantly smaller than they are. This led to a significantly

higher false negative rate. When finding an optimal threshold, this method is very sensitive to the threshold parameter, meaning small changes in the threshold value can lead to large changes in porosity representation. Based on the F1 scores, 0.75 appeared to perform the best for these conditions. While the exact value may vary slightly based on process conditions, these results show that optimizing this threshold requires a balance of increasing the threshold to minimize false positives and decreasing the threshold to minimize false negatives.

Additionally, this method appears to be very sensitive to external factors. Overall image brightness significantly affected the optimal threshold value for each image, as the optimal threshold for one layer may result in an entirely blacked-out or white-out result in others. Therefore, this method may be best for analyzing a small number of layers with minimal change in conditions external to the analysis itself.

5.2 Statistical Thresholding Segmentation

When observing the raw IR images, increasing the N value can significantly change the look of the segmented images. This is most prevalent between lower values of N, such as between N = 0.0 and N = 1.5. However, this change is less prevalent as N increase, as can be seen between N = 4.5 and N = 6.0. Additionally, while contours were filtered out of the analysis in this work, it can be noted that the contours naturally filter out consistently between N = 1.5 and N = 3.0, unaffected by porosity levels. This notes a potential relationship in the differences in detected emissivity between contours and pores. This relationship can be researched further in future studies, but this was outside the scope of this work.

When analyzing the colorized comparison images, the N value appears to influence detection accuracy. However, it seems to have the largest impact on the false positive and true negative values. When N was below 2, the false positive rate was highest, meaning that if the N value is too low, the detection algorithm will detect pores that are not actually present. While the false negative rate continues to go down as N increases, the decrease in false negative rate also decreases, indicating that an increase in N after 3 does not result in significant increase in detection accuracy. Based on the plotting of F1 score, the statistical method has the performance nearly double at N = 2 from N = 0, but the performance increases less as N is increased up to 6. From a binary pore/no pore perspective, N > 1.5 accurately detect the existence of pores in the parts, and further increase of N primarily results in the decreasing of false positive values. This shows that while an increase in N does result in an increase in performance, the rate of increase in performance will decrease over time. This builds off previously done work using statistical methods, such as Barlett et al. [17], which just used N = 1 for porosity detection. Adjusting N can result in greater porosity detection to best optimize for the given set of process conditions.

However, this method does seem to struggle slightly with dimly lit images (such as the ones seen in Sample 11). Based on the build layout, these images were likely dimmer due to being beamed first, therefore having more time to cool down before being imaged (as each image was taken after the layer had finished being beamed). This means the intensity of the pixels in the contours likely played a greater role in skewing the average and SD pixel intensity values. In these methods, the contour pixels were filtered out after segmentation and alignment, and therefore they did not have a direct impact on the classification accuracy. However, as stated, the contours may have had a more indirect impact on the classification.

Additionally, this method is computationally expensive. The intensity value of each pixel in the image is considered when determining the mean and SD. While this was not a

problem in this work (as the larger images were cropped to only contain a single sample), this could prove problematic with larger image sizes. Calculating the mean and SD for a given set of values is still simple, so larger image sizes will likely not be too problematic.

5.3 Gradient Segmentation

As previously stated, analyses for Central, Intermediate, and Roberts method were not performed due to these types not detecting large enough gradients. For the remaining two gradients, the Sobel method appears to perform better than Prewitt. Prewitt missed several pores, and the pores that were detected were not accurately represented. Meanwhile, the Sobel method appears to detect all the pores present. However, it struggles to accurately represent pores that are next to each other. This likely has to do with nearby pores having reduced emissivity gradients compared to stand-alone pores, meaning the change in gray values will be lower and more difficult to detect. This is seen best in the samples with higher porosity levels (FO = 75 mA) compared to the sample with moderate porosity levels (FO = 65 mA). The large amount of porosity results in much of the relevant area, so the change between pixels that are part of a pore have less change than pixels that are not part of a pore. Additionally, this method appears to be very susceptible to the effect of changes in overall brightness. Despite having less porosity, the gradient methods were not able to identify pores. This may be due to a similar reasoning as with the high porosity samples. When the overall image is dimmer, the difference between the gray values of pores and solid material will be less, leading to a decreased gradient magnitude.

While both analyzed gradient types received high classification accuracy, this number alone is slightly misleading. In most samples, most pixels will be for true negative, which means those pixels have values of 0. And since classification accuracy takes both true positive and true negative into account when considering accuracy, the apparent “true

negatives" skew the results, making them appear to perform better than samples with false positive results. Both these methods, especially the Prewitt, performed fairly poorly.

However, while it did not represent the pores accurately, the Sobel gradient method was fairly accurate with detecting the existence of porosity in components. The Sobel pores had significantly smaller area and often were just along the edges of the actual pores, the method can still be used in a more binary detection method. This could be used in applications that just need to know if pores are present in parts, but the size and shape of the porosity is irrelevant.

5.4 Segmentation Method Comparison

When evaluating these methods, one aspect that must first be addressed is the difference between the appearance of pores in the IR images and the CT images. Since the IR images are taken in situ, the temperature of the parts is higher due to having just been beamed. Therefore, the parts will contract between the times the IR images and CT images were taken due to the significant change in temperature. This has been documented in literature [17] [26] and is why many detection algorithms focus on detecting the presence and special distribution of porosity instead of the exact size [15]. This is likely a factor in why the F1 score magnitudes are lower than other work, as these methods were evaluated on a pixel-by-pixel level instead of a pore-by-pore level. This phenomenon has yet to be fully researched and detailed quantitatively. However, the trends from the F1 score can still be utilized to see the influence each method's parameter has on its performance.

Out of the three methods, the manual segmentation method appears to be the most sensitive to its adjustable parameter, while the statistical segmentation appears to be less sensitive. While the statistical segmentation varied from changing the N parameter, the change from

each adjustment is less significant than the manual segmentation. In the tested range, the manual segmentation results ranged from fully whited-out to fully blacked-out on most slices. While there were still instances of this happening with the statistical segmentation, the occurrence was significantly less, happening only on sporadic slices. This implies the statistical method is more robust and reliable than the manual method, as the tighter window of possible results means that outside influences are less likely to affect the detection of porosity.

One such example of required robustness is performing with a significant change in overall brightness. These changes in brightness occur for a variety of reasons from porosity density to time since being beamed. Even within the image stacks of a single sample, some layers were significantly brighter than others. Set thresholds, like the ones in the manual method, do not take this into account as they simply look at pixels above a certain threshold. Statistical methods can account for this, as each layer has its own threshold based on the intensity of the pixels. Additionally, because of this, the statistical method is easier to optimize as factors external to the method itself play less of a role in determining optimal parameters.

In terms of overall ranking, the statistical method performed best out of the three methods. The robustness of the method to external conditions allowed it to more accurately model porosity in a larger variety of samples. The manual method performed second best. While not as robust as the statistical method and containing significantly more change between samples, it was still able to detect most pores in the images. The gradient method performed worst. Out of the 5 methods used, 3 of the methods produced no results, 1 produced minimal results, and the last produced moderate results at best.

However, these methods could be used in tandem with each other. For example, since the gradient method could detect the presence of pores, it could be used to detect layers of

interest. Then, a more robust method, such as the statistical method, could be used to get a more quantitative and accurate understanding of the quantity, size, and shape of the porosity on those layers. This example would reduce the overall required computational power, as the Sobel gradient method only considers the surrounding 8 pixels while the statistical consider all pixels in the image.

CHAPTER 6. CONCLUSION

6.1 Conclusion

In this work, three different types of basic segmentation methods (manual thresholding, statistical thresholding, and gradient thresholding) were analyzed for their effectiveness in detecting and modeling porosity in in situ IR images in EB-PBF processes. Each of these methods were analyzed and optimized individually, using XCT scans as ground-truth verification, then they were compared to each other for their accuracy. Some conclusions from the work include:

1. How are manual segmentation methods affected by their threshold parameter? The manual method, while being susceptible to influence of external conditions, produces best results when set between 65 – 80% of the maximum intensity value. With an optimal threshold value, porosity can be accurately detected and modeled. However, the tolerance on this value is small, as small changes in the threshold value can result in significant decrease in classification accuracy.
2. How are statistical segmentation methods affected by their N value parameter? The statistical method produced the best results overall, with an optimal N number of SD above the mean set to at least 3. This method showed the least amount of variance between results due to both internal and external factors and most accurately detected and model porosity, and the tolerance for optimal N was much larger than the other methods.
3. How are gradient segmentation methods affected by their gradient type? The gradient methods are significantly affected by their gradient type. While still usable for detecting the presence of pores, performs the worst compared to the other

methods presented. Most methods analyzed did not produce results, and the remaining methods did not produce as accurate results as the other segmentation methods.

4. How do factors external to the analysis methods (such as porosity levels and overall image brightness) affect porosity detection in these methods? Overall brightness of the images, which can be due to a variety of factors such as porosity density or time between beaming and imaging, plays a significant role in detection accuracy and finding optimal parameters. However, while this influenced the results in each of these methods, the statistical method showed the most resilience and still performed best through a variety of brightness conditions.

6.2 Contributions

This work adds to the field of AM by providing an objective evaluation of commonly used segmentation techniques used for detecting porosity. This will allow future researchers to utilize these techniques with their most optimal settings for this set of build condition settings without having to figure them out for themselves. Additionally, it will allow for industry to use these techniques to properly detect porosity in their own components using these techniques. Finally, this evaluation framework can be used to evaluate the performance of other segmentation methods.

6.3 Limitations and Future Work

This work looked at how the main parameter for each of the 3 segmentation methods affects how well it detects and models porosity. However, there are other factors that could affect the detection of pores. One such factor is pre-processing of the images. This analysis used an anisotropic diffusion filter, but there are a variety of filters that could be used (such as

gaussian and median). Therefore, future work could investigate what affect these filters have on the detection of porosity and which ones can be used to improve each segmentation method.

Another limitation was the difference in resolution between the CT and IR images. The IR images used a pixel size of 70 µm, which is nearly 6 times the size of the 12 µm voxel size used in the CT scans. Therefore, any pores smaller than 70 µm will not appear in the IR imaging, and any pores smaller than 12 µm will not appear in the CT. Additionally, the pores will naturally appear larger in the IR images than in the CT. However, there are other factors that can affect how the pores will change from when the IR images are taken (immediately post beaming) and when they are CT scanned (such as shrinkage due to cooling). Future work can be done to gain a better understanding of how much of an effect each of these factors have on the changes in representation and how to best compensate for that when developing more accurate models.

Additionally, the analysis of this work was limited to 3 basic segmentation methods. However, the evaluation framework can be applied to any segmentation method (such as convolution neural networks or even machine learning). Once the analysis method has generated a stack of segmented images, they can be compared to the CT ground-truth data to objectively evaluate performance.

Finally, this analysis was limited to only a single AM process. All parts were made solely out of Ti-6Al-4V via one model of EB-PBF, and analysis was only performed on IR images with the same imaging settings. Therefore, future work could look at how different segmentation methods perform with different materials (such as aluminum or stainless steel), processes (such as laser PBF or directed energy deposition), and in-situ imaging types (such as optical) and settings (such as adjusting the aperture). For materials, the biggest factor in changing how these parameters will change from one material to another

is the thermal emissivity. The IR cameras measure thermal emissivity, so this material property will directly affect the gray scale values in the images and their contrast with each other. Therefore, this will influence the effectiveness of these segmentation methods. For processes, porosity forms differently in EB-PBF compared to other processes, giving them a significantly different appearance in the in-situ images. For imaging types and settings, since different imaging types measure different types of light, the appearance of objects will differ. For example, optical images measure visible light instead of IR. Adjusting most imaging settings will change how much light is measured. Both changes may affect their appearance in the images, most relevant being their brightness (which was established to influence the performance of these methods. However, this same analysis framework can be applied to any of these changes and allow the analysis methods to be optimized for their specific conditions.

REFERENCES

[1] H. Kodama, "Automatic method for fabricating a three-dimensional plastic model with photo-hardening polymer," *Review of Scientific Instruments*, vol. 52, no. 11, pp. 1770–1773, Aug. 1981.

[2] O. Abdulhameed, A. Al-Ahmari, W. Ameen, and S. H. Mian, "Additive manufacturing: Challenges, trends, and applications," *Advances in Mechanical Engineering*, vol. 11, no. 2, pp. 3429–3435, Feb. 2019.

[3] F. H. Kim and S. P. Moylan, "Literature review of Metal Additive Manufacturing Defects," *NIST Advanced Manufacturing Series*, May 2018.

[4] P. Edwards, A. O'Conner, and M. Ramulu, "Electron beam additive manufacturing of titanium components: Properties and performance," *Journal of Manufacturing Science and Engineering*, vol. 135, no. 6, pp. 1–7, Nov. 2013.

[5] M. Bisht, N. Ray, F. Verbist, and S. Coeck, "Correlation of selective laser melting-melt pool events with the tensile properties of ti-6al-4v Eli processed by laser powder bed fusion," *Additive Manufacturing*, vol. 22, pp. 302–306, May 2018.

[6] A. du Plessis, I. Yadroitsava, and I. Yadroitsev, "Effects of defects on mechanical properties in metal additive manufacturing: A review focusing on X-ray tomography insights," *Materials and Design*, vol. 187, Feb. 2020.

[7] R. McCann, M. A. Obeidi, C. Hughes, É. McCarthy, D. S. Egan, R. K. Vijayaraghavan, A. M. Joshi, V. Acinas Garzon, D. P. Dowling, P. J. McNally, and D. Brabazon, "In-situ sensing, process monitoring and machine control in Laser Powder Bed Fusion: A Review," *Additive Manufacturing*, vol. 45, May 2021.

[8] J. Reijonen, A. Revuelta, T. Riipinen, K. Ruusuvuori, and P. Puukko, "On the effect of shielding gas flow on porosity and melt pool geometry in laser powder bed fusion additive manufacturing," *Additive Manufacturing*, vol. 32, 2020.

[9] M. Aminzadeh and T. Kurfess, “Layerwise automated visual inspection in Laser Powder-bed additive manufacturing,” *Volume 2: Materials; Biomanufacturing; Properties, Applications and Systems; Sustainable Manufacturing*, 2015.

[10] R. J. Williams, A. Piglione, T. Rønneberg, C. Jones, M.-S. Pham, C. M. Davies, and P. A. Hooper, “In situ thermography for laser powder bed fusion: Effects of layer temperature on porosity, microstructure and mechanical properties,” *Additive Manufacturing*, vol. 30, p. 100880, Oct. 2019.

[11] E. Rodriguez, J. Mireles, C. A. Terrazas, D. Espalin, M. A. Perez, and R. B. Wicker, “Approximation of absolute surface temperature measurements of powder bed fusion additive manufacturing technology using in situ infrared thermography,” *Additive Manufacturing*, vol. 5, pp. 31–39, Jan. 2015.

[12] Z. C. Cordero, R. B. Dinwiddie, D. Immel, and R. R. Dehoff, “Nucleation and growth of chimney pores during electron-beam additive manufacturing,” *Journal of Materials Science*, vol. 52, no. 6, pp. 3429–3435, 2016.

[13] Ridler, T. W., & Calvard, S. (1978). Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man and Cybernetics*, *smc-8*(8), 630–632.

[14] C. S. Lough, X. Wang, C. C. Smith, R. G. Landers, D. A. Bristow, J. A. Drallmeier, B. Brown, and E. C. Kinzel, “Correlation of SWIR imaging with LPBF 304L stainless steel part properties,” *Additive Manufacturing*, vol. 35, p. 101359, May 2020.

[15] M. Mahmoudi, A. A. Ezzat, and A. Elwany, “Layerwise anomaly detection in laser powder-bed fusion metal additive manufacturing,” *Journal of Manufacturing Science and Engineering*, vol. 141, no. 3, pp. 1–13, Jan. 2019.

[16] Otsu, N., Smith, P. L., Reid, D. B., Environment, C., Palo, L., Alto, P., & Smith, P. L. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *C*(1), 62–66.

[17] J. L. Bartlett, F. M. Heim, Y. V. Murty, and X. Li, “In situ defect detection in selective laser melting via full-field infrared thermography,” *Additive Manufacturing*, vol. 24, pp. 595–605, Jul. 2018.

[18] S. Yoder, S. Morgan, C. Kinzy, E. Barnes, M. Kirka, V. Paquit, P. Nandwana, A. Plotkowski, R. R. Dehoff, and S. S. Babu, “Characterization of topology optimized ti-6al-4v components using electron beam powder bed fusion,” *Additive Manufacturing*, vol. 19, pp. 184–196, Jan. 2018.

[19] D. Kozjek, F. M. Carter, C. Porter, J.-E. Mogonye, K. Ehmann, and J. Cao, “Data-driven prediction of next-layer melt pool temperatures in laser powder bed fusion based on co-axial high-resolution planck thermometry measurements,” *Journal of Manufacturing Processes*, vol. 79, pp. 81–90, Apr. 2022.

[20] L. Scime and J. Beuth, “Using machine learning to identify in-situ melt pool signatures indicative of flaw formation in a laser powder bed fusion additive manufacturing process,” *Additive Manufacturing*, vol. 25, pp. 151–165, Oct. 2019.

[21] H. C. de Winton, F. Cegla, and P. A. Hooper, “A method for objectively evaluating the defect detection performance of In-Situ Monitoring Systems,” *Additive Manufacturing*, vol. 48, p. 102431, Nov. 2021.

[22] G. Mohr, S. J. Altenburg, A. Ulbricht, P. Heinrich, D. Baum, C. Maierhofer, and K. Hilgenberg, “In-situ defect detection in laser powder bed fusion by using thermography and optical tomography—comparison to computed tomography,” *Metals*, vol. 10, no. 1, p. 103, Jan. 2020.

[23] G. Croset, G. Martin, C. Josserond, P. Lhuissier, J.-J. Blandin, and R. Dendievel, “In-situ layerwise monitoring of electron beam powder bed fusion using near-infrared imaging,” *Additive Manufacturing*, vol. 38, p. 101767, Jan. 2021.

[24] S. Coeck, M. Bisht, J. Plas, and F. Verbist, “Prediction of lack of fusion porosity in selective laser melting based on melt pool monitoring data,” *Additive Manufacturing*, vol. 25, pp. 347–356, Jan. 2019.

[25] Z. Ren, L. Gao, S. J. Clark, K. Fezzaa, P. Shevchenko, A. Choi, W. Everhart, A. D. Rollett, L. Chen, and T. Sun, “Machine learning–aided real-time detection of keyhole pore generation in laser powder bed fusion,” *Science*, vol. 379, no. 6627, pp. 89–94, Jan. 2023.

[26] Arnold, C., Breuning, C., & Körner, C. (2021). Electron-optical in situ imaging for the assessment of accuracy in electron beam powder bed fusion. *Materials*, *14*(23). https://doi.org/10.3390/ma14237240