Journal of Manufacturing Processes 149 (2025) 1066–1077

Contents lists available at ScienceDirect

Journal of Manufacturing Processes

journal homepage: www.elsevier.com/locate/manpro

Effect of threshold parameters on infrared segmentation methods for porosity detection in electron beam powder bed fusion

Brian Johnstone [a], Jaime Berez [b], Caroline Massey [a], Elliott Jost [a], Christopher Saldana [a], Katherine Fu [c,*]

[a] Georgia Institute of Technology, George W. Woodruff School of Mechanical Engineering, 801 Ferst Dr., Atlanta, GA 30332, USA
[b] University of North Carolina at Charlotte, Mechanical Engineering and Engineering Science, 9201 University City Blvd., Charolette, NC 28223, USA
[c] University of Wisconsin at Madison, College of Engineering, 1415 Engineering Drive, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Metal additive manufacturing
Process monitoring
Infrared imaging
X-Ray computed tomography

ABSTRACT

In-situ process monitoring has seen significant interest in additive manufacturing to address qualification and certification goals. This is especially prevalent in metal powder bed fusion processes such as electron beam powder bed fusion (PBF-EB), with layer-wise infrared imaging being commonly used to detect defects. This work compares two different segmentation methods (static thresholding and statistical thresholding) used for detecting porosity from in-situ infrared imaging data for PBF-EB. Samples were manufactured at a variety of focus offset values to induce porosity. Then, 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, the influential parameters, static threshold and N-value (number of standard deviations above the mean pixel value), respectively, for both image segmentation methods were analyzed and compared for their effects on porosity detection. With optimal parameter settings, the two methods had similar porosity detection performance, but the statistical method performed better under a larger variety of parameter settings.

1. Background

Additive manufacturing (AM) has seen an increase in use in industry and research to create components with features such as complex geometries, increased customizability, and weight optimization that cannot be accomplished via other manufacturing methods. However, guaranteeing component quality in AM processes can be difficult due to the process understanding of AM being less established than that of manufacturing methods like machining or casting. Therefore, many researchers have focused efforts on detecting and measuring factors related to part quality [1]. Without proper process control, powder bed fusion (PBF) parts are prone to defects, such as dross, high surface roughness, and porosity. Porosity primarily occurs in electron beam powder bed fusion (PBF-EB) when sections inside the solid component are not fused together with the rest of the part, leading to voids in the component [2]. Previous researchers have shown porosity can impact mechanical performance by producing premature fatigue failure [3] and decreased ultimate tensile strength [4]. Because of this, detecting, predicting, and reducing the likelihood of these defects appearing in AM components has become a major focus of current metal AM research.

While standard evaluation methods for determining material properties (such as tensile and hardness testing) have been used to evaluate the performance of AM parts, non-destructive evaluation (NDE) methods offer the ability to use parts after evaluation. One NDE method that has seen high use in AM is X-ray computed tomography (XCT) [5]. While XCT and other ex-situ NDE methods have become more capable and widely used, they can be time and cost intensive. This has led to an increased interest in utilizing in-situ evaluation methods that qualify a component in-process rather than adding subsequent inspection processes [6]. These methods incorporate a variety of sensors to detect anomalies in the build process and predict the presence of defects. Sensors have been used to detect anomalies in machine functionality, such as gas flow [7] or recoater blade vibrations [8], and in the manufactured parts, such as porosity [9] or high surface roughness [10]. Both visible-light spectrum [11] and infrared (IR) cameras [12] have been used to detect anomalies in PBF processes and have been established to effectively detect certain defects. IR cameras have been especially effective at detecting porosity in electron beam powder bed fusion

* Corresponding author.
E-mail address: kfu26@wisc.edu (K. Fu).

https://doi.org/10.1016/j.jmapro.2025.06.027
Received 18 November 2024; Received in revised form 14 May 2025; Accepted 5 June 2025
Available online 17 June 2025

because pores have greater thermal emissivity than the solid material and therefore appear as bright spots in IR images [13].

Many researchers have used IR imaging to measure the effect of interlayer cooling time on porosity generation [14], while others have used it to detect the presence [15] and development of [16] porosity defects in both PBF-EB and laser powder bed fusion (PBF-LB) processes. To detect these pores, a variety of image processing techniques have been used to segment pores from the rest of the image. A commonly used method is image binarization with a static threshold [17]. This method is the easiest to implement, as the threshold is set to a single pre-specified pixel value, usually set by the user, for all images. Any pixel values above this threshold will be designated as part of a pore or air, while any pixel value below will not. Due to this ease of use, this method has also been used in a few different contexts, including to determine areas with temperatures above a desired value during a build process [18] and to determine regions of interest (ROI) for further analysis [19]. Another method is a statistical thresholding method. This method calculates the threshold for each image using the mean and standard deviation (σ) of the image pixel values. These statistical methods work like the Otsu threshold selection method [20], 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 less subjective. This method has been previously used with the threshold set to one σ above the mean pixel value to evaluate the effectiveness of using IR imaging for defect detection in PBF processes [21].

Finally, many researchers have developed machine learning (ML) methods to detect pores in IR images. However, these methods have also been typically used to detect a variety of other part defects (such as high surface roughness [22]) or have tried to predict the likely cause of the pores (such as lack of fusion or keyholing [23]). Some ML methods have tended to use more feature-based methods (such as identifying certain shapes) in addition to the more basic segmentation methods, such as thresholding or gradient detection [24]. Additionally, it should be noted that ML methods require significant computing power and training time.

Little research has been done to optimize and study the performance of the aforementioned segmentation approaches for porosity detection. Static thresholds were typically used for analyzing individual layers, as the threshold is easy to adjust, but the optimal threshold value was usually determined qualitatively and adjusted manually [14,18,19]. Statistical methods have only set their threshold to 1 σ above the mean pixel value and have not looked at how changing this parameter can affect detection accuracy [21]. Additionally, little research has been conducted comparing the effectiveness of these methods using ex-situ evaluation as ground truth [25]. Such evaluations have still only been qualitative in nature or used single-layer comparisons to evaluate the defect detection accuracy for whole components.

Some researchers have used XCT scanning to evaluate the validity of IR in situ imaging, but this has primarily been used qualitatively without objective, quantifiable metrics. XCT has been used to verify the presence of pores and other defects detected via IR images [26], as well as to observe how areas of increased thermal emissivity have led to the formation of an increased number of defects [27]. However, while this has been thoroughly and quantitatively researched for PBF-LB, these instances have been relatively qualitative and subjective in their evaluation for PBF-EB, basing their performance entirely on the operator's perspective (i.e. manual alignment, simple detection). There have been very few instances of XCT being used to objectively, quantitatively evaluate in-situ layer-wise IR imaging (i.e. measuring the size of pores, numerical comparison). One such instance involved overlaying point clouds of pores found in IR images over a point cloud of pores derived via XCT scans. This was then used to train a prediction model to detect lack of fusion porosity [28]. In other work, X-rays have been used in-situ to the manufacturing process instead of ex-situ to evaluate in-situ IR imaging to train an ML model for real-time porosity detection [29]. Nevertheless, these comparisons have either only been done qualitatively or using too coarse of a resolution 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 in PBF-EB. Finally, while there has been ample work with PBF-LB, the pore formation is significantly different between the two processes. In PBF-LB, pores are typically formed from valleys collapsing and trapping gas in the solid part [30]. Meanwhile in PBF-EB, pores are typically formed due to successive layers or improper heat dissipation [16]. Therefore, in IR images, the PBF-LB pores appear more as darker spots while the PBF-EB pores appear more as brighter spots.

This work will address how two low-level segmentation methods are affected by their primary parameter. The first method used a static threshold method for all images, where the threshold was set to various percentages of the maximum bit value. The second method used a statistical threshold for each image based on the mean and standard deviation of the grayscale values. Both methods were evaluated based on their accuracy of detecting pixels of pores and non-pores. Additionally, observations on how some external factors, such as porosity density and overall image brightness, affect detection ability were made. XCT scans with a sufficiently small voxel size were used as a ground-truth comparison to objectively evaluate the effectiveness of each method via precision and accuracy scores, and a rigorous, feature-based registration method is used to align the IR and XCT image data sets. The insights presented give future researchers greater understanding of the effectiveness of these methods when applied to in-situ IR data.

2. Methods

2.1. Sample design

The specimens used for this experiment were manufactured out of Ti-6Al-4 V powder on an *ARCAM Spectra L* PBF-EB machine. In total, 27 samples were manufactured with a layer height of 70 μm. While this height may vary due to various spreading inconsistencies, this is a common assumption in general PBF usage and literature. Future work can explore the effect these variations have on layer-wise defect detection methods. Porosity was created at varying levels by adjusting the focus offset (FO), as adjusting FO has a strong relationship to porosity formation [16]. Of these samples, 9 samples had an FO of 55 mA (labeled A1-A9), 9 had an FO of 65 mA (labeled B1-B9), and 9 had an FO of 75 mA (C1-C9). They were placed in the build volume in sets of 9 containing three of each FO (Fig. 1). The A samples were beamed first, followed by B and C. All samples were designed as cylinders with a flat vertical face (see Fig. 2) to facilitate registering the in-situ IR images to the ex-situ XCT scans. Registration refers to aligning special data from different sources for direct comparison (in this case, IR imaging and XCT scans). This process is further explained in Section 2.3.

2.2. IR image processing and segmentation methods

The in-situ images were taken at the end of each layer via the original equipment manufacturer (OEM) supplied IR camera (Prosilica GT). This camera used a prime lens with a focal length of 350 mm, an infrared-radiation sensor with a bit depth of 8 bits, a gamma value set to 1, an image size of 5120 × 5120 pixels, and a resolved pixel size of approximately 70 μm within the field of view. The IR images were distortion-corrected in MATLAB 2022b using the program's pinhole camera model for distortion correction. Once corrected, each image was cropped to a region of approximately 137 × 137 pixels (9.59 × 9.59 mm) 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. Finally, each segmentation method was performed on each image stack to generate a stack of binary-segmented images. For the static 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 (T)

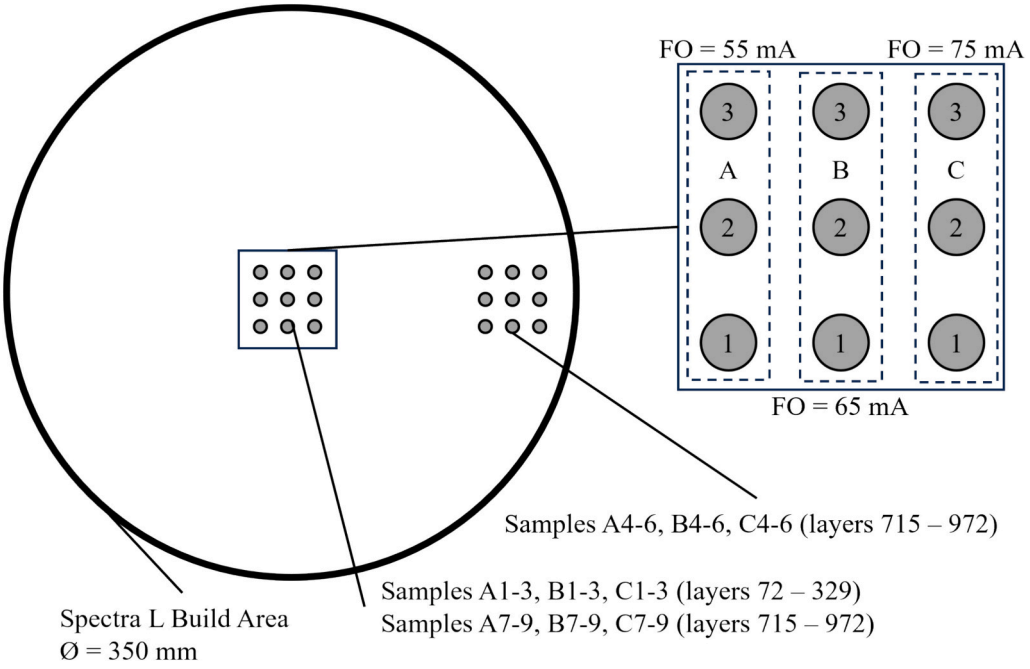

Fig. 1. Build layout.

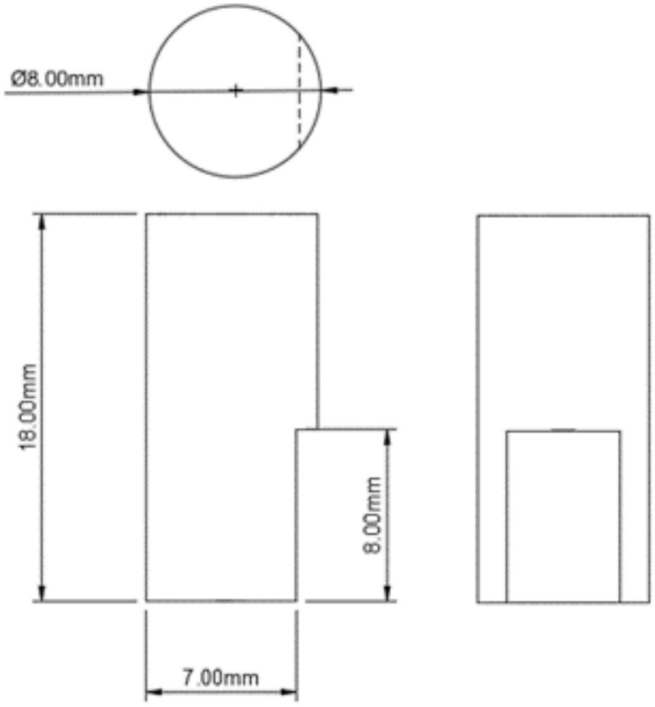

Fig. 2. Sample design.

varying from 0.5 to 0.95. For the statistical cutoff, the mean (μ) and standard deviation (σ) of the pixel grayscale values for each image were calculated and used to determine the threshold being μ + N * σ (with N ranging from 0 to 6), then they were binarized using the MATLAB imbinarize function. These ranges are listed in Table 1. Once the segmented image stacks were generated, they were cropped to be within the contours of each individual part. Finally, the images were resampled to match the smaller pixel size from the two-dimensional slices of the XCT reconstruction to allow for direct pixel-to-pixel comparisons between the XCT and IR data.

Table 1
Segmentation methods, parameters, and ranges.

Method	Parameter	Tested values/types
Static	Threshold (Percentage of Maximum Bit Depth, T)	0.5–0.95 (0.05 step size)
Statistical	Number of Standard Deviations Above Mean (N)	0.0 to 6.0 (0.5 step size)

2.3. XCT volume processing

Once manufactured, the samples were scanned in a *Zeiss Metrotom 800 XCT* machine. To ensure semi-repeatable sample placement within the CT volume, samples were placed in a nylon (low X-ray attenuation material) fixture during scanning. All samples were scanned using a voltage of 130 kV, current of 61 μA, a 0.25 mm copper prefilter, 1600 projections. Scans were reconstructed using a Feldkamp-David-Kress reconstruction algorithm and Shepp-Logan digital filter [31]. The XCT scan volumes were registered using MATLAB code developed by the research team to register the scan volume to the build coordinate system. First, volumes were converted to surfaces by thresholding the volume grayscale values via an ISO-50 threshold between the material and air grayscale value peaks. Registration was accomplished using datum features, such as points, axes, planes, etc., derived from surface data geometries [32] and can be used to align different surfaces for direct comparison (in this case, the surfaces created by IR and XCT data). The primary datum feature was the cylindrical section, the secondary datum was the vertical face, and the tertiary datum feature was the top face. For each of these datum features, surface points were sampled within a region of interest (ROI) that captured most of the feature, then basic geometric shapes were associated with the features via orthogonal least squares minimization to determine the datums. A cylinder was fit to the cylindrical sections, while planes were fit to the flat faces. A rotation matrix was determined that would align the axis of the primary datum, the cylinder, with the +Z direction. After applying this rotation, the normal vector of the secondary datum, the vertical face, was projected onto to the X-Y. A second rotation about the Z axis was then determined such that the normal vector would be aligned with the +X direction. Finally, a translation was applied to the twice rotated data

that would make the axis of the primary datum coincident with (X, Y) = (0,0) and the centroid of the tertiary datum, the top face, coincident with Z = 0. The two rotation matrices were composited and along with the translation matrix, the overall transformation was applied to the volumes via MATLAB imwarp3 function. Then, the registered volumes were sliced into image stacks along the *Z*-axis in MATLAB at each voxel.

To detect and segment porosity, these image stacks of the XCT data were processed using Weka Segmentation 3D, a machine learning image segmentation plug-in tool in ImageJ [33]. This program allows the user to directly label voxels in the input images to classify voxels for each class (in this case, air, part fixture, solid material, and pore classes, as shown in Fig. 3). A Gaussian blurring convolution was used to reduce noise influence and was calculated with window sizes of 1, 2, 4, and 8 voxel width. These convolutions were computed at each level as a series of feature volumes (difference of Gaussians, derivatives, structure, Laplacian, Hessian, mean, median, and variance) to expand the data set. The model was trained using a selection of images from each sample's image stack and used the integrated *FastRandomForest* algorithm, which is a re-implementation of a Random Forest classifier with speed and memory improvements. After each classification run, model predictions were manually examined and evaluated for accuracy. Errors were addressed by additional manual relabeling of voxels before retraining, and this process was repeated until satisfactory predictions, according to the operator, were produced by the model [34]. Finally, the XCT images were cropped to be within the part boundaries to eliminate edge segmentation errors.

2.4. XCT to IR comparison analysis

Once all the image stacks were processed, the segmented IR image stacks for each sample were compared directly to their XCT counterpart on a pixel-by-pixel level via a marker pixel placed in the center of the XCT and IR volumes for initial registration. This comparison was performed by image subtraction, where the pixel values of one image stack are subtracted from the value of the same pixel in the other. Based on the resulting image stack, each pixel was given a classification accuracy value represented by a new red-green-blue (RGB) color value. These color values were based on the following detection criteria: true positive (TP, green, IR detected pore where XCT detected pore), true negative (TN, gray, IR did not detect pore where XCT did not detect pore), false positive (FP, blue, IR detected pore where XCT did not detect pore), and false negative (FN, red, IR did not detect pore where XCT did detect pore), as shown in Fig. 8 and Fig. 12 in Sections 3.2 and 3.3, respectively. The black pixels represent pixels outside the sample boundary.

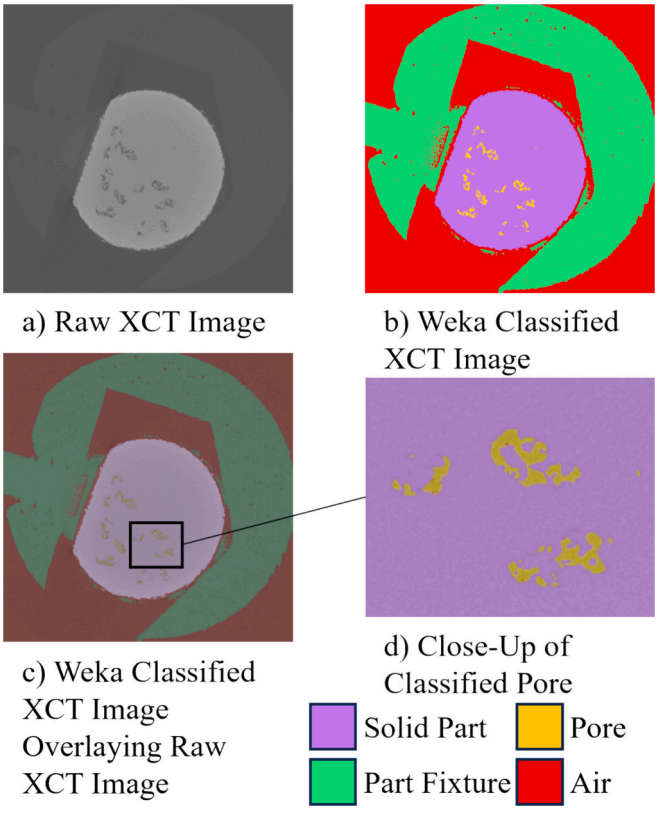

Fig. 3. Weka classification example with a) raw XCT image, b) classified XCT image, c) Weka classified image overlaying raw XCT image, and d) close-up view of classified pore overlaying raw XCT image.

3. Results

3.1. Infrared and X-Ray computed tomography imaging results

Since both segmentation methods depend directly on the grayscale values of the IR images, histograms were used alongside the raw IR images to gain preliminary understanding of the images. In Fig. 4, three examples of raw IR images are shown alongside their image histograms. These images were selected from specimens that represent the three different FO values, sample A5 in Fig. 4a) with FO = 55 mA, Sample B2 in Fig. 4b) with FO = 65 mA, and sample C5 in Fig. 4c) with FO = 75 mA). As expected, the IR data shows an increase in FO resulted in an increase in porosity. The bright dot in the center of each image is the reference pixel, which was used to properly register and align the IR and XCT images. Additionally, another aspect to note is that the overall brightness varied between the samples due to the thermal cycling inherent to the additive process, which may have influenced the effectiveness of the segmentation methods. This can be seen via both the average pixel values and the distribution of gray values on the histograms. Brighter images were distributed on the higher end of the histograms and therefore had a higher average pixel value. Finally, the pores appeared to be the brightest parts of the image, which agrees with the previously discussed literature. The next brightest parts appear to be the contours, followed by the solid material and lastly the surrounding semi-sintered powder. This corresponds to the grayscale value peaks in the histograms, with surrounding semi-sintered being the leftmost peak, followed by the solid material being the second, and the contour and pore peak being last. However, these peaks vary in their magnitude and variation from the rest of the grayscale values, making simple peak detection (such as in the Otsu method) difficult. Since a Blackbox calibration was not performed, this analysis is a comparative study, not absolute.

As described in Section 2.3, XCT scanning was used as a ground-truth measurement to evaluate the effectiveness of each segmentation method. Cross sections of Samples A5, B2, and C5 are shown as examples of results of these scans in Fig. 5. The additional object in the image from the scans for Samples B2 and C5 is the fixture used for scanning. The changes in FO also resulted in varying levels of detectable porosity, which allowed the XCT scans to be used as ground-truth reference for the IR images. As expected from the literature [16] and confirmed in both the IR and XCT datasets, the increase in FO resulted in an increase in porosity, allowing for analysis at different porosity levels.

Before performing the objective comparison, the rescaled IR images and registered XCT images were qualitatively compared to verify the success of the IR segmentation, XCT segmentation, and registration processes. Fig. 6 shows an example of this comparison for a single cross section of Sample B2. It is important to note that the pores may have had a slightly different appearance in the IR and XCT image stacks. In the IR images, the pores were larger and more bulbous, due to heat dissipation and coarser image resolution. Meanwhile in the CT images, the pores were more irregular in shape due to the finer resolution. However, while appearing different in shape, the pores in both image sets were still in similar locations. Nonetheless, evaluating the performance of the

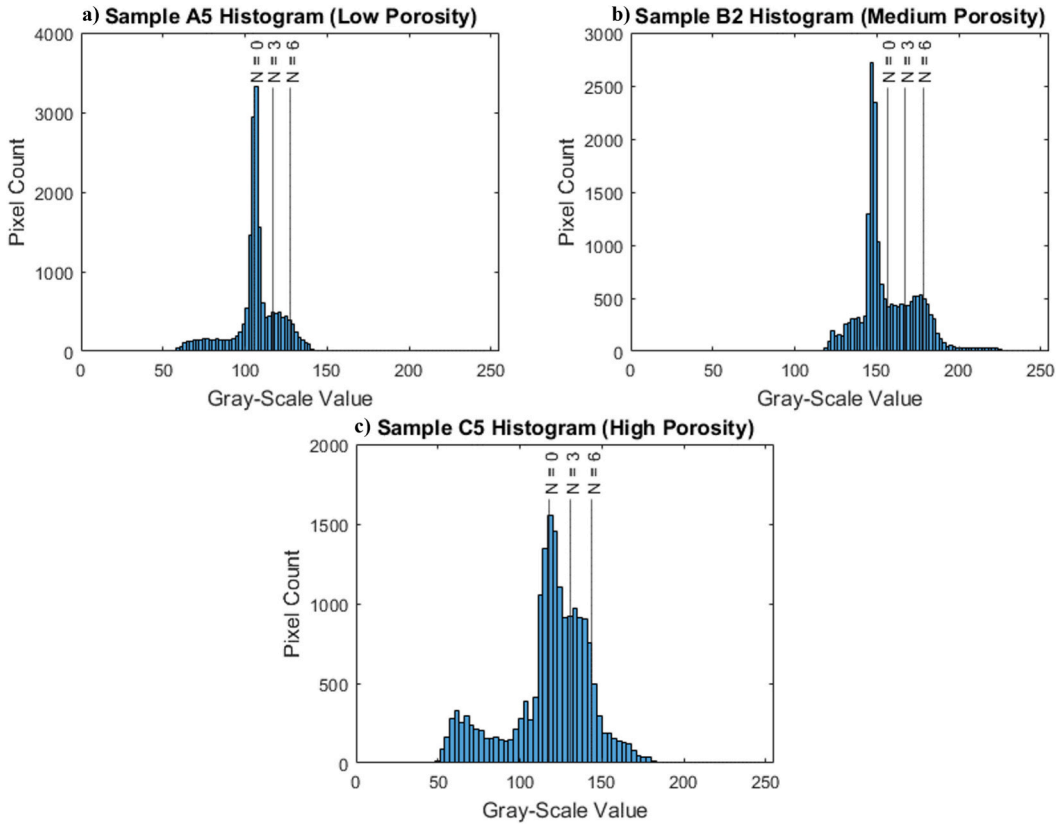

Fig. 4. Distortion-Corrected IR images and histograms for a) sample A5 (FO = 55 mA, low porosity, average gray value = 106), b) sample B2 (FO = 65 mA, medium porosity, average gray value = 156), and c) sample C5 (FO = 75 mA, high porosity, average gray value = 118).

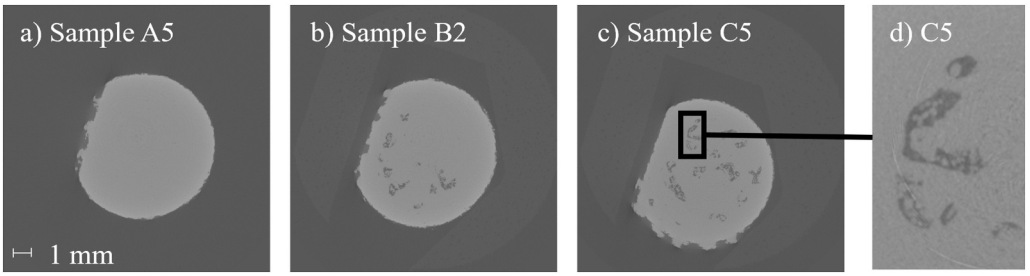

Fig. 5. Raw XCT images for a) sample A5 (low porosity), b) sample B2 (medium porosity), and c) sample C5 (high porosity). d) Detail image of a pore from sample C5.

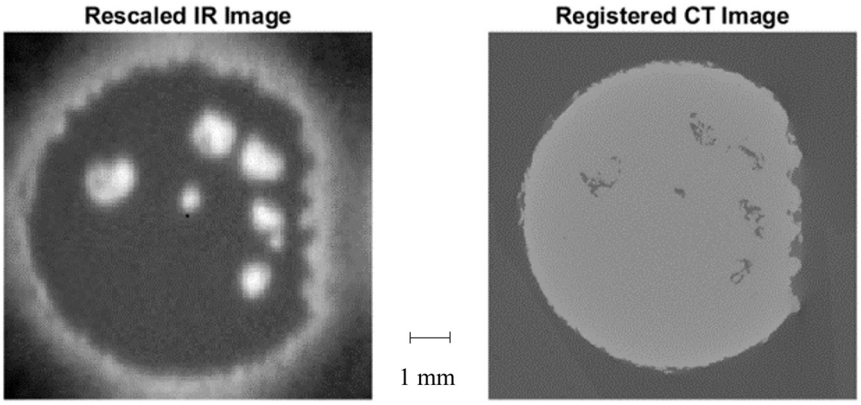

Fig. 6. Side-by-side comparison of rescaled IR image and registered CT image for sample B2.

analysis methods needed to account for this difference. This is more thoroughly discussed in the Discussion section.

3.2. Static thresholding segmentation

Fig. 7 shows the results of performing the static thresholding segmentation analysis on a single cross section of Samples A5, B2, and C5. In these resulting images, pixels classified as a pore are white, while pixels classified as not a pore are black. The pixel in the center of each image is the reference pixel, which was used to align the XCT and IR image stacks. As can be seen, the threshold value influenced the results of the segmentation. When the threshold was set high enough, the resulting segmentation classified everything as not a pore (called blacking-out, seen in $T = 0.80$ of Fig. 7a). When the threshold was set low enough, the resulting segmentation classified everything as a pore (called whiting-out, seen in $T = 0.50$ of Fig. 7b). Since these blacking-out and whiting-out conditions typically happened below T = 0.5 and above $T = 0.85$, respectively, results outside the T = 0.5 to 0.85 range were not analyzed further.

To perform the evaluation of the IR images using the CT images, each pair of images from each stack were overlayed, and each pixel was colored based on which of the 4 evaluation conditions (described in Section 2.4) applied. Fig. 8 shows an example of these colorized results comparing static IR and XCT images for a single cross section of Sample B7. As can be seen, the lowest threshold values were more likely to oversize the pores, which led to increased FP and decreased TN values and decreased performance. However, the opposite can be seen in comparison images with the highest threshold values. These high-threshold images were more likely to undersize (or not even detect) pores, which lead to increased FN and decreased TP values and decreased performance.

To get an understanding of how each evaluation condition, i.e., TP, FP, TN, FN, varied with the parameter value, the average area of all samples for each evaluation condition was plotted for each parameter value with bounds of 1 standard deviation above and below (Fig. 9). This was determined by counting the pixels from the colorized comparison images for each condition, then multiplying that count by the area of each pixel (12 μm × 12 μm). One observation to note is the TN area was at least an order of magnitude higher than the TP. This was likely due to how most of the part in the image is material, meaning there was significantly more area associated with non-pore material than pore. This resulted in significantly imbalanced classes. Therefore, the G-Mean metric, which uses TP and TN proportions for calculations, was chosen to measure classification performance over other metrics (such as classification accuracy). Otherwise, the quantity of TN pixels would skew the results if other performance metrics were used.

For the static segmentation method, the threshold value had a significant impact on the area of detected pores for each condition. Both TP and FP decreased significantly over the range of $T = 0.5$ to $T = 0.85$, with both conditions reaching 0 mm^2 by T = 0.85 after starting at 0.45 mm^2 and 10 mm^2, respectively. Meanwhile, the TN and FN values both increased over the range, with TN going from 19 mm^2 to 30 mm^2 and the FN going from 0.05 mm^2 to 0.5 mm^2. Additionally, the standard deviation bounds for the TP, TN, and TP all converged to the average values by T = 0.8, whereas the standard deviation bounds for the FN did the opposite and increased to the same magnitude as the average. It should be noted these metrics are averages over all images of all samples, averaging-out environmental aspects (such as brightness). The standard deviations partially incorporated the effects of these environmental aspects, as these aspects result in higher variance and therefore larger standard deviations, but there were other factors that also resulted in higher variance. However, these metrics were still good summary metrics alongside the qualitative results, such as those in Fig. 7 and Fig. 8, since the qualitative results provided explanations for the higher variances in the metric values.

To quantitatively evaluate how threshold (T) affected porosity detection, a G-Mean metric was calculated at each T value and plotted in Fig. 10. G-Mean is a metric calculated using the Sensitivity and Specificity, shown in Eqs. (1)–(3), which are proportions of positive and negative classes. This made the G-Mean metric ideal for this case because of the large imbalance between the negative and positive classes. For this calculation, the number of pixels for each of the four evaluation conditions (true positive, true negative, false positive, false negative) was counted for each image slice of each sample, and the totals for each parameter were summed for the plot. These scores range from 0 to 1, with 1 meaning perfect detection. For the static cutoff, the G-Mean increased from $T = 0.5$ to $T = 0.55$, maximizing at 0.79. After T = 0.55, the G-Mean steadily decreased as T continued to increase, reaching

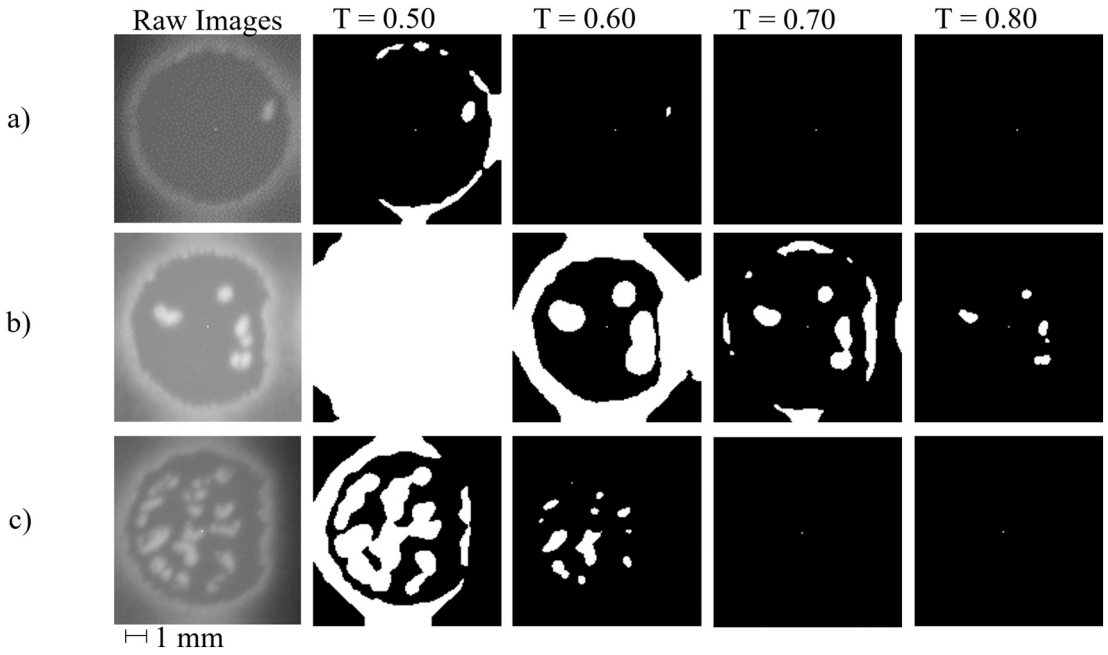

Fig. 7. Static segmentation results for a) layer 118 (height = 8.26 mm) of sample A5 (FO = 55 mA, low porosity), b) layer 100 (height = 7.0 mm) of sample B2 (FO = 65 mA, medium porosity), and c) layer 40 (height = 2.8 mm) of sample C5 (FO = 75 mA, high porosity) compared to their raw IR images.

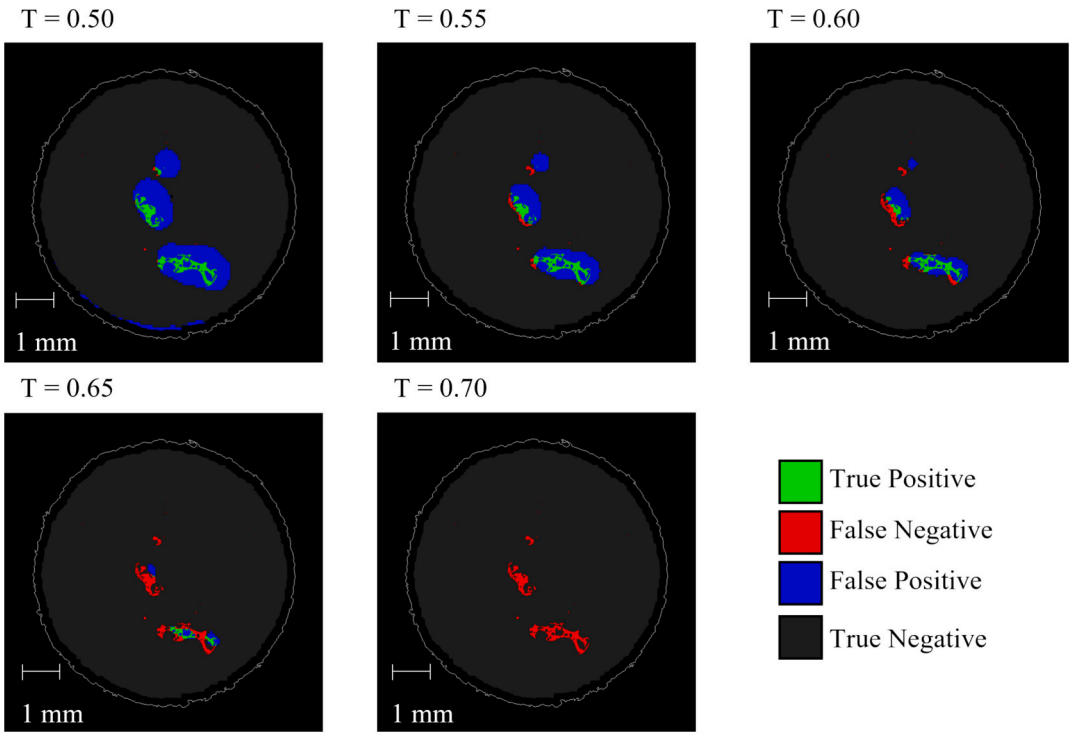

Fig. 8. Colorized comparison of static segmentations applied at height = 10.92 mm on sample C7 with various threshold values.

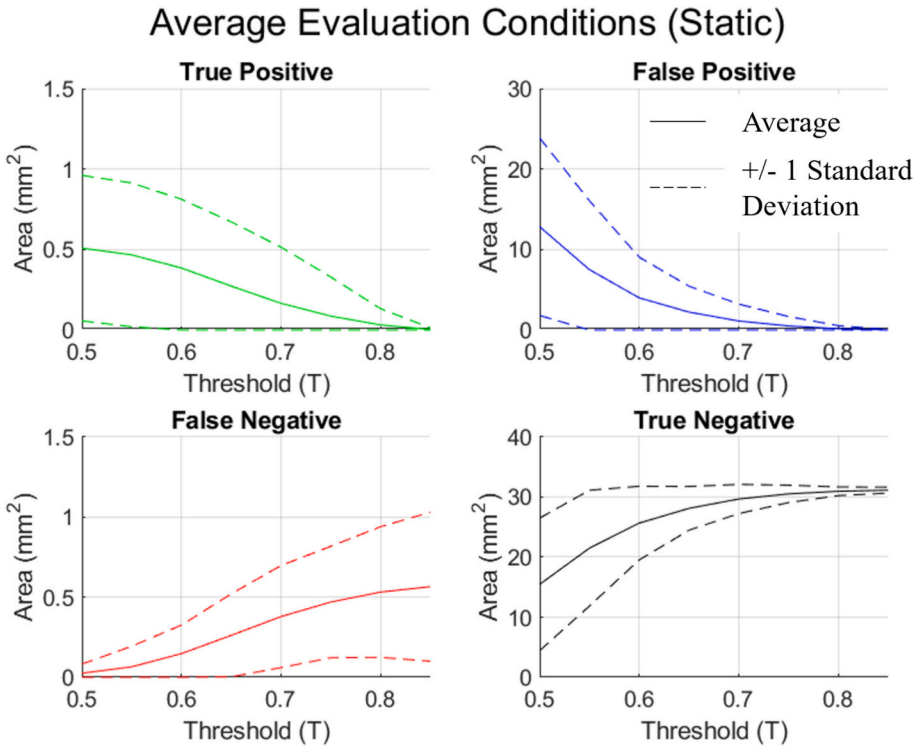

Fig. 9. Average evaluation condition count (solid) +/− 1 standard deviation (dashed) for static segmentation methods.

0.07 at $T = 0.85$. This matched what was qualitatively observed in Fig. 7 and Fig. 8, as the increase in T from 0.5 removed any whiting-out over-classifying as well as some pore oversizing. After 0.55, the pores began to be undersized, eventually leading to blacking-out under-classifying. This also quantitatively justified the previous decision to not analyze images with T greater than 0.85, as the trend showed those G-Mean values would have been close to 0.

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

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

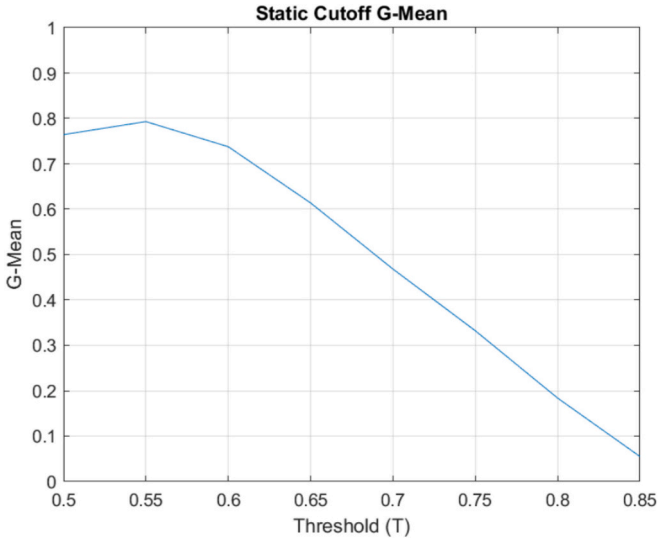

Fig. 10. G-Mean for static segmentation method.

$$G-Mean = \sqrt{Sensitivity \times Specificity} \quad (3)$$

3.3. Statistical thresholding segmentation

Fig. 11 shows the results of performing the statistical segmentation on Samples A5, B2, and C5. Similar to the static thresholding segmentation method, the tested parameter (in this case, N) influenced the detection of porosity. If the value of N was too low (such as with $N = 0$ in Fig. 11a), the segmentation misidentified some material pixels in the image as pore pixels, and if the value of N was too high (such as with $N = 6$ in Fig. 11c), the segmentation misidentified pore pixels in the image as material pixels. While there were noticeable differences between the different values of N, the differences were less obvious compared to the other methods. Compared to the static method, the thresholds were closer together for statistical methods than static and located between the material and pore peaks, shifting with the location and distribution of the gray values. Therefore, determining the optimal threshold qualitatively is more difficult, and quantitative evaluation is required.

As with the static segmentation method, a set of colorized comparison images were generated comparing statistical IR and XCT image stacks (Fig. 12). As can be seen, when the N value was set closer to 0, the images were overclassified which increased the FP and decreased the TN values. This was like the static method comparison images (Fig. 8). However, unlike the static method, within the N value range, the statistical method comparison images never reach a point of full underclassification due to the smaller change in threshold value per change in parameter value. Additionally, while there was significant difference between the images with $N = 0$ and $N = 1.5$, the difference between images became less significant as N increased (such as between $N = 4.5$ and $N = 6$).

To see how the N-value affected the average area (+/− 1 standard deviation) for each evaluation condition, these conditions were plotted in Fig. 13. As with the static segmentation, the true negative area was significantly higher than any of the other conditions. However, unlike the static methods, the true positive area always remained above the false positive count. Additionally, the TP and FN values decreased and increased, respectively, linearly but shallowly. The TP values decreased from 0.4 to 0.26 mm^2, while the FN values increased from 0.1 to 0.24 mm^2 in the range of $N = 0$ to $N = 6$. Meanwhile, the TN and TP values had very different behavior. The TN increased logarithmically, starting at 15 mm^2 and leveling out at 27 mm^2 by $N = 2$, while the FP decreased logarithmically, starting at 15 mm^2 and leveling out at 3 mm^2 by $N = 2$. Finally, the standard deviation bounds were mostly constant with respect to the average throughout the full range.

As was done with the static segmentation, G-Mean values were calculated for the statistical segmentation at each N value (plotted in Fig. 14). Results were plotted on the same axes as Fig. 10 for direct comparison. The relationship between N and G-Mean values were relatively concave down, though it was not symmetric. From N = 0 to N = 2, the G-Mean increased from 0.6 to a maximum of 0.78 before it decreased to 0.7 at N = 6. While somewhat parabolic, the relationship is not symmetric, as the increase from N = 0 to N = 2 is significantly larger and more parabolic than the decrease from N = 2 to N = 6, which is also

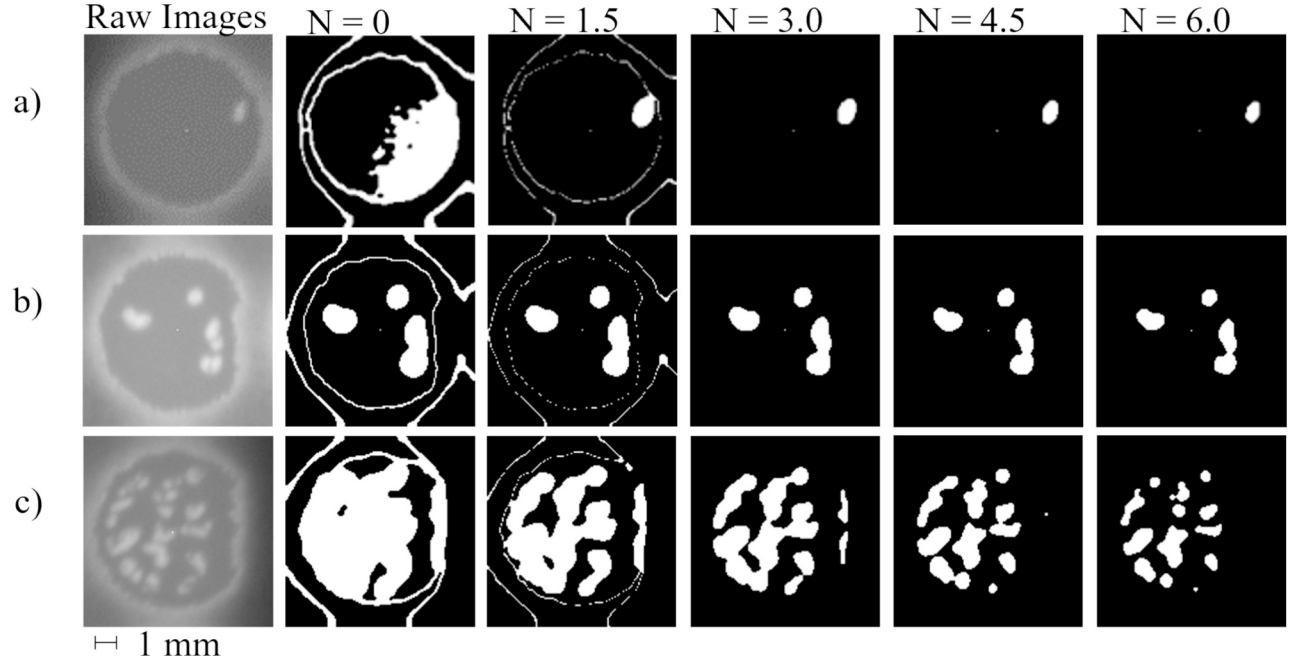

Fig. 11. Statistical segmentation results for a) layer 118 (height = 8.26 mm) of sample A5 (FO = 55 mA), b) layer 100 (height = 7.0 mm) of sample B2 (FO = 65 mA), and c) layer 40 (height = 2.8 mm) of sample C5 (FO = 75 mA) compared to their raw IR images.

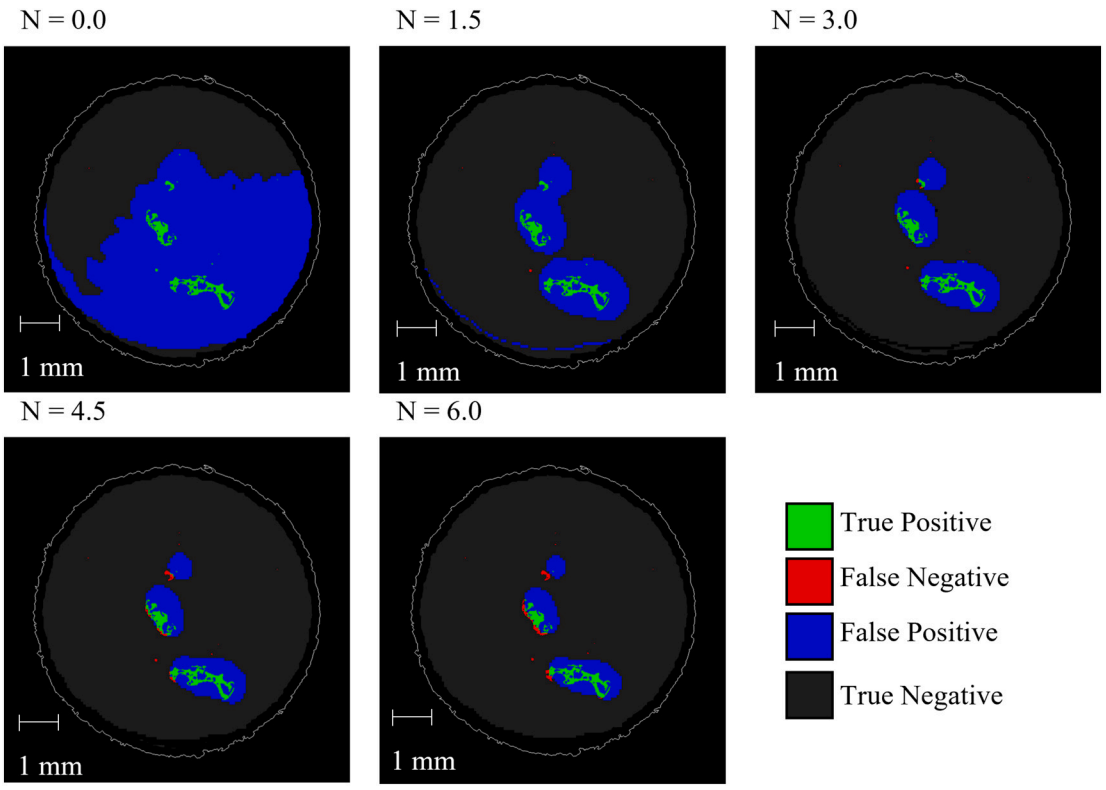

Fig. 12. Colorized comparison of statistical segmentations applied at height = 10.92 mm on sample B7 with various N-values.

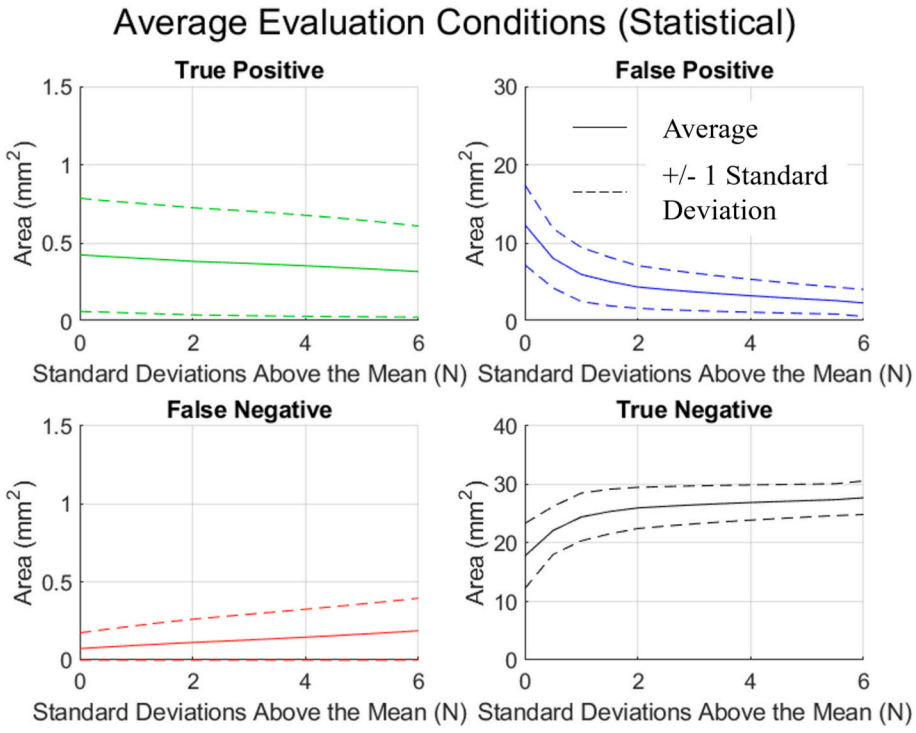

Fig. 13. Average evaluation condition area (mm^2) (solid) +/− 1 standard deviation (dashed) for statistical segmentation methods.

significantly more linear.

4. Discussion

When evaluating these methods, one aspect that must be addressed first is the difference between the appearance of pores in the IR images and the XCT images. This can be seen in the side-by-side comparison in Fig. 6. As previously stated, in IR images, pores appear as bright spots because the unmelted powder in the pore areas absorbed less heat energy than the solid regions, therefore emitting more heat radiation.

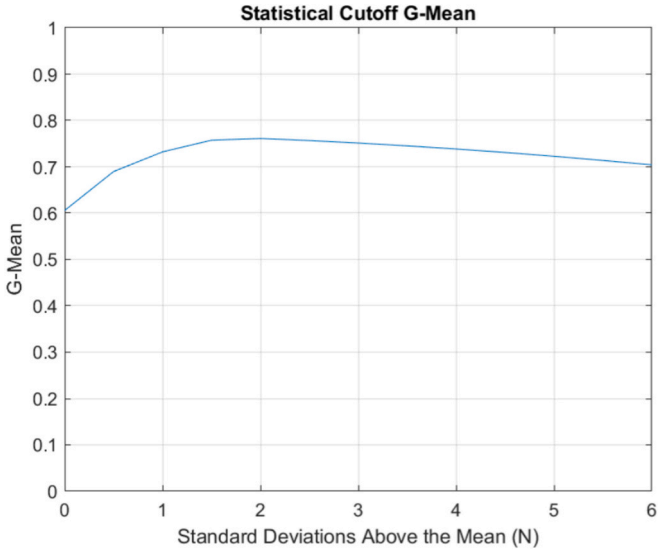

Fig. 14. G-Mean of statistical segmentation method.

Since heat spreads from areas of high temperature to areas of low temperature, the heat energy from the pore-forming areas conducts into neighboring areas without pores. This causes the surrounding non-pore areas to irradiate more heat and the pores to appear larger in the IR images. Therefore, the pores may have changed apparent size and shape by the time they were XCT scanned. This has been documented in literature [10] [35] and is why many detection algorithms focus on detecting the presence and spatial distribution of porosity instead of the exact size [19]. This is likely a factor as to why the G-Mean magnitudes were 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 for PBF-EB. However, the magnitudes are still representative of performance and the trends from the G-Mean can still be utilized to see the influence each method's parameter had on its performance.

For the static threshold segmentation method, the threshold value appeared to have a significant effect on detection accuracy (Fig. 8). When the threshold value was too low (as seen when set to 0.50), the cutoff was below the gray values of the fused part. This led to the segmented image being whited-out and over-classified, resulting in a significantly higher FP rate. When the threshold value was too high (as seen when set to 0.80 or higher), the segmented image was blacked-out and under-classified. Therefore, some pores were either not detected or were represented as significantly smaller than they were, resulting in a significantly higher FN rate. When finding an optimal threshold, this method was very sensitive to the threshold parameter, meaning small changes in the threshold value led to large changes in porosity representation. Based on the G-Mean, a threshold of 0.55 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 FP and decreasing the threshold to minimize FN.

Additionally, this method appeared 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 sometimes resulted 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.

When observing the statistically-thresholded images, such as those in Fig. 11, increasing the N value significantly changed the resulting segmented images. This was most prevalent between lower values of N, such as between $N = 0.0$ and $N = 1.5$. However, this change was less prevalent as N increases, 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 filtered out consistently between $N = 1.5$ and $N = 3.0$, unaffected by porosity levels. This suggests 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 appeared to influence detection accuracy. With smaller N values, specifically less than two, the thresholded images oversized the pores, with increasing N leading to less oversizing. However, after an N value of two, the images began to underestimate the size of the pores. The quantitative results from the G-Mean plotted in Fig. 14 match these qualitative findings. The detection accuracy increased from $N = 0$ to $N = 2$ as the FP rate decreased from significantly oversizing the pores. After $N = 2$, the detected pores began to undersize and lead to a higher FN rate. However, since the number of TN pixels was significantly larger than the number of TP pixels, the increased FN rate had less of an impact than the increased FP on the porosity detection, which is seen by the smaller decrease from $N = 2$ to $N = 6$ compared to $N = 0$ to $N = 2$. This has expanded the insights of previous work using statistical methods [10], which just used $N = 1$ for porosity detection. As these results show, adjusting N can result in greater porosity detection to best optimize for the given set of process conditions. Based on the results, N should be adjusted to be equal to 2 for optimal results detecting porosity.

The statistical thresholding did seem to struggle slightly with dimmer images (such as the ones seen in Sample A5). Based on the build layout, these images were likely dimmer due to the corresponding samples having been beamed before others. Therefore, the samples in these images had more time to dissipate heat 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 determining the average and σ 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 still had an impact on the classification.

Additionally, this method is more computationally expensive than a static threshold because a new threshold is calculated for each image instead of using the same threshold for all. The intensity value of each pixel in the image is considered when determining the μ and σ. 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.

To compare the two methods directly, the parameters for each method were normalized using the initial parameter value and relevant ranges for each parameter using Eq. (4). Since the static thresholds were based on the range of possible bit values, the total range was 0 to 1. However, since the pores were brighter in the images, only the top half of the range was used, making the static parameter range 0.5. For the statistical method, the standard deviations are based on the distribution of pixel values on each image. Unfortunately, since the distribution varied between images and a standard deviation assumes a unimodal, normal distribution, the number of standard deviations between the average and maximum pixel values varied significantly between images. The average difference was 10, so the total range of N was set to −10 to 10 and therefore the parameter range was 0–10, with an initial N value of 0.

$$\text{Normalized Parameter} = \frac{\text{Parameter Value} - \text{Initial Parameter Value}}{\text{Parameter Range}} \quad (4)$$

The G-Mean values for each method with the normalized parameters were plotted together in Fig. 15. Out of the two methods, the static segmentation method appears to be more sensitive to its adjustable parameter, while the statistical segmentation appears to be less sensitive. While the statistical segmentation varied from changing the N

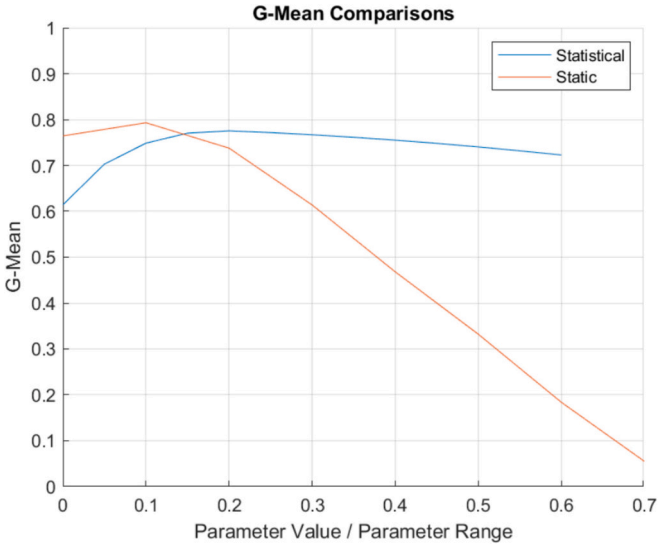

Fig. 15. Normalized parameter plots of both static and statistical methods.

parameter, the change from each adjustment resulted in a smaller change in threshold value and is less significant than the static segmentation. This can be quantitatively observed by the static segmentation G-Mean values ranging from 0 to 0.80 across the parameter range, while the statistical stayed within the range of 0.6 to 0.8 throughout the entire parameter range. This matched the qualitative observations from Fig. 7 and Fig. 11, as the static 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 static method, as the tighter window of possible results means that outside influences are less likely to affect the detection of porosity.

One reason for having a robust detection algorithm is that images can vary in overall brightness. These changes in brightness can 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 had images that were significantly brighter than others. Set thresholds, like the ones in the static method, do not take this into account, as they simply look at pixels above a certain cut-off value. Statistical methods can account for this, as each layer's image has its own threshold based on the intensity of the pixel values. Additionally, because of this, the statistical method is more adaptable to these external factors, making it easier to determine the highest-performing values for a variety of conditions.

Overall, both methods performed similarly at optimal parameter settings, with the static method's best performance at $T = 0.55$ and the statistical method's best performance at $N = 2$. This means the pore pixel values on the histogram were around 55 % of the bit depth regardless of different external factors (porosity density, overall brightness, etc.) for this set of build conditions. Therefore, it can be inferred that detecting porosity for a set of build conditions has an optimal static threshold that can be easily implemented. However, the statistical method performed better at a larger variety of parameter settings. This makes this method more robust and reliable under a variety of different N values, meaning the optimal parameter value is less critical for optimization. Based on these results, it can be inferred this method would be better for detecting porosity for an untested set of build conditions.

5. Conclusions

In this work, two different types of basic segmentation methods (static thresholding and statistical thresholding) were analyzed for their effectiveness in detecting and modeling porosity in in-situ IR images in PBF-EB processes. Both 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:

- The static method produced the best results when the threshold was set between 55 % of the maximum pixel intensity value. With an optimal threshold value, porosity was most accurately detected and modeled. However, the tolerance on this value was small, as small changes in the threshold value resulted in a significant decrease in classification accuracy.
- The statistical method produced the best results when the imaging threshold was set to 2σ above the mean. This method showed the least amount of variance between results from both internal and external factors. Of the methods assessed, this method most accurately detected and modeled porosity, and the tolerance for optimal N was much larger than the parameters for the other 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, played a significant role in detection accuracy and finding optimal parameters. However, while this influenced the results in both methods, the statistical method showed the most resilience and still performed best through a variety of brightness conditions.

Along with showing how each method performed with different parameter settings, this work also provided some insight into the distribution of grayscale values in these infrared images. As previously discussed, this distribution was affected by a variety of factors. The next step would be to look at changing the distribution of grayscale values, such as through image filtering. The goal of this would be to give more distinguished classification peaks, pushing intermediate pixels towards one or the other. These may provide additional information on factors that affect grayscale distribution in these images and how to mitigate any detrimental effects they have on the accuracy of each method to detect porosity.

CRediT authorship contribution statement

Brian Johnstone: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jaime Berez:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Caroline Massey:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Elliott Jost:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Christopher Saldana:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Katherine Fu:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

Funding

This work was funded by the United States Department of Energy, under the Enhanced Preparation for Intelligent Cybermanufacturing Systems grant (DE-EE-0008303), and by Eaton Aerospace.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kim FH, Moylan SP. Literature review of metal additive manufacturing defects. NIST Adv Manuf Ser 2018:1–17 [Online]. Available: http://nvlpubs.nist.gov/nistpubs/ams/NIST.AMS.100-16.pdf.

[2] Edwards P, O'Conner A, Ramulu M. Electron beam additive manufacturing of titanium components: properties and performance. J Manuf Sci Eng 2013;135(6):1–7. https://doi.org/10.1115/1.4025773.

[3] Bisht M, Ray N, Verbist F, Coeck S. Correlation of selective laser melting-melt pool events with the tensile properties of Ti-6Al-4V ELI processed by laser powder bed fusion. Addit Manuf 2018;22(May):302–6. https://doi.org/10.1016/j.addma.2018.05.004.

[4] du Plessis A, Yadroitsava I, Yadroitsev I. Effects of defects on mechanical properties in metal additive manufacturing: a review focusing on X-ray tomography insights. Mater Des 2020;187:108385. https://doi.org/10.1016/j.matdes.2019.108385.

[5] Cunningham R, Narra SP, Montgomery C, Beuth J, Rollett AD. Synchrotron-based X-ray microtomography characterization of the effect of processing variables on porosity formation in laser power-bed additive manufacturing of Ti-6Al-4V. Jom 2017;69(3):479–84. https://doi.org/10.1007/s11837-016-2234-1.

[6] McCann R, Obediди M, Hughes C, McCarthy E, Egan D, Vijayaraghavan R, et al. In-situ sensing, process monitoring and machine control in laser powder bed fusion: a review. Addit Manuf 2021;45(November 2020). https://doi.org/10.1016/j.addma.2021.102058.

[7] Reijonen J, Revuelta A, Riipinen T, Ruusuvuori K, Puukko P. On the effect of shielding gas flow on porosity and melt pool geometry in laser powder bed fusion additive manufacturing. Addit Manuf 2020;32. https://doi.org/10.1016/j.addma.2019.101030.

[8] Craeghs T, Clijsters S, Kruth J. Online quality control of selective laser melting. Phys Rev E 2011:212–26 [Online]. Available: http://www.ainfo.inia.uy/digital/bitstream/item/7130/1/LUZARDO-BUIATRIA-2017.pdf.

[9] Forien J-B, Calta NP, DePond PJ, Guss GM, Roehling TT, Matthews MJ. Detecting keyhole pore defects and monitoring process signatures during laser powder bed fusion: a correlation between in situ pyrometry and ex situ X-ray radiography. Addit Manuf 2020:101336. https://doi.org/10.1016/j.addma.2020.101336.

[10] Yang T, Liao W, Wei H, Zhang C, Chen X, Zhang K. Effect of processing parameters on overhanging surface roughness during laser powder bed fusion of AlSi10Mg. J Manuf Process 2021;61(June 2020):440–53. https://doi.org/10.1016/j.jmapro.2020.11.030.

[11] Layerwise Automated Visual Inspection in Laser Powder-Bed Additive Manufacturing. 19, 2015. doi:https://doi.org/10.1115/MSEC2015-9393.

[12] Moylan S, Whitenton E, Lane B, Slotwinski J. Infrared thermography for laser-based powder bed fusion additive manufacturing processes. In: 40TH ANNUAL REVIEW OF PROGRESS IN QUANTITATIVE NONDESTRUCTIVE EVALUATION: Incorporating the 10th International Conference on Barkhausen Noise and Micromagnetic Testing; Oct. 2014. p. 1191–6. https://doi.org/10.1063/1.4864956.

[13] Krauss H, Zeugner T, Zaeh MF. Layerwise monitoring of the selective laser melting process by thermography. Phys Procedia Sep. 2014;56:64–71. https://doi.org/10.1016/j.phpro.2014.08.097.

[14] Gong H, Rafi K, Karthik NV, Starr T, Stucker B. "Defect morphology in Ti-6Al-4V parts fabricated by Selective Laser Melting and Electron Beam Melting," *24th Int. SFF Symp. - An Addit. Manuf. Conf. SFF 2013*, no. July 2015. 2013. p. 440–53. https://doi.org/10.26153/tsw/15566.

[15] Williams RJ, Pigline A, Ronneberg T, Jones C, Pham M, Davies C, et al. In situ thermography for laser powder bed fusion: effects of layer temperature on porosity, microstructure and mechanical properties. Addit Manuf 2019;30(October):100880. https://doi.org/10.1016/j.addma.2019.100880.

[16] Cordero ZC, Dinwiddie RB, Immel D, Dehoff RR. Nucleation and growth of chimney pores during electron-beam additive manufacturing. J Mater Sci 2017;52(6):3429–35. https://doi.org/10.1007/s10853-016-0631-z.

[17] Russ JC, Neal FB. The Image Processing Handbook. 2017.

[18] Rodriguez E, Mireles J, Terrazas CA, Espalin D, Perez MA, Wicker RB. Approximation of absolute surface temperature measurements of powder bed fusion additive manufacturing technology using in situ infrared thermography. Addit Manuf 2015;5:31–9. https://doi.org/10.1016/j.addma.2014.12.001.

[19] Mahmoudi M, Ezzat AA, Elwany A. Layerwise anomaly detection in laser powder-bed fusion metal additive manufacturing 2019;141(March):1–13. https://doi.org/10.1115/1.4042108.

[20] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979;C(1):62–6.

[21] Lough CS, Wang X, Smith C, Landers R, Bristow D, Drallmeier J, et al. Correlation of SWIR imaging with LPBF 304L stainless steel part properties. Addit Manuf 2020;35(May):101359. https://doi.org/10.1016/j.addma.2020.101359.

[22] Bartlett JL, Heim FM, Murty YV, Li X. In situ defect detection in selective laser melting via full-field infrared thermography. Addit Manuf 2018;24(July):595–605. https://doi.org/10.1016/j.addma.2018.10.045.

[23] Kozjek D, Carter FM, Porter C, Mogonye JE, Ehmann K, Cao J. Data-driven prediction of next-layer melt pool temperatures in laser powder bed fusion based on co-axial high-resolution Planck thermometry measurements. J Manuf Process 2022;79(April):81–90. https://doi.org/10.1016/j.jmapro.2022.04.033.

[24] Scime L, Beuth J. Using machine learning to identify in-situ melt pool signatures indicative of flaw formation in a laser powder bed fusion additive manufacturing process. Addit Manuf 2019;25(October 2018):151–65. https://doi.org/10.1016/j.addma.2018.11.010.

[25] de Winton HC, Cegla F, Hooper PA, de Winton HC, Cegla F, Hooper PA. A method for objectively evaluating the defect detection performance of in-situ monitoring systems. Addit Manuf 2021;48:102431. https://doi.org/10.1016/j.addma.2021.102431.

[26] Croset G, Martin G, Josserond C, Lhuissier P, Blandin JJ, Dendievel R. In-situ layerwise monitoring of electron beam powder bed fusion using near-infrared imaging. Addit Manuf Feb. 2021;38. https://doi.org/10.1016/j.addma.2020.101767.

[27] Mohr G, Altenburg S, Ulbricht A, Heinrich P, Baum B, Maierhofer C, et al. In-situ defect detection in laser powder bed fusion by using thermography and optical tomography—comparison to computed tomography. Metals (Basel) Jan. 2020;10(1). https://doi.org/10.3390/met10010103.

[28] Coeck S, Bisht M, Plas J, Verbist F. Prediction of lack of fusion porosity in selective laser melting based on melt pool monitoring data. Addit Manuf 2019;25(October 2018):347–56. https://doi.org/10.1016/j.addma.2018.11.015.

[29] Ren Z, Gao L, Clark S, Fezzaa K, Shevchenko P, Choi A, et al. Machine learning-aided real-time detection of keyhole pore generation in laser powder bed fusion. Science 2023:89–94. https://doi.org/10.1126/science.add4667.

[30] Martin AA, Calta NP, Khairallah SA, Wang J, Depond PJ, Fong AY, et al. Dynamics of pore formation during laser powder bed fusion additive manufacturing. Nat Commun 2019;10(1):1–10. https://doi.org/10.1038/s41467-019-10009-2.

[31] Shepp LA, Logan BF. Fourier reconstruction of a head section. IEEE Trans Nucl Sci 1974;NS-21(3):21–43. https://doi.org/10.1109/tns.1974.6499235.

[32] Anon. Dimensioning and tolerancing. In: ANSI Stand, 2018, no. Y14, 5; 1973. https://doi.org/10.3139/9781569908167.017.

[33] Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri K, Schindelin J, Cardona A, et al. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 2017;33(15):2424–6. https://doi.org/10.1093/bioinformatics/btx180.

[34] Miers JC, Moore DG, Saldana C. Defect evolution in tensile loading of 316L processed by laser powder bed fusion. Exp Mech 2022;62(6):969–83. https://doi.org/10.1007/s11340-021-00815-5.

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