XCTPore: An Open Source Database for Porosity in X-ray CT scanned Components

4 th Singapore International Non-destructive Testing Conference and Exhibition (SINCE2022)

XCTPore: An Open Source Database for Porosity in X-ray CT scanned Components

Bisma Mutiargo[1*] Andrew A.Malcolm[1], Zheng Zheng Wong[1],
Bryan Wei Kang Siow[2], Richie Rui Xin Goh[1]
[1] Advanced Remanufacturing and Technology Centre, A*STAR (Agency for Science, Technology and Research)
[2] National University of Singapore

[†]mutiargob@artc.a-star.edu.sg

Abstract

The fourth industrial revolution has brought many benefits to the mechanical engineering world. Through data revolution, defect segmentation in complex XCT images can now be automated in a shorter time while achieving more accurate results. Prior work [1] proves that a deep learning approach can extract pores from every voxel of 3D XCT data and calculate its porosity with high accuracy. However, it was established that training a deep learning model with limited data can cause the model to overfit or have inferior segmentation performance than models that are trained with a larger dataset. This overfit issue is a common issue in all sorts of data-driven inspections. Often, obtaining raw data and annotating these raw data for machine learning inspection can take significant resources, shifting the focus away from the network design and training.

We present XCTPore, an open source database of X-ray CT images containing 2D slices and 3D volumetric data of X-ray CT scanned additively manufactured components with varying porosity and image quality. This database is currently maintained by the Advanced Remanufacturing and Technology Centre (ARTC) and is open to industry practitioners and academic contributors. The content of the database can be queried through our python code found in our GitHub link[1].

[1] https://github.com/BismaMutiargo/XCTPore

1 Introduction

Advanced manufacturing methods, such as additive manufacturing, often leave traces of pores inside the printed components. The cause of these voids is generally caused by rapidly cooled gas or unoptimized printing parameters. Mechanically, it is known that pores are a form of material discontinuity and can turn into cracks under cyclic load.

Pores are typically measured during the process development stage with optical microscopy or Archimedes density, both destructive measurements. Optical microscopy requires the part to be cut, polished and placed under the microscope. The measurement assumes an idealized equal distribution of pores along the sample, which rarely occurs. Archimedes density measures the total amount of pores by calculating the weight of the sample in a dense medium and air and factoring it with the theoretical density of the material, thus contributing to the uncertainty of the measurement [2]. It assumes that the theoretical density of the material is accurate, and more often than not, Archimedes' density setup requires the component to be no larger than 2-3 cm in size.

X-ray CT is the only established method that can capture the most information about the component, from the distribution of pores, the characteristics of each pore and the total pore percentage against materials inside the sample.

We introduced the application of deep learning in the previous work [1]. The result suggests that deep learning can segment pores and materials in X-ray CT slices to a satisfactory quality. However, it was summarised that deep learning models might tend to overfit or underfit due to the lack of data. Furthermore, the segmentation performance may differ in images that are entirely new to the deep learning model. It was then suggested that the dataset was the first problem to overcome to build a more robust deep learning approach.

In this paper, we introduce the initial version of XCTPore, a publicly-available X-ray CT dataset containing CT slices of additively manufactured components with their corresponding masks. The XCTPore dataset consists of 3300 grayscale images in 11 stacks. Each stack refers to a different set of XCT slices with its own unique characteristics Each class has 300 images and their corresponding masks; it also differs in their porosity level and material shape. Refer to Annex A for sample images of the currently available classes and a class description.

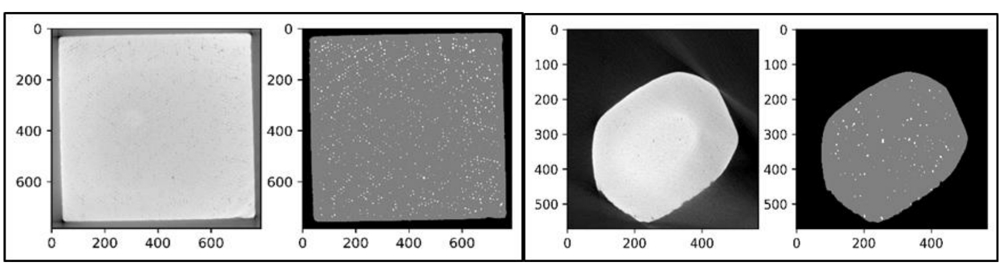


Figure 1: Class MA001 (left), Class MC001 (right)

2 Dataset Format

The dataset is stored in public Google Cloud Bucket and can be downloaded with the instructions in the GitHub link [***https://github.com/BismaMutiargo/XCTPore)***. The Bucket contains files like *MA001_train, MA001_mask, MA002_train, MA002_mask …. TC002_train, and TC002_mask*. Each file is a *.npy* file and can be loaded with the *np.save* function. The prefix represents key information such as its geometry, material, and porosity classification. Refer to Annex A for the information table.

Each *.npy* file has either a *"_train"* or *"_mask"* suffix and the class it belongs to is written as its prefix. Files with *"_train"* suffix contain the training images in a *(300, h, w) uint16* NumPy array. The bit-depth of each pixel is 16-bits (uint16).

Files with *"_mask"* contain the mask for the training image in a *(300, h, w) uint8* NumPy array. The mask at index *i* represents the mask for *ith* training image in the training data. The mask is segmented in the following format:

0 – Foreground 1 – Material 2 – Pore

For example, a file named "*MA001_train"* is the training image for class MA001. The corresponding mask can be found in the file named "*MA001_mask".*

3 Applications of the XCTPore Dataset

XCTPore dataset may be used for image training surrounding porosity segmentation, noise reduction and material segmentation in X-ray CT slices. It is free to be redistributed amongst researchers and developers and made into commercial models sold for a cost. It adopts a similar concept to other major open source datasets such as ImageNet[5]. A permissive license is similar to the BSD 2-Clause License.

Table 1. Permission and limitations to XCTPore dataset

Permission	**Limitations**
Commercial use ✓	Liability ✗
Modifications ✓	Warranty ✗
Distribution ✓	
Private use ✓	

4 Download Instructions

Detailed download instructions can be found in our GitHub link [***https://github.com/BismaMutiargo/XCTPore***]; note that egress fees are chargeable to the requestor, and they will be charged to the requestor's google cloud platform account. As of the writing of this paper, downloading the whole dataset (around 6GB) will cost approximately USD 0.50, and this price may change over time depending on google cloud's charge rates and the current state of the database.

5 Contributing to the dataset

The XCTPore database is open for contributions from academia and industry practitioners. Contributors may upload new CT images containing pores to the database by following instructions found in the last section of the readme.md file of the GitHub link [***https://github.com/BismaMutiargo/XCTPore***]. The current contribution structure is as follows;

1. The user submits to the branch bucket through remote ingress (instructions found in GitHub link's ReadMe.md)
2. The image is uploaded onto the branch bucket.
3. ARTC will monitor the quality and the type of images that the contributor uploaded. it is recommended that the uploaded image should be in 16-bit grayscale format.
4. ARTC will perform semi-supervised labelling using a random forest classifier [3] to the uploaded image, to create a mask pair.
5. ARTC will merge the dataset into the primary bucket, and the total size and image indexing will be updated in the GitHub link.

6 Example usage of XCTPore dataset on an unsupervised classification model

Example usage of the database is demonstrated in the following experiment, using the open source implementation of Unsupervised Semantic Segmentation by Distilling Feature Correspondences (STEGO) [3]. This short experiment aims to demonstrate the XCTPore dataset's usability in training unsupervised segmentation models. The objective of the STEGO experiment is to obtain an automated classification of pores based on the input image presented to the model. For this example, every stack in the XCTPore dataset is used.

Five random images from stack MC001, MC002, TC001 and TC002 were taken to train the model. Each image is upscaled to 3072 x 3072 and then tiled into 256 tiles 192 x 192. The total sample images amount to 5120 training images.

A total of 240 cropped images were derived from original 30 images from stack MA001, MA002, MB001, MB002, TA001, TA002, TB001, TB002 as inference dataset. Similar to image preparation for model training, each image was upscaled to 3072 x 3072 and then tiled into 256 tiles of 192x192. Each tile is sent for inferencing before being reconstructed back into 3072 x 3072. The inference images were then downscaled to 1024 x 1024 to conserve drive memory.

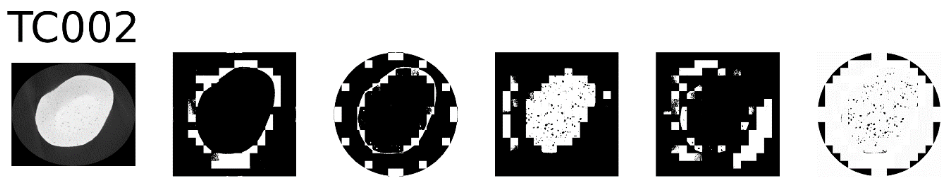


Figure 2: Inference result from STEGO of class TC002. From left to right, Original Image, Class 1, Class 2, Class 3, Class 4, Combination of Class 1-4.

Figure 2 shows the inference result from STEGO trained with the XCTPore dataset. It can be seen from the automatically-generated classes (from left to right) that the STEGO model managed to classify pore as an image feature class. However, the automated classification still contains unwanted noise, and thus further image processing was needed to clean the image.

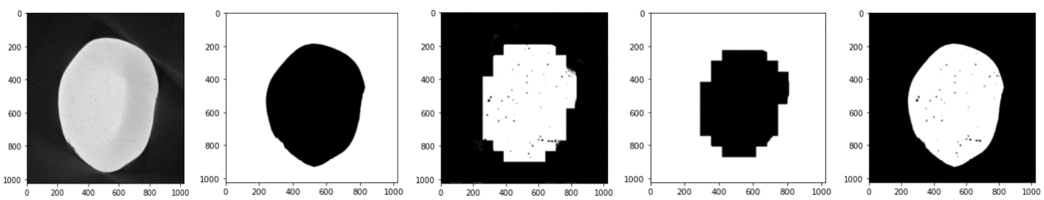


Figure 3: Breakdown of processing step. From left to right, Original Image, Original Image background, Class 3, Class 3 background, Final Mask.

Figure 3 shows the post-processing method used to clean the STEGO output. The first step was extracting the original image's background using a threshold algorithm. Then, the background of the most relevant class (class 3 in this case) was extracted using the same threshold algorithm. The last step was to combine the original background, class 3 background, and class 3 itself by addition. A clipping process was used to clip any values above 255 to 255 to maintain the uint8 format. The output of this additional processing is shown in the rightmost figure and shows the dataset's improvement. This process can be repeated to a new set of inference images.

7 Discussions

This project kickstarts a prototype database relevant to the NDT industry in the long term after various fine-tuning work. As the AI industry is flourishing and has made leaps of improvements throughout the past years, such a centralised database should reduce the cost and burden of developing new models to segment porosity in X-ray CT slices to developers.

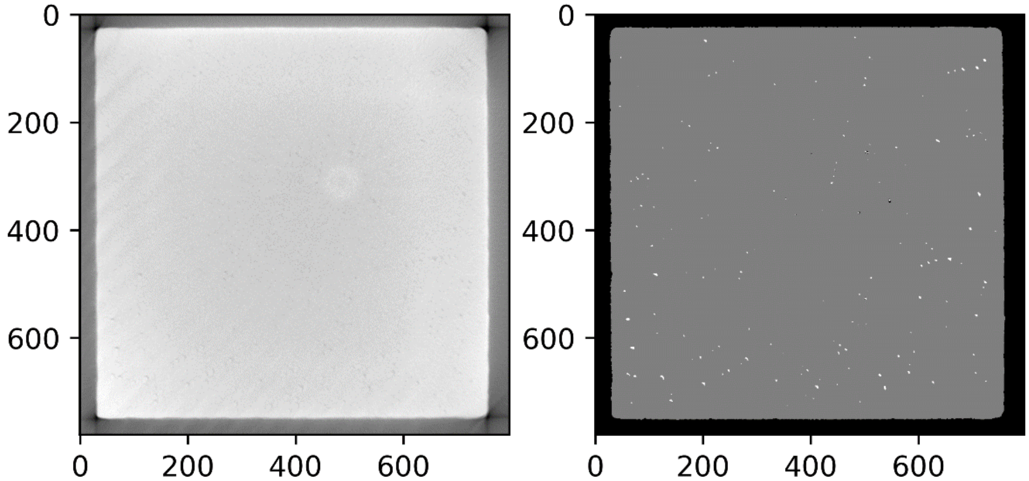
This centralised database will also mean that further development work on classification models can be standardised, and model performance can be compared to the same dataset, thus leading to better precision and accuracy overall, making it scalable for long-term development work.

8 Annex A

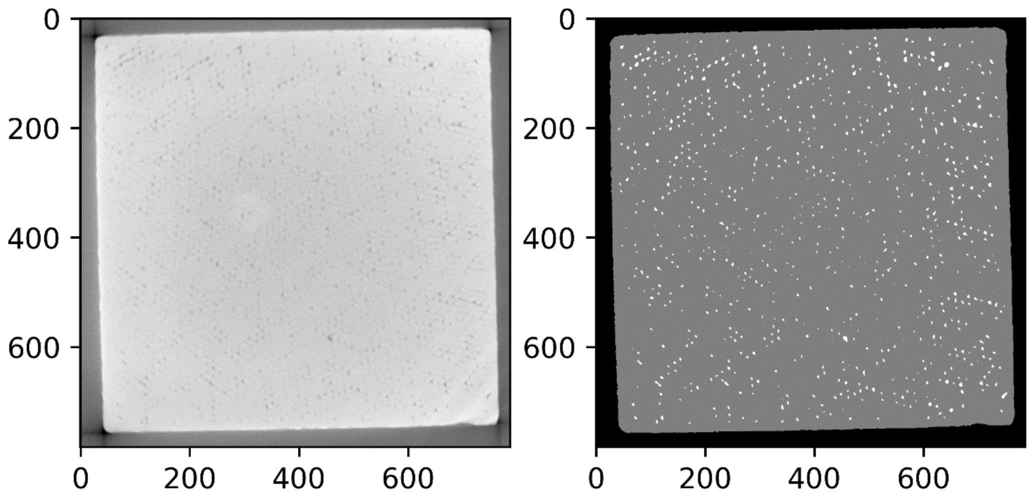
Table A.1. List of files available for download under XCTPore

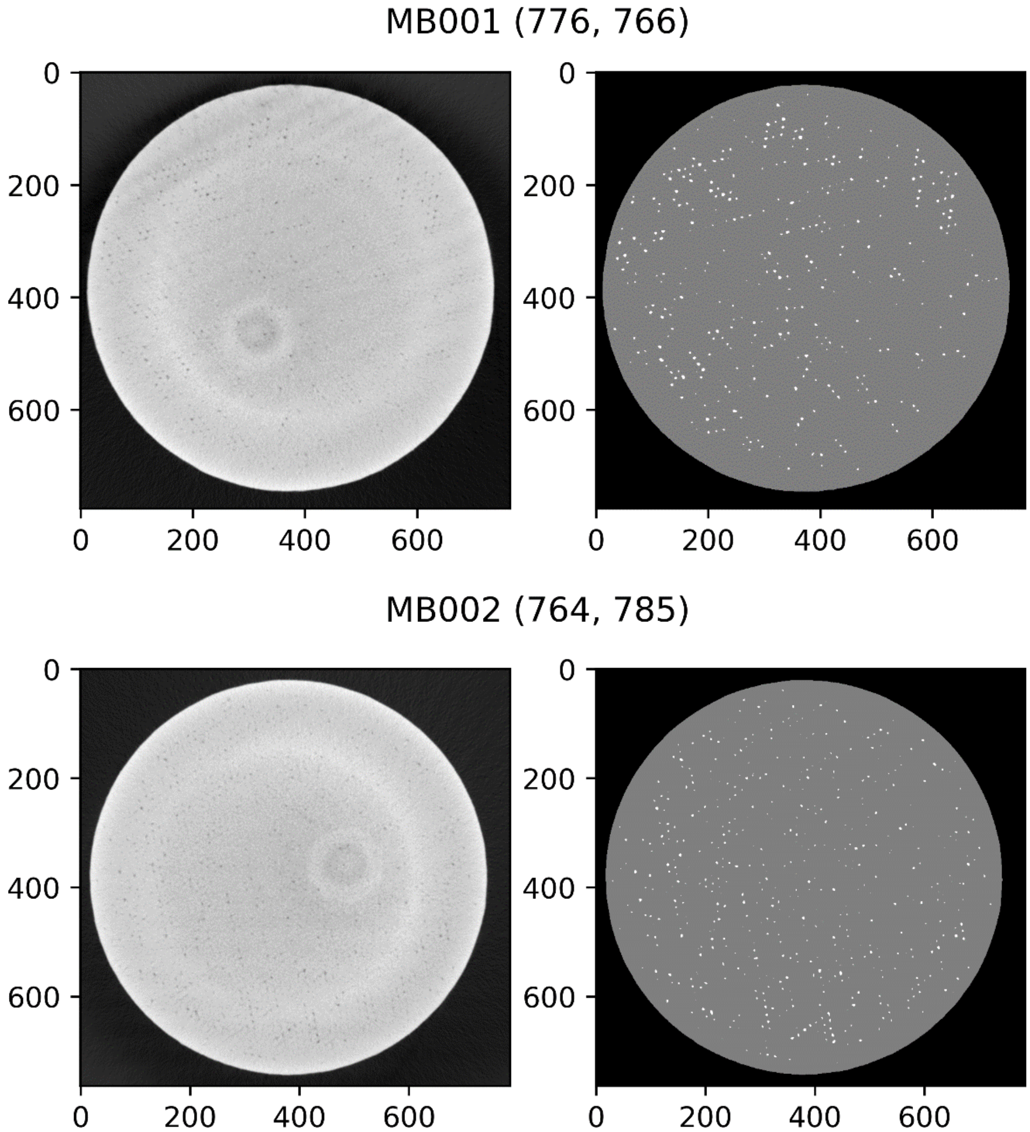
Name	Geometry	Material	Porosity
TA001	Cube	Titanium	Low
TA002	Cube		High
TB001	Cylinder		Low
TB002	Cylinder		High
TC001	Large Bracket		Low
TC002	Large Bracket		High
TD001	Small Bracket		Low
TD002	Small Bracket	Cobalt Chrome	High
MA001	Cube		Low
MA002	Cube		High
MB001	Cylinder		Low
MB002	Cylinder		High
MC001	Large Bracket		Low
MC002	Large Bracket		High
MD001	Small Bracket		Low
MD002	Small Bracket		High

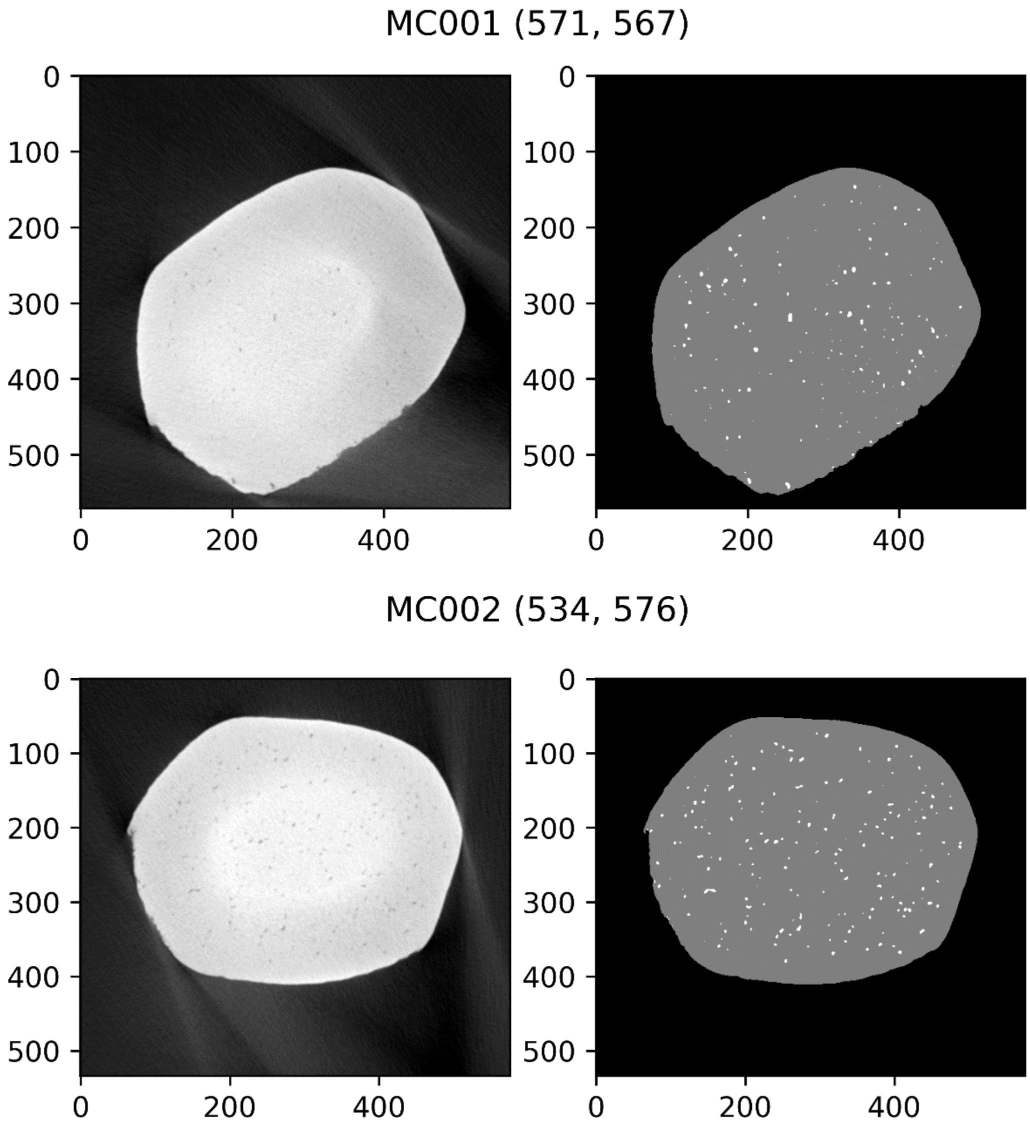
MA001 (780, 797)



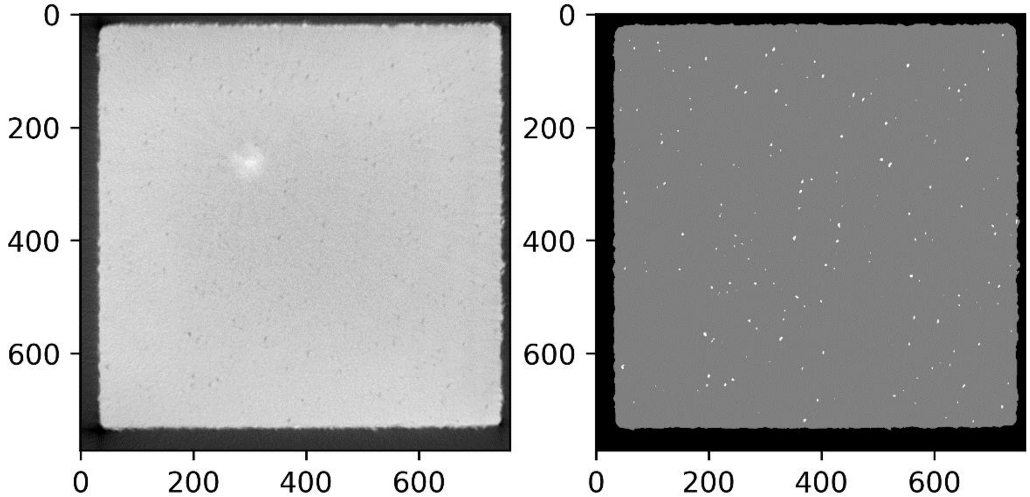
MA002 (783, 787)



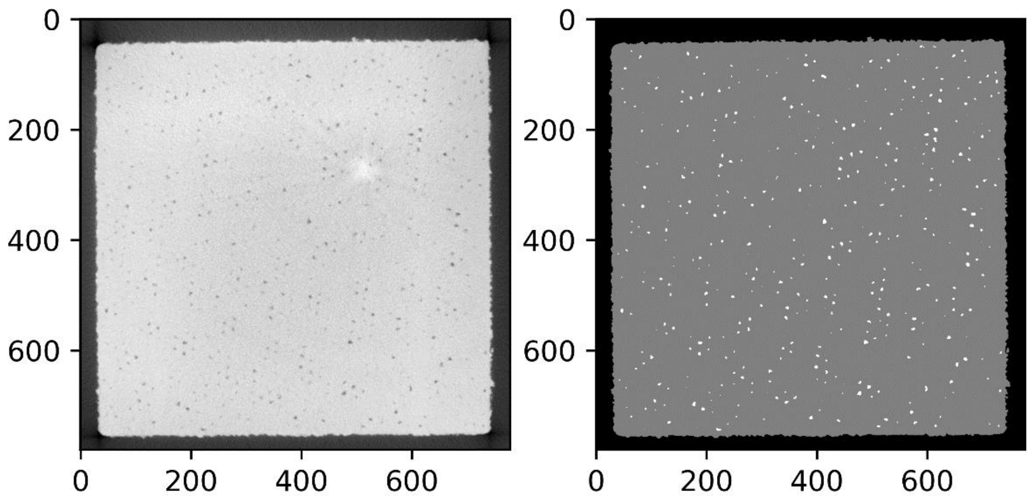

MB001 (776, 766)
0
200
400
600
MB002 (764, 785)
0
200
400
600

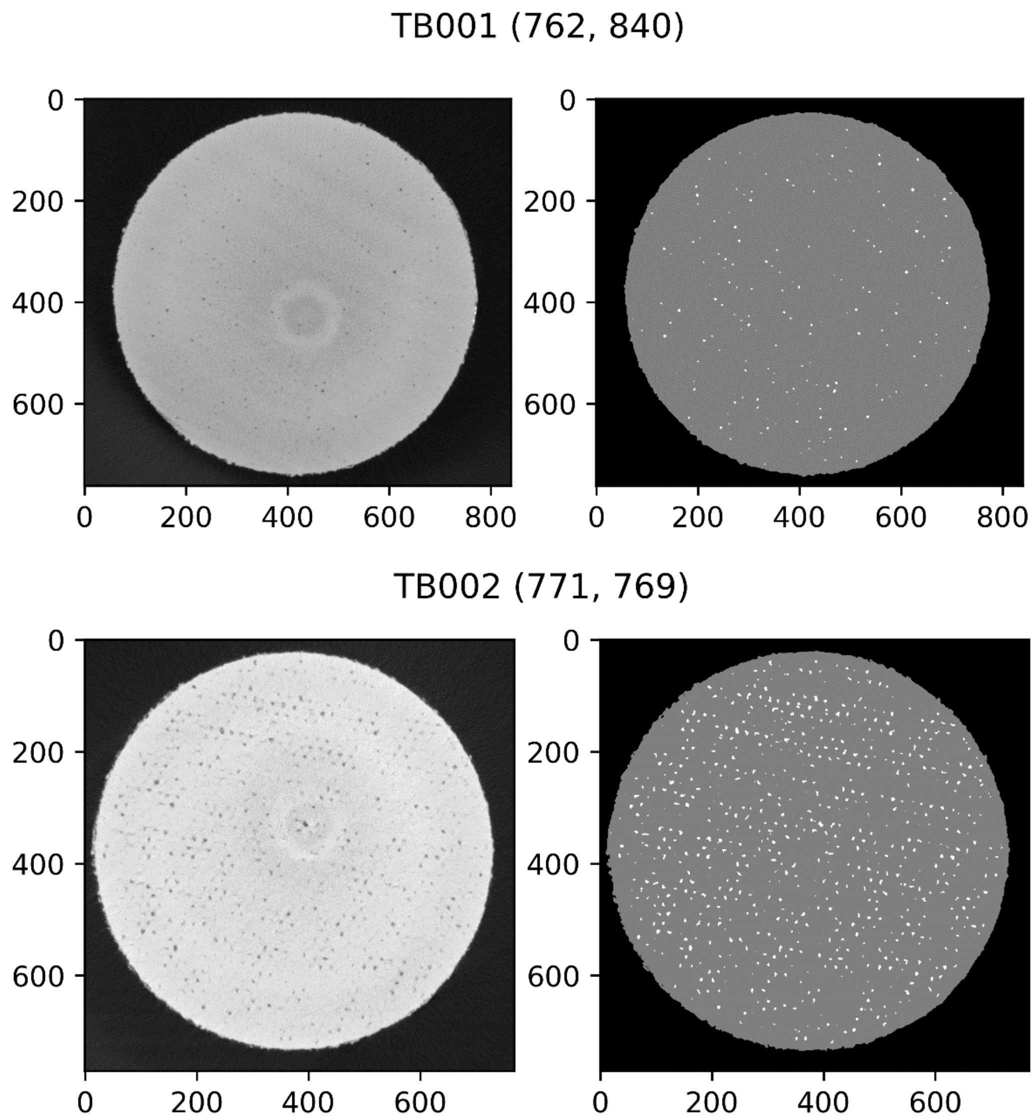

MC001 (571, 567)
0
100
200
300
400
500
0
200
400
MC002 (534, 576)
0
100
200
300
400
500
0
200
400

TA001 (772, 762)

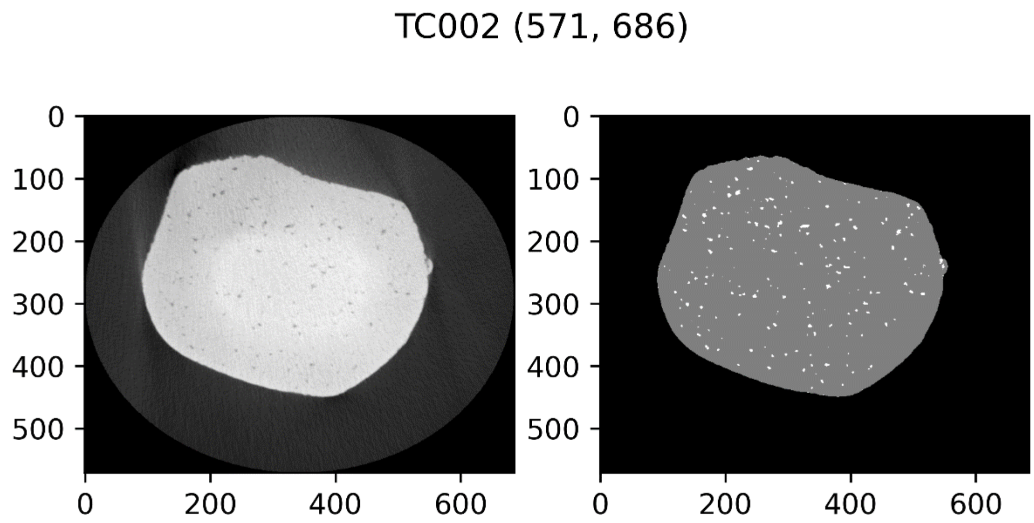


TA002 (780, 778)





TB001 (762, 840)
0
200
400
600
0
200
400
600
800
TB002 (771, 769)
0
200
400
600
0
200
400
600



References

B. Mutiargo, M. Pavlovic, B. Goh, k. Manickavasagam, S. Hussain, T. Shota, A. A. Malcolm, and A. Jhinaoui, "Evaluation of X-Ray Computed Tomography (CT) Images of Additively Manufactured Components Using Deep Learning," in *Singapore International Non-destructive Testing Conference (SINCE) 2019*, Singapore, 2019.

J. J. Lifton, Z. J. Tan, B. Goh, and B. Mutiargo, "On the uncertainty of porosity measurements of additively manufactured metal parts," *Measurement*, 2022.

B. Mutiargo, A. Garbout and A. A. Malcolm, "Defect detection using trainable segmentation," in *Joint International Workshop on Advanced Image Technology (IWAIT) and International Forum on Medical Imaging in Asia (IFMIA)*, Singapore, 2019.

M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely and W. T. Freeman, "Unsupervised Semantic Segmentation by Distilling Feature Correspondences," *arXiv*.

J. Deng, W. Dong, R. Socher, L.-j. Li, K. Li and F.-F. Li, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009.