Article title
Image-based dataset of artifact surfaces fabricated by additive manufacturing with applications in machine learning

Authors
Jiaqi Lyu

Javid Akhavan

Souran Manoochehri*

Affiliations
Stevens Institute of Technology
1 Castle Point Terrace,
Hoboken, NJ 07030, USA

Corresponding author's email address
smanooch@stevens.edu

Abstract
Fused Deposition Modeling (FDM), also known as Fused Filament Fabrication (FFF), is the most widely used type of Additive Manufacturing (AM) technology at the consumer level. This technology severely suffers from a lack of online quality assessment and process adjustment. To fill up this gap, a high-speed 2D Laser Profiler KEYENCE LJ-V7000 series is equipped above an FDM machine and performs a scan after each print layer. The point cloud of the upper surface will be processed and transformed into a 2D depth map to analyze the in-plane anomalies during the FDM fabrication process. The details of the experimental setup and data processing method can be found in [1]. The author used the above data to categorize the surface quality into four categories: under printing, Overprinting, Normal, and empty regions. The author showed the effectiveness of data in detecting print anomalies, and further work can be done to show the application of more advanced algorithms towards a better detection accuracy or to present a novel way to approach the data and detect a broader range of anomalies.

Specifications table

Table 1: Specification Table

Subject	*Engineering, Computer Science*
Specific subject area	*Mechanical Engineering, Manufacturing Engineering, Image processing, Image identification, Image classification, computer vision*
Type of data	*Image*

1

How the data were acquired	A Laser Surface Profilometer acquired point clouds by the KEYENCE LJ-V7000 series. Experts labeled a portion of them for anomalies.
Data format	Analyzed
Description of data collection	A 2D Laser Profiler KEYENCE LJ-V7000 series is equipped above an FDM machine. After each layer deposition, the laser scanner acquires the 3D point cloud of the upper surface. Then the 3D point cloud is pre-processed to achieve the target surface of the part. Then the top surface is segmented from the point cloud and converted to a 2D depth image provided with this article.
Data source location	• Institution: Stevens Institute of Technology • City/Town/Region: Hoboken NJ • Country: USA
Data accessibility	With the article
Related research article	For an article that has been accepted and is in press: J. Lyu, J.A.T. Boroujeni, S. Manoochehri, In-situ laser-based process monitoring and in-plane surface anomaly identification for additive manufacturing using point cloud and machine learning, Proceedings of Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf., 2021, In Press

Value of the data

- *The data in this article provide some background on point cloud processing and anomaly detection in FDM additive manufacturing machines.*
- *The dataset provided can be used by researchers to develop and compare new algorithms' efficiency and prediction accuracies.*
- *The dataset provided and the techniques used can be combined with many 3D printing machines to study the undergoing print.*
- *The dataset can be further utilized for the development of more capable networks with more comprehensive detection capacities.*
- *This dataset is expected to benefit scientists working on process advancement and online control for AM machines.*
- *This paper's idea and the dataset can be utilized to design a control algorithm for quality advancement.*

Data description

The dataset presented in this article consists of 434 scan files from the top surface of the artifacts under study. Each scan file is pre-analyzed and threshold to generate color map data to represent the surface quality better. All the scan files are summarized and stored in a struct file with branches as follows:
1) *Heightmap image of size [300×300×3 uint8]*
2) *Labels [10×10 double]*
3) *Part name or number as a string*

Each height map data is divided into a 10by10 grid (100 segments in total), and professionals labeled each segment into four categories: (a) Over Printing Situation, (b) Normally Printed Situation, (c) Under Printing Situation, (d) Empty. After acquiring labels, each image was cropped into 100 smaller images and then stored in a separate folder based on the label. Table 1: Specification Table *Table 2 summarizes the amount of data available for each category and the proportion used for training, validation, and testing. Four directories containing all the segments are also attached to the dataset. On top of the data, we have also attached a MATLAB-based User Interface (U.I.) developed for labeling this dataset.* Figure 1: Structure of the dataset *Figure 1 shows the structure of the dataset.*
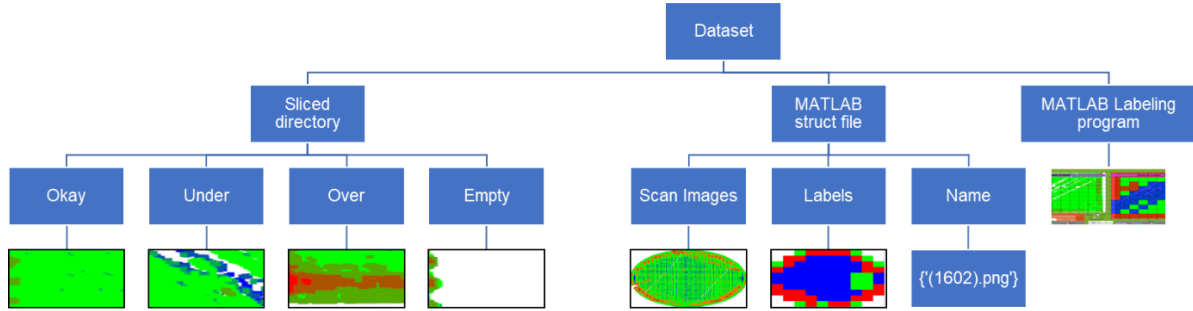

Figure 1: Structure of the dataset

Table 2: Number of training, validating, and testing images before data augmentation.

CATEGORY	NAME	TRAINING	VAL	TEST	TOTAL
1	Over	8,587	2,454	1,227	12,268
2	Okay	9,259	2,646	1,323	13,228
3	Under	10,308	2,945	1,473	14,726
4	Empty	2,225	636	318	3,179
TOTAL		30,379	8,681	4,341	43,401

Experimental design, materials, and methods

The dataset presented in this study is captured by an online laser-based process monitoring and anomaly identification system for the FDM machine. This system equips with a non-contact laser scanner to obtain the top surface profile of each layer during the FDM fabrication process. This section describes the experimental setup, the pre-processing steps for the dataset, and the data labeling steps.
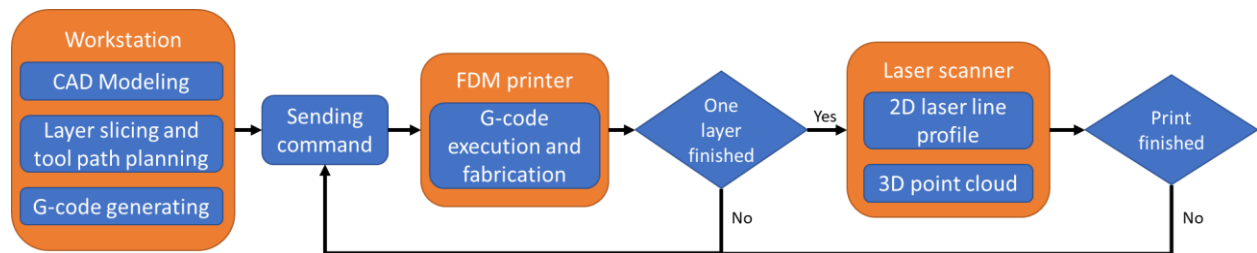

Figure 2: Schematic of the online laser-based process monitoring system

3

1. Experimental Setup

In this study, the FDM machine used is Creality ender 5 with 220mm × 220mm × 300mm building volume. The schematic of the online laser-based process monitoring system is shown in Figure 2. A high-speed 2D Laser Profiler KEYENCE LJ-V7000 series is equipped above the FDM machine. The laser scanner includes a laser line emitter and built-in camera packaged with a fixed relative position. The scanner is calibrated from the factory and has a 1 μm resolution in the Z direction. The measurement range of the laser scanner is ±48 mm from the reference position, which is 200 mm below the bottom of the sensor. An artifact is designed in SOLIDWORKS 3D CAD design software for each print and converted into Standard Tessellation Language (STL) with American Standard Code for Information Interchange (ASCII) codes. The Simplify3D Software is used to slice the STL model and generates the G-code of the artifact. Then the FDM machine starts printing after receiving the G-code from the workstation. After each layer is completed, the scanner frame motion mechanism will drive the laser scanner to move above the fabricated layer with constant speed. At the same time, an angle encoder continuously triggers the scanner to perform measurements. The limit switches are there to make sure the scan is limited to the platform area of the FDM machine. Then the 3D point cloud of the upper surface is acquired from the laser scanner. The experimental setup is shown in Figure 3.
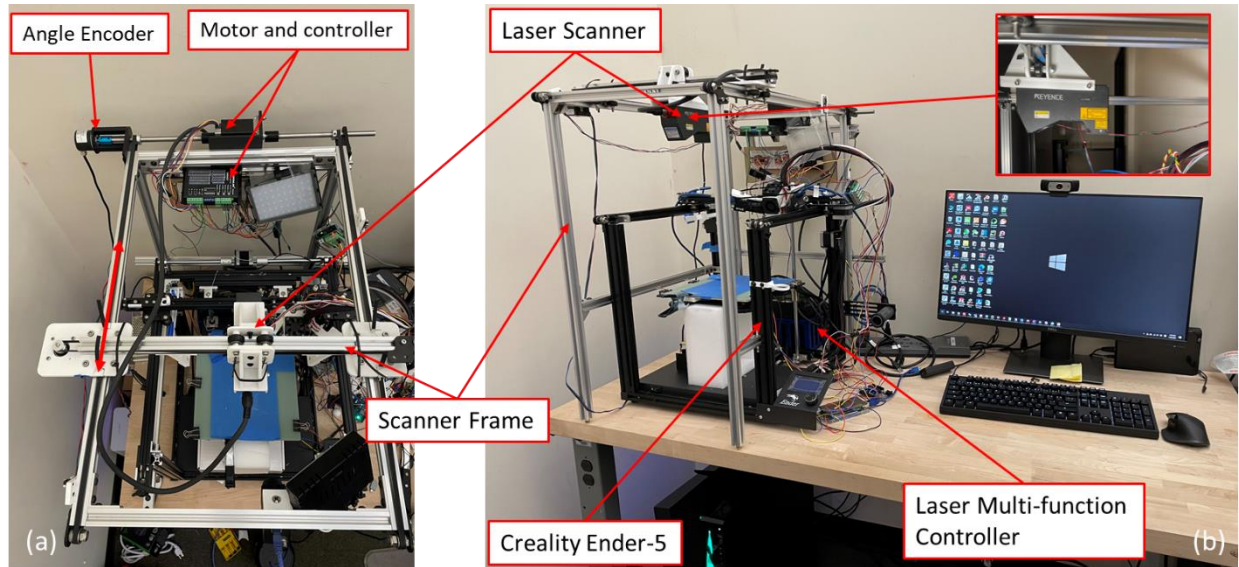

Figure 3: Experimental setup (a) the top view (b) all the components of the 3D printer-scanner system

2. Point Cloud Processing

The 3D point cloud acquired from the laser scanner includes noise that needs to be removed before further analysis. The mechanical vibration commonly causes noise during the scanning process and the shadow effect characteristics of the optical scanning sensor. This study uses three steps to pre-process the 3D point cloud, including replacing missing measurement, removing measurement noise, and transformation.

Step 1: replace missing measurement values

4

When the target is out of detecting range during the laser scanning process, or does not capture enough information, the laser scanner generates missing measurements values, flagged as Nan in this laser scanner. The raw data obtained from the laser scanner is a 2D depth matrix in the form of $m \times n$. To replace the missing measurement values, the median operator with a 3×3 window is utilized [2]. If missing values surround the missing value, the pixel will be removed. Then filled 2D depth dataset is converted to point cloud form represented as (x, y, z) spatial coordinates.

Step 2: remove measurement noise

Due to the design of the laser scanner, the shadow effect can cause measurement noise in the raw data. When the scanner moves above areas of rapid change in height, these areas could be identified as outliers algorithmically according to the surrounding topographies. The statistical outlier removal algorithm and region growing algorithm in the point cloud library (PCL) [3] are customized and utilized to reduce the noise in the point cloud dataset. The effect of applying these algorithms is shown in Figure 4**Error! Reference source not found.**. Figure 4(a) highlights the noise point cloud and concentrates on the step area. In Figure 4(b), to better display the point cloud after using a statistical outlier removal algorithm, the point cloud with different heights is rendered as different colors. It shows that the algorithm can effectively remove the noise between steps. Figure 4(c) shows the noise point cloud as an island apart from the main body. The extra deposit plastic is as the print head travels through the open air and the laser scanner captures it. This phenomenon is commonly caused by incorrect retraction settings or high extruder temperatures. Since the study focuses on the upper place surface, the point cloud of extra material is treated as noise data and can be removed by using a region-growing algorithm.

Step 3: point cloud transformation

After removing the noise data, the point cloud should be transformed to the FDM printer coordinate to keep the scan results in the same spatial coordinate. Therefore, a calibration process is necessary to find the spatial relationship between the FDM printer and the laser scanner. The details of the calibration process are described in a previous study [4].
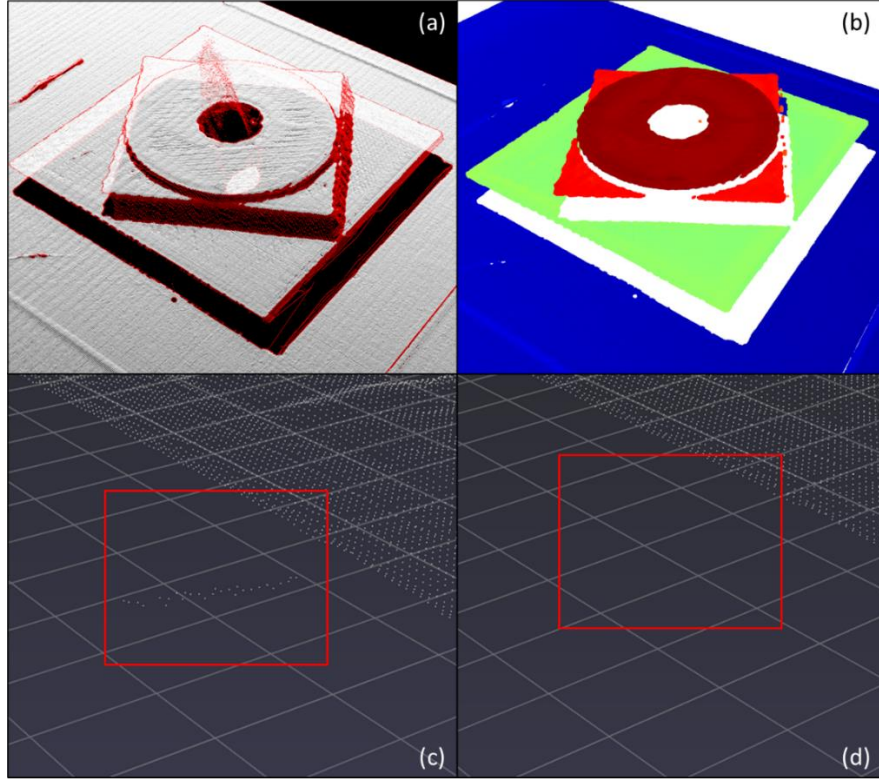
Figure 4: Effects of filtering for point cloud dataset (a) and (b) point cloud before and after applying statistical outlier removal algorithm, (c) and (d) point cloud before and after using region growing algorithm

Then the interested point cloud of the upper surface is segmented from the pre-processed point cloud. As shown in Figure 5, the Random Sample Consensus (RANSAC) algorithm [5] *is used to segment the planes of the point cloud. Using this algorithm, only the points between two parallel planes within distance threshold δ are considered as inliers. In this study, the δ is set as half of the layer thickness. The majority of the point cloud in the upper surface can be selected, shown as green points in Figure 5. The plane segmented by RANSAC is identified as the upper surface virtual plane. Then the threshold δ is increased to the layer thickness value to include more point clouds, which are used as points of upper surface for the following processing.*
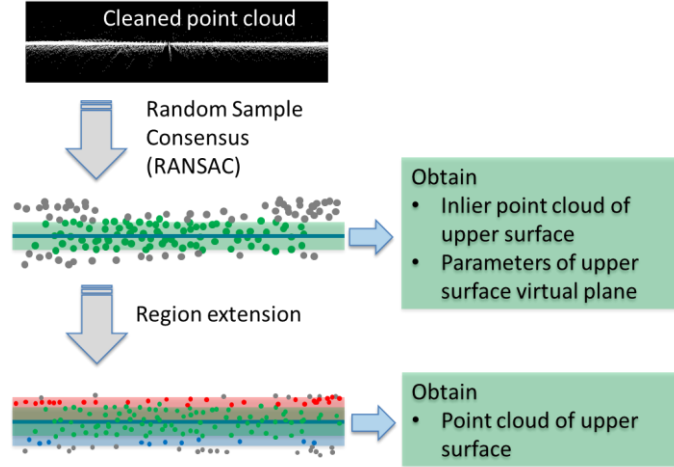
Figure 5: Upper surface segmentation

Finally, to generate a 2D depth image dataset shared in this paper, the point cloud of the upper surface needs to be projected on a 2D plane. The rasterization method is implemented[4]. *Then the color of each pixel in the 2D depth image is determined by the ratio of the accumulated depth of pixel and layer thickness. The color mapping rule is shown in Figure 6.*
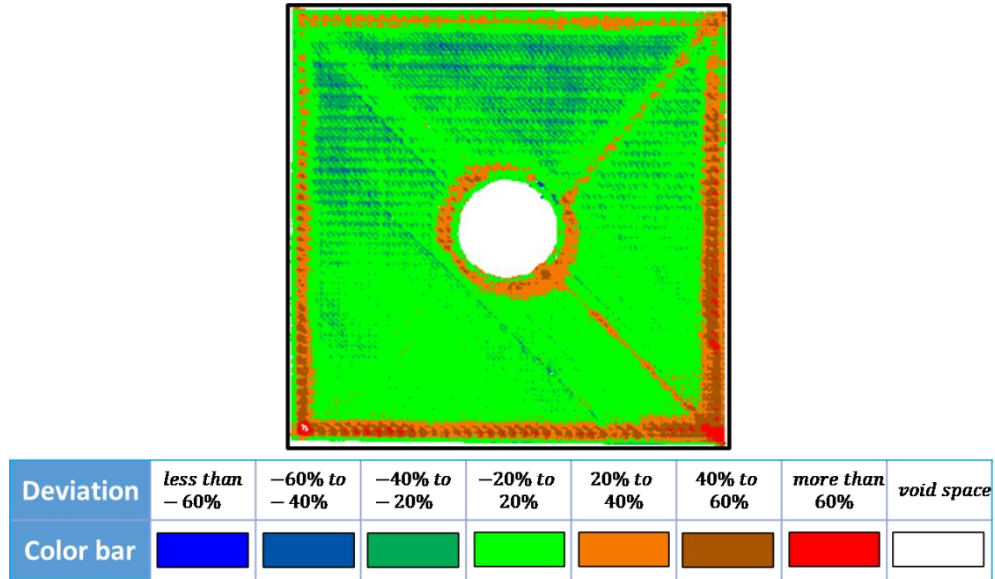

Deviation	less than −60%	−60% to −40%	−40% to −20%	−20% to 20%	20% to 40%	40% to 60%	more than 60%	void space
Color bar								

Figure 6: Color mapping rule for depth image
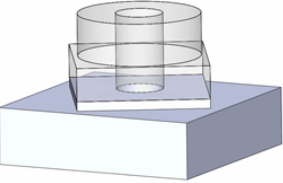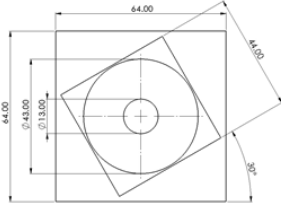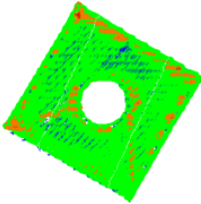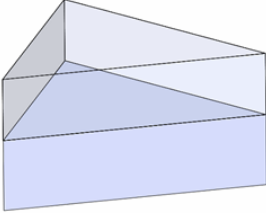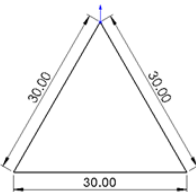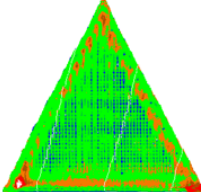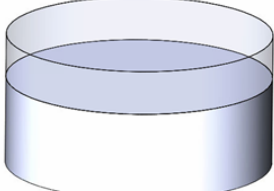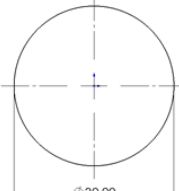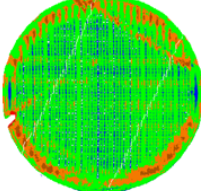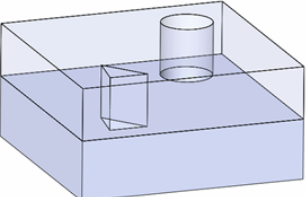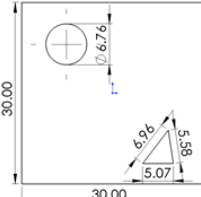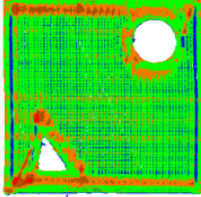
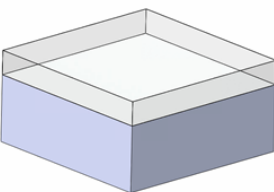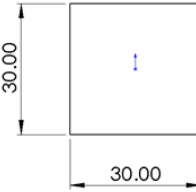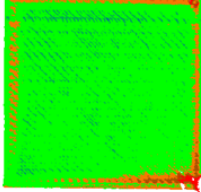| CAD model	2D drawing	Depth image

Figure 7: Artifacts fabricated by FDM machine

3. Database Structure:

The database is generated after the printing of multiple artifacts, shown in Figure 7. During the printing of each artifact, with different settings, various qualities were observed. The process parameters are shown in Table 3. All the other parameters are kept constant. Later, each observation was studied and processed. In total, we gathered 434 scans, and by dividing each image into 10by10 segments, we generated 43400 labeled data points.

8

Table 3: FDM machine parameters

FDM machine parameters	Value
Layer thickness	0.3 mm
Print speed	3500 mm/min
Infill density	100%
Feed rate	50% ~ 150%
Extruder temperature	180°C ~ 250°C

Labeling the dataset:

Labeling a dataset in many cases requires a significant investment and effort of professionals to make sure the data is labeled correctly and in a repeatable nonbiased manner. In this dataset, to simplify the process, four categories for labels are selected: 1) Over Printing Situation, 2) Normally Printed Situation, 3) Under Printing Situation, 4) Empty Region. To better visualize the labeled data, solid color has been given to each grid based on its label respectively: Red, Green, Blue, white. Due to the complexity of the labeling task and the massive data, a User Interface based on MATLAB was developed to minimize the user error in labeling and store the labels by taking advantage of visual representation of data and labels. This Interface allows researchers to load their data, define the number of segments and labels needed, and ease the labeling process. Professionals can go over the previously labeled data with the review functionality to double-check and verify their accuracy visually. In Figure 8, a sample of the labeling U.I. is shown in the Testing mode. This U.I. is also shared along with the data.
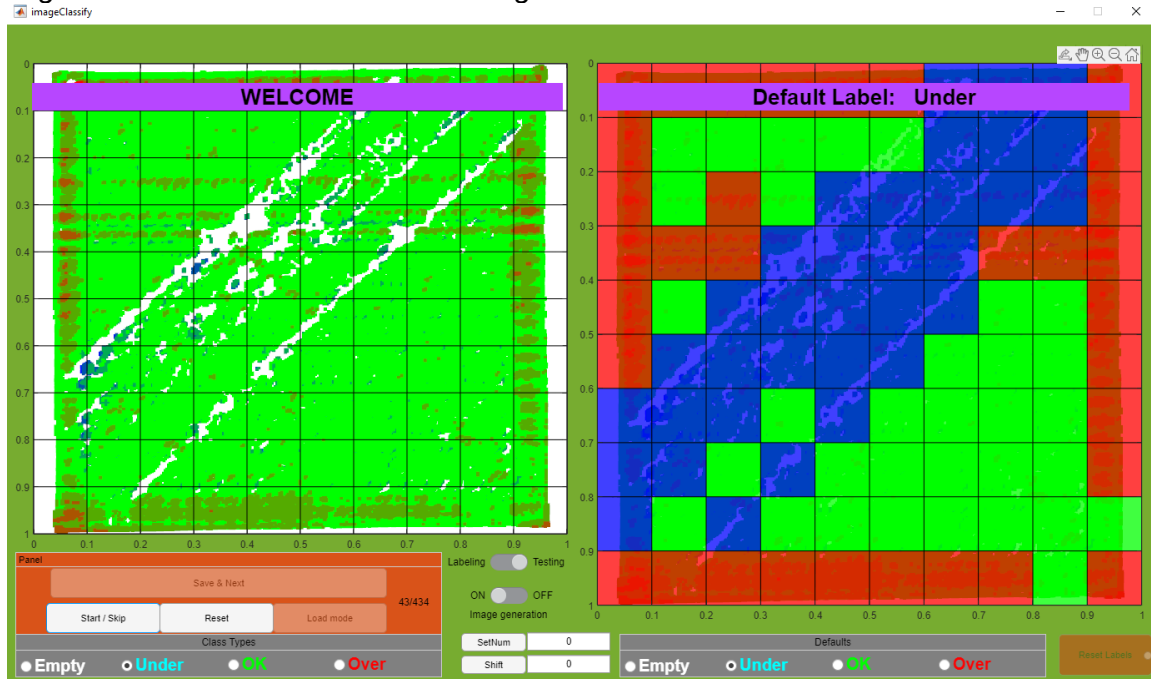

Figure 8: View of MATLAB Labeling application in review mode to visualize the labels recorded.

CRediT author statement

Jiaqi Lyu: *Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation*
Javid Akhavan: *Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation*
Souran Manoochehri: *Supervision*

Acknowledgments

Declaration of interests
*Please **tick** the appropriate statement below and declare any financial interests/personal relationships which may affect your work in the box below.*

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

References

[1]   J. Lyu, J.A.T. Boroujeni, S. Manoochehri, In-situ laser-based process monitoring and in-plane surface anomaly identification for additive manufacturing using point cloud and machine learning, in: Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf., 2021.

[2]   N. Senin, L. Blunt, M. Tolley, The use of areal surface topography analysis for the inspection of micro-fabricated thin foil laser targets for ion acceleration, Meas. Sci. Technol. 23 (2012) 105004. https://doi.org/10.1088/0957-0233/23/10/105004.

[3]   R.B. Rusu, S. Cousins, 3d is here: Point cloud library (pcl), in: 2011 IEEE Int. Conf. Robot. Autom., IEEE, 2011: pp. 1–4. https://doi.org/10.1109/ICRA.2011.5980567.

[4]   J. Lyu, S. Manoochehri, Online Convolutional Neural Network-based anomaly detection and quality control for Fused Filament Fabrication process, Virtual Phys. Prototyp. 16 (2021) 160–177. https://doi.org/10.1080/17452759.2021.1905858.

[5]   R. Schnabel, R. Wahl, R. Klein, Efficient RANSAC for point-cloud shape detection, in: Comput. Graph. Forum, Wiley Online Library, 2007: pp. 214–226. https://doi.org/10.1111/j.1467-8659.2007.01016.x.