


# Simplified Object Detection for Manufacturing: Introducing a Low-Resolution Dataset

Jonas Werheid  <sup>1</sup>

Shengjie He  <sup>1</sup>

Tobias Hamann  <sup>1</sup>

Anas Abdelrazeq  <sup>1</sup>

Robert H. Schmitt  <sup>1</sup>


1. Chair of Intelligence in Quality Sensing (WZL-IQS), RWTH Aachen University, Aachen.



**Date Submitted:**

2024-10-16

**Licenses:**

This article is licensed under: 

**Keywords:**

Dataset, Computer Vision, Object Detection, Quality Classification, Manufacturing

**Data availability:**

Data can be found here: <https://zenodo.org/records/10731976>

**Software availability:**

Software can be found here: [https://git.rwth-aachen.de/zukip-ro/yolov5\\_for\\_plastic\\_brick\\_quality\\_classification](https://git.rwth-aachen.de/zukip-ro/yolov5_for_plastic_brick_quality_classification)

## Abstract.

Machine learning (ML), particularly within the domain of computer vision (CV), has established solutions for automated quality classification using visual data in manufacturing processes. Object detection as a CV method for quality classification provides a distinct advantage in enabling the assessment of items within the manufacturing environment, regardless of their location in images. However, substantial challenges remain regarding labeled data availability in manufacturing contexts, training examples, data imbalance, and the complexity of incorporating these methods into real-world applications. Furthermore, real-world datasets often lack adherence to FAIR principles, which limits their accessibility and interoperability, especially for small- and medium-sized enterprises (SMEs) working to integrate object detection into their manufacturing processes. In this article, we present a low-resolution 640x640 dataset based on plastic bricks for object detection, featuring two quality labels to identify minor surface defects as an example of quality classification. We analyze the dataset using a YOLOv5 model on three different dataset sizes, while accounting for class imbalance, to demonstrate the accuracy of an object detection model in a simple manufacturing use case. The mean Average Precision mAP@0.5 for correctly identifying instances in our testing dataset ranges from 0.668 to 0.774, depending on dataset size and class imbalance. While our focus is on demonstrating object detection with low-resolution images and limited data availability, the generated data and trained model also adhere to FAIR principles. Therefore, these resources are made available with proper metadata to support their reuse and further investigation into object detection tasks for similar quality classification use cases in manufacturing.

## 1 Introduction

Object detection and pose estimation are key capabilities in unstructured or less structured environments to enable smart manufacturing applications, such as autonomous robots or process monitoring [1]. However, these areas in computer vision (CV) including advanced machine

5 learning (ML) techniques are still in their infancy [2]. Although research reveals a robust  
6 understanding of ML and applications, notably small- and medium-sized enterprises (SMEs)  
7 show low maturity with only 8 percent of SMEs in Germany having deployed ML technologies  
8 in a questionnaire done in 2020 [3]. Also, a further study with 368 German SMEs revealed in  
9 2021 that just 5.8 percent of them developed AI solutions by themselves [4]. The governmental  
10 project "Mittelstand Digital" identified insufficient data as the second most significant obstacle  
11 among nine barriers to AI adoption in SMEs. Furthermore, the preparation of best practices and  
12 examples was highlighted as the most suitable public measure among 16 factors that support  
13 SMEs in AI integration [5].

14 These challenges and circumstances underscore the critical necessity for open-source ML datasets  
15 and pre-trained models, serving as illustrative examples to articulate best practices and facilitate  
16 the transfer of research into the industry for SMEs to deploy ML techniques such as object  
17 detection and foster their manufacturing processes. Additionally, such open-source publications  
18 must encourage FAIR principles to ensure efficient integration and interoperability of presented  
19 best practices for SMEs and stakeholders [6].

20 Recent approaches introduced various object detection datasets, in diverse domains, such as  
21 for detection of industrial tubes or safety helmets in different scenarios [7],[8]. Moreover, the  
22 existing research contributes to datasets provided with a focus on object detection in the context  
23 of defect detection or quality classification of industrial goods, such as metal parts, printed  
24 circuit boards, or insulator components for electricity supply [9], [10], [11]. Also, datasets  
25 incorporating plastic bricks are available as artificial use cases [12], [13]. These serve as learning  
26 resources and provide realistic synthetic image datasets for training object detection methods in  
27 an understandable context [12].

28 However, the literature does not describe object detection datasets as best practices for SMEs  
29 in the context of exemplary manufacturing applications and their adherence to FAIR principles  
30 for easy reuse. Demonstrating a tangible object detection use case in manufacturing with  
31 low-resolution image data and development showcases considering limited data availability  
32 is not addressed in the literature. Exemplary model development showcases, illustrating best  
33 practices for developing algorithms of the corresponding datasets, are either not provided or lack  
34 description. Also, findability and descriptions of access licenses are not described, indicating an  
35 insufficient fulfillment of FAIR principles. For example, Digital Object Identifiers or Metadata  
36 are typically not provided within these resources. FAIRness evaluation software, such as F-UJI,  
37 evaluates the FAIRness of the cited resources with a score below 65 percent [14]. This highlights  
38 a significant gap in FAIR datasets and showcases that could offer tailored best practices for  
39 SMEs in manufacturing to foster their AI integration.

40 Building upon the context of research challenges and existing approaches, we develop a simple  
41 low-resolution object detection dataset based on plastic bricks with some having minor surface  
42 defects. Furthermore, we train a current ML model of the YOLO series to detect the bricks  
43 and whether they show defects. Different sizes of datasets are used to assess how performance  
44 varies depending on the availability of data. Moreover, class imbalance, a common challenge in  
45 manufacturing, is considered to highlight its impact on detection precision [15]. Our primary  
46 discovery centers around achieving good accuracy levels despite limited data availability, class

47 imbalance, and suboptimal camera resolutions, emphasizing the critical interplay between data,  
48 resolution, and the specific use case under consideration.

49 We structure these by presenting the dataset and its properties first, then explaining its creation  
50 and methods in Section 2. In Section 3, we analyze the dataset with the open-source object  
51 detection model YOLOv5 and provide a pre-trained architecture including insights and analytics  
52 of the training with varying dataset sizes and class imbalances. Hence, the data and model are  
53 published regarding FAIR principles with metadata ensuring the transferability of this publication  
54 to stakeholders, such as developers in SMEs. Finally, the contribution and its limitations will be  
55 discussed in the conclusions.

## 56 **2 Dataset**

57 The dataset encapsulates the complexities of surface defect detection with plastic toy bricks as  
58 objects. It comprises multiple plastic bricks of different colors and sizes within a single frame,  
59 that are either defective or valid. Defective bricks have indentations and deformations on the  
60 surface, aiming to resemble common surface defects in industrial manufacturing. The following  
61 section provides a comprehensive overview of the dataset, including insights into the collection  
62 methods and employed tools. Section 2.1 delves into the fundamental details and properties of  
63 the dataset, while Section 2.2 outlines the process of image collection and annotation creation.

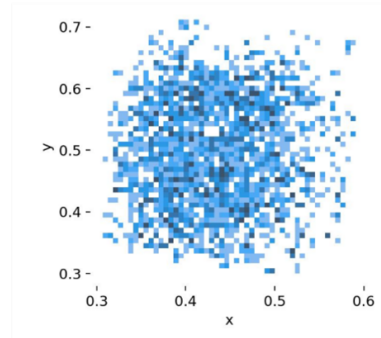
### 64 **2.1 Data Description**

65 The dataset provides images of plastic toy bricks with surface damage caused by a hammer as  
66 objects to inspect. While the bricks occur in multiple colors and sizes, the labels are provided  
67 binary with valid bricks and defective ones having damages on their surfaces. The dataset  
68 consists of 1500 images containing a total of approximately 4400 objects. Among these objects,  
69 there are roughly 2000 instances representing defects and 2400 representing valid instances. This  
70 balanced distribution of labels within the dataset serves to counteract possible biases and prevent  
71 models from learning disproportionately toward any particular class and therefore simplify the  
72 object detection task. Nevertheless, the dataset can be manipulated to introduce class imbalance  
73 by utilizing the metadata on class distribution to select a subset of the data, thereby making the  
74 task more challenging, as demonstrated in Section 3.

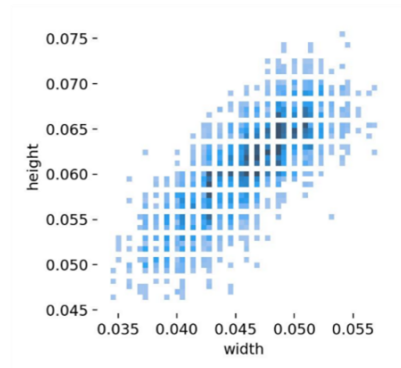
75 Each image has a corresponding label. Table 1 shows all information provided by a label. The  
76 coordinates x-center and y-center are normalized and refer to the coordinates of the center point  
77 of a bounding box, that labels an object to inspect. Width and height represent the dimensions of  
78 the bounding box in normalized pixels, where pixel values are scaled between 0 and 1, relative  
79 to the image dimensions. Lastly, the label indicates the two classes valid and defective. Figure  
80 1 overlaps the labels of each image. Figure 1a shows x-center and y-center. The uniform  
81 distribution counteracts any specific patterns in the locations of objects. Further, Figure 1b  
82 represents the height and width of each bounding box center and indicates the dimension of an  
83 object. The linear distribution occurs due to the quadratic geometry of all plastic bricks used.  
84 Defective instances in the testing set do not appear in the training or validation set images.

Class	X-Center	Y-Center	Width	Height
Defective	0.43984375	0.43125	0.0375	0.0546875
Valid	0.44765625	0.5921875	0.0390625	0.05625

**Table 1:** The content of the label file corresponding to the example image in Figure 3b



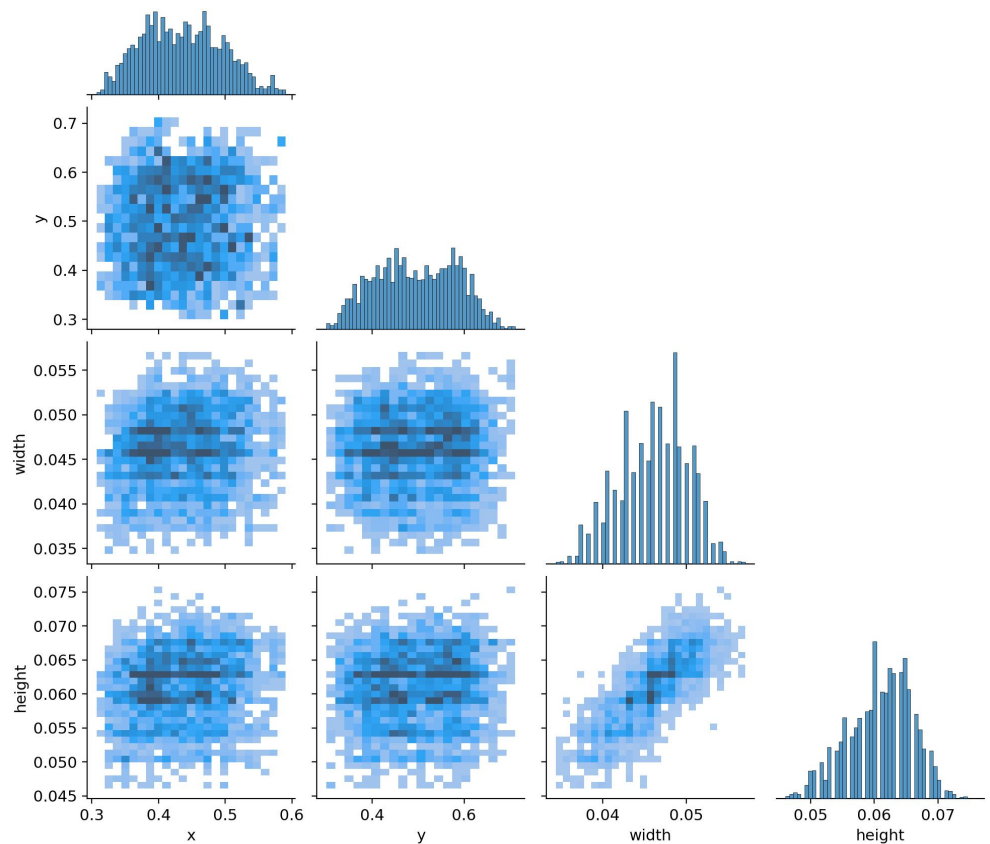
**(a)** X-center and y-center are the normalized coordinates of the bounding boxes around objects



**(b)** Height and width represent the size of a bounding box indicating the dimension of an object and its distance to the camera

**Figure 1:** The distribution in both figures is nearly uniform and therefore counteracts specific patterns in object locations

85 The correlogram in Figure 2 shows a detailed correlation of all data properties. It is a group of  
 86 2-dimensional histograms showing each axis of the data against each other axis. The correlation  
 87 statistics indicate the position, width, and height of the bounding boxes of the objects. The figure  
 88 indicates that the dataset properties are balanced in each label combination with no clusters  
 89 visible. The distributions of single labels present approximately normal distribution. Notably,  
 90 outliers are infrequent, and those present are rare points rather than data values that significantly  
 91 deviate from the expected pattern.



**Figure 2:** The correlation of all labels to each other shows an approximately normal distribution and balance in the data

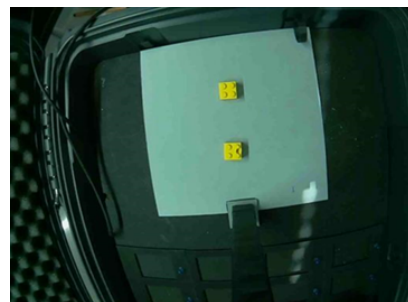
92 Each image is saved in JPG/.jpg format with a size ranging between 35 and 40 kilobytes. These  
 93 images maintain a consistent shape of 640x640 pixels. The corresponding labels for these  
 94 images are stored in a separate file in TXT/.txt format, which also includes metadata on how  
 95 many instances of each class are present in the image, allowing for the creation of imbalanced  
 96 subsets of the data. The file paths for both the images and the labels are specified within a file  
 97 in YAML/.yaml format. As a result, all files collectively occupy a total size of 58.2 megabytes.  
 98 The files are available on Zenodo and linked in Section 4: [Usage Notes](#).

99 The dataset offers a wide range of possibilities for diverse tasks, including object localization,  
 100 object classification, object counting, semantic segmentation, and scene understanding. However,  
 101 the dataset's provided labels and the identifiable damages on the objects make it particularly well-  
 102 suited for tasks related to object detection and quality classification, specifically in identifying  
 103 surface defects often encountered in manufacturing industries.

## 104 2.2 Data Collection

105 The data collection was done in a defined procedure. Images were captured with a microcontroller  
106 board and a compatible camera. An Arduino UNO was chosen with an OV7670 300KP VGA  
107 Camera. Arduino embedded systems are widely available and used for prototype purposes.  
108 They benefit from an active online community helping to lower development challenges [16].  
109 Moreover, the setup includes fluorescent lighting directed toward the objects under inspection,  
110 with the camera positioned on a tripod to maintain a fixed distance from the ground, where all  
111 the objects are placed. The distance between the camera and the objects is determined by the  
112 angle each object has relative to the camera. On the software side, Python code controls the  
113 capturing process. The collection started from single objects with different colors, angles, and  
114 positions, as well as defects on some objects. Later on, multiple objects were placed in one  
115 image with the same differences described. Each defect is generated by a hammer manually and  
116 therefore individual with a varying degree of surface damage. This supports the diversity of  
117 surface damages that are labeled as defective.

118 The annotation of the images is based on the software Roboflow [17]. Features, polygon bounding  
119 boxes, and labels are provided with this software. Besides, Roboflow is used for auto-orient to  
120 discard common rotations by metadata and standardize pixel ordering, as well as resizing the  
121 images to a frame of 640x640 pixels from the original camera resolution of 1640x1232 pixels.  
122 This resolution of 640x640 is often suggested to facilitate the convenient use of object detection  
123 models, such as YOLOv5 [18]. Figure 3a shows an exemplary image before annotation and  
124 Figure 3b shows the same image after annotation. The purple box indicates the valid object,  
125 while the red box indicates the defective one. Table 1 shows the corresponding label information  
126 of Figure 3b. All boxes are applied comprehensively around the relevant objects, ensuring that  
127 occluded objects are always fully included. Besides, we aimed to minimize the spaces between  
128 the bounding box borders and the objects to ensure that only the relevant objects are enclosed  
129 within the box.



(a) Original image



(b) Labeled image

**Figure 3:** Exemplary image of the dataset consisting of two objects with one valid and one defective instance

130 Finally, the captured images and labels are stored in Zenodo and saved with a Data Management  
131 Plan (DMP) created with RDMO [19]. The DMP includes information about metadata, data  
132 formats, as well as technical insights to enhance scientific reuse within FAIR principles. F-UJI  
133 scored the resource with a FAIRness of 75 percent.

### 134 3 Object Detection and Quality Classification Showcase

135 While the presented dataset provides possibilities to perform various tasks, this section aims to  
 136 demonstrate the dataset's suitability for object detection and quality classification through binary  
 137 defect detection of the surface damages occurring on the objects. This showcase shall be a best  
 138 practice to learn and facilitate additional exploration. Additionally, training is conducted on  
 139 varying dataset sizes and class imbalances to demonstrate the performance and its relationship  
 140 with both the quantity of data and the degree of imbalance used during training. The variation in  
 141 dataset size and class imbalance is intended to address challenges faced by SMEs with limited  
 142 and imbalanced data. Therefore, we first explain the metrics used for this task, introduce the  
 143 algorithm trained, and then present its results across different dataset sizes and class imbalances.

#### 144 3.1 Metrics

145 As the task consists of binary defect detection on objects that need to be detected first, several  
 146 metrics need to be used. The object detection is measured by Intersection over Union ( $IoU$ ), as  
 147 suggested by literature [20]. This metric is based on the ratio of the area of intersection of two  
 148 bounding boxes to the area of union of two bounding boxes as shown in the Formula

$$IoU = \frac{\text{Area of Intersection of two bounding boxes}}{\text{Area of Union of two bounding boxes}} \quad (1)$$

149 Therefore, greater  $IoU$  values signify increased overlap and an improved prediction. To elim-  
 150 inate redundant boxes encompassing the same object,  $IoU$  typically employs Non-Maximum  
 151 Suppression. This method operates on the criterion that predictions with  $IoU$  lower than the  
 152 confidence threshold are ignored, while only boxes with  $IoU$  values exceeding this threshold  
 153 are retained. Here, the confidence threshold denotes the minimum score at which the model  
 154 considers a prediction to be valid. Furthermore, Precision ( $P$ ) and Recall ( $R$ ) as classification  
 155 metrics are applied to measure the accuracy of fault detection within detected objects. Generally,  
 156 an image typically contains a wealth of information, including both relevant and irrelevant  
 157 objects. To clarify this,  $P$  is introduced to only indicate relevant ones. It measures the proportion  
 158 of correctly recognized objects out of all detected objects.  $R$ , on the other hand, measures the  
 159 proportion of relevant objects that were correctly recognized by the model out of all relevant  
 160 objects. The mathematical definitions of  $P$  and  $R$  are shown in Formula 2 and Formula 3. True  
 161 Positive ( $TP$ ) represents correct detections ( $IoU \geq \text{confidence threshold}$ ), False Positive ( $FP$ )  
 162 represents a wrong detection ( $IoU < \text{confidence threshold}$ ), and False Negative ( $FN$ ) represents  
 163 a wrong misdetection.

$$\text{Precision}(P) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall}(R) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{TP}{TP + FN} \quad (3)$$

164  $P$  and  $R$  offer a trade-off that it is graphically represented in the PR curve by varying the  
 165 classification threshold. The area under this curve provides the average precision for each class

166 ( $AP_i$ ) for the trained model. The average of this value across all classes is referred to as the mean  
 167 Average Precision ( $mAP$ ), which is used to evaluate performance in object detection and quality  
 168 classification in this showcase, as it combines all introduced metrics. The equation is shown in  
 169 Formula 4. The equation is shown in Formula 4.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

170  $N$  corresponds to the total number of object classes.  $mAP$  has different categories, varying  
 171 in their parameter settings. We select the most common ones  $mAP@0.5$  and  $mAP@0.5:0.95$ .  
 172  $mAP@0.5$  is used across several benchmark challenges on datasets such as Pascal VOC or  
 173 COCO. It interpolates with 101 recall points with ( $IoU$ ) threshold = 0.5, which means that  $IoU$   
 174 values greater than or equal to 0.5 are considered  $TP$ , while values less than 0.5 are considered  
 175  $FP$  predictions.  $mAP@0.5:0.95$  uses the same interpolation method as  $mAP@0.5$ , but averages  
 176 the APs obtained from using ten different  $IoU$  thresholds (0.5, 0.55, ..., 0.95). The introduced  
 177 metrics  $P$ ,  $R$ ,  $mAP@0.5$  and  $mAP@0.5:0.95$  measure the performance of the algorithm during  
 178 training and in tests after training in this showcase.

### 179 3.2 Algorithm and Training

180 An algorithm of YOLO series is selected as an example real-time object detection algorithm  
 181 commonly used in research and industry. YOLO series object detection algorithms use a one-  
 182 stage neural network to directly complete detection object localization and classification without  
 183 using pre-generated region proposals [21], [22]. They are widely used for their good balance be-  
 184 tween high speed and high accuracy, easy implementation, and low-cost maintenance. YOLOv5,  
 185 proposed by Jocher Glenn [18], is selected as the YOLO version after consideration of com-  
 186 puting resources, layers of the network, model parameters, detection accuracy, inference time,  
 187 deployment ability, and algorithm practicability. The specific model YOLOv5 is used for its  
 188 properties of lightweight and relatively high speed. Since the size of the dataset in this showcase  
 189 is relatively small and the background information is fixed, real-time detection and high accuracy  
 190 can be ensured by YOLOv5s at the same time.

191 Training is conducted on smaller subsets of the dataset, as well as with class imbalances, to  
 192 demonstrate the model's performance to the number of images and the degree of class imbalance  
 193 used for training. Three different dataset sizes, with a class imbalance in the first two, are used  
 194 as shown in Table 2. To create the imbalanced datasets, images with fewer instances of the  
 195 'valid' class were selectively removed, resulting in a final dataset with around 65% of images  
 196 containing valid parts. The sizes of the training datasets are 310, 378, and 1050, respectively.  
 197 The validation and testing set sizes are 16% and 20% for the 1st and 2nd datasets, and 20% and  
 198 10% of the total data for the complete dataset. The algorithm is trained 300 epochs with a batch  
 199 size of 32 using YOLOv5s default hyperparameters.



	Training Set	Validation Set	Testing Set
1st(normal)	310	78	97
1st(imbalanced)	310	78	97
2nd(normal)	378	94	118
2nd(imbalanced)	378	94	118
3rd	1050	300	150

**Table 2:** Split of Training set, Validation set, and Testing set for all dataset sizes used

### 200 3.3 Evaluation

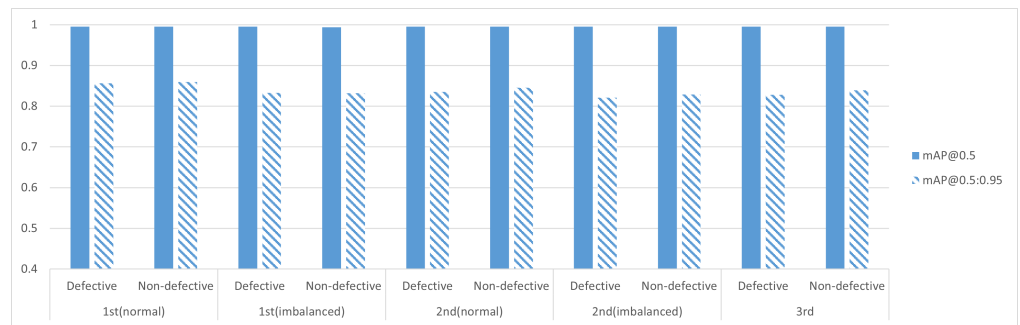
201 As introduced, the results are presented with  $P$ ,  $R$ ,  $mAP@0.5$  and  $mAP@0.5:0.95$  for validation  
 202 and testing set of the dataset and visualized in Table 3 and Table 4. The performance on the  
 203 validation set exceeds that of the testing set, indicating overfitting during training. Overall, the  
 204 performance on the testing data varies depending on dataset size and class imbalance. Regarding  
 205 the entire dataset, the trained model achieves a validation  $mAP@0.5$  of 0.995 and test a  $mAP@0.5$   
 206 of 0.668. The visualized comparison between the size of the dataset can be seen in Figure 4  
 207 for the validation data and in Figure 6 for the testing data. Despite this, there is no significant  
 208 performance increase, suggesting that even with the smallest dataset, satisfactory performance  
 209 in training, but not in testing is achieved. However, it is possible that more advanced models,  
 210 such as newer versions of YOLO, could achieve better test performance.

	Class	Precision	Recall	$mAP@0.5$	$mAP@0.5:0.95$
1st(normal)	All	0.99	1	0.995	0.858
	Defective	1	0.999	0.995	0.856
	Valid	0.98	1	0.995	0.859
1st(imbalanced)	All	0.994	0.988	0.994	0.833
	Defective	1	0.975	0.995	0.833
	Valid	0.988	1	0.994	0.832
2nd(normal)	All	0.989	0.983	0.995	0.84
	Defective	0.978	0.99	0.995	0.835
	Valid	1	0.976	0.995	0.845
2nd(imbalanced)	All	0.995	0.996	0.995	0.825
	Defective	1	0.992	0.995	0.821
	Valid	0.99	1	0.995	0.829
3rd	All	0.998	0.999	0.995	0.833
	Defective	0.997	1	0.995	0.828
	Valid	1	0.998	0.995	0.839

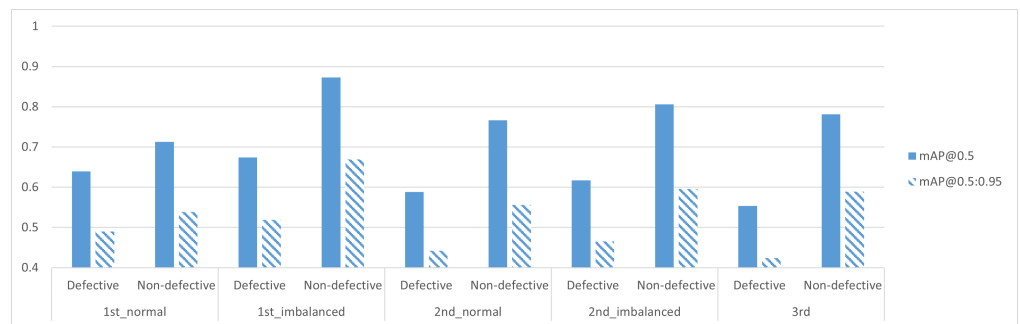
**Table 3:** Precision, Recall,  $mAP@0.5$  and  $mAP@0.5:0.95$  for the Validation Set

	Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
1st(normal)	All	0.763	0.708	0.676	0.514
	Defective	0.547	0.986	0.639	0.49
	Valid	0.979	0.43	0.713	0.539
1st(imbalanced)	All	0.733	0.794	0.774	0.594
	Defective	0.541	0.995	0.647	0.519
	Valid	0.926	0.592	0.873	0.669
2nd(normal)	All	0.687	0.714	0.677	0.499
	Defective	0.477	1	0.588	0.442
	Valid	0.897	0.429	0.766	0.556
2nd(imbalanced)	All	0.762	0.712	0.711	0.531
	Defective	0.543	1	0.617	0.465
	Valid	0.982	0.425	0.806	0.596
3rd	All	0.707	0.708	0.668	0.507
	Defective	0.473	0.997	0.554	0.424
	Valid	0.941	0.419	0.781	0.589

**Table 4:** Precision, Recall, mAP@0.5 and mAP@0.5:0.95 for the Testing Set



**Figure 4:** mAP@0.5 and mAP@0.5:0.95 metrics of validation set of all five dataset



**Figure 5:** mAP@0.5 and mAP@0.5:0.95 metrics of Testing set of all five dataset

## 211 4 Conclusion

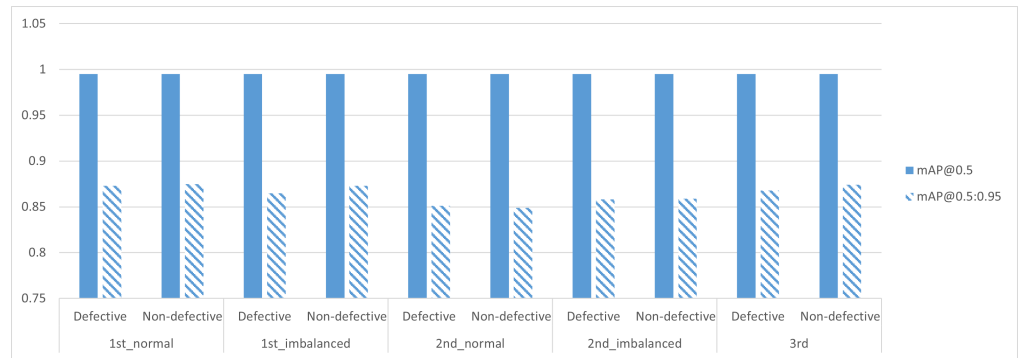
212 SMEs in the manufacturing sector lag behind their larger counterparts in the adoption of ML  
213 technologies like object detection. This is influenced by factors including insufficient data, high  
214 complexity, and a scarcity of tangible examples. We presented a simple low-resolution dataset  
215 based on plastic bricks with different surface defects to address a typical use case of object  
216 detection in manufacturing. By using a low-resolution dataset with a limited number of instances  
217 and accounting for class imbalances, we aimed to address typical challenges faced by SMEs.  
218 A showcase provided with a YOLOv5 model indicated sufficient performance with different  
219 metrics. Our findings show that maintaining simplicity does not compromise performance,  
220 demonstrating the effectiveness of straightforward open-source object detection methods and  
221 achieving an  $mAP@0.5:0.5$  score up to 0.995 in training and 0.774 in testing. These findings  
222 were published ensuring FAIR principles and achieved an FAIR score of 75 percent in F-UJI.  
223 The provided data and YOLO model can be reused for learning purposes and establish the  
224 groundwork for transferring knowledge to object detection tasks with similar surface damages  
225 on the objects to inspect. However, it's important to note that the limitation lies in the inability to  
226 directly apply such models or data to unrelated tasks. The consideration of the specific context  
227 is fundamental for the transferability of the presented methods. Additionally, it is important  
228 to recognize that industrial damages can significantly differ in the complexity of their defects.  
229 Future research should focus on investigating more universally applicable resources, facilitating  
230 direct transfer for use cases at SMEs through interoperable research approaches.

## 231 5 Usage Notes

232 The dataset generated for this research is accessible on Zenodo via DOI ([https://zenodo.o](https://zenodo.org/records/10731976)  
233 [rg/records/10731976](https://zenodo.org/records/10731976)). The dataset is licensed under the Creative Commons Attribution 4.0  
234 International License (CC BY 4.0). The developed algorithm is available on RWTH Aachen  
235 Gitlab ([https://git.rwth-aachen.de/zukipro/yolov5\\_for\\_plastic\\_brick\\_qualit](https://git.rwth-aachen.de/zukipro/yolov5_for_plastic_brick_quality_classification)  
236 [y\\_classification](https://git.rwth-aachen.de/zukipro/yolov5_for_plastic_brick_quality_classification)) and licensed under GNU Affero General Public License v3.0.

237 **6 Appendix**

	Class	mAP@0.5	mAP@0.5:0.95
1st (normal)	Defective	0.995	0.873
	Non-defective	0.995	0.875
1st (imbalanced)	Defective	0.995	0.865
	Non-defective	0.995	0.873
2nd (normal)	Defective	0.995	0.851
	Non-defective	0.995	0.849
2nd (imbalanced)	Defective	0.995	0.858
	Non-defective	0.995	0.859
3rd	Defective	0.995	0.868
	Non-defective	0.995	0.874

**Table 5:** mAP@0.5 and mAP@0.5:0.95 for the Training set**Figure 6:** mAP@0.5 and mAP@0.5:0.95 metrics of Training set of all five dataset238 **7 Acknowledgements**

239 The project "ZUKIPRO" is funded as part of the "Future Centers" program by the Federal  
 240 Ministry of Labour and Social Affairs and the European Union through the European Social  
 241 Fund Plus (ESF Plus).

242 **8 Roles and contributions**

243 **Jonas Werheid:** Conceptualization, Writing – original draft, Writing – review & editing

244 **Shengjie He:** Conceptualization, Writing – original draft

245 **Tobias Hamann:** Writing – review & editing

246 **Anas Abdelrazeq:** Writing – review & editing

247 **Robert H. Schmitt:** Funding acquisition & Supervision

248 **References**

- 249 [1] M. Rudorfer, *Towards Robust Object Detection and Pose Estimation as a Service for*  
250 *Manufacturing Industries*. Fraunhofer IRB Verlag, 2021.
- 251 [2] L. Malburg, M.-P. Rieder, R. Seiger, P. Klein, and R. Bergmann, “Object detection for  
252 smart factory processes by machine learning,” *Procedia Computer Science*, vol. 184,  
253 pp. 581–588, 2021, The 12th International Conference on Ambient Systems, Networks  
254 and Technologies (ANT) / The 4th International Conference on Emerging Data and  
255 Industry 4.0 (EDI40) / Affiliated Workshops, ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2021.04.009>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921007821>.
- 258 [3] M. Bauer, C. van Dinther, and D. Kiefer, “Machine learning in sme: An empirical study  
259 on enablers and success factors,” 2020.
- 260 [4] P. Ulrich and V. Frank, “Relevance and adoption of ai technologies in german smes –  
261 results from survey-based research,” *Procedia Computer Science*, vol. 192, pp. 2152–2159,  
262 2021, Knowledge-Based and Intelligent Information Engineering Systems: Proceedings  
263 of the 25th International Conference KES2021, ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2021.08.228>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921017245>.
- 266 [5] C. M. Martin Lundborg. “Künstliche intelligenz im mittelstand: Relevant, anwendungen,  
267 transfer.” (), [Online]. Available: [https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?\\_\\_blob=publicationFile&v=5](https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5).
- 270 [6] H. van Vlijmen, A. Mons, A. Waalkens, *et al.*, “The Need of Industry to Go FAIR,” *Data*  
271 *Intelligence*, vol. 2, no. 1-2, pp. 276–284, Jan. 2020, ISSN: 2641-435X. DOI: [10.1162/dint\\_a\\_00050](https://doi.org/10.1162/dint_a_00050). [Online]. Available: [https://doi.org/10.1162/dint%5C\\_a%5C\\_00050](https://doi.org/10.1162/dint%5C_a%5C_00050).
- 274 [7] T. DATA. “Safety helmet detection dataset.” (), [Online]. Available: <https://www.kaggle.com/datasets/trainingdatapro/helmet-detection/>.
- 276 [8] TubeData1. “Tube object detection image dataset.” (), [Online]. Available: <https://universe.roboflow.com/tube/tube-object-detection/dataset/1>.
- 278 [9] Ruben. “Aughmanity<sub>v2.0</sub>computervisionproject.” (), [Online]. Available: [https://universe.roboflow.com/ruben-8bqya/aughmanity\\_v2.0](https://universe.roboflow.com/ruben-8bqya/aughmanity_v2.0).
- 280 [10] H. Z. Jianfeng Zheng Hang Wu. “Insulator-defect detection dataset.” (), [Online]. Available:  
281 <https://datasetninja.com/insulator-defect-detection>.
- 282 [11] G. L. Runwei Ding Linhui Dai. “Augmented pcb defect dataset.” (), [Online]. Available:  
283 <https://datasetninja.com/augmented-pcb-defect>.
- 284 [12] M. Gribulis. “Synthetic lego brick dataset for object detection.” (), [Online]. Available:  
285 <https://www.kaggle.com/datasets/mantasgr/synthetic-lego-brick-dataset-for-object-detection/>.
- 286

- 287 [13] DREAMFACTOR. “Largest lego dataset (600 parts).” (), [Online]. Available: <https://www.kaggle.com/datasets/dreamfactor/biggest-lego-dataset-600-parts/data>.  
288  
289
- 290 [14] “F-uji.” (), [Online]. Available: <https://www.f-uji.net/?action=test>.
- 291 [15] A. de Giorgio, G. Cola, and L. Wang, “Systematic review of class imbalance problems in  
292 manufacturing,” *Journal of Manufacturing Systems*, vol. 71, pp. 620–644, 2023, ISSN:  
293 0278-6125. DOI: <https://doi.org/10.1016/j.jmsy.2023.10.014>. [Online].  
294 Available: [https://www.sciencedirect.com/science/article/pii/S02786125](https://www.sciencedirect.com/science/article/pii/S0278612523002157)  
295 [23002157](https://www.sciencedirect.com/science/article/pii/S0278612523002157).
- 296 [16] H. K. Kondaveeti, N. K. Kumaravelu, S. D. Vanambathina, S. E. Mathe, and S. Vappangi,  
297 “A systematic literature review on prototyping with arduino: Applications, challenges,  
298 advantages, and limitations,” DOI: [https://doi.org/10.1016/j.cosrev.2021.10](https://doi.org/10.1016/j.cosrev.2021.10.0364)  
299 [0364](https://doi.org/10.1016/j.cosrev.2021.10.0364).
- 300 [17] B. Dwyer, J. Nelson, and J. e. a. Solawetz. “Roboflow (version 1.0) [software].” (),  
301 [Online]. Available: <https://roboflow.com>.
- 302 [18] G. Jocher, *Ultralytics yolov5*, version 7.0, 2020. DOI: [10.5281/zenodo.3908559](https://doi.org/10.5281/zenodo.3908559).  
303 [Online]. Available: <https://github.com/ultralytics/yolov5>.
- 304 [19] “Research data management organiser.” (), [Online]. Available: [https://rdmorganise](https://rdmorganiser.github.io/)  
305 [r.github.io/](https://rdmorganiser.github.io/).
- 306 [20] H. Rezatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized  
307 intersection over union: A metric and a loss for bounding box regression,” in *Proceedings*  
308 *of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun.  
309 2019.
- 310 [21] X. Xie, K. Chen, Y. Guo, B. Tan, L. Chen, and M. Huang, “A flame-detection algorithm us-  
311 ing the improved yolov5,” *Fire*, vol. 6, no. 8, 2023, ISSN: 2571-6255. [Online]. Available:  
312 <https://www.mdpi.com/2571-6255/6/8/313>.
- 313 [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep con-  
314 volutional neural networks,” in *Advances in Neural Information Processing Systems*, F.  
315 Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25, Curran Associates, Inc.,  
316 2012. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/pape](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)  
317 [r/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf).