

Rail Line Surfaces Defect Monitoring using YOLO Architecture: Case Study on Madiun-Magetan Track, East Java

Hanum Arrosida¹

¹Department of Computer Control Engineering,
Politeknik Negeri Madiun, Indonesia

Agus Susanto², Adiratna Ciptaningrum², Tyan Rudianti², Masayu Nazar Surya Kencana²

²Department of Rolling Stock Engineering, Politeknik Negeri Madiun, Indonesia

Rizal Mahmud³

³Department of Mechanical Engineering and Informatics,
School of Science and Technology, Meiji University, Japan

Abstract:- In Indonesia, trains are one of the most popular means of transportation for Indonesians to help with the mobility of passengers and goods. However, train derailment is also something that happens quite frequently. The train derailment was caused by several factors, the rail line damage is one of the biggest possible causes. For this reason, it is necessary to carry out inspections on it to detect and find defects on the rails for subsequent repairs. Manual inspections, as is still often done by officers in Indonesia, have shortcomings such as low efficiency, human error, and danger. Automatic inspection can shorten inspection time, reduce maintenance costs, and data can be real time. The aim of this research is to create an automatic inspection system using You Only Look Once (YOLO) algorithm to rail line detect in Indonesia by taking case studies in train operational areas along tracks that pass through in two cities, namely from the Madiun Station to West Station, in East Java Province. This area is known as Daerah Operasi 7 (DAOP 7) with the 14 km in distance. The result showed that the detection system using the YOLO model had mAP value of 99.41%, a precision value of 99%, a recall value of 99%, an f-score value of 99%, and an average IoU value of 85.84%. The YOLO model can detect railway track surface abnormalities accurately and optimally. Therefore, it can be used an automatic inspection for monitoring rail line in Indonesia generally and rail line in East Java Province, especially.

Keywords:- Train; popular means of transportation; rail line damage monitoring; DAOP 7 rail track; YOLO.

I. INTRODUCTION

Rail transportation currently plays an important role in the daily lives of the Indonesian society, both for passenger and goods movement. Every year, train users in Indonesia increase significantly. Indonesia Railway Operator (PT. KAI) officially recorded a volume of train users in 2022 of 119.8 million passenger and it is increasing 42% compared to 2021 [1]. This increase also has a direct impact on increasing the frequency of train travel for passengers and

goods so that the number of train transportation accidents increases relatively [2]. Based on railway accident investigation data from 2017 to 2021 issued by the Indonesia Transportation Safety Committee (KNKT) [3], the majority of accidents were train derailments, the number of which reached 21 accident cases with the main cause coming from infrastructure factors.

Train derailment is a condition when a train runs off the track caused by mechanical disturbances to the rail such as wear, broken rails, cracked rails, and supporting components of the railroad structure such as rail tethers and rail sleepers that do not function properly [4]. Whereas, good track conditions will make train travel safe and comfortable, so that train tracks become one of the important aspects supporting smooth train travel [5]. Therefore, it is very important to detect defects on the rail line surface.

It must be acknowledged that the method for checking defects on the railroad tracks surface in operational work areas in Indonesia is still carried out by means of visual inspection, where the inspectors is tracing the rail track [6, 7]. Of course, this method is very ineffective in Indonesia, considering the length of Indonesia's railway track reaches 6.32 million km which makes it the longest rail track among ASEAN countries and ranks 26th in the world [8,9]. Apart from that, the presence of human error, low efficiency, danger, and subjective nature mean that this method must also be updated [10].

There are several sophisticated methods recommended for detecting surface abnormalities of railway tracks, for example based on signal analysis. This is because several features will be generated to provide information about the condition of a mechanical structure by applying various signal processing tools, such as fast Fourier (FFT), short-time Fourier (STFT), and Hilbert-Huang (HHT) transforms. Lei et al. proposed wavelet transform based train track monitoring [11]. Zhai et al. identified the vibration behavior of the car body, bogie frame and axle box of a Chinese high-speed (CRH) train operating at a speed of 350 km/hour due to track abnormalities [12].

Apart from applying signal processing, vibration responses can also be analyzed using statistical methods. In this case, principal component analysis (PCA) is generally applied to detect train track damage. This is because PCA is a standard multivariate technique that is able to reduce the dimensions of the data so that it is computationally light [13]. However, PCA was weak if the raw data is filled with noise. This is because the majority of raw data acquired from the field, including vibrations measured from trains, are backgrounded with noise that is almost impossible to eliminate with conventional filters, such as low pass filters (LPF) or high pass filters (HPF) [14]. This is because LPF and HPF only cut raw data at low and high frequencies which results in the loss of some important information [15-19].

Other recommended methods are ultrasonic flaw detection, eddy current detection, parallel laser, and ground penetration radar. However, due to various external interferences, the resulting signal is difficult to handle and the use of this method does not achieve maximum performance results [20]. Apart from that, the jigs and fixtures and the reliability of using equipment in this method are still in the development stage considering that the equipment must deal with extreme weather, costs and limited facilities in an area, for example train tracks in rural areas [21]. Therefore, these methods still need further improvement and development so that a method for detecting rail surface defects is needed that can be carried out quickly and efficiently, and reduces inspection costs in order to improve the safety of rail transportation [6].

One method of monitoring train track abnormalities can be done by detecting objects through image processing. Convolutional Neural Network (CNN) as a method of deep learning has high accuracy when used for image recognition [22]. Some examples of algorithms derived from CNN include R-CNN, Faster R-CNN, Mask R-CNN [23]. Zhang et al. applied Faster R-CNN for drainage pipe damage detection which reached 83% mAP [24]. Li, et al. adopted ZF-Net as the Faster-RCNN backbone and added a maxpooling layer in the head network to detect defects on a large scale, and they have achieved 80.7% mAP [25]. However, this algorithm has a drawback, namely the high computational complexity that makes this possible bottleneck because this method uses a two-stage object detector type algorithm [26].

On the other hand, You Only Look Once (YOLO) can shorten computing time and is lighter because it uses a one stage object detector algorithm, where the regression and classification processes are combined in one network [27]. Zhang, et al. uses YOLOV3 with batch regularization and focus loss to detect bridge surface damage, which achieves good performance [28]. Yin, et al. used YOLOV3 to detect sewer pipe damage defects and obtained 85.37% mAP [29]. Deng, et al. uses YOLOV2 with graffiti interference to detect cracks and defects in concrete surface with complex background [30]. They claim that the accuracy of this method is even higher than RCNN, namely 77% versus 74.5% mAP, and has higher performance, namely 0.17 seconds versus 0.23 seconds. Thus, the single-stage method

is simpler and faster, which is more suitable for end-to-end online defect detection in various industrial fields.

Although deep learning-based object detection methods, including YOLO, have been partially studied in industry, most of them are still in the laboratory stage and are difficult to implement for two reasons. First, the surface defects of industrial products are very complex and vary in scale, it is difficult to detect and locate defects of various sizes in a large background area. Second, online detection has a very high demand for real-time performance, but most studies ignore its speed. However, there is still potential for improvement when applied to surface defect inspection in industrial products.

In this research, YOLOV3 will be used as a method to detect surface abnormalities on Indonesian railway tracks with high accuracy and speed. This is because Indonesian railways have unique standards for train track width (gauge length) compared to train tracks in other countries in the world. The width of the rail is only 1.04 m so it is called a narrow-gauge rail line, which has implications for the projection of the track geometry of each rail line or the central axis of the track according to the horizontal plane. Due to the narrowness of this rail, mechanical and dynamic strength of rail structures can degrade more quickly over time due to the load and large amounts of energy carried by trains. This will certainly have an impact on comfort and safety for passengers and goods.

First, the DarkNet-53 network is implemented as a backbone network. Second, an enhanced feature pyramid network (EFPN) structure is built specifically for small-sized object detection. At the same time, a new feature fusing module (FFM) was designed to better integrate EFPN cross-scale features. In addition, an Intersection Over Union (IoU) loss branch is added to improve the bounding box positioning accuracy and narrow the detection gap between one-stage and two-stage methods. In the end, we evaluate the proposed method on NEU-DET, and the results can show clear superiority over other methods.

II. PROPOSED DEFECT DETECTION METHOD

You Only Look Once (YOLO) is a part of the Convolutional Neural Network (CNN) that can be utilized for detecting various properties of objects in the form of photos through image processing in real time with high accuracy. He first introduced by Redmon, J. et al. [30] eight years ago. In contrast to its competitors in the same class, for example R-CNN which uses two detection stages, YOLO only uses a single network structure to detect object properties. So, it is very applicable for use in detecting surface defects in industries that require fast recognition times, such as in the railway industry.

YOLO then underwent several versions in its development, namely YOLO-v1, YOLO-v2, and YOLO-v3. In YOLO-v1, each grid cell can only detect one object whose center point falls on that grid cell. Because of this, the number of detected adjacent objects is limited, especially small objects that appear in a group [30]. Additionally, YOLO-v1 only learns predictions based on

the data provided, making it difficult to detect new objects that have different aspect ratios and configurations. Based on these reasons, YOLO-v2 appears to improve the limitations contained in YOLO-v1. YOLO-v2 focuses more on developing recall and using the Darknet-19 framework which can improve its performance. Furthermore, YOLO technology is always being developed so that YOLO-v3

appears which has faster performance with higher accuracy than before. Different from the two versions, YOLO-v3 has a degenerate anchor box mechanism so that it allows the detection of two different objects even though the centers of both objects fall on the same grid cell, the object is small, and the computational process is light [31].

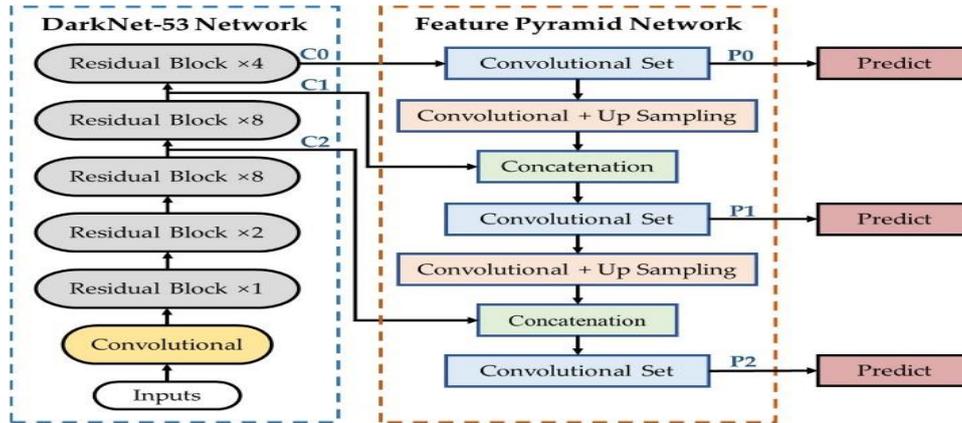


Fig. 1: Flowchart of Yolo-v3 process [30,31]

To be able to detect objects well, YOLO-v3 uses logistic regression technology in its bounding box and uses an independent logistic classifier, DarkNet-53 as a backbone network consisting of 53 convolutional layers, utilizing a Graphics Processing Unit (GPU) and residual blocks for feature extraction and filters measuring 3x3 and 1x1 as shown in Fig. 1. DarkNet-53 Network is a combination of sampled feature maps. The constructed feature pyramid network produces three feature maps for bounding box regression and target classification, respectively, to obtain outputs of different scales.

YOLO-v3 to perform multilabel classification, which indicates that an object can be recognized with more than one class prediction at the same time. Apart from that, YOLO-v3 also applies three different strides at each scale with respective sizes of 32, 16, and 8. When there is an input object in the form of an image measuring 416 x 416, the system will carry out detection at a scale of 13 x 13, 26 x 26, and 52 x 52. At each convolutional layer, YOLO-v3 applies the Leaky ReLU activation function and Batch Normalization which is a learning technique that standardizes the input to the layer for each mini batch. This can certainly speed up the training process and stabilize the learning process. In object detection, YOLO-v3 adds 53 layers after the feature extractor for a total of 106 layers.

Fig. 2 features the YOLO-v3 architecture for predicting classes. Based on this architecture, YOLO-v3 uses independent logistic classifiers, making it possible for

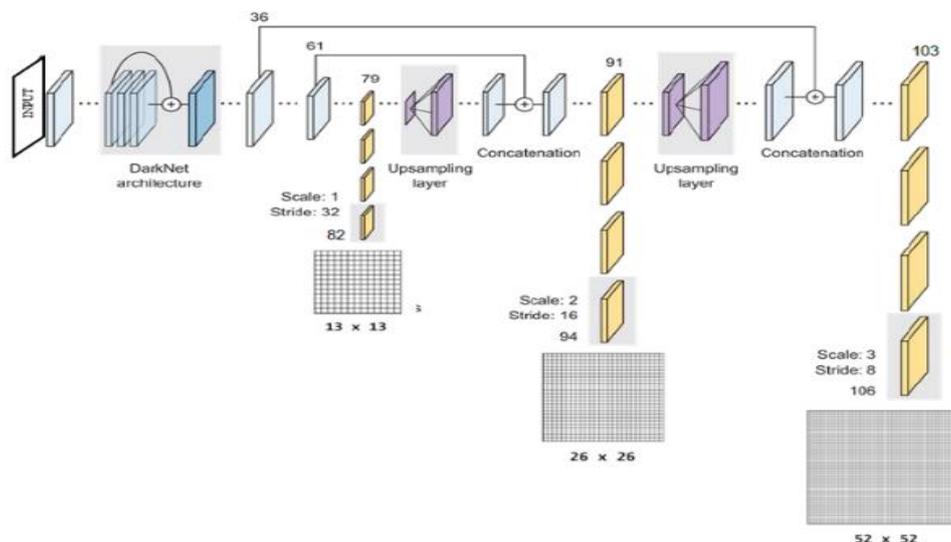


Fig. 2: YOLO-v3 architecture for class prediction



Fig. 3: YOLO-v3 performance in detecting living and non-living creatures



Fig. 4: YOLO-v3 performance in detecting three different types of things



Fig. 5: YOLO-v3 performance in detecting object collections

The YOLO-v3 model itself has been trained and can detect objects of 80 classes in the Common Objects in Context (COCO) dataset. Some examples of objects that can be detected with the same weight are person, bicycle, car, motorbike, dog, cat, horse, bird, train, truck, airplane, bus, etc. This model is often called a pre-trained model. By using this model, object detection can be carried out directly without carrying out training first. However, for custom datasets other than the 80 class COCO dataset, it is necessary to train the model first to get the weights for the custom dataset.

To test YOLO-v3's ability to detect objects, several different case examples are given as shown in **Fig. 3 to Fig. 5**. **Fig. 3** shows that the YOLO-v3 model is able to detect

living and non-living creatures, namely bicycle and person, even though it uses the same weights and models. This shows that YOLOv3 is able to detect multilabel classification objects correctly. **Fig. 4** displays the results of detecting students and chair objects in our Lab. with correct and accurate detection results. Meanwhile, **Fig. 5** shows the ability of YOLO-v3 to detect objects correctly even if they are in a collection or more than one, of different colors and genders, in this case it is "spoon". To measure the performance of the classification model, the confusion matrix is used because it contains information that compares the classification results of the model (prediction) with the actual values (ground truth). **Table 1** shows the confusion matrix table with four different combinations.

Table 1: Four combinations of predicted values using the Confusion Matrix

		Actual Values	
		1	0
Predicted Values	1	TP (True Positive)	FP (False Positive)
	0	FN (False Negative)	TN (True Negative)

Determining whether a classification model is good or not can be seen from the system performance measurement parameters including precision, recall, F1-score, mean average precision (mAP), and accuracy. Precision is used to measure the percentage of correct predictions by how accurate the predictions from the model are with the following equation;

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall or sensitivity (True Positive Rate) shows how good the model is at finding all true positives for all classes which are formulated as follows;

$$recall = \frac{TP}{TP + FN} \tag{2}$$

F1-score is a comparison of the average precision and recall. The highest value of the F1-score is 1, while the lowest value is 0. The F1-score is written as follows;

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{3}$$

Mean Average Precision(mAP) is used to evaluate the object detection model. The higher the mAP value, the more accurate the model detection will be. In other words,

mAP shows how good the model is at detecting objects. mAP is the average of average precision (AP) for the entire class which is formulated as follows;

$$mAP = \frac{1}{K} \sum_{n=1}^K AP_n \tag{4}$$

where AP_n is AP class n, K is the number of classes.

Accuracy is a comparison between True Positive and True Negative with the entire data. Accuracy shows the level of closeness of predicted values to actual values.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

III. RESEARCH METHODS

A. Research design

The research will utilize the advantages of YOLO-v3 to detect abnormalities in train tracks in Indonesia. This is to build a system for detection defects on railway tracks based on deep learning so that it can be developed for automatic rail inspection. The system was built using the YOLOv3 algorithm with the Darknet53 framework. Fig. 6 shows the flow diagram of the detection system.

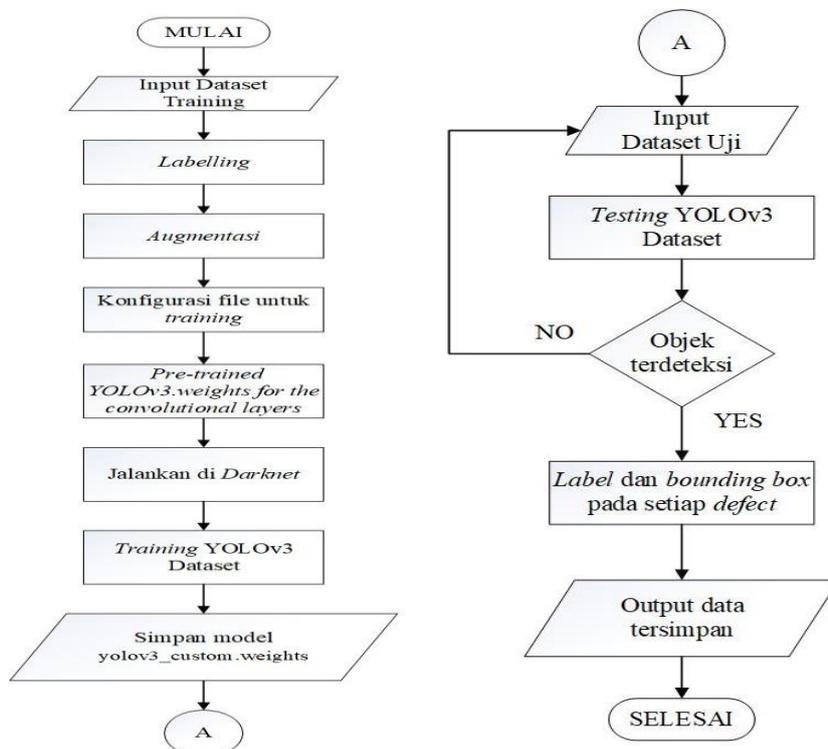


Fig. 6: Flow diagram of the railway track surface abnormality detection system based on YOLO-v3.

B. Datasets

A custom dataset is needed to train the desired model. A dataset of 600 original photos and 840 augmented images of deformed train tracks has been prepared. The types of train track images chosen include photos containing spalling, squat, cracking and broken off. These four types of defects will become classes in this research. These images were taken directly from rail line tract surface in the train operational area along the Madiun Station to Magetan Station, in East Java, Indonesia. This area is known as

Daerah Operasi 7 (DAOP 7) with a train track distance is about 14 km. The rail track where the data was collected is shown in **Fig. 7**.

The dataset that has been collected is a 2250 x 4000-pixel photo with ".jpg" format displayed in **Fig. 8** as a whole. As previously explained, the types of train track defects to be analyzed include pictures of squat, spalling, crack and break defects which are shown in **Fig. 9**.



Fig. 7: Dataset Retrieval Map in DAOP 7, along 14 km from Madiun to Magetan Station, in East Java, Indonesia

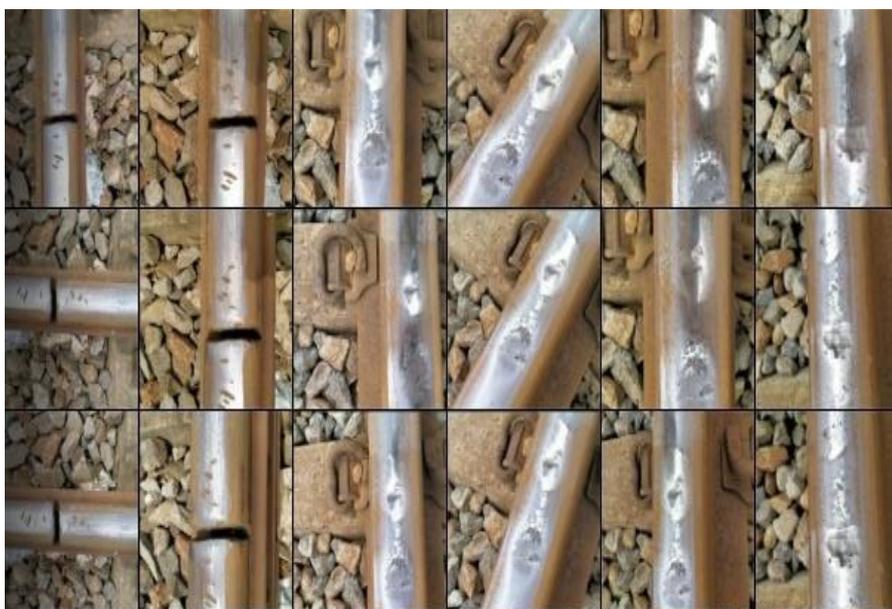


Fig. 8: Photos of various types of train track damage as a dataset

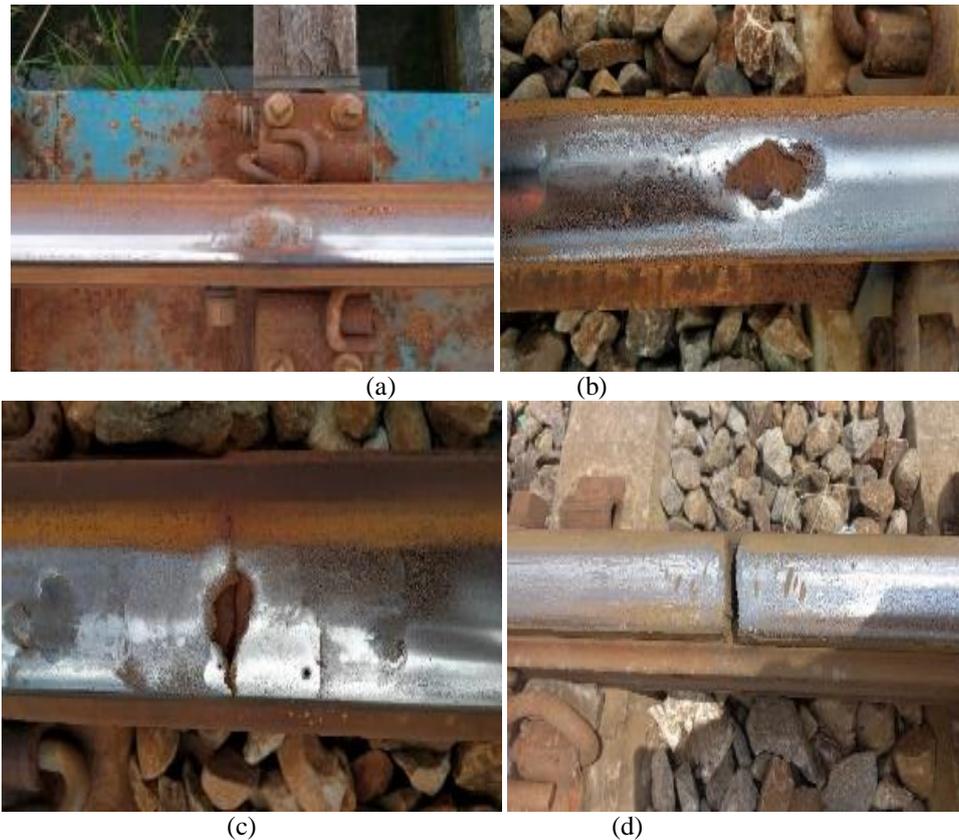


Fig. 9: Dataset consists of squat, spalling, crack and broken-off

As a requirement for correct input data, the dataset needs to undergo preprocessing which includes sorting data, resizing data, and dividing data. The image size is equalized to 416 x 416-pixels and divided into 3 division ratios, namely 70% data set for training, 20% set for validation, and 10% set for testing. This division is done automatically using the roboflow feature. The training set is used to train the model so that it requires more images for optimal learning results. The validation set is used to optimize model parameters and select the best model. Having a validation set can make it easier to measure performance and compare model performance with different configurations. Meanwhile, the testing set is used to evaluate model performance so that it can be used to calculate evaluation matrices such as accuracy, precision, recall, etc. The data used in the testing set is not used in the

training set and validation set so that the three data sets have different data and are new data.

IV. RESULTS AND DISCUSSION

This research uses four classes, namely squatting, spalling, cracking, and breaking up. The data is resized to match the overall image size, namely to 416 x 416 pixels. The dataset is then labeled to mark the ground truth area of the object that will be used to train the network model. Labeling is done using the annotation tool robo flow. Labeling results are saved in a .text file. The file contains five values, each representing the class, x coordinate, y coordinate, bounding box width, and bounding box height. **Fig. 10** shows the labeling process using roboflow on one of the datasets where the labeling process has been carried out.



Fig. 10: Labeling process and labeling results files

The labeling results are in the form of two files, namely the annotated .jpg file and the .txt file. The .txt file contains five values, each of which represents the class, x coordinate, y coordinate, bounding box width, and bounding box height. The .txt file of each image has

different values even though they are of the same class because these values are influenced by the labeling process. **Fig. 11** shows one of the .txt files resulting from image labeling.

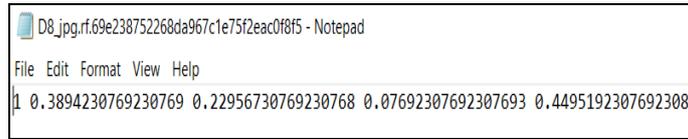


Fig. 11: Coordinate of labeling result files

After that, to increase the number and variety of datasets, the augmentation method was carried out. **Table 2** shows the data augmentation method used in this research.

Table 2: Data Augmentation Method

No	Data Augmentation
1	Crop: 0% min zoom, 50% max zoom
2	Grayscale 15%
3	Hue between -20° and +20°
4	Brightness between -30% and +30%
5	Exposure between -20% and +20%
6	Blur up to 1px
7	90 rotate: Clockwise, counter clockwise, upside down
8	Bounding box 90 rotate
9	Bounding box brightness
10	Bounding box exposure

Fig. 12 shows the augmentation results with a crop setting of 0% shown in **Fig. 12 (a)** and a crop of 50% in **Fig. 12 (b)**. The results of this augmentation are to increase the dataset used for the rail defect detection system. Apart from that, the crop augmentation results are useful for simplifying the detection system for various possible

conditions in rail defects, especially in setting the distance between the camera and the detected rail. In **Fig. 13**, there are augmentation results with a grayscale setting of 15%. The results of this augmentation setting aim to facilitate the detection of rail defects in various possible situations, including gray level conditions.



Fig. 12: Augmentation crop 0% (a) and crop 50% (b)



Fig. 13: 15% grayscale augmentation

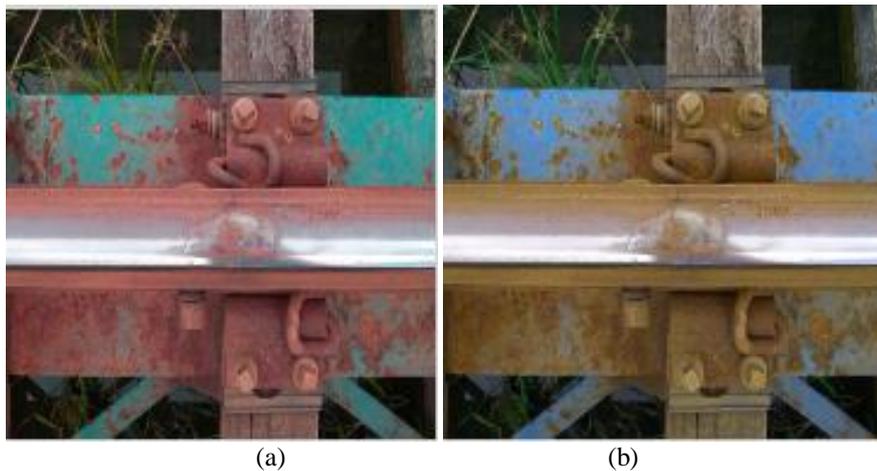
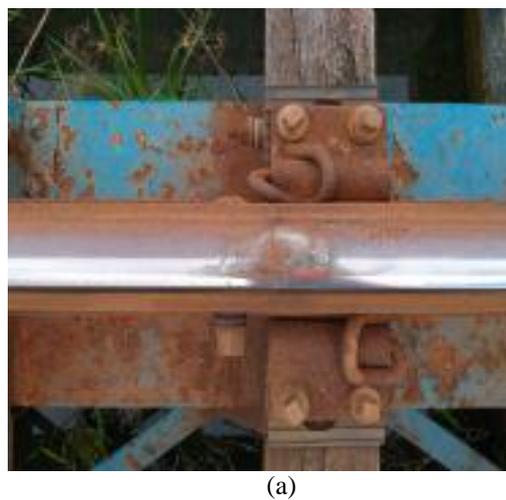


Fig. 14: Augmentation of *hue* -20° (a) and *hue* 20° (b)

Fig. 14 shows the augmentation results of the basic hue or hue settings, where in **Fig. 14(a)** the hue setting is -20° and **Fig. 14(b)** the hue setting is 20°. The results of this

augmentation setting aim to make it easier to detect rail defects in various conditions that occur, including basic color conditions.



(a)



(b)

(c)

Fig. 15: Brightness augmentation 0% (a), -30% (b), 30% (c)

Fig. 15 shows the augmentation results of the brightness or light intensity settings, whose **Fig. 15(a)**, **Fig. 15(b)**, and **Fig. 15(c)** are 0% for the brightness setting, -30% for the brightness setting, and 30% for the brightness

setting. The results of this augmentation setting aim to make it easier to detect rail defects in various conditions that occur, including conditions where light intensity is always changing.



(a)



(b)



(c)

Fig. 16: Exposure augmentation 0% (a), -20% (b), 20% (c)

Fig. 16 shows the augmented results of the exposure settings or light exposure, where in **Fig. 16(a)** the exposure setting is 0%, **Fig. 16(b)** the exposure setting is -20% and **Fig. 16(c)** the exposure setting is 20%. The results of this

augmentation setting aim to facilitate the detection of rail defects in various conditions that occur, including light exposure conditions.



(a)



(b)

Fig. 17: Blur augmentation 0 px (a), 1 px (b)

Fig. 17 shows the augmentation results of the blur settings, where in **Fig. 17(a)** the blur setting is 0 px and **Fig. 17(b)** the blur setting is 1 px. The results of this

augmentation setting aim to make it easier to detect rail defects in various conditions, including blur conditions.

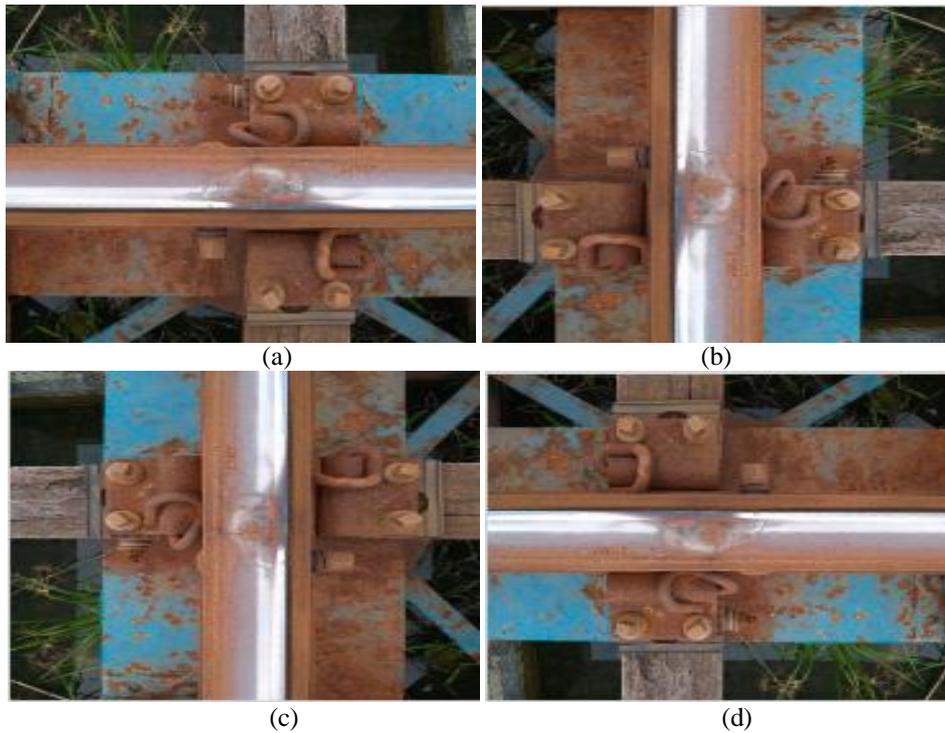


Fig. 18: Rotate 90°; (a) preprocessed, (b) clockwise, (c) counter clockwise, and (d) upside down

Fig. 18 shows the augmentation results of the rotation or corner position settings, where in **Fig. 18(a)** the preprocessed image position settings, 18(b) the clockwise position settings, 18(c) the clockwise counter position settings and 18(d) the upside-down image position settings. The results of this augmentation setting aim to facilitate the

detection of rail defects in various conditions that occur, including image corner conditions.

Before carrying out the training process, the configuration file needs to be set. Table 3 shows the settings required in the training process.

Table 3: Data settings required in training process

Hyperparameters	Mark
Batches	64
Subdivisions	16
Learning rate	0.001
Max batches	8000
classes	4
Width, height	416, 416
steps	6400, 7200
Filter convolutional layers before YOLO	30

The files required for setup include "file.cfg", "obj.names", "obj.data", and "generating_train.txt". The "obj.names" file contains the classes used in the research including crack, spalling, squat, and breaking. Meanwhile, the "obj.data" file shows the data storage path that will be used during the training process.

The result of this research is a system built using the YOLOv3 algorithm with a darknet framework to detect abnormalities on railway tracks. The trained model

obtained a Mean Average Precision of 99.41% with a confidence threshold of 0.25. Based on these values, the previously trained model has high performance in detecting defects on railway tracks. **Table 4** shows the average precision values for each class and the True Positive (TP) and False Positive (FP) values. Several other assessment parameters such as precision, recall, and F1-score of the entire class are also shown in **Table 5**.

Table 4: AP scores for each class

Class	Rail Defect	Average Precision	True Positive	False Positive
0	Crack	98%	481	23
1	Broken off	100%	123	0
2	Spalling	99,62%	855	9
3	Squat	100%	1162	7

The table shows the average precision, True Positive, and False Positive values of each rail defect detection class. The average precision range for the four classes is between 98% and 100% and there are also several TP and FP detected. A high TP value indicates that the system really identifies objects in the dataset with the correct conditions. Meanwhile, a low FP value indicates that the system made an error in detecting objects of small value. TP and FP are used to measure system performance and evaluate object detection accuracy values. The comparison of AP scores from each class is more clearly visualized in the graph in Fig. 19.

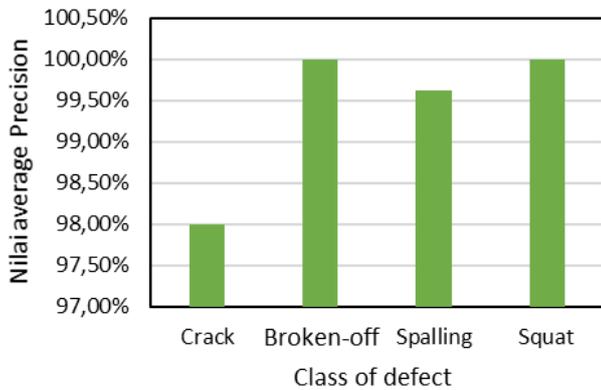


Fig. 19: Data average precision for each class

Based on this graph, it can be concluded that detection in the squat and breaking defect classes has the most optimal AP value compared to the detection of two other rail defects, namely; cracking and spalling. Next, let's look at the results of the analysis of the number of iterations, mAP, precision, recall value, F1 Score, and average IoU as shown in Table 5. This table shows the model values consisting of 8000 iteration values, mAP of 99.41%, precision with a value of 99%, recall value of 99%, F1 Score 99%, and average IoU with a Fig. of 85.84%. These values have basically achieved the desired results in the training process.

During the training process, data.weights will be stored in Google Drive after every 1000 iterations which have previously been connected. This is another advantage of the rail defect detection system which is being built to avoid weight loss due to internet network instability or other problems which are usually experienced by train monitoring stations in Indonesia so there is no need to repeat training from the start.

Table 5: Assessment Parameters Model

Parameters Model	Assessment
Max batches/iterations	8000
Folder	99.41%
Precision	99%
Recall	99%
F1-score	99%
Average IoU	85.84%

Figure 20 shows a chart of the relationship between average loss and iteration which is the result after the training stage. Based on this curve, losses start above 18, then drop sharply. When the iteration reaches 800 times, the loss reduces to 1.2. With the continuous increase of the learning rate, the loss continues to decrease in a small range, reducing to 0.1795 when repeated it becomes 8000 times.

The evaluation matrix of the trained model is also presented in Fig. 21 which shows the model performance between the training and validation processes. It can be seen that the box loss, obj loss and class loss graphs show a decreasing trend, which means that the relative loss during training decreases with learning. Inversely proportional to the precision, recall and mAP values, these curves show an increasing trend. In terms of initial accuracy and precision it is 0, then increases drastically as the learning rate increases. When training is repeated 800 times, accuracy reaches a value of 1, which indicates that the model has high precision and recall as well as the mAP curve. The results of the matrix evaluation show that the model has good performance.

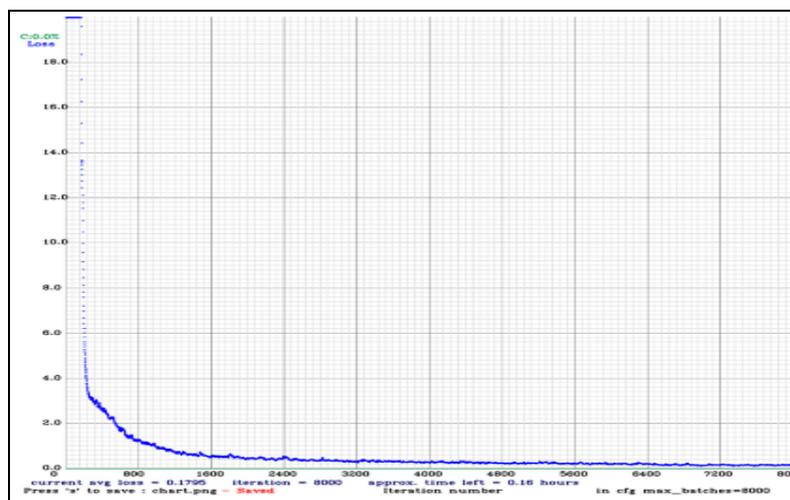


Fig. 20: Average loss vs iteration curve

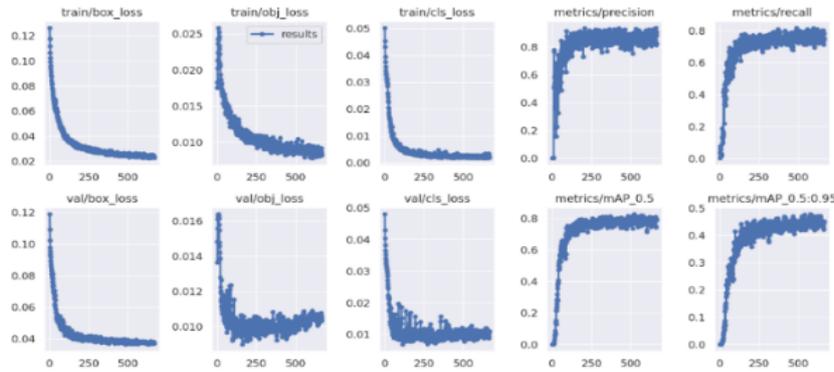


Fig. 21: Training graphs

The training results are in the form of model weights in the format.weights. This model will be used for data testing to run object detection with custom datasets. The following are the results of object detection with a custom dataset.

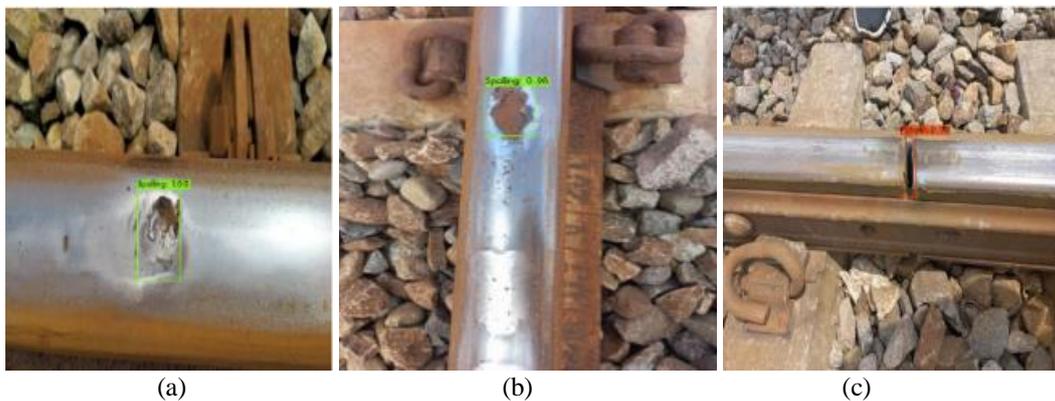


Fig. 22: Railway track defect detection

Fig. 22 predicting detection of railroad track defects in only 1 class, where in the image there are spalling and broken rail defects. This detection system is able to detect rail defects well, as evidenced by the confidence value which is quite high and is able to detect rail defects correctly. Fig. 22(a) shows the model in detecting spalling

defects which shows the prediction confidence value reaches 1.00. This means that the image is 100 correct in the form of a spalling defect. Meanwhile, Figs. 22(b) and 22(c) respectively show spalling and breaking defects with respective prediction confidence values of 0.96 and 0.99.

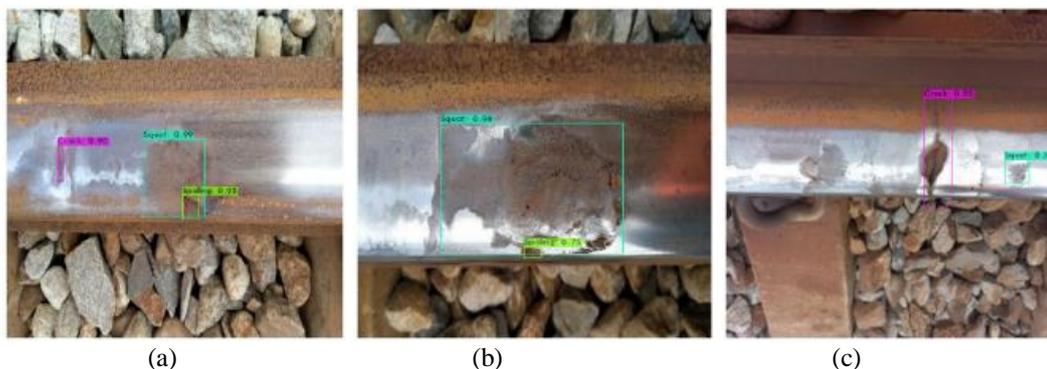


Fig. 23: Multi Detection of railway track defects

Fig. 23 There is multiple detection of railway track defects in several classes, where in the image there are various types of defects at the same detection time. As in Fig. 23(a), there are rail defects in one detection, namely crack, spalling and squat. This also applies to Figs. 23(b) and 23(c) which are capable of carrying out multiple detection of rail defects. Therefore, the system is capable of carrying out multiple detections at the same time properly and in accordance with the type of rail defects that exist.

According to the image above, it can be found very well that the recognition accuracy is very high even though it is used because of the limitations of graphic cards where the GPU computing power will affect the detection time. Although the time spent on image detection is very small, overall satisfactory results have been achieved.

V. CONCLUSION

Railway tracks are a very vital piece of infrastructure because they are directly related to passenger comfort and safety. Therefore, monitoring railway track abnormalities is very crucial in the improvement system in service institutions. This research uses YOLO-v3 as a surface defect detector on railway tracks along operational area 7 (DAOP 7) which passes between 2 cities, namely from Madiun Station to Magetan Station.

The detection process using YOLO includes; preprocessing of a number of data sets for resizing, image division, labeling and augmentation. Next is the training process which produces a model with a MAP of 99.41%. The model that has been trained is then tested against the data. The average precision range for the four classes is between 98% and 100% and there are also several True Positives (TP) and False Positives (FP) from each class of rail flaw detection. A high TP value indicates that the system really identifies objects in the dataset with the correct conditions. Based on the average loss chart, losses continue to decrease in a small range, reducing to 0.179502 when repeated it becomes 8000 times as the learning rate continues to increase.

According to the image above, it can be found very well that the recognition accuracy is very high even though it is used because of the limitations of graphic cards where the GPU computing power will affect the detection time. Although the time spent on image detection is very small, overall satisfactory results have been achieved. Overall, the results of the detection system using the YOLO model has mAP value of 99.41%, a precision value of 99%, a recall value of 99%, an f-score value of 99%, and an average IoU value of 85.84%. The YOLO model can detect railway track abnormalities accurately and optimally.

ACKNOWLEDGMENT

We would like to express our gratitude to Indonesia Railway Operator (PT. KAI) which has permitted the retrieval of the dataset along 14 km rail tract, from Madiun to Magetan Station. Those datasets were crucial part of resource in conducting this research.

REFERENCES

- [1]. PT KAI. (2022). Train User Volume Increases 42% in Semester I 2022 [Press Release – in Indonesian]. Indonesia Railway Company PT KAI. https://www.kai.id/information/full_news/5379-volume-pelanggan-kereta-api-naik-42-pada-semester-i-2022. Accessed October 15th, 2022.
- [2]. Luthfi, K. M., Sugiana, A., & Suratman, F. Y. (2021). Broken rail detection system using laser. IOP Conference Series: Materials Science and Engineering, 1098(4), 042045. <https://doi.org/10.1088/1757-899X/1098/4/042045>
- [3]. Komite Nasional Keselamatan Transportasi. (2021). Buku Statistik Investigasi Kecelakaan Transportasi KNKT 2021. Pelayanan Investigasi dan Kerjasama

- Bagian Data, Informasi, dan Humas Komite Nasional Keselamatan Transportasi.
- [4]. Indonesia Transportation Safety Committee (KNKT). Book of Transportation Accident Investigation Statistics in 2021. Report, July 2022.
- [5]. Lai, J., Xu, J., Wang, P., Yan, Z., Wang, S., Chen, R., & Sun, J. (2021). Numerical investigation of dynamic derailment behavior of railway vehicle when passing through a turnout. *Engineering Failure Analysis*, 121, 105132. <https://doi.org/10.1016/j.engfailanal.2020.105132>.
- [6]. Yuliang Zhao, Zhiqiang Liu, Dong Yi, Xiaodong Yu, Xiaopeng Sha, Lianjiang Li, Hui Sun, Zhikun Zhan, and Wen Jung Li. A Review on Rail Defect Detection Systems Based on Wireless Sensors. *Sensors* 2022, 22, 6409. <https://doi.org/10.3390/s22176409>
- [7]. Aydin, I., Akin, E., & Karakose, M. (2021). Defect classification based on deep features for railway tracks in sustainable transportation. *Applied Soft Computing*, 111(107706), 1–14. <https://doi.org/10.1016/j.asoc.2021.107706>
- [8]. Fahmi. New Measuring Train from Switzerland Arrives in Jakarta, Here's What It Looks Like [Internet-in Indonesian]. *Berita Trans.com*. 2022. Available from: 01 April 2023
- [9]. Rizaty MA. Development of Railway Track Length in Indonesia for 2016-2020 [Internet-in Indonesian]. 2021 [cited 2023 Apr 1]. Available from: <https://databoks.katadata.co.id/datapublish/2021/10/26/indonesia-miliki-rel-kereta-sepanjang-632-juta-meter-pada-2020>
- [10]. Wikipedia. List of countries by rail transport network size - Wikipedia [Internet]. [cited 2023 Apr 1]. Available from: https://en.wikipedia.org/wiki/List_of_countries_by_rail_transport_network_size#cite_note-40
- [11]. James, A., Jie, W., Xulei, Y., Chenghao, Y., Ngan, N. B., Yuxin, L., Yi, S., Chandrasekhar, V., & Zeng, Z. (2018). TrackNet—A Deep Learning Based Fault Detection for Railway Track Inspection. 2018 International Conference on Intelligent Rail Transportation (ICIRT), 1–5. <https://doi.org/10.1109/ICIRT.2018.8641608>
- [12]. Xu L, Zhai W, Chen Z. On use of characteristic wavelengths of track irregularities to predict track portions with deteriorated wheel/rail forces. *Mechanical System and Signal Processing*, 2018;104:264–78.
- [13]. Zhai W, Liu P, Lin J, Wang K. Experimental investigation on vibration behaviour of a CRH train at speed of 350 km/h. *Int. J. of Rail Transportation*. 2015;3(1):1–16.
- [14]. Ginanjar G, Choi IM, Kim SM. PCA exchange method for compensation of error sources in pressure balance calibration. *Measurement*, 2018;124:179–83.
- [15]. Cao H, Lei Y, He Z. Chatter identification in end milling process using wavelet packets and Hilbert-Huang transform. *Int. J. of Machine Tools and Manufacture*. 2013;69:11–9.
- [16]. Agus S., Yamada K, Tanaka R, Handoko YA, Subhan MF. Chatter identification in turning process based on

- vibration analysis using Hilbert- Huang transform. *J. of Mechanical Engineering & Science*. 2020;14(2):6856–68.
- [17]. Agus S., Liu C-H, Yamada K, Hwang Y-R, Tanaka R, Sekiya K. Milling Process Monitoring Based on Vibration Analysis Using Hilbert-Huang Transform. *Int. J. Automation Technology*. 2018;12(5):688–98.
- [18]. Agus S., Yamada K, Mani K, Tanaka R, Sekiya K. Vibration Analysis in Milling of Thin-Walled Workpieces Using Hilbert-Huang Transform. In: 9th International Conference on Leading Edge Manufacturing in 21st century: LEM21. 2017. p. 1–6.
- [19]. Agus S., Liu CH, Yamada K, Hwang YR, Tanaka R, Sekiya K. Application of Hilbert–Huang transform for vibration signal analysis in end-milling. *Precision Engineering*, 2018;53(April):263–77.
- [20]. Agus S., Artono B, Khonjun S, Mahmud R. Denoising of Disturbed Signal using Reconstruction Technique of EMD for Railway Bearing Condition Monitoring. *Int. J. Computation Science Issues*. 2020;17(5).
- [21]. Yongzhi Min, Benyu Xiao, Jianwu Dang, Biao Yue and Tiandong Cheng. Real time detection system for rail surface defects based on machine vision. . *EURASIP Journal on Image and Video Processing* (2018) 2018:3. DOI 10.1186/s13640-017-0241-yXXX
- [22]. Danyang Zheng, Liming Li, Shubin Zheng, Xiaodong Chai, Shuguang Zhao, Qianqian Tong, Ji Wang, and Lizheng Guo. A Defect Detection Method for Rail Surface and Fasteners Based on Deep Convolutional Neural Network. *Computational Intelligence and Neuroscience Volume 2021*, Article ID 2565500, 15 pages <https://doi.org/10.1155/2021/2565500>
- [23]. Guo, F., Qian, Y., Rizos, D., Suo, Z., & Chen, X. (2021). Automatic Rail Surface Defects Inspection Based on Mask R-CNN. *Transportation Research Record: Journal of the Transportation Research Board*, 2675(11), 655–668. <https://doi.org/10.1177/036119812111019034>
- [24]. Zhang, Jiawei, Xiang Liu, Xing Zhang, Zhenghao Xi, and Shuohong Wang. 2023. "Automatic Detection Method of Sewer Pipe Defects Using Deep Learning Techniques" *Applied Sciences* 13, no. 7: 4589. <https://doi.org/10.3390/app13074589>
- [25]. Unified Vision-Based Methodology for Simultaneous Concrete Defect Detection and Geolocalization. *Computer-Aided Civil and Infrastructure Engineering* 00 (2018) 1–18
- [26]. Quang Ngoc The Ho, Thanh Trung Do, Pham Son Minh, Van-Thuc Nguyen, and Van Thanh Tien Nguyen, . Turning Chatter Detection Using a Multi-Input Convolutional Neural Network via Image and Sound Signal. *Machines* 2023, 11, 644. <https://doi.org/10.3390/machines11060644>
- [27]. He, Y., Zeng, H., Fan, Y., Ji, S., & Wu, J. (2019). Application of Deep Learning in Integrated Pest Management: A Real-Time System for Detection and Diagnosis of Oilseed Rape Pests. *Mobile Information Systems*, 2019, 1–14. <https://doi.org/10.1155/2019/4570808>.
- [28]. Teng, Shuai, Zongchao Liu, and Xiaoda Li. 2022. "Improved YOLOv3-Based Bridge Surface Defect Detection by Combining High- and Low-Resolution Feature Images" *Buildings* 12, no. 8: 1225. <https://doi.org/10.3390/buildings12081225>
- [29]. Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., & Kurach, L. (2020). A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in construction*, 109, [102967]. <https://doi.org/10.1016/j.autcon.2019.102967>
- [30]. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition. <http://dx.doi.org/10.1109/CVPR.2016.91>
- [31]. Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement (arXiv:1804.02767). arXiv. <http://arxiv.org/abs/1804.02767>.