Metal Surface Defect Detection Using Deep Learning

1st Harshitha.P
Automotive Electronics
Vellore Institute of Technology
Vellore, India
harshirajtp@gmail.com

2nd Monisha.S
Automotive Electronics
Vellore Institute of Technology
Vellore, India
monishasaravanan2607@gmail.com

3rd Sahithi Chalimila
Automotive Electronics
Vellore Institute of Technology
Vellore, India
csahithi2000@gmail.com

Dr.Hemprasad Yahswant Patil
School of Electronics Engineering (SENSE)
Vellore Institute of Technology
Vellore, India
hemprasadpatil@gmail.com

***Abstract*—The metal surface of machinery plays a role in numerous industries, including automotive, in today bustling environment. A defective metal surface lowers the performance and quality of the product. The surface examination is done manually or using crude automated techniques, but the issue cannot be entirely resolved due to external variables. This study proposes a novel method for metal surface recognition utilizing deep learning approaches, namely Convolution Neural Network (CNN) combined with self-attention mechanisms. It aims to enhance the ability of the model to recognize minute details that are essential for precise detection. The self-attention modules enable the network to capture complex patterns and spatial relationships by choosing focusing on specific areas within metal surface dataset, enhancing sensitivity to difficult conditions.**

***Index Terms*—Metal Surface Detection, Deep Learning, CNN, Self-attention**

I. Introduction

Metal surfaces play a vital role in the manufacturing of machinery utilized in industry, agriculture, and transportation, including automobiles, commercial vehicles, airplanes, and rockets. The surface defects in metal lead to low quality and performance of the products [1].Numerous steps have been taken to check for defects and ensure the product's quality during the manufacturing stage. The inspection takes place using the traditional method, i.e., manual assessment or rudimentary automated systems. The drawbacks of traditional methods lie in their dependence on predefined rules and thresholds. Manual inspections are time-consuming, susceptible to human errors, and may not be efficient to handle the mass number of productions. Automated systems are hindered by limitations in speed, precision, and adaptability to various environmental conditions, like the complexities of varying surface conditions, lighting scenarios, and geometric intricacies. Deep learning has become a viable option for automated metal surface detection as a result of these obstacles being acknowledged. Deep learning techniques have been used in modern days to overcome defects in manual detection including low accuracy, poor real time performance etc [2]. They also address many limitations by enabling machines to autonomously learn and adapt to complex patterns and variations, ultimately overcoming the shortcomings of traditional approaches. Many industrial applications, like fabric surface defect detection, steel strip defect detection, aluminum strip defect detection, etc., utilise these deep learning systems. Figure 1 depicts a picture of a metallic surface flaw in an industrial setting. As illustrated in Figure 1, The existence of defects is very complex, and there are multiple types as shown in Figure 1. such as scars (1), holes (2), bubbles (3), inclusion (4), iron oxide scale (5), rolled printing (6), edge cracking (7), scratches (8), and scrapes (9). Other than these defects, some other defects are glue marks or spots, damage marks, or dust from the industrial

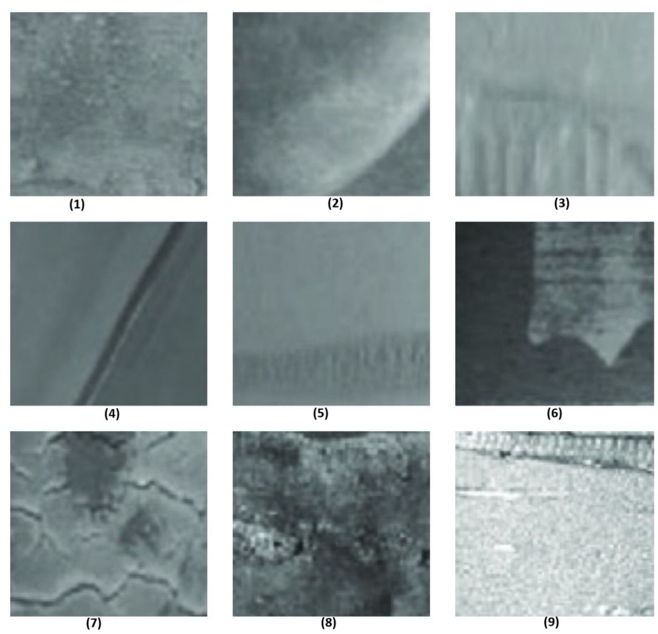

Fig. 1. An image of a defective metallic surface in industry

environment. Deep learning is a contemporary technology that can be used to detect defects in the surface of metal using the Convolution Neural Network (CNN) algorithm. CNN symbolize a significant breakthrough in deep learning that is especially suited for visual data-intensive problems like computer vision and image identification. Driven by the inadequacies of conventional neural networks in handling grid-structured data, such as images, CNNs have a unique design that radically changes the process of feature extraction [3]. Convolutional layers, which use filters to identify local patterns and hierarchies in the input, are at the center. They are excellent at hierarchical feature learning, enhanced by pooling layers for spatial down sampling and fully connected layers for advanced feature processing. This enables them to recognize complicated patterns on their own, ranging from simple textures and edges in lower layers to more complex objects and structures in deeper layers [4]. By utilizing transfer learning, CNNs that have been pretrained on extensive datasets can be optimized for certain tasks using a small amount of labeled data, demonstrating their versatility and effectiveness. As we explore further into their uses and developments, it becomes clear that CNNs have an immense effect on how machines understand and analyze visual data [5]. They also open up opportunities for innovation in a wide range of fields, reshaping the field of artificial intelligence today. In spite of its successes, research is still being conducted to solve issues such as interpretability and robustness, so that CNNs can keep paving the way in novel developments for visual perception tasks.

II. RELATED WORK

In manufacturing sectors, visual inspection of metal surfaces is a crucial step. This is done to reduce losses and build customer trust. This process is time-consuming, to overcome problem the automated systems are developed to detect defects in metals using different algorithm approaches [6]. Pranoto et al. proposed a model using VGG16-CNN along with SSD to detect sharp edges and burrs on metal surfaces. In this research a total of 360 images of an aluminium metal was used which was classified into three classes namely normal edges, burr edges, sharp edges. This model achieved in analysing defects in high resolution images with great accuracy in short span [7]. Evstafev et al. have worked on detecting surface flaws on folded metal sheets utilizing CNN. In this paper, they detected the liability in metal sheet under various lightening conditions with higher classification accuracy [8]. Xin et al. proposed a novel automatic workpiece surface defect detection system. The primary goal of this model is to take numerous pictures of uneven metal workpieces in various lighting circumstances and from different angles. To detect metal flaws, binary Bayesian hypothesis test is performed using a multi-angle image processing technique. According to the testing findings, recommended defect-detection methods perform noticeably more accurate [9]. Demir et al. presented a new deep learning-based strategy for surface defect detection and classification. For feature extraction, the Parallel Attention Residual-Convolutional Neural Network model was used. The recovered feature is selected using the NCA-Relief method. The dual and multi-class classification has been excueted by the machine technique [10]. Selamet et al. presented an innovative solution to address the environmental issues by using ShapeFromShading (SFS) and Faster R-CNN in the Automatic Detection and Classification of Defective Areas on Metal Parts. Environmental issue includes illumination conditions that affect the surface of metal. The dataset was compiled from two sources: the tagged NEU dataset and the unlabelled KSDD2 dataset [11]. Usamentiaga et al. has put out a methodology that addresses an issue that has been present for many years. Obtaining a fast and accurate output for a large and complicated dataset is the most challenging problem. For automated surface detection, they employed semantic segmentation and object detection. Defecting inclusion, patches, scratches, crazing, and rolling into which uses YOLOv5 and U-Net have all been investigated in this study [12]. Shah and Patil employed three object detection algorithms, namely YOLO V5, YOLO V7, and Detectron2, to automatically recognise and pinpoint the various kinds of metal faults, such as wrinkles and oil spots. NEU-DET and GC10-DET, two pre-existing datasets, were combined to produce their own dataset during this process. Three models mAP scores were compared to determine how well each could find and identify flaws [13]. Litvintseva et al. developed a method to identify and categorize defects on metal surfaces using images in real-time. To determine which model is most effective in identifying flaws on a flat steel surface, researchers compared three models: E-Net, DeepLabV3, and U-Net. A dataset including 12,568 photos divided into four distinct fault groups is used to train this model. According to the trial findings, the DeepV3 model and ResNet34 architecture were the most effective at precisely identifying and categorising the flaws [14]. Typically, hot-dip galvanising is applied to steel metals in order to prevent corrosion. Galvanised items can have faults as a result of technological challenges and a complex procedure. To detect defects on surface of zinc-coated steel material, Xiao et al. proposed a unique model called YOLOv5 Transformer-BiFPN. This model integrates components from the transformer structure, Bi-FPN, and YOLO-v5. According to the experimental results, the model effectively detects sprangles on steel surfaces, enhancing the production process' overall quality control [15].

III. FRAMEWORK

A. A Novel Deep-Learning Mechanism

Self-attention is a proposed mechanism in machine learning and computer vision for capturing dependencies.It will serve as a link between the neural network's first layer and the input sequence. The incorporation of self-attention into convolutional neural networks (CNN) presents an approach to enhance the model's ability to detect a wide range of dependencies, contextual understanding, and parallel computation information in image data. Generally, the self-attention mechanism will allow inputs from each position into the other position that describe a non-local relationship. By implementing CNN, their performance in computer vision tasks improves. In a CNN model, the convolutional layer concentrates on the local region for each input image through particular fields; each neuron will have information about a specific area. CNN is very good at extracting local features, but it fails to extract long-range features. By combining both self-attention and CNN, features in both the short and long ranges can be explored. Self-attention has a limited receptive field due to the high resolution of the input image. CNN is dependent on fixed-size kernels, but self-attention will focus on any element in the input feature map, which increases the receptive field without adding any computational complexity. Working with the self-attention layer in CNN involves a few steps. The first step is to reshape the input features, represented as at position i, to form a new 2D matrix (reshaped) representing the input features, given by

$$y \in \mathbb{R}^{C \times K} \tag{1}$$

Where C represents count of channels whereas K represents the product of other dimensions. Next action is to acquire a, b, and c by performing 1 x 1 convolutions on y. This will change the count of channels in C and is given as

$$a(y) = W_a y \tag{2}$$

$$b(y) = W_b y \tag{3}$$

$$c(y) = W_c y \tag{4}$$

Where a(y), b(y), and c(y) are transformed representations obtained through 1 x 1 convolution. The next step is to compute the SoftMax weights based on dot products between a (y) and b (y) to produce attention maps, which are given as

$$\gamma_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})} \tag{5}$$

The above equation represents attention weight between positions of i and j, which is used to quantify pixel j in the image relative to pixel i. Now, because these weights (gamma) are computed over the entire height and breadth of the feature set, the receptive field is not limited to the size of a small kernel. The output of the self-attention layer is given as

$$O_i = x\left(\sum_{i=1}^{N} \gamma_{j,i} c(y_i)\right) \tag{6}$$

$$x(y) = W_x y \tag{7}$$

$$W_x \in \mathbb{R}^{C^* \times C} \tag{8}$$

Where x is the output of another 1 x 1 convolution on the weighted sum. The number of channels in output features is the same as the count of channels in input features in self-attention. The final step is to enhance input features of y to the weighted output, which is given by

$$z_i = \alpha O_i + y_i \tag{9}$$

Here, α is another learnable parameter. CNN, combined with the self-attention mechanism, has a wide range of applications in computer vision tasks, including image classification and object detection. By incorporating self-attention with CNN, we demonstrate improved performance in situations where global context influences the interpretation of visual contents. Increased computational demands should be taken into consideration as they have quadratic complexity, which can impact the overall efficiency of the model.

B. *Convolution Neural Network*

This work explores the use of a novel hybrid architecture based on deep learning for the detection of flaws in metal parts. The dataset is first divided into five classes, and then a feature extractor called Convolutional Neural Network (CNN) is applied to it. One subclass of machine learning is the convolutional neural network (CNN, or convnet). This particular artificial neural network type is employed for diverse purposes and data kinds. CNN architecture consists of several layers that are arranged in a sequential manner. The fundamental building block in CNN architecture is the convolution layer. Convolution involves the application of filters or kernels to input data to extract low-level features like edges and textures. The features extracted are utilized in the classification process.

$$H_j = f\left(\sum_{i=1}^{N} K_i \cdot G_{i,j} + B_j\right) \tag{10}$$

Where represents the output matrix, represents the input matrix. Kernel is denoted by and denotes bias. The next stage is the pooling stage, which is used to reduce the size of feature maps to cut down on computational costs. The most common operation applied here is max pooling. Usually, a filter of size 2 x 2 with stride 2 is commonly used. After convolution and pooling operations are applied, Non-linear activation functions like sigmoid and hyperbolic tangential functions are used to introduce non-linearity into the network. This is done to build complex mappings between input and output. One or more fully connected layers are connected to the flattened one-dimensional (1D) array arising from the convolution and pooling layers. In fully connected layers, the neurons from two different layers are connected. In fully connected layers, the neurons from two different layers are connected. The CNN model is trained using the dataset through a process called the backpropagation algorithm, where the loss function and gradient descent operations play important roles. The discrepancy between the expected and actual outputs is measured using the loss function. Reducing the loss by adjusting the network's weights is the aim of model training. The gradient descent algorithm acts as an optimization algorithm that is used to update kernels and weights. The gradient of the loss function indicates the steepest uphill direction, showing how much the loss would increase if we moved in each parameter's direction. To reduce the loss, each learnable parameter (weight or kernel) is updated by moving in the negative direction of the gradient. It is mathematically represented as

$$\alpha = \alpha - \theta \frac{\partial L}{\partial \alpha} \tag{11}$$

Where α stands for learnable parameter, θ stands for learning rate, and L stands for loss function. The learning rate is the most important parameter, which is to be set before the training starts. The gradient of the loss function is computed by using a training dataset called mini-batch due to memory limitations.

IV. DATASET

A dataset is a collection of images. It is important to be trained and evaluated for image processing algorithms and models. Nick Lin has created a metal surface defects dataset. The dataset is a combination of the GC10-DET dataset and online metal-defected images. The dataset contains 649 images, which describe six metal surface defect types. These six metal defect types are crease, inclusion, oil spot, rolled pit, waist folding, and burr. We have created six folders each belonging to one defect type consists images of a particular defect. The label helps learn and predict accurately. The label provides accurate information that will lead us to the desired output. The dataset consists of labels for each image that will mark the defective portion.

- Crease- Occurs during uncoiling processes, looks like a vertical fold across metal strip.
- Inclusion- Region where the surface is loose and pressed into metal
- Oil Spot- Affects the appearance of the product, caused by mechanical lubricant
- Rolled Pit- Bulges on metal surface due to tension roll damage.
- Waist Folding- Folds due to low carbon.
- Burrs- Flaws in metal surface.

V. DATA AUGMENTATION

Data augmentation is a common method used in machine learning and computer vision for increasing the density of a dataset by applying different transformations to each original image. By this technique, single image will get multiple images of different variations [16]. We can enhance robustness, and performance by creating different images of same data. This is useful for model to learn steady features and decreases overfitting. Data Augmentation is done by using MATLAB R2023b software. Where each image in the dataset will be transformed. They go through a number of changes, including scaling, flipping, and rotation. These images are shifted horizontally, we can decide number of times an image can be iterated. For example, if we give 50 times the image will be iterated to produce 50 augmented images. Shifting the image horizontally by a specific number of times will be done by the Image Processing Toolbox in MATLAB. The resulting set will be the slightly shifted versions of the original images. This augmented dataset will be exposed to the model which will detect the defective parts in metal surfaces.

VI. ARCHITECTURE

Our model uses the Keras deep learning framework to implement a convolutional neural network (CNN) for the purpose of classifying metal surface defects into five categories. The primary goal of our model is to classify input images into one of these predefined classes [17]. The model architecture is constructed in such a way that a linear stack of layers is arranged in a step-by-step fashion. The architecture begins with a convolution layer comprising 48 filters, each of size 3x3. This layer utilizes the ReLU activation function to introduce non-linearity so as to capture complex patterns and features from the input images. Following that, a max pooling layer of pool size 2 x 2 is used to reduce the spatial dimensions of the extracted features. To prevent overfitting of the image during training, a dropout layer with a dropout rate of 0.45 is introduced. The architecture further includes another convolution layer with 80 filters each of size 3 x 3, followed by another max-pooling layer. This is done to capture high-level features and spatial hierarchies in the data [18]. The flattened output from these layers is sent to a dense layer comprising 120 neurons. Again, the ReLU activation function is applied and sent to the dropout layer with a dropout rate of 0.55 to promote better generalization of the data. A novel aspect of this model involves the utilization of the self-attention layer along with sigmoid activation to capture crucial features and improve overall performance [19]. A flattened layer and a dense layer with softmax activation are used to convert extracted features into probability distributions across different classes. This allows the neural network to assign a likelihood score for each class, indicating the model's confidence in its prediction. The model has been trained using 50 epochs. In each epoch, the model looks at a set of images taken for training and adjusts itself to get better at the task. The Adam optimization technique is used to train the model

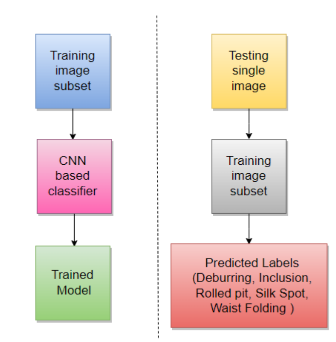

Fig. 2. Block Diagram of Training and Testing Phase

to make accurate predictions. This technique is frequently used for training classification models [20]. This optimization algorithm adjusts the model's parameters to minimize cross-entropy.The primary goal of the model is to accurately identify images and correctly assign labels, ultimately improving its ability to generalize and make accurate predictions on new, unseen data.

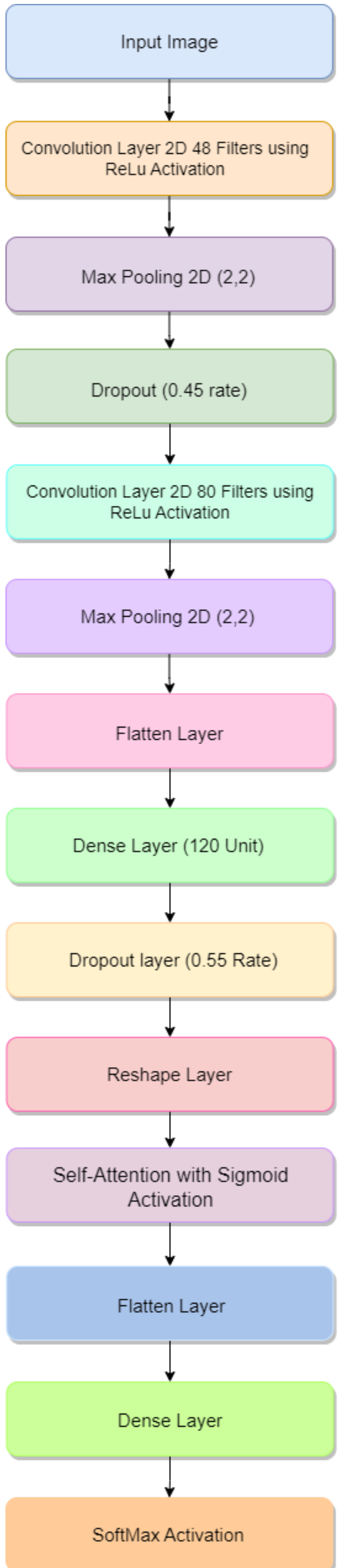

Fig. 3. Proposed CNN Architecture

VII. RESULTS

The dataset of 5 classes, which has 649 images each, was prepared by a data augmentation algorithm. All the datasets have been trained in layers of self-attention; they are Convolution2D, MaxPooling2D, Flatten, Dense, Dropout, and Reshape. The model has been trained with 50 epochs. A trained dataset was obtained with a validation rate of 95.24

MODEL	VALIDATION	TRAINING ACCURACY
PROPOSED CNN MODEL	95.24%	96.89%

Fig. 4. Result

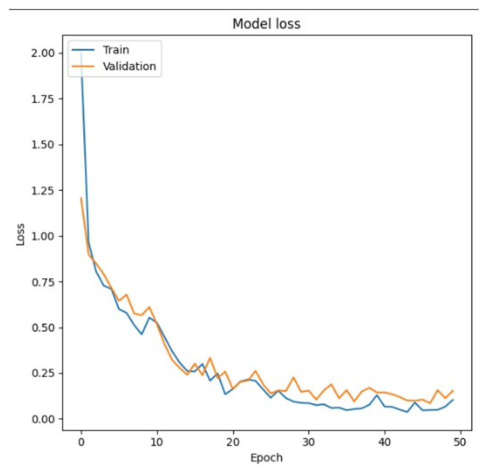

Fig. 5. Training and Validation Loss

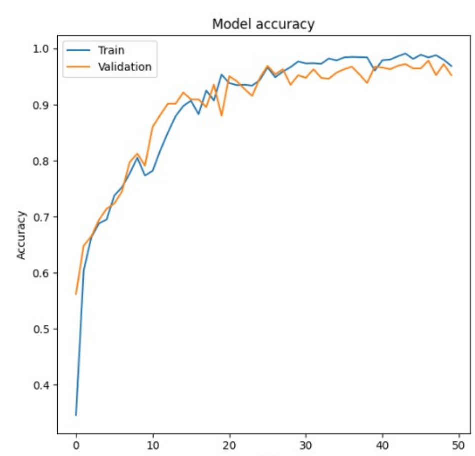

Fig. 6. Training and Validation Accuracy

VIII. CONCLUSION

The integration of CNN and self-attention mechanisms has proven to be a powerful and effective approach for metal surface detection. This model can be useful while screening a large number of datasets with metal surface defects. The algorithm proposed to identify defects like deblurring, inclusion, rolled-pit, silk spot, and waist folding. The deep learning technique is used to identify defects in many surfaces, like fabric, steel, aluminium, etc. This model helps manufacturers inspect the surface of metal machinery faster in complex environments. It marks a significant stride towards more advanced and reliable metal surface detection systems in the field of computer vision and industrial automation.

REFERENCES

[1] Zhao Zhang, Weibo Wang, and Xiaoyan Tian. Semantic segmentation of metal surface defects and corresponding strategies. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023.

[2] Mingshan Liu, Yifei Huang, and Yuanli Xie. Design of steel surface defect detection system based on deep learning. In *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE, June 2022.

[3] Hsien-I. Lin and Fauzy Satrio Wibowo. Image data assessment approach for deep learning-based metal surface defect-detection systems. *IEEE Access*, 9:47621–47638, 2021.

[4] Shaotao Guo and Zequan Xing. Automatic visual pit detection system for bottom surface of cylindrical lithium battery. *IEEE Access*, 11:56538–56549, 2023.

[5] Xuejin Hu, Jing Yang, Fengling Jiang, Amir Hussain, Kia Dashtipour, and Mandar Gogate. Steel surface defect detection based on self-supervised contrastive representation learning with matching metric. *Applied Soft Computing*, 145:110578, September 2023.

[6] Gadamsetty Pranav, T Sonam, and T Sree Sharmila. Defect detection with less training samples using deep neural networks. In *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*. IEEE, April 2023.

[7] Kirana Astari Pranoto, Wahyu Caesarendra, Iskandar Petra, Grzegorz Królczyk, Mochamad Denny Surindra, and Putri Wulandari Yoyo. Burrs and sharp edge detection of metal workpiece using cnn image classification method for intelligent manufacturing application. In *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*. IEEE, July 2023.

[8] Oleg Evstafev, Sergei Shavetov, and Anton Pyrkin. Detecting and classifying surface defects in rolled steel sheets using deep learning techniques. In *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, July 2023.

[9] Wenkai Xin, Wei Liu, and Kun Yan. Novel automatic workpiece surface defects detection system. In *2022 22nd International Conference on Control, Automation and Systems (ICCAS)*. IEEE, November 2022.

[10] Kursat Demir, Mustafa Ay, Mehmet Cavas, and Fatih Demir. Automated steel surface defect detection and classification using a new deep learning-based approach. *Neural Computing and Applications*, 35(11):8389–8406, December 2022.

[11] Feyza Selamet, Serap Cakar, and Muhammed Kotan. Automatic detection and classification of defective areas on metal parts by using adaptive fusion of faster r-cnn and shape from shading. *IEEE Access*, 10:126030–126038, 2022.

[12] Ruben Usamentiaga, Dario G. Lema, Oscar D. Pedrayes, and Daniel F. Garcia. Automated surface defect detection in metals: A comparative review of object detection and semantic segmentation using deep learning. *IEEE Transactions on Industry Applications*, 58(3):4203–4213, May 2022.

[13] Harshil Shah and Vaishnavi Patil. Metal surface defect detection using object detection models. In *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. IEEE, January 2023.

[14] Anna Litvintseva, Oleg Evstafev, and Sergey Shavetov. Real-time steel surface defect recognition based on cnn. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, August 2021.

[15] Dong Xiao, Fei Tong Xie, Yang Gao, Zhen Ni Li, and Hong Fei Xie. A detection method of spangle defects on zinc-coated steel surfaces based on improved yolo-v5. *The International Journal of Advanced Manufacturing Technology*, 128(1–2):937–951, July 2023.

[16] Jong Pil Yun, Woosang Crino Shin, Gyogwon Koo, Min Su Kim, Chungki Lee, and Sang Jun Lee. Automated defect inspection system for metal surfaces based on deep learning and data augmentation. *Journal of Manufacturing Systems*, 55:317–324, April 2020.

[17] Yasir Aslam, N. Santhi, N. Ramasamy, and K. Ramar. Retracted article: Localization and segmentation of metal cracks using deep learning. *Journal of Ambient Intelligence and Humanized Computing*, 12(3):4205–4213, February 2020.

[18] Ioannis D. Apostolopoulos and Mpesiana A. Tzani. Industrial object and defect recognition utilizing multilevel feature extraction from industrial scenes with deep learning approach. *Journal of Ambient Intelligence and Humanized Computing*, 14(8):10263–10276, January 2022.

[19] Maisha Maliha and Vishal Pramanik. Vehicle detection and identification system: Convolutional neural network with self-attention. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, July 2023.

[20] Xianxin Shao, XiaoJun Xia, and Jia Song. Design and research of metal surface defect detection based on machine vision. In *2022 11th International Conference of Information and Communication Technology (ICTech))*. IEEE, February 2022.