Journal of Scientific & Industrial Research
Vol. 82, February 2023, pp. 178-184
DOI: 10.56042/jsir.v82i2.69945

Edge Intelligence with Light Weight CNN Model for Surface Defect Detection in Manufacturing Industry

Shobha Rani D[1]*, Lakshmi Ramani Burra[2], Kalyani G[3] & Narendra Kumar Rao B[4]

[1]Institute of Aeronautical Engineering College, Hyderabad 500 043, Telangana, India

[2]Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada 520 007, Andhra Pradesh, India

[3]Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada 520 007, Andhra Pradesh, India

[4]Sree Vidyanikethan Engineering College, Tirupati 517 102, Andhra Pradesh, India

Received 07 July 2022; revised 22 September 2022; accepted 07 October 2022

Surface defect identification is essential for maintaining and improving the quality of industrial products. However, numerous environmental factors, including reflection, radiance, light, and material, affect the defect detection process, considerably increasing the difficulty of detecting surface defects. Deep Learning, a part of Artificial intelligence, can detect surface defects in the industrial sector. However, conventional deep learning techniques are heavy in terms of expensive GPU requirements to support massive computations during the defect detection process.CondenseNetV2, a Lightweight CNN-based model, which performs well on microscopic defect inspection, and can be operated on low-frequency edge devices, was proposed in this research. It provides sufficient feature extractions with little computational overhead by reusing a set of the existing Sparse Feature Reactivation module. The training data are subjected to data augmentation techniques, and the hyper-parameters of the proposed model are fine-tuned with transfer learning. The model was tested extensively with two real datasets while running on an edge device (NVIDIA Jetson Xavier Nx SOM). The experiment results confirm that the projected model can efficiently detect the faults in the real-world environment while reliably and robustly diagnosing them.

Keywords: CondenseNet, Convolutional neural networks, Deep learning, Edge device, Industrial products

Introduction

Industry 4.0 automated the manufacturing processes by providing customizable and adaptive mass production technology to the industries.[1] Industry 4.0 includes several technologies, including cloud computing[2], big data and analytics[3], Internet of Things[4], artificial intelligence & machine learning[5,6], and automation[7] (such as intelligent robots in product assembly lines). The quality of manufactured goods is very readily impacted during the industrial production process because of the short comings and constraints of working conditions, current technology, and other factors. Traditional industrial techniques are being replaced by new strategies based on artificial intelligence to make decisions on their own, work independently, and also for continuous learning.[8] Delivering products with high quality are most important in manufacturing industries. The most prominent issue which affects the rate is surface defects in the product. So, product surface defect detection[9] is required to guarantee the qualification ratio and reliable quality.

Surface defect detection denotes the discovery of faults, color contamination, external body, holes, scratches, etc., on the sample surface to be verified to gather a variety of essential data, such as the contour, category, size, and surface defects location to be tested.[10] Surface defect inspection is often done manually, which is very subjective, time-consuming, and unable to satisfy the demands of real-time detection. Automated surface detection approaches can be used to supplement or take the place of human judgment to overcome the limitations of manual examination. Modern technologies are used in surface defect identification research, including ultrasonic inspection, machine learning, and deep learning.[11]

Deep learning has recently shown exceptional performance in various image-based applications, such as defect/medical diagnosis[12], facial detection, object detection[13] and classification[14], pattern recognition and many more. As a result, the detection of surface defects is made possible using deep learning technology, but their application in particular

*Author for Correspondence
E-mail: d.shobharani@iare.ac.in

industries is only in the progression stage. Minor surface defects throughout the manufacturing process could result in many recalled products, seriously affecting the industry's reputation and company sales. An automatic surface defect detection system can be employed to meet the demands of the manufacturing sector, such as low latency response, storage, and high accuracy. Since edge computing offers several advantages, it is helpful for manufacturing industries. It is a sort of decentralized processing that allows data to be handled directly by the device that generated it or a local server.

The proposed work suggests a casting fault detection technique based on edge computing to satisfy industry demands. Practical approaches, like data augmentation, are adopted to make the model more adaptable and fine-tune. CondenseNetV2's hyper-parameters are fine-tuned with transfer learning. Finally, the trained model was deployed to the NVIDIA Jetson Xavier Nx SOM as an edge device to speed up detection.

Literature Review

Numerous recent research initiatives have been taken to suggest intelligent machine vision algorithms for evaluating defective goods. These systems would use a variety of deep learning algorithms to benefit from the data produced by various integrated technologies into contemporary industrial processes. A list of some notable achievements in this area is provided below:

An innovative and lightweight detection technique based on the attention mechanism is proposed by Zhuxi *et al.*[15], focusing on using aluminum strip defect inspection in industries. The YOLO- DCSAM backbone network is built using the YOLOv4 architecture and is designed with a parallel dual-channel attention module and depth-wise separable convolution. It reduces the network scale and improves how the various channels affect the feature map. The neck network is also rebuilt and made lighter for feature blending, which broadens the network's receptive field and streamlines using the SPPM-PANet segment. Additionally, improvements to the cluster's loss function and anchor box size boost the model's applicability to defective items.

Wan *et al.*[16] offers an upgraded YOLOv5-based detection method. First, the network layer in the backbone network is extended, and the attention mechanism is added. Then, by including a small-scale detection layer, the model is amplified from a three-output forecast layer to a four-output estimation layer. Thirdly, network feature fusion has been enhanced by the neck network. Lastly, depth-wise separable convolution is used in place of the original convolution to create a light weight ceramic tile recognition system.

Surface imperfections that appear overtime or as a result of incidents like collisions with flying debris can significantly reduce the strength of glass. The count and dimension of scratches are to be counted to determine whether the glass components will shatter throughout their lifespan. Zhufeng, *et al.*[17] developed a pixel-level instance segmentation approach utilizing a region-based convolution neural network with a mask to recognize scratches on transparent glass surfaces.

The machine vision inspection system powered by artificial intelligence is best suited for examining the degree of defects in bottle quality. Sahoo *et al.*[18] proposed a dynamic bottle inspection framework using segmentation techniques. The bottle image was separated from the background of the picture. The predicted adaptive characteristics were then used to create a dataset using mathematical methods such as the wavelet transform, Principal Component Analysis (PCA), and the average gray scale two-dimensional feature vector. An artificial neural network trained by back propagation, differential evaluation, and support vector machine algorithms is employed to categorize the images.

Ren *et al.*[19] projected a general method with a small training dataset. This approach is associated with building a classifier employing the characteristics of image patches extracted from a trained deep learning network. The trained classifier is then convolved to create pixel-wise prediction across the input. The experiment uses three open datasets and one industry dataset. The investigation entails two tasks: Fault segmentation, and Image classification.

Liu *et al.*[20] presented a framework for detecting the defects on the surface of the steel strip by incorporating a self-attention mechanism to mitigate the detrimental effects of information redundancy and improve feature discrimination. The authors developed a model that can recognize the same class and the position of six types of typical surface flaws on steel strips when performing the defect detection task. The self-attention mechanisms effectively capture spatially semantic linkages among the given two positions of the feature maps, and global contextual interdependencies. The suggested framework has more excellent localization capabilities and improved detection accuracy compared to Faster R-CNN with FPN.

Jiang *et al*.[21] present a coaxial bright-field and a low-angle bright-field imaging system. Images with a resolution of 16,000 × 8092 are acquired employing 8K line-scan complementary metal oxide semiconductor cameras. When there are weak scratches and coloration issues, the coaxial bright-field method is used; however, when there is a dent issue, the Flow-angle bright-field imaging system is used. Based on U-net, a proposed symmetric convolution neural network with encoder and decoder architecture generates semantic segmentation similar to the source image. In the glass surface defect collection, more than 30,000 faulty and non-defective photos were manually annotated from more than 10,000 source photographs. The authors claimed that the results confirmed the experimental findings, which revealed that the average recall rate and precision exceeded 95%.

Jingwen *et al*.[22] presented a CNN-based network approach with two parts. The designed network consists of two parts feature extractor and classifier. The authors use a portion of the pre-trained network VGG16 as a feature extractor, and the authors defined CNN as a classifier to identify the defects on the surface of the steel strips. The classifier works efficiently based on the feature maps formed by the feature extractor of the designed network. An optical inspection platform that combines parallel image processing with a high-resolution optomechanical module has been proposed to analyze touch panel glass for faults. Parallel image processing on a combination of CPU and GPU platforms was used to examine the key characteristics of the glass surface. After examination, a high rate of surface scratch detection on the touch panel glass was achieved with a minimal defect size.

Proposed Methodology

Intelligent machine vision systems help to assure excellent quality before a product is delivered to the user by enabling early error detection in the production line. However, while creating exemplary machine vision architecture for defective product inspection, it is essential to consider the various technologies used in the production process, starting with the raw material to the finished product. Leveraging all the data produced throughout the production process is also essential to improve the system and find the primary reasons for failures to enhance product quality. The architecture of the proposed methodology with edge devices is shown in Fig. 1.

The architecture starts with surface images of the products as the input. Data augmentation allows the process to detect defects in the products irrespective of the orientation of the product. As part of data processing, augmentation is applied to the dataset to increase the size and the images from different angles. After the augmentation, the dataset is separated into training and testing data. Training data is given to lightweight CondenseNet-V2 to retrain the transfer learning network. The trained model is subjected to the evaluation, in which the model is validated and tested with test data. The validated model is then deployed into edge computing devices that detect the defect on the product surfaces. The recognized label is the final output of the proposed architecture.

CondenseNet

Innovative and effective convolutional network architectures called CondenseNet improve the parameter and Floating-Point Computation

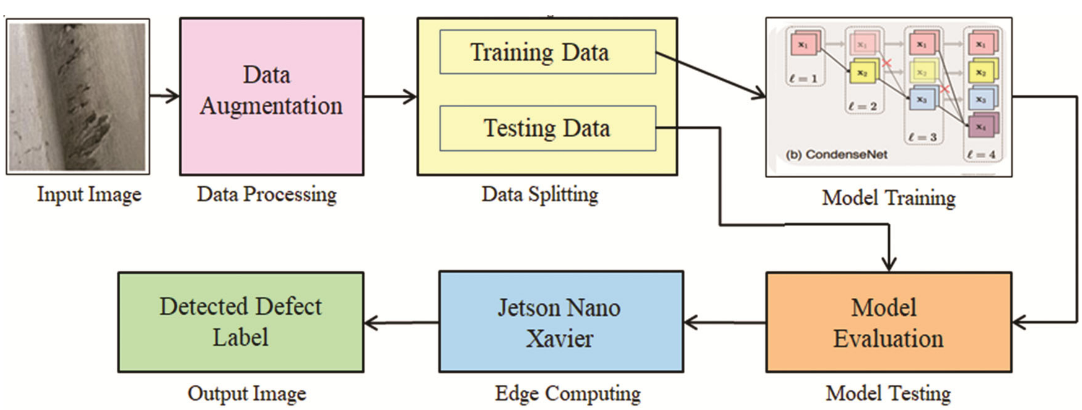

Fig. 1 — A flow process model of proposed methodology with edge device

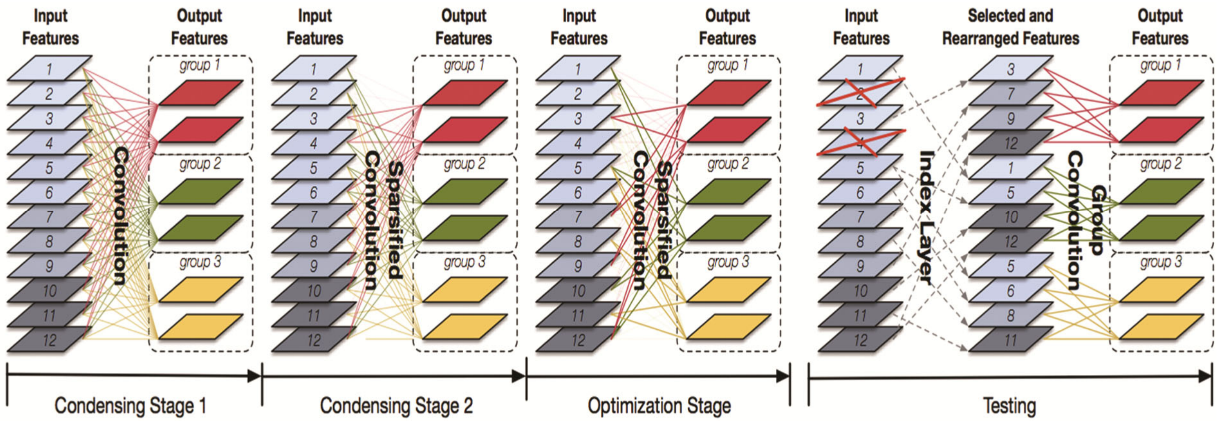

Fig. 2 — Learned Group Convolution

Performance of DenseNet (FLOPs). CondenseNet was proposed as a DenseNet based method for automatically learning a good connectivity pattern.[23] The authors develop a brand-new essential building block known as Learned Group Convolution by segmenting the filters of the bottleneck layer into multiple groups and gradually erasing less crucial features throughout the training, as shown in Fig. 2. CondenseNet's main distinguishing feature is its ability to transform the final pre-trained model in to typical group convolutions, accelerating deployment.

Deep networks with dense connectivity can achieve high computational efficiency by reusing features depicted in Fig 3. A method called sparse feature activation is utilized to actively increase the relevance of features for reuse. The technique can be further enhanced if redundant characteristics are eliminated, as demonstrated by the CondenseNet. CondenseNetV2 is the network here each layer is capable of learning to 1) selectively reuse a subset of the essential features from earlier levels and 2) actively update a subset of those features to make them more relevant for forthcoming layers.

Transfer Learning

Transfer learning is a technique for performing image classification tasks using a trained model's feature representations. Transfer learning is typically used when the data set is too large to start from scratch and to train a complete model. Large datasets are generally used to train pre-trained models, which is a common practice in the arena of computer vision. Different computer vision applications can use the weights that the models determine. These models can be used to train new models or to apply straight to new task predictions. When pre-trained models are merged into new models, the training time and generalization error are reduced. A transfer learning phase with fine-tuning parameters increases the model's functionality. While validating the model to provide the desired outputs during fine-tuning, the parameters of a trained model are adjusted and tailored to fit precisely. When retraining the model or a subset, use a low learning rate and stay away from over-fitting.

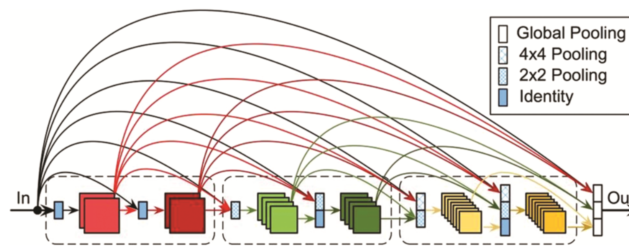

Fig. 3 — CondenseNet with Fully Dense Connectivity

Results and Discussion

Dataset

Identifying surface defects on industrial products currently needs a significant and coherent dataset. The foundation for the research is the dataset. A good dataset make site as easier to find problems and summarize them to create solutions. This section classifies the standard industrial datasets based on the different objects and application scenarios. Two datasets are considered to assess the effectiveness of the planned methodology.

The surface of the Ball Screw Drives Dataset[24] contains 21835 images in the .png format. The classification for each image is "P, N," where P intends for pitting (also known as pitting(s)), which are surface failures, and N intends for no pitting (No Surface failures). The collection comprises 11 075 images without surface failure and 10 760 images with surface failure. A 20% of the dataset is used for validation & testing, and the residual 80% for training.

The NEU surface defect dataset[25] contains 1800 images of the defects on the surface of hot-rolled steel strips. It has a total of six kinds of surface defects, Crazing (Cr), Patches (Ps), Rolled_in_Scale (Rs), Pitted surface (Ps), Inclusion (In), and Scratches (Sc). The description of the class labels is as follows: Inclusion (One common type of metal surface imperfection is inclusion. Some inclusions are pushed into the plate while others are loose and easily slide off), Crazing (The phenomenon that causes inevitable surface cracks in a material is known as crazing), Patches (A distinctive feature that distinguishes a piece of metal from the rest), Pitted surface (Pitting is a type of rust that concentrates on a relatively small area of metal surfaces and seeps into the inside of the metal. Pitting often has a minor diameter with a deep depth), Scratches (An abrasion mark on a surface is called a scratch), and Rolled_in_scale (mill scale is rolled into the metal during the rolling procedure). Every individual class has 300 images; 240 are utilized for training and 60 for testing.

The characteristics of the two datasets in terms of training and testing instanced for class label-wise are shown in Table 1. Sample images without and with defects in the surface of the Ball Screw Drives Dataset are shown in Fig. 4 and Fig. 5 respectively. The model images of each class in the NEU dataset are shown in Fig. 6.

Table 1 — Characteristics of the Surface defect datasets

Dataset	Class label	Instances in Training	Instances in Testing	Total
Surface of the Ball Screw Drive	Surface failure (P)	8608	2152	10760
	No Surface failure (N)	8860	2215	11075
	Total	17468	2215	21835
NEU Surface Defect dataset	Crazing (Cr)	240	60	300
	Patches (Ps)	240	60	300
	Rolled_in_Scale (Rs)	240	60	300
	Pitted_surface (Ps)	240	60	300
	Inclusion (In)	240	60	300
	Scratches (Sc)	240	60	300
	Total	1440	360	1800

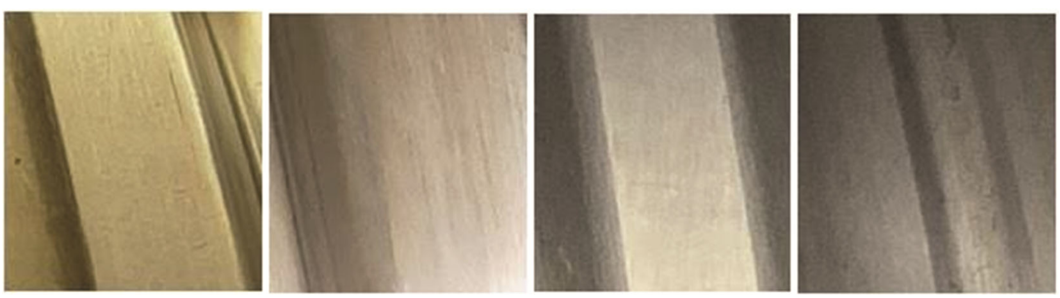

Fig. 4 — Sample of images without Surface failure

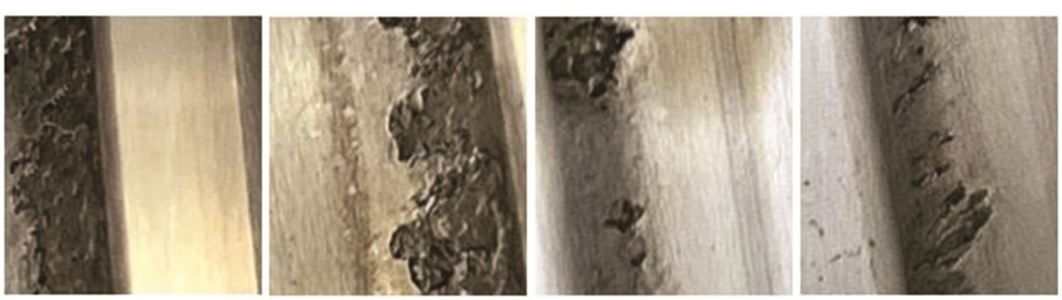

Fig. 5 — Sample of images with Surface failure

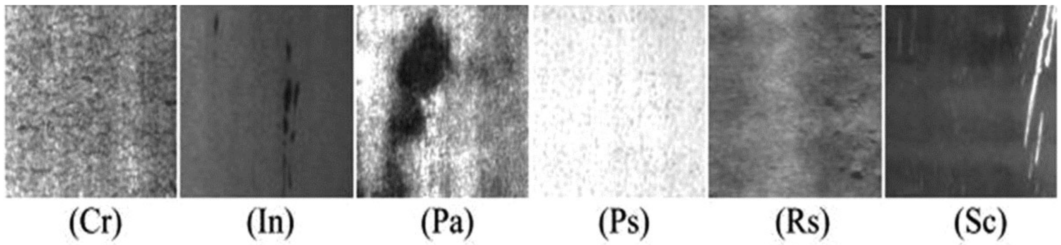

Fig. 6 — NEU dataset-six classes of surface defects sample images

Experimental Environment

We conduct a series of experiments using two real datasets. This experiment is performed on the DELL Power Edge Server, and the system hardware configuration is shown in Table 2. The computer software configuration includes the Ubuntu operating system and the PyTorch deep learning framework.

Performance Evaluation

Standard performance measures, including accuracy, recall, precision, and F1-score, are considered to evaluate the developed models. The equations for precision, recall, and F1-score and accuracy are specified in Eqs (1) through (4), respectively. Precision estimates the percentage of detections that contain defects, whereas recall indicates the ratio of defects that are identified. The harmonic mean of recall and precision is the F1-score, which illustrates the trade-off between the two-component metrics. Images with defects belong to the positive class, whereas those without defects belong to the negative class. The formulas of evaluated measures are shown in Eqs (1–4).

$$Pecision = \frac{TP}{FP+TP} \quad \dots (1)$$

$$Recall = \frac{TP}{FN+TP} \quad \dots (2)$$

$$F1_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \dots (3)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad \dots (4)$$

The notations TN, TP, FN, and FP, stand for the number of true positives, false positives, and false negatives, respectively.

We examined the pre-trained model and used the fine-tuning technique on the surface of the Ball Screw Drives Dataset. When using n pre-train weights as initial network weights for the dataset, the full CondenseNetV2 model can be trained and achieve an accuracy of 98.08%, and the classification results are shown in Table 3. However, for the NEU Surface defect dataset, the proposed model retrains some layers of the model while leaving others frozen in training. The experiment exhibits 93.33% accuracy, with the trainable of the last four levels and the first eight layers being frozen. The classification outcomes of the NEU Surface defect dataset are shown for each type of defect in Table 3. The results of the projected method are compared with the standard CNN-based model. The standard CNN model achieved 96.46% accuracy; while the proposed method performed 98.08% on the Ball screw drives dataset. The standard CNN model gained 90.59% accuracy with the NEU dataset, and the proposed model got 93.33% accuracy for surface defects detection.

Table 2 — System hardware configuration

Hardware	Product Specification
CPU	2 × Intel Xeon Gold6226R-2.9G
GPU	NVIDIA QuadroRTX6000, 24GB GDDR6
RAM	2 × 64GB RDIMM
Edge Device	NVIDIA Jetson Xavier Nx SOM

Table 3 — Classification results of ball screw drives and NEU surface defect dataset

Dataset	Class Label	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Ball Screw Drives Dataset	Surface failure (P)	98	99	98	98.08
	No Surface failure (N)	99	98	98	
NEU Surface Defect dataset	Crazing (Cr)	93	93	93	93.33
	Inclusion (In)	95	95	95	
	Patches (Ps)	97	97	97	
	Pitted_surface (Ps)	92	92	92	
	Rolled in scale (Rs)	93	93	93	
	Scratches (Sc)	90	90	90	

Conclusions

Surface defect detection is essential for intelligent production. Nowadays, most industrial product surface defects are manually inspected, which is time-consuming, expensive due to high labor expenses, and error-prone. Therefore, research on industrial product surface defect identification has significant practical implications. The proposed method has an autonomous visual inspection mechanism, lower power consumption, lower latency, more efficiency, and easy upgradeability. The usage of image processing methods as data augmentation boosted the surface fault dataset's adaptability. A lightweight Condense NetV2-based transfer learning model for defect identification deployed on the NVIDIA Jetson Xavier Nx SOM edge device to accelerate the detection process. The experiment proved that the proposed methodology achieved good results compared to the standard Convolutional Neural Network. The trained condenseNetV2 model improves surface defect detection efficiency in industries and is quick and accurate at defect detection and in future it can be extended to detect the surface defects of the products manufactured with different materials not only metal.

References

1 Kergroach S, Industry 4.0: New challenges and opportunities for the labourmarket, *Форсайт*, **11(4)** (2017) 6–8.

2 Rana A K & Sharma S, Industry 4.0 manufacturing based on IoT, cloud computing, and big data: Manufacturing purpose scenario, *Adv Commun Comput Technol*, **668** (2021) 1109–1119.

3 Rehman M H U, Yaqoob I, Salah K, Imran M, Jayaraman P P & Perera C, The role of big data analytics in industrial Internet of Things, *Future Gener Comput Syst*, **99** (2019) 247–259.

4 Nord J H, Koohang A & Paliszkiewicz J, The Internet of Things: Review and theoretical framework, *Expert Syst Appl*, **133** (2019) 97–108.

5 Lee, J, Davari H, Singh J & Pandhare V, Industrial Artificial Intelligence for industry 4.0-based manufacturing systems, *Manuf Lett*, **18** (2018) 20–23.

6 Janakiramaiah B, Kalyani G, Prasad L V, Karuna A & Krishna M, Intelligent system for leaf disease detection using capsule networks for horticulture, *J Intell Fuzzy Syst*, **41(6)** (2021) 6697–6713.

7 Hashemi-Petroodi S E, Thevenin S, Kovalev S and Dolgui A, Operations management issues in design and control of hybrid human-robot collaborative manufacturing systems: a survey, *Annu Rev Control*, **49** (2020) 264–276.

8 Singh H, Big data, industry 4.0 and cyber-physical systems integration: A smart industry context, *Mater Today Proc*, **46** (2021) 157–162.

9 De Vitis G A, Foglia P & Prete C A, Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase, *IET Image Process*, **14(12)** (2020) 2911–2921.

10 Jain S, Seth G, Paruthi A, Soni U & Kumar G, Synthetic data augmentation for surface defect detection and classification using deep learning, *J Intell Manuf*, (2020) 1–14.

11 Zhuxi M A, Li Y, Huang M, Huang Q, Cheng J & Tang S, A light weight detector based on attention mechanism for aluminum strip surface defect detection, *Comput Ind*, **136** (2022) 103585.

12 Janakiramaiah B, Kalyani G, Karuna A, Prasad L V & Krishna M, Military object detection in defense using multi-level capsule networks, *Soft Comput*, **27** (2023) 1045–1059.

13 Kalyani G, Janakiramaiah B, Karuna A & Prasad L V, Diabetic retinopathy detection and classification using capsule networks, *Complex Intell Syst*, (2021) 1–14, https://doi.org/10.1007/s40747-021-00318-9.

14 Subramanian M, Lv N P & Ve S, Hyperparameter optimization for transfer learning of VGG16 for disease identification in corn leaves using Bayesian optimization, *Big Data*, **10(3)** (2022) 215–229.

15 Wan G, Fang H, Wang D, Yan J & Xie B, Ceramic tile surface defect detection based on deep learning, *Ceram Int*, **48 (8)** (2022) 11085–11093.

16 Pan Z, Yang J, Wang X.E, Wang F, Azim I & Wang C, Image-based surface scratch detection on architectural glass panels using deep learning approach, *Construction and Building Materials*, **282** (2021) 122717.

17 Sahoo S K, Mahesh Sharma M & Choudhury B B, A dynamic bottle inspection structure, in *Computational Intelligence in Data Mining, Advances in Intelligent Systems and Computing,* vol 711, edited by H Behera, J Nayak, B Naik, A Abraham (Springer, Singapore) 2019, 873–884, https://doi.org/10.1007/978-981-10-8055-5_77.

18 Ren R, Hung T & Tan K C A generic deep-learning-based approach for automated surface inspection, *IEEE Trans Cybernet*, **48(3)** (2017) 929–940.

19 Ye R, Chang M, Pan C S, Chiang C A & Gabayno J L, High-resolution optical inspection system for fast detection and classification of surface defects, *Int J Optomechatronics*, **12(1)** (2018) 1–10.

20 Liu R, Huang M, and Cao P, An end-to-end steel strip surface defects detection framework: considering complex background interference, In *33rd Chinese Control and Decis Conf (CCDC) China*, 2021, 317–322.

21 Jiang J, Cao P, Lu Z, Lou W & Yang Y, Surface defect detection for mobile phone back glass based on symmetric convolutional neural network deep learning, *Appl Sci*, **10(10)** (2020) 3621.

22 Fu J, Zhu X & Li Y, *Recognition of Surface Defects on Steel Sheet using Transfer Learning* (Xian Jiaotong University) 2020, 25–34.

23 Huang G, Liu S, van der Maaten L & Weinberger K Q, Condensenet: An efficient densenet using learned group convolutions, In *Proc IEEE Conf Comput Vis Pattern Recognit* (Salt Lake City, UT, USA) 2018, 2752–2761.

24 Schlagenhauf T & Landwehr M, Industrial machine tool component surface defect dataset, *Data in Brief*, **39** (2021) 107643.

25 Schlagenhauf T, *Ball Screw Drive Surface Defect Dataset for Classification* (Karlsruher Institut für Technologie) 2021.