

Glove Defect Detection Via YOLO V5

How Yong Chen¹, Ahmad Fakhri Ab. Nasir^{2,3}, Khairul Fikri Muhammad¹, Anwar P.P. Abdul Majeed^{1,3}, Mohd Azraai Mohd Razman¹, and Muhammad Aizzat Zakaria¹

¹Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

²Faculty of Computing, Universiti Malaysia Pahang, 26600 Pekan Pahang, Malaysia.

³Centre for Software Development & Integrated Computing, Universiti Malaysia Pahang, 26600 Malaysia.

ABSTRACT – Malaysia is one of the biggest producers and exporters of gloves in the world. To meet and exceed the customer's expectation, a predictive defect model is necessary to minimize the defect glove. There are three crucial parts to develop an effective defect glove detection model, which are data collection, model development and model evaluation. The data provided should be good quality, the algorithm for developing the model should reach high accuracy and high inference time due to the fast glove production line, and the developed model must compare to the other quality model to prove its robustness and effectiveness. This paper focuses on employing the YOLO V5 model for glove defect detection as well as investigating the efficiency of other several deep learning approaches. The dataset collected in this research was 493 images with three classes which are normal glove, tear glove and unstripped glove. To avoid overfitting due to the small amount of dataset, argumentation processes such as saturation, exposure and noise were applied to increase the dataset number to 1148 images. Data were then split to 70:20:10 for the training-validation-test ratio. The parameter setup was 100 epochs with 129 iterations. The YOLO V5 was compared with Scaled YOLO V4, Detectron2 and EfficientDet by the training time, model size, accuracy, and inference time. In conclusion, the best model was YOLO V5 because it reached the lowest training (0.259 hour) and inference time (0.0095 seconds), smallest model size (14418kb) and highest accuracy (mAP = 0.9951).

ARTICLE HISTORY

Received: 4th April 2021

Revised: 23rd June 2021

Accepted: 30th July 2021

KEYWORDS

Deep learning

YOLO V5

Detection

Glove defect

INTRODUCTION

Since January 2020 with the COVID-19 pandemic, rubber glove producers in Malaysia have been rapidly expanding their production capacity, from the medical glove to the normal rubber glove to meet the increasing global demand. In the current market, the human visual-based approach for defect detection in the production line is the standard method to assess a product's quality. But the traditional human visual inspection is an unmeasured process with the variable and subjective result. To solve this issue, computer vision with a deep learning approach is necessary. Computer vision by deep learning-based neural networks is a complex system that can memorise and learn from previously trained data to solve the problems of large datasets. Computer vision is famous in this decade, and it is implemented for various tasks in defect detection such as tomato detection [1], road defect detection [2], textile fabrics [3], ceramic tiles [4], wafer defects [5], leaf diseases [6], eye diseases [7], human presence [8], etc. Object detection with the YOLO family is used widely in the defect detection algorithm. Several searches had proven the effectiveness of YOLO V5 in its accuracy and inference time [9][10].

The current famous technique of detecting the defect glove is based on region of interest (ROI) by using image processing. The research of automated detection of glove defects using vision control had proved that this method was able to achieve an overall accuracy of 81% with extracting the glove image from its background by image processing technique. Image processing used include grey-scale, morphology, thresholding, hole filtering and noise removal. The area of the object could be determined by converting the pixels to the other units so it can be used for classification. The average range for normal gloves is between 10000 and 20000 pixels. The research is also conducted without the light source and the accuracy can achieve up to 70.5% [11].

Andersen et al. in [12] evaluates the popular existing architectures in the area of corrosion and defect detection on the marine vessel with the hierarchical learning system with VGG16, VGG19, ResNet50, Inception V3 and Inception ResNet V2, for object image segmentation is Mask-RCNN and Yolact. 820 images of corrosion and 494 of non-corrosion in the marine vessel are used as the dataset, which is divided as 70% of the training set and 30% of validation. ResNet achieved the highest recall score of 0.99 in image classification while Faster-RCNN is the best performing architecture for object detection and Mask-RCNN is the best performing architecture, for instance, segmentation [12].

Besides, Shi et al. in [13] developed the research on the deep learning in defect detection of industry wood veneer in the fast speed has proved that deep learning gives a great accuracy in the high-speed production line. Wood defects account for only a small part of the overall industrial wood detection process, whereas a simple glance network was used

to identify the defective images because the glance network has a simpler structure than a mask network model, so the processing speed is higher. Glance network is created based on Convolutional Neural Network (CNN). Moreover, Network Attached Storage (NAS) is used to design a glance network to achieve accurate detection performance. To determine the form and precise position of wood defects, the defect images are then entered into the mask-RCNN network. Mask R-CNN uses ResNet50 and FPN for further feature extraction. This proposed model has reached the detection accuracy of up to 98.7%, and the mean average precision of the model reached 95.31%.

Lastly, Yang et al. [14] proposed a defect detection algorithm for tiny parts that are based on a single short detector network (SSD) and deep learning. This research is using a different algorithm which are YOLO V3, Faster-RCNN, FPN and their method, and the object detected is small, which is similar to our investigation. For this research, the data is collected from the high-speed industrial camera to capture the image of 0.8cm darning needle. The research results are that accuracy can be achieved 82.45% for YOLO V3, 82.5% for Faster-RCNN, 89.45% for FPN and 93.55% for their proposed method. To sum up, the accuracy of the algorithm proposed is higher than the other method because the proposed method uses an anchor mechanism to extract the features.

From the gathered literature, it is clearly shown that the development of automated glove defect detection models is still limited. Thus, the objective of this paper is to establish the deep learning model using YOLO V5 and compare it against other models such as YOLO V4, Detectron2 and EfficientDet for efficient detection of glove defects. It is worth noting that the proposed performance evaluations i.e., training time, model size, inference time and accuracy of different deep learning models has not yet been reported in the literature for glove defect detection. The finding of the research is important since the establishment of a baseline model could guide researchers and practitioners in this domain.

METHODOLOGY

The study was divided into two stages which are data preparation (first three sections) and model development (the final section). The main model was using the YOLO algorithm and trained by using google Collab. The evaluation of the performance is evaluated in the training set and test set. The training set was evaluated based on training time, accuracy and model size and the test set was evaluated based on inference time.

Data Collection

The dataset consists of three classes which are normal gloves (Figure 1 (a)), tear (Figure 1 (b)) and unstripped gloves (Figure 1 (c)). It was gathered from a glove company around Klang valley, Malaysia. The class of the normal glove consists of 701 images, 46 images for the class of the tear glove and 11 images for the unstripped glove. The total amount of images in this dataset is 493 images. Then, the dataset undergoes the labelling process before it starts to be trained.

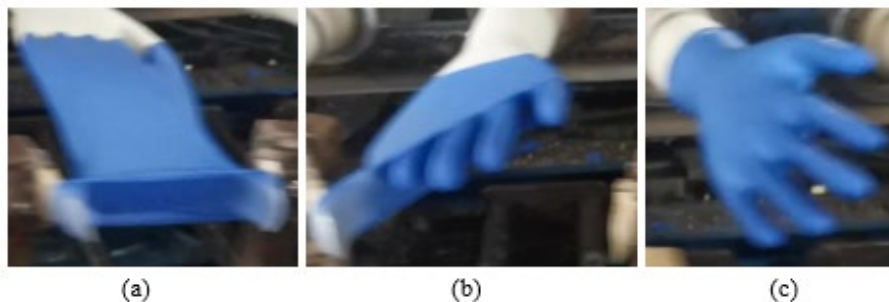


Figure 1. Sample images for each class (a) normal glove, (b) tear glove, (c) unstripped glove.

Data Pre-processing and Augmentation

Before the dataset is sent for developing, auto-orientation and resize is necessary to fasten the training process. The images are undergone auto-orient to make sure the photo is in portrait mode. For resize, the original size of the input image is 1280 pixels in width and 720 pixels in height. After downsizing, it becomes 416 pixels in width and 416 pixels in height.

There are three image augmentations used in this research which are saturation, exposure and noise augmentation. After the data argumentation, the total image for the dataset is increased from 493 images to 1148 images which the class for normal gloves rose to 2103 images, 138 images for the class of tear glove and 33 images for the class of unstripped glove.

Data Splitting

Train-test split ratio is taking the dataset and splitting it into three subsets which are training set, validation set and test set. In order to choose the most suitable training-test split ratio, the dataset is tested in several ratios and the result for the accuracy and the training time for the training set and validation set are recorded in Table 1. Based on Table 1, the train-validation-test split ratio of 70-20-10 reaches the highest accuracy and the lowest training time. Hence, this train-test ratio is chosen.

Table 1. Performance comparison of different train-test split ratio.

Training set (%)	Validation set (%)	Test set (%)	Accuracy (mAP)	Training time
70	20	10	0.9951	15min 55sec
67	17	17	0.9938	30min 04 sec
56	19	25	0.9906	21min 54 sec

Model Development

For YOLOv5, the network architecture can be divided into 4 parts which are input, backbone, neck and prediction as shown in Figure 2. The main difference of YOLOv5 with the other model is because of CSP structure in its backbone structure and FPN with PAN structure in its neck structure.

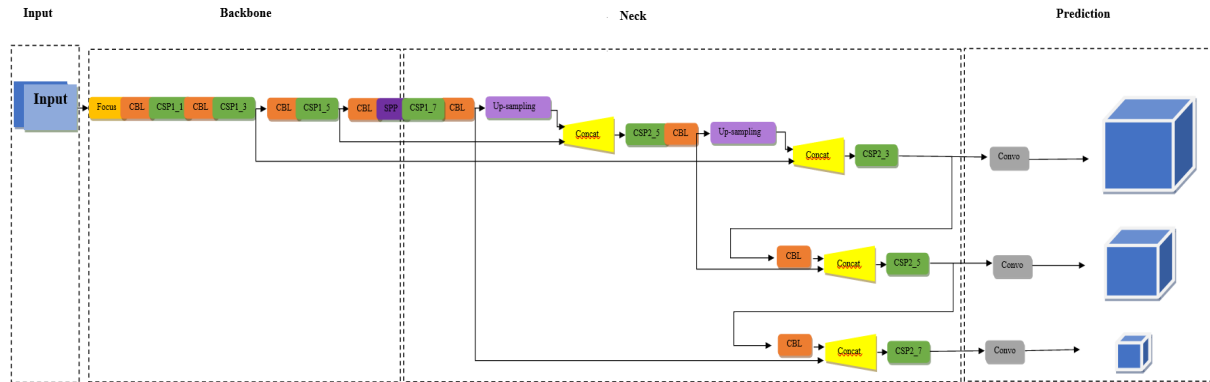


Figure 2. Network architecture of YOLOv5.

During the training process, the result may be varying if the parameter used to train the model is different, therefore the parameter has to be consistent. The number of training epochs will directly affect the result of the training process. Thus, several training epochs are tested and the result for the accuracy and the training time is recorded as in Table 2. Based on Table 2, the training epochs of 100 reach the shortest training time and highest accuracy, therefore this parameter is used as the training parameter for all the deep learning models in this research.

Table 2. Comparison of training epochs.

Training epochs	Accuracy (mAP)	Training time
100	0.9951	15min 55sec
500	0.9950	1h 19min 23sec
1000	0.9948	2h 36min 2sec
2000	0.9803	6h 14min 12sec

The whole process for developing the training model of YOLOv5 is done in Google Collab. The process starts by installing the YOLOv5 dependencies to create the environment. Then, import the dataset which is in YOLOv5 format. Next, specify the training configuration by selecting YOLOv5s from the model of YOLOv5 in this process because it is the fastest base model. Then, define the parameter of training the model. In this research, the input image size is 415 pixels because the input image has been downsized into 415 pixels so it can fasten the training process, batch size for training the model is 16 and the number of training epochs is 100. After setting the configuration, run the training model. Next, evaluate the performance of the trained model by Tensor Board once the training mode has been completed and visualise the training data. In this research, the training time, model size and accuracy of the training set is recorded. Then, run the trained model on the test image. The inference time of the model on the test set is recorded. After training has completed, export the developed YOLOv5 weight for future inference. Lastly, the trained YOLOv5 model is compared with Scaled-YOLOv4, Detectron2 and EfficientDet with the same parameter setup to prove its effectiveness and robustness on detecting the defect gloves.

RESULTS AND DISCUSSION

To evaluate the performance of the model, the training time for training the models are recorded and compared. The training time may be varying due to the different size or quality of the data input or training model, but in this research, the data is fixed for all models. In this research, the data is fixed to go through 100 epochs with 129 iterations per epoch. The result is recorded and shown in Figure 3. From Figure 3, we have found out that YOLO V5 uses the least time to train.

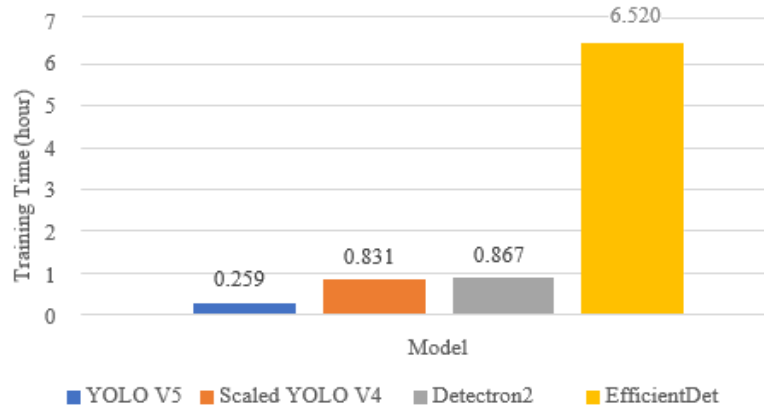


Figure 3. Comparison of training time.

The size of the model represents the file size of the weights. To evaluate the robustness of the model, the model size is calculated. The output trained model for YOLOv5 and Scaled YOLOv4 is in the form of .pt while Detectron2 and EfficientDet is in the form of .pth. Based on Figure 4, the largest model size is Scaled YOLOv4 with 410940kb and the smallest model size is YOLOv5 with 14418kb.

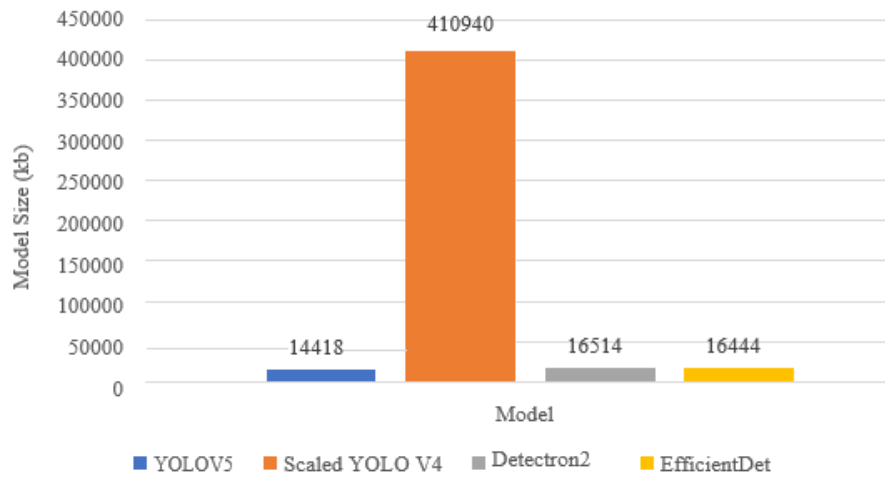


Figure 4. Comparison of model size.

In order to estimate and compare the accuracy of the model, the mAP (mean Average Precision) metric of the model is calculated. The accuracy of the model shows the correctness of prediction of the object. Therefore, the higher the accuracy, the higher the correctness of prediction of the object. The high accuracy represents the better model. The mAP computes the average precision value for recall value over 0 to 1. The mAP is collected from the training set, Based on Figure 5, the YOLOv5 reaches the highest accuracy which is up to 0.9951 and the lowest accuracy of the model is Detectron2.

To analyse the performance of the model, the inference time of the model is compared. The inference time is the time taken for each image of the test set passing through the model. In this research, 346 images are used to pass through the four models and the inference time is recorded and compared. From Figure 6, it shows that the shortest inference time used is EfficientDet with 0.763s and the longest inference time used is Detectron2 which is 15.96s.

The purpose for this research is to develop and determine the most suitable object detection model for defect glove detection. The dataset used for this object detection is normal glove, defect glove and unstripped glove. The input data is being pre-processed to increase the effectiveness of the model then sent to undergo augmentation to increase the size of the model. Due to the imbalance of the class which the number of images for defect glove and unstripped glove is limited compared to the normal glove, the augmentation is necessary in order to overcome this problem.

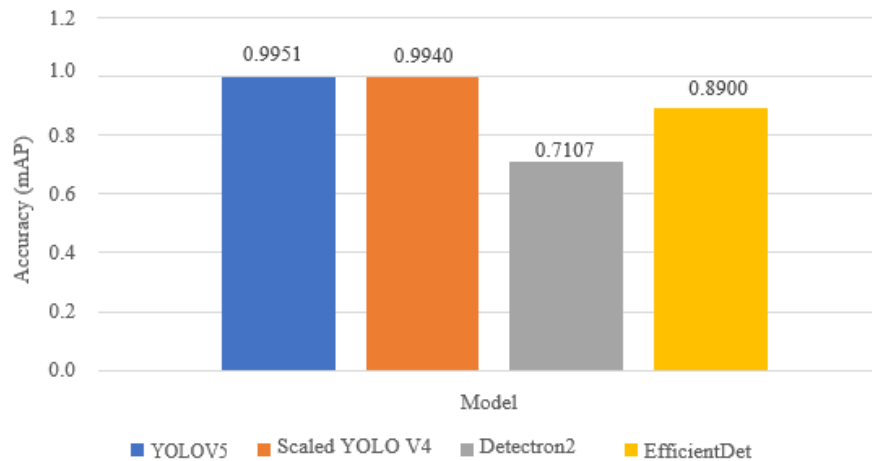


Figure 5. Comparison of accuracy.

To fulfil the objective of this research, the models are compared by its training time, model size, inference time and accuracy. Based on the results collected, the best model for defect glove detection is YOLOv5, because it has the lowest training time, model size, inference time and highest accuracy which is up to 0.9951. That is because YOLOv5 created CSPDarknet as the backbone of Darknet by integrating Cross Stage Partial Network (CSPNet) into it. With CSPNet, it can help to resolve the problem of repeating gradient information which will increase the training time and inference time in large-scale backbones. Besides, it also helps to incorporate the gradient changes into the feature map which will decrease the parameter and fps (frame per second), resulting to decrease the inference speed and accuracy, but also reduce the model size. To further improve the speed of training and inference time, YOLOv5 uses Path Aggregation Network (PANet) in its middle part of architecture to improve the information flow. PANet uses a redesigned FPN structure with a better bottom-up approach which will enhance the low-level feature propagation. Simultaneously, the adaptive feature propagation connects the feature grid to all the feature levels. It is employed to ensure that the important information from each feature level can reach the next sub-network. By means of this, it can improve the location accuracy of the object by improving the utilisation of accurate localisation signals in the lower layers. Thus, it can reach high accuracy.

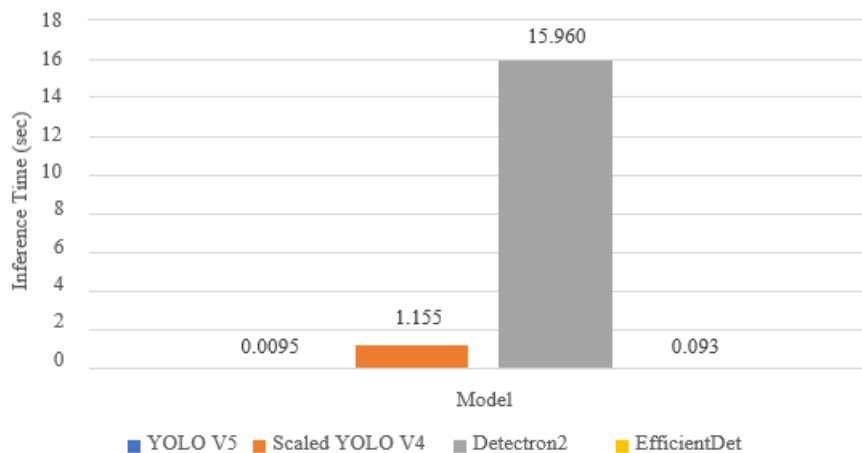


Figure 6. Comparison of inference time.

CONCLUSION

In conclusion, the proposed defect glove detection using computer vision and deep learning approach with YOLOv5 algorithm has successfully detected three different types of gloves which are normal glove, tear glove, unstripped glove and able to reach lowest training time (0.259 hours), model size (14418kb), inference time (0.0095s) and highest accuracy (mAP = 0.9951). Besides, this model can be further improved by increasing the class of defect gloves. In the real market, the defect glove mainly can be divided into tearing, knocking, incomplete beading, dirty, stain, hole, double dipping, finger not enough and touching. But in this research, there are only 3 classes due to the limitation of the dataset obtained. With the purpose of solving this issue, the classes for the glove dataset can move forward a single step to increase the number of the classes. Moreover, to maintain or improve the accuracy of the model if the number of classes and the size of the dataset is increased, the model needs to be fine-tuned because the dataset has become complex and the object detection model is hard to recognize if they are too similar.

REFERENCES

- [1] O. P. Toon, M. A. Zakaria, A. F. A. Nasir, A. PP A. Majeed, C. Y. Tan, and L. C. Y. Ng, "Autonomous Tomato Harvesting Robotic System in Greenhouses: Deep Learning Classification," *MEKATRONIKA*, vol. 1, issue 1, pp. 80–86, 2019, doi: 10.15282/mekatronika.v1i1.1148.
- [2] B. Akarsu, M. Karakose, K. S. Parlak, E. Akin, and Alisan Sarimaden, "A Fast and Adaptive Road Defect Detection Approach Using Computer Vision with Real Time Implementation," *International Journal of Applied Mathematics, Electronics and Computers*, vol. 4, pp. 290–290, 2016, doi: 10.18100/IJAMEC.270546.
- [3] K. L. Mak, P. Peng, and H. Y. K. Lau: "A Real-time Computer Vision System for Detecting Defects in Textile Fabrics," *2005 IEEE International Conference on Industrial Technology*, pp. 469-474, 2005, doi: 10.1109/ICIT.2005.1600684.
- [4] Z. Hocenski, T. Matic, and I. Vidovic: "Technology transfer of computer vision defect detection to ceramic tiles industry," *2016 International Conference on Smart Systems and Technologies (SST)*, pp. 301-305, 2016, doi: 10.1109/SST.2016.7765678.
- [5] J. A. Mat Jizat, A. P.P. Abdul Majeed, A. F. Ab. Nasir, Z. Taha, and E. Yuen, "Evaluation of the machine learning classifier in wafer defects classification", *ICT Express*, 2021, doi: 10.1016/j.icte.2021.04.007.
- [6] B. S. Bari, M. N. Islam, M. Rashid, M. J. Hasan, M. A. M. Razman, R. M. Musa, A. F. A. Nasir, and A. P.P. A. Majeed, "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework", *PeerJ Comput. Sci.* 7:e432, 2021, doi: 10.7717/peerj-cs.432.
- [7] F. N. Mohd Noor, W. H. Mohd Isa, I. Mohd Khairuddin, M. A. Mohd Razman, J. A. Mat Jizat, A. F. Ab Nasir, R. Muazu Musa, A. PP Abdul Majeed, "The Diagnosis of Diabetic Retinopathy: A Transfer Learning with Support Vector Machine Approach," *Advances in Robotics, Automation and Data Analytics: Selected Papers from ICITES 2020*, vol. 1350, pp. 391-398, 2021.
- [8] V. Balachandran, M. N. A. Shapiee, A. F. A. Nasir, M. A. M. Razman, and A. PP A. Majeed, "Deep Learning Based Human Presence Detection," *MEKATRONIKA*, vol. 2, issue 2, pp. 55-61, 2020, doi: 10.15282/mekatronika.v2i2.6768.
- [9] V. A. Adibhatla, H.-C. Chih, C.-C. Hsu, J. Cheng, M. F. Abbod, and J.-S. Shieh, "Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once," *Mathematical Bio-sciences and Engineering*, vol. 18, issue. 4, pp. 4411–4428, 2021, doi: 10.3934/mbe.2021223.
- [10] S. Li et al.: "Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm," *Construction and Build-ing Materials*, vol. 273, p. 121949, 2021, doi: 10.1016/j.conbuildmat.2020.121949.
- [11] K. Thang and N. S. Lai: "Automated Detection of Glove Defects Using Vision Control," *International Journal of Engineering & Technology*, vol. 16, no. 3, pp. 18-23, 2016.
- [12] R. Andersen, L. Nalpantidis, O. Ravn, and E. Boukas: "Investigating Deep Learning Architectures towards Autonomous Inspection for Marine Classification," *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 197-204, 2020, doi: 10.1109/SSRR50563.2020.9292621.
- [13] J. Shi, Z. Li, T. Zhu, D. Wang, and C. Ni, "Defect Detection of Industry Wood Veneer Based on NAS and Multi-Channel Mask R-CNN," *Sensors*, vol. 20, no. 16, p. 4398, 2020, doi: 10.3390/s20164398.
- [14] J. Yang, S. Li, Z. Wang, and G. Yang, "Real-Time Tiny Part Defect Detection System in Manufacturing Using Deep Learning," *IEEE Access*, vol. 7, pp. 89278–89291, 2019, doi: 10.1109/ACCESS.2019.2925561.