Multiclass Classification and Defect Detection of Steel Tube Using Modified YOLO

Chapter · November 2023

DOI: 10.1007/978-981-99-8184-7_32

5 authors, including:

Deepti Raj
VIT University
16 PUBLICATIONS 119 CITATIONS
SEE PROFILE

Prabadevi. B
VIT University
76 PUBLICATIONS 2,511 CITATIONS
SEE PROFILE

Thippa Reddy Gadekallu
Jiaxing University
422 PUBLICATIONS 18,912 CITATIONS
SEE PROFILE

Surbhi Bhatia
University of Salford
198 PUBLICATIONS 2,955 CITATIONS
SEE PROFILE

Multiclass Classification and Defect Detection of Surfaces Using Modified-YOLO

1st Deepti Raj G
School of Information Technology and Engineering
Vellore Institute of Technology
Vellore, India
deeptiraj.g2020@vitstudent.ac.in

2nd Dr.Prabadevi.B
School of Information Technology and Engineering
Vellore Institute of Technology
Vellore, India
prabadevi.b@vit.ac.in

2023 12th International Conference on Advanced Computing (ICoAC) | 979-8-3503-1821-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICoAC59537.2023.10249530

***Abstract*—Industrial products based on glass fiber are widely used in various industries due to their excellent properties. Steel tubes known for its strength and durability are widely used in hazardous high pressure environments such as the petroleum, chemical, natural gas and shale gas industries. Deep learning greatly improves inspection efficiency in object detection and defect detection. In this work, we use a well-known You Only Look Once version 7(YOLOv7) model to achieve accurate defects detection of steel tube and glass tube images. First, the classification of the dataset is checked using a VGG16(Visual Geometry Group) and EfficientNet. A Convolutional Block Attention Module (CBAM) mechanism is then integrated into the YOLOv5 backbone network to improve feature expression ability. Additionally, the Generalized Intersection over Union (GIoU) loss function is used to enable the model to learn precise object localization. Experimental results show that the modified YOLOv7 is better than the other models.**

***Index Terms*—Classification, Defect Detection, YOLO, YOLOv7, CBAM Attention, GIoU Loss.**

I. Introduction

A. Need for Defect classification and Detection

The process of classifying and identifying defects is very important in many industries such as manufacturing, software development and quality control [12]. This critical process helps detect and classify errors and defects in systems, processes and products. Ensuring the quality of goods and services is highly dependent on accurate defect detection and classification. Organizations that can effectively identify and categorize errors are able to analyze and fix the underlying problem, resulting in better service. The benefits of early error detection and classification are enormous. Businesses can ensure customer satisfaction, comply with regulatory requirements, and improve their overall reputation. By quickly identifying and resolving issues in your production or development process, you can avoid costly rework, recalls, and customer complaints, and significantly reduce your costs [8]. In addition, efficient defect classification and detection helps reduce defective, defective, or damaged products. Through this process, the root causes of errors can be understood, allowing companies to take targeted corrective action and improve overall operational efficiency.

By analyzing the data collected, companies can uncover patterns, trends and common causes of defects so that they can take preventive actions, improve overall efficiency and make informed decisions regarding process improvement. Defects pose risks to the end user as well as the company, and especially in critical industries such as medical, aerospace, and automotive, defects can lead to safety concerns and life-threatening situations. To mitigate these risks, organizations must implement appropriate corrective actions and ensure compliance with security standards through effective error detection and detection. By systematically collecting and analyzing data about errors, businesses can learn from their mistakes, gain insight into recurring problems, and identify root causes [13]. This information can be used to improve product design, improve manufacturing techniques, and provide effective employee training. Defects can lead to customer dissatisfaction, negative reviews, and loss of brand trust. So, organizations should quickly address consumer concerns and implement rapid remediation to maintain reputation and customer satisfaction.

B. Classification and Object detection in Glass tube and steel tube

Computer vision technology is a valuable tool for object recognition, allowing algorithms to analyze images and videos to accurately identify and locate specific objects [11]. In various industrial contexts such as manufacturing, construction, and infrastructure, classification and object recognition techniques play an important role in ensuring the quality and integrity of surfaces [5]. The classification process involves classifying surfaces based on their inherent characteristics and properties. Defect detection is another important aspect of maintaining surface quality. Detection of defects, anomalies, or imperfections is essential to ensure the safety and reliability of systems in which these tube surfaces are used. The most common defects on steel tube surfaces include cracks, rust, dents, bends, welding defects, surface irregularities, etc., while glass tube surfaces have defects such as airlines, scratches, knots, cracks, etc. Various non-destructive inspection methods are used to detect these defects. These methods include optical inspection, ultrasonic inspection, magnetic particle inspection, and X-ray inspection. These techniques enable early detection of defects, facilitate timely maintenance, and prevent potential failures that can compromise structural integrity.

The classification and object detection process typically involves a combination of manual inspection and automated methods. Human inspectors play an important role in visually inspecting and classifying tube surfaces based on their unique properties. However, to increase efficiency and accuracy, automated techniques such as computer vision and machine learning algorithms are used to detect errors and identify specific objects. The automated method features the precise location of defects, allowing manufacturers, inspectors and end users to determine if a surface is suitable and safe for its intended use. Leveraging the expertise of human inspectors and the power of automated technology, this collaborative approach enables comprehensive assessment of tube surfaces.

C. Object Detection-YOLO

YOLO or ”You Only Look Once” is a very popular and efficient deep learning algorithm used for object recognition and detection in computer vision. Its outstanding speed and real-time processing capabilities isuseful for a variety of applications such as robotics, video surveillance, and autonomous vehicles. YOLO's architecture is simple compared to other object detection methods. It uses a unified approach to perform object detection and classification in one step, as opposed to the two-step techniques used by other algorithms. By looking at the entire image instead of focusing on a specific area, YOLO collects comprehensive information for accurate object detection. YOLO uses anchor boxes to handle objects of various sizes and shapes and predicts offsets to these anchor boxes for effective detection. Large datasets can be used to train efficiently, eliminating the need for extensive fine-tuning and costly computation for local recommendations.
YOLO's grid-based strategy divides the input image into grids and allows prediction of objects that fit into each assigned grid cell. This approach properly detects multiple instances of the same object class and effectively handles duplicate objects. Due to its versatility, YOLO can be applied to industrial inspection, quality control, medical imaging, and more. The latest versions of YOLO are YOLOv7 and YOLOv8, with YOLOv7 gaining popularity in the industry for its efficient error detection and its balance of speed and accuracy. YOLOv8 is considered the state of the art in object recognition, but there is no official paper on it yet.

D. Contribution of the paper

Conventional methods have limitations in achieving and effectively utilizing all the characteristics for classification process. This study first analyses two classification techniques on the datasets. To address this issue, they used classification techniques on their dataset and made improvements to the YOLOv7 base model. Two major improvements were made to the YOLOv7 model to find a balance between accuracy and detection time. First, we integrated the CBAM attention module to improve the model's feature extraction ability. Second, we integrated the GIoU loss function to further improve model to learn about precise localization. The resulting model, proved to perform better than original YOLOv7 model when

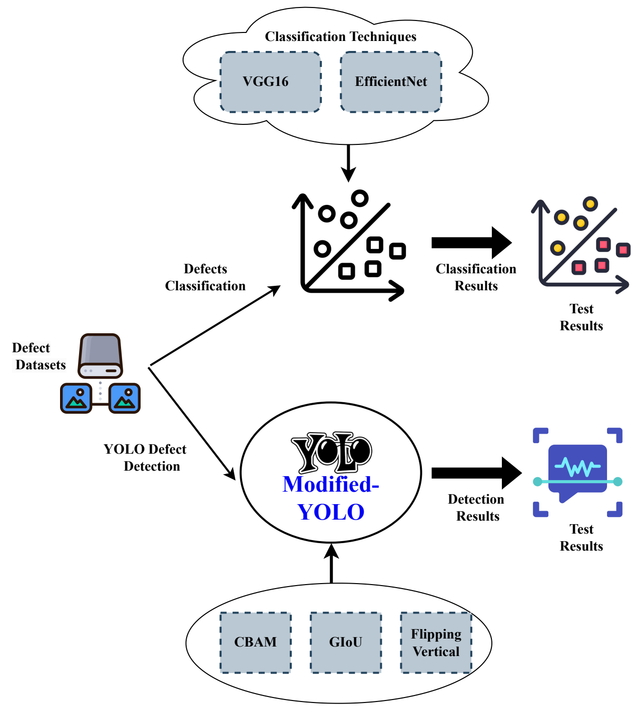

Fig. 1. Flow of the paper

tested on steel and glass tube datasets. Experimental results confirmed the superiority of the proposed method in detecting and classifying surface defects in these materials.

The contributions of this paper are as follows:

1) Two classification techniques such as: VGG16 and EfficientNet are used to check the classification of the classes in the dataset.
2) To improve detection accuracy, CBAM is integrated to YOLOv7. GIoU is used instead of traditional Complete-IoU (CIoU) to improve model convergence and localization process.
3) The model is trained with vertical flipping technique.

The rest of the paper is structured as follows. The second section presents a comprehensive literature review. In the third section, two classification techniques are described with an overview of the YOLOv7 detection framework, a description of the CBAM mechanism, and a description of the GIoU loss function. The fourth section presents the experimental datasets used in the study and the metrics used to evaluate the model's performance. The fifth section presents test results obtained from experiments. Finally, this paper is concluded with the results and draws conclusions based on the results of the Modified-YOLOv7 method.

II. LITERATURE SURVEY

In a study by Yang et al., [19], YOLOv5 was used to detect defects in steel surfaces. The effectiveness of single-

shot object detection was demonstrated by comparing the results with a faster region-based CNN (R-CNN). Wang et al. [18] proposed a method to automatically detect steel tube defects by combining advanced Faster R-CNN with advanced ResNet50. The goal was to reduce average execution time while improving accuracy. In another study by Li et al., [10], the authors built latest YOLO series, YOlOX, and introduced SCED-Net (Steel Coil End Surface Detection Network). By improving the data augmentation method, incorporating the proposed multi-information fusion module, and utilizing focus loss and soft NMS, SCED-Net has achieved significantly improved results. Zen et al. [20] used a convolutional neural network (CNN) to extract features from an augmented image set using transfer learning methods. They then used a hierarchical model ensemble for location-based fault detection. Rani et al. [14] introduced CondenseNetV2, a lightweight CNN-based model optimized for small defect inspection. This model is designed to work efficiently on low frequency edge devices. Using the current set of sparse feature reactivation modules, the authors achieved effective feature extraction with minimal computational effort. The model was thoroughly tested on an edge device (NVIDIA Jetson Xavier Nx SOM) using two real-world data sets.
The authors [3] present an algorithm based on moving average filters operating at the row level that can minimize the effects of glass tube rotation, vibration, and irregularities and reduce processing time during the defect detection stage. This algorithm has better accuracy and outperforms in terms of processing time. The authors [1] improve MobileNetV2 and propose a multi-scale cross-fusion attention module that combines it with a computer vision image feature extraction method for hairiness detection. Finally, modified MobileNetV2 network is used as the backbone of the YOLOX network to achieve more efficient defect detection on glass tubes. A CBAM-based target detection method for YOLOv7 to detect defects in transparent glass tubes is proposed by [15]. All pooling layers in YOLOv7 are replaced by CBAM to better capture the target's features.

These studies demonstrate different approaches and models for detecting defects in steel tubes and glass tube surfaces, each offering unique contributions and advantages in terms of accuracy, speed and suitability for edge devices.

III. Methodology

A. Classification Methods

Classification algorithms play an important role in computer vision and defect detection because they enable automatic identification and classification of objects, patterns, and anomalies in image and video data. These methods are based on machine learning algorithms that learn from labeled training data and use that knowledge to classify new unlabeled data based on learned patterns and features. Various computer vision tasks such as object detection, object recognition, image segmentation, and scene understanding utilize classification techniques [2]. To achieve this, a classification model is trained on a large collection of labeled images so that it can recognize different elements and classes in the input images. After training, these models can accurately classify new images, enabling automatic object recognition in real-world applications. This ability is especially valuable in areas such as self-driving cars, surveillance systems, and robotics, where the ability to recognize and understand environmental influences is critical. Convolutional neural networks (CNN) and other feature extraction techniques are widely used to automatically learn discriminating features from images to improve the performance of classification models in these areas [16]. The classification process involves recognizing patterns, correlations, and trends in datasets. This makes it easier to organize and classify data for effective analysis and decision making. A variety of computer vision classification techniques are available, including supervised learning, unsupervised learning, transfer learning, deep learning, and ensemble techniques. These techniques can be tailored to specific problem domains. In this study, we used the VGG16 model and the EfficientNet model to validate the classification performance of different classes in the dataset. These models are widely used in the computer vision community and have shown promising results on various classification tasks.

1) VGG16: VGG16 was the foundational model for computer vision, paving the way for deeper and more complex architectures. VGG16 outperforms the latest models in terms of performance and efficiency. VGG16 gained popularity due to its excellent performance in the 2014 ImageNet Large Scale Visual Recognition Competition (ILSVRC). The VGG16 architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. A key feature of VGG16 is the use of small 3x3 convolution filters throughout the network [9]. This allows deeper networks without excessively increasing the parameters. The VGG16 architecture consists of four layers: an input layer, a convolutional layer, a fully connected layer, and a softmax layer. When training VGG16 for image classification, labeled data (input images with good class labels) is used to optimize the network weights and biases. This is typically done using stochastic gradient descent (SGD) or some other optimization algorithm. The goal is to minimize the cross-entropy loss, which measures the difference between predicted probabilities and true class labels. After training, VGG16 can be used for image classification tasks by passing input images through the network. The class with the highest probability in the output softmax layer is taken as the predicted class of the input image.

2) EfficientNet: The key idea of EfficientNet is to tune the depth, width and resolution of the network in a principled way to find a balance between model size and accuracy. This is achieved using a compound scaling method that systematically scales the model dimensions based on a user-defined scaling parameter, usually called phi [7]. EfficientNet architecture is based on a relatively small and efficient core network. Applying compound scaling increases the depth, width, and resolution of the base network, creating a larger and more powerful model. Scaling is done to balance increased model complexity to avoid overfitting and excessive computational costs. The main components of EfficientNet architecture are baseline network, composite scaling, depth scaling, width scaling, and resolution scaling. EfficientNet achieved the best performance while maintaining computational efficiency on various image classification benchmarks. It is commonly used in computer vision tasks such as image recognition, object recognition, and segmentation. EfficientNet can adjust the model size using the phi scaling factor, making it a versatile choice for a variety of applications and computing resources.

B. Background of YOLOv7

The YOLOv7 architecture offers significant architectural improvements, including additive scaling, an enhanced efficient layer aggregation network (EELAN), a large number of freebies with planned and redesigned convolutions, coarseness of auxiliary loss and fine lead loss. [17] [6] . These changes improve support for both mobile and conventional GPUs. Considering these advances, the backbone of YOLOv7 is chosen as the feature extraction network for the model.

1) Backbone: The initial backbone of the YOLOv7 model consists of four CBS blocks, each block consisting of a convolution layer, a batch normalization, and a SiLU (Sigmoid Linear Unit) layer. These blocks are stacked to process the input image and generate intermediate features. Subsequently four convolution operations are applied to extract the underlying features from these intermediate representations. YOLOv7 uses Max Pooling (MP) and E-ELAN (Extended Efficient Layer Aggregation Network) modules to extract detailed features from previously obtained intermediate features. These modules help the model capture more detailed and meaningful patterns, improving its ability to detect objects and anomalies in the input data.

2) Feature Fusion Zone: The network feature fusion layer is designed to improve the network's ability to learn features from the underlying network. This is achieved by learning features separately at different resolutions and centralizing the combination of those features. The goal is to enable the network to learn a wide range of image features and integrate them effectively to improve overall performance in feature extraction and object detection tasks.

3) Detection Head: The detection heads of the YOLOv7 algorithm is responsible for processing the various feature scales produced by the model, specifically 20x20, 40x40 and 80x80 [21]. Each of these three feature scales is designed to detect different sized objects in the input image. The 20x20 scale is especially suitable for detecting small targets, the 40x40 scale is suitable for medium-sized targets, and the 80x80 scale is specialized for detecting larger targets. This multi-scale approach enables YOLOv7 to effectively handle objects of different sizes, greatly improving its object detection capabilities in different scenarios.

C. Modified-YOLOv7

In this work, we use transfer learning to get the weights from the original YOLOv7 model pretrained on the COCO dataset. Modified-YOLOv7 is formed by integrating the CBAM attention mechanism and the GIoU loss function into YOLOv7 backbone. Additionally, the model has been optimized through the use of vertical flip augmentation to further improve the performance of detecting surface defects in steel and glass tubes.

1) Attention Mechanism: The CBAM attention module is designed to specifically emphasize informative features in CNNs and suppress irrelevant features. The CBAM module consists of two main components. Channel Attention Module (CAM) is responsible for obtaining per-channel attention by considering interdependencies between different functional channels. Spatial Attention Module (SAM) obtains spatial attention by considering interdependencies between spatial locations in feature maps. By combining channel and spatial attention mechanisms, the CBAM module enables CNNs to adaptively focus on important features, resulting in various computer vision tasks such as image classification, object detection, and segmentation [19]. This leads to better performance and better feature representation. The CBAM attention module has been shown to effectively improve the performance of CNNs while adding relatively few parameters, making it a valuable tool for enhancing the capabilities of modern deep learning models [4].

2) GIoU Loss Function: The GIoU loss function is a single differentiable loss function used in the context of object detection tasks, especially anchor-based object detection algorithms such as YOLO and SSD (Single Shot Multibox Detector). This is an extension of the classic Intersection over Union (IoU) loss that measures the overlap between the predicted bounding box and the ground truth bounding box. The main problem with using IoU alone as the loss function is that it does not take into account the localization accuracy of the predicted boundary frames. As a result, it can be difficult for the model to learn accurate localization, which is essential for accurate object detection. GIoU addresses this limitation by including an additional term IoU calculation that facilitates better localization. The GIoU loss function is defined as

$$L_{GIoU} = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|} \tag{1}$$

In the GIoU loss formula above, C is the smallest box that covers the prediction box and the ground truth bounding box, acting like a penalty term to bring the prediction box closer to the target ground truth box. The GIoU loss function provides a more comprehensive measure of predicted bounding box quality compared to traditional IoU-based loss. By considering both spatial overlap and spatial extent, GIoU encourages the model to generate more accurate and local bounding boxes, resulting in better object detection performance. It has been widely adopted by various object detection frameworks and has proven effective in training models with better localization skills.

3) Optimization: Optimizing deep learning models is important for improving performance, achieving faster convergence, efficient utilization of resources, scalability, edge device deployment, interpretability, and enabling transfer learning. It

TABLE I
EVALUATION INDICES FOR CLASSIFICATION

Parameter	Value
Input Image Size	180*180
Epochs	100
Batch	16
Optimizer	Adam
Initial, Final learning rate	1*10-3, 1*10-2

employs compression, parameter and architectural optimization, and task-specific tuning to create more accurate models that train faster and deploy more effectively to resource-constrained devices. To optimize the YOLO model, data augmentation techniques such as random scaling, rotation, flipping, and cropping are applied during training. These enhancements increase the data coverage, improve the model's ability to generalize to new cases, and improve the efficiency of object detection. Additionally, data augmentation helps address class imbalance and overfitting issues, resulting in more reliable and accurate YOLO models. In this work, we generalized YOLOv7 model using the vertically flipped data augmentation technique, contributing to an overall improvement in model performance and robustness.

IV. DATASETS DESCRIPTION, EVALUATION INDICATORS AND METRICS

The datasets used in this study are the steel tube dataset and glass tube. This steel tube dataset consists of a total of 3344 images classified into 8 defect classes: air hole, air hole hollow, bite edge, broken arc, crack, overlap, slag inclusion, and unfused defects. The glass tube dataset consists of 2 defects: blot and spot. This work was performed on a desktop with an Intel(R) Core(TM) i5-1035G CPU and 12 GB of memory in Windows 10 operating system. The implementation is done on a Google Colaboratory Notebook that enables GPU computation acceleration with NVIDIA T4.

In this study, the performance of the Modified-YOLOv7 model was evaluated using multiple measures to demonstrate its improved capabilities. The primary metrics used for evaluation were mean Average Precision (mAP) with precison, recall and F1 score. Table I refers to the different evaluation indices used for classification. The evaluation indices for detection are: epochs-100, batchsize-16, input image size-640*640, weight decay-5*10-4, Momentum-937*10-3, weights yolov7.pt, SGD optimizer, initial and final learning rates are 0.01 and 0.001 respectively.

V. RESULTS

A. Classification

Two classification models VGG16 and EfficientNet models are used in this study. The classification recognition accuracy for the steel tube and glass tube dataset is shown in the table II.

Both models provide virtually the same classification accuracy, but the EfficientNet model classifies 0.54% more accurately on steel tube dataset and VGG16 models has 0.75% more accuracy on glass tube dataset. This indicates that the

TABLE II
CLASSIFICATION RESULTS

Dataset Name	Model Name	
	VGG16	EfficientNet
Steel Tube	97%	97.54%
Glass Tube	95.75%	95%

TABLE III
METRICS COMPARISON FOR DIFFERENT MODELS

Model Name	Glass Tube Dataset				Steel Tube Dataset			
	mAP@0.5 (%)	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)	Precision (%)	Recall (%)	F1-Score (%)
YOLOv7	82.1	78.9	75	76.9	88.5	86.8	85.7	85.9
YOLOv7-CBAM	82.6	77	75.6	76.2	90.4	86.7	85.3	85.9
YOLOv7-GIoU	80.2	72	78.8	75.2	90.8	87.2	85.9	86.5
Modified-YOLOv7	**83.7**	**74.5**	**80**	**77.1**	**92**	**90.5**	**88.8**	**89.6**

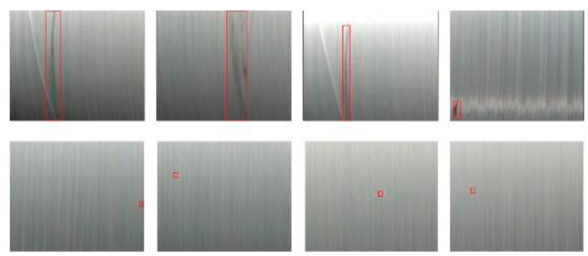

Fig. 2. Visual detection results on Glass tube Dataset. In sequence, the defects marked are: Blot and Spot

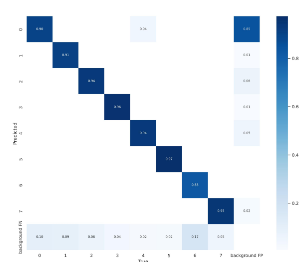

Fig. 3. Confusion Matrix of Modified-YOLOv7 on Steel tube

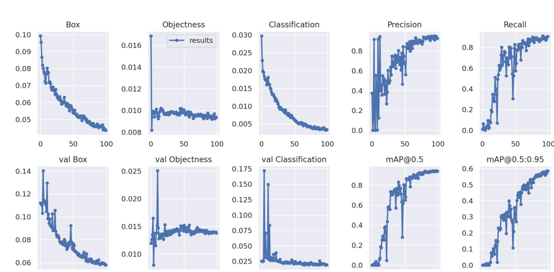

Fig. 4. Results Graph for Modified-YOLOv7 on Steel tube

dataset has been classified correctly and can be used to search for defects.

B. Detection

Table III compares mAP, precision, recall, F1 score between models. YOLOv7-CBAM integrates the CBAM mechanism into the YOLOv7 network structure. In YOLOv7-GIoU, GIoU loss function is used instead of traditional CIoU, and Modified-YOLOv7 is formed by taking CBAM, GIoU, hyper-parameter vertical flipping. As can be seen when compared with the original YOLOv7 model, on the steel tube dataset, the revised Modified-YOLOv7 model improved the average accuracy by 3.95%, improved the precision by 4.26%, and recall by 3.61%. On glass tube dataset we observed Modified-YOLOv7 model has increased accuracy by 0.97%, recall by 6.66%. The visual detection results on steel tube dataset with Modified-YOLOv7 are shown in Figure 2.
A confusion matrix provides an overview of a classifier's performance by tabulating the number of correct and incorrect predictions for each class. Figure 3 shows the confusion matrix specific to the performance of the modified YOLOv7 model in the steel tube dataset. Figure 4 show the results graph for Modified-YOLOv7 on the Glass tube dataset. Based on the comparison and analysis of the experiments performed, Modified-YOLOv7, an improved version of the YOLOv7 algorithm proposed in this study, shows obvious advantages in terms of detection accuracy.

VI. Conclusion

In this work, we present an extension of the YOLOv7 base model called Modified-YOLOv7. It is specifically designed to identify defects in steel and glass tube surface datasets with complex backgrounds. To validate the classification of the dataset, we used two classification methods, VGG16 model and EfficientNet. Our analysis revealed that both classification models have similar accuracy. However, the EfficientNet model outperformed VGG16 by 0.54% in classifying the steel tube dataset, and VGG16 showed superior accuracy in classifying the glass tube dataset. To enhance the model's feature extraction capabilities, a CBAM (Convolutional Block Attention Module) mechanism is integrated into the head of YOLOv7. Furthermore, replacing the original CIoU loss function with the GIoU loss function resulted in improved model performance. To further refine the model, we also introduced a vertical flip augmentation technique during training. This contributes to the model's ability to generalize and perform well in real-world scenarios. Experimental results show that by incorporating the above tactics, the Modified-YOLOv7 has improved accuracy on both the datasets compared to the original YOLOv7. Grayscale images are currently used for the dataset in this study. To broaden the scope of our investigation, we plan to explore the use of hyperspectral images and evaluate how well the model performs on this type of data.

References

[1] Junmin Bao, Junfeng Jing, and Yaohua Xie. A defect detection system of glass tube yarn based on machine vision. *Journal of Industrial Textiles*, 53:15280837231152878, 2023.

[2] Jason Brownlee. A gentle introduction to object recognition with deep learning. *Machine Learning Mastery*, 5, 2019.

[3] Gabriele A De Vitis, Pierfrancesco Foglia, and Cosimo A Prete. Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase. *IET Image Processing*, 14(12):2911–2921, 2020.

[4] Jian Duan, Xi Zhang, and Tielin Shi. A hybrid attention-based paralleled deep learning model for tool wear prediction. *Expert Systems with Applications*, 211:118548, 2023.

[5] Islam M El-Galy, Bassiouny I Saleh, and Mahmoud H Ahmed. Functionally graded materials classifications and development trends from industrial point of view. *SN Applied Sciences*, 1:1–23, 2019.

[6] Ignazio Gallo, Anwar Ur Rehman, Ramin Heidarian Dehkordi, Nicola Landro, Riccardo La Grassa, and Mirco Boschetti. Deep object detection of crop weeds: Performance of yolov7 on a real case dataset from uav images. *Remote Sensing*, 15(2):539, 2023.

[7] Emma Genders and Serestina Viriri. Plant diseases detection and classification using transfer learning. In *Pan-African Artificial Intelligence and Smart Systems Conference*, pages 150–166. Springer, 2021.

[8] Don R Hansen, Maryanne M Mowen, and Dan L Heitger. *Cost management*. Cengage Learning, 2021.

[9] Zichao Jiang. A novel crop weed recognition method based on transfer learning from vgg16 implemented by keras. In *IOP Conference Series: Materials Science and Engineering*, volume 677, page 032073. IOP Publishing, 2019.

[10] Yunlong Li, Shu Lin, Chang Liu, and Qingjie Kong. The defects detection in steel coil end face based on sced-net. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2022.

[11] Chinthakindi Balaram Murthy, Mohammad Farukh Hashmi, Neeraj Dhanraj Bokde, and Zong Woo Geem. Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—a comprehensive review. *Applied sciences*, 10(9):3280, 2020.

[12] Minwoo Park and Jongpil Jeong. Design and implementation of machine vision-based quality inspection system in mask manufacturing process. *Sustainability*, 14(10):6009, 2022.

[13] Mohammad Farhad Peerally, Susan Carr, Justin Waring, and Mary Dixon-Woods. The problem with root cause analysis. *BMJ quality & safety*, 26(5):417–422, 2017.

[14] D Shobha Rani, Lakshmi Ramani Burra, G Kalyani, B Rao, et al. Edge intelligence with light weight cnn model for surface defect detection in manufacturing industry. *Journal of Scientific & Industrial Research*, 82(02):178–184, 2023.

[15] Zeyu Sheng, Haiguang Chen, and Zifeng Qi. Cbam-based method in yolov7 for detecting defective vacuum glass tubes. In *Proceedings of the 2023 2nd Asia Conference on Algorithms, Computing and Machine Learning*, pages 413–418, 2023.

[16] Sushreeta Tripathy and Rishabh Singh. Convolutional neural network: an overview and application in image classification. In *Proceedings of Third International Conference on Sustainable Computing: SUSCOM 2021*, pages 145–153. Springer, 2022.

[17] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7464–7475, 2023.

[18] Shuai Wang, Xiaojun Xia, Lanqing Ye, and Binbin Yang. Automatic detection and classification of steel surface defect using deep convolutional neural networks. *Metals*, 11(3):388, 2021.

[19] Dingming Yang, Yanrong Cui, Zeyu Yu, and Hongqiang Yuan. Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence*, 35(15):1237–1249, 2021.

[20] Wei Zeng, Zhiyuan You, Mingyue Huang, Zelong Kong, Yikuan Yu, and Xinyi Le. Steel sheet defect detection based on deep learning method. In *2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 152–157. IEEE, 2019.

[21] Yuan Zhang, Youpeng Sun, Zheng Wang, and Ying Jiang. Yolov7-rar for urban vehicle detection. *Sensors*, 23(4):1801, 2023.