An Automatic Defect Detection System for Synthetic Shuttlecocks Using Transformer Model

CHING-SHENG LIN[ID][1] AND HAN-YI HSIEH[2]
[1]Master Program of Digital Innovation, Tunghai University, Taichung 40704, Taiwan
[2]Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10608, Taiwan

Corresponding author: Ching-Sheng Lin (cslin612@thu.edu.tw)

ABSTRACT With an estimation of 220 million people playing badminton on a regular basis, it was particularly popular in Asia but has growing popularity in different regions of the world. The demands of the relevant products, such as shuttlecocks and rackets, are also increasing in the sports industry. Synthetic shuttlecock, produced to offer similar experience and feel as feather shuttlecocks to players, is a more economical alternative to feather shuttlecocks. In addition to maintaining high throughput production for synthetic shuttlecocks with cost reduction, a more substantial improvement in quality control is desired as well. Since the defect detection of synthetic shuttlecocks is a challenging task, it heavily relies on human visual inspection at present. The existing manual quality-inspection process is not only error-prone but also considerably less efficient. In this paper, we propose an intelligent system to overcome these difficulties and bridge the gap between research and practice. Two cylinder grippers are designed to automatically deliver the shuttlecocks, a camera is used for capturing images and an end-to-end objection detection approach based on the Transformer model is investigated to recognize defects. Empirical results show that the proposed system obtains encouraging performance with AP_{50} value of 87.5% and outperforms other methods. Ablation studies demonstrate that our approach can considerably boost the detection performance of synthetic shuttlecocks. Moreover, the processing speed is much faster than human operators and suitable for industrial applications.

INDEX TERMS Synthetic shuttlecocks, defect detection, intelligent system, transformer model, cylinder gripper.

I. INTRODUCTION

The beginnings of badminton can be traced back to mid-19th century and the origin of the game was played with feather shuttlecocks [1]. However, it was until 1950s that the synthetic shuttlecocks were invented and badminton had a much wider appeal. With an estimation of 220 million people playing badminton on a regular basis, its increasing popularity is not confined to Asia but has been expanded to different areas. Badminton is one of the top ten most popular participation sports in the world [2] and made its Olympic debut as an official medal sport at the 1992 Summer Olympics. Although both feather and synthetic shuttlecocks contain sixteen leaves and one cork head (shown in Figure 1), feather shuttlecocks are made from goose or duck and the synthetic shuttlecocks comprise of plastic or nylon materials.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhigang Liu[ID].

Because feathers are more expensive and brittle, synthetic shuttlecock is increasingly demanding for its high durability and low cost.

There are two different types of synthetic shuttlecocks, one is single-piece injection-moulded (right-handed side of Figure 1) and another is two-part skirt design [3]. Since single-piece injection-moulded synthetic shuttlecock has been the mainstream design and dominated the market for the past 50 years, the discussion of this paper will focus on this type. During the manufacturing process, several types of damages and defects might happen to affect quality and decrease performance. Therefore, in addition to providing mass production to ensure the competition in the synthetic shuttlecock industry, there is an urgent requirement to develop new techniques with high quality and low cost. In the production of synthetic shuttlecocks, items require to be inspected for any defects before shipment to customers. Currently, many companies still heavily rely on operators or

quality assurance experts to perform the task. Manual visual inspection is inefficient and problematic which can't assure stable and accurate operation. The incapability of identifying the flaws could result in a significant loss to the company. Figure 2 displays four commonly seen defects in the synthetic shuttlecock which include short shots, redundant filaments, stains and air traps. These images present the following characteristics and pose the corresponding challenges [4]:

- Multiple defects in the same image: There are some defects, which may belong or not to the same defect type, existing in the same image with intersection or interference. These variations are challenging and may lead to erroneous recognition.
- Various size defects: It is rather difficult to identify an irregular defect pattern for the same type [5] and detect small-size defects on the curved surface of synthetic shuttlecocks.

With the progressive development of neural networks, Artificial Intelligence (AI) has been applied with great success in various fields and the manufacturing industry is no exception. Many companies have already used AI related techniques to overcome their existing limitations and bottlenecks for the purpose of fitting into the Industry 4.0 paradigm. However, to the best of our knowledge in the synthetic shuttlecock industry, there are not enough research works which focus on the defect detection. In this paper, we establish an automatic inspection system based on machine vision techniques for shuttlecock defect detection. An integrated equipment is designed to preform vision inspection and reduce manual workloads. Our end-to-end defect detection algorithm is based on the Transformer model [6]. Since the learning situation of this study is from multi-class imbalanced and small data sets, we also apply data augmentation techniques to improve the performance of our model in detecting defects.

Specifically, there are two main contributions in this paper:

- We devise an automatic defect detection system for synthetic shuttlecocks covering hardware design and software development. The hardware design ensures quick integration with existing industrial processes and the software development improves the effectiveness of defect detection. Companies, especially for small and medium-sized enterprises (SME), can greatly benefit from the development of this economical and technical design.
- We propose an end-to-end objection detection approach based on the Transformer model to simultaneously detect and classify the defects for synthetic shuttlecocks. The experimental study indicates that our model can achieve satisfying results and possess much higher efficiency in contrast to human inspection.

The remainder of this paper is organized as follows. Section 2 reviews several techniques and research methods related to the work of this paper. The proposed equipment design and network architecture are described in Section 3. We discuss the experiments and report the results in Section 4.
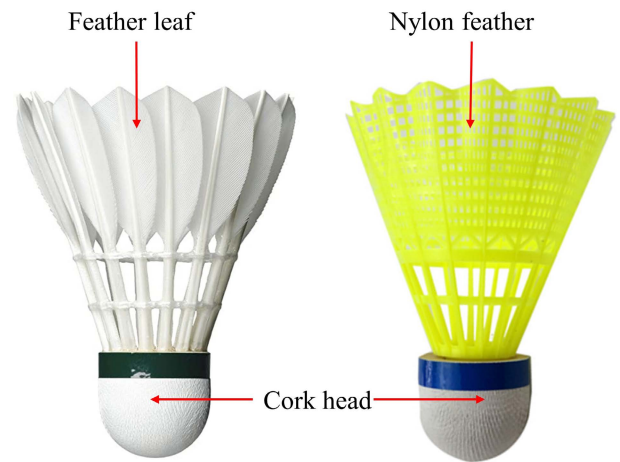

FIGURE 1. A feather shuttlecock (left) and synthetic shuttlecock (right).
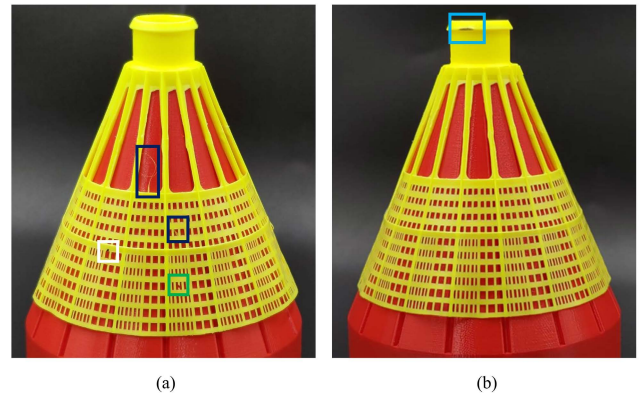

(a)   (b)

FIGURE 2. Four types of defects in synthetic shuttlecocks.
Figure (a) contains stains (white box), short shots (green box) and redundant filaments (black box). Figure (b) contains air traps (blue box).

In Section 5, we present conclusions and discuss future research directions.

II. RELATED WORK

In this section, we will review several works related to our research including deep learning based detectors and defect detection tasks.

A. DEEP LEARNING BASED DETECTORS

Object detection, one of the most important and challenging problems in computer vision, aims to detect object instances of predefined classes in images [7], [8]. Recently, deep learning based approaches have become the most popular solution in this research field and can be divided into one-stage detector and two-stage detector [9]. The YOLO series [10]–[13] and RetinaNet [14] are two well-known one-stage detectors. The representative work for two-stage detectors is the R-CNN (Regions with CNN) series including R-CNN [15], Fast R-CNN [16], Faster R-CNN [17] and Mask R-CNN [18]. One-stage detectors apply a single network to directly predict the object classification and position. On the other hand,

two-stage detectors first generate several candidate regions and then refine the classification and localization of the regions in the second step.

The YOLO network uses a single CNN to process input images and directly calculates the classification confidences and position coordinates of the object. With its end-to-end network structure, the detection speed can be greatly improved. RetinaNet is another one-stage object detector which contains two building blocks, feature pyramid networks (FPN) and focal loss. The use of FPN could improve multi-scale object predictions. The focal loss is designed to handle the class imbalance by reshaping the standard cross entropy loss.

R-CNN, a two-step object detection framework, uses an external selective search to generate candidate object regions which are then fed into individual CNN to extract fixed length feature vectors for the purpose of object classification and bounding box regression. Unlike R-CNN which feeds about 2k candidate object regions to a CNN for each image, Fast R-CNN only needs the original input image to generate Region of Interests (RoI) by selective search for speeding up the computation. After that, a ROI pooling layer and two full connection layers are appended to classify object category and fine-tune ROI position. To overcome the computational burden of selective search algorithm, Faster R-CNN introduces a novel region proposal network (RPN), which is a fully convolutional network (FCN), to extract candidate regions. Instead of using only bounding boxes, Mask R-CNN adds an additional mask branch based on Faster R-CNN to locate exact pixels of each object instance. Therefore, the model is able to provide three outputs which include a class label, a bounding box and a mask for each candidate region.

B. VISION BASED DEFECT DETECTION

Defect detection plays an essential role to ensure product quality in a broad range of manufacturing industries, such as civil, energy and plant. The defects on the product not only can affect the appearance but also could cause safety issues. Traditionally, defect detections are conducted by human experts but manual visual inspection is usually error-prone and costly. With the rapid development of computer vision and machine learning techniques, nowadays vision based defect detection has attracted more attention and gradually been adopted by industries.

To address the surface-anomaly detection problem [19], a two-stage architecture is proposed to learn from a small number of defected training data. The first stage called the segmentation network is used to perform at the level of individual image pixels. The second staged is a decision network which uses the output from the segmentation network as the input to learn the probability of anomaly presence in the image. Pavement crack detection is a critical task due to the complicated pavement conditions and has been studied for decades. A CNN-based architecture is modelled as a multi-label classification problem to predict the crack at the pixel level [20]. Moreover, a novel strategy to modify the ratio of positive-to-negative training data is also proposed to address the severely imbalanced problem.

Since the demand for wind power has considerably increased due to environmental concerns, the quality assurance of wind turbine blade (WTB) has become an imperative issue as well. A YOLO-based small object detection approach (YSODA) supports the multiscale feature pyramid to inspect WTB defects by amalgamating features in the layers of CNN [21]. The resulting detection accuracy of YSODA reaches 91.3% which outperforms YOLO model (88.7%). However, the detection speed is 24 fps which is slower than YOLO (30 fps) due to the additional feature extraction. To locate four different types of damages on WTB, the faster R-CNN algorithm with Inception-ResNet-v2 architecture is proposed and achieves 81.10% mean average precision [22]. The detection speed is 2.11 seconds and it is faster than human-based analysis which requires 20 seconds to 3 minutes depending on the difficulties.

As a valuable and rare natural resource, the identification of wood defects can reduce the waste of wood materials and improve the automated processing in the wood industry. An improved SSD algorithm with a DenseNet backbone is proposed to detect three types of wood defects including live knots, dead knots and checking [23]. The transfer learning method is applied on the ImageNet data set to address the labelled data scarcity issue and the mean average precision value is about 96.1%. In addition, Faster R-CNN with the ResNet pre-training model is used to find wood panel surface defects and achieves an average accuracy of 80.6% on a synthetically augmented dataset [24].

Local binary pattern (LBP) is one of the most powerful local feature extraction methods by estimating the local contrast of an image between pixels [25]. An improved version of LBP features is employed for porosity detection on stone images [26]. To handle the color images, a multi-resolution and noise-resistant of LBP is introduced to extract color/texture features and identify surface defects [27]. Completed Local Quartet Patterns (CLQP) operator, which is rotation invariant and gray-scale invariant, extracts fabric image local texture features for localizing surface defects with 97.66% detection rate [28].

Since defect data set is inherently difficult to obtain, data augmentation of training samples is also an important research topic in this area. A generative adversarial network (GAN) is proposed to exaggerate the small defects within the images and also expand the defected samples [29]. Another GAN-based approach, cycle GAN, takes pairs of defect images to exchange their colors and textures to generate new defective data without changing the distribution of color and grain in the dataset [30].

III. AUTOMATIC DEFECT DETECTION SYSTEM FOR SYNTHETIC SHUTTLECOCKS

The framework of the proposed intelligent system for shuttlecocks inspection is shown in Figure 3. During the training stage, the images are collected by the camera device followed
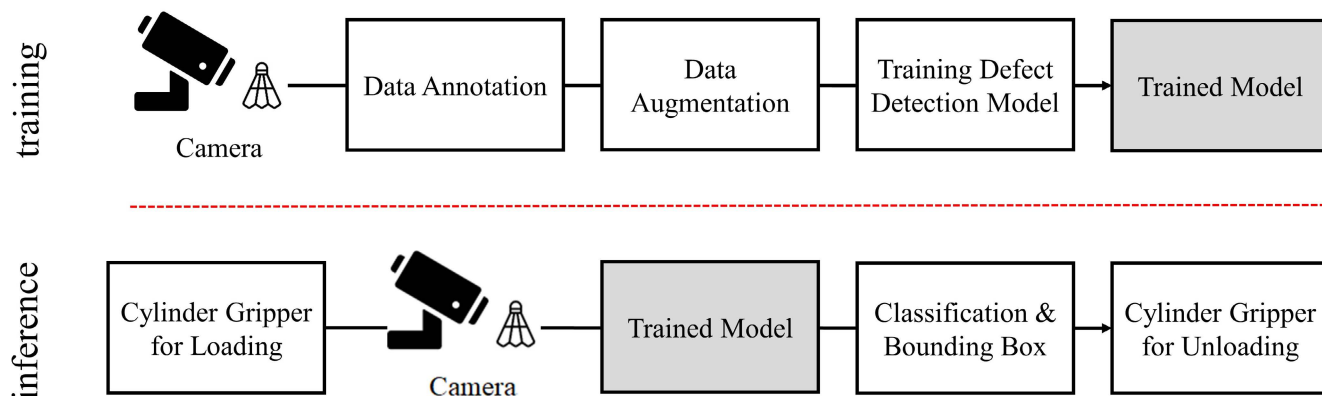
FIGURE 3. The framework of our proposed automatic defect detection system for synthetic shuttlecocks where the top is the training process while the bottom represents the inference stage.
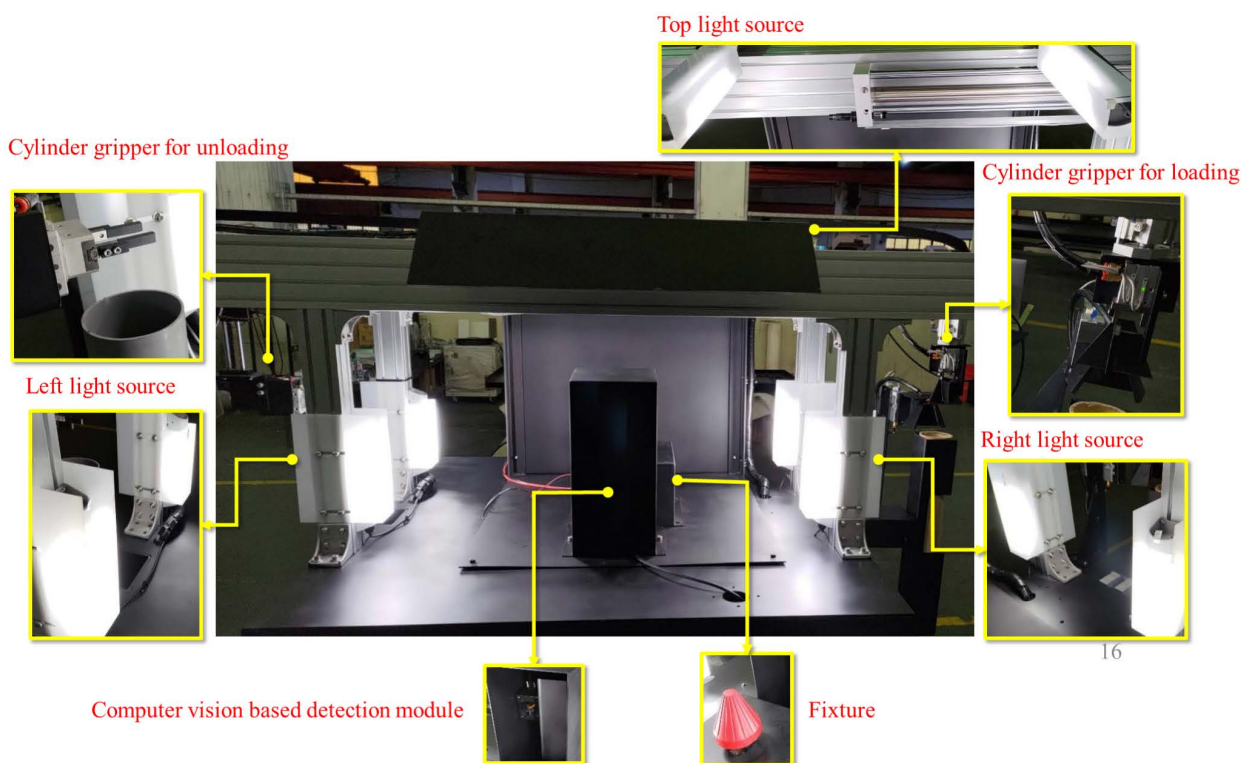

FIGURE 4. The experimental equipment layout of the shuttlecock defect detection system and close-up images of main components.

by human annotation. In order to make our model more robust and prevent overfitting, we adopt data augmentation to obtain more training data without extra labor cost. An end-to-end objection detection approach based on the Transformer model with multi-layer representations is proposed to detect the defects of shuttlecocks effectively. In the inference phase, the on-line inspection is conducted on our designed experimental equipment which is integrated with the trained model.

A. EXPERIMENTAL EQUIPMENT LAYOUT

In order to address the shuttlecocks inspection, we develop a machine vision system to automate the whole process as

shown in Figure 4. The system mainly consists of a fixture, two cylinder grippers, three lighting sources and a computer vision based detection module. The fixture which is made of Polyoxymethylene is used to hold the shuttlecock with rotating capability. Within the computer vision based detection module, there is an IDS LE AF camera with M12 liquid lens installed to capture images and a defect detection model to identify anomalous conditions. To ensure a full coverage of 360 degrees, the fixture will automatically rotate 45 degrees for eight times in every inspection and the camera will take eight images per shuttlecock accordingly. A stepper motor (17PM-K405-P3VS) is used to rotate the fixture after an
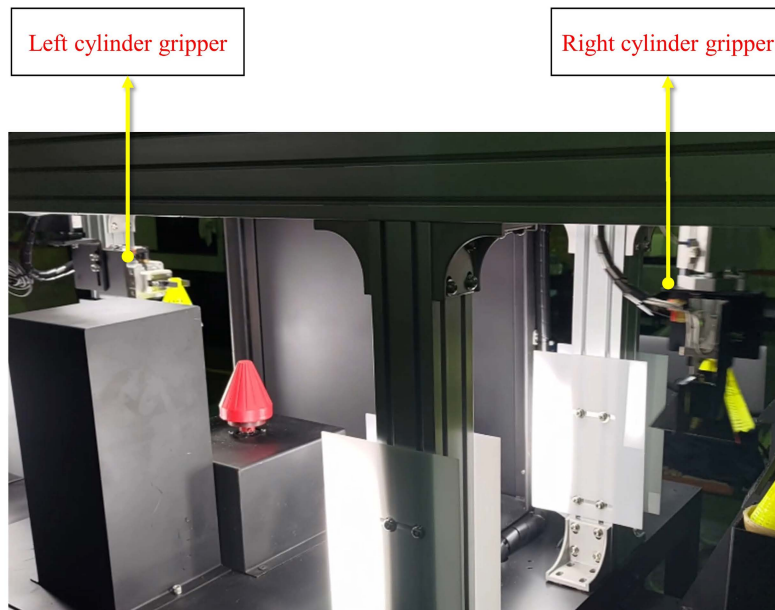
FIGURE 5. The Left cylinder gripper removes the shuttlecock from the fixture after inspection and the right cylinder gripper moves the next shuttlecock to be inspected.

image has been inspected by our defect detection model. It usually takes 0.4 seconds for one rotation. Two cylinder grippers are used to deliver the shuttlecock where the right gripper takes the shuttlecock to the texture for inspection and the left one will send the shuttlecock to the designated position based on the examined result (Figure 5). The distance between right (left) cylinder gripper and the fixture is about 45 (60) cm. Due to the light conditions inside the equipment layout, it is necessary to have carefully-placed lights to ensure the quality for camera shooting. We setup three LED lights (top, right and left lights) to illuminate the shuttlecock.

B. DATA ACQUISITION
The defects of the shuttlecock during production can appear in different forms. Some of them are due to the quality of raw materials, while others are caused by machine malfunctions. There are four types of defects:

1. Short shots: The main reason of short shot defect is that molten plastic does not completely fill the cavity. A remedy to avoid short shots is to adjust the mold temperature.
2. Redundant filaments: If the injection pressure is too high, melt plastics may overflow and flash defect will occur. As a result, the defective items either will be discarded or require additional process.
3. Stains: Material is a critical cause of defects in injection moulding process. When material is contaminated with foreign particles, it could induce black-specks and affect the quality of finished product.
4. Air traps: This defect is caused by lack of vents and can result in incomplete filling and packing. The issue is usually resolved by adding air vents.

In order to train the proposed neural network to accurately detect the shuttlecock defects by vision based approaches, it is important to acquire images from the production line. Once the images have been collected, they will be labelled to form the ground truth by experts from the shuttlecock industry. We use LabelMe software [31] to annotate images for object detection task and convert the generated json file to standard COCO format [32].

Since the training data from the shop floor is often highly scarce especially in defect detection field, insufficient labelled data can't generalize well and may lead to overfitting problems. Data augmentation is a technique to alleviate limited training dataset issue without collecting new data [33]. In addition, because the collection and labelling of defect samples in shuttlecock manufacturing industries can be tedious and time-consuming, data augmentation provides an effective approach to diversifying the data distribution. There are various basic image augmentation techniques such as brightness adjustment, random cropping, flipping, scaling, rotation and adding noises. Since the shuttlecock is put on the fixture, we do not consider those augmentation techniques with deformation strategy. Therefore, we propose the use of horizontal flipping and adding noise. Flipping is one of the most common augmentation approach and has been proven useful on popular datasets such as CIFAR-10 and ImageNet. To reduce the network's tendency of learning high frequency features, adding Gaussian noise essentially could have data points in all frequencies to distort the high frequency features. It is worthwhile to note that adding Gaussian noise will not change the labelled results but flipping technique modifies the label by mirroring the image along the central axis. The augmented data are expected to represent a more comprehensive

dataset which minimizes the difference between the training data and new observations.

C. TRANSFORMER-BASED DEFECT DETECTION BY MULTI-LAYER REPRESENTATIONS

Driven by the development of deep neural networks, the performance of object detection has been significantly improved and used in many applications. In this paper, we formulate the defect detection of shuttlecocks as an object detection problem by directly predicting a set of bounding boxes and the corresponding defect labels. DETR [34], a simple end-to-end architecture with CNNs and Transformer, is proposed to solve the object detection task. Motivated by DETR, our proposed model is built upon it by following a similar practice and further exploring the multi-layer representations of Transformer decoders to enhance the performance. In this section, we first introduce the architecture of the proposed method followed by the optimization objective.

Our network architecture to address the defect detection of shuttlecocks is depicted in Figure 6. It mainly contains three components: a CNN backbone to extract feature representations of the input image, a transformer framework to model one-to-one set prediction utilizing all layers of decoder representations and feed forward networks (FFNs) to predict the bounding boxes and class labels.

1) BACKBONE

A conventional CNN backbone is used to extract the original pixel-level feature sequence for a given input image $x \epsilon R^{3 \times H0 \times W0}$ where H_0 is height, W_0 is width and there are three channels. The typical lower-resolution feature map $f \epsilon R^{C \times H \times W}$ is set C = 2048, H = H_0/32 and W = W_0/32. In this paper, convolutional layers from ResNet-50 model [35] are taken as the backbone. It can be divided into five stages. The stage one, which starts with a convolutional layer followed by a batch normalization and activation function, is the input stage to compute the initial feature maps. The second to fifth stages contain a set of convolution blocks and identity blocks where there are three convolution layers in each block.

2) TRANSFORMER

The Transformer model has become the main architecture for many natural language processing tasks and has recently been adapted to computer vision tasks. The self-attention mechanism of the transformer is capable of modelling the interactions on all pairwise elements and is able to eliminate duplicate predictions for set prediction. It is composed of an N-layer encoder and decoder. The encoder first applies a 1×1 convolution to convert the dimension of the feature map from C to a smaller dimension d, resulting a new feature map $z_0 \epsilon R^{d \times H \times W}$. Since the encoder requires a sequence as input, the feature map z_0 is flatten and embedded to form a one-dimensional sequence $f_e \epsilon R^{d \times HW}$ with embeddings of size d. Then f_e with positional encodings is fed into encoder layers to get encoded sequence features as the encoder output.

Afterward, with the multi-head attention mechanism in the decoder, N decoder output embeddings are generated from encoder outputs and N learned object queries. The output embeddings will be supplied to the next FFN layer to yield predictions.

3) FFN LAYER

To make the final detection prediction after the decoder layer, the last component is the FFN layer. The multi-layer perceptron is the regression branch to predict center coordinates, height and width of the bounding box. The linear projection layer is the classification branch which applies a softmax function to predict the class labels. Additionally, there is a special class label ø used to indicate that no target class is detected within the box. Existing DETR only takes the last layer output of the decoder to perform object detection. Since only relying on the final layer representation could lead to information loss, we explore the potential of adopting multi-layer representations of decoder outputs for both bounding box regression and class prediction. Given the output representation of each decoder layer $\{h_i\}_{i=1}^N$, we concatenate all h_i to form a representation h_c for class prediction and h_b for bounding box regression. Subsequently, h_c is sent to the Class FFN and h_b is fed to Bounding Box FFN to finish the final detection work.

There are two steps in optimization objective where the first step is to produce an optimal bipartite matching and the second step is to minimize the loss between the matched pairs obtained from the first step. Given a fixed-size set of N predictions, we perform a bipartite matching to align the system predictions $\hat{y} = \{\hat{y}_i\}_{i=1}^N$ to ground-truth labels y. The best permutation of N elements between $\hat{y}$ and y is $\hat{\sigma}$ which minimizes the matching cost defined below:

$$\hat{\sigma} = \underset{\sigma \in \partial_N}{\text{argmin}} \sum_{i=1}^N L_{\text{match}} \left(y_i, \hat{y}_{\sigma(i)} \right) \tag{1}$$

where $L_{\text{match}} \left(y_i, \hat{y}_{\sigma(i)} \right)$ is a matching cost between the ground-truth y_i and the predicted result at index $\sigma(i)$. Since the objection detection task needs to consider both the class and bounding box prediction, the L_{match} is defined as

$$L_{\text{match}} \left(y_i, \hat{y}_{\sigma(i)} \right) = -\mathbb{1}_{\{c_i \neq \emptyset\}} \hat{p}_{\sigma(i)} \left(c_i \right) + \mathbb{1}_{\{c_i \neq \emptyset\}} L_{\text{box}} \left(b_i, \hat{b}_{\sigma(i)} \right) \tag{2}$$

where c_i is the target class label and b_i is the bounding box of the ground truth. Accordingly, $\hat{b}_{\sigma(i)}$ is the predicted box and the prediction probability of class c_i is $\hat{p}_{\sigma(i)}(c_i)$. The $\mathbb{1}_{\{c_i \neq \emptyset\}}$ is an indicator function where the value will be 1 if $\{c_i \neq \emptyset\}$ is true and 0 otherwise. Hungarian algorithm could be employed here to find the one-to-one matching [36]. Note that bounding box loss L_{box} is defined as the linear combination of the generalized IoU loss L_{iou} [37] and the L_1 loss. The introduction of L_{iou} is used to mitigate the relative scaling issue occurred in L_1 loss.

$$L_{\text{box}} \left(b_i, \hat{b}_{\sigma(i)} \right) = \lambda_{iou} L_{\text{iou}} \left(b_i, \hat{b}_{\sigma(i)} \right) + \lambda_{L_1} \left\| b_i - \hat{b}_{\sigma(i)} \right\|_1 \tag{3}$$
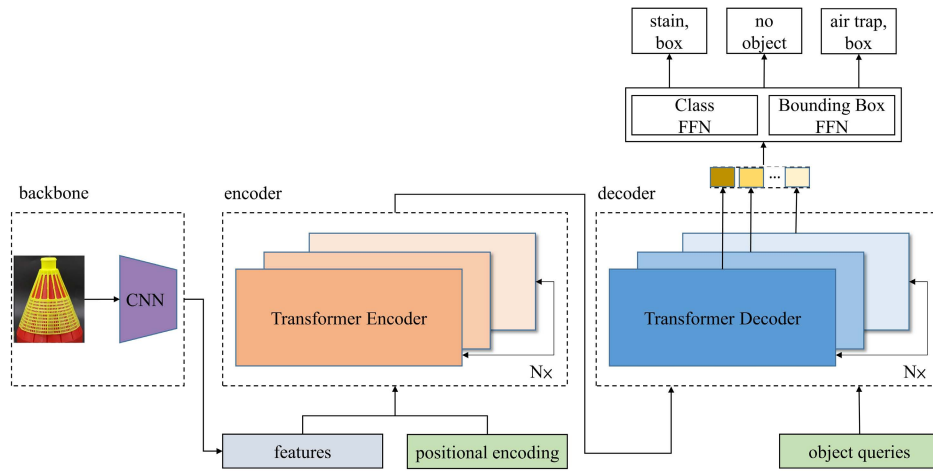
FIGURE 6. The network architecture.

TABLE 1. Distribution of defects in training / validation / testing datasets.

Defect Type	Training		Validation	Testing
	Original	Augment		
Short shots	387	774	127	129
Redundant filaments	472	944	157	154
Stains	169	338	52	48
Air traps	134	268	44	44

After establishing the optimal assignment $\hat{\sigma}$ of bipartite matching, the second step is to calculate the loss function (Hungarian loss) for all matched pairs consisting of class probability loss and bounding box loss:

$$L_{\text{Hungarian}}\left(y, \hat{y}\right) = \sum_{i=1}^{N} -\log(\hat{p}_{\hat{\sigma}(i)}(c_i))$$
$$+ \mathbb{1}_{\{c_i \neq \varnothing\}} L_{\text{box}}\left(b_i, \hat{b}_{\hat{\sigma}(i)}\right) \quad (4)$$

IV. EXPERIMENT AND RESULTS

In this section, we perform the evaluation of the proposed method to detect the shuttlecock defects including: (1) dataset construction; (2) evaluation metrics; (3) comparison with other algorithms; and (4) ablation experiments of using data augmentation and multi-layer representations.

A. DATASET

The total number of raw images is 857 and the dataset is labelled by domain experts. It usually takes 5-10 seconds to annotate a defect depending on the difficulty level. Each image has one or more defects with the resolution of 3240 × 1833. The dataset consists of four types of defects,
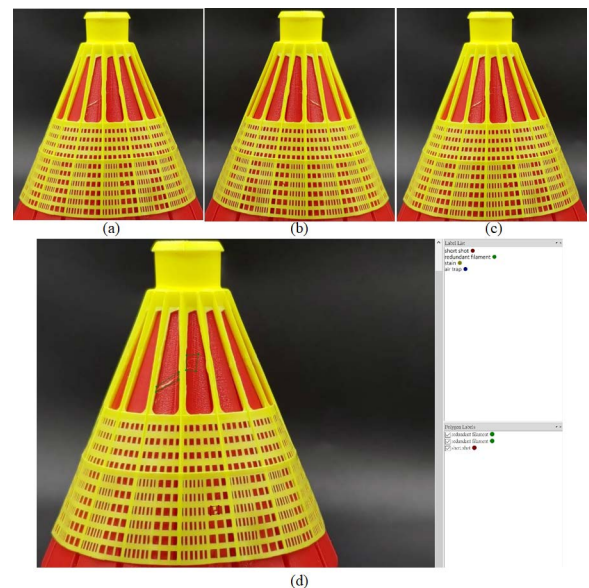

FIGURE 7. (a) The example image; (b) the horizontal flipping image; (c) the image with Gaussian noise; (d) the image labelled by LabelMe.

short shots, redundant filaments, stains and air traps. The images are divided into 512 training images, 172 validation images and 173 testing images, respectively. In Table 1, we display the statistics of each defect type. In order to have more number of training data to avoid overfitting, horizontal flipping and adding noise are employed to increase the number of training images from 512 to 1536. The quantity of augmented defect types is also added to Table 1. In Figure 7, we display an example image with data augmentation results and its annotation by LabelMe.

B. EVALUATION METRIC

Intersection over Union (IoU) is the intersection area of a predicted bounding box (B_p) and a ground-truth box (B_{gt})

divided by the union area.

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \qquad (5)$$

A detection is considered to be correct if the IoU is greater than a predefined threshold. To evaluate the proposed approach, we measure the performance by the evaluation matrix of COCO with different IoU threshold settings. Precision and recall are two well-known metrics to evaluate correctness and effectiveness. Recall measures how efficient the proposed approach is for retrieving correct regions, while precision indicates how many predicted regions are correct. However, since a single point value of recall or precision is not good enough to measure the quality [38], average precision (AP) is proposed to calculate with precision results on different values of recall given by a specific IoU threshold and is defined as

$$AP_{th} = \int_0^1 P(R)dR, \text{ where th is the threshold} \qquad (6)$$

Moreover, to have a more complete evaluation, we also use mean average precision (mAP) which computes the average AP across different IoU thresholds from 0.5 to 0.95 with an interval of 0.05. AP_{th} and mAP are two indicators in our empirical studies.

C. EXPERIMENTAL RESULTS

To validate the proposed approach for the defect detection of shuttlecocks, we make comparison with other detection approaches including one-stage detectors (RetinaNet and DETR) and two-stage detectors (Faster R-CNN and Mask R-CNN). The pre-trained ResNet-50 is employed as our backbone network and the hyper-parameter setting of the model is shown in Table 2. For the other compared algorithms, we also follow similar hyper-parameter settings. To carry out the comparison experiments, we adopt the released codes of DETR [39] and the Detectron2 API [40] is utilized for Faster R-CNN, Mask R-CNN and RetinaNet. The experiments are conducted on the Windows system with an 11th Gen Intel(R) Core(TM) i9-11900K @ 3.50GHz, 128 GB RAM, and NVIDIA GeForce RTX 3090 GPU 24G.

Table 3 shows the performance comparisons of AP for all defect classes with IoU threshold set to be 0.5 which is a typical value to use in detection work. The results show that the proposed method achieves the highest AP in all classes except for short shots where our model is the second-best (84.5%). Compared with the second best competitor, our method increases the detection of redundant filaments by nearly 3.3% and stains by 1.6%. In the last column of Table 3, we also demonstrate the inference speed comparison in terms of frames per second (FPS) for each approach. Our approach can reach 7.336 FPS and performs better than most other methods. We observe that our method only introduces moderate additional time costs compared with DETR. In general, it usually takes 4 to 5 seconds for an operator to finish

a shuttlecock inspection. With the accurate detection results and a much faster processing speed than human experts, our approach is sufficient for industrial applications.

To have a more complete comparison, we also list the comparison in terms of AP_{50}, AP_{75} and mAP to show the feasibility of our approach shown in Table 4. On the primary metric of interest in this research (AP_{50}), we push the score to 87.5%. On the strict metrics (AP_{75} and mAP), our performance is almost on par with other approaches. Compared with DTER, we increase the AP_{50} score by 1.9% (from 85.6% to 87.5%), the AP_{75} score by 2.5% (from 63.0% to 65.5%) and the mAP score by 3.1% (from 59.5% to 62.6%).

We analyze the prediction results for our approach and have some observations. First, air trap defect is almost solved because of its frequent occurrence in the specific area which is on top of the shuttlecock. Second, several false positives are caused by the effects of the fixture. The flashes may be left on the surface of the fixture occasionally and our model may falsely recognize these flashes as the redundant filament defect type (Figure 8(a)). Avoiding this mistake requires further investigation of materials which are less likely to leave flashes on the fixture surface or needs additional procedure before uploading the shuttlecock to the fixture each time. Note that the current fixture is made of Polyoxymethylene. Last, there is some misclassification between the stain defect and short shot defect (Figure 8(b)). The major reason is that the black-speck position is ambiguous and our model wrongly predicts this situation as the short shot defect.

TABLE 2. Hyper-parameter configuration.

| Parameter	Value
Batch size	8
Dropout rate	0.1
Learning rate	0.0001
Backbone learning rate	0.00001
weight decay	0.0001

D. ABLATION STUDIES

Finally, to further verify the value of the proposed modules, we discuss the impact of multi-layer representations and data augmentation. Three types of experiments are carried out by excluding: (1) the data augmentation (DA); (2) the multi-layer representations of classes (MLR-C); (3) the multi-layer representations of bounding boxes (MLR-BB). The results are shown in Table 5 and we find that all proposed components have contributed to the task. More specifically, without applying augmentation techniques, the AP_{50} score loses 3.4% (from 87.5% to 84.1%) and mAP drops 1.9% (from 62.6% to 60.7%). When removing the multi-layer representations

TABLE 3. The performance comparison of AP with IoU = 0.5 for all defect types and inference speed comparison listed in the last column.

model	Short shots	Redundant filaments	Stains	Air traps	FPS
Faster R-CNN	78.1%	70.3%	65.2%	100%	5.870
Mask R-CNN	77.2%	71.1%	68.0%	100%	5.217
RetinaNet	**85.9%**	73.4%	56.4%	100%	6.407
DETR	81.7%	78.9%	81.8%	100%	**7.402**
Our method	84.5%	**82.2%**	**83.4%**	**100%**	7.336

TABLE 4. The performance comparison in terms of AP_{50}, AP_{75} and mAP.

model	AP_{50}	AP_{75}	mAP
Faster R-CNN	78.4%	65.2%	60.4%
Mask R-CNN	79.1%	66.1%	**63.1%**
RetinaNet	78.9%	**66.5%**	62.3%
DETR	85.6%	63.0%	59.5%
Our method	**87.5%**	65.5%	62.6%
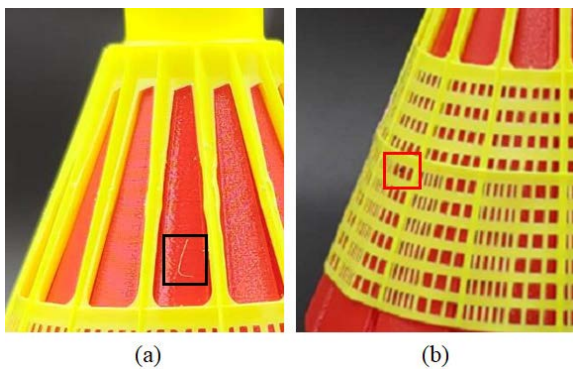
FIGURE 8. (a) A false positive example due to the fixture; (b) a misclassification between the stain defect and short shot defect.

TABLE 5. Ablation experimental results in terms of AP_{50}, AP_{75} and mAP.

model	AP_{50}	AP_{75}	mAP
Our method	**87.5%**	**65.5%**	**62.6%**
W/o DA	84.1%	64.6%	60.7%
W/o MLR-C	85.0%	64.7%	61.8%
W/o MLR-BB	86.5%	63.6%	61.1%

of bounding boxes, the AP_{75} score decreases 1.9% (from 65.5% to 63.6%). In summary, from these ablation studies, the proposed method is effective on the defect detection of shuttlecocks.
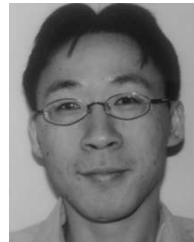
V. CONCLUSION

The end customers in the sports industry have high expectations for the quality of shuttlecocks. Nowadays, the defect inspection is still conducted by human operators and there are not enough research works in this field. To this end, in this study, we present a hardware design and vision based approach to automatically detect defects of shuttlecocks for industrial application. Two cylinder grippers are used to deliver shuttlecocks for expediting the production process without human intervention. An end-to-end network based on Transformer model is adopted to perform defect detection. Empirical study shows that the proposed method achieves promising results with AP_{50} value of 87.5% and is superior to other state-of-the-art techniques. Ablation studies show that the data augmentation, the multi-layer representations of classes and bounding boxes all contribute to increase the performance of defect detection. With the introducing of multi-layer representations, our method only brings little additional time costs and is better than most other existing models.

In the future work, we intend to extend the study further in several directions. First, to reduce the false positive alarms caused by the fixture, we will explore the material which is less prone to leave stains and filaments on the fixture. Second, to reduce the misclassification between the stain defect and others, we will investigate the use of high-resolution feature maps to increase the performance. Third, although the current platform is already feasible for real-time application, it is still important to compress the model, reduce the parameters, and further accelerate the computation. Last, it would be interesting to apply our detection network to other surface inspection tasks to validate the robustness.

REFERENCES

[1] A. Cooke, "Computer simulation of shuttlecock trajectories," *Sports Eng.*, vol. 5, no. 2, pp. 93–105, May 2002.

[2] R. Vandekerkhove, I. Moons, and E. Du Bois, "Introducing repair in sports' consumables: Investigation of repairability of badminton shuttles," *J. Cleaner Prod.*, vol. 326, Dec. 2021, Art. no. 129229.

[3] T.-M.-T. Woo and F. Alam, "Comparative aerodynamics of synthetic badminton shuttlecocks," *Sports Eng.*, vol. 21, no. 1, pp. 21–29, Mar. 2018.

[4] B. Wei, K. Hao, and L. Gao, "MLMA-Net: Multi-level multi-attentional learning for multi-label object detection in textile defect images," 2021, *arXiv:2102.00376*.

[5] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 196–210, 2015.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[7] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019, *arXiv:1905.05055*.

[8] C. Xu, C. Shi, H. Bi, C. Liu, Y. Yuan, H. Guo, and Y. Chen, "A page object detection method based on mask R-CNN," *IEEE Access*, vol. 9, pp. 143448–143457, 2021.

[9] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sens.*, vol. 13, no. 1, p. 89, Dec. 2020.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

The image shows two black-and-white portrait photographs of the authors.

[11] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[16] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[19] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, 2019.

[20] Z. Fan, Y. Wu, J. Lu, and W. Li, "Automatic pavement crack detection based on structured prediction with the convolutional neural network," 2018, *arXiv:1802.02208*.

[21] Z. Qiu, S. Wang, Z. Zeng, and D. Yu, "Automatic visual defects inspection of wind turbine blades via YOLO-based small object detection approach," *J. Electron. Imag.*, vol. 28, no. 4, p. 1, Aug. 2019.

[22] A. Shihavuddin, X. Chen, V. Fedorov, A. N. Christensen, N. A. B. Riis, K. Branner, A. B. Dahl, and R. R. Paulsen, "Wind turbine surface damage detection by deep learning aided drone inspection analysis," *Energies*, vol. 12, no. 4, p. 676, Feb. 2019.

[23] F. Ding, Z. Zhuang, Y. Liu, D. Jiang, X. Yan, and Z. Wang, "Detecting defects on solid wood panels based on an improved SSD algorithm," *Sensors*, vol. 20, no. 18, p. 5315, Sep. 2020.

[24] A. Urbonas, V. Raudonis, R. Maskeliūnas, and R. Damaševičius, "Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning," *Appl. Sci.*, vol. 9, no. 22, p. 4898, Nov. 2019.

[25] M. Topi, P. Matti, and O. Timo, "Texture classification by multi-predicate local binary pattern operators," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Sep. 2000, pp. 939–942.

[26] F. Tajeripour and S. H. Fekri-Ershad, "Porosity detection by using improved local binary pattern," in *Proc. 11th WSEAS Int. Conf. Signal Process., Robot. Automat. (ISPRA)*, vol. 1, Feb. 2012, pp. 116–121.

[27] S. Fekri-Ershad and F. Tajeripour, "Multi-resolution and noise-resistant surface defect detection approach using new version of local binary patterns," *Appl. Artif. Intell.*, vol. 31, nos. 5–6, pp. 395–410, Jul. 2017.

[28] Z. Pourkaramdel, S. Fekri-Ershad, and L. Nanni, "Fabric defect detection based on completed local quartet patterns and majority decision algorithm," *Expert Syst. Appl.*, vol. 198, Jul. 2022, Art. no. 116827.

[29] J. Lian, W. Jia, M. Zareapoor, Y. Zheng, R. Luo, D. K. Jain, and N. Kumar, "Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1343–1351, Feb. 2020.

[30] D. Li, W. Xie, B. Wang, W. Zhong, and H. Wang, "Data augmentation and layered deformable mask R-CNN-based detection of wood defects," *IEEE Access*, vol. 9, pp. 108162–108174, 2021.

[31] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, May 2008.

[32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Sep. 2014, pp. 740–755.

[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Aug. 2020, pp. 213–229.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[36] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.

[37] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

[38] X. Yang, A. Chen, G. Zhou, J. Wang, W. Chen, Y. Gao, and R. Jiang, "Instance segmentation and classification method for plant leaf images based on ISC-MRCNN and APS-DCCNN," *IEEE Access*, vol. 8, pp. 151555–151573, 2020.

[39] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. (2020). *DETR*. [Online]. Available: https://github.com/facebookresearch/detr

[40] Y. Wu, A. Kirillov, F. Massa, W. Y. Lo, and R. Girshick. (2019). *Detectron2*. [Online]. Available: https://github.com/facebookresearch/detectron2

CHING-SHENG LIN received the B.S. and M.S. degrees in applied mathematics from the National Chung Hsing University, Taiwan, in 1998 and 2001, respectively, and the Ph.D. degree in computer science from the State University of New York at Albany, in 2016.

He was a Postdoctoral Research Fellow with Tunghai University, Taiwan, in 2018, where he is currently an Assistant Professor with the Master Program of Digital Innovation. His research interests include natural language processing, artificial intelligence, and machine learning.

HAN-YI HSIEH received the B.E. and M.S. degrees in computer science from Tunghai University, Taichung, Taiwan, in 2013 and 2017, respectively. He is currently pursuing the Ph.D. degree in computer science with the National Taipei University of Technology, Taipei, Taiwan.

Since 2018, he has been a Research Assistant with the Software Engineering and Technology Center, Tunghai University. His research interests include the machine learning, digital image processing, and digital signal processing.

● ● ●