An AR-Assisted Deep Learning-Based Approach for Automatic Inspection of Aviation Connectors

Shufei Li, Pai Zheng, *Member*, *IEEE*, and Lianyu Zheng*

***Abstract*— The mismatched-pins inspection of complex aviation connector is a critical process to ensure the correct wiring harness assembly, of which the existing manual operation is error-prone and time-consuming. Aiming to fill this gap, this work proposes an augmented reality (AR)-assisted deep learning-based approach to tackle three major challenges in aviation connector inspection, including small pins detection, multi-pins sequencing, and mismatched pins visualization. Firstly, the proposed spatial-attention pyramid network approach extracts image features in multilayers and searches for their spatial relationships among images. Secondly, based on the cluster-generation sequencing algorithm, these detected pins are clustered into annuluses of expected layers and numbered according to their polar angles. Lastly, AR glass as the inspection visualization platform, highlights mismatched pins in the augmented interface to warn operators automatically. Compared with other existing methodologies, the experimental result shows that the proposed approach can achieve better performance accuracy, and support operator's inspection process efficiently.**

***Index Terms*—Industrial inspection, deep learning, augmented reality, spatial-attention pyramid network, aviation connector**

I. Introduction

With digitalization services permeating manufacturing industries, the modern industries are characterized by increased automatization and reconfigurability [1]. In this context, the digital innovation enables the flexible planning and managing of production process cost-effectively, and hence is capable of meeting customized requirements, achieving improved quality and reduced time to market [2].

Industrial inspection has always been playing a critical role to enable the final production quality of the aviation industry, due to its complex structure design and stringent quality requirements. Nevertheless, the conventional manual operation is error-prone and time-consuming, where introducing novel digital approaches has attracted many concerns during the past decade. For instances, Boeing introduced a fault isolation system [3] to train operators to quickly learn troubleshooting techniques of the engine air bleed system. In the avionics industry, a new robotic endoscope [4] was introduced to automatically inspect different residuals (e.g., sand and metallic dust) inside the oil ducts of the metallic gearbox. Despite these achievements, however, due to the complex scenarios in aviation, there still lies several challenges to realize the automatic industrial inspection.

One major challenge is that existing industrial inspection systems, such as industrial robots and specialized vision equipment [5], fail to present good flexibility and adaptability in those narrow spaces of aerospace products. To address this, mobile and wearable augmented reality (AR) systems allow co-inspection [6] between remotely connected customers and experts inside the complex structures of products. With the characteristic of mobility and intuitionistic visualization, the development of AR-assisted environment provides the possibility of automatic inspection at manufacturing sites.

Nevertheless, since the AR-based implementation collects vision information suffering from complex onsite background noises, most existing visual inspection approaches cannot guarantee effective performance of detection accuracy. Conventional methods, such as template matching [7] and filter-based feature selection [8], perform well on specialized vision-inspection platforms [9] with stable environmental conditions, but shows poor performance on dynamic industrial scenarios. To overcome this challenge, a recent work of deep learning-based inspection have shown unique advantages to identify product status and quality in a complex manufacturing environment [10]. In this context, the powerful deep neural networks can achieve reliable results for industrial inspection of defects from visual information with heavily textured backgrounds [11]. Despite its good performance, to the authors' knowledge, few works leverage deep learning technologies and AR-based tools together to conduct the autonomous industrial inspection for components inside aerospace products with narrow spaces and complex structures. Most existing works of industrial inspection only look for the start-of-art performance accuracy by introducing high-precision visual sensors and optical systems, which fail to provide flexibility and adaptability services for inspection tasks that may be dynamically changed.

Aiming to fill this research gap, this paper proposes a deep learning-based inspection approach with AR-assisted

This research is funded by the Civil Airplane Technology Development Program (MJ-2017-G-70), the Beijing Key Laboratory of Digital Design and Manufacturing Project, and the Departmental General Research Fund (G-UAHH) from the Hong Kong Polytechnic University, Hong Kong (*Corresponding author: Lianyu Zheng).

Mr. Shufei Li was with the Department of Industrial and Manufacturing Systems Engineering, Beihang University, Beijing, China, 100191 (e-mail: lsfshufei@buaa.edu.cn).

Dr. Pai Zheng is with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China, 999077 (e-mail: pai.zheng@polyu.edu.hk).

Prof. Lianyu Zheng is with the Department of Industrial and Manufacturing Systems Engineering, Beihang University, Beijing, China, 100191 (email: lyzheng@buaa.edu.cn).

visualization platform to automatically detect the mismatched pins of aviation connector in wiring harness (ACWH). The rest of this paper is organized as follows. Section II introduces related works of AR-assisted and deep learning-based technologies in manufacturing. Section III illustrates the pre-requisite information of ACWH that are collected from the on-site workplace. The flowchart of the ACWH inspection process is described in section IV, where the proposed spatial-attention pyramid network and post-processing algorithm are presented in detail. Section V compares the proposed deep learning model with existing start-of-art detectors to evaluate the performance. Meanwhile, processing results of major steps of the automatic inspection method are presented as well. Section VI discusses the experimental results of the inspection work and other potential implementations, while conclusions and future works are summarized in Section VII at last.

II. Related Works

This section summarizes the fundamentals of ACWH inspection, and gives a comprehensive review of recent development and implementation of AR systems and deep learning approaches to automate the industrial inspection.

A. Fundamentals of ACWH inspection

The industrial inspection of ACWH is a very critical operation to guarantee the transmission of electrical signals all around the system of products. The field of ACWH inspection can be mostly categorized into two subfields: position inspection and inspection of mismatched pins.

The main idea behind the former is to diagnose the position error of pins in the stage of mating connectors [12]. With force sensor signals, a robotic system was applied to detect and recovery these errors [13]. Based on machine vision, a measurement system consisting of CCD cameras was developed to detect pins of electric connectors [14]. The system can be applied to replace the manual inspection, which is inefficient and susceptible to subjective factors. To verify the position of pins both in height and in the 2D plane where they lie, a stereo vision system [15] was designed to measure the position of pins that are present in various shapes. Then, deep neural network and pattern matching based on prior knowledge registration were introduced to a 3D vision measurement platform, which can detect abnormal positions of pins for a total of 33 types of connectors [16].

The objective of the latter one is to ensure the correct wiring harness assembly. Wiring harness assembly for aviation connectors is a complicated operation [18], including the recognition of wires, the connection of wires and pins based on identified text, and the mismatched-pins inspection which presents the consistency between pins and wires. The operators are only allowed to connect consistent optical fiber cords at the connector for distribution work in central offices [17]. Similarly, not all the pins of connectors are allowed to be connected with wires, especially among the aerospace products. Nevertheless, scarcely any existing work provides autonomous inspection for mismatched pins of ACWH. A critical challenge that hinders the inspection task is that connectors are fixed with wiring harness in the narrow space of these aerospace products.

B. AR technologies and tools for manufacturing

In the era of advanced manufacturing, AR technologies characterized by digital information is increasingly applied in the production of complex products to achieve visual guidance, interactive collaboration, information management, etc. Based on 3D object registration methods, an assembly training and guidance system [19] was developed to render assembly planning, status of inspection and quality evaluation of products on the model of cabin parts in an AR environment.

To enhance the interaction between human cognition and manual operation, an AR guidance system [20] was established to render appropriate information on a bare-hand interface by evaluating users' behavioral experiences during different phases of tasks. Aiming to develop AR systems in an interaction-free manner, gesture-based behavior recognition method [21] was proposed via SURF feature matching. With a head-worn system, operators can also share on-site information with remote managers. Normally, most existing AR-based platforms are passive due to lack of feedback. To aware operation steps, a haptic module [22] was introduced to track motions of users' wrist.

For highly customized production, AR glasses are employed to transfer dynamically changing information and instructions through a cloud environment to station operators [23]. With tasks assigning to production schedules, intuitionistic instructions are transmitted to human operators on demand, increasing its flexibility. Therefore, the AR technologies and tools can provide a flexible, adaptable and visualizing platform for automatic inspection in manufacturing.

C. Deep learning approach for industrial inspection

One of the major stages for industrial inspection is to detect potential objects from visual information. Recently, deep learning-based object detectors have shown unique advantages on large-scale image databases, such as Pascal VOC [24] and MS COCO [25]. Faster R-CNN [26] was proposed to detect objects in images with two stages including candidate extraction and verification. For objects with varied scales, feature pyramid network (FPN) [27] was introduced to extract the intrinsic information from multi-level feature maps. To avoid the time-consuming process caused by two-stage detection, a single-stage detector YOLO [28] was proposed to meet the real-time need. The major difference is that YOLO divides the entire image into multiple grids, which are used to directly identify objects' positions and types. But YOLO damages performance to some extent. The successive version, i.e., YOLO v2 [29] and YOLO v3 [30], greatly improve performance accuracy and remain efficient. Another one-stage detector SSD [31] achieves a higher detection performance for objects of small sizes, as it adopts the trick of the anchor proposed in Faster R-CNN. To achieve competitive performance equal to two-stage detectors, the focal loss was applied to one-stage detector RetinaNet [32] to tackle the problem of extreme class imbalance during the training.

With vast amounts of data generating in smart factories, deep learning-based methods facilitate the development of analytic and inferential capability for industrial inspection. For instances, a fog computing-based inspection system [33] was introduced to detect defects from images of the production line. The system can cope with extremely large data in real time by offloading the computation burden from the central server to fog nodes. For defect detection of carbon fiber reinforced polymer (CFRP), an unsupervised machine learning method [34] was introduced to extract the weaker deboned defect from noise and background. Moreover, with ultrasonic thermography (UT) visualizing the barely visible impact damage (BVID), a damage inspection method based on manifold learning [35] was proposed to enhance the detectability for CFRP damage.

From the above literature, one can find that most existing studies implement either AR-assisted technologies or deep learning approaches for industrial inspection, respectively. Nevertheless, few concerned about the integration of both, let alone to facilitate the flexible and adaptable inspections in the complex aviation scenarios (e.g. ACWH inspection). Motivated by this, a novel approach should be provided.

III. Pre-requisite Information of ACWH

This section illustrates the basic information of ACWH and its image data acquisition and analysis method, as the pre-requisite information to achieve automatic inspection.

A. The structural information of ACWH

As shown in Fig. 1 (a), the structural information of ACWH includes types of connectors, assembly instructions and assembly results, and it is normally stored in digital forms (XML files) in today's aviation industry. Each pin of connector owns its specific serial number. Outside of all the pins, there are one main dowel and four vice dowels distributed, and the main dowel indicates the position of a pin of the initial serial number (e.g. No. 1). For different types of connectors, assembly instructions and assembly results assigning a series of pins to be connected with specific wires should be provided during the assembly, as illustrated in Fig. 1 (b)-(c). In this context, inspection of mismatched pins is an indispensable step to guarantee the correct assembly.

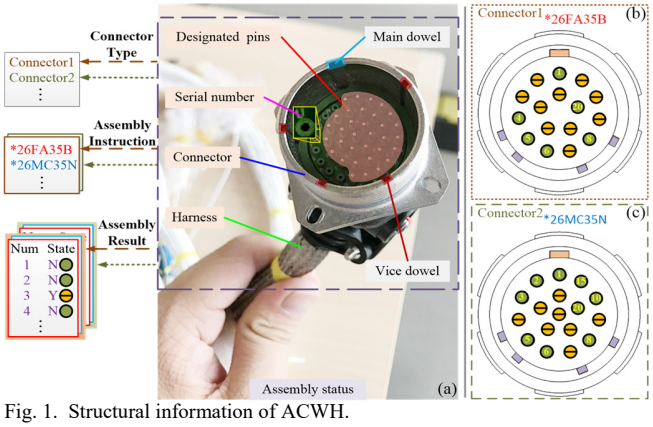


Fig. 1. Structural information of ACWH.

B. Collection of ACWH image dataset

To train effective classifier for the connectors and pins, training dataset should contain enough image samples and all potential scenarios. For dataset acquisition, 1724 images are captured from manufacturing factories by the high-definition digital camera of AR glasses. These images are shooting from different angles and contain complex industrial backgrounds, as shown in the upper part of Fig. 2. Meanwhile, these images contain imaging effects caused by the change of light intensity to ensure the universality of the image dataset. As shown in the lower part of Fig. 2, pins and dowels are classified into eight types to indicate actual assembly status. J1 and J2 represent pins with wires or without wires for female connectors respectively, while P1 and P2 are related examples for male connectors. Dowels on the female connectors are denoted by DP1 and DP2, with the latter signature representing the main dowel. Similarly, DH1 and DH2 carry the same implication for male connectors. Observed in detail, dowels suffer contradictory features (e.g. colors and shapes). Pins of female connectors have only few characteristics and graphic features, while the pins of male connectors are extremely small.

Fig. 2. Image acquisition and categories of pins and dowels.

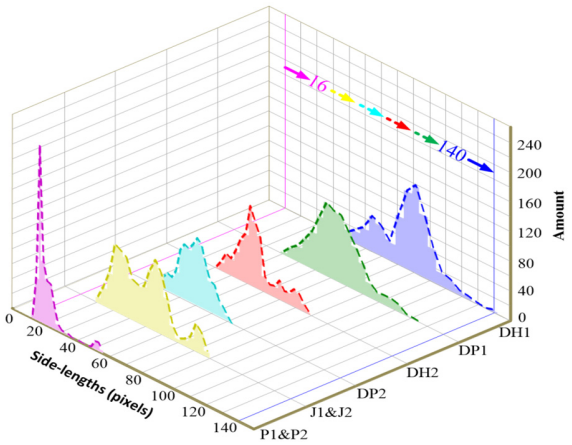


Fig. 3. Statistics analysis of pins and dowels dataset.

C. Statistical analysis of ACWH images dataset

To analyze geometric sizes of pins and dowels, connector regions are scaled to standard images of 1024×1024 pixels. Fig.

3 shows the relationship between geometric lengths of different pins and their amounts in the dataset. It can be found that pins and dowels of connectors present a huge variation range of sizes, of which the maximum length is 140 pixels and the minimum value is only 16 pixels. Also, there is a large disparity of sizes between the largest and the smallest sample of each category.

IV. Proposed Inspection Approach

The work proposes an AR-assisted deep learning-based inspection approach to automatically inspect mismatch pins of ACWH. With the support of AR glasses, visualized information can be transmitted to the onsite operators, where the proposed 4-step inspection flowchart is depicted in Fig. 4.

Step 1. Structural information (XML files) and wireframe images (e.g. CAD models) stored in industrial systems are shared with AR glasses. The onsite connector image is captured by the digital camera of AR glasses.

Step 2. With computing resource from digital manufacturing, the connector detection model is employed to locate the connector region and recognize its type from pre-proposing images (e.g. normalizing and channel modifying). The connector region is cropped from the input as a separate image. Pin detector, i.e. spatial-attention pyramid network is applied to locate pins and dowels from the cropped image and classify their categories, e.g. P1 or P2.

Step 3. Pin-distribution recovering approach is employed to obtain their original distribution. Pins are then sorted by the cluster-generation sequencing algorithm with serial numbers.

Step 4. Since assembly results from structural information are the ground truth (GT), mismatched pins can be identified by comparing the assembly status with the GT in order. The AR glass delivers warning information to operators with an intuitionistic interface to highlight those mismatched pins.

A. Spatial-attention Pyramid Network

Object detector, RetinaNet is utilized to acquire coordinates and types of connectors from images that contain complex manufacturing background. According to the output coordinates from RetinaNet, the connector region is cut down as a single image to detect pins and dowels thereafter.

To detect pins and dowels of extremely small sizes and various scales, a network of FPN architecture is built to extract both large features and small features from images by integrating the feature information between high-level layers and shallow layers. Meanwhile, BiLSTM [36] is adopted to replace convolution operation in deep-level layers, so as to extract spatial rules of pins and dowels distributed on connectors. In this way, the proposed spatial-attention pyramid network not only focuses on the graphic features of images (e.g. edge, color and corner) in various field of views, but also searches for the spatial relationships of extracted features.

The architecture of the spatial-attention pyramid network is shown in Fig. 5. The model mainly consists of feature extractor stem, feature pyramid net and output layer. ResNet50 is employed as the backbone of the feature extractor branch. In the feature pyramid branch, features of high-level layers and low-level layers are gradually integrated as follows:

$$b_i = \begin{cases} conv_{1\times1}(f_i) & if\ i \le 2 \\ B\_C(f_i) & otherwise \end{cases} \quad (1)$$

$$g_i = \begin{cases} add\big(b_i, upsamble(b_{i+1})\big) & if\ i = 3 \\ add\big(b_i, upsamble(g_{i+1})\big) & otherwise \end{cases} \quad (2)$$

, where the *B_C* layer contains BiLSTM components, as the detailed structure is shown in Fig. 6. For each pixel of feature maps, a sliding window (3×3 pixels) is introduced to gather values of this pixel and surrounding eight pixels, and these nine values are output to the final channel of feature maps. Hence, this processing, named extract patch, increases information for the feature map by expanding its dimension from (B, H, W, C) to (B, H, W, 3×3×C) without new training parameters introducing into detection model. Hence, it does not aggravate the difficulty of training. To meet requirements of input data of BiLSTM, feature maps that output form extracting patch is reshaped to (B×H, W, 3×3×C) as sequence data. The sequential data is sent to BiLSTM to generate feature maps of (B×H, W, C), which contains information of position correlations of pins and dowels. Finally, the output feature map is recovered to (B, H, W, C) for afterwards deeper layers of the network, where C is defined as the number of channels of feature maps.

The *B_C* layer is only applied to deeper layers of *Res5c_relu* and *Res4f_relu*, which are output from ResNet50. By comparison, shallow layers are not enough effective to extract the intrinsic information of images. More importantly, disadvantages of shallow feature maps, such as large sizes and fewer channels, can lead to a short data sequence which requires more basic units in the chained BiLSTM components. As these problems make the training process more difficult and even degenerate the accuracy performance of networks, the *B_C* layer is not introduced to the low-level feature maps.

The output layer is simple and efficient, as subnets of classification and regression are only applied to the final two upper feature maps of the feature pyramid branch. Each pixel (i.e. sliding position) of feature maps of these two subnets is related to one anchor. More specifically, anchors on the top subnet and the bottom subnet respectively represent 16^2 and 32^2 pixels of input images. Values of sliding steps of anchors are set to 4 and 8 for these two subnets, respectively. Finally, nine anchors are yield on every sliding position by multiplying three scales ($[2^0, 2^{1/3}, 2^{2/3}]$) and three aspects rations ($[2^0, 2^{1/3}, 1+2^{2/3}]$) to each anchor. These anchors cover scale ranges from 16 to 82 pixels of input images. Every anchor is assigned to a length *K* one-hot vector for classification targets and a 4-vector for box regression targets. *K* is the number of classes of pins and dowels and is set as eight.

In classification subnet, the probability of object presence at each sliding position is indicated by nine anchors and *K* object classes. In the end, sigmoid activations are attached to output 9*K* binary predictions for each sliding location.

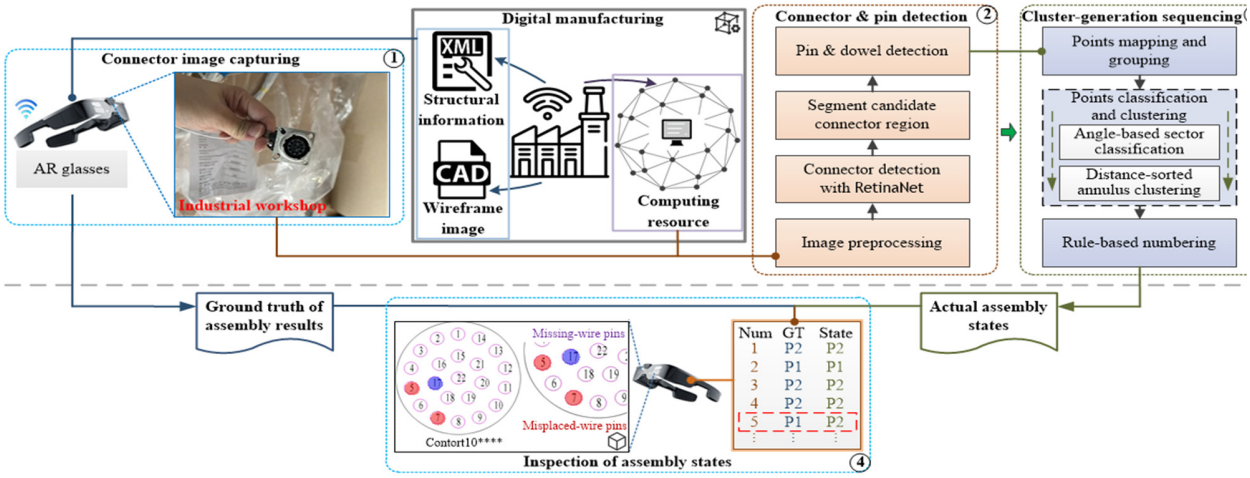


Fig. 4. Flowchart of inspecting mismatched pins.

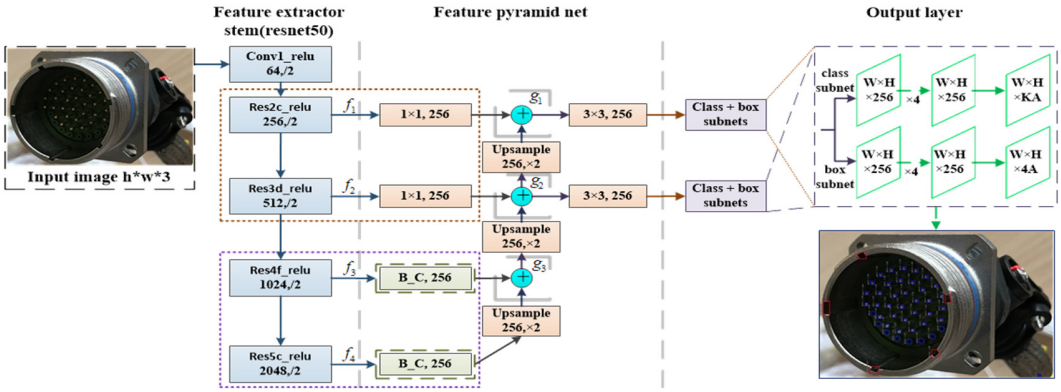


Fig. 5. The architecture of spatial-attention pyramid network.

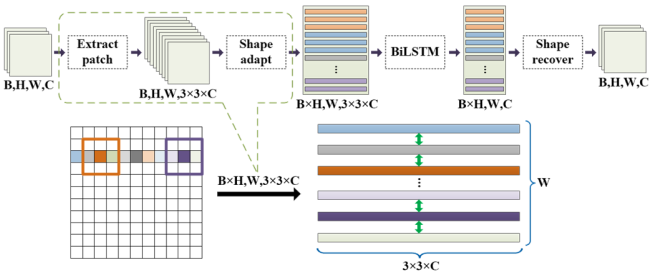


Fig. 6. Spatial-attention layer with BiLSTM component.

In box regression subnet, there are four outputs predicting the relative coordinates v of the bounding box location of one anchor. The relative coordinates v is given by:

$$v_x = (c_x - c_x^a)/w^a,\ v_y = (c_y - c_y^a)/h^a$$
$$v_w = log(w - w^a),\ v_h = log(h - h^a) \quad (3)$$

, where v=$\{v_x, v_y, v_w, v_h\}$ are the relative predicted coordinates. $\{c_x^a, c_y^a, w^a, h^a\}$ denote the center coordinates and its width and height of the anchor box, which can be pre-computed from an input image. $\{c_x, c_y, w, h\}$ are the predicted coordinates of the bounding box. To jointly optimize parameters of pin detector, two loss functions, L_s^{cl} and L_v^{re}, are introduced to calculate errors of subnets of classification and box regression, respectively. The loss is formulated as

$$L = L_s^{cl} + \lambda_{re} L_v^{re} \quad (4)$$

, where λ_{re} balances the influence of two losses on the training and is set to 1. In the image dataset, there is a large disparity of the number of pins and dowels from different classes. To solve the problem of class imbalance in the classification task during training, L_s^{cl} adopts focal loss which is expressed as follows

$$L_s^{cl} = \begin{cases} -\alpha(1-\hat{y})^{\gamma}\log\hat{y}, & if\ y = 1 \\ -(1-\alpha)\hat{y}^{\gamma}\log(1-\hat{y}), & otherwise \end{cases} \quad (5)$$

, where y is the label of instances, α balances the importance of positive/negative examples (i.e. candidates generated by the network is potential objects or not) and is set to 0.25 in the experiment. The γ trades weights between easy/hard examples and enable training to focus on hard examples (i.e. losses of examples with errors) and is set to 2.

The loss function of regression tasks in object detection models normally applies the L_n-norms, which is defined by the distance between the corners of two boxes. As Intersection over Union (IoU) is a metric used in evaluation benchmarks, metrics between the training and testing are not consistent. Another problem is that L_n -norms make detection performance sensitive to the choice of IOU threshold, especially the detection model may present instability for objects of various sizes. Hence, L_v^{re} adopts the form of GIOU [37] loss that enables models to present invariant accuracy against different choices of IOU thresholds. With $\{x_1, y_1, x_2, y_2\}$ indicating corners of the predicted bounding box $\hat{R}$ and $\{x_1^*, y_1^*, x_2^*, y_2^*\}$ denoting corners of the GT bounding box R^*, L_v^{re} is attained by:

$$L_v^{re} = 1 - \left({|\hat{R} \cap R^*|}/{|\hat{R} \cup R^*|} - {(R^C - |\hat{R} \cup R^*|)}/{R^C} \right) \quad (6)$$

, where R^C is the area of the smallest enclosing box B^C, and coordinates of B^C are:

$$x_i^c = min(x_i, x_i^*),\ y_i^c = min(y_i, y_i^*),\ i=1,2 \quad (7)$$

$$R^C = (x_2^c - x_1^c) \times (y_2^c - y_1^c) \quad (8)$$

B. Pin-distribution Recovering and Cluster-generation Sequencing Algorithm

As shooting angles and focal distances of cameras are uncontrolled, pins may be distributed on irregular annuluses in images, which affect the subsequent work to obtain serial numbers of pins. Homographic matrix H (i.e., perspective transformation matrix) is utilized to recover the original distribution of pins. The formula used to transform two points of (x, y, z) and (x', y', z') with matrix H is given:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (9)$$

More than four optimal matching points are required to estimate parameters of H due to its nine freedom degrees. One registration image, in which pins are distributed on regular multi-layer annuluses is selected from every type of connector image. Then five pairs of optimal matching points, i.e. one pair of *DP1/DH1* and four pairs of *DP2/DH2*, are searched over the real-time test image and the registration image to calculate the H. Finally, coordinates of all pins on test images are recovered by H to acquire their original distribution on the connector. As illustrated in Fig. 7 (a)-(b), pins of connectors are distributed on annuluses from the outer layer to inner layer anticlockwise. The number of annuluses' layer is determined by the type of a connector. Specifically, the central pin P^e is the last serial number if it exists on connectors. Based on the distribution rule, the strategy of cluster-generation sequencing algorithm contains the following stages:

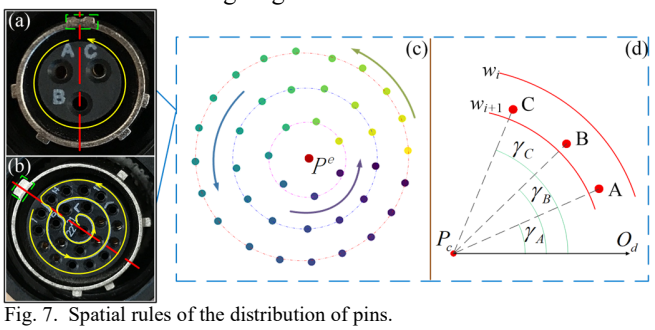


Fig. 7. Spatial rules of the distribution of pins.

1) *Points mapping and grouping*. All detected pins and dowels are mapped to discrete points according to coordinates of their center. The discrete points related to dowels are grouping into set d, of which the main dowel (marked as O_d) is a reference of the initial position to sequence all pins. Then these discrete points related to pins are put into group S. The central point Pc of all points in S is calculated.

2) *Points classification and clustering*. The connector type obtained from the connector detection model presents information about the number of annuluses' layer n and whether the given connector has the central pin P^e. With the information of n and P^e, points in set S are clustered into different annuluses from the outer layer to the inner layer. The detailed process is described in Algorithm 1.

3) *Rue-based numbering*. For all points in set S, a polar coordinate system is built with Pc as the pole point and $\boldsymbol{P_cO_d}$ as the polar axis. Based on the polar-angle value γ_i of points, serial numbers of all pins are acquired by sorting points of every annulus from outside to inside (see Fig. 7(c)-(d)).

Algorithm 1: Points classification and clustering

Input: Existing flag f_c of P^e, Number of annuluses' layer n, Point set of pins S
Output: Point sets in every annulus $\{C_1,..., C_n, C_{n+1}\}$

1: **function** CLUSTERING (f_c, n, S)
2. $P_c(\bar{x}, \bar{y}) \leftarrow \frac{1}{N}\sum_i^N P(x_i, y_i)\ \forall P_i \in S$
3. **if** f_c **then**
4. The nearest point to P_c in S is moved to C_{n+1}
5. **end if**
6. $u = \lfloor(|S| - 2*n)/(2*n)\rfloor+1$ // The number of sectors
7. r =(|S|-2*n)%(2*n)+2*n // The number of points in the last sector
8. Points in S are sorted based on their polar angles to pole point P_c
9. The last r points in S is moved to C_u
10. The remaining points in S are split equally into sets $C_1,..., C_{u-1}$
11. Points in $C_1,..., C_u$ are sorted by their distance to P_c
12. Calculating the index burst (*IB*) for points in $C_1,..., C_u$
13. Points in $C_1,..., C_u$ are split into n sets $C_1,..., C_n$ with the largest n-1 *IB*
14. **return** $\{C_1,..., C_l, C_{l+1}\}$
15. **end function**

The processing of points classification and clustering contains two stages, i.e. angle-based classification and distance-sorted annulus clustering. In angle-based sector classification, the first consideration is that the number of points in each sector should be greater than annuluses' layers n to ensure that every sector contains points (i.e. pins) from all annulus of the connector. Another one is to generate as many sectors as possible to split all points into fan-shaped distributions of smaller angles, which reduce the difficulty to classify pins to different layers of annuluses. Based on these two aims, the number of sectors u is given in Line 6 of the algorithm 1, and the amount of points r for the last sector is calculated in Line 7. Then, the last r points in set S are moved to the last sector while the remaining points in S are divided equally to other sectors, as shown in Fig. 8 (a).

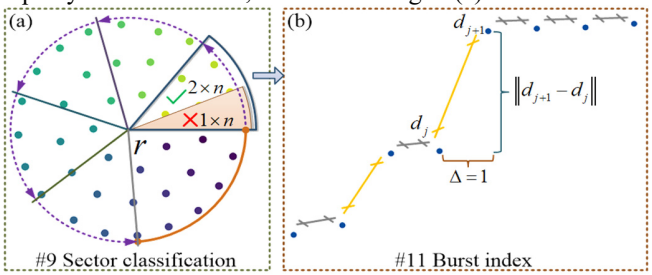


Fig. 8. Point classification and clustering based on angle and distance.

In the process of distance-sorted annulus clustering, burst index (Fig. 8 (b)) that indicates the distance between points inside each sector to central point P_c is first calculated in Line 12 of the Algorithm 1. Then the largest n-1 burst index in every sector is used to split points of the corresponding sector into n different annuluses. Finally, serial numbers of all pins are obtained by sorting these points in the subsequent process of

rule-based numbering. Traversing all numbered pins in order, mismatched pins are identified by comparing the GT of assembly result and its actual status.

V. Case Study and Experimental Result

In this section, several comparison experiments are conducted to evaluate the performance of detection accuracy for the spatial-attention network. Processing results of each step for inspection of mismatched pins are visualized as well.

A. Pin detection

There are 11 different types of connectors in the dataset which contains 1346 training images and 378 testing images. RetinaNet is trained as the connector detector to locate positions of the connector region in an image. The testing images are used to quantify the detection result with the evaluation criterion of mean average precision (mAP), which is the metric for object detection in Pascal VOC. The performance of detection accuracy is presented in Table I.

The connector region is cropped down from the above 1346 training images and 378 testing images respectively to develop the dataset for pin detection model. Finally, 1009 training images and 360 testing images are acquired after removing images which suffer extremely poor quality, such as blur and false labels. Similarly, these two groups of images are fed into the spatial-attention pyramid network to train the pin detector and test performance of the trained model, respectively. Without any augmented image, our proposed model is trained with the stochastic gradient descent optimizer over two GPUs (RTX 2080) with a total of two images per minibatch, i.e. each GPU processes one image of 1024×1024 pixels. The model is trained for 80k iterations with an initial learning rate of 0.01, which is then divided by 10 at 40k and again at 60k iterations. The weight decay is set to 0.0001 and the momentum is set to 0.9. The training loss is the sum of the focal loss for the classification and GIOU loss for the box regression.

To verify the performance of detecting pins and dowels in images, the proposed spatial-attention pyramid network is compared with the first-class 2D object detectors, including Faster R-CNN, SSD, YOLO v2, RetinaNet and YOLO v3. These existing object detectors achieve top performance in the most popular public databases, e.g. PASCAL VOC and MS-COCO, etc. Based on the same computing resource (one RTX 2080 GPU, i7-9700k CPU and 16G RAM) and the connector dataset leveraged, Table II presents the performance accuracy and the computation efficiency of our proposed model and these existing detectors on detecting pins and dowels from on-site images. Some examples and detection results of the pin detector are illustrated in Fig. 9.

YOLO v3 normally presents a high performance of detection accuracy for small objects. Compared with YOLO v3, our detector shows a dramatic improvement to detect pins, such as J1, J2, P1 and P2, which are distributed with some spatial rules. It means that the BiLSTM structure in the network is activated to extract the spatial features during training and testing. Besides, J2 represents negative hard examples which are not easy to distinct with J1, but our proposed method presents the most significant improvement of performance on detecting this kind of confused instances. The reason is that our detector focuses on extracting features of such negative hard samples during training with the implement of the focal loss.

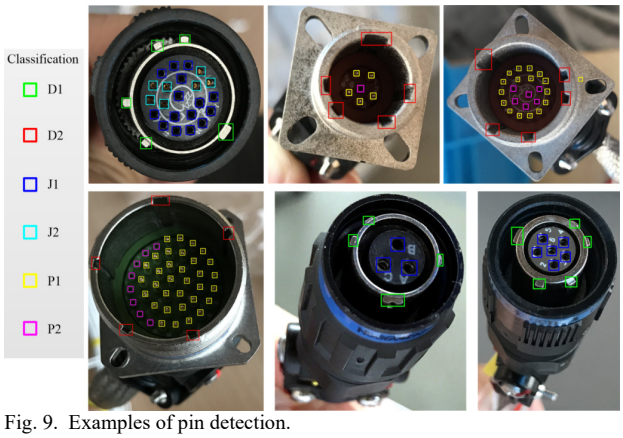


Fig. 9. Examples of pin detection.

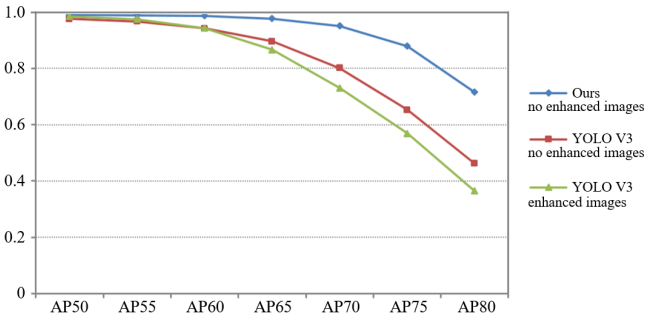


Fig. 10. mAP in multiple IoU thresholds.

In the MS COCO challenge, the widely reported accuracy measure for detection, averages mAP, is calculated based on multiple IoU values, e.g. 0.5, 0.6, 0.7 and 0.8, etc. Taking this more reasonable evaluation criterion into consideration, the detection results are quantified with mAP of the different threshold values to verify the advantaged performance of the proposed spatial-attention network. Five different ways of image expansion are adopted to enhance the original 1009 training images to 6054 images. Our model is trained by the original 1009 images, and the two models of YOLO v3 are trained by 1009 images and enhanced 6054 images respectively. As shown in Fig. 10, the gap of detection accuracy between our model and YOLO v3 enlarges as the IoU threshold rises. Our proposed model is less sensitive to the choice of IoU threshold and maintain higher detection accuracy even in a large IoU value.

TABLE I
Performance of Connector Detection Accuracy

mAP	CONN1	CONN2	CONN3	CONN4	CONN5	CONN6	CONN7	CONN8	CONN9	CONN10	CONN11
0.9961	0.9996	0.9846	0.9925	0.9956	0.9992	0.9979	0.9941	0.9996	0.9977	0.9975	0.9985

TABLE II
PERFORMANCE OF PIN DETECTION

Method	mAP	DH1	DH2	DP1	DP2	J1	J2	P1	P2	time
Faster R-CNN	0.7268	0.9044	0.9078	0.8984	0.9047	0.8325	0.8248	0.2193	0.3225	--
SSD	0.5112	0.8956	0.8595	0.8305	0.5188	0.3432	0.4619	0.0961	0.0625	--
YOLO V2	0.4982	0.8060	0.7849	0.8143	0.7147	0.3430	0.3414	0.0909	0.0909	--
RetinaNet	0.7483	0.9696	0.9892	0.9943	0.9795	0.9870	0.8888	0.0633	0.1152	237ms
YOLO V3	0.9757	0.9935	0.9980	0.9840	0.9653	0.9843	0.9084	0.9834	0.9893	109ms
Ours	0.9900	0.9912	0.9964	0.9829	0.9858	**0.9965**	**0.9709**	**0.9973**	**0.9986**	318ms

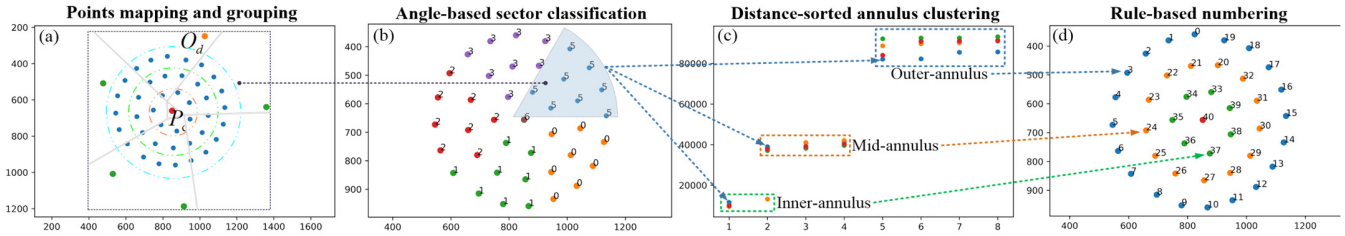


Fig. 12. Visualization of cluster-generation sequencing algorithm.

TABLE III
Accuracy of Models trained by Different Parameters

γ	$\propto$	AP50	AP65	AP75
1.0	0.50	0.9480	0.8853	0.7736
1.0	0.25	0.9771	0.9261	0.8054
2.0	0.25	0.9900	0.9772	0.8794
5.0	0.25	0.9891	0.9419	0.8217

As mentioned earlier, by introducing parameter α and parameter γ in the loss function of the classification task, the proposed model can put attention to hard, misclassified examples during training. The effects on the accuracy of models trained by other values of the parameters are illustrated in Table III. To maximize the performance of the model, the parameter α and the parameter γ are set to 0.25 and 2.0, respectively.

B. Visualization for inspection of mismatched pins

With five pairs of matching points (i.e. dowels of connectors) establishing a homographic matrix *H*, the pin-distribution recovering algorithm recovers pins' distribution of every annulus from irregular ellipses to more regular circles (see Fig. 11). It can against the coordinate deviation of pins caused by weird shooting angle of a camera. After the processing of the pin-distribution recovering, detected pins and dowels are grouped into discrete points and then classified into separated sectors, as illustrated in Fig. 12 (a)-(b). Points from every sector are clustered into different annuluses from outside to inside, i.e. outer-annulus, mid-annulus and inner-annulus, as presented in Fig. 12 (c). Finally, 40 points related to pins obtain their serial numbers correctly by sorting from the outer layer to the inner layer in a counterclockwise direction (see Fig. 12 (d)). Mismatched pins are identified from the inconsistency of comparisons between the GT of assembly results and actual assembly status of these numbered pins. The AR glass, as an important interactive device, presents a wireframe image which highlights mismatched pins to onsite operators. For example, in Fig. 13, the highlight blue pin of No. 6 of Connector 5 reminds operators to install an aviation wire into this pin, while the red pin of No. 4 denotes a misplaced one to be removed.

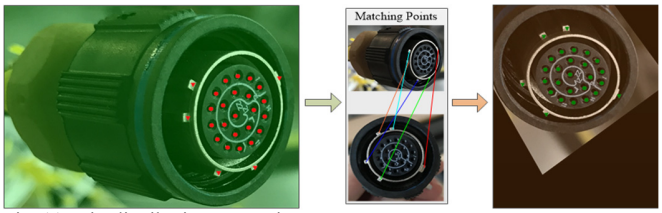


Fig. 11. Pin-distribution recovering.

With the deployed system, one cycle inspection starts from that a connector image is captured by AR glasses in real time to the end when the AR glasses display the wireframe image. During the cycle, the majority of time spends on the calculation of the detection model, where connectors are detected from images in 224ms and pins are recognized from images in 318ms. Coupled with the rest of little time consuming, such as image transmission and information interchange, the AR-assisted deep learning-based inspection of mismatched pins can be performed within 1s between manufacturing systems and onsite operators automatically.

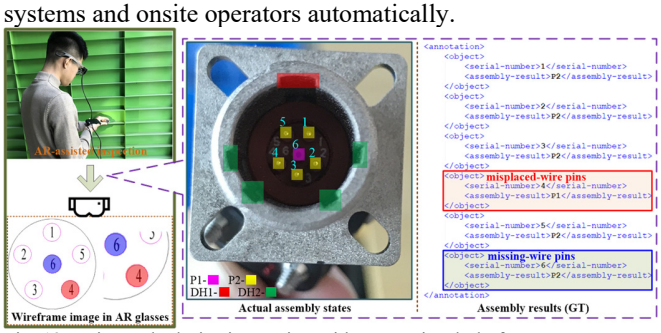


Fig. 13. Mismatched-pins inspection with AR-assisted platform.

VI. DISCUSSION

To detect small pins of various scales of the ACWH, the proposed spatial-attention pyramid network extracts feature information in a diverse field of view, and search for their spatial relationships. Also, the cluster-generation sequencing algorithm can resist position deviation of pins in images to some extent. By integrating on AR glasses, the automatic inspection method can be applied to any complex industrial scenarios to provide digital inspection services with flexibility, and to present an intuitionistic visualization effect. Meanwhile, compared to other existing industrial camera-based inspection

systems (Refs. 14-16) or robotic systems (Refs. 12-13), the deployed system has prominent advantages in terms of competitive price and powerful computation efficiency, as shown in Table IV. As computing serves are commonly essential facilities existing in the aviation industries, the only cost of purchasing hardware for the AR-based inspection system can be less than $1,700. It is envisioned that the proposed approach can also be readily extended to many other applications, such as target tracking in human-machine collaboration systems. Despite those advantages, the computation efficiency of the proposed approach (318ms) is rather slower than the YOLO V3 (109ms). Therefore, there is still room for further improving its computing efficiency without sacrificing the accuracy.

TABLE IV
COSTS OF INSPECTION SYSTEMS

Platform	Industrial Camera-based Inspection System		Robotic System	Ours
	Nanovea ST400 Profilometer	Ref. [16]		
Hardware	Color Video Zoom Camera Integration, Adjustable Height experimental platform	Two Basler Industrial Cameras, Light Source, Experimental Platform	Robot, Industrial Cameras, Light Source	HiAR G200
Computing Power	PC	PC with GTX 1080 GPU	PC	RTX 2080 GPU, i7-9700k CPU and16G RAM
Pricing	Over 50,000$	Around 4,200$	Over 30,000$	Around 3,000$

VII. CONCLUSION

In this research, an automatic aviation connector inspection method was proposed, which is in line with today's digital transformation in the aviation industry. With inferred knowledge transferring from the learning model, AR glass was leveraged to support onsite operators' inspection work effortlessly in a visualized manner. To summarize, the main scientific contribution of this research lies in in two aspects:

1) Proposed a spatial-attention pyramid network combining architectures of FPN and BiLSTM to address the challenge of detecting small objects of various scales, which exposure to complex and nature industrial scenarios. Since most machine parts and components have spatial correlations, the proposed model shows great advantages of adaptability and flexibility that can be adopted in various applications.

2) Integration of AR visualization and deep learning technologies to provide automatic inspection services with flexibility and mobility. This work opens the new way to conduct visualized industrial inspection automatically inside product regardless of its complex structure or narrow spaces.

Moreover, to validate the proposed approach, an experimental analysis of mismatched pins of 11 different ACWH was conducted, and the results show that the pin detection model achieves competitive performance with an accuracy up to 99.00%. Apart from the abovementioned advantages, several potential future research directions are also highlighted here, including 1) optimizing the loss function of the model to improve the performance of detection accuracy, and verifying the feature extraction ability of the model for other important parts of aerospace products, 2) extending the proposed AR-assisted data-driven inspection approach to other industrial scenarios, and 3) deploying cloud-edge computing service platform to facilitate the on-demand inspection service.

REFERENCES

[1] L. Thames and D. Schaefer, "Software-defined Cloud Manufacturing for Industry 4.0," in *Procedia CIRP*, 2016, vol. 52, pp. 12–17.

[2] P. Zheng *et al.*, "Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives," *Front. Mech. Eng.*, vol. 13, no. 2, pp. 137–150, 2018.

[3] H. Rios, E. González, C. Rodriguez, H. R. Siller, and M. Contero, "A mobile solution to enhance training and execution of troubleshooting techniques of the engine air bleed system on Boeing 737," *Procedia Comput. Sci.*, vol. 25, pp. 161–170, 2013.

[4] S. Martelli *et al.*, "Deep Endoscope: Intelligent Duct Inspection for the Avionic Industry," *IEEE Trans. Ind. Informatics*, vol. 14, no. 4, pp. 1701–1711, 2018.

[5] R. Maglietta, A. Milella, M. Caccia, and G. Bruzzone, "A vision-based system for robotic inspection of marine vessels," *Signal, Image Video Process.*, vol. 12, no. 3, pp. 471–478, 2018.

[6] R. K. Gupta *et al.*, "Gathering, evaluating and managing customer feedback during aircraft production," *Comput. Ind. Eng.*, vol. 115, no. May 2017, pp. 559–572, 2018.

[7] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun, and K. Liu, "A Simple Guidance Template-Based Defect Detection Method for Strip Steel Surfaces," *IEEE Trans. Ind. Informatics*, vol. 15, no. 5, pp. 2798–2809, 2019.

[8] Y. Park and I. S. Kweon, "Ambiguous Surface Defect Image Classification of AMOLED Displays in Smartphones," *IEEE Trans. Ind. Informatics*, vol. 12, no. 2, pp. 597–607, 2016.

[9] C. F. J. Kuo, T. ying Fang, C. L. Lee, and H. C. Wu, "Automated optical inspection system for surface mount device light emitting diodes," *J. Intell. Manuf.*, vol. 30, no. 2, pp. 641–655, 2019.

[10] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, 2018.

[11] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Ann. - Manuf. Technol.*, vol. 65, no. 1, pp. 417–420, 2016.

[12] J. Huang, P. Di, T. Fukuda, and T. Matsuno, "Robust model-based online fault detection for mating process of electric connectors in robotic wiring harness assembly systems," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 5, pp. 1207–1215, 2010.

[13] F. Chen, F. Cannella, J. Huang, H. Sasaki, and T. Fukuda, "A Study on Error Recovery Search Strategies of Electronic Connector Mating for Robotic Fault-Tolerant Assembly," *J. Intell. Robot. Syst. Theory Appl.*, vol. 81, no. 2, pp. 257–271, 2016.

[14] G. S. and Y. P. Daxing Zhao, Wei Feng and P. Yu, "High Precision Measurement System of Micro-electronic Connector based on Machine Vision," *J. Appl. Sci.*, vol. 13, no. 22, pp. 5363–5369, 2013.

[15] L. Stroppa and C. Cristalli, "Stereo Vision System for Accurate 3D Measurements of Connector Pins' Positions in Production Lines," *Exp. Tech.*, vol. 41, no. 1, pp. 69–78, 2017.

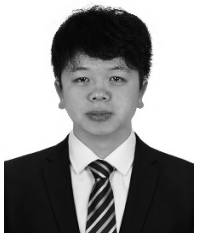
[16] D. Zhao, F. Kong, and F. Du, "Vision-based adaptive stereo measurement of pins on multi-type electrical connectors," *Meas. Sci. Technol.*, 2019.

[17] I. Ogushi, K. Katayama, and Y. Azuma, "A novel navigation system using augmented reality technology for distribution work at optical connector panels," *14th Asia-Pacific Netw. Oper. Manag. Symp. "Management Big Data IoT Era", APNOMS 2012 - Final Progr.*, no. Idm, pp. 1–7, 2012.

[18] L. Zheng, X. Liu, Z. An, S. Li, and R. Zhang, "A smart assistance system for cable assembly by combining wearable augmented reality with portable visual inspection," *Virtual Real. Intell. Hardw.*, vol. 2, no. 1, pp. 12–27, 2020.

[19] Y. Liu, S. Q. Li, J. F. Wang, H. Zeng, and J. P. Lu, "A computer vision-based assistant system for the assembly of narrow cabin

products," *Int. J. Adv. Manuf. Technol.*, vol. 76, no. 1–4, pp. 281–293, 2014.
[20] X. Wang, S. K. Ong, and A. Y. C. Nee, "Multi-modal augmented-reality assembly guidance based on bare-hand interface," *Adv. Eng. Informatics*, vol. 30, no. 3, pp. 406–421, 2016.
[21] X. Yin, X. Fan, W. Zhu, and R. Liu, "Synchronous AR assembly assistance and monitoring system based on ego-centric vision," *Assem. Autom.*, vol. 39, no. 1, pp. 1–16, 2019.
[22] C. Y. Siew, S. K. Ong, and A. Y. C. Nee, "A practical augmented reality-assisted maintenance system framework for adaptive user support," *Robot. Comput. Integr. Manuf.*, vol. 59, no. March, pp. 115–129, 2019.
[23] D. Mourtzis, V. Zogopoulos, and F. Xanthi, "Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling," *Int. J. Adv. Manuf. Technol.*, 2019.
[24] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2015.
[25] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014, pp. 740–755.
[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
[29] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
[30] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv Prepr. arXiv1804.02767*, 2018.
[31] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21–37.
[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
[33] L. Li, K. Ota, and M. Dong, "Deep Learning for Smart Industry: Efficient Manufacture Inspection System with Fog Computing," *IEEE Trans. Ind. Informatics*, vol. 14, no. 10, pp. 4665–4673, 2018.
[34] J. Ahmed, B. Gao, and W. L. Woo, "Wavelet-Integrated Alternating Sparse Dictionary Matrix Decomposition in Thermal Imaging CFRP Defect Detection," *IEEE Trans. Ind. Informatics*, vol. 15, no. 7, pp. 4033–4043, 2019.
[35] X. Zhang *et al.*, "CFRP Impact Damage Inspection Based on Manifold Learning Using Ultrasonic Induced Thermography," *IEEE Trans. Ind. Informatics*, vol. 15, no. 5, pp. 2648–2659, 2019.
[36] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, and A. Stolcke, "The Microsoft 2017 conversational speech recognition system," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2018, pp. 5934–5938.
[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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.



Shufei Li received the M.S. degree in the Department of Industrial and Manufacturing Systems Engineering from Beihang University, Beijing, China, in 2020, and the B.E. degree in Mechatronic Engineering from Shandong Jianzhu University, Jinan, China, in 2017.

He is currently working towards the Ph.D. degree in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. His research interests include AR-assisted smart manufacturing, human-robot collaboration, deep learning and computer vision.

PAI ZHENG (M'17) is currently an Assistant Professor in the Department of Industrial and Systems Engineering, at the Hong Kong Polytechnic University. He has been a Research Fellow in the Delta-NTU Corporate Laboratory for Cyber-Physical Systems, in the School of Electrical and Electronic Engineering, at Nanyang Technological University (NTU), Singapore from Jan. 2018 to Sep. 2019. Dr. Zheng received the dual Bachelor's Degrees in Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2010, the Master's Degree in Mechanical Engineering from Beihang University, Beijing, China in 2013, and the Ph.D. degree in Mechanical Engineering at University of Auckland, Auckland, New Zealand in 2017.

His research interest includes smart product-service systems, engineering informatics, and manufacturing servitization. He is a member of IEEE, CMES, and ASME, and serves as the editorial board member for the journal of Advanced Engineering Informatics, managing guest editor for the Journal of Cleaner Production, and referee for several high impact international journals in the manufacturing/industrial engineering field.

LIAN-YU ZHENG is currently a professor and the Head of the Department of Industrial and Manufacturing Systems Engineering, School of Mechanical Engineering and Automation, at Beihang University. He received the B.S., M.S. and Ph.D. degrees in mechanical engineering from Beihang University, Beijing, China, in 1989, 1993, 2001 respectively.

His current research interests include digital and intelligent manufacturing, reconfigurable and flexible manufacturing, industrial artificial intelligence, manufacturing modelling and simulation. Professor Zheng is an editorial board member of the Journal of Intelligent Manufacturing and of the Journal of Engineering Sciences. He serves as the academic members of the Group Technology and Intelligence Integration Committee, and the Mechanical Automation Committee of CSME as well as the Intelligent Manufacturing Committee of Chinese Association for Artificial Intelligence (CAAI). He is also an academic board member of the Beijing Key Lab of Digital Design and Manufacturing.