International Journal of Innovative Computing, Information and Control
Volume **22**, Number **2**, April **2026**

ICIC International ©2026 ISSN 1349-4198
pp. **549–564**

DEVELOPMENT OF A SURFACE DEFECT INSPECTION SYSTEM FOR 3D GLOSSY PARTS USING DEEP LEARNING

Zhong Zhang[1], Keisuke Matsuoka[2], Jin Qi[1] and Toshihiro Okuyama[3]

[1]Department of Smart Design
Aichi Sangyo University
12-5 Harayama, Okamachi, Okazaki-shi, Aichi 444-0005, Japan
{ tadashis; saisusumu }@asu.ac.jp

[2]Department of Intelligent Mechanical Engineering
Hiroshima Institute of Technology
2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193, Japan
ad20114@cc.it-hiroshima.ac.jp

[3]Department of Commercial Claims
Tokio Marine & Nichido Fire Insurance Co., Ltd.
2-6-4 Otemachi, Chiyoda-ku, Tokyo 100-8050, Japan
toshihiro.okuyama3@tmnf.jp

Received July 2025; revised November 2025

Abstract. *In the manufacturing of industrial products, mirror finishing is commonly applied to giving the product a glossy appearance. However, such surfaces reflect light specularly, making automatic defect inspection challenging, and therefore limiting its adoption in practice. Currently, many quality inspections are still performed manually, but issues such as rising labor costs, low efficiency, and inconsistent inspection accuracy have increased the demand for automation. This study aims to develop an automatic inspection system for detecting surface defects on glossy parts. Two types of inspection devices were developed: one for cylindrical parts (Part A) and the other for three-dimensional curved parts (Part B). For Part A, we successfully constructed an automatic inspection system that completes defect detection from image capture to result output in approximately 18.53 seconds for the first part, and an average of about 9.60 seconds for subsequent parts. Furthermore, optimizing the lens configuration for each case enabled high defect detection accuracy for both Part A and Part B.*
Keywords: Surface inspection, Surface defect, Automatic inspection system, Deep learning

1. **Introduction.** The process of mirror-finishing, commonly used in industrial products to enhance their surface gloss, often requires rigorous quality inspection to identify and remove any defects during manufacturing. While automation of inspection is now commonplace for non-mirror-finished products, the reflective property of such products renders traditional methods of automated inspection ineffective. Consequently, visual inspection by human operators remains the norm, though it suffers from several problems such as reliance on operator skill, labor intensity, and high cost. An urgent need exists, therefore, to establish an effective method for automating the inspection of mirror-finished products.

DOI: 10.24507/ijicic.22.02.549

Numerous researchers both in Japan and overseas have dedicated considerable effort to the development of effective surface inspection methods for glossy parts. For example, Kanno [1] introduced a small defect detection device that uses slit light for mirror-coated products, while Nakamura [2] proposed ring-lighting as a means of detecting defects based on variance of the surface normal direction. Höfer et al. [3] put forward an infrared ellipsometry-based approach to inspecting mirror defects, and Wakisako and Mori [4] developed appearance inspection technology suitable for glossy plastic parts using a stripe pattern projection method. Similarly, Hoshino et al. [5] proposed an appearance inspection method for glossy parts that involved using a striped pattern coaxial light source. Mengesha et al. [6] used scanning electron microscopy (SEM) to generate grayscale images and 3D height maps of scratched nickel plating, enabling precise characterization using machine learning clustering algorithms. Wang et al. [7] developed a multimodal defect detection system (MDDS) by integrating 2D imaging and 3D point clouds to overcome the depth limitations of 2D imaging, but this system also relies on high-precision industrial equipment. As such, while high-accuracy scratch detection systems exist, they typically require expensive measurement instruments, making them less practical for general on-site use.

Deep learning has shown promising results in solving problems that require high generalization performance in image recognition, with convolutional neural networks (CNNs) being particularly effective. For instance, Zhang et al. [8] developed an ensemble CNN by combining multiple CNNs. However, one limitation of this approach was the inability to specify the position of detected defects based on a binary classification output. To address this issue, Girshick et al. [9] proposed a region-based CNN (Regions with CNN features: R-CNN), which identifies object candidates in an image, extracts features using a CNN, and then classifies objects using multiple support vector machines (SVMs). This method allows for more precise localization of defects. The introduction of a Mask R-CNN [10] further improves accuracy by performing pixel-level classification within the object regions generated by the R-CNN stage. Bolya et al. [11] evaluated the speed and performance of various instance segmentation methods on the COCO dataset, showing that while Mask R-CNN achieves high accuracy, it is relatively slower than single-stage detectors. Zhang et al. [12] developed a specular surface inspection technique using Mask R-CNN and a low-cost planar LED light source. However, due to issues caused by specular reflections, the F-score was limited to 0.78. Furthermore, since the system was implemented on Google Collaboratory (Google Collab), full automation of the inspection process was not feasible.

On the other hand, YOLOv4, proposed by Bochkovskiy et al. [13] in April 2020, is a one-stage detector that directly predicts bounding boxes and class labels from feature maps using CNNs, achieving high computational speed. For example, Liu et al. [14] reported real-time defect detection on metal surfaces using an improved YOLOv4; Xie et al. [15] applied YOLOv4 with an integrated attention mechanism for metal surface defect detection; and Li et al. [16] employed YOLOv4 in a self-supervised framework for triaxial sensing, including localization, angle measurement, and defect detection. Although YOLOv4 has been widely adopted, it is less effective at detecting small objects and is primarily used in applications where real-time performance is critical, such as surveillance, autonomous driving, and drones.

In contrast, Mask R-CNN is a two-stage detector in which a region proposal network (RPN) first generates candidate regions, followed by classification, bounding box regression, and mask generation. While it is computationally slower than YOLOv4, Mask R-CNN excels at detecting small objects and complex shapes. Therefore, in applications where accuracy and detailed shape information are essential, such as medical image analysis and defect inspection in manufacturing, Mask R-CNN is often preferred. Given that

the aim of this study is to generate detailed defect masks, Mask R-CNN is considered more suitable than YOLOv4.

In this study, we develop an automated surface defect inspection system for glossy components that captures images using a coaxial fringe projection light source to address the issue of specular reflection. This system generates defect mask images and determines the presence or absence of defects using Mask R-CNN. Furthermore, to reduce the processing time required for defect inspection, we propose an improved inspection process in which the AI model weights are loaded in advance and the system is prepared before image acquisition begins. The effectiveness of the proposed approach was experimentally validated, and promising results were obtained.

The paper is organized as follows. Section 2 provides a detailed description of the surface defect inspection system, including its configuration, the principles behind the defect inspection process, image measurement unit attitude control, and photography. Additionally, we discuss the use of Mask R-CNN in the defect inspection unit. In Section 3, we present our experimental results and provide a discussion. Finally, Section 4 gives conclusions of our study.

2. Configuration of Surface Defect Inspection System.

2.1. Principle of surface defect inspection system. Figure 1 shows a schematic diagram of the surface defect inspection system for glossy parts created in this research. The system consists of an image measurement unit and a defect inspection unit. The inspection workstation runs Ubuntu 18.04 as its operating system. It is equipped with an Intel Core i7-9700 CPU, 16GB of memory, and an NVIDIA GeForce RTX 2060 GPU with 12GB of VRAM.

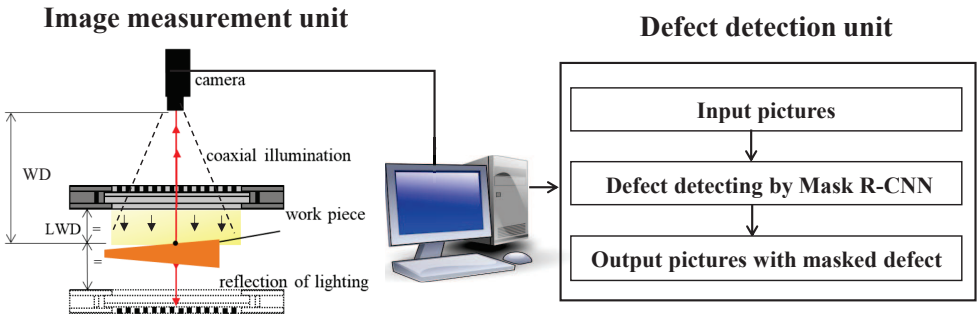


FIGURE 1. Schematic diagram of the surface defect inspection system

The image measurement unit consists of a camera, a coaxial fringe projection light source, and a component attitude control device. During the image measurement process, a light-dark pattern is projected onto the surface of the parts using the light source, and the reflected pattern is continuously captured by the camera as the parts are moved into different postures. If the surface is defect-free, a regular light-dark pattern is obtained. On the other hand, if there is a defect, such as a scratch, bump, or dent, a distorted light-and-dark pattern will be captured at the location of the defect.

The defect inspection unit employs Mask R-CNN, a type of deep learning algorithm. To perform defect inspection, a set of images taken from the part's surface is fed into the unit, which uses its learned Mask R-CNN model to identify any defects present. If no defects are detected, the input image is output as-is. However, if a defect is found, the image is output with a mask attached to the location of the defect. If only one masked image appears in the output inspection result images, the parts are deemed defective.

2.2. **Image measurement unit attitude control and photography.** Figure 2 shows examples of the parts measured in this study. As illustrated in the figure, Part A has a cylindrical shape, enabling complete surface imaging simply by rotating the part. In contrast, Part B has a complex three-dimensional geometry, requiring vertical and horizontal translations in addition to rotation to capture all surfaces. Since the inspection requirements differ depending on the type and function of each part, their acceptable defect sizes also vary. Part A is a gear shift lever knob located in the vehicle cabin. According to the visual inspection standards of automobile manufacturers, defects smaller than 0.2 mm are regarded as acceptable, while those larger than this threshold are considered defective. On the other hand, Part B is a decorative component of a side mirror, for which defects smaller than 0.5 mm are accepted under the same standards. Therefore, Part A requires the detection of smaller defects and consequently demands higher spatial resolution.

(a) A-parts (b) B-parts

FIGURE 2. Example parts to be measured

Figure 3 shows the flow of operation control for the image measuring unit used for Part A. The unit comprises a camera, a coaxial light source (always on), an Arduino Mega 2560, and a stepping motor, and operates in the sequence indicated by the arrows. Commands from the workstation are transmitted to the Arduino Mega 2560, which controls the stepper motors and camera accordingly via serial communication. For example, if the Arduino Mega 2560 receives the command "r" followed by a two-digit number, it stops the stepping motor at the angles obtained by dividing 360° by the number. Then, it sends a trigger signal to the camera to capture images. If "r10" is received, the system captures 10 images at every 36° rotation of the parts.

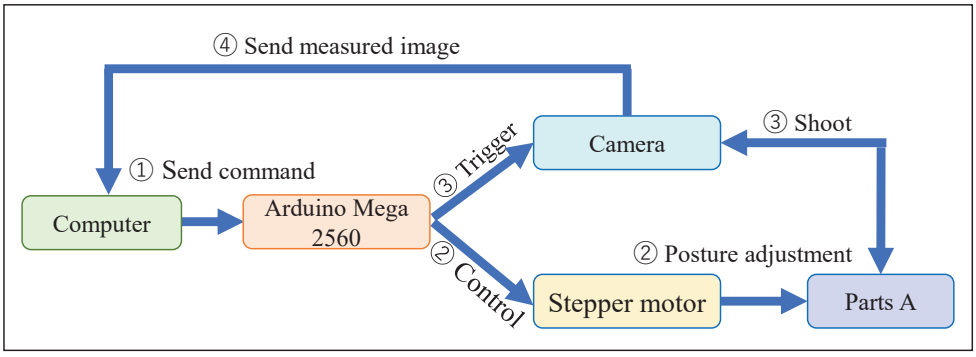


FIGURE 3. Flow of operation control of the image measurement unit for Part A

Figure 4 shows the operation control flow of the image measuring unit for surfaces of parts with 3D curved surfaces, such as the Part B. The unit consists of a camera, a coaxial light source, two Arduino Mega 2560s, six stepping motors, and six end stops. The stepping motors and end stops provide flexible attitude control.

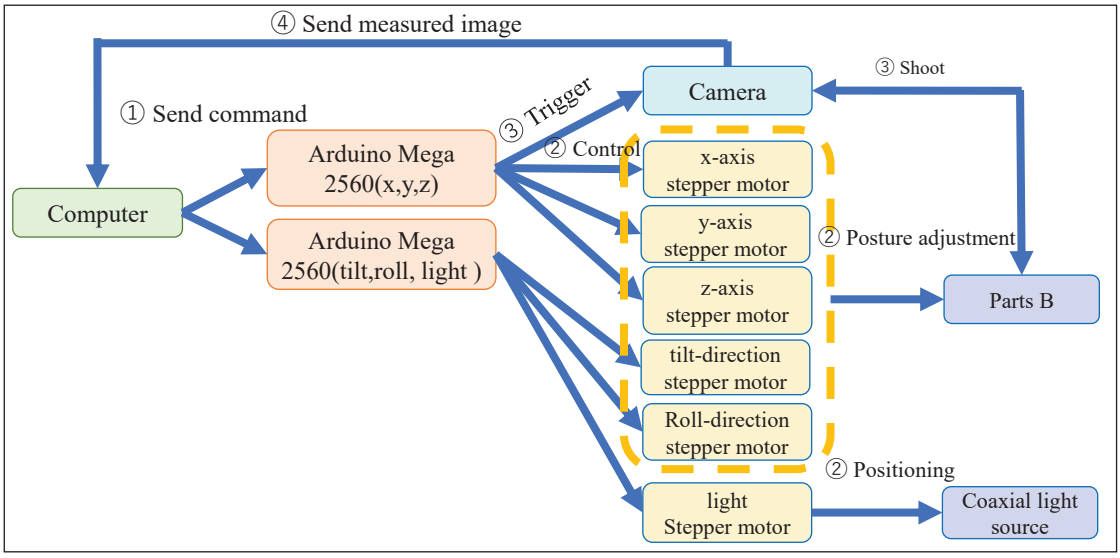


FIGURE 4. Operation control flow of the image measuring unit of the Part B

Translation in the x, y, and z directions, as well as rotation on the tilt and roll axes, enables positioning of the parts. The position of the coaxial light source is also adjustable. One Arduino Mega 2560 controls the motor that adjusts the (x, y, z) position, while the other controls the motor that adjusts the tilt, roll angle of the parts, and height of the coaxial light source. Both Arduinos are also used to calculate coordinates.

The movable limits of each axis are 300[mm] for x, 240[mm] for y, 200[mm] for z, 150[mm] for tilt, 300[mm] for roll, and 170[mm] for light source height. Motion control commands are specified using G-code, commonly used for controlling 3D printers. Table 1 presents the motion control commands.

TABLE 1. Command table for motion control of the image measurement unit of Part B

Command	Command description
t	Command to move the motor a little for operation test
s	Command to return the moving parts of the device to its initial position using the end stop
e	Command to exit control mode
p	After receiving the command, receive the coordinates of x, y, z (or tilt, roll, height of light) with 3 digits, and command the movable parts to move to the specified coordinates
b	Command to send camera trigger and shoot

To operate the system, commands are sent from the workstation to the Arduino Mega 2560. For instance, after receiving the "p" command, the Arduino Mega 2560 would receive coordinates representing the desired position and orientation of the parts, and move the parts accordingly. Once the "b" command was received, the Arduino Mega 2560 would trigger the camera to take a picture.

2.3. **Mask R-CNN used in the defect inspection unit.** Mask R-CNN [10] is a method derived from the CNN that has been developed following the evolution of several variants including R-CNN, Fast R-CNN, and Faster R-CNN. R-CNN [9] employs the selective search algorithm [17] to extract 2000 object-like rectangular regions from the input image. Each region is resized into a 227×227 pixel square and is fed into the CNN. The CNN comprises five convolutional layers and two fully-connected layers to produce a 4096-dimensional feature vector. A support vector machine (SVM) is then used for image classification to identify the objects within each region. However, to reduce this method's long processing time, Fast R-CNN and Faster R-CNN variants were introduced, the latter of which replaces selective search [17] with a CNN variant called RPN [18].

Mask R-CNN builds on Faster R-CNN by adding a segmentation task, resulting in more accurate object region estimation at a pixel-level classification. The object categories are identified within the regions obtained by the Faster R-CNN part. Figure 5 illustrates the Mask R-CNN architecture [10].

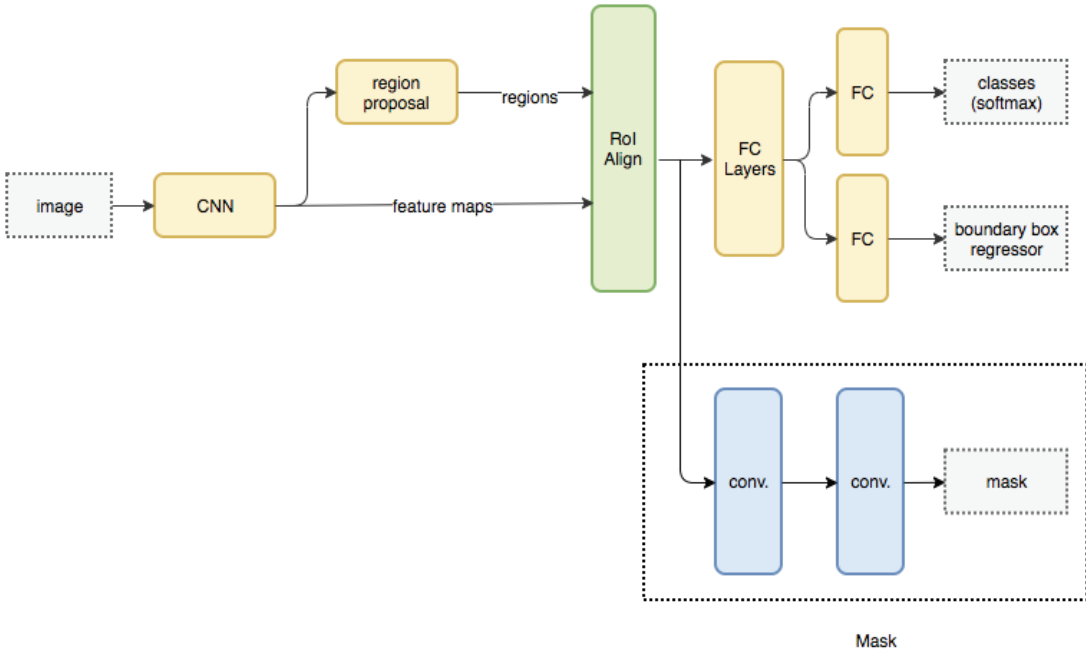


FIGURE 5. The structure of the Mask R-CNN [10]

Bolya et al. [11] investigated the speed and performance of various instance segmentation methods on the COCO dataset and presented the results shown in Figure 6. The horizontal axis of the figure represents the number of frames processed per second (FPS), while the vertical axis indicates segmentation accuracy, expressed as mean average precision (mAP). Generally, a segmentation accuracy of 30% or more in terms of mAP, as measured by COCO evaluation criteria, is considered to represent high prediction performance. As shown in Figure 6, there is a trade-off between processing speed and prediction accuracy. YOLACT++, proposed by Bolya et al. [11], offers almost the same functionality as Mask R-CNN, but achieves faster processing speed at the cost of slightly lower accuracy. In contrast, Mask R-CNN achieves higher accuracy but has the drawback of relatively slower processing speed. In this study, we take advantage of the high accuracy of Mask R-CNN, use a sample mode [19], which identifies defective regions on glossy surfaces by masking balloons, and train our own model accordingly. We then explore how to improve processing speed while maintaining the high accuracy of Mask R-CNN.

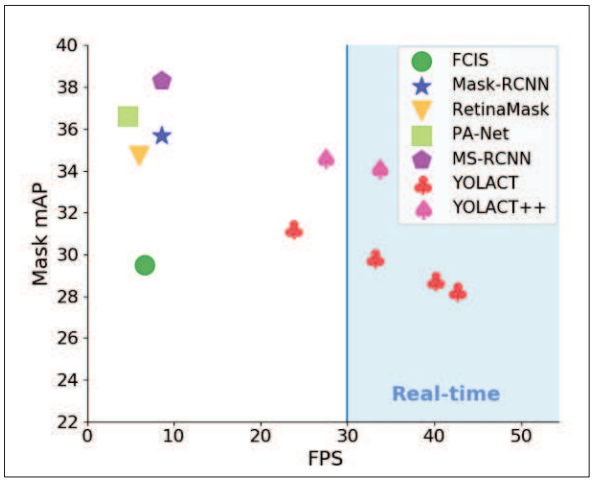


FIGURE 6. Speed-performance trade-off for various instance segmentation methods [11]

3. Experimental Results and Discussion.

3.1. **Configuration of defect inspection equipment.** Figure 7 depicts the defect inspection equipment used for Part B. The camera utilized was a Baumer HXG20 (Release 2), manufactured by Baumer, with the Lens 1: PL5014 lens made by Pixco shown in Table 2. Additionally, a coaxial fringe projection light source was used, specifically the LFX3-150SW-PT-A produced by CCS Corporation.

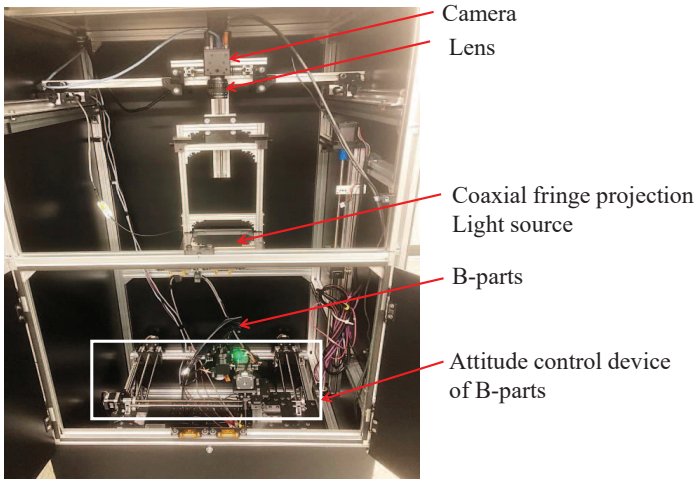


FIGURE 7. Appearance of defect inspection equipment for Part B

As shown in Figure 7, the coaxial light source is positioned on the same axis as the camera and the object, with the light source installed between them. The product has a striped pattern printed on the light guide diffusion plate with a pitch width of 1 mm, which receives light from the LED in the black edge part and illuminates as a light source. This mechanism projects a striped pattern onto the surface of the glossy component. An example photograph of Part B is shown in Figure 8(a), while Figure 8(b) shows a comparison image taken with a planar LED light source. Table 3 lists the optical parameters of the measurement unit shown in Figure 1, where Lens 1 was used for capturing images of Part B. As can be seen by comparing Figures 8(a) and 8(b), the coaxial fringe projection light source generates fringe patterns that highlight defects that were not visible under the planar light source.

TABLE 2. Specifications of the two lenses used in the experiments, showing the nominal optical and mechanical parameters provided by the manufacturers

Parameter	Lens 1	Lens 2
Focal length	50 mm (fixed)	28-80 mm
Maximum aperture (open f-number)	f/1.4	f/3.5-5.6
Minimum aperture	f/16	f/22
Minimum focusing distance (MFD)	300 mm	700 mm
Zoom ratio	1 (fixed focal length)	1:08
Mount type	C-mount	C-mount
Filter diameter	–	–
Lens type	Manual focus/iris	Manual focus/iris

* Lens 1: Pixco PL5014; Lens 2: TAMRON AF 28-80 mm F/3.5-5.6 Aspherical.

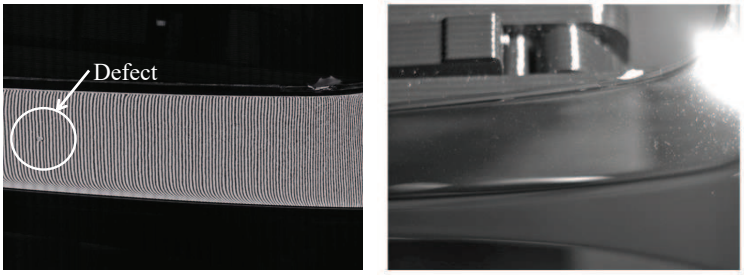


FIGURE 8. Comparison of Part B images taken using Lens 1: (a) With a fringe projection light source and (b) with a planar LED light source

TABLE 3. Optical parameters of lenses during the experiment

Parameter	Lens 1	Lens 2
Focal length [mm]	50	70
Working distance (WD) [mm]	720	640
Aperture (f-number) [-]	4.0	6.0
Magnification M [-]	0.0746	0.1429
Spatial resolution [μm/px]	73.7	38.5
Depth of field (DOF) [mm]	8.49	3.70

* Pixel pitch of the Baumer HXG20 sensor is 5.5 μm.
* DOF was estimated using the formula $DOF \approx 2Nc(1+M)/M^2$, where c denotes the circle of confusion equal to the pixel pitch.
* WD was measured from the front element of the lens to the target surface.

Figure 9 shows the posture control unit of the defect inspection system for Part A. Although this device uses a different posture control mechanism from that of Part B, it employs the same camera as used for Part B. An example image of Part A captured using the system is shown in Figure 10(a). For comparison, Figure 10(b) shows an image taken with a planar LED light source. In both cases, the Lens 2 specifications are provided in Table 2, and the corresponding optical parameters and shooting conditions are listed under the "Lens 2" entry in Table 3. As seen in the comparison between Figures 10(a) and 10(b), the coaxial fringe projection light source generates fringe patterns that make the defects more clearly visible.

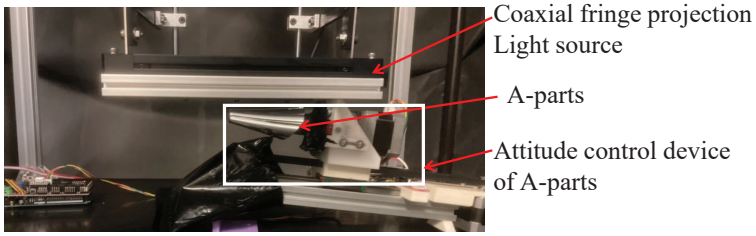


FIGURE 9. Component attitude control device for Part A

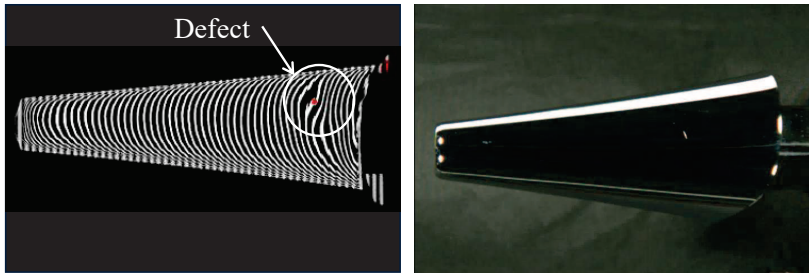


FIGURE 10. Comparison of Part A images taken using Lens 2: (a) With a fringe projection light source and (b) with a planar LED light source

To quantitatively evaluate the applicability of each lens, their spatial resolutions were compared based on the parameters listed in Table 3. As shown in Table 3, the spatial resolution of Lens 1 (with a focal length of 50 mm) is 0.0737 mm/px, while that of Lens 2 (with a focal length of 70 mm) is 0.0385 mm/px. Based on these values, defects of 0.2 mm and 0.5 mm correspond to approximately 3 pixels and 7 pixels for Lens 1, and about 5 pixels and 13 pixels for Lens 2, respectively. Since defect detection requires a minimum width of at least 5 pixels, Lens 2 is considered suitable for inspecting Part A, which involves smaller defects, whereas Lens 1 is appropriate for Part B, which contains relatively larger defects.

3.2. **Defect inspection part construction by Mask R-CNN using transfer learning.** This study employs transfer learning to mitigate the data requirements for training a Mask R-CNN based defect detection system. Transfer learning leverages knowledge from a source domain (where abundant data is available) to a target domain (often with limited data), thereby reducing the required amount of training data and accelerating convergence. Our model was initialized with weights from a Mask R-CNN model pre-trained on the large-scale Microsoft COCO dataset, which comprises over 330000 images. This pre-trained model was subsequently fine-tuned using our limited dataset of glossy surface defect images.

Our transfer learning protocol was specifically customized to address the primary challenge of detecting minute surface defects. The model was adapted as follows: the initial layers of the backbone network were frozen to retain generic, low-level feature extractors learned from COCO. In contrast, the task specific heads including the region proposal network (RPN), the classifier, and the regression modules were fine-tuned to recognize the new defect categories. Crucially, to improve the sensitivity to small defects, we reconfigured the RPN's anchor sizes to be commensurate with the scale of the target objects, which often measure only a few pixels in diameter.

To construct and evaluate a defect inspection system based on Mask R-CNN, it is necessary to prepare a dataset consisting of training, validation, and test images. In this

study, two types of datasets, Case 1 and Case 2 were prepared, as shown in Tables 4 and 5.

Case 1, presented in Table 4, is a mixed dataset composed of images of Part A and Part B: 372 images of Part A and 238 images of Part B, totaling 610 images. In this case, Lens 1 was used, and the parameters of the measurement unit and shooting conditions are listed under the "Lens 1" entry in Table 3. The dataset was divided into training and validation sets at a 7 : 3 ratio. In both sets, the number of defective and non-defective images was balanced at a 1 : 1 ratio. For the test data, 200 images were used for Part A (184 non-defective, 16 defective), and 28 images for Part B (14 non-defective, 14 defective).

TABLE 4. Dataset specifications of Case 1

Dataset	Part A	Part B	Total
Training data	282	154	436
Verification data	90	84	174
Total	372	238	610

Case 2, shown in Table 5, consists of four datasets of Part A with varying ratios of defective and non-defective images. Each dataset was split into training and validation sets at a 7 : 3 ratio. The test dataset contained a total of 88 images, comprising 44 non-defective and 44 defective images. In this case, Lens 2 was used, and the corresponding parameters of the measurement unit and shooting conditions are listed under the "Lens 2" entry in Table 3.

TABLE 5. Dataset specifications of Case 2

Dataset	Dataset-1	Dataset-2	Dataset-3	Dataset-4
Defective	162	168	171	168
No defects	107	84	52	0
Proportion of no defects	0.40	0.33	0.23	0.0

An example of inspection results obtained by the inspection system proposed using dataset Case 1 in this study is shown in Figure 11. Figure 11(a) shows the inspection result for Part A, while Figure 11(b) shows the result for Part B. As can be seen from the figures, defects were detected in both Part A and Part B, and are indicated by red masks.

3.3. **Evaluation index.** Quantitative evaluation is essential in defect inspection. To this end, we use a pre-defined dataset as the correct answer set and obtain predictions from the learning model. The resulting defect inspection outcomes can be classified into four categories using the evaluation index presented in Table 6.

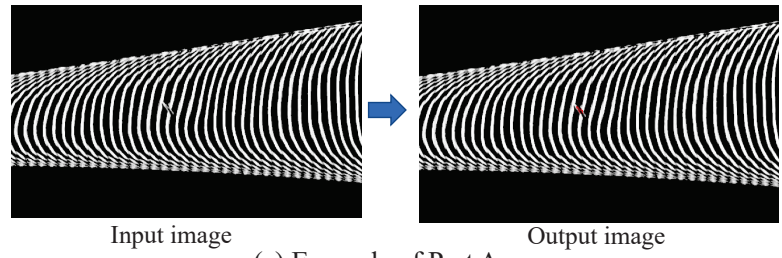
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

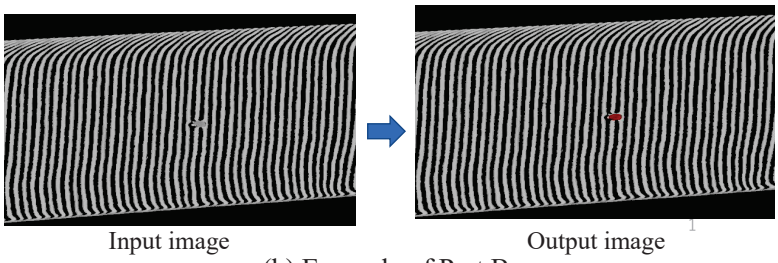
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Accuracy is the percentage of correctly predicted results out of the total results. The precision refers to the percentage of correct positive predictions out of all positive cases.



(a) Example of Part A



(b) Example of Part B

FIGURE 11. Example of defect detection results by Mask R-CNN using dataset Case 1

TABLE 6. Evaluation index

Correct label/prediction result	Defective	Normal
Defective	True Positive: TP	False Negative: FN
Normal	False Positive: FP	True Negative: TN

The recall, on the other hand, measures the percentage of correct positive predictions out of all actual positive cases. The F-score is a harmonic mean of precision and recall. In the context of defect detection, achieving high precision is essential to minimize the risk of missing defects. Therefore, models with high precision are preferred over those with lower precision.

3.4. **Experimental results and discussion.** As defined earlier, accuracy was used as the evaluation parameter. Using this parameter, the inspection system's performance was evaluated across different datasets, and the final evaluation system was constructed under the condition corresponding to the highest accuracy (epoch).

We first present the performance evaluation of the defect inspection system built using Mask R-CNN trained on Case 1 (the mixed dataset of Parts A and B), as shown in Table 7. The training was conducted on Google Colab with the number of epochs set to 50, which was confirmed to be sufficient for the convergence of the loss function. The inspection system was constructed under the best accuracy condition (Epoch = 49).

The results indicate that Part B achieved perfect scores of 1.00 in accuracy, precision, and F-score, demonstrating exceptionally high detection performance. In contrast, for

TABLE 7. Experimental results of dataset Case 1

	Part A	Part B
Accuracy	0.92	1.00
Precision	0.40	1.00
Recall	0.12	1.00
F-score	0.19	1.00

Part A, while the accuracy was 0.92, both precision and F-score remained low at 0.40 and 0.19, respectively. This reveals a significant issue with a high rate of false positives for Part A.

An analysis of the images used in Case 1 revealed that the defects in Part A were significantly smaller than those in Part B. The spatial resolution of the lens used in Case 1 (focal length: 50 mm) was 0.0737 mm/px. At this resolution, the minute defects in Part A were difficult to distinguish from the background curved stripe patterns. This likely led to an increase in false positives, resulting in low precision and F-score. Therefore, employing an optical system capable of higher-resolution imaging is essential for accurately detecting minute defects such as those found in Part A.

Next, we describe the experimental setup and results for Case 2. In this case, a high-resolution lens (Lens 2) with a focal length of 70 mm and a spatial resolution of 0.0385 mm/px was employed. The model was trained on Google Colab, and the number of epochs was set to 50, as this value was confirmed to be sufficient for the convergence of the loss function. Figure 12 illustrates the transitions of loss and accuracy during the training process for the four datasets. From Figure 12, it can be observed that each dataset achieved its highest accuracy at a different epoch.

The performance evaluation results of the inspection system under the best accuracy conditions are summarized in Table 8 (performance metrics) and Table 9 (confusion

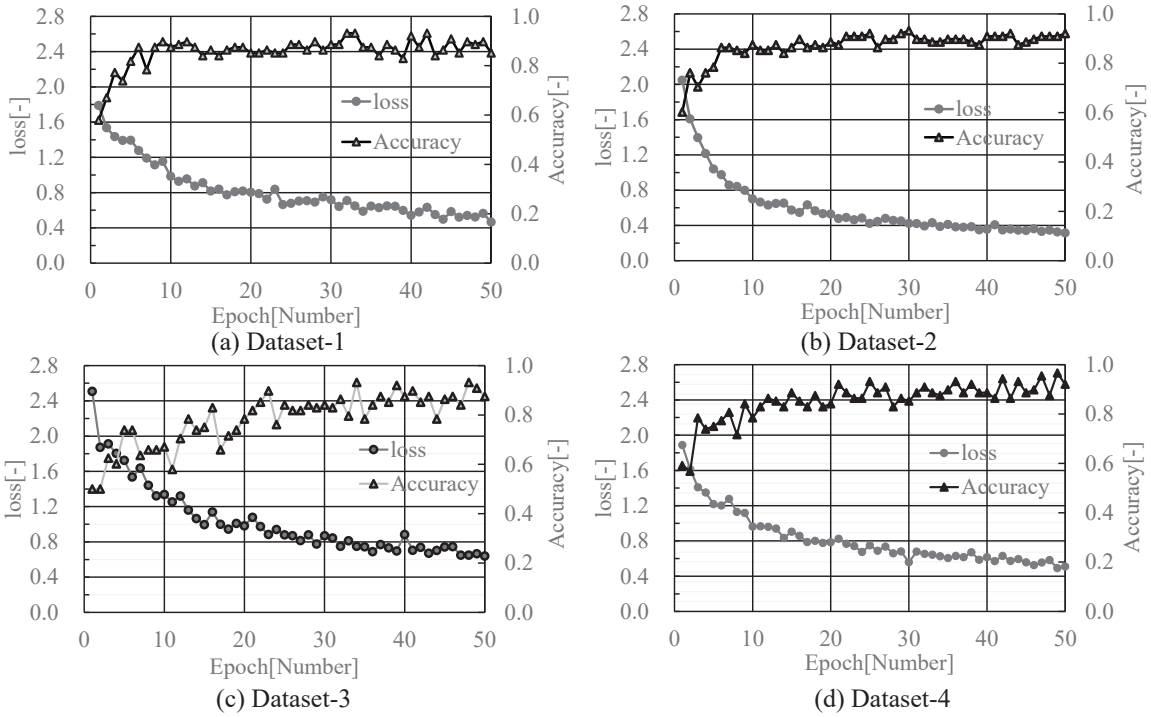


FIGURE 12. Transition of the loss and accuracy during the training process for the four datasets

TABLE 8. Experimental results of dataset Case 2 obtained at best accuracy condition

	Dataset-1	Dataset-2	Dataset-3	Dataset-4
Accuracy	0.93	0.93	0.93	0.96
Precision	0.95	0.97	0.97	1.00
Recall	0.91	0.89	0.89	0.93
F-score	0.93	0.93	0.93	0.96

TABLE 9. Confusion matrices of the four experimental conditions

(a) Dataset-1 (Epoch = 42)

Correct/Prediction	Defective	Normal
Defective	40	4
Normal	2	42

(b) Dataset-2 (Epoch = 30)

Correct/Prediction	Defective	Normal
Defective	39	5
Normal	1	43

(c) Dataset-3 (Epoch = 34)

Correct/Prediction	Defective	Normal
Defective	39	5
Normal	1	43

(d) Dataset-4 (Epoch = 49)

Correct/Prediction	Defective	Normal
Defective	41	3
Normal	0	44

matrices). As shown in Table 8, all datasets in Case 2 exhibited consistently high performance, with accuracy exceeding 0.93, precision above 0.95, and F-scores greater than 0.93. These results strongly indicate that the high spatial resolution of Lens 2 was effective in improving inspection performance, particularly for detecting minute defects in Part A that were difficult to identify in Case 1.

Notably, Dataset 4, which excluded non-defective images, achieved the best performance (accuracy: 0.96, precision: 1.00, F-score: 0.96). The confusion matrix in Table 9(d) shows that all 44 defective images were correctly classified as "Defective" (TP = 44, FN = 0), resulting in a perfect precision of 1.00. Moreover, the number of false positives (FP) for normal images was lower than in the other datasets, leading to a recall of 0.93. This suggests that removing normal images from the training data enhanced model specialization and reduced false positives. A plausible explanation is that by focusing learning solely on defective features, the model became less susceptible to confusion caused by background noise.

A remaining challenge is the improvement of recall. As indicated in Table 8, recall was relatively lower than the other metrics across all datasets, preventing perfect accuracy from being achieved. Since recall reflects the system's ability to detect all actual defective instances without omission, augmenting the training data with a greater variety and number of defective samples is expected to improve the model's generalization capability and detection sensitivity, ultimately leading to enhanced overall inspection accuracy.

To shorten the processing time required for defect inspection, we revised the inspection procedure. Figure 13(a) shows the conventional inspection process, while Figure 13(b) illustrates the improved version. As shown in Figure 13(a), the conventional method keeps the workstation idle until the imaging of surface defects on a Part is completed. Subsequently, the image undergoes binarization, and the AI model weights are loaded, followed by defect inspection. The same procedure is repeated for the second and subsequent Parts. In contrast, the improved method shown in Figure 13(b) preloads the AI model weights and completes system initialization before imaging begins. For the first Part, images are captured and each image is processed through binarization and then analyzed for defects using AI. From the second Part onward, since the AI model has already been loaded and the system is fully prepared, the process can proceed directly to image capture and inspection.

As a result, in the conventional method shown in Figure 13(a), the average processing time required to inspect ten images per Part was approximately 18.53 seconds. In contrast, with the improved method shown in Figure 13(b), although the initial Part still took about 18.53 seconds, the average processing time for subsequent Parts was reduced to

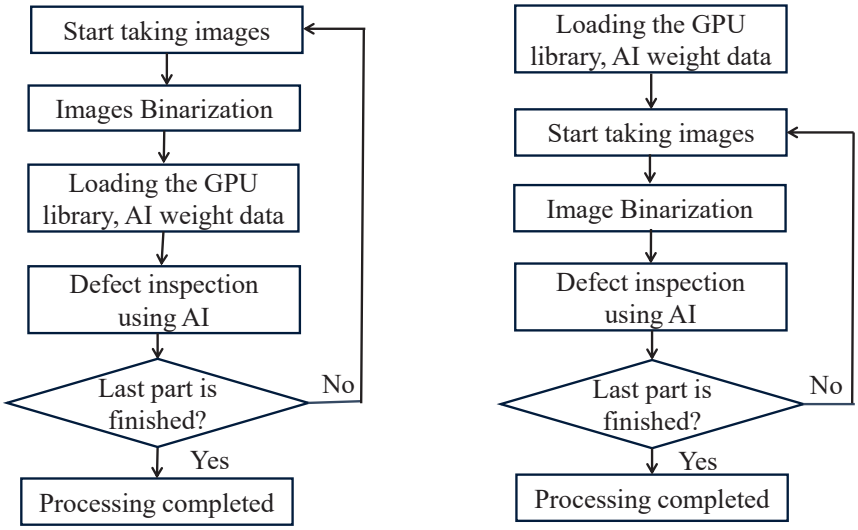


FIGURE 13. Comparison between the conventional and improved defect inspection processes

approximately 9.60 seconds per ten images. This improvement effectively doubled the inspection speed.

4. **Conclusion.** In this study, we proposed an automated inspection system for glossy parts that utilizes a coaxial fringe projection light source to capture surface images, applies masking to defects, and uses Mask R-CNN to classify the presence or absence of defects. The system was evaluated using two different types of parts: Part A and Part B, with the following results.

1) For Part A, the automated defect inspection system successfully completed defect detection with an average processing time of 18.53 seconds for the first Part (consisting of 10 images) and approximately 9.60 seconds for each subsequent Part. This improvement in speed is attributed to the strategy of preloading the AI model weights and preparing the inspection system before initiating image capture.
2) In the dataset used in Case 1, Part A showed a very low F-measure of 0.19, whereas Part B achieved an F-measure of 1.00, indicating high precision. This discrepancy is thought to result from the fact that the defects in Part A were smaller than those in Part B. For the same Part A, the dataset used in Case 2 achieved a significantly higher F-measure of 0.96. The key difference between the two cases was the lens used, and the results indicate that a telephoto lens capable of capturing magnified images is effective for detecting small defects.
3) With regard to the training dataset, it was found that using only defective images yielded higher diagnostic accuracy than using a mixed dataset of defective and non-defective images. This is likely because non-defective images act as noise during the Mask R-CNN training process. Since non-defective images lack significant features, they may reduce training efficiency.

For future work, it is necessary to perform a stress test on 10000 components under an ISO-standard factory environment to quantify the impact of ambient light, contamination, and vibration on the structural similarity index (SSIM) and the false alarm rate.

REFERENCES

[1] N. Kanno, Defect detection technology for mirror-coated products: Development of new defect detection technology by "variable curve matching method", *Journal of Japan Plastic Industry Federation*, vol.64, no.7, pp.22-25, 2013 (in Japanese).

[2] Y. Nakamura, Defect detection based on variance of the surface normal direction using a ring-lighting system, *Proc. of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)*, Paper no.336, pp.1-5, 2015.

[3] S. Höfer, J. Burke and M. Heizmann, Infrared deflectometry for the inspection of diffusely specular surfaces, *Advanced Optical Technologies*, vol.5, nos.5-6, pp.377-387, 2016.

[4] H. Wakizako and Y. Mori, Study of visual inspection for glossy parts, *The Japanese Journal of the Institute of Industrial Applications Engineers*, vol.4, no.2, pp.45-49, 2016 (in Japanese with English abstract).

[5] M. Hoshino et al., Study on examination for appearance of the luster part using the stripe pattern, *Proc. of Tokai Engineering Complex 2019*, pp.1-2, 2019 (in Japanese).

[6] B. N. Mengesha, A. C. Grizzle, W. Demisse, K. L. Klein, A. Elliott and P. Tyagi, Machine learning-enabled quantitative analysis of optically obscure scratches on nickel-plated additively manufactured (AM) samples, *Materials*, vol.16, no.18, 6301, pp.1-12, 2023.

[7] R. Wang, W. Du and Q. Jiang, Quantitative estimation method for complex part surface defects based on multimodal information fusion, *Complex & Intelligent Systems*, vol.11, no.260, pp.1-27, 2025.

[8] Z. Zhang, B. Zhang, T. Akiduki, T. Mashimo and T. Yu, Research on surface defects detection of reflected curved surface based on convolutional neural networks, *ICIC Express Letters, Part B: Applications*, vol.10, no.7, pp.627-634, 2019.

[9] R. Girshick, J. Donahue, T. Darrell and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proc. of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, pp.580-587, 2014.

[10] K. He, G. Gkioxari, P. Dollar and R. Girshick, Mask R-CNN, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.42, no.2, pp.386-397, 2020.

[11] D. Bolya, C. Zhou, F. Xiao and Y. J. Lee, YOLACT++: Better real-time instance segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.44, no.2, pp.1108-1121, 2022.

[12] Z. Zhang, T. Shirai and T. Akiduki, Development of a surface defect inspection method and system by deep learning, *ICIC Express Letters, Part B: Applications*, vol.13, no.8, pp.827-835, 2022.

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

[14] Y. Liu, Q. Wang, H. Zhang, Y. Liu and K. Zhao, Real-time defect detection of metal surface based on improved YOLOv4, *International Journal of Innovative Computing, Information and Control*, vol.18, no.4, pp.1329-1338, 2022.

[15] X. Xie, C. Li, Y. Liu, J. Song, J. Ahn and Z. Zhang, Application of YOLOV4 algorithm with integrated attention mechanism in metal surface defect detection, *International Journal of Innovative Computing, Information and Control*, vol.19, no.2, pp.447-463, 2023.

[16] X. Li, D. Zhou and X. Shao, A self-supervised method for trilateral sensing: Localization, angular measurement, and defect detection, *International Journal of Innovative Computing, Information and Control*, vol.21, no.2, pp.457-467, 2025.

[17] J. R. R. Uijlings, K. E. A. Sande, T. Gevers and A. W. M. Smeulders, Selective search for object recognition, *International Journal of Computer Vision*, vol.104, no.2, pp.154-171, 2013.

[18] 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 28 (NIPS 2015)*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett (eds.), Curran Associates, Inc., 2015.

[19] C. Q. Tan et al., A survey on deep transfer learning, *Artificial Neural Networks and Machine Learning (ICANN 2018)*, pp.270-290, 2018.

Author Biography

Zhong Zhang received his Ph.D. degree in Engineering from Okayama University, Japan, in 1993. He worked at Okayama Prefectural University, Japan, and the Industrial Technology Center of Okayama Prefecture, Japan, and served as a visiting researcher at the University of Melbourne, Australia, from 1998 to 1999. He was a Professor at Toyohashi University of Technology, Japan, from 2004 to 2019 and at Hiroshima Institute of Technology, Japan, from 2020 to 2024. His research interests include vibration and noise measurement, signal processing, fault diagnosis, wavelet transform, and AI-based intelligent systems. He is currently a Professor at Aichi Sangyo University, Japan.

Keisuke Matsuoka received his B.E. degree from the Department of Intelligent Mechanical Engineering, Hiroshima Institute of Technology, Japan, in 2014. He is engaged in research on image processing. He is currently working as an engineer at DISCO Corporation.

Jin Qi received his M.E. degree from Beijing University of Posts and Telecommunications, China, in 2009 and completed the doctoral coursework in the Graduate School of Global Information and Telecommunication Studies at Waseda University, Japan, in 2013. Since 2023, he has been a Lecturer in the Department of Smart Design, Faculty of Design Engineering, Aichi Sangyo University, Japan. His research interests include wireless communications, computer networks, and artificial intelligence.

Toshihiro Okuyama is with Tokio Marine & Nichido Fire Insurance Co., Ltd. His responsibilities include research, investigation, and improvement activities related to accident prevention in the fields of import-export logistics and insurance.