ICIC Express Letters
Part B: Applications
Volume **13**, Number **8**, August **2022**

ICIC International ©2022 ISSN 2185-2766
pp. **827–835**

DEVELOPMENT OF A SURFACE DEFECT INSPECTION METHOD AND SYSTEM BY DEEP LEARNING

Zhong Zhang[1], Takahiro Shirai[2] and Takuma Akiduki[2]

[1]Department of Intelligent Mechanical Engineering
Hiroshima Institute of Technology
2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193, Japan
t.sho.g4@cc.it-hiroshima.ac.jp

[2]Department of Mechanical Engineering
Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku-cho, Toyohashi 441-8580, Japan
{ shirai; akiduki }@is.me.tut.ac.jp

Received December 2021; accepted February 2022

Abstract. *There are many product parts with strong specular reflection such as metal-plated parts. However, specular highlighting by illumination has a brightness and shape similar to the reflected light generated by the unevenness of defects, which hinders defect inspection. Currently, automation of inspections has not progressed, and visual inspection by inspectors is the mainstream. Such inspectors make full use of their senses which have been trained by learning, and they are able to distinguish specular highlights and defects that are seemingly similar to a very complicated and flexible response. In this research, we develop a surface defect inspection system for mirror-finished parts that imitates the learning process of inspectors and employs deep learning. Then, we construct a surface defect inspection system for mirror surface parts using a Mask R-CNN, which can utilize transfer learning with a small number of samples, and which is also a type of deep learning. As a result of evaluating the inspection system with 1000 sample data images, the defect detection rate (Recall rate) is 0.816, the conformance rate (Precision rate) is 0.753, and the F-measure is 0.783.*
Keywords: Deep learning, Specular reflection, Surface defect, Defect inspection system

1. **Introduction.** In the inspection of product surfaces with high mirror reflectance represented by metal-plated parts, it is required to make the imaging conditions uniform while the surrounding environment is reflected. Therefore, in most cases, it is necessary to reduce specular reflection from the light source and set the environment individually for each product. In particular, specular highlighting by illumination has a brightness and shape similar to the flicker caused by the unevenness of the defect area, which hinders the detection of defects on the surface to be inspected.

Many researchers have been enthusiastically researching the surface inspection of glossy parts. Kanno [1] detailed a small defect detection device for mirror-coated products using slit light, Nakamura [2] proposed ring-lighting to detect defects based on variance of the surface normal direction, Höfer et al. [3] proposed a mirror defect inspection device using infrared ellipsometry, Wakisako and Mori [4] proposed appearance inspection technology for glossy plastic parts using a stripe pattern projection method, and Hoshino et al. [5] proposed an appearance inspection method for glossy parts using a striped pattern coaxial light source. However, these methods have problems such as increased cost due to the introduction of special lighting and equipment compared to ordinary parts. In small and medium-sized enterprises, the current situation is that visual inspection by inspectors is prioritized in consideration of introduction costs and issues. Cao et al. [6] proposed a

DOI: 10.24507/icicelb.13.08.827

method that combined visual saliency detection and RPCA for detecting surface defects of wind turbine blades. Uneven illumination and Gaussian noise can be suppressed effectively by adding a noise term and a Laplacian regularization term to the basic Robust Principal Component Analysis (RPCA) model. However, this method has not yet been confirmed to be effective for surface inspection of glossy parts with strong specular reflection. It is thought that the inspector uses visual or other senses to discriminate specular highlights and defects that are seemingly similar to a very complicated and flexible response.

On the other hand, in deep learning, there are many cases where the method represented by the Convolutional Neural Network (CNN) can solve image recognition problems in which high generalization performance is required. Authors [7] constructed a check model by ensemble CNN using multiple Convolutional Neural Networks (CNN). However, there was a problem whereby the location of the found defect could not be identified by executing the inspection as a classification task based on the presence or absence of defects. Here, the Region based Convolutional Neural Network (R-CNN) [8] and the Faster R-CNN [9] were proposed, which searched for object candidates in images, applied a CNN to extracting features, and identified categories and objects position by a multiple Support Vector Machine (SVM). Furthermore, a great method, called Mask R-CNN [10] has been proposed and widely applied. It only extends the Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition, and is widely applied because it has excellent properties. Therefore, if the Mask R-CNN is used, defect detection can be regarded as an object search task, and because the defect region and location can be discovered, it is thought that conventional problems can be solved.

In this study, we selected a learning model and adopted a Mask R-CNN. Using it, we build an inspection system that automatically performs flexible surface inspections which have been performed by inspectors. Therefore, the correct answer data for the defective area is created and expanded, and used for learning and evaluation of the proposed inspection system. Then, using 1000 unseen evaluation data, the proposed inspection system is evaluated and the improvement of the F-measure indicating the inspection accuracy is confirmed.

The remainder of the paper is organized as follows. In Section 2, the principle of image feature extraction by deep learning is introduced; Section 3 details a new surface defect inspection system using Mask R-CNN; Section 4 is concerned with verification experiment results and discussion; Section 5 explains results and discussion. Finally, Section 6 gives conclusions and closing remarks.

2. **Principle of Image Feature Extraction by Deep Learning.**

2.1. **Review of CNN.** A Convolutional Neural Network (CNN) is constructed by connecting multiple layers. An example of the configuration is shown in Figure 1 [10]. The first half is composed of convolution and pooling operations, and performs the optimum

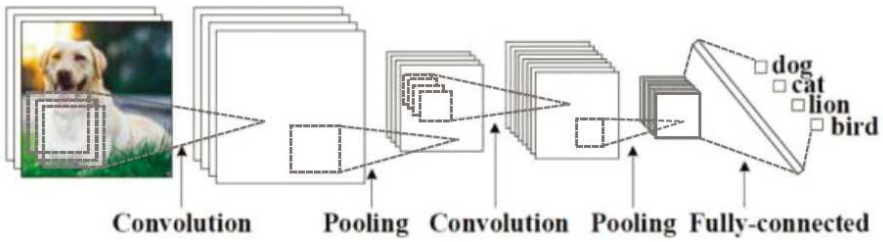


FIGURE 1. The structure of the CNN [10]

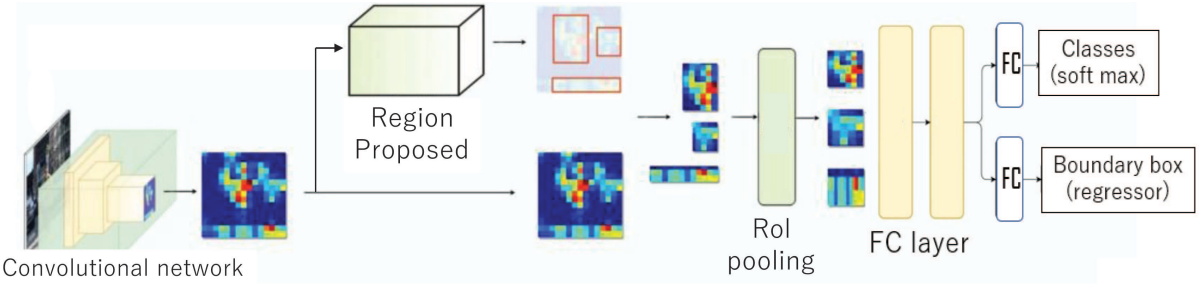
feature vector operation for identifying input data. After that, the feature vectors obtained by the first half are linked and identified by the classifier of the fully connected layer placed in the second half. The identification result is output for each class. Since the accuracy of each class is calculated at the end of the network, nodes for the class to be identified are required. For characteristic of the CNN, features can be extracted while maintaining the relationship of each part of data with a shape such as images and 3D data, and it is possible to have robustness against position movement.

Image recognition has the following two tasks.

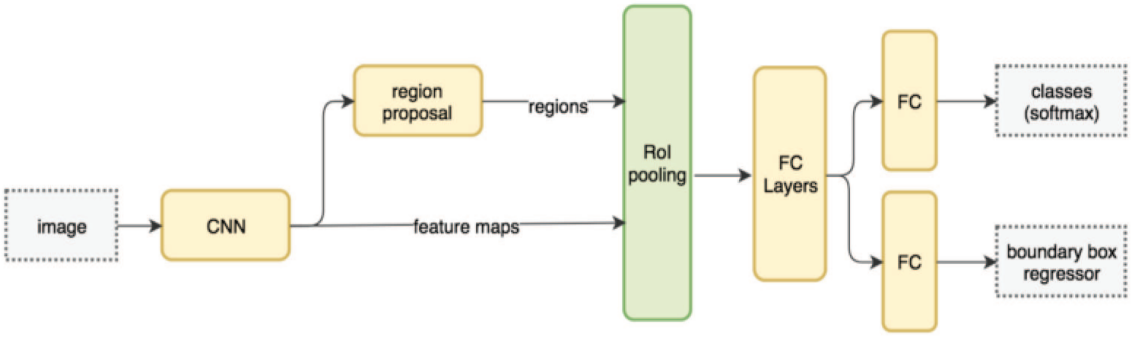
1) Image recognition task: What kind of object is it if it is an object?
2) Object detection task: Is an image area a background or an object?

The CNN only performs the first task, that is, it recognizes the entire input image. On the other hand, the Faster Region based Convolutional Neural Network (Faster R-CNN) based on the R-CNN performs the first task and then the second task [10].

2.2. **Faster R-CNN** [9]**.** Figure 2 [10] shows the (a) conceptual diagram and (b) flowchart of Faster R-CNN. As shown in the figure, the input image is first subjected to CNN to extract features, and then a feature map is output. Next, the region with the object is extracted by the region proposal based on the feature map. The actual region proposal can be composed of 3 to 4 layers of CNN and is small in size. Then, by the Region-of-Interest pooling layer (RoI pooling) that connects the region proposal and the CNN output, the feature map of the output from the CNN is cut out in the feature area extracted by the region proposal, and the area size is also adjusted. Finally, as with Faster R-CNN, each feature region is categorized by applying a classifier, and the position of the object is estimated by regression. This approach contributes to the improvement of accuracy and calculation efficiency. However, one issue mentioned of Faster R-CNN is the RoI pooling. When performing RoI pooling, the object detection task did not support pixel-to-pixel, but the purpose was to infer the bounding box, so there was no problem with some deviation. However, with segmentation, a slight deviation causes a problem.



(a) Conceptual diagram of Faster R-CNN



(b) Flow chart of Faster R-CNN

FIGURE 2. The structure of the Faster R-CNN [9]

2.3. **Mask R-CNN** [10]**.** The flowchart of the Mask R-CNN is shown in Figure 3. As shown in the figure, Mask R-CNN is an expanded model, which adds a branch that predicts the mask of an object to the Faster R-CNN in addition to the branch for detecting the existing bounding box. A new RoI Align layer is also proposed so that the mask boundary can be estimated accurately in order to overcome the deviation of the object position caused by RoI pooling. The RoI Align layer uses bilinear interpolation sampling learned from the Spatial Transformer [10] layer, and pools the features coordinated with sub-pixel accuracy by bilinear interpolation. The RoI Align layer is not a simple pooling, but instead it normalizes the image size. In this way, Mask R-CNN is easy to train, and the processing speed is not much different from Faster R-CNN. In addition, Mask R-CNN is easy to generalize to other tasks such as human pose estimation. So, in this study, we selected a learning model and adopted Mask R-CNN.

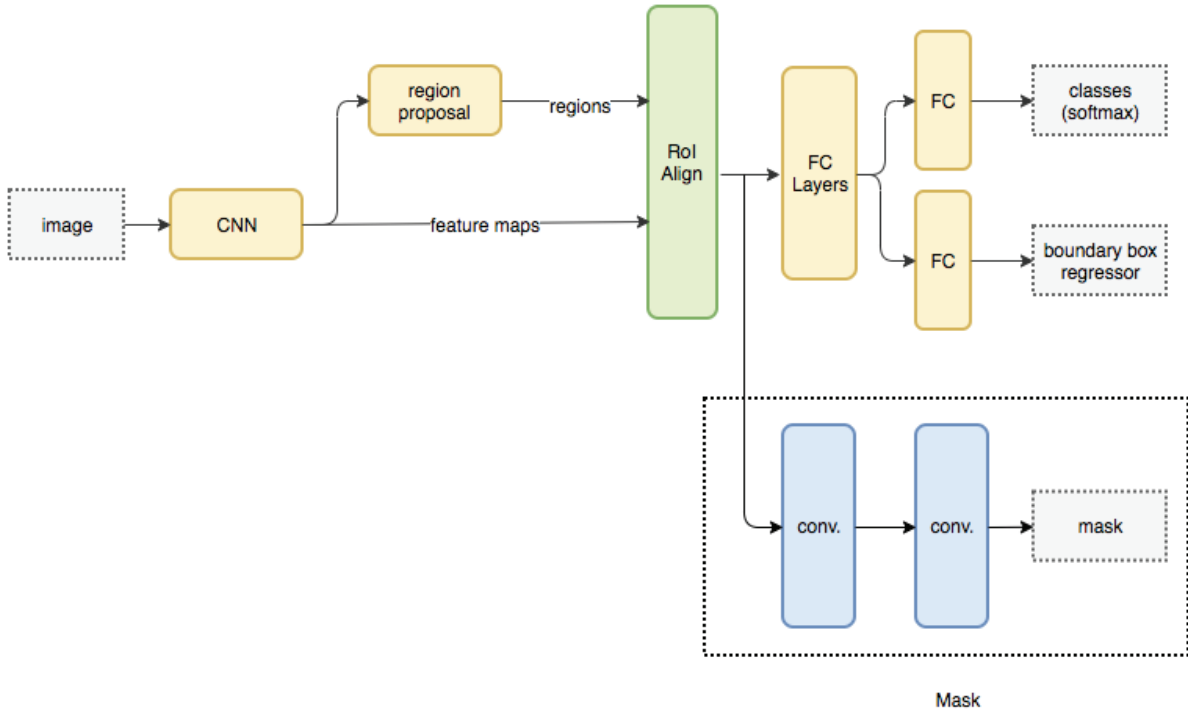


FIGURE 3. The structure of the Mask R-CNN [10]

2.4. **Transfer learning.** In order to make highly accurate predictions in deep learning, it is necessary to learn using a large amount of high-quality data, but in reality, only limited data may be available. Here, transfer learning is proposed to obtain highly accurate predictions by performing additional learning with a small amount of data and in a short time, using a trained model that has been fully trained in the same field [11].

This method has been successfully applied in the field of image recognition [12]. There are "features that should be captured in common" in various tasks in images. Therefore, the features in the data can be extracted by using the part without the final layer of the pre-trained model. At this time, the trained model is treated as a "feature extractor". In this case, the weight of the network that learned the features remains fixed, and only the weight of the new layer for solving the target task is determined by transfer learning, and image recognition is performed. This technique is called network-based transfer learning, and it is used in this study.

3. Proposal of a New Surface Defect Inspection System Using Mask R-CNN.

3.1. A new surface defect inspection system using Mask R-CNN.

So far, the authors have regarded the visual inspection of metallic luster parts as a classification task that classifies the presence or absence of defects from the entire inspection image [7]. However, it is not easy to construct the data so that only the defects are extracted from the common features of the sample image. Tao et al. [13] considered the defect inspection of metal produce surfaces using deep learning as an object detection task and achieved an Intersect over Union (IoU) score of 89.6[%]. Therefore, in this research, we consider defect detection as an object detection task and aim to recognize defects efficiently by enabling humans to indicate the defect area.

In addition to distinguishing between specular highlights and surface defects, the inspection system should have functions such as displaying the defect judgment position, capturing camera images, and listing images. The proposed inspection system is shown in Figure 4. As is shown in Figure 4, this system mainly consists of an image measurement unit and an intelligent defect judgment unit. In addition, the intelligent defect judgment unit by using the Mask R-CNN [10] consists of defect feature extraction, defect region identification, image list output including defect position and so on.

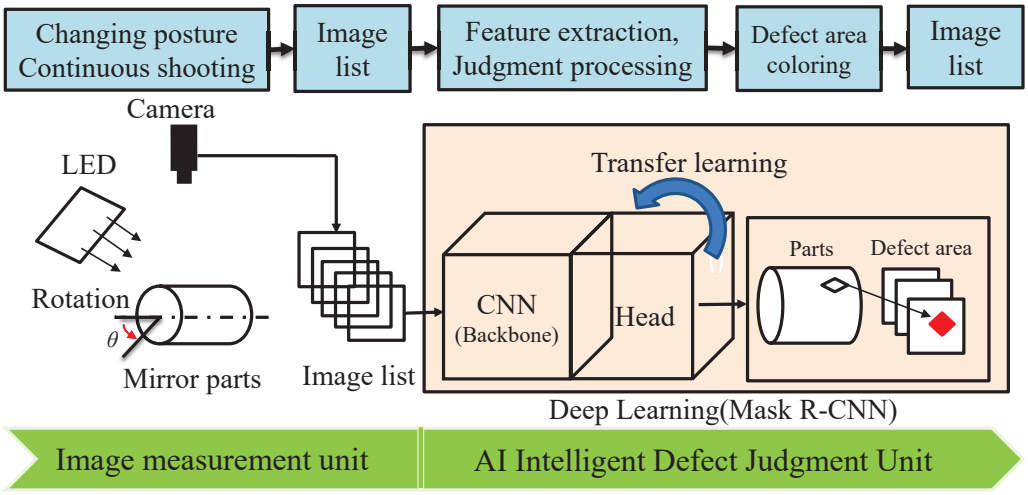


FIGURE 4. Schematic diagram of the proposed system

To construct the inspection system, learning was performed using a data set extended by image processing. In this case, transfer learning is adopted to reduce the number of learning cycles. As for the data set, it needs to be expanded in order to improve the detection accuracy of the defective area with a small number of samples. In addition, by adopting the Mask R-CNN as the learning model, it is necessary to consider the evaluation index for the area and select an evaluation method that can expect accurate feedback. In this study, in order to evaluate the effectiveness of the proposed method, by following the reference paper [16], the three indicators Recall rate (Sensitivity), Precision rate and F-measure were used.

3.2. Creating data set.

Therefore, in order to train the Mask R-CNN, a data set was created for object recognition focusing on defects on the surface of mirror-finished parts. Here, the shift lever knob shown in Figure 5 was coated with chrome plating. The defect area was indicated for 360 RGB images with surface defects with an image size of 480×640. In order to prevent the training model from excessively recognizing unintended patterns, it was decided that no instruction is given for defects whose representative length does not exceed 0.5 [mm] and whose pixel value in the image is 3 [pixel] or less. In addition, in

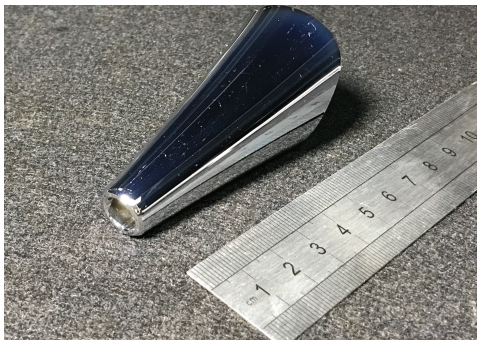


FIGURE 5. Target parts

TABLE 1. Data set specifications

Data set	Training (defect)	Verification (defect)	Total
Before expansion	360 (91)	72 (22)	432
After expansion	5400 (1365)	1080 (330)	6480
For evaluation	–	5400 (1521)	5400

TABLE 2. Learning condition

Model	Backbone Network (Layer)	Initial-Weight	Data-Number	Case
Mask R-CNN	ResNet-101	COCO	360	a
			5400	
		ImageNet	360	b
			5400	
	ResNet-50	COCO	360	c
			5400	
		ImageNet	360	d
			5400	

order to cope with overfitting, the image was rotated at random angles and the brightness was adjusted to expand the data set to 5400 sheets. Table 1 shows the specifications of the constructed training data set and verification data.

3.3. **Inspection system construction by transfer learning.** Here, the Mask R-CNN is trained using the data sets of before and after expansion processing, and the difference in the degree of learning is investigated. The progress of learning due to changes in the backbone networks were also investigated. In this study, transfer learning [12] is adopted to reduce the number of learning cycles of the Mask R-CNN. That is, based on the initial weight obtained by training the Mask R-CNN in advance with a general large-scale data set, additional training is performed with a small-scale data set. In this study, the number of learning cycles (Epoch) is set to 300. The learning conditions for the number of data set images, the total number of layers in the backbone network, and the initial weights are shown in Table 2.

As shown in Table 2, in order to investigate whether each condition from (a) to (d) is affected by the data set expansion, training is performed on the data sets before and after the expansion. In conditions (a) and (b), ResNet-101 was adopted as the backbone network. Compared to ResNet-50 under conditions (c) and (d), there are more network parameters in ResNet-101, so it is expected that extraction can be performed for more complex features due to the characteristics of CNN. However, on the contrary, there is a concern that learning will end early and overfitting will occur due to the large number of parameters.

For the initial weights, we used Mask R-CNN with each backbone network, which has been trained using two publicly available object search data sets, namely Microsoft COCO [14] which was used for conditions (a) and (c), and ImageNet [15] which was used for conditions (b) and (d). There are differences in the number of samples and the number of categories between the two, which is thought to cause differences in the formation of feature filters in the input-side layer of the backbone and in the learning status of the RPN (Region Proposal Network) [9].

4. **Verification Experiment Results and Discussion.**

4.1. **Evaluation index.**

Defect inspection requires an evaluation value to be evaluated quantitatively. Here, when the correct answer set is given in advance as a data set and the prediction by the learning model is obtained, the defect inspection result can be divided into four classifications by the confusion matrix shown in Table 3.

$$\text{Precision rate} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

The Precision rate is expressed as in Equation (1), and shows the ratio of the predictions made by the learning model when they are predicted correctly. It is an index to evaluate the accuracy of classification prediction in the checking model. The Recall rate (Sensitivity) is expressed by Equation (2), and represents the ratio that covers the correct answer data in the prediction output by the learning model. It is an index to evaluate the small number of oversights in the checking model. The learning model used in the inspection system needs to learn the features corresponding to unknown defects from a limited data set and acquire appropriate generalization performance. If it is overfitted to specific defect data, the Recall rate will decrease even if the Precision rate increases. On the contrary, when detailed data is adapted to reduce oversight, even if the Recall rate increases, misjudgment increases and the accuracy rate decreases. Here, the learning model to be adopted in the inspection system needs to be well-balanced between misjudgment and oversight tendencies, and to have both properties at a high level. The F-measure is the harmonic mean of the precision and recall, and is shown by Equation (3), which shows the balance between the two values.

Table 3. Confusion matrix

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

First, the Mask R-CNN-based inspection system is configured under each learning condition in which the number of layers in the backbone network, the initial weight, and the number of transfer learning data are set as shown in Table 2. Then, the F-measure of the configured system is calculated and evaluated, and the optimal combination of learning conditions is selected.

Transfer learning under each condition was performed 300 times. Thereafter, an inspection system was constructed using weights every 10 times, and 1000 evaluation data images with randomly selected defect areas were inspected, and the system was evaluated from the inspection results. At this time, the average value of the Precision rate and the Recall rate, and the F-measure were calculated for the predicted region in all the images of the evaluation data. The weight at the time when the F-measure showed a maximum

was selected, because the inspection requires a high level of precision and recall. The condition with the highest F-measure was obtained when ResNet-101 used the MS COCO pre-trained weights, and was trained 220 times by transfer learning using the extended data set. In this study, therefore, the weight obtained in this condition is adopted for the inspection system.

5. **Results and Discussion.** In order to confirm the effectiveness of the method using the Mask R-CNN proposed in this study, a comparison of conventional methods using an Ensemble Convolutional Neural Network (Ensemble CNN) [7] under the same conditions was conducted, and the differences in the precision, recall, and F-measure between the two methods were examined. The conditions and procedure for comparison are as follows.

1) The same pre-expansion data set is expanded by each method, and learning processing is performed using the obtained data set.
2) From the unknown evaluation data set, 500 images without defect regions and 500 images with one or more defect regions are inspected by the learning model when the optimal weighting factor for each method is selected.
3) Considering that the conventional method is a task of classifying one image according to the presence or absence of defects, the proposed method also judges the result based on the presence or absence of defects in one image. In other words, even with the proposed method, it is judged that there is a defect when there are one or more defective areas in the image, and it is judged as normal only when there is no defective area.
4) The evaluation of the inspection result is expressed as a confusion matrix when one image is judged by the presence or absence of defects, and the evaluation is performed by comparing the precision, recall, and F-measure of each method.

Table 4 shows the evaluation results of each method. From the evaluation index shown in Table 4, it can be seen that each value of the proposed method was improved compared to the conventional research method, and the defect detection performance was improved. An improvement of 0.173 was seen in the F-measure. This is thought to be because the proposed method selectively learns defects as regions, so that specular highlights and defect regions can be distinguished more clearly.

Table 4. Test results for each method

Method	Precision	Recall	F-measure
Using Mask R-CNN (Our method)	0.753	0.816	0.783
Using Ensemble CNN	0.632	0.590	0.610

6. **Conclusion and Remarks.** In this study, we constructed a surface defect inspection system for mirror-finished parts using a Mask R-CNN, which can utilize transfer learning with a small number of samples, and which is also a type of deep learning. The results obtained are as follows.

1) When using transfer learning to build a test system, it is necessary to select a model that has been trained with big data. In this study, the searching optimal learning model for the test system was evaluated by the F-measure. The condition showing the highest F-measure was pre-trained with MS COCO on ResNet-101 and 220 times with the expanded data set in our case of learning.
2) As a result of evaluation using 1000 unknown evaluation data images in the verification experiment of the configured inspection system, by using the proposed method, each evaluation value was improved compared to the conventional method (Ensemble CNN), and the F-measure was improved by 0.173.

3) The proposed surface defect inspection system using Mask R-CNN is effective for inspection of surface defects. It is considered that this is because the specular highlight and the defect region are more clearly distinguished by selectively learning the defect as a region.

However, in the current inspection system, the Recall rate is higher than the Precision rate, and it is considered that the part that is not the defect area tends to be over-detected. In the future, it will be necessary to investigate the factors individually for data with many over-detections and improve the data set.

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 (EECSS2015)*, 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. Wakizakoa 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*, pp.1-2, 2019 (in Japanese).

[6] J. Cao, G. Yang, X. Yang and J. Li, A visual surface defect detection method based on low rank and sparse representation, *International Journal of Innovative Computing, Information and Control*, vol.16, no.1, pp.45-61, 2020.

[7] 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.

[8] 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.

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

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

[11] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowledge and Data Engineering*, vol.22, no.10, pp.1345-1359, 2010.

[12] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, A survey on deep transfer learning, in *Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science*, V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis and I. Maglogiannis (eds.), Cham, Springer, 2018.

[13] X. Tao, D. Zhang, W. Ma and X. Liu, Automatic metallic surface defect detection and recognition with convolutional neural networks, *Applied Sciences*, vol.8, no.9, DOI: 10.3390/app8091575, 2018.

[14] T. Y. Lin et al., Microsoft COCO: Common objects in context, in *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars (eds.), Cham, Springer, 2014.

[15] O. Russakovsky, J. Deng, H. Su et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.*, vol.115, pp.211-252, 2015.

[16] Y. Fujita, H. Nakamura and Y. Hamamoto, Automatic and exact crack extraction from concrete surfaces using image processing techniques, *Journal of Japan Society of Civil Engineers*, vol.66, no.3, pp.459-470, 2010 (in Japanese with English abstract).