Automatic Mackerel Sorting Machine Using Global and Local Features

doi: 10.1109/ACCESS.2019.2917554

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 11, 2019, accepted May 6, 2019, date of publication May 17, 2019, date of current version May 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917554

Automatic Mackerel Sorting Machine Using Global and Local Features

YOSHITO NAGAOKA, TOMO MIYAZAKI, (Member, IEEE), YOSHIHIRO SUGAYA, (Member, IEEE), AND SHINICHIRO OMACHI, (Senior Member, IEEE)
Graduate School of Engineering, Tohoku University, Sendai 9808579, Japan

Corresponding author: Tomo Miyazaki (tomo@iic.ecei.tohoku.ac.jp)

This work was supported in part by the JSPS KAKENHI under Grant 16H02841 and Grant 18K19772.

ABSTRACT In Japan, blue and chub mackerels are often caught simultaneously, and their market prices are different. Humans need to sort them manually, which requires heavy labor. The demand for automatic sorting machines is increasing. The aim of this paper is to develop an automatic sorting machine of mackerels, which is a challenging task. There are two required functions. First, it needs localization of mackerels on a conveyor belt so that mackerels can be transported to destinations. Second, species classification is needed, but it is difficult due to similar appearance among the species. In this paper, we propose an automatic sorting machine using deep neural networks and a red laser light. Specifically, we irradiate red laser to the abdomen, and the shape of the laser will be circle and ellipse on the blue and chub mackerels, respectively. We take images and use neural networks to locate the whole body and irradiated regions. Then, we classify mackerels using features extracted from the whole body and irradiated regions. Using both features makes the classification accurate and robust. The experimental results show that the proposed classification is superior to the methods using either feature of irradiated or whole body regions. Moreover, we confirmed that the automatic mackerel-sorting machine performs accurately.

INDEX TERMS Convolutional neural networks, fish classification, fish localization.

I. INTRODUCTION

In Japan, mackerel is a popular fish used as a food ingredient, and a large amount of mackerels are caught every year. Blue and chub mackerels are often caught simultaneously. Because these two species are traded at different market prices, they should be sorted manually before shipment. However, there is a social problem that the number of workers capable of performing the classification decreases as their age increases. In addition, sorting the two species is not easy even for specialists. In general, the blue mackerels have a spotted pattern on their abdomen, while the chub mackerels do not. However, it is possible that the spotted patterns of the blue mackerels may be quite thin, or some of the chub mackerels may have a spotted pattern. Therefore, the ones with features of both these mackerels are categorized as *Hybrid* for convenience. We show these three species in Fig. 1.

Fish image classification has been studied since decades [1]–[7]. Generally, the existing methods address how to extract features from fish images and how to train classifiers [8]–[11]. To the best of the authors' knowledge, most existing methods are in software [5]–[7] without actual hardware implementation, whereas, in this paper, we addresse both software and hardware to develop an automatic mackerel-sorting machine.

Developing automatic sorting machine of mackerels is a challenging task. There are two required functions. Firstly, it needs to automatically detect mackerels on a conveyor belt so that mackerels can be separated from the belt. Consequently, we can transport mackerels to destinations. Secondly, species classification is needed. Unfortunately, we discovered that a simple convolutional neural network, CNN, cannot obtain favorable results. One obstacle is similar appearance among the species. Although there are distinguishable features between species: texture on abdomen and mackerel shape,[1] it is hard to train CNN to capture such features.

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

[1]Generally, body shapes are round and ellipse for blue and chub mackerels, respectively.

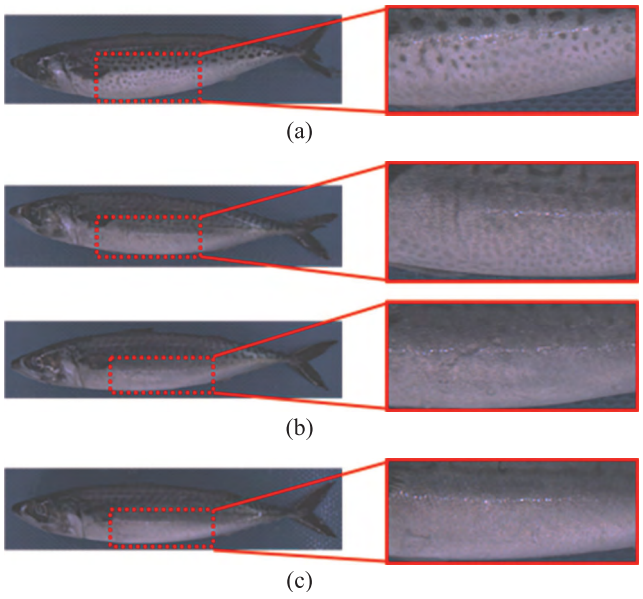

FIGURE 1. Mackerel species. Top row shows a blue mackerel, the next two rows show the hybrid type, and the last row shows a chub mackerel. The right column shows the pattern on their abdomen. (a) Blue mackerel. (b) Hybrid type. (c) Chub mackerel.

We propose an automatic sorting machine using a red laser light and Faster R-CNN [12]. The overall sorting machine architecture is depicted in Fig. 2. The machine is equipped with conveyor belt to transport the mackerels, cameras with red laser for image capture, and a computer that controls the sorting. The key mechanism is the irradiating red laser to abdomen from diagonally above. Since the shapes around abdomen area are different between blue and chub mackerels, the laser irradiated on blue mackerels appears circle, whereas it will be ellipse on chub mackerels. Thus, we can capture both the body shape and texture information from the irradiated region.

In this paper, we define feature extracted from abdomen region as *local feature*, whereas feature extracted from whole mackerel region is defined as *global feature*. We propose to use both the global and local features for species classification. Therefore, we construct a novel module, *GLCC* (Global and Local features Conjunct Classifier). We place this module on top of the Faster R-CNN. The use of both these features achieves a higher performance than the vanilla Faster R-CNN that uses only global or local features. It is worth noting that the processing of the GLCC module is extremely fast because it uses shared convolutional feature maps of the Faster R-CNN.

The actual process of the sorting mackerels is shown in Fig. 3. First of all, the machine receives a mackerel and aligns it horizontally. Then, as shown in (a), a mackerel is transferred from the back to the front side and the cameras above it took pictures. Subsequently, as shown in (b), the species is predicted. We use Faster R-CNN to detect the two regions where red laser is irradiated and a mackerel exists in the image. We extract global and local features from both the regions and classify the mackerel into one of the three species. Then, as shown in (c), the mackerel is shifted to one of the three destinations. Finally, as shown in (d), the mackerel was sorted correctly. We confirmed that the

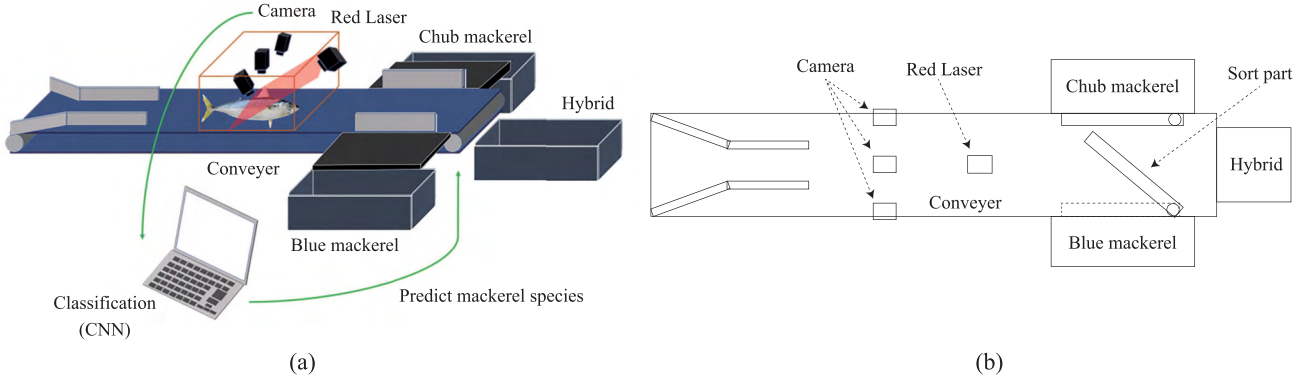

FIGURE 2. Overall architecture of sorting machine. (a) Overview of sorting machine. (b) Components of sorting machine.

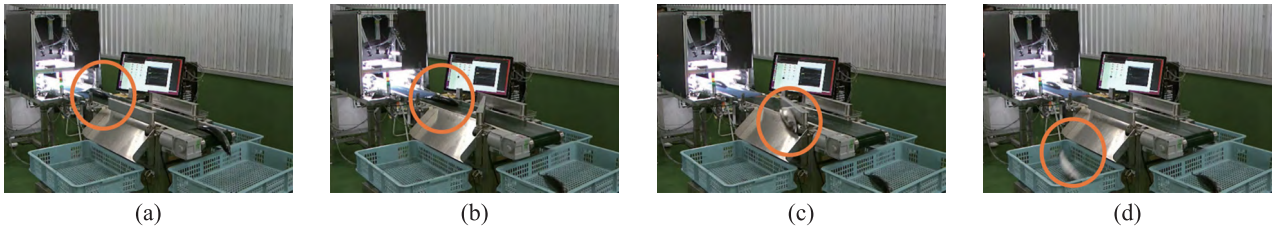

FIGURE 3. The sorting machine and the processes of sorting mackerels by the proposed method. The conveyed mackerel (encircled in orange) was classified as a blue mackerel by the proposed CNN, and shifted to the left (indicating the blue mackerel category) of the conveyor belt.

sorting machine could sort mackerels successfully with less human labor.

The contributions of this study are as follows:

- We develop a mechanism to obtain the body shape feature in image by irradiating red laser. We can easily get the three-dimensional feature of an image by this method. To the best of the authors' knowledge, this study is the first to attempt three-dimensional feature extraction for mackerel classification.
- We propose GLCC that is a novel CNN module uses global and local features for species classification. This module requires almost no additive computation cost, because it uses a shared convolutional feature map with the object detection CNN.
- We confirmed that a CNN can be trained not only on the texture of an image but also on its shape information indirectly. In this paper, we irradiated red laser to extract features of the body shape of mackerels. Consequently, the trained CNN can focus on the shape information.

This paper is an extended version of our conference paper [13]. We presented the mechanism to obtain body shape feature by irradiating red laser beam and overall sorting machine. Moreover, we considered about the effectiveness of using not only global feature but also local feature by the activations of CNN.

The remainder of this paper is organized as follows. In section II, we present some related works. In section III, we present our proposed method. In section IV, we compare the proposed method with the conventional method. In section V, we conclude this study.

II. RELATED WORKS

In this section, first, we mention the literature on fish classification. Second, we address the recent object detection methods because our proposed method is based on Faster R-CNN. Finally, we mention the FGIC (fine-grained image classification) task, because our objective is to classify mackerels based on their class.

A. FISH CLASSIFICATION

Fouad *et al.* [1] classified tilapia and non-tilapia using image processing techniques and machine learning. They used scale invariant feature transformation [14] and SURF (speeded up robust features) [15] in the feature extraction phase, and a SVM (support vector machine), artificial neural networks, and K-nearest neighbor in the classification stage. They concluded that the SVM with SURF achieved the best classification performance. Khotimah *et al.* [2] proposed classification algorithms to categorize frozen bigeye, yellowfin, and skipjack tuna. They used shape information and texture features for classification. In their study, circular rate of tuna's head, ratio of head area and circular area and so on are used as shape information, and texture feature is the gray level co-occurrence matrix (GLCC) [16] of abdomen region. They used a decision tree for the classification. Kitasato *et al.* [17] classified blue and chub mackerels based on their geometric features and several textures as discriminative features using SVM as the classifier. They measured the ratio of the base length between the first and ninth spines of the dorsal fins to the fork length as the geometric feature, and considered the texture features based on the GLCM of the abdomen region. They classified successfully and achieved a high performance, but their method is highly dependent on a mackerel's condition and outer environment. Hasija *et al.* [3] used a graph matching with subspace technique to classify fish, and achieved a high classification accuracy compared to other methods. Chuang *et al.* [4] used several features of a fish such as head size, eye texture, and tail ratio for the classification of seven fish species. Chuang *et al.* [5], [6] proposed an unsupervised machine learning strategy for feature extraction. Hsiao and Chen [7] proposed an over-atoms accumulation orthogonal matching pursuit for fish recognition. There are some studies that use engineered feature or sparse representation and machine learning for fish classification [8]–[11].

Recently, there have been many studies using CNN for classification owing to its high performance. Siddiqui *et al.* [18] used a CNN to classify underwater fishes and proved that CNN is very effective in fish classification in an underwater environment containing noise and blur. Ge *et al.* [19] also used CNN to extract the feature representation of image and used a GMM (Gaussian mixture model) to classify fine-grained fish images. In addition, other CNN-based methods also exist [20]–[22]. Apart from fish classification, fish detection under a restricted environment is also performed using a CNN [23]–[27].

Classifying fish based on local features such as abdomen or head shape has been studied intensively. Moreover, a CNN has been applied in each of the mentioned studies. Inspired by these studies, we determine the use of a CNN to achieve a high classification performance with a fast processing speed.

B. OBJECT DETECTION

Here, we present an object detection method, which is the core idea of our proposed method. Object detection is a popular research subject under computer vision, and the recent trend is a CNN based method for object detection. Faster R-CNN is one of the recent state-of-the-art methods and is derived from R-CNN (Region based CNN) [28] and Fast R-CNN [29]. R-CNN is composed of a proposal generation and classification stage. Proposals from a given image are generated using other method modules such as a selective search [30]. Proposal regions cropped from the original input image are fed into the classification stage, which uses the CNN to classify proposals into objects or background. In addition, a bounding-box regression process adjusts proposal rectangles to the object size accurately. The problem of R-CNN is its high computation cost because the CNN computes a feature map for each proposal. In Fast R-CNN, an RoI-pooling is introduced to share convolution features. Given an input image, the CNN computes feature maps of the entire image. By using the RoI-pooling, the feature maps

of the proposal regions are cropped and pooled to a fixed size, thereby reducing the computation cost. However, Fast R-CNN requires another pipeline to generate the proposals, hence, it cannot process end-to-end consistently. Faster R-CNN uses an RPN (region proposal network) to generate proposals with convolutional layers. This realizes end-to-end processing and improves detection speed and accuracy.

Apart from Faster R-CNN, YOLO [31]–[33] and SSD [34] are also CNN-based methods and achieve a comparable performance. These methods are a single shot detector unlike Faster R-CNN; hence, they can detect fast at the cost of a little reduction in the detection accuracy.

In this paper, the input image we considered contains a fish in a part region of input image. To detect the region containing mackerel features accurately, we use Faster R-CNN as the backbone.

C. FINE-GRAINED IMAGE CLASSIFICATION

FGIC is a more challenging task than general classification because the objects to be classified are similar to each other. In this task, two problems occur: a large intra-class and a small inter-class variance. To solve these problems, there are some strategies using local features. Zhang *et al.* [35] used R-CNN to detect proposals of object parts and classified them using an SVM. Xiao *et al.* [36] used a selective search to generate proposal patches and they were classified into objects and parts using a CNN. Ge *et al.* [19] used a CNN to extract feature representations and used the GMM to classify fine-grained images such as those of fishes and food. Fu *et al.* [37] proposed a recurrent attention model without the need for bounding-box annotations.

The recent trend includes a weakly supervised training [38]–[40] that does not require much effort for annotation. These methods are effective by using local features of objects for fine-grained classification, but have a multi-stage architecture and complex pipelines. Our idea of using global and local features of mackerels is inspired by these studies. Moreover, our proposed method is an end-to-end trainable CNN architecture.

The proposed GLCC module extracts features from global and local regions of mackerels. Lisin *et al.* [41] also uses global and local features. The difference between the proposed method and [41] is the number of classifiers. Several classifiers were used in [41], whereas one classifier is used in the proposed method. Lisin *et al.* [41] trained the classifiers for each global and local feature. Then they used an ensemble algorithm called stacking to combine the outputs of the separate classifiers. On the other hand, the proposed method combines the two features with regarding to channel axis, resulting in one feature vector. Then we predict scores using the one feature vector.

III. PROPOSED METHOD

In this section, we describe the proposed feature extraction and mackerel classification methods based on a CNN and sorting machine. First, we describe the mechanism of the sorting machine based on above concept. Second, we introduce the mackerel features that form the core of the proposed CNN architecture and the mechanism to obtain those features. Third, we describe the proposed classification method. Finally, the training strategy for the proposed method is noted.

A. SORTING MACHINE

The overall architecture of the sorting machine is depicted in Fig. 2. In order to identify the discriminative feature of mackerels using red laser and thus, the sorting machine consists of a red laser irradiator located diagonally above.

First, mackerels are transported from the conveyor belt from the left to the right and are lined horizontally. When mackerels pass the sensor, cameras take photographs using irradiated red laser. The photograph is inputted to the proposed CNN model, and it outputs the predicted mackerel species signal. Using the predicted results, the sorting machine sorts the conveyed mackerels into three directions.

This machine realizes automatic mackerel sorting without human labor.

B. EXTRACTING BODY SHAPE FEATURES

We emphasize the importance of local features extracted from the abdomen of mackerels. In practice, fisheries distinguish mackerels by inspecting the patterns on their abdomens. Moreover, different types of mackerels have different body shapes.

The blue mackerel has an approximately circular body shape, while the shape of the chub mackerel is close to an ellipse. This difference is not trivial and an important factor when distinguishing types of mackerels. The body shape appears on a local region toward overall mackerel body and is smaller than the texture feature; thus, we use the body shape as a local feature.

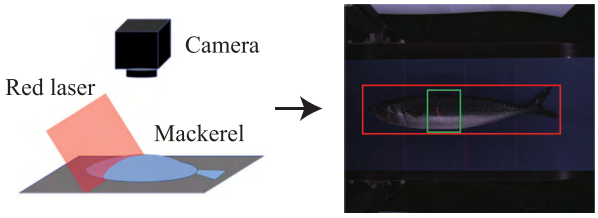

FIGURE 4. Mechanism to obtain local features. Red and green rectangles represent the body and abdomen regions of the mackerel, respectively. Faster R-CNN detects each region.

The texture is clearly visible on the abdomen, and it can hence be used for image processing. However, we cannot use body shape information for image processing because that feature does not ordinarily appear on the surface of the mackerel. To extract features from the abdomen, the proposed mechanism irradiates a red laser beam toward the mackerel from diagonally above as shown in Fig. 4, to obtain the body shape feature image. We can detect the abdomen through the

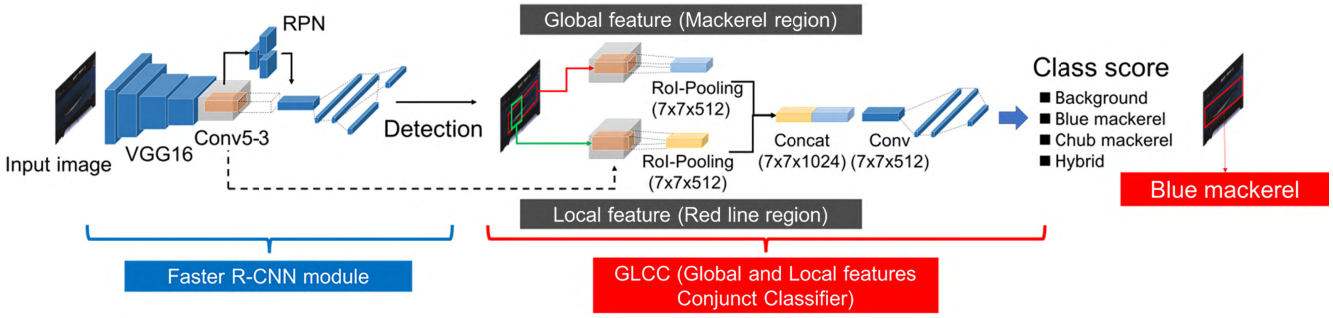

FIGURE 5. Architecture of the proposed method. The most important concept of this architecture is the feature extraction from both the whole body (mackerel region) and abdomen (red line region), which are depicted as global and local features after detection. Specifically, we extract features of the detected regions from conv5-3 feature map. Then, we apply RoI-Pooling to the features, resulting in the same size features. Finally, both of the features are concatenated and used to predict class scores.

regions of red laser light. Because mackerels lie horizontally, we can obtain continuous information on their abdomen by using red laser. Through this mechanism, we can obtain the body shape information in the form of an image. When a mackerel flows on the conveyor belt and reaches a specific position, the camera automatically captures images so that a series of similar images is obtained.

This mechanism is simple but effective to identify the object's shape. Almost all studies on fish classification focus only their texture; thus, to the best of the authors' knowledge, this study is the first to attempt three-dimensional feature extraction for mackerel classification. In this paper, we used the whole mackerel region as the global feature and the abdomen region with red laser as the local feature mentioned above.

C. PROPOSED CLASSIFICATION ALGORITHM

CNNs are frequently used in the field of computer vision and achieve the best performance in several tasks such as object classification, object detection, and semantic segmentation. Therefore, we used this powerful tool for image classification. However, vanilla CNNs do not use mechanisms that focus on the local descriptor and overall image information evenly. Based on these reasons, we proposed a new CNN module directing attention to the overall image and the local region of the object. To realize both regions, we adopted a Faster R-CNN for convenience. In addition, we stress that reliability and applicability of the Faster R-CNN are validated through many applications: road damage recognition [42], action analysis [43], pedestrian detection [44], document analysis [45], domain adaptation [46], [47], image captioning [48], and scene analysis [49].

The overall architecture of the proposed method is shown in Fig. 5. It consists of two modules: Faster R-CNN and global and local features conjunct classifier (GLCC) that is the proposed module. We summarize the steps of the proposed method in Fig. 6 to provide implementation details.

The Faster R-CNN module detects the body and abdomen regions of the mackerel. Broadly, the algorithm of the Faster R-CNN is composed of three parts. Firstly, it extracts features

Input: image
Output: Predicted class scores

1: Apply Faster R-CNN to the image
- 1.1: Extract features by VGG16
- 1.2: Generate proposal regions
- 1.3: Predict body and abdomen scores for the regions
- 1.4: Take the most confident region for each of body and abdomen
- 1.5: Provide body region, abdomen region, and features extractd in Step 1.1 to GLCC

2: Apply GLCC
- 2.1: Get features of the body and abdomen regions from features extracted in Step 1.1
- 2.2: Apply RoI-Pooling to the features
- 2.3: Concatenate the features
- 2.4: Predict class scores using the concatenated feature

FIGURE 6. Detailed steps of the proposed method.

from the input image by CNN, which is called as backbone. In this paper, we used VGG16 [50] as the backbone. Secondly, RPN produces proposal regions. Thirdly, the proposals are classified into the body, abdomen, or background.

The GLCC predicts the scores for four classes: blue mackerel, chub mackerel, hybrid, and background. The detailed procedures are following. Given the global and local regions by the Faster R-CNN, the GLCC extracts the global feature and the local feature by applying RoI-Pooling to the regions in the conv5-3 feature map of VGG16, resulting in two same size features: f_{global} and f_{local}. Then, we merge the two features into one feature by concatenating them along channel axis: $f_{concat} = concat(f_{global}, f_{local})$, where $concat()$ is a function that concatenates two feature matrices along a channel axis. In this paper, we determine the third axis of the features as the channel axis. Finally, we predict class scores using convolution and fully connected layers.

By using the Faster R-CNN, we can detect the global and local regions of the mackerel. Moreover, the algorithm to extract both features is added to the CNN for

end-to-end processing. Owing to both global and local features, the proposed method can accurately distinguish the mackerel species.

D. TRAINING STRATEGY

To train the proposed network, we used bounding boxes of the mackerel and red laser regions as the ground truth. We pre-trained the Faster R-CNN module to detect seven regions: background, blue mackerel, chub mackerel, hybrid, blue mackerel red line, chub mackerel red line, and hybrid red line. Then, we fine-tuned the overall network (Faster R-CNN + GLCC). The overall training loss is summarized in (1).

$$L_{total} = L_{rpn} + \lambda_{fastrcnn} L_{fastrcnn} + \lambda_{glcc} L_{glcc} \tag{1}$$

L_{rpn}, $L_{fastrcnn}$, and L_{glcc} represent the loss function of the RPN, Fast R-CNN, and GLCC, respectively. $\lambda_{fastrcnn}$ and λ_{glcc} are hyper parameters to balance each loss, and both these parameters are set to 1. L_{rpn} and $L_{fastrcnn}$ are described in [12], [29], and thus a detailed explanation is not provided here. L_{glcc} is the softmax cross entropy loss for the GLCC module and is defined by (2).

$$L_{glcc} = -\frac{1}{|M|} \sum_{i \in M} \ln p_u^i \tag{2}$$

M denotes a minibatch subset, and p_u is the output for the correct label. The GLCC is trained with a pair of inputs and true labels, i.e., $(R^i_{mackerel}, R^i_{redline}, u^i) \in M$. We selected a pair of mackerel $R_{mackerel}$ and red line regions $R_{redline}$ as the input of the GLCC based on their higher probabilities. We assign a mackerel class label if both the IoU overlap ratio between $R_{mackerel}$ and the ground truth of mackerel and the ratio between the $R_{redline}$ and the ground truth of red line are higher than 0.5; the others are labeled as background (non-mackerel). We set the value 0.5 by following [12], [28], [29].

IV. EXPERIMENTS

In this section, we evaluated the proposed method (Faster R-CNN + GLCC) and compared it with other methods. The proposed method is a detection-based method, but the aim of this study is to use image recognition for image classification. Therefore, we compared our method to simple image classification. We applied vanilla Faster R-CNNs to use the global (mackerel region) or local regions (redline region) to confirm the effectiveness of using both of them, and we denoted these as the Faster R-CNN - G and Faster R-CNN - L, respectively.

In addition to these, we used a simple VGG16 CNN to recognize mackerel images. This VGG16 CNN uses object recognition and not object detection. This can help demonstrate that the object detection-based method is more effective than the image recognition method for this task. The backbone of the Faster R-CNN used in this paper was VGG16 and so a fair comparison could be made.

For both training and testing, we used GPU NVIDIA TITAN X (Pascal).

A. DATASETS AND EVALUATION METRICS

We prepared 417 mackerel images for training (blue mackerel: 81 images, chub mackerel: 258 images, and hybrid: 78 images), and 534 images for testing (blue mackerel: 48 images, chub mackerel: 264 images, and hybrid: 222 images). We splitted the images by dates of collection: 417 images are captured in December 2017, and 534 images in January 2018. This split ensures variability of mackerels. Each image has a mackerel region and red line region as the ground truth. The Faster R-CNN - G was trained with only mackerel regions, while the Faster R-CNN - L was trained with only red line regions. The Faster R-CNN + GLCC was trained using both of them. The same ground truths were evaluated.

To compare the proposed method with other methods, we used accuracy, detection, and inference speed as evaluation metrics. Accuracy is defined in (3).

$$Accuracy = \frac{N_{truepositive}}{N_{total}} \tag{3}$$

$N_{truepositive}$ is the number of images that are correctly categorized by the methods, and N_{total} is the total number of images for the testing dataset.

Detection is defined in (4). The prediction by object detection methods is dependent on the detected mackerel or redline region in the image.

$$Detection = \frac{N_{truepositiveBB}}{N_{totalBB}} \tag{4}$$

$N_{truepositiveBB}$ is the number of bounding boxes that are correctly detected as mackerel or redline regions by the methods. In this work, the region in which the IoU overlap with the ground-truth region is greater than 0.5 represents correct detection. $N_{totalBB}$ is the total number of ground-truth bounding boxes. IoU is computed by (5).

$$IoU = \frac{|R_{overlap}|}{|R_{detect} \cup R_{groundtruth}|} \tag{5}$$

R_{detect} and $R_{groundtruth}$ denote the area of detection and ground truth, respectively. $R_{overlap}$ is the overlap area between R_{detect} and $R_{groundtruth}$.

Inference speed indicates the processing time per image. The sorting machine requires a fast processing speed and thus, the methods are compared from the view point of speed.

B. EXPERIMENTAL RESULTS

Numerical results are shown in Table 1, and some successful resulting images are shown in Fig. 7.

The inference speed of VGG16 was much faster than the Faster R-CNN methods because it is a simple feedforward CNN. However, its accuracy was much lower than the Faster R-CNN methods. This is because that the VGG16 used overall image information containing large background regions

TABLE 1. Accuracy of fish classification. The speed is the average processing time of the architecture shown in Fig. 5. We measured the average processing time over the test images.

Method	Accuracy	Detection	Speed (sec/img)
VGG16	67.98	-	0.010
Faster R-CNN - G	87.08	100	0.095
Faster R-CNN - L	82.02	100	0.099
Proposed	90.26	100	0.086

for recognition. A discussion on the Faster R-CNN methods is presented as follows.

All Faster R-CNN methods can detect mackerel or red line regions successfully. The proposed method detected mackerels and recognized their species more accurately than the other methods. In terms of the processing time, the proposed method is almost equivalent to Faster R-CNN method - L and - G. As shown in Fig. 7, the recognition results are incorrect in both the Faster R-CNN - G and Faster R-CNN - L, while the proposed method successfully recognized the species. From this aspect, using both global and local features is effective in order to distinguish the mackerel species because the two features can complement each other. In addition, the proposed method took almost the same processing time as the vanilla Faster R-CNN because the GLCC module shares convolution feature maps with the Faster R-CNN module. Hence, the GLCC requires less computation cost and can achieve improved recognition accuracy.

In this experiment, the Faster R-CNN - G detects the mackerels with red laser, but this recognition result is lower than that of the proposed method. From this view, it is important for the algorithm to consider the local descriptor.

We carefully considered applying the proposed method to a general image classification, such as ImageNet. However, the method needs the special equipment to irradiate red laser physically. It is not feasible to develop the equipment for general targets. Therefore, the method works in the particular situation, and it cannot be applied to general classification tasks.

C. FAILURE RESULTS

Some failure results are shown in Fig. 8. These images were classified as the incorrect species by the proposed method. This is because these mackerels exhibited more similar features than others.

For example, let us consider the top row of results in Fig. 8; the correct species is the chub mackerel. This chub mackerel

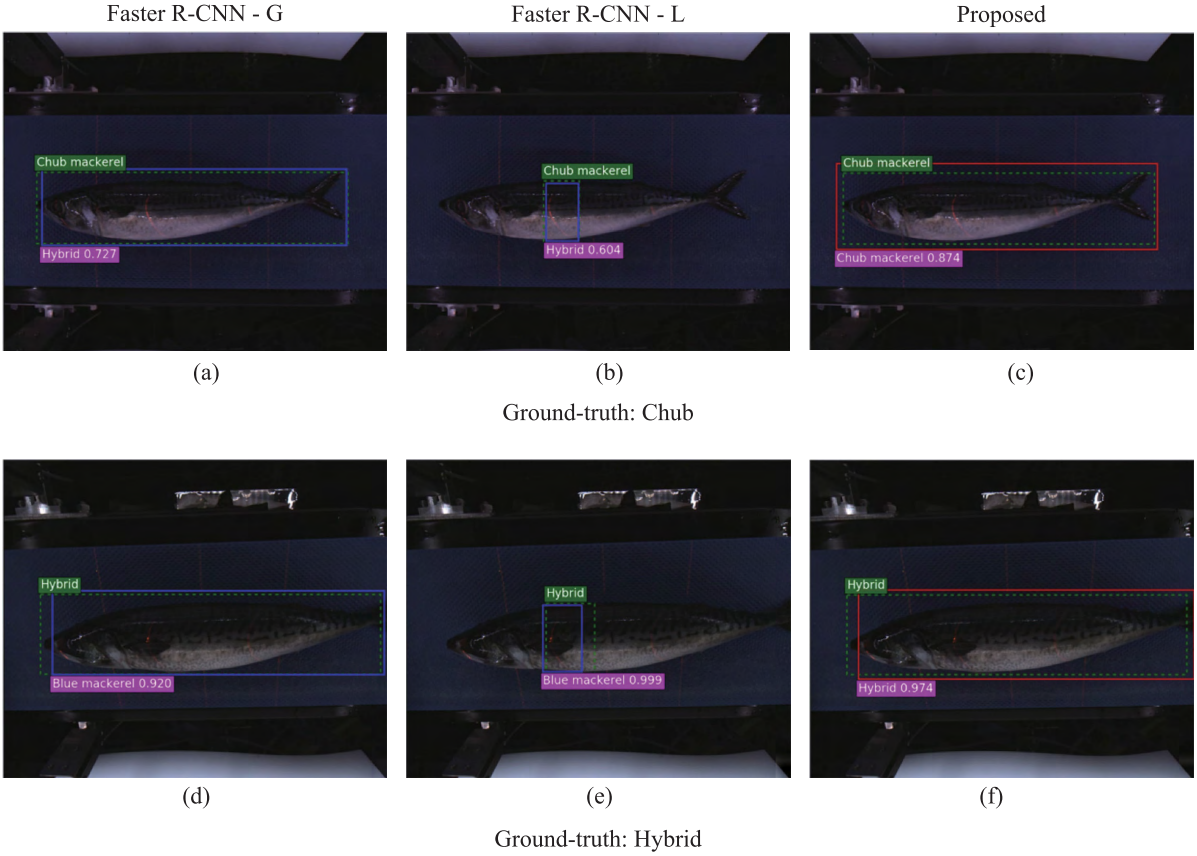

FIGURE 7. Examples of results. The green dashed rectangle and text in the images are ground-truth bounding-box and species, respectively. The red rectangle represents success result, while the blue rectangle is failure. Purple text indicates predicted species and its confidence score. (a) Hybrid (score: 0.72). (b) Hybrid (score: 0.60). (c) Chub (score: 0.87). (d) Blue (score: 0.92). (e) Blue (score: 0.99). (f) Hybrid (score: 0.97).

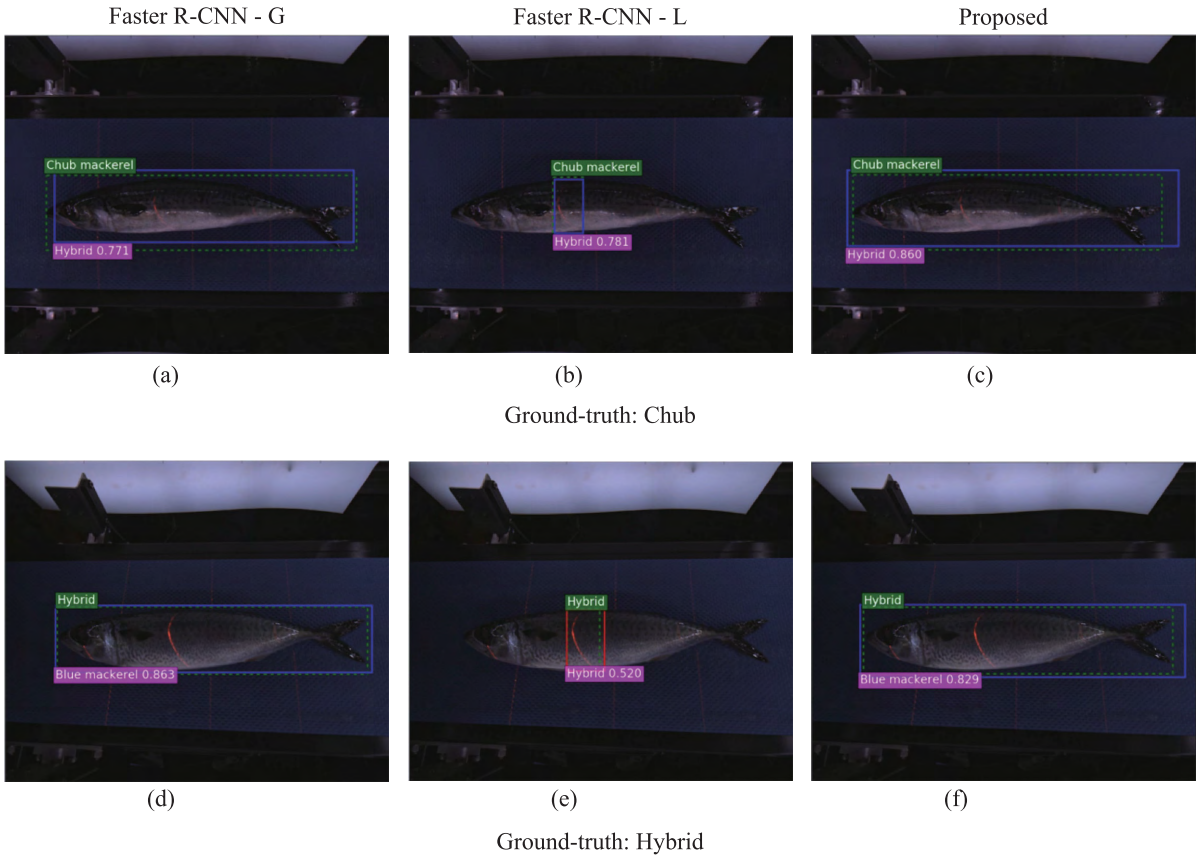

FIGURE 8. Examples of failure results. (a) Hybrid (score: 0.77). (b) Hybrid (score: 0.78). (c) Hybrid (score: 0.86). (d) Blue (score: 0.86). (e) Hybrid (score: 0.52). (f) Blue (score: 0.82).

has almost no spotted pattern in the abdomen and although its body shape is of the chub mackerel type, some hybrid mackerels exhibit shapes similar to it. Therefore, although, all methods may make mistakes, the proposed method predicts with higher accuracies than the other methods. This shows the effectiveness of interpolation between the global and local features and thus, the confidence in the hybrid class increases.

Next, considering the bottom row of results in Fig. 8, the correct species is the hybrid mackerel case; the proposed method predicted blue mackerel, and the Faster R-CNN - G also predicted blue mackerel. On the other hand, the Faster R-CNN - L predicted the species successfully. This hybrid mackerel exhibits a spotted pattern on the abdomen that is specific to blue mackerels; its body shape was close to an ellipse that is specific to the chub mackerel. The proposed method predicted it as the blue mackerel with a probability of 0.829; this confidence score was lower than the result of the Faster R-CNN - G. The Faster R-CNN - G predicted it as the blue mackerel with a probability of 0.863 and the Faster R-CNN - L predicted it as the hybrid with probability of 0.520; thus, the local features taught the CNN information about hybrid species. As the result of this, the proposed method predicted incorrect labels with a lower probability than the Faster R-CNN - G, but it cannot obtain enough information to predict correct labels.

Although using both the global and local features complements each other, if both features do not match these species, classification becomes more difficult. In this study, we used features that seemed effective, but there may be other discriminative features that are not based on specific appearances. To address this issue, we need to examine discriminative feature specific to mackerel, and use another feature without our supervision.

We carried out an experiment using ResNet101 [51] as the backbone of the Faster R-CNN. The accuracy was 76.8 (%), which was less than VGG16. Generally, ResNet is superior to VGG in image classification tasks. However, the experimental results show that the accuracy is lower than expected. We think this result was due to the amount of our dataset. Actually, ResNet101 is three times bigger than VGG16. The number of parameters is 14.7M in VGG, whereas 42.5M in ResNet101. Note that the numbers do not contain FC layers since those models are used as feature extractors. In order to train big models such as ResNet, we need to expand our dataset. Overall, VGG is a suitable model for our dataset. Hence, the proposed method works well with limited datasets such as ours.

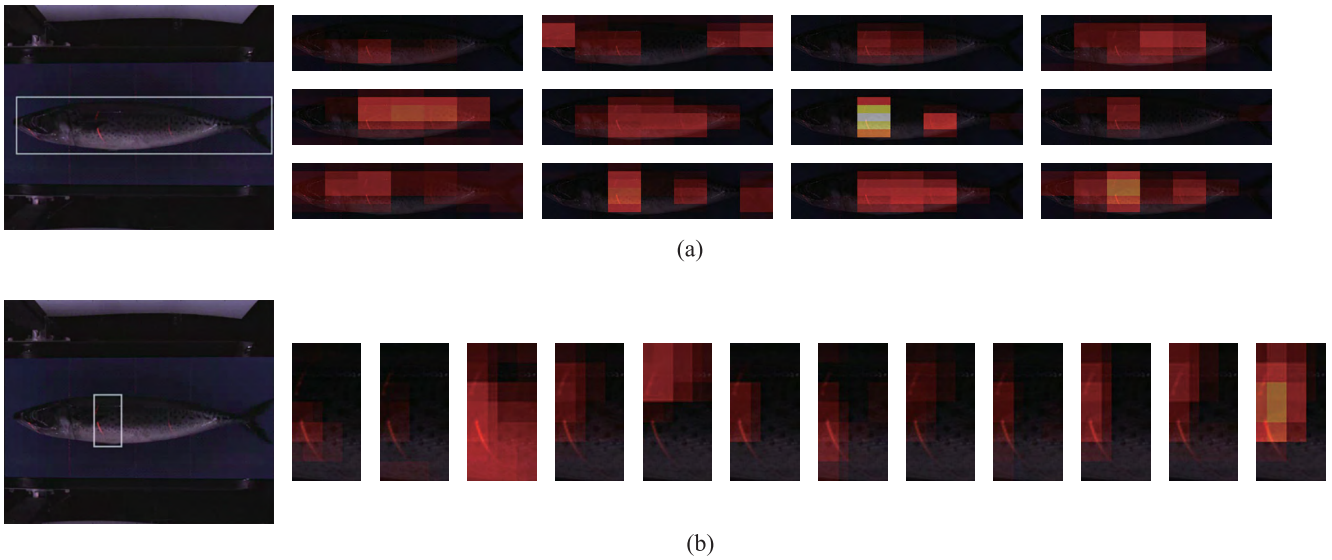

FIGURE 9. Each region detected by proposed method in one image of test dataset. Top row is global region (overall mackerel region) and bottom one is local region (red line region). Left column shows the results of detection and right one shows some examples of the activation feature maps cropped from RoI-pooling layer inside detected region. The red region indicates high activation. This mackerel is classified to blue mackerel successfully with a probability of 0.996. (a) Mackerel region (Global region). (b) Red line region (Local region).

D. FEATURE ACTIVATION

Here, we discuss the activation of feature regions in mackerels. Fig. 9 shows the detection results of the global and local regions by the proposed method and activation feature maps cropped from RoI-pooling layer inside the detected regions. For visibility, we show the activation maps on the detection regions. As seen from Fig. 9 (a), some partial regions of the mackerel are activated. However, the activated regions are coarse due to RoI-pooling.

On the other hand, (b) shows that fine regions are activated. Moreover, we can see the activations along the irradiated red laser region. This verifies that the CNN can be trained to focus on three-dimensional feature indirectly. The local region is much smaller than the overall mackerel size, but activation occurs at every position in feature maps of local regions. Because of this, even if this feature map is pooled to be smaller, it significantly contributes to recognition accuracy. These results indicate that important fine regions may not be activated if the feature map is large. RoI-pooling to large regions extracts coarse features, thus, it may lose fine activation. RoI-pooling to small regions extracts fine features. Hence, global and local features can compensate each other. Therefore, feature extraction from global and local regions is an important factor in mackerel classification.

V. CONCLUSION

We proposed an automatic mackerel sorting machine. This is based on the Faster R-CNN and the GLCC that uses global and local features extracted from the whole body and abdomen regions, respectively. In addition, we extract both the features from shared convolution maps so that we can reduce computational cost. The GLCC improved the classification accuracy by using global and local features as discriminative features. Moreover, we validate that the local region is necessary because of its fine activation in the convolutional feature map. The experimental results show that the strategy using both features is very effective. In this work, we used the Faster R-CNN as a feature region-based detector; however, this can be replaced with other object detection methods based on CNN such as YOLO and SSD. We will focus on improving the classification accuracy considering other discriminative features and using other efficient object detection techniques in future studies. Another important future work is developing a own neural network model using better models such as YOLO to improve the performance.

ACKNOWLEDGMENT

The data used for the experiments was provided with the cooperation of LATEST-SYSTEM Inc. The machine for sorting fish was constructed by Sea Tech Co., Ltd.

REFERENCES

[1] M. M. M. Fouad, H. M. Zawbaa, N. El-Bendary, and A. E. Hassanien, "Automatic nile tilapia fish classification approach using machine learning techniques," in *Proc. HIS*, Dec. 2013, pp. 173–178.
[2] W. N. Khotimah, A. Z. Arifin, A. Yuniarti, A. Y. Wijaya, D. A. Navastara, and M. A. Kalbuadi, "Tuna fish classification using decision tree algorithm and image processing method," in *Proc. IC3INA*, Oct. 2015, pp. 126–131.
[3] S. Hasija, M. J. Buragohain, and S. Indu, "Fish species classification using graph embedding discriminant analysis," in *Proc. CMVIT*, Feb. 2017, pp. 81–86.
[4] M.-C. Chuang, J.-N. Hwang, F.-F. Kuo, M.-K. Shan, and K. Williams, "Recognizing live fish species by hierarchical partial classification based on the exponential benefit," in *Proc. ICIP*, Oct. 2014, pp. 5232–5236.
[5] M.-C. Chuang, J.-N. Hwang, and K. Williams, "A feature learning and object recognition framework for underwater fish images," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1862–1872, Apr. 2016.
[6] M.-C. Chuang, J.-N. Hwang, and K. Williams, "Supervised and unsupervised feature extraction methods for underwater fish species recognition," in *Proc. ICPR Workshop Comput. Vis. Anal. Underwater Imag.*, Aug. 2014, pp. 33–40.

[7] Y.-H. Hsiao and C.-C. Chen, "Over-atoms accumulation orthogonal matching pursuit reconstruction algorithm for fish recognition and identification," in *Proc. ICPR*, Dec. 2016, pp. 1071–1076.
[8] A. Rova, G. Mori, and L. M. Dill, "One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video," in *Proc. MVA*, 2007, pp. 404–407.
[9] F. Shafait *et al.*, "Fish identification from videos captured in uncontrolled underwater environments," *ICES J. Mar. Sci.*, vol. 73, no. 10, pp. 2737–2746, Nov. 2016.
[10] Y.-H. Shiau, F.-P. Lin, and C.-C. Chen, "Using sparse representation for fish recognition and verification in real world observation," presented at the Int. Workshop Visual Observ. Anal. Animal Insect Behav. (VAIB), conjunction Int. Conf. Pattern Recognit., 2012.
[11] K. Blanc, D. Lingrand, and F. Precioso, "Fish species recognition from video using SVM classifier," in *Proc. ACM Int. Workshop Multimedia Anal. Ecological Data*, Nov. 2014, pp. 1–6.
[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[13] Y. Nagaoka, T. Miyazaki, Y. Sugaya, and S. Omachi, "Mackerel classification using global and local features," in *Proc. ETFA*, vol. 1, Sep. 2018, pp. 1209–1212.
[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[15] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. ECCV*, May 2006, pp. 404–417.
[16] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *Int. J. Sci. Res.*, vol. 3, no. 5, p. 1, May 2013.
[17] A. Kitasato, T. Miyazaki, Y. Sugaya, and S. Omachi, "Automatic discrimination between scomber japonicus and scomber australasicus by geometric and texture features," *Fishes*, vol. 3, no. 3, no. 3, p. 26, Sep. 2018.
[18] S. A. Siddiqui *et al.*, "Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data," *ICES J. Mar. Sci.*, vol. 75, no. 1, pp. 374–389, Jan./Feb. 2017.
[19] Z. Ge, C. McCool, C. Sanderson, and P. Corke, "Modelling local deep convolutional neural network features to improve fine-grained image classification," in *Proc. ICIP*, Sep. 2015, pp. 4112–4116.
[20] G. Ding *et al.*, "Fish recognition using convolutional neural network," in *Proc. OCEANS*, Sep. 2017, pp. 1–4.
[21] D. Rathi, S. Jain, and D. S. Indu. (2018). "Underwater fish species classification using convolutional neural network and deep learning." [Online]. Available: https://arxiv.org/abs/1805.10106
[22] L. Meng, T. Hirayama, and S. Oyanagi, "Underwater-drone with panoramic camera for automatic fish recognition based on deep learning," *IEEE Access*, vol. 6, pp. 17880–17886, 2018.
[23] W. Xu and S. Matzner. (2018). "Underwater fish detection using deep learning for water power applications." [Online]. Available: https://arxiv.org/abs/1811.01494
[24] S. Choi, "Fish identification in underwater video with deep convolutional neural network: SNUMedinfo at LifeCLEF fish task 2015," in *Proc. CEUR Workshops*. [Online]. Available: http://ceur-ws.org/Vol-1391/110-CR.pdf
[25] X. Li, M. Shang, H. Qin, and L. Chen, "Fast accurate fish detection and recognition of underwater images with Fast R-CNN," in *Proc. OCEANS*, Oct. 2015, pp. 1–5.
[26] R. Mandal, R. M. Connolly, T. A. Schlacher, and B. Stantic, "Assessing fish abundance from underwater video using deep neural networks," in *Proc. IJCNN*, Jul. 2018, pp. 1–6.
[27] D. Zhang, G. Kopanas, C. Desai, S. Chai, and M. Piacentino, "Unsupervised underwater fish detection fusing flow and objectiveness," in *Proc. WACVW*, Mar. 2016, pp. 1–7.
[28] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Jun. 2014, pp. 580–587.
[29] R. Girshick, "Fast R-CNN," in *Proc. ICCV*, Dec. 2015, pp. 1440–1448.
[30] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.
[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, Jun. 2016, pp. 779–788.
[32] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. CVPR*, Jul. 2017, pp. 6517–6525.
[33] R. Joseph and F. Ali. (2018). "YOLOv3: An incremental improvement." [Online]. Available: https://arxiv.org/abs/1804.02767
[34] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. ECCV*, Oct. 2016, pp. 21–37.
[35] N. Zhang, J. Donahue, R. B. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Proc. ECCV*, Sep. 2014, pp. 834–849.
[36] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proc. CVPR*, Jun. 2015, pp. 842–850.
[37] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," in *Proc. CVPR*, vol. 2, Jul. 2017, pp. 4476–4484.
[38] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018.
[39] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian, "Picking deep filter responses for fine-grained image recognition," in *Proc. CVPR*, Jun. 2016, pp. 1134–1142.
[40] X. He, Y. Peng, and J. Zhao, "Fast fine-grained image classification via weakly supervised discriminative localization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1394–1407, May 2019.
[41] D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining local and global image features for object class recognition," in *Proc. CVPR Workshops*, Sep. 2005, p. 47.
[42] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with Faster R-CNN," in *Proc. IEEE Big Data*, Dec. 2018, pp. 5220–5223.
[43] Y. Chao, S. Vijayanarasimhan, B. Seybold, D. A. Ross, J. Deng, and R. Sukthankar, "Rethinking the Faster R-CNN architecture for temporal action localization," in *Proc. CVPR*, Jun. 2018, pp. 1130–1139.
[44] Y. Xu, X. Liu, Y. Liu, and S. Zhu, "Multi-view people tracking via hierarchical trajectory composition," in *Proc. CVPR*, Jun. 2016, pp. 4256–4265.
[45] J. Ma *et al.*, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3111–3122, Nov. 2018.
[46] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. CVPR*, Jun. 2018, pp. 3339–3348.
[47] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proc. CVPR*, Jun. 2018, pp. 5001–5009.
[48] J. Lu, J. Yang, D. Batra, and D. Parikh, "Neural baby talk," in *Proc. CVPR*, Jun. 2018, pp. 7219–7228.
[49] D. Xu, Y. Zhu, C. B. Choy, and L. Fei-Fei, "Scene graph generation by iterative message passing," in *Proc. CVPR*, Jul. 2017, pp. 3097–3106.
[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the ICLR, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.

YOSHITO NAGAOKA received the B.E. and M.E. degrees from Tohoku University, in 2017 and 2019, respectively. His research interest includes image processing using deep neural networks.

TOMO MIYAZAKI (M'09) received the Ph.D. degree from the Graduate School of Engineering, Tohoku University, in 2011. He joined Hitachi, Ltd., in 2011. He was with the Graduate School of Engineering, Tohoku University, as a Researcher, from 2013 to 2014. Since 2015, he has been an Assistant Professor with Tohoku University. His research interests include pattern recognition and image processing. He is a member of the Institute of Electronics, Information and Communication Engineers.

YOSHIHIRO SUGAYA received B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1995, 1997, and 2002, respectively. He is currently an Associate Professor with the Graduate School of Engineering, Tohoku University. His research interests include the computer vision, pattern recognition, image processing, and parallel processing and distributed computing. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Information Processing Society of Japan.

SHINICHIRO OMACHI (M'96–SM'11) received B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1988, 1990, and 1993, respectively. He was a Research Associate with the Education Center for Information Processing, Tohoku University, from 1993 to 1996. Since 1996, he has been with the Graduate School of Engineering, Tohoku University, where he is currently a Professor. From 2000 to 2001, he was a Visiting Associate Professor with Brown University. His research interests include pattern recognition, computer vision, image processing, image coding, and parallel processing. He is a member of the Institute of Electronics, Information and Communication Engineers. He received the IAPR/ICDAR Best Paper Award, in 2007, the Best Paper Method Award of the 33rd Annual Conference of the GfKl, in 2010, the ICFHR Best Paper Award, in 2010, and the IEICE Best Paper Award, in 2012.