

Automated Detection of Hydrothermal Emission Signatures From Multibeam Echo Sounder Images Using Deep Learning

Kazuhide Mimura , Kentaro Nakamura , Kazuhiro Takao, Kazutaka Yasukawa, and Yasuhiro Kato

Abstract—Seafloor massive sulfide deposits have attracted attention as a mineral resource, as they contain a wide variety of base, precious, and other valuable critical metals. Previous studies have shown that signatures of hydrothermal activity can be detected by a multibeam echo sounder (MBES), which would be beneficial for exploring sulfide deposits. Although detecting such signatures from acoustic images is currently performed by skilled humans, automating this process could lead to improved efficiency and cost effectiveness of exploration for the seafloor deposits. Herein, we attempted to establish a method for automated detection of MBES water column anomalies using deep learning models. First, we compared the “Mask R-CNN” and “YOLO-v5” detection model architectures, wherein YOLO-v5 yielded higher F1 scores. We then compared the number of training classes and found that models trained with two classes (signal and noise) exhibited superior performance compared with models trained with only one class (signal). Finally, we examined the number of trainable parameters

and obtained the best model performance when the YOLO-v5 model with a large trainable parameters was used in the two-class training process. The best model had a precision of 0.928, a recall of 0.881, and an F1 score of 0.904. Moreover, this model achieved a low false alarm rate (less than 0.7%) and had a high detection speed (20–25 ms per frame), indicating that it can be applied in the field for automatic and real-time exploration of seafloor hydrothermal deposits.

Index Terms—Deep learning, massive sulfide deposits, multibeam echo sounder (MBES), object detection, seafloor exploration, seafloor mineral resource.

I. INTRODUCTION

AS discovery of high-grade terrestrial metal deposits has become difficult, increasing attention has been paid to seafloor mineral resources such as ferromanganese nodules, corich ferromanganese crusts, seafloor massive sulfides, rare-earth element, and yttrium-rich mud [1], [2], [3], [4], [5], [6]. Among these, seafloor massive sulfide deposits, which are composed of sulfide chimneys and mounds, form in close proximity to submarine hydrothermal vents. Since these deposits are rich in various base, precious, and critical metals [1], [2], [7], [8], [9], they have attracted attention as potential sources of minerals that are essential for sustainable society in the future [2], [10], [11]. Moreover, seafloor massive sulfide deposits also attract attention as an analogue of metal deposits on land that formed over a geological time scale of hundreds of millions of years [7]. Thus, the study of submarine hydrothermal deposits is important both to thoroughly understand the resources that humankind utilizes at present and to secure resources that humankind will require in the future.

In addition to the values in resource engineering, understanding hydrothermal activities is also important for understanding the Earth and the origins of life [12], [13], [14]. Hydrothermal fluids form when cold seawater seeping into ocean crust is heated by magma beneath the seafloor. This process transfers heat and various elements (including alkali metals, Fe, Mn, and P) from the solid earth to the ocean [12], [15], [16]. It is also known that a unique ecosystem also forms around hydrothermal vents [17], [18], [19]. Unlike the photosynthesis-supported ecosystems that are widespread on land and at the ocean surface, hydrothermal systems support ecosystems that produce energy using the redox potential difference between hydrothermal fluid and seawater, which is considered to be an analogue for early earth ecosystems

Manuscript received 21 November 2022; revised 24 January 2023; accepted 16 February 2023. Date of publication 24 February 2023; date of current version 22 March 2023. This work was supported in part by the Council for Science, Technology, and Innovation through its Cross-ministerial Strategic Innovation Promotion Program “Next-generation technology for ocean resources exploration” (Lead agency: JAMSTEC) and in part by the Japan Society for the Promotion of Science KAKENHI under Grant 20H05658 to YK, Grant 17H01361 to KN, Grant 19J14560, and Grant 21K20354 to KM. (Corresponding author: Kentaro Nakamura.)

Kazuhide Mimura is with the Ocean Resources Research Center for Next Generation, Chiba Institute of Technology, Chiba 275-0016, Japan, and also with the Frontier Research Center for Energy and Resources, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: kazuhide.mimura@p.chibakoudai.jp).

Kentaro Nakamura is with the Frontier Research Center for Energy and Resources, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan, also with the Department of Systems Innovation, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan, and also with the Ocean Resources Research Center for Next Generation, Chiba Institute of Technology, Chiba 275-0016, Japan (e-mail: kentaron@sys.t.u-tokyo.ac.jp).

Kazuhiro Takao is with the Department of Systems Innovation, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan.

Kazutaka Yasukawa is with the Frontier Research Center for Energy and Resources, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan, and also with the Department of Systems Innovation, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: k-yasukawa@sys.t.u-tokyo.ac.jp).

Yasuhiro Kato is with the Department of Systems Innovation, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan, also with the Ocean Resources Research Center for Next Generation, Chiba Institute of Technology, Chiba 275-0016, Japan, and also with the Submarine Resources Research Center, Research Institute for Marine Resources Utilization, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa 237-0061, Japan (e-mail: ykato@sys.t.u-tokyo.ac.jp).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSTARS.2023.3247467>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3247467

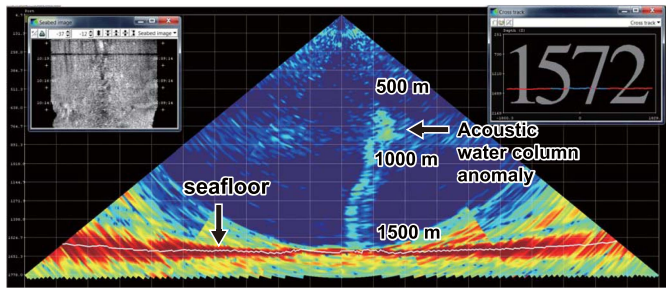


Fig. 1. Representative image of an acoustic water column anomaly observed by the R/V *Yokosuka* with an EM122 (12 kHz) system.

[13], [20], [21], [22], [23]. Therefore, exploration of seafloor massive sulfide deposits is important, not only as resources, but also for various scientific purposes.

Generally, exploration for seafloor resources is technically challenging and costly. In these circumstances, for seafloor massive sulfides, multibeam echo sounders (MBES) have attracted attention as a method for detecting hydrothermal signatures [24], [25]. MBES is an equipment that transmits multiple fan-shaped acoustic beams to the seafloor and measures the arrival time and intensity of backscatter waves. Although it was originally developed to measure bathymetry by detecting waves reflected from seafloor, it has been used for imaging the water column above the seafloor to detect and monitor various targets such as gas seep, shipwrecks, and marine organisms [26], [27], [28]. Surveys around Okinawa Trough revealed that vertical acoustic reflections were observed in the water column image above hydrothermal vents (see Fig. 1). Since the depth at which the acoustic reflections disappear corresponds to the phase-changing points of CO_2 from hydrate or liquid to gas, it is considered that CO_2 droplets released by hydrothermal activity are the cause of acoustic reflections [24].

In previous studies, water column backscatter data observed by the MBES were converted into color images, and potential hydrothermal signatures were detected by skilled observers. If this process was automated, it would not only improve the exploration efficiency for seafloor hydrothermal deposits but could also be applied to future fully autonomous exploration with autonomous underwater vehicles (AUVs) and autonomous surface vehicles (ASVs).

Recently, image recognition technology has demonstrated remarkable progress. Deep learning-based models can not only classify images but can also identify and locate various objects included within an image. This latter technique, called object detection, has been applied in various academic fields, such as medical image diagnosis [29], cell detection [30], and fossil observation [31], and can perform an automatic advanced image diagnosis with high accuracy, which was previously only possible for skilled observers. Therefore, we aimed to develop a deep learning model that could automatically and accurately detect signals of seafloor hydrothermal activity from MBES images. For this purpose, we compared several learning conditions and determined the optimal conditions.

II. EXPERIMENTAL METHODS

A. Data Description

MBES images acquired during research cruises YK14-17 and YK15-14 onboard the R/V *Yokosuka* were used as the dataset. Cruise YK14-17 was conducted in the area north of Okinawa Island in the Central Okinawa Trough (see Fig. 2), from August 19 to September 19, 2014, and YK15-14 was conducted west of Okinawa Island in the Central Okinawa Trough (see Fig. 2) from August 13 to 26, 2015.

An EM122 system (Kongsberg Maritime) was used to acquire the MBES data. The acoustic beams were emitted every 5 s. The colored heat map of reflection intensity, generated by the MBES system, were captured as images of 1600×740 pixels.

The previous study [24] defined the signal of hydrothermal activities in MBES images as one to several streams of acoustic scatters that rise vertically from seafloor to 500–1000 m of water depth without significant change in the width of stream. MBES images also include some patterns of “noise,” which are similar to signals, but are not caused by hydrothermal activities. Examples of noise include layering noise, which occurs horizontally, sidelobe, semicircular patterns that always occur at the sides of the water column, and background noise, which represents turbulence or reflections of particles in seawater [33] (supplementary material 1). Using these definitions, 1180 images containing at least one signal of hydrothermal activity were selected for the dataset in this article. We note here that the signals in the datasets were related to the actual hydrothermal sites confirmed by a manned submersible or a remotely operated vehicle. These images were randomly divided into three subsets: 800 images were used for the training dataset, 100 were used for validation, and 280 were used for testing. Images in each subset were annotated with the classes “signal,” indicating signs of hydrothermal activity, and “noise,” which has color patterns similar to those of “signal,” but with different shapes. The datasets used in this article are available on Mendeley Data [34].

B. Model Training Conditions

A total of 10 trainings and evaluations were performed by changing the three training conditions described below (see Table I).

1) *Architecture Selection*: Image recognition techniques that use deep learning include image classification; which determines the class of an image, object detection; which predicts the boxes that surround objects in images, semantic segmentation; which predict the position of objects at the pixel level, and instance segmentation; which is similar to semantic segmentation but can distinguish different objects. All of these techniques seem to be capable of determining whether an image contains signals of hydrothermal activities. However, considering the application of the deep learning system for actual exploration, the system should be able to determine the position and number of signals within an image. The techniques that meet this purpose are object detection and instance segmentation. Here, we selected Mask R-CNN [see Fig. 3(a)] [35] and YOLO-v5 [see Fig. 3(b)] [36] as representative architectures of instance-segmentation and object

TABLE I
TRAINING CONDITIONS AND EVALUATION RESULTS

Experimental No.	1	2	3	4	5	6	7	8	9	10
Conditions	Mask R-CNN				YOLO-v5					
Architecture										
Noise learning	×	○	×	○	×	○	×	○	×	○
layers/model	Heads	Heads	3+	3+	YOLOv5s	YOLOv5s	YOLOv5m	YOLOv5m	YOLOv5l	YOLOv5l
Trainable parameters	21.0M	21.0M	63.4M	63.4M	7.2M	7.2M	21.2M	21.2M	46.5M	46.5M
Performances										
Precision	0.119	0.338	0.307	0.539	0.833	0.896	0.899	0.901	0.885	0.928
Recall	0.714	0.629	0.862	0.83	0.841	0.868	0.863	0.878	0.835	0.881
F1 score	0.204	0.44	0.453	0.654	0.837	0.882	0.881	0.889	0.859	0.904

detection, respectively. In this article, Mask R-CNN was used for Cases 1–4, while YOLO-v5 was used for Cases 5–10 (see Table I).

2) *Learning Classes*: To investigate the importance of learning noise, we compared the model performances based on the dataset in which only signals were annotated [see Fig. 4(b)] and a dataset in which both signals and noise were annotated [see Fig. 4(c)]. In Cases 1, 3, 5, 7, and 9, the models learned only the signal information, whereas the models learned both signal and noise information in Cases 2, 4, 6, 8, and 10.

3) *Number of Trainable Parameters*: In general, machine learning models perform better when the number of trainable parameters increases. However, as the number of trainable parameters increases, a larger dataset is required to suppress overtraining [37]. Therefore, selecting the appropriate number of trainable parameters according to the size of the prepared dataset is effective for achieving a high performance.

In Mask R-CNN, the number of trainable parameters can be changed by selecting the trainable layers. The Mask R-CNN model can be divided into three major parts: the backbone, which extracts features from images; the region proposal network (RPN), which suggests candidate regions for objects; and the head, which predicts the classes and contours of the objects proposed by the RPN (see Fig. 5) [35]. In the “Heads” learning condition, the weights of layers in only the RPN and head are updated during training, while the “3+” condition trains a portion of the backbone in addition to those trained in “Heads.” For layers that were not trained, the weights were fixed to those of the model pretrained by COCO [38], which is a large dataset of general images. In this article, trainable layers were set to “Heads” for Cases 1 and 2, and to “3+” for Cases 3 and 4.

In YOLO-v5, several models with different numbers of parameters were provided. In this article, we trained the YOLO-v5s model with a small number (~ 7.2 million) of parameters in Cases 5 and 6. Similarly, we trained YOLO-v5m with a medium number (~ 21.2 million) of parameters in Cases 7 and 8, and YOLO-v5l with a large number (~ 46.5 million) of parameters in Cases 9 and 10.

C. Experimental Procedures

For all cases, we trained the model for up to 80 epochs using the training dataset. The training of Mask R-CNN was conducted

using ResNet-101 [39] as the backbone of the model. We used the default loss functions implemented for Mask R-CNN and YOLO-v5, respectively, both of which sum up the losses in locating objects and in determining classes of the objects [35], [36]. Based on preliminary examination [31], [36], the initial learning rate was set at 0.001 for Mask R-CNN and 0.01 for YOLO-v5. In all cases of training, the images were randomly flipped, scaled, and moved horizontally and vertically to prevent overlearning.

After training for each epoch, the model was evaluated using the validation dataset. In Mask R-CNN, the mean average precision (mAP) was monitored to determine the best epoch. In YOLO, the best model is automatically determined by a built-in function that measures the fitness between the predicted and original data, which basically uses mAPs.

Finally, the best-epoch model was utilized to detect signals in the test dataset. The evaluation was based on whether the signal was detected correctly, regardless of whether the model learned the noise. For both Mask R-CNN and YOLO, an intersection over union (IoU) of greater than 50% was considered a “correct” answer. To compare the performance of the two architectures under the same conditions, the IoU of both Mask R-CNN and YOLO models were evaluated based on the bounding box. The number of true positives (TPs), false positives (FPs), and false negatives (FNs) were counted, and precision, recall, and F1 scores were calculated according to the following equations:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{F1 score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}. \quad (3)$$

Precision represents the extent to which a model does not misclassify nonsignals as signals, recall represents the extent to which a model does not fail to detect an actual signal, and the F1 score is the harmonic mean of precision and recall, indicating the overall balance of the model.

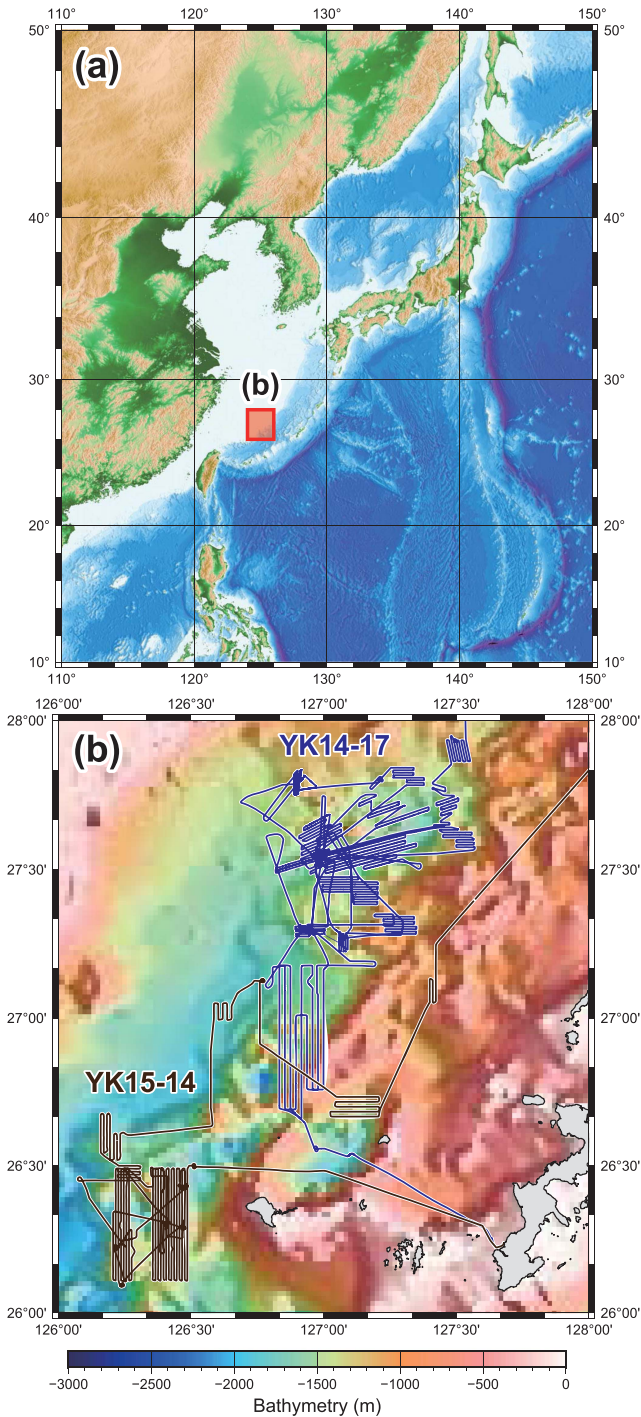


Fig. 2. (a) Map of the Western North Pacific Ocean. The red rectangle indicates the area surveyed in this article [enlarged in (b)]. (b) Map of the surveyed area with ship tracks. The YK14-17 track is indicated in blue and that of YK15-14 is in brown. The ETOPO1 dataset [32] was used to illustrate the bathymetry.

III. RESULTS AND DISCUSSION

A. Performance of the Trained Models

1) *Architectures*: The precisions, recalls, and F1 scores for all training cases are listed in Table I. Examples of the detected images are shown in supplementary material 2. The F1 scores of Mask R-CNN ranged from 0.20 to 0.65, which were much lower than those of YOLO-v5 (see Fig. 6, 0.84–0.90). This

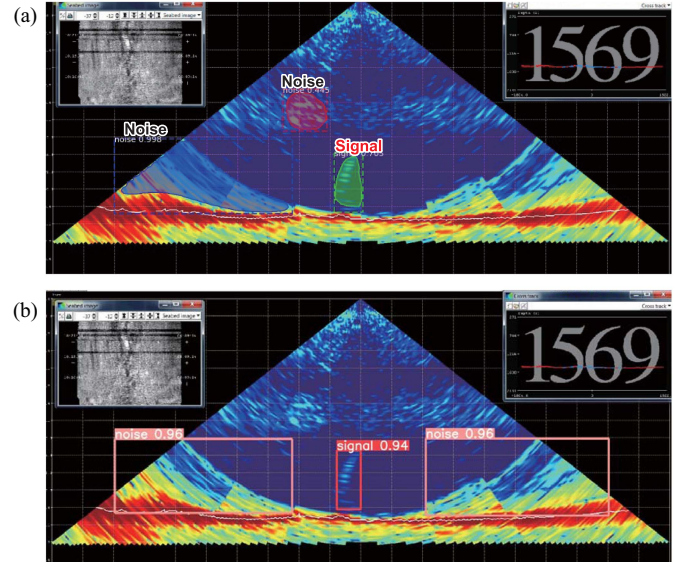


Fig. 3. Detection examples of (a) Mask R-CNN and (b) YOLO-v5 models, showing the differences in predicting regions of objects.

resulted from the low precision (0.119–0.539) of Mask R-CNN compared to YOLO-v5 (0.833–0.928). This means that the Mask R-CNN models made a number of errors of detecting nonsignals as signals. In contrast, the recall of Mask R-CNN ranged from 0.62 to 0.86, which was slightly lower or similar to that of YOLO-v5 (0.84–0.87).

The difference between the F1 scores of the Mask R-CNN and YOLO-v5 models may have been due to the differing complexity of the tasks. While YOLO performs a simple task of determining boxes surrounding objects, Mask R-CNN performs a more complex task of determining the region of the object at the pixel level (see Fig. 3). Therefore, with the limited dataset of approximately 1000 images in this article, YOLO might be able to achieve higher accuracy.

2) *Effects of Noise Learning*: In Mask R-CNN, models that learned noise had higher precisions and F1 scores by more than 20% and had lower recall by several percent [see Fig. 7(a)]. As the total balance of the model indicated by the F1 score increased, noise learning was considered to be effective. Cases 1 and 3, in which the model did not learn noise, were characterized by very low precisions and slightly higher recalls compared to Cases 2 and 4. This suggests that all suspicious regions were detected as signals. In this scenario, learning negative (nonsignal) objects would suppress detection and yield a more balanced model.

In YOLO, models that learned noise had higher evaluation parameters by several percent, suggesting that noise learning was also effective for YOLO [see Fig. 7(b)]. However, the models without noise learning also had relatively high precisions. This suggests that the models that learned only signals were already able to distinguish between the signal and noise to some extent. Therefore, the effect of noise learning was not as high as that of Mask R-CNN. Interestingly, recalls also increased with noise learning in YOLO, suggesting that learning noise may have refined the criteria between the signal and noise more accurately.

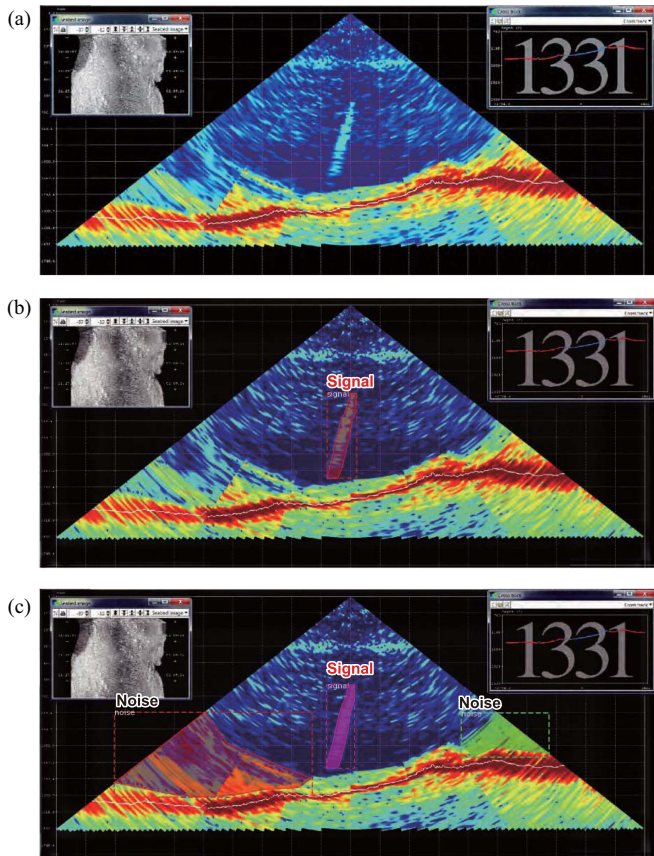


Fig. 4. Example of an (a) original and (b), (c) annotated images. While the locations of only signals were annotated in the “signal only” dataset (b), locations of both signals and noise were annotated in the “with noise” dataset (c).

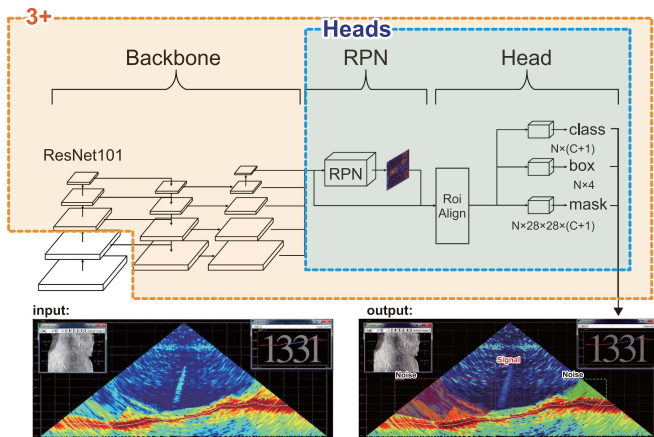


Fig. 5. Schematic illustration of the mask R-CNN architecture. Layers that are trainable in conditions “3+” are highlighted in orange, while those trainable in condition “heads” are highlighted in blue. N indicates the number of objects detected in an image and C indicates the total number of classes.

3) *Effects of the Numbers of Trainable Parameters:* For Mask R-CNN, increasing the number of trainable parameters resulted in increases in all evaluation indices [see Fig. 8(a)]. Under the condition of only training the “Heads” layers, the weights of the backbone that extract features of the image are fixed to the weight pretrained by the huge dataset of general

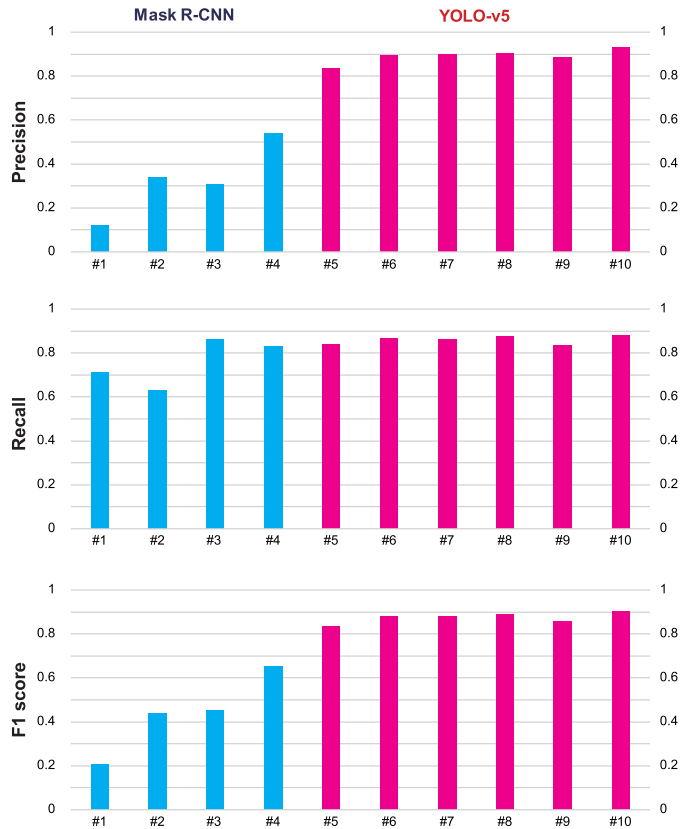


Fig. 6. Bar charts showing precisions, recalls and F1 scores for all the training cases. The performances of mask R-CNN models are colored in cyan, while those of YOLO-v5 models are colored in magenta.

images (cf., Fig. 5). Owing to the heterogeneity of the MBES image from such general images the Mask R-CNN model trained only on “Heads” may not have been able to extract features from the MBES images effectively.

In YOLO, the model YOLO-v5m demonstrated the highest F1 score when only learning signals, whereas YOLO-v5l had the highest F1 score when learning both signals and noise [see Fig. 8(b)]. This suggests that YOLO-v5m and YOLO-v5l are suitable for a dataset with a scale of approximately 1000 images. The best number of parameters differed depending on the number of learning classes. A model with relatively few parameters (YOLO-v5 m) may be suitable for simple single-class detection, whereas a model with a large number of parameters (YOLO-v5l) may be suitable for multiclass detection.

B. Applicability to Actual Exploration

Of the ten cases examined in this article, Case 10 had the highest precision, recall, and F1 score, which were 0.928, 0.881, and 0.904, respectively. Here, we further investigate the applicability of this model for actual exploration of hydrothermal activities.

First, signals rarely occur upon practical exploration, which requires the model to minimize the generation of false alerts from non-signal images. To evaluate this perspective, the existence of the signal was inferred from 1016 images that actually do not contain a signal by the best model. As a result, false signals

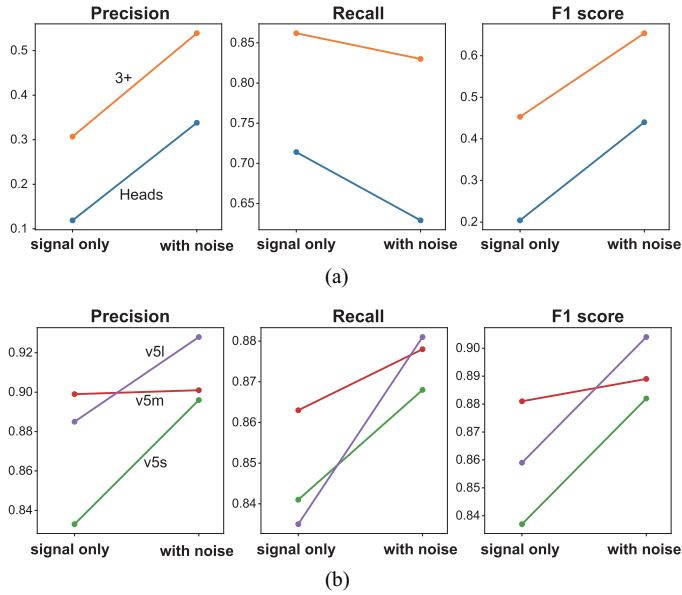


Fig. 7. Comparisons of precision, recall and F1 scores of the models that learned only signals and models that learned both signals and noises. Architectures are (a) Mask R-CNN and (b) YOLO-v5.

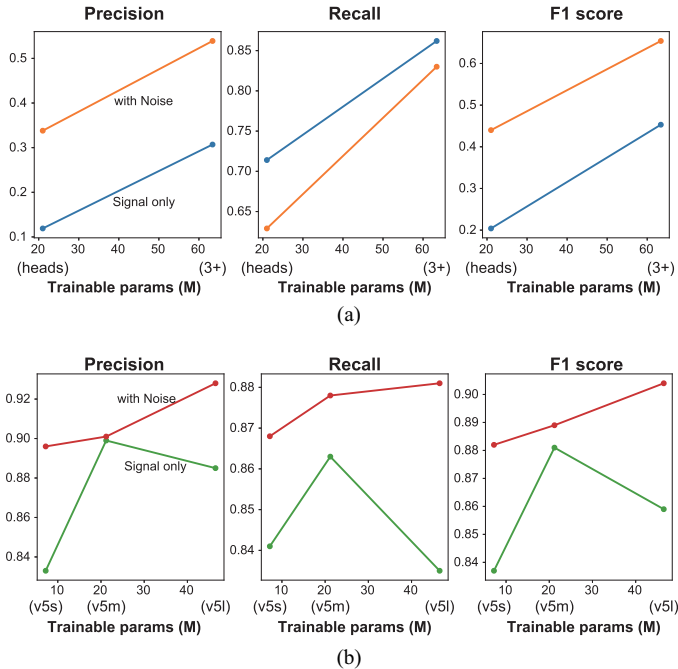


Fig. 8. Precisions, recalls, and F1 scores of the models with different numbers of trainable parameters. Architectures are (a) Mask R-CNN and (b) YOLO-v5.

were generated from only seven images, with the false alarm rate being 0.69%. Therefore, by using this model, the exploration of hydrothermal activity will not be severely hampered by a large number of false positives.

Second, in terms of the prediction output format, YOLO only encloses the region of the acoustic anomaly with a box, whereas Mask R-CNN can obtain the contours of the anomaly at the pixel level. This means that Mask R-CNN can identify the shapes of the anomalies more accurately. However, the main purpose

of the MBES survey is to detect the locations of hydrothermal vents and determining the shapes of the anomalies is not a high priority. Therefore, the output of YOLO is sufficient for detecting hydrothermal activity.

Finally, to determine the detection speed, we performed a detection test from a movie file using the model from Case 10 (supplementary material 3). On the Google Collaboratory platform, the average detection time using a Tesla-P100 GPU (NVIDIA Corporation) was 20–25 ms/frame, which is generally faster than the pace of typical MBES image generation. Therefore, the model from Case 10 can instantly recognize the signs of hydrothermal activity from images generated by MBES and can be used for real-time detection.

Since a single commercial GPU card is sufficiently fast, the detection system can be built in a compact and power-efficient manner without a huge machine learning server or power supply unit. This suggests that detecting seafloor massive sulfide deposits can be achieved using a relatively small-scale platform. AUVs and ASVs are small-scale platforms that have made remarkable progress recently [40], [41]. If the detection system proposed herein could be loaded onto AUVs, they would not only be able to automatically predict the locations of hydrothermal vents by detecting acoustic anomalies from the sea surface but would also be able to dive to the prospective area and perform a camera survey to confirm the presence of a deposit. Such fully automated exploration is very cost-effective compared to conventional exploration using research vessels and piloted submersibles and has the potential to accelerate future exploration of seafloor massive sulfide deposits.

IV. CONCLUSION

We trained deep learning models to automatically detect signals of seafloor hydrothermal activity from MBES images under ten training conditions and evaluated their performances. When comparing the Mask R-CNN and YOLO-v5 model architectures, YOLO-v5 presented a better performance. For the performances of models that learned only signals and those that learned both signals and noises, the F1 scores of the cases with noise learning were higher than those of cases without noise learning in all comparison pairs, indicating that noise learning was effective. When considering the number of trainable parameters, the YOLO-v5l model with a large number of trainable parameters was the most suitable. The best model had high performance scores, with a precision of 0.928, a recall of 0.881, and an F1 score of 0.904. In addition, this model achieved a low false alarm rate and had a high detection speed (20–25 ms per frame), indicating that the model is applicable for automated exploration of seafloor massive sulfide deposits.

ACKNOWLEDGMENT

The authors would like to thank the crew of the *R/V Yokosuka* for their technical support during the research cruises. And also like to thank H. Okamoto for providing invaluable insight into signal detection through preliminary discussions.

REFERENCES

- [1] M. Hannington, J. Jamieson, T. Monecke, S. Petersen, and S. Beaulieu, "The abundance of seafloor massive sulfide deposits," *Geology*, vol. 39, no. 12, pp. 1155–1158, Dec. 2011, doi: [10.1130/G32468.1](https://doi.org/10.1130/G32468.1).
- [2] J. R. Hein, K. Mizell, A. Koschinsky, and T. A. Conrad, "Deep-ocean mineral deposits as a source of critical metals for high-and green-technology applications: Comparison with land-based resources," *Ore Geol. Rev.*, vol. 51, pp. 1–14, Dec. 2012, doi: [10.1016/j.oregeorev.2012.12.001](https://doi.org/10.1016/j.oregeorev.2012.12.001).
- [3] J. R. Hein, A. Koschinsky, and T. Kuhn, "Deep-ocean polymetallic nodules as a resource for critical materials," *Nature Rev. Earth Environ.*, vol. 1, no. 3, pp. 158–169, Feb. 2020, doi: [10.1038/s43017-020-0027-0](https://doi.org/10.1038/s43017-020-0027-0).
- [4] Y. Kato et al., "Deep-sea mud in the Pacific Ocean as a potential resource for rare-earth elements," *Nature Geosci.*, vol. 4, no. 8, pp. 535–539, Jul. 2011, doi: [10.1038/ngeo1185](https://doi.org/10.1038/ngeo1185).
- [5] S. Machida, K. Fujinaga, T. Ishii, K. Nakamura, N. Hirano, and Y. Kato, "Geology and geochemistry of ferromanganese nodules in the Japanese exclusive economic zone around Minamitorishima Island," *Geo-Chem. J.*, vol. 50, no. 6, pp. 539–555, 2016, doi: [10.2343/geochemj.2.0419](https://doi.org/10.2343/geochemj.2.0419).
- [6] Y. Takaya et al., "The tremendous potential of deep-sea mud as a source of rare-earth elements," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, Apr. 2018, doi: [10.1038/s41598-018-23948-5](https://doi.org/10.1038/s41598-018-23948-5).
- [7] T. Nozaki et al., "Subseafloor sulphide deposit formed by pumice replacement mineralisation," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Apr. 2021, doi: [10.1038/s41598-021-87050-z](https://doi.org/10.1038/s41598-021-87050-z).
- [8] T. Nozaki et al., "Formation of highly Zn-enriched sulfide scale at a deep-sea artificial hydrothermal vent, Iheya-North Knoll, Okinawa Trough," *Mineralium Deposita*, vol. 56, no. 5, pp. 975–990, Oct. 2020, doi: [10.1007/s00126-020-01022-3](https://doi.org/10.1007/s00126-020-01022-3).
- [9] L. Toffolo, P. Nimis, G. A. Tret'yakov, I. Y. Melekestseva, and V. E. Beltenev, "Seafloor massive sulfides from mid-ocean ridges: Exploring the causes of their geochemical variability with multivariate analysis," *Earth-Sci. Rev.*, vol. 201, Feb. 2020, Art. no. 102958, doi: [10.1016/j.earscirev.2019.102958](https://doi.org/10.1016/j.earscirev.2019.102958).
- [10] K. A. Miller, K. F. Thompson, P. Johnston, and D. Santillo, "An overview of seabed mining including the current state of development, environmental impacts, and knowledge gaps," *Front. Mar. Sci.*, vol. 4, Jan. 2018, Art. no. 418, doi: [10.3389/fmars.2017.00418](https://doi.org/10.3389/fmars.2017.00418).
- [11] T. Nozaki et al., "Rapid growth of mineral deposits at artificial seafloor hydrothermal vents," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, Feb. 2016, doi: [10.1038/srep22163](https://doi.org/10.1038/srep22163).
- [12] H. Elderfield and A. Schultz, "Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean," *Annu. Rev. Earth Planet. Sci.*, vol. 24, pp. 191–224, May 1996, doi: [10.1146/annurev.earth.24.1.191](https://doi.org/10.1146/annurev.earth.24.1.191).
- [13] K. Takai, K. Nakamura, K. Suzuki, F. Inagaki, K. H. Nealson, and H. Kumagai, "Ultramafics-Hydrothermalism-Hydrogenesis-HyperSLiME (UltraH3) linkage: A key insight into early microbial ecosystem in the Archean deep-sea hydrothermal systems," *Paleontological Res.*, vol. 10, no. 4, pp. 269–282, Jan. 2009, doi: [10.2517/prpsj.10.269](https://doi.org/10.2517/prpsj.10.269).
- [14] K. Takai and K. Nakamura, "Archaeal diversity and community development in deep-sea hydrothermal vents," *Curr. Opin. Microbiol.*, vol. 14, no. 3, pp. 282–291, Jun. 2011, doi: [10.1016/j.mib.2011.04.013](https://doi.org/10.1016/j.mib.2011.04.013).
- [15] C. G. Wheat, J. McManus, M. J. Mottl, and E. Giambalvo, "Oceanic phosphorus imbalance: Magnitude of the mid-ocean ridge flank hydrothermal sink," *Geophys. Res. Lett.*, vol. 30, no. 17, Sep. 2003, doi: [10.1029/2003GL017318](https://doi.org/10.1029/2003GL017318).
- [16] D. Hasterok, D. S. Chapman, and E. E. Davis, "Oceanic heat flow: Implications for global heat loss," *Earth Planet. Sci. Lett.*, vol. 311, no. 3–4, pp. 386–395, Nov. 2011, doi: [10.1016/j.epsl.2011.09.044](https://doi.org/10.1016/j.epsl.2011.09.044).
- [17] R. A. Lutz and M. J. Kennish, "Ecology of deep-sea hydrothermal vent communities: A review," *Rev. Geophys.*, vol. 31, no. 3, pp. 211–242, Aug. 1993, doi: [10.1029/93RG01280](https://doi.org/10.1029/93RG01280).
- [18] C. R. Fisher, K. Takai, and N. L. Bris, "Hydrothermal vent ecosystems," *Oceanography*, vol. 20, no. 1, pp. 14–23, 2007, doi: [10.5670/oceanog.2007.75](https://doi.org/10.5670/oceanog.2007.75).
- [19] G. J. Dick, K. Anantharaman, B. J. Baker, M. Li, D. C. Reed, and C. S. Sheik, "The microbiology of deep-sea hydrothermal vent plumes: Ecological and biogeographic linkages to seafloor and water column habitats," *Front. Microbiol.*, vol. 4, May 2013, Art. no. 124, doi: [10.3389/fmicb.2013.00124](https://doi.org/10.3389/fmicb.2013.00124).
- [20] D. M. Karl, "The microbiology of deep-sea hydrothermal vents," in *Ecology of Free-Living, Hydrothermal Vent Microbial Community*. Boca Raton, FL, USA: CRC Press, 1995, pp. 35–124.
- [21] W. Martin, J. Baross, D. Kelley, and M. J. Russell, "Hydrothermal vents and the origin of life," *Nature Rev. Microbiol.*, vol. 6, no. 11, pp. 805–814, Nov. 2008, doi: [10.1038/nrmicro1991](https://doi.org/10.1038/nrmicro1991).
- [22] K. Nakamura and K. Takai, "Theoretical constraints of physical and chemical properties of hydrothermal fluids on variations in chemolithotrophic microbial communities in seafloor hydrothermal systems," *Prog. Earth Planet. Sci.*, vol. 1, no. 1, pp. 1–24, Apr. 2014, doi: [10.1186/2197-4284-1-5](https://doi.org/10.1186/2197-4284-1-5).
- [23] K. Nakamura and K. Takai, "Geochemical constraints on potential biomass sustained by subseafloor water-rock interactions," in *Subseafloor Biosphere Linked to Hydrothermal Systems*. Berlin, Germany: Springer-Verlag, 2014, doi: [10.1007/978-4-431-54865-2_2](https://doi.org/10.1007/978-4-431-54865-2_2).
- [24] K. Nakamura, S. Kawagucci, K. Kitada, H. Kumagai, K. Takai, and K. Okino, "Water column imaging with multibeam echo-sounding in the mid-Okinawa Trough: Implications for distribution of deep-sea hydrothermal vent sites and the cause of acoustic water column anomaly," *Geo-Chem. J.*, vol. 49, no. 6, pp. 579–596, Dec. 2015, doi: [10.2343/geochemj.2.0387](https://doi.org/10.2343/geochemj.2.0387).
- [25] T. Kasaya, H. Machiyama, K. Kitada, and K. Nakamura, "Trial exploration for hydrothermal activity using acoustic measurements at the North Iheya Knoll," *Geo-Chem. J.*, vol. 49, no. 6, pp. 597–602, Dec. 2015, doi: [10.2343/geochemj.2.0389](https://doi.org/10.2343/geochemj.2.0389).
- [26] J. H. Clarke, "Applications of multibeam water column imaging for hydrographic survey," *Hydrographic J.*, vol. 120, no. 3, 2006.
- [27] K. Colbo, T. Ross, C. Brown, and T. Weber, "A review of oceanographic applications of water column data from multibeam echosounders," *Estuarine, Coastal Shelf Sci.*, vol. 145, pp. 41–56, May 2014, doi: [10.1016/j.ecss.2014.04.002](https://doi.org/10.1016/j.ecss.2014.04.002).
- [28] A. W. Nau, B. Scoulding, R. J. Kloser, Y. Ladroit, and V. Lucieer, "Extended detection of shallow water gas seeps from multibeam echosounder water column data," *Front. Remote Sens.*, vol. 3, pp. 1–18, Jul. 2022, doi: [10.3389/frsen.2022.839417](https://doi.org/10.3389/frsen.2022.839417).
- [29] Y. Sakai et al., "Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 4138–4141, doi: [10.1109/EMBC.2018.8513274](https://doi.org/10.1109/EMBC.2018.8513274).
- [30] I. Suleymanova et al., "A deep convolutional neural network approach for astrology detection," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 12878, doi: [10.1038/s41598-018-31284-x](https://doi.org/10.1038/s41598-018-31284-x).
- [31] K. Mimura, S. Minabe, K. Nakamura, K. Yasukawa, J. Ohta, and Y. Kato, "Automated detection of microfossil fish teeth using combined deep-learning models," *Appl. Comput. Geosci.*, vol. 16, Dec. 2022, Art. no. 100092, doi: [10.1016/j.acags.2022.100092](https://doi.org/10.1016/j.acags.2022.100092).
- [32] NOAA National Geophysical Data Center, "ETOPO1 1 arc-minute global relief model. NOAA national centers for environmental information," 2009. Accessed: Nov. 08, 2022. [Online]. Available: <https://www.ngdc.noaa.gov/mgg/global>
- [33] J. Zhao, J. Meng, H. Zhang, and S. Wang, "Comprehensive detection of gas plumes from multibeam water column images with minimisation of noise interferences," *Sensors*, vol. 17, no. 12, Nov. 2017, Art. no. 2755, doi: [10.3390/s17122755](https://doi.org/10.3390/s17122755).
- [34] K. Mimura and K. Nakamura, "Datasets for hydrothermal plume detection," *Mendelay Data*, vol. 1, 2022, doi: [10.17632/dg2595f68b.1](https://doi.org/10.17632/dg2595f68b.1).
- [35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [36] G. Jocher et al., "Ultralytics/yolov5: V6.2 - yolov5 classification models, apple M1, reproducibility, clearml and deci.ai integrations," 2022, doi: [10.5281/zenodo](https://doi.org/10.5281/zenodo).
- [37] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
- [38] T. Y. Lin et al., "Microsoft coco: Common objects in context," in *Proc. Comput. Vis. – ECCV 13th Eur. Conf.*, 2014, pp. 6–12, doi: [10.1007/978-3-319-10602-1_48](https://doi.org/10.1007/978-3-319-10602-1_48).
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [40] T. Matsuda, T. Maki, and T. Sakamaki, "Accurate and efficient seafloor observations with multiple autonomous underwater vehicles: Theory and experiments in a hydrothermal vent field," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2333–2339, Jul. 2019, doi: [10.1109/LRA.2019.2902744](https://doi.org/10.1109/LRA.2019.2902744).
- [41] T. Wu et al., "A hydrothermal investigation system for the Qianlong-II autonomous underwater vehicle," *Acta Oceanologica Sinica*, vol. 38, no. 3, pp. 159–165, 2019, doi: [10.1007/s13131-019-1408-4](https://doi.org/10.1007/s13131-019-1408-4).

Kazuhide Mimura received the bachelor's, master's, and Ph.D. degrees in engineering from The University of Tokyo, Tokyo, Japan, in 2016, 2018, and 2021, respectively.

He is currently a Chief Research Scientist with the Ocean Resource Research Center for Next Generation, Chiba Institute of Technology, Japan.

Kentaro Nakamura received the bachelor's and master's degrees of science from Yamaguchi University, Yamaguchi, Japan, in 1997 and 1999, respectively, and the Ph.D. degree in engineering from The University of Tokyo, Tokyo, Japan, in 2004.

He is currently an Associate Professor with the Frontier Research Center for Energy and Resources, The University of Tokyo.

Kazuhiro Takao received the master's degree in engineering from The University of Tokyo, Tokyo, Japan, in 2020.

He is currently with a company in Japan.

Kazutaka Yasukawa received the bachelor's, master's, and Ph.D. degrees in engineering from The University of Tokyo, Tokyo, Japan, in 2008, 2010, and 2015, respectively.

He is currently an Associate Professor with the Frontier Research Center for Energy and Resources, The University of Tokyo.

Yasuhiro Kato received the bachelor's, master's, and Ph.D. degrees of science from The University of Tokyo, Tokyo, Japan, in 1985, 1987, and 1990, respectively.

He is currently a Professor with the Department of Systems Innovation, School of Engineering, The University of Tokyo.