https://doi.org/10.2352/EI.2024.36.15.COIMG-162

Open Set Domain Adaptation for Image Classification with Multiple Unknown Labels Using Unsupervised Clustering in a Target Domain

Daichi Nishihara; Osaka University; Suita, Osaka, Japan
Yoshihiro Midoh; Osaka University; Suita, Osaka, Japan
Youyang Ng; Kioxia Corporation, Kanagawa, Japan
Osamu Yamane; Kioxia Corporation, Kanagawa, Japan
Maasa Takahashi; Kioxia Corporation, Kanagawa, Japan
Shuhei Iijima; Kioxia Corporation, Kanagawa, Japan
Jun Shiomi; Osaka University; Suita, Osaka, Japan
Goh Itoh; Kioxia Corporation, Kanagawa, Japan
Noriyuki Miura; Osaka University; Suita, Osaka, Japan

Abstract

Domain adaptation, which transfers an existing system with teacher labels (source domain) to another system without teacher labels (target domain), has garnered significant interest to reduce human annotations and build AI models efficiently. Open set domain adaptation considers unknown labels in the target domain that were not present in the source domain. Conventional methods treat unknown labels as a single entity, but this assumption may not hold true in real-world scenarios. To address this challenge, we propose open set domain adaptation for image classification with multiple unknown labels. Assuming that there exists a discrepancy in the feature space between the known labels in the source domain and the unknown labels in the target domain based on their type, we can leverage clustering to classify the types of unknown labels by considering the pixel-wise feature distances between samples in the target domain and the known labels in the source domain. This enables us to assign pseudo-labels to target samples based on the classification results obtained through unsupervised clustering with an unknown number of clusters. Experimental results show that the accuracy of domain adaptation is improved by re-training using these pseudo-labels in a closed set domain adaptation setting.

Introduction

AI-based image classification is widely used in advanced measurement areas such as semiconductor defect inspection [1]. A large amount of data acquisition and data annotation through manual labeling by human annotators, are required to train accurate models. To reduce the cost of training data creation and increase development throughput, it is desirable to be able to transfer machine learning models constructed in the existing equipment to the next-generation system as much as possible [2].

In advanced measurement instruments, image quality to be tested may be improved in terms of signal-to-noise ratio and image resolution by improving an existing system or may be degraded by modifying the equipment closer to the detection limit. For example, noise will increase due to high-speed scanning for dynamic observation in electron microscopy. It is also highly likely that unknown labels or new defects that did not exist in the training data set will arise due to new manufacturing processes and different equipment applications.

Therefore, open set domain adaptation (DA) with unknown labels in the target domain is attracting attention. The domain is a set of data with certain characteristics. When there are differences in the frequency of label occurrence or the distribution of pixel values in an image sample, it is considered to be a different domain. DA is a technique to transfer a machine learning model trained on a data set (source domain) to other small data sets (target domain) [3]. Data in the source domain is usually abundant and has labels serve as teachers, but data in the target domain has no labels (although labels exist for benchmark data, they are not used for training but only for evaluating classification performance). When the source and target domains share the same label and the target domain contains unknown labels, it is called closed set DA and open set DA, respectively. Figure 1 shows a schematic diagram of open set DA. DA is performed by aligning each sample in the target domain closer to a similar known label in the source domain while target samples with the unknown label are separated. As shown in Figure 1(b), conventional methods assume that there is only one unknown label [4]. Target samples that do not fit any of the known labels in the source domain are judged as unknown label. However, there can be multiple unknown labels in actual operation.

We propose a DA method for image classification with multiple unknown labels as shown in Figure 1(c). If unknown labels could be assigned to each target sample without human labeling, it would be possible to improve the accuracy of image classification and streamline the process of checking the prediction results and providing feedback. T. Jing et al. have also proposed a method to detect unknown labels in the target domain using

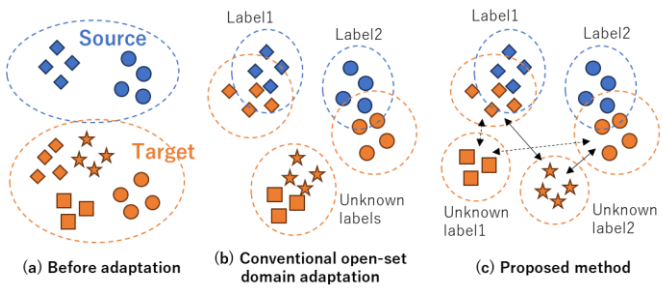


Figure 1. *A schematic illustration of open set domain adaptation.*

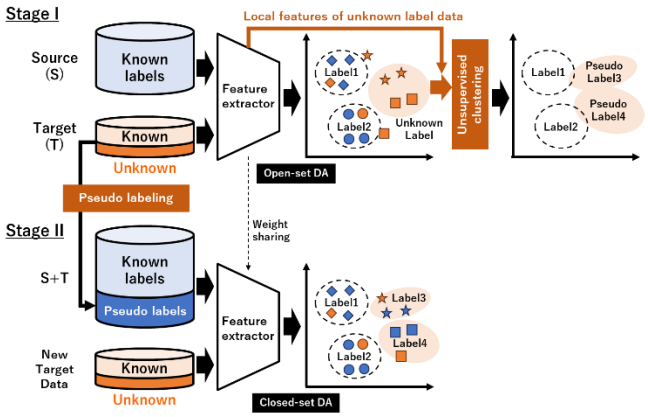


Figure 2. *An overview of our proposed method. Our approach consists of two stages, open set DA and closed set DA implemented in sequence. Target samples that are determined as unknown label in the first open set DA are given pseudo-labels by unsupervised clustering based on local features, and then closed set DA is applied.*

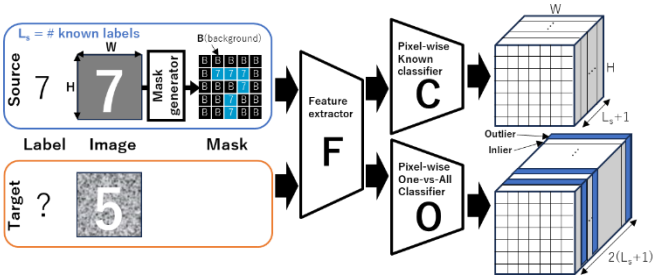
structure preserving partial alignment [5]. Our method utilizes unsupervised clustering based on segmentation-based local features to assign each unknown label.

This paper gives an overview of the proposed method and shows its effectiveness using modified MNIST dataset [6]. The outline of this work is organized as follows. Methods section introduces processing steps of the proposed method using pseudo labeling based on unsupervised clustering in a target domain. Results section is devoted to experimental results. Finally, the conclusion is drawn in Concluding Remarks section.

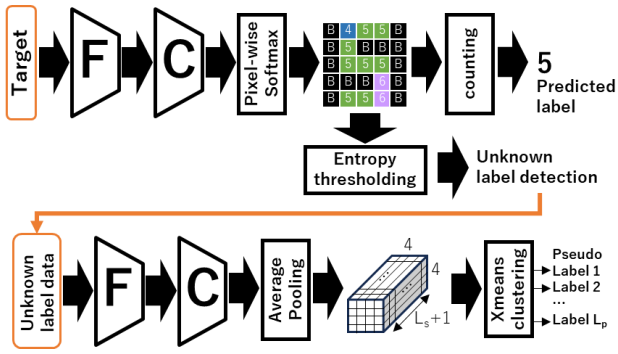
Methods

We extend the previously proposed DA for single unknown label called OVANet [7] to achieve image classification for multiple unknown labels. Figure 2 shows an overview of our proposed method. OVANet has two classifiers, one known classifier (C) that predicts which of the known labels in source domain it fits into and the other One-vs-All classifier (O) that determines whether it is inside or outside the known labels. During training, C is trained to be more likely to predict the correct label, while O is more likely to classify whether a target sample is an inlier or an outlier of each known label. Here, the inlier and outlier probabilities add up to 1 and are trained so that one of them approaches 1. The target samples are classified as either a known label or a single unknown label. In the proposed method, both classifiers C and O are replaced with a pixel-wise configuration used widely for image segmentation to extract local features proactively. Such local features are considered effective in separating multiple unknown labels. Therefore, the target samples determined as unknown label by open set DA in stage I are classified by unsupervised clustering with an unknown number of labels. Based on the obtained clustering results, pseudo-labels are given to the initial target samples and DA in stage II is performed as a closed set DA again.

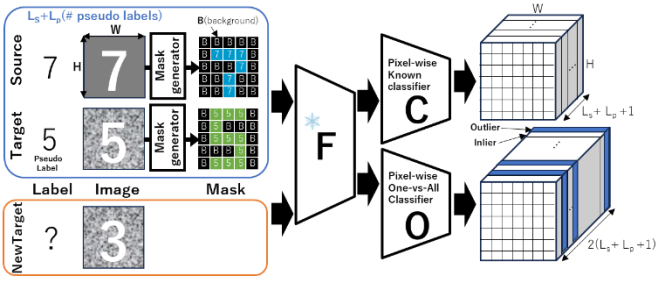
Figure 3 illustrates the processing steps of the proposed method. Firstly, a ground truth mask is created by a mask generator so that loss functions can be calculated on a pixel-by-pixel basis as in image segmentation (Fig. 3 (a)). Here, we use simple thresholding as the mask generator. If a target pixel has a pixel value above the threshold, a ground truth label is given to the pixel, and otherwise a background label is given. Classifier C is trained to correctly classify each pixel on source samples as either one of the known labels (L_s) or the background label by minimizing the cross-entropy, and classifier O is trained to make inlier of positive (known label) target samples and outlier of hardest negative class (possible unknown label) closer to 1 using the hard negative classifier sampling of OVANet. O has two outputs, inlier and outlier probabilities, for each label in C. Therefore, output dimensions of classifiers C and O are [W, H, L_s+1] and [W, H, 2(L_s+1)], respectively. Secondly, image segmentation is performed based on local features obtained from



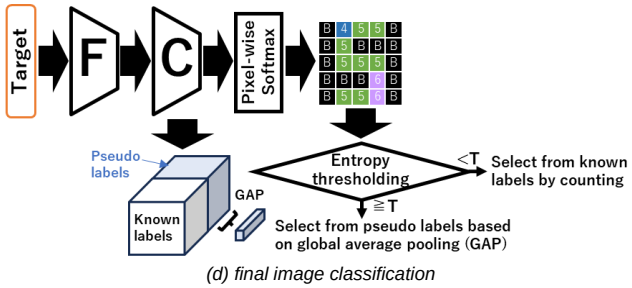
(a) open set DA in the stage I



(b) unsupervised clustering



(c) closed set DA in the stage II



(d) final image classification

Figure 3. *Illustration of the processing steps of the proposed method.*

Figure 4. *Modified MNIST dataset for evaluation of DA image classification: (a) original image, (b) gray scaled image (source domain), (c) noisy gray scaled image (target domain).*

classifier C and the entropy of the segmented labels is calculated (Fig. 3 (b)). If it is above a threshold, an unknown label is assigned, otherwise, the most frequent label is used as the predicted label. Local features of target samples determined as the unknown label is dimensionally compressed by average pooling.

Unsupervised clustering is applied to the compressed feature vectors obtained from all unknown label target samples. The number of unknown labels L_p is automatically determined (by X-means [8] etc.) and pseudo-labels are assigned according to the characteristics of the unknown labels. Thirdly, target samples given pseudo-labels are added to the source domain, and closed set DA is applied to the new samples in the target domain (Fig. 3 (c)). Here, network weights of feature extractor (F) are frozen during the training process. Output dimensions of classifiers C and O in stage II are [W, H, L_s+ L_p+1] and [W, H, 2(L_s+ L_p+1)], respectively. Finally, image classification is performed based on the entropy of segmented labels (Fig. 3 (d)). If the entropy is above the threshold, one of pseudo-labels is given to the target sample so that maximizes the value of the global average pooling of local features.

Results

We applied the proposed method to modified hand-written number image dataset, MNIST (# labels = 10) [6] as shown in Figure 4. Two different domain image datasets were created through image processing. Gray scaled images in source domain were converted from black and white into the contrast range as often seen in semiconductor metrology using scanning electron microscope. Noisy gray scaled images in target domain were also generated by adding uniform random noise. We evaluated these image datasets as different domains. The number of samples per label in source and target domains were 350 and 100, respectively. When the number of unknown labels in the target domain is 2, the sample sizes of source and target domains in the stage I are 2,800 (350 x 8) and 1,000 (100 x 10), and then target samples classified as unknown labels in the stage I are added to the source domain in the stage II. The target domain in the stage II includes new unused samples. Image classification of open and closed set DA was applied to common test samples in the target domain that were not used in the training process.

In the following, we compared the classification performance of open and closed sets DA changing the pairs of unknown labels in target domain.

Unknown labels: 0 and 1

It was assumed that the gray scaled image dataset in the source domain has 8 known labels (2-9) and the gray scaled image dataset in the target domain has the same known labels and two unknown labels (0 and 1). Figure 5 shows experimental results.

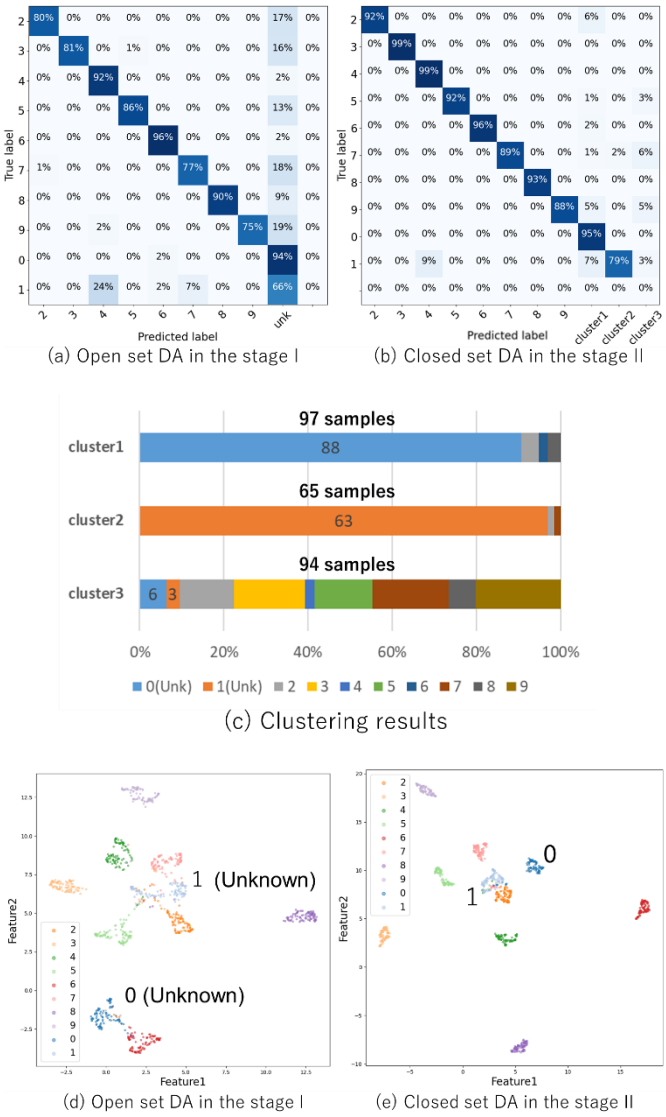


Figure 5. *Experimental results of image classification with unknown labels 0 and 1: confusion matrix of (a) open set DA in the stage I and (b) closed set DA in the stage II, (c) clustering results by X-means, and UMAP feature map of known classifier outputs for (d) open and (e) closed set DA.*

Figures 5 (a) and (b) represents confusion matrices classified by open set DA in the stage I and closed set DA in the stage II, respectively. The average classification accuracies of known labels using open and closed set DA were 84.6% and 93.5%. Also, clustering results for unknown determined target samples are shown in figure 5 (c). Unsupervised unknown label clustering was applied to the 256 samples classified as unknown labels by open set DA. The result classified 97, 65 and 94 samples into cluster 1, 2 and 3, respectively. Although the number of unknown labels was mistakenly estimated at 3, clusters 1 and 2 contain the majority of 0 and 1, respectively and cluster 3 include misclassified known labels. These results suggest that the proposed method succeeded in roughly classifying unknown labels. Figures 5 (d) and (e) show dimension-compressed feature maps of outputs of pixel-wise known classifier trained by (d) open set DA in the stage I and (e) closed set DA in the stage II using UMAP [9]. It can be seen from

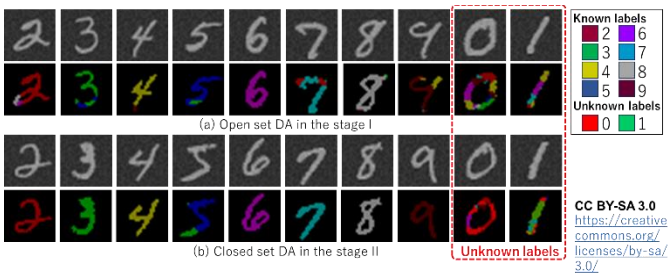


Figure 6. *Comparison of segmentation results obtained by pixel-wise softmax of known classifier C for (a) open and (b) closed set DA with unknown labels 0 and 1.*

figure 5 (d) that unknown labels 0 and 1 form separate clusters. When open set DA assumed a single unknown label is applied to data with multiple unknown labels, it is thought that the feature map of the unknown label samples will expand, and it will be easier to misclassify known labels as unknown. From figure 5 (b), it can be seen that using the pseudo labeling, the two unknown labels 0 and 1 can be separated with a correct answer rate of 95% and 79%, respectively. Although several target samples with known labels were misclassified by clustering, the classification accuracy of known labels did not decrease, but rather improved (8.9%). Very few unknown labeled samples were classified into fake cluster 3. We can see that UMAP features of both known and unknown labels are clustered compared with open set DA (Figs. 5 (d) and (e)).

Figure 6 compares segmentation results obtained by pixel-wise softmax of known classifier C in the unsupervised clustering step based on open set DA (Fig. 3 (b)) and the final image classification step based on closed set DA (Fig. 3 (d)), respectively. For known label samples, segmentation results reflect local features according to individual numeric parts. Even with open set DA, as shown in figure 6 (a), there are some degree of differences in local features between 0 and 1 which leads to high entropy. These local differences can also have contributed to the separation in unsupervised clustering. The closed set DA given pseudo-labels promotes learning of the shape of the unknown labels and better segmentation results (Fig. 6 (d)). These results suggest that the proposed method is effective for DA of multiple unknown labels.

Unknown labels: 3 and 5

As in the previous section, it was assumed that the gray scaled image dataset in the source domain has 8 known labels (0-2, 4, 6-9) and the noisy gray scaled image dataset in the target domain has the same known labels and two unknown labels (3 and 5). Figure 7 shows experimental results. As shown by the confusion matrices of figures 7 (a) and (b), the average classification accuracies of known labels using open and closed set DA were 89.6% and 95.5%, respectively. Although several target samples with known labels were misclassified by clustering, the number of unknown labels was correctly estimated at 2, and the classification accuracy of known labels improved by 5.9%. The two unknown labels 3 and 5 were also separated with a correct answer rate of 78% and 72%, respectively. From figures 7 (b) and (d), it can be seen that cohesion and separation of clusters in the feature map have also been improved. Figure 8 compares segmentation results obtained by pixel-wise softmax of known classifier C for open and closed set DA. It can be seen that the number of pixels classified as unknown labels increases in images 3 and 5 by closed set DA.

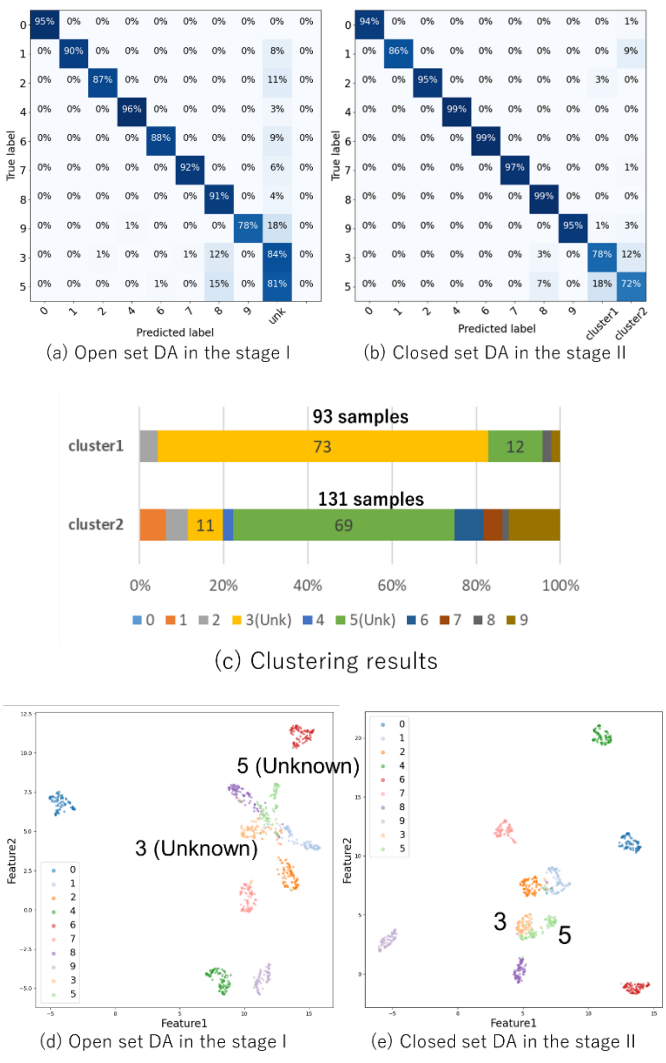


Figure 7. *Experimental results of image classification with unknown labels 3 and 5: confusion matrix of (a) open set DA in the stage I and (b) closed set DA in the stage II, (c) clustering results by X-means, and UMAP feature map of known classifier outputs for (d) open and (e) closed set DA.*

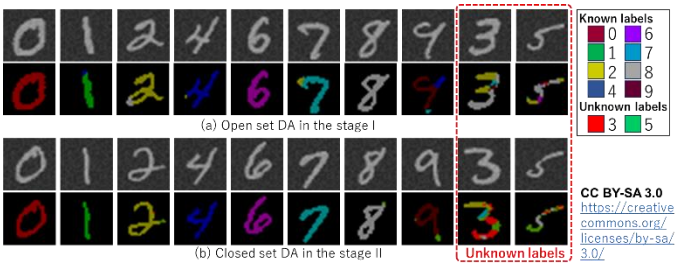


Figure 8. *Comparison of segmentation results obtained by pixel-wise softmax of known classifier C for (a) open and (b) closed set DA with unknown labels 3 and 5.*

Concluding remarks

In this paper, we propose a domain adaptation method for image classification with multiple unknown labels utilizing unsupervised clustering of pixel-wise local features. Our method involves two stages: (1) We apply separation of unknown labels and extraction of local features through open set domain adaptation. Subsequently, we employ unsupervised clustering of the extracted features to estimate the number of unknown labels and provide pseudo-labels to the target samples. (2) We apply closed set domain adaptation for final image classification. The experimental results on the modified MNIST dataset demonstrate that our proposed method improves classification accuracy while effectively separating unknown labels.

References

[1] B. Dey, et al. “Deep learning based defect classification and detection in SEM images: a mask R-CNN approach," in Proc. SPIE PC12053, Metrology, Inspection, and Process Control XXXVI, PC120530K, 2022.

[2] H. Zhou, et al. "Intelligent bearing fault diagnosis method based on a domain aligned clustering network," Measurement Science and Technology, vol. 34, 044001, 2023.

[3] P. Singhal, et al. "Domain Adaptation: Challenges, Methods, Datasets, and Applications," IEEE Access, vol. 11, pp. 6973-7020, 2023.

[4] J. H. Jang, et al. "Unknown-aware domain adversarial learning for open-set domain adaptation", Advances in Neural Information Processing Systems, vol. 35, pp. 16755-16767, 2022.

[5] T. Jing, et al. "Towards novel target discovery through open-set domain adaptation," in Proc. the IEEE/CVF International Conference on Computer Vision, pp. 9322-9331, 2021.

[6] Y. LeCun, et al. "Gradient-based learning applied to document recognition, " in Proc. the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[7] D. Saito and S. Kate, "OVANet: One-vs-all network for universal domain adaptation, " in Proc. the IEEE/CVF international conference on computer vision, pp. 9000-9009, 2021.

[8] D. Pelleg and W. M. Andrew, "X-means: Extending k-means with efficient estimation of the number of clusters," in Proc. the 17th International Conference on Machine Learning, vol. 1. 2000.

[9] L. McInnes, et al. “UMAP: Uniform manifold approximation and projection for dimension reduction,” arXiv preprint arXiv:1802.03426, 2018.

Author Biography

Daichi Nishihara received his BE in electronic engineering from Osaka University, Suita, Japan. He is a graduate student at the Graduate School of Information Science and Technology, Osaka University. His work has focused on the image processing for electron microscopes.

Yoshihiro Midoh received his BE and ME in electronic engineering from Osaka University, Suita, Japan and his PhD in information science and technology from Osaka University. He is a specially appointed associate professor at the Graduate School of Information Science and Technology, Osaka University. His work has focused on the signal and image processing for advanced measuring equipment such as electron microscopy.

Youyang Ng received his BE in electrical engineering from Universiti Malaya, Malaysia. He is a researcher at the Institute of Memory Technology Research and Development, Kioxia Corporation. His work has focused on the computer vision and natural language processing for intelligent systems.

Osamu Yamane received his BS and MS in electrical engineering from Keio University, Yokohama, Japan. He is a researcher at the Institute of Memory Technology Research and Development, Kioxia Corporation. His work has focused on the computer vision and digital transformation technology.

Maasa Takahashi received his BE and ME in Intermedia Art and Science from Waseda University, Tokyo, Japan. He is a researcher at the Institute of Memory Technology Research and Development, Kioxia Corporation. His work has focused on the computer vision and image processing.

Shuhei Iijima received his BE and ME in Engineering System from Tsukuba University, Japan. He is a researcher at the Institute of Memory Technology Research and Development, Kioxia Corporation. His work has focused on the computer vision and digital transformation technology.

Jun Shiomi received his BE in electrical and electronics engineering from Kyoto University, Kyoto, Japan and his ME and PhD in informatics from Kyoto University. He is an associate professor at the Graduate School of Information Science and Technology, Osaka University, Suita, Japan. His work has focused on design and optimization of low-power integrated circuits.

Goh Itoh received his BE and ME in Science and Technology from Keio University, Kanagawa, Japan. He is an assistant to general manager at Digital Transformation Technology R&D Center, Institute of Memory Technology Research & Development, Kioxia Corporation. His work has focused on the image processing and digital transformation technology.

Noriyuki Miura received the BS, MS, and PhD in electrical engineering all from Keio University, Yokohama, Japan. He is a professor at the Graduate School of Information Science and Technology, Osaka University, Suita, Japan. His work has focused on hardware security/safety and next-generation heterogeneous computing systems.