Cell segmentation without annotation by unsupervised domain adaptation based on cooperative self-learning

Shintaro Miyaki[1], Shori Nishimoto[1], Yuta Tokuoka[2],
Takahiro G Yamada[2], Takashi Morikura[1] & Akira Funahashi[1,2,*]

[1]*Center for Biosciences and Informatics, Graduate School of Fundamental Science and Technology, Keio University, Kanagawa, Japan.*

[2]*Department of Biosciences and Informatics, Keio University, Kanagawa, Japan.*

**funa@bio.keio.ac.jp*

Abstract

Cell segmentation, which extracts cells from microscopic images, is essential for quantitative evaluation of cell morphology. Recently, supervised deep-learning-based models have been shown to achieve highly accurate segmentation. However, the performance of these supervised models is often degraded when the models infer unknown cell types that are not included in the train dataset. One approach to overcoming the performance degradation is acquiring new annotated data for each cell type. However, constructing datasets for all cell types is not feasible because labeling every single pixel, rather than each image, is required in the segmentation task. Learning methods that can achieve highly accurate segmentation without annotation is strongly required.

Here, we developed a cell segmentation method based on unsupervised domain adaptation with cooperative self-learning (CULPICO: Cooperative Unsupervised Learning for PIxel-wise COloring). The proposed method consists of two independent segmentation models and a mutual exchange mechanism of inference data. For the data with labels, the models are trained through supervised learning. For the data without labels, the models infer a label probability at each pixel and generate a pseudo-label as unsupervised learning. The pseudo-labels created by each model are mutually used as ground-truth in the other model. Loss function is corrected by considering pixel-level discrepancies between the label probabilities inferred by the two models. The proposed method, despite being an unsupervised learning method, can segment the unknown cell types without labels

with an accuracy comparable to supervised learning models. Our method, which could solve the performance degradation problem without constructing new datasets, is expected to accelerate life science by reducing the cost of extracting quantitative biological knowledge.

Introduction

Recent advances in microscopy have made it possible to obtain large numbers of cellular microscopic images. For the acceleration of life science, it is essential to construct mathematical techniques to extract quantitative biological knowledge from such a huge amount of data[1]. In particular, cell segmentation, which extracts cellular regions from microscopic images, is an important technique for quantifying morphological features of cells and gaining new insights into biological phenomena[2]. Traditional cell segmentation has been supported by image processing techniques such as image filtering and morphological operations[3]. However, those approaches require pre-determination of processing parameters for each image, resulting in a lack of objectivity and low throughput. Recently, a deep-learning-based algorithms that can automatically determine the parameters have been proposed[4, 5, 6, 7]. Deep-learning models that automate image processing operations while eliminating arbitrariness can solve the objectivity and throughput issues. Such models include convolutional neural networks (CNNs), which use convolutional processing, that are trained with microscopic images as input and a label matrix annotated with whether each pixel is a cellular region or not as output. These models generally consist of an encoder model that maps features from the input microscopic image to the latent space and a decoder model that outputs a label matrix from the latent representation. Previous studies have reported that supervised CNNs trained on certain cell types show better segmentation accuracy than the traditional image processing methods[4]. However, the segmentation performance of such supervised trained models is often degraded for unknown data not included in the training data[8]. A naive approach to this problem is to annotate the unknown cell types with additional correct labels. However, the cost of constructing a new dataset for each unknown cell type is extremely high because the segmentation task requires pixel-level labeling in one image while general classification tasks require only image-level labeling. Learning methods that can achieve highly accurate segmentation of unknown cell types without additional annotation are required.

In deep learning techniques, it is generally assumed that feature distributions (domains) of training and inference data follow the same distribution. By contrast, if each domain was significantly different from the others, the inference performance would be degraded[8]. This phenomenon is usually called domain shift. Toward solving this problem, some algorithms based on unsupervised domain adaptation have been developed[9, 10, 11, 12]. These algorithms learn to align the feature distributions of the unknown target domain with those of the known source domain with the correct labels. Saito et al. developed an architecture with two decoder models branching from a single encoder model[13]. They tried to align the feature distributions by adversarial training based on the discrepancy between the feature distributions inferred by each decoder model. Zou et al. developed a single encoder-decoder model to generate pseudo-labels for unknown data without labels. They tried to implicitly align the feature distributions by reusing the pseudo-labels as correct labels[14]. Han et al. proposed a cooperative

learning method by preparing two independent networks and propagating inference errors in one model back to the other model[15]. Previous methods have tried to overcome domain shift by correcting the feature distribution at the image level. However, it is still difficult to correct the feature distribution at the pixel level, and therefore achieving segmentation as accurate as that achieved with supervised learning is a challenging task.

Here, we developed an unsupervised domain adaptation method that can correct the feature distributions at the pixel level. Our architecture consists of two independent encoder-decoder models and constructed a cooperative self-learning framework that consider the pixel-level discrepancy between the pseudo-label inferred by each model. To assess the performance of our method, the segmentation accuracy was evaluated by a public dataset of cell images acquired with phase-contrast microscopy. The proposed method robustly improved the segmentation performance even for unknown cell types. Notably, the proposed method, despite being an unsupervised learning method, outperformed the performance of supervised learning model for some cell types. A model analysis indicated that the simultaneous updating of pseudo-labels and models considering the pixel-level discrepancies may have contributed to improving the performance by suppressing false negatives. The proposed method, which achieves segmentation accuracy as high as supervised learning even for unknown cell types without correct labels, is expected to accelerate the development of life science by reducing the cost of extracting quantitative biological knowledge.

Methods

Dataset

To assess the performance of our method, we used a public dataset of cell images published by Edlund et al.[16]. The dataset consists of phase-contrast microscopic images of eight cell types showing a variety of cell morphologies, and label images annotated with cellular and background regions by experts. In this study, the models were trained with the phase-contrast microscopic images as input and the labeled images as output. The representative microscopic images of eight cell types are shown in Supplementary Fig.S1. The cell types are human glioblastoma cell line A172, human breast cancer cell line BT-474, mouse macroglial cell line BV-2, human hepatocellular carcinoma cell line Huh7, human breast cancer cell line MCF7, human neuroblastoma cell line SH-SY5Y, human breast cancer cell line SK-BR-3 and human ovarian cancer cell line SK-OV-3 . Detailed information on image acquisition can be found in the paper by Edlund et al.[16].

Because the purpose of this study is to develop a cell morphology segmentation algorithm, data with missing cell region labels or data with extremely high cell density were excluded from the dataset in this study (Supplementary Fig.S2). In addition, to minimize the effects of well plate derived characteristic image patterns on learning, we constructed train and test datasets using microscopic images of cells cultured in different well plates, respectively (Supplementary Table S1). The train data were split 3:1 into training data and validation data.

To analyze morphological features of each cell type, microscopic images were cropped to 224 $\times$ 224 pixels from the image center. After the cropping, feature representations with 4096 dimensions were extracted using the VGG16 model[17], which was trained for the image classification task on ImageNet[18], a benchmark dataset for natural images. To qualitatively evaluate the morphological feature distribution, representations were reduced to two dimensions using t-SNE[19] and then plotted in two-dimensional space.

Overview of the propose method

The proposed method (CULPICO: Cooperative Unsupervised Learning for PIxel-wise COloring) consisted of two independent encoder-decoder models with a mutual exchange mechanism (Fig. 1). To achieves cooperative self-learning, we designed a loss function using the pixel-level discrepancy. The encoder-decoder model was constructed by a U-Net architecture consisting of different initial value parameters (Supplementary Fig. S3). In the U-Net architecture, batch normalization was performed after each convolution or inverse convolution process.

In this study, we defined the datasets consisting of microscopic images with labels as the source dataset, while the datasets consisting of microscopic images without labels as the target dataset. The proposed method achieves unsupervised domain adaptation by simultaneously performing supervised learning on the source dataset and

cooperative self-learning on the target dataset.

For the supervised learning on the source dataset, we trained the models using the the microscopic image as input and the labels as output. The loss function L_s of the supervised learning is defined as follows based on the cross-entropy error between the inferred label probability and the correct label.

$$\begin{aligned} L_s = & \frac{1}{N}\sum_{i=1}^{N} -\{(y_s)_i \cdot \log({p_s}^m)_i + (1-(y_s)_i)\cdot \log(1-({p_s}^m)_i)\} \\ & + \frac{1}{N}\sum_{i=1}^{N} -\{(y_s)_i \cdot \log({p_s}^n)_i + (1-(y_s)_i)\cdot \log(1-({p_s}^n)_i)\} \end{aligned} \tag{1}$$

where i is the index number of each pixel and N is the total number of pixels in the image. The $(p_s)_i$ is i-th pixel in the inferred label probability for the image and $(y_s)_i$ is i-th pixel in the correct label. The m and n denote different encoder-decoder models.

For the cooperative self-learning on the target dataset, the models infered label probabilities representing the probability of cellular regions in the target image, and pseudo-labels were generated by binarizing them with a predefined threshold value. The pseudo-label generated in one model was then used as the pseudo-correct label in the other model. The loss function L_t of the cooperative self-learning was designed based on the weighted cross-entropy error according to pixel-level discrepancy between the inference label probabilities in the two models (Eq. (2)). The loss function L_t was expected to actively learn the error in pixels with low discrepancy, while carefully learning the error in pixels with high discrepancy.

$$\begin{aligned} L_t = & \frac{1}{N}\sum_{i=1}^{N} -(1-(d^{m,n})_i)\cdot\{(\hat{y_t}^n)_i \cdot \log({p_t}^m)_i + (1-(\hat{y_t}^n)_i)\cdot \log(1-({p_t}^m)_i)\} \\ & + \frac{1}{N}\sum_{i=1}^{N} -(1-(d^{m,n})_i)\cdot\{(\hat{y_t}^m)_i \cdot \log({p_t}^n)_i + (1-(\hat{y_t}^m)_i)\cdot \log(1-({p_t}^n)_i)\} \end{aligned} \tag{2}$$

where $(p_t)_i$ is i-th pixel in the label probability inferred for the target image, the $(\hat{y_t})_i$ is the i-th pixel in the pseudo-correct label (Eq. (3)), and the $(d^{m,n})_i$ is the pixel-level discrepancy between the label probabilities inferred from the two models m and n (Eq. (4)).

$$(\hat{y_t})_i = \begin{cases} 1, \text{ if } (p_t)_i > 0.5 \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

$$(d^{m,n})_i = |({p_t}^m)_i - ({p_t}^n)_i| \tag{4}$$

where label 1 is the cellular region and label 0 is the background region.

Using the above loss functions L_s and L_t, the total loss function L_{total} was defined as follows (Eq. (5))

$$L_{total} = L_s + L_t \tag{5}$$

By performing the cooperative self-learning in the target dataset simultaneously with supervised learning in the source dataset, the feature distribution of the target dataset is expected to adapt to that of the source. Although previous methods were developed to improve the performance of segmentation using errors at the image level[15], the proposed method was designed using the errors at the pixel level.

Evaluation metrics of segmentation

To evaluate the effectiveness of the proposed method, we developed a U-Net model as a lower-bound model that was trained using only the source dataset. In addition, we developed a U-Net model as an upper-bound model that was trained using only the target dataset. The U-Net models were constructed with the architecture shown in Supplementary Fig.S4. The lower-bound model is expected to degrade inference performance at the target because the target cell types are not included in the train data. The upper-bound model is equivalent to a general supervised learning model, and the inference performance at the target is expected to be relatively high because the target cell types are included in the train data. To compare our model to previous unsupervised domain adaptation models, we also trained Saito et al.'s model [13], which is an adversarial learning method, and Zou et al.'s model [14], which is a self-learning method. The Saito et al. model was constructed with the architecture shown in Supplementary Fig.S5. In the architecture of the previous models, the batch normalization was performed after each convolution or inverse convolution process.

The training was performed using a GPU (NVIDIA V100 or A100) with a mini-batch size of 16 for 400 epochs. The microscopic images were normalized to [0,1] by performing min-max normalization. In addition, the microscopic images in which cells were located at the periphery of the image so that their entire cellular shape could not be captured were interpolated by mirror padding. The data were augmented by inverting left and right, flipping up and down, and random cropping. The image size after cropping was fixed at 352 $\times$ 272 pixels. The model at the epoch with the smallest loss to the data for validation was selected as the best model. We evaluated the performance of the trained model with the test data that were not used in training.

As an evaluation metric, we calculated the intersection-over-union metric (IoU), which is widely used to evaluate the performance of segmentation (Eq. (6)).

$$IoU = \frac{TP}{TP + FP + FN} \tag{6}$$

where TP, FP, and FN are the number of pixels counted as True Positive, False Positive, and False Negative, respectively. The range of IoU is from 0 to 1, and closer to 1 indicates the better segmentation.

Results

Proposed method robustly improved the segmentation performance for unknown cell types

To establish an unsupervised domain adaptation method that can achieve highly accurate segmentation of unknown cell types without labels, we developed a cooperative self-learning method using the pixel-level discrepancy between different model inferences (Fig. 1). To evaluate the proposed method, we used the public dataset of cell images published by Edlund et al.[16]. The dataset consists of phase-contrast microscopic images of eight different cell types. In this study, we defined the source as the cell type for which the correct label is annotated, and the target as an unknown cell type for which the correct label is not annotated. We evaluated the performance of the proposed model using 56 different combinations of cell types used as source and target. To evaluate the effectiveness of the proposed method, we used as the lower-bound model a supervised U-Net model in which only the source cell type was used for training, and as the upper-bound model one in which the target cell type was used for training. In this study, the performance of the lower-bound model is the minimum one that must be achieved, and that of the upper-bound model is the target one.

To analyze the morphological characteristics of each cell type, the result of dimensionality reduction of the microscopic images is shown in Fig. 2. In the reduced 2D space, it was shown that clusters were formed for each cell type. The upper-bound model, which infers the same cell types as in training, has the potential to achieve highly accurate segmentation, while the lower-bound model, which infers cell types different from those in training, may degrade the model performance due to the domain shift.

To evaluate the effect of cell-type differences during training and inference on the performance of the models, we trained 56 supervised U-Net models with different combinations of cell types. Representative learning curves of the U-Net models are shown in Supplementary Fig.S6. At the end of the learning process, the convergence of the error was observed, indicating that the U-Net model was well-trained. The performance of each U-Net models is shown in Table 1. The results showed that the upper-bound model has high accuracy for most of the combinations except for SH-SY5Y cells, but the lower-bound model showed lower accuracy than the upper bound model. The combinations with Huh7 cells as the target showed the most significant performance degradation among all cell types. Notably, the IoU of the lower-bound model with Huh7 cells as the target and BV-2 cells as the source was reduced by a maximum of 0.806 (0.860 $\rightarrow$ 0.0540) relative to the upper bound model. The lower-bound model also showed that there were variations in IoU for each source cell type even if the target cell type was the same.

To evaluate the effectiveness of the proposed method, the proposed models were trained with 48 different combinations of source and target cell types. Representative learning curves of the proposed method are shown in Supplementary Fig.S7. At the end of the learning process, the error differences between the two models converged

as the training progressed, suggesting that the two models were able to learn cooperatively. The performance of the proposed models is shown in Table 2, and the difference in performance between the proposed models and the lower bound models is shown in Fig.3. The proposed models had higher accuracy than the lower-bound model in 42 combinations. It should be noted that the proposed models improved the IoU by 0.64 compared with the lower-bound model in the combination of Huh7 cells as target, which showed the largest performance degradation between the upper-bound and lower bound models, and SH-SY5Y cells as source, which showed the lowest accuracy in the upper-bound model (Table 1). In addition, the proposed method showed higher accuracy than the lower-bound model when Huh7, MCF7, or SK-OV3 cells were used as the target cells, regardless of which cell type was used as the source. Furthermore, the proposed method showed higher accuracy than the lower-bound model when SH-SY5Y or BV-2 cells were used as the source cells, regardless of which cell type was used as the target. In contrast, the proposed method tended to show lower accuracy than the lower-bound model, especially when SH-SY5Y cells were used as the target. These results suggested that SH-SY5Y and BV-2 cells contain necessary shape information to construct cell shape features, while Huh7, MCF7, and SK-OV3 cells contain sufficient shape information to construct cell shape features.

Taken together, these results indicate that the proposed method robustly improves the segmentation performance for unknown cell types compared with the lower-bound model.

Proposed method, despite being an unsupervised learning model, outperforms segmentation of some unknown cell types compared with supervised learning

To evaluate the effectiveness of the proposed method, we compared and assessed the model performance of lower-bound models, upper-bound models, the self-learning model by Zou et al. [14] and the adversarial learning model by Saito et al. [13]. The results in the previous section showed that SH-SY5Y cells may contain necessary shape information to construct cell shape features but were the most difficult cell type to segment in the upper bound model. Huh7 cells had the potential to contain sufficient shape information to construct cell shape features but were the cell type with the most significant performance degradation between the upper- and lower-bound models. MCF7 cells may have contained sufficient shape information to construct cell shape features but were the cell type whose cell shape differs significantly from the other cell types (Supplementary Fig.S1). Based on these findings, we compared the segmentation performance of models using the combinations of these three characteristic cell types.

The proposed method showed higher accuracy than the previous models except for the combinations with SH-SY5Y cells as the target (Table 3). Most importantly, even though the proposed method is an unsupervised learning model, there was a combination of cell types that showed accuracy comparable to the upper-bound model (combinations with Huh7 and MCF7 cells as the target) (Table 3). It was also noteworthy that the proposed method showed higher accuracy than the upper-bound model when the source and target cell type combinations were Huh7 $\rightarrow$ MCF7 and SH-SY5Y $\rightarrow$ MCF7, while the previous models showed lower accuracy than the upper-bound model. In contrast, when SH-SY5Y cells were used as the target, the proposed method showed lower accuracy than the upper-bound model, as in the previous models. Although the proposed method learned sufficient features for cell shape in Huh7 and MCF7 cells, like the previous models it had difficulty learning the necessary features for cell shape in SH-SY5Y cells.

Simultaneous updating of the pseudo-label and model to consider the pixel-level discrepancy might have improved the accuracy of detecting cellular regions

To verify the factors that contributed to the performance improvement of the proposed method, we qualitatively evaluated the segmentation results and verified the transition of the pixel-level discrepancy and pseudo-label accuracy during the training.

Representative segmentation results are shown in Fig. 4. In the combination of SH-SY5Y → MCF7 (Lower bound: 0.390 ± 0.267, Ours: 0.863 ± 0.070) and SH-SY5Y → Huh7 (Lower bound: 0.171 ± 0.141, Ours: 0.808 ± 0.061), the true-positive regions increased and the false-negative regions decreased, indicating the performance improvement. In contrast, the combination of Huh7 → SH-SY5Y (Lower bound: 0.679 ± 0.076, Ours: 0.672 ± 0.074) did not suppress false-negative errors in the regions where protrusion-like structures were present. Taken together, these results suggested that the Huh7 and MCF7 cells contained sufficient features to construct cell shape, whereas the SH-SY5Y cells contained the necessary features to construct cell shape.

A heat map showing the pixel-level discrepancy before and after learning is shown in Fig.5a. In the combination of SH-SY5Y → Huh7 (IoU:0.808 ± 0.061), the high-discrepancy regions tended to be decreased, whereas in the combination of Huh7 → SH-SY5Y (IoU:0.672 ± 0.074), the high-discrepancy regions remained even at the end of learning.

To evaluate the effects of pseudo-label updates on the process of self-learning, we compared the IoU transition of pseudo-labels in the learning process of the proposed method and Zou et al.'s method[14] (Fig. 5b). In the SH-SY5Y → Huh7 combination (IoU:0.808 ± 0.061), the IoU of the pseudo-labels in the proposed method improved significantly in the first 10 epochs of learning, whereas the accuracy of the Zou et al. method stopped the accuracy improvement at 300 epochs after a gradual improvement. However, for the combination of Huh7 → SH-SY5Y (IoU: 0.672 ± 0.074), the IoU of the pseudo-labels in the proposed method seemed to decrease gradually during the learning process. This indicates that the proposed method achieved the performance improvement by updating the pseudo-label and model simultaneously, while Zou et al.'s method updates pseudo-labels step by step[14].

These results suggested that the simultaneous updating of the pseudo-label and model with consideration of the pixel-level discrepancy improved the segmentation performance by suppressing false negatives in some cell type combinations, such as Huh7 and MCF7 cells. Although the cell types used in the validation of this study were only a subset of the many cell types, it was suggested that we can develop a model that can effectively suppress false-negative errors in unknown cell types by training with the cell images that have protrusion-like structures as a source (Supplementary Fig. S1), such as SH-SY5Y cells, which may have contained the necessary features to constitute cell shape.

Discussions

To overcome the performance degradation of segmentation for an unknown cell type, we developed an unsupervised domain adaptation method based on a cooperative self-learning method that uses two independent models for aligning feature distributions at the pixel-level discrepancy (Fig. 1). As a result, for the unknown cell types that belong to a different domain from the learned cell type (Fig. 2), the proposed method robustly improved the segmentation performance compared with the lower-bound model (Fig. 3). In addition, the proposed method using the pixel-level discrepancy not only showed higher segmentation performance than previous approaches using image-level discrepancy, but also showed accuracy comparable to supervised learning models for some cell type combinations such as Huh7 and MCF7 cells, despite the unsupervised learning (Table 3). Furthermore, the proposed method might have improved the accuracy of detecting cellular regions by suppressing false-negative regions through simultaneous updating of pseudo-labels and models considering the pixel-level discrepancy (Fig. 4 and Fig. 5).

Based on the performance evaluation of the supervised U-Net models, the upper-bound model achieved high segmentation accuracy, whereas the lower-bound model showed low segmentation accuracy (Table 1). In comparison, in the results of dimensionality reduction using t-SNE, the segmentation accuracy within the domain of the learned cell type was high, while the segmentation accuracy for cell types in a different domain from the one at the time of learning was low (Fig. 2). These results suggested that, as in the previous studies[8], this phenomenon occurred by a domain shift.

For the Huh7 → MCF7 and SH-SY5Y → MCF7 combinations, the proposed method outperformed the segmentation performance of the upper-bound model (Table 3). The qualitative evaluation of the segmentation results suggested that the performance degradation due to domain shift was caused by an increase in false-negatives, and the proposed method, which cooperatively learns pseudo-labels using the pixel-level discrepancy, improved the segmentation performance by suppressing the false negatives (Fig. 4). In these combination of these cell types, the pixel-level discrepancy decreased with learning (Fig. 5a), and the accuracy of the pseudo-label generated based on the discrepancy increased with learning (Fig.5b). These findings suggest that generating highly accurate pseudo-labels by using the pixel-level discrepancy resulted in high segmentation performance even for unknown cell types.

In contrast, for the combination of Huh7 → SH-SY5Y and MCF7 → SH-SY5Y, the proposed method showed lower segmentation accuracy than the lower-bound model. Huh7 and MCF7 cells show rounded cell morphology, whereas SH-SY5Y cells have characteristic protrusion-like structures (Supplementary Fig. S1). The qualitative evaluation of the segmentation results indicated that the proposed models trained on the Huh7 cells failed to detect the protrusion-like structures of SH-SY5Y cells, and the false-negative errors seemed to increase (Fig. 4).

In addition, the pixel-level discrepancy could not be reduced in these combinations (Fig. 5a), and the accuracy of the pseudo-labels seemed to decrease gradually during the learning process (Fig. 5b). In the case of combinations of cell types with extremely different cell morphologies, the models could not correct the errors by examining the discrepancy, which may have caused the performance degradation as independent models proceeded to learn incorrect labels from each other. Considering these results, we propose that cells with protrusion-like structures, such as SH-SY5Y cells, contain the necessary features to constitute cell shape, while rounded cells, such as Huh7 and MCF7 cells, contained sufficient features to constitute cell shape. Although the models were only evaluated on a subset of the many cell types, a model trained on cells with characteristic structures as a source using the proposed method can achieve effective unsupervised segmentation, even on unknown cell types. By comparing the dimensionality-reduced feature map using t-SNE and the cell types that show characteristic cell shapes, it might be possible to construct an efficient segmentation technique by selecting cell types that should be prioritized for constructing datasets.

In summary, the proposed method using the pixel-level discrepancy, despite being an unsupervised learning method, was able to segment the unknown cell types with accuracy comparable even to the supervised learning models for some combinations of cell types. The proposed method, which enables automatic extraction of cellular regions for unknown cell types without additional annotation, is expected to accelerate the acquisition of new biological knowledge.

References

[1] Peng, H. Bioimage informatics: a new area of engineering biology. *Bioinformatics* **24**, 1827–1836 (2008).

[2] Meijering, E. Cell segmentation: 50 years down the road [life sciences]. *IEEE Signal Processing Magazine* **29**, 140–145 (2012).

[3] Liao, Q. & Deng, Y. An accurate segmentation method for white blood cell images. In *Proceedings IEEE International Symposium on Biomedical Imaging*, 245–248 (IEEE, 2002).

[4] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105 (Curran Associates, Inc., 2012).

[5] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer, 2015).

[6] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**, 203–211 (2021).

[7] Tang, Y. *et al.* Self-supervised pre-training of swin transformers for 3D medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20730–20740 (2022).

[8] Zhou, K., Liu, Z., Qiao, Y., Xiang, T. & Loy, C. C. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[9] Toldo, M., Maracani, A., Michieli, U. & Zanuttigh, P. Unsupervised domain adaptation in semantic segmentation: a review. *Technologies* **8**, 35 (2020).

[10] Oza, P., Sindagi, V. A., Sharmini, V. V. & Patel, V. M. Unsupervised domain adaptation of object detectors: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[11] Zhang, Y., Miao, S., Mansi, T. & Liao, R. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 599–607 (Springer, 2018).

[12] Chen, C., Dou, Q., Chen, H. & Heng, P.-A. Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation. In *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*, 143–151 (Springer, 2018).

[13] Saito, K., Watanabe, K., Ushiku, Y. & Harada, T. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3723–3732 (IEEE, 2018).

[14] Zou, Y., Yu, Z., Kumar, B. & Wang, J. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, 289–305 (Springer International Publishing, 2018).

[15] Han, B. *et al.* Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in Neural Information Processing Systems*, vol. 31 (Curran Associates Inc., 2018).

[16] Edlund, C. *et al.* Livecell—a large-scale dataset for label-free live cell segmentation. *Nature methods* **18**, 1038–1045 (2021).

[17] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[18] Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (IEEE, 2009).

[19] van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

Code availability

The source code of this study is available from `https://github.com/funalab/CULPICO`.

Data availability

The datasets including all training and testing datasets during the current study are available from `https://drive.google.com/file/d/1knrykZ8aOwNKbOUzQ8a1PvGQodOkY1-V/view?usp=sharing`. The datasets are part of a public cell image dataset LIVECell published by Edlund et. al. [16] . Original LIVECell dataset are published under Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

Acknowledgement

We are grateful to Yuya Kobayashi for valuable discussions. The research was funded by MEXT/JSPS KAKENHI Grant Number JP18H04742 "Resonance Bio" and JST CREST, Japan Grant Number JPMJCR2331 to A.F.to A.F.

Author contributions

S.M., S.N and A.F. designed the conceptual idea and the study. S.M. implemented the proposed algorithm. S.M., S.N, Y.T., T.G.Y., T.M and A.F. wrote the manuscript, with suggestions from the other authors.

Competing interests

The authors declare no competing interests.

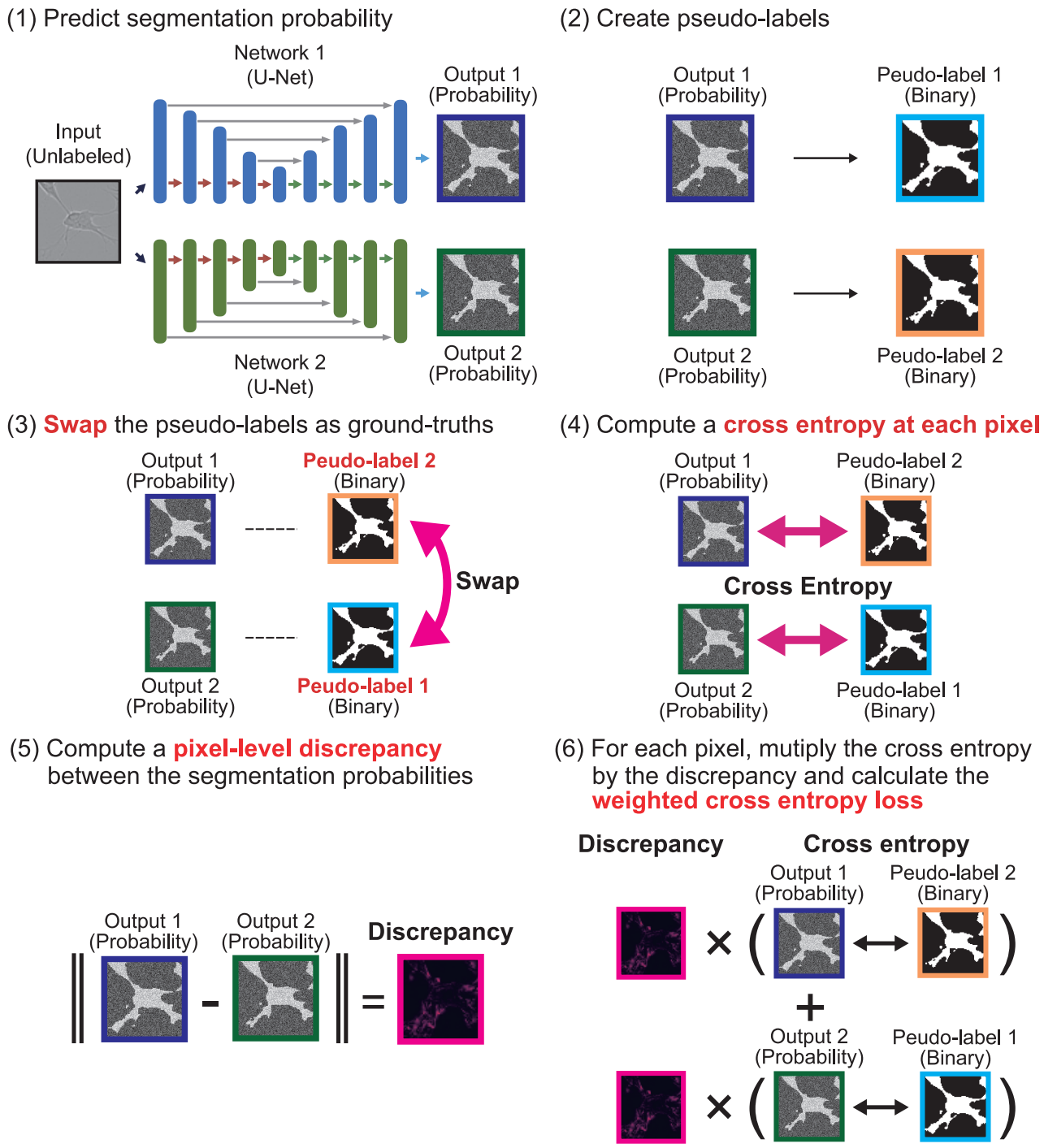

Figure 1: Schematic diagram of the proposed method for performing collaborative self-learning on the target dataset.
(1) Infer label probabilities for the input images on Network 1 and Network 2, respectively. (2) Generate pseudo-labels from the inferred label probabilities by binarizing based on a pre-set threshold value. (3) The pseudo-label generated in one model is used as the pseudo-correct label in the other model. (4) Compute the cross entropy between the inferred label probability and the pseudo-correct label for each pixel. (5) Compute the pixel-level discrepancy using the absolute difference between the inferred label probabilities at each pixel. (6) Compute a weighted cross entropy multiplied by the discrepancy for each pixel. The specific formulas of the loss function are explained in the Methods section.

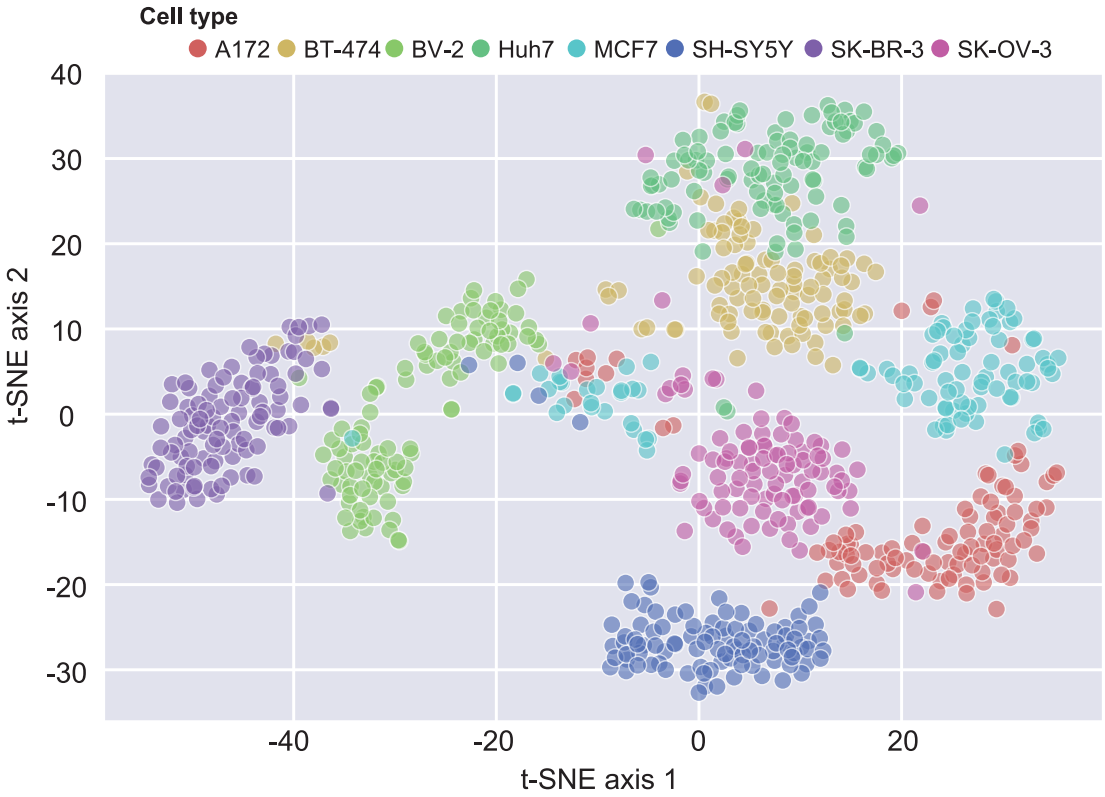

Figure 2: Dimensionality reduction of microscopic cell images into two-dimensional space using t-SNE. The horizontal and vertical axes represent the values in a two-dimensional vector reduced by t-SNE. The plotted points represent the dimension-reduced microscopic images, and the respective colors represent differences in cell types.

Table 1: Segmentation performance for each cell type using the supervised U-Net model
The vertical series represents the cell types used to train the model (source), and the horizontal series represents the cell types used to test the model (target). The values in the table show the mean IoU for the test data, and those in parentheses show the standard deviation. The IoUs of the upper bound models are underlined.

Source \ Target	A172	BT-474	BV-2	Huh7	MCF7	SH-SY5Y	SK-BR-3	SK-OV-3
A172	<u>0.890</u> (0.033)	0.814 (0.091)	0.748 (0.107)	0.727 (0.094)	0.853 (0.100)	0.688 (0.084)	0.846 (0.034)	0.577 (0.105)
BT-474	0.912 (0.032)	<u>0.818</u> (0.084)	0.782 (0.075)	0.786 (0.114)	0.811 (0.103)	0.744 (0.071)	0.890 (0.043)	0.854 (0.052)
BV-2	0.241 (0.162)	0.241 (0.246)	<u>0.840</u> (0.060)	0.0540 (0.0570)	0.434 (0.232)	0.460 (0.106)	0.761 (0.082)	0.858 (0.061)
Huh7	0.913 (0.030)	0.752 (0.157)	0.744 (0.197)	<u>0.860</u> (0.089)	0.723 (0.196)	0.697 (0.076)	0.878 (0.037)	0.869 (0.069)
MCF7	0.904 (0.031)	0.802 (0.086)	0.810 (0.121)	0.650 (0.095)	<u>0.849</u> (0.097)	0.739 (0.072)	0.866 (0.041)	0.738 (0.088)
SH-SY5Y	0.533 (0.149)	0.312 (0.221)	0.834 (0.069)	0.171 (0.141)	0.390 (0.267)	<u>0.787</u> (0.059)	0.798 (0.082)	0.671 (0.093)
SK-BR-3	0.904 (0.033)	0.833 (0.092)	0.804 (0.076)	0.833 (0.099)	0.855 (0.102)	0.671 (0.125)	<u>0.904</u> (0.036)	0.904 (0.037)
SK-OV-3	0.900 (0.052)	0.792 (0.118)	0.796 (0.093)	0.817 (0.108)	0.849 (0.101)	0.756 (0.074)	0.832 (0.046)	<u>0.910</u> (0.042)

Table 2: Segmentation performance for each cell type using the proposed method
The vertical series represents the cell types used to train the model (source), and the horizontal series represents the cell types used to test the model (target). Values in the table show the mean IoUs for the test data, and those in parentheses show the standard deviation.

Source \ Target	A172	BT-474	BV-2	Huh7	MCF7	SH-SY5Y	SK-BR-3	SK-OV-3
A172	-	0.784 (0.074)	0.518 (0.079)	0.792 (0.069)	0.857 (0.068)	0.617 (0.083)	0.837 (0.052)	0.914 (0.030)
BT-474	0.852 (0.038)	-	0.779 (0.068)	0.861 (0.058)	0.861 (0.067)	0.712 (0.078)	0.891 (0.031)	0.875 (0.054)
BV-2	0.860 (0.035)	0.815 (0.079)	-	0.579 (0.055)	0.852 (0.076)	0.692 (0.071)	0.875 (0.031)	0.888 (0.050)
Huh7	0.909 (0.030)	0.820 (0.078)	0.686 (0.142)	-	0.856 (0.072)	0.672 (0.074)	0.879 (0.041)	0.912 (0.031)
MCF7	0.820 (0.046)	0.819 (0.080)	0.828 (0.088)	0.806 (0.060)	-	0.630 (0.080)	0.890 (0.030)	0.826 (0.059)
SH-SY5Y	0.907 (0.028)	0.814 (0.078)	0.827 (0.065)	0.808 (0.061)	0.863 (0.070)	-	0.881 (0.029)	0.813 (0.061)
SK-BR-3	0.912 (0.029)	0.824 (0.079)	0.826 (0.088)	0.850 (0.059)	0.850 (0.071)	0.710 (0.079)	-	0.910 (0.032)
SK-OV-3	0.911 (0.037)	0.806 (0.080)	0.800 (0.092)	0.844 (0.061)	0.854 (0.071)	0.710 (0.075)	0.868 (0.048)	-

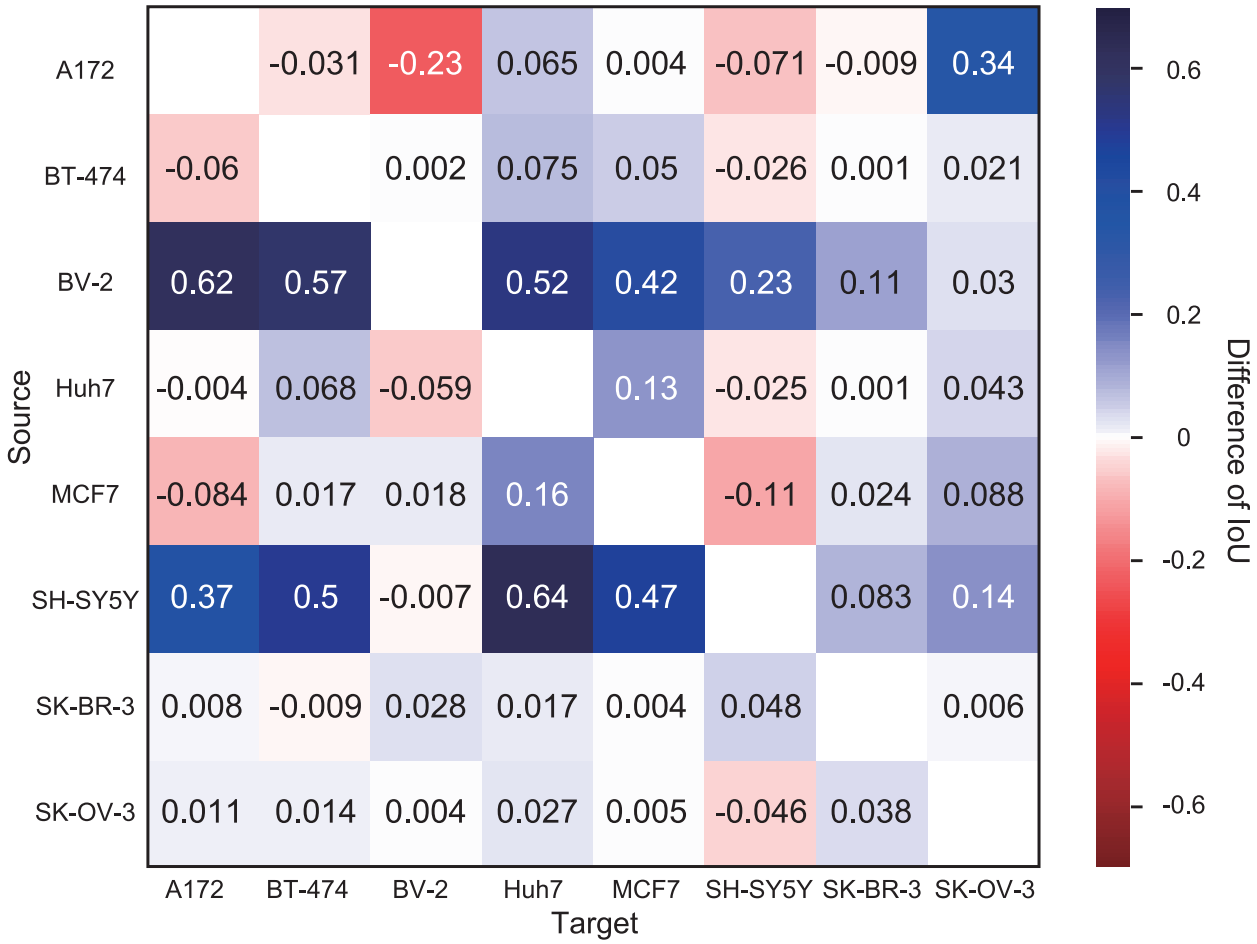

Figure 3: Difference in segmentation performance for each cell type between the proposed method and the lower-bound model

The values represent the difference of IoU between the proposed method and the lower-bound model. The horizontal series represent the cell types used as the target and the vertical series represent the cell types used as the source. Combinations with high segmentation accuracy are shown in blue, and those with low segmentation accuracy are shown in red.

Table 3: Comparison of segmentation performance between the proposed and previous methods
Values in the table represent the mean IoU for the test data, and those in parentheses show the standard deviation. Those in bold type represent the model with the highest IoU, except for the upper-bound model. Underlines represent the models with higher IoU than the upper-bound model. The lower-bound model is the model trained with only source data. The upper-bound model is the model trained with only target data.

Source→Target	Lower-bound	Zou, et al.[14]	Saito, et al.[13]	Our model	Upper-bound
MCF7→Huh7	0.650 (0.095)	0.710 (0.091)	0.662 (0.061)	**0.806** (0.060)	0.860 (0.089)
SH-SY5Y→Huh7	0.171 (0.141)	0.558 (0.069)	0.663 (0.055)	**0.808** (0.061)	0.860 (0.089)
Huh7→MCF7	0.723 (0.196)	0.826 (0.064)	0.828 (0.065)	**<u>0.856</u>** (0.072)	0.849 (0.097)
SH-SY5Y→MCF7	0.390 (0.267)	0.757 (0.070)	0.801 (0.062)	**<u>0.863</u>** (0.070)	0.849 (0.097)
Huh7→SH-SY5Y	0.697 (0.076)	**0.701** (0.075)	0.688 (0.078)	0.672 (0.074)	0.787 (0.059)
MCF7→SH-SY5Y	**0.739** (0.072)	0.711 (0.075)	0.693 (0.079)	0.630 (0.080)	0.787 (0.059)

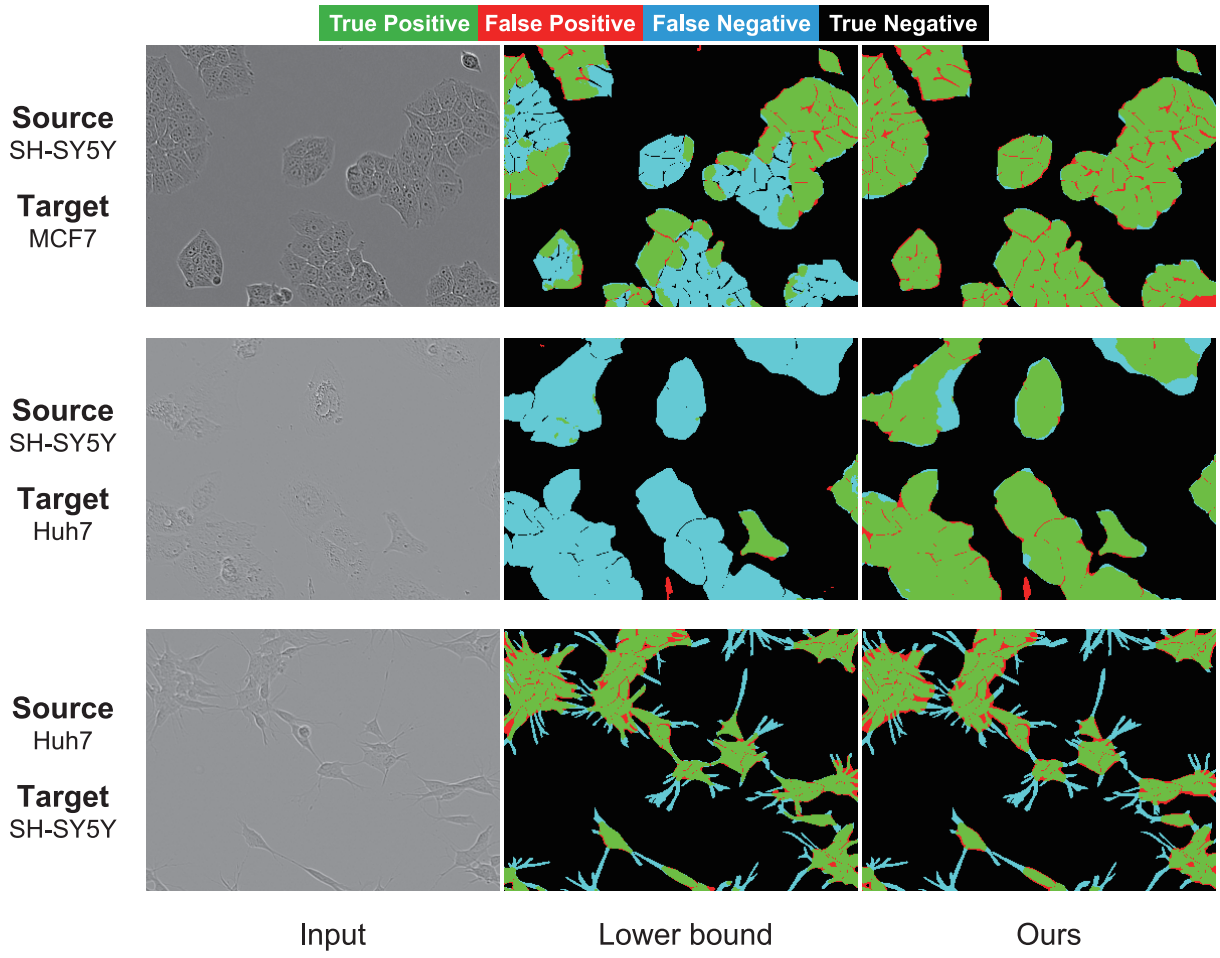

Figure 4: Representative segmentation results
The segmentation image is created by a merge of the model output and the correct data. The color-coded regions in the image correspond to True Positive, False Positive, False Negative, and True Negative regions, respectively, as shown in the upper part of the figure. The upper, middle, and lower rows show the results when the source and target cell type combinations are SH-SY5Y → MCF7, SH-SY5Y → Huh7, and Huh7 → SH-SY5Y, respectively.

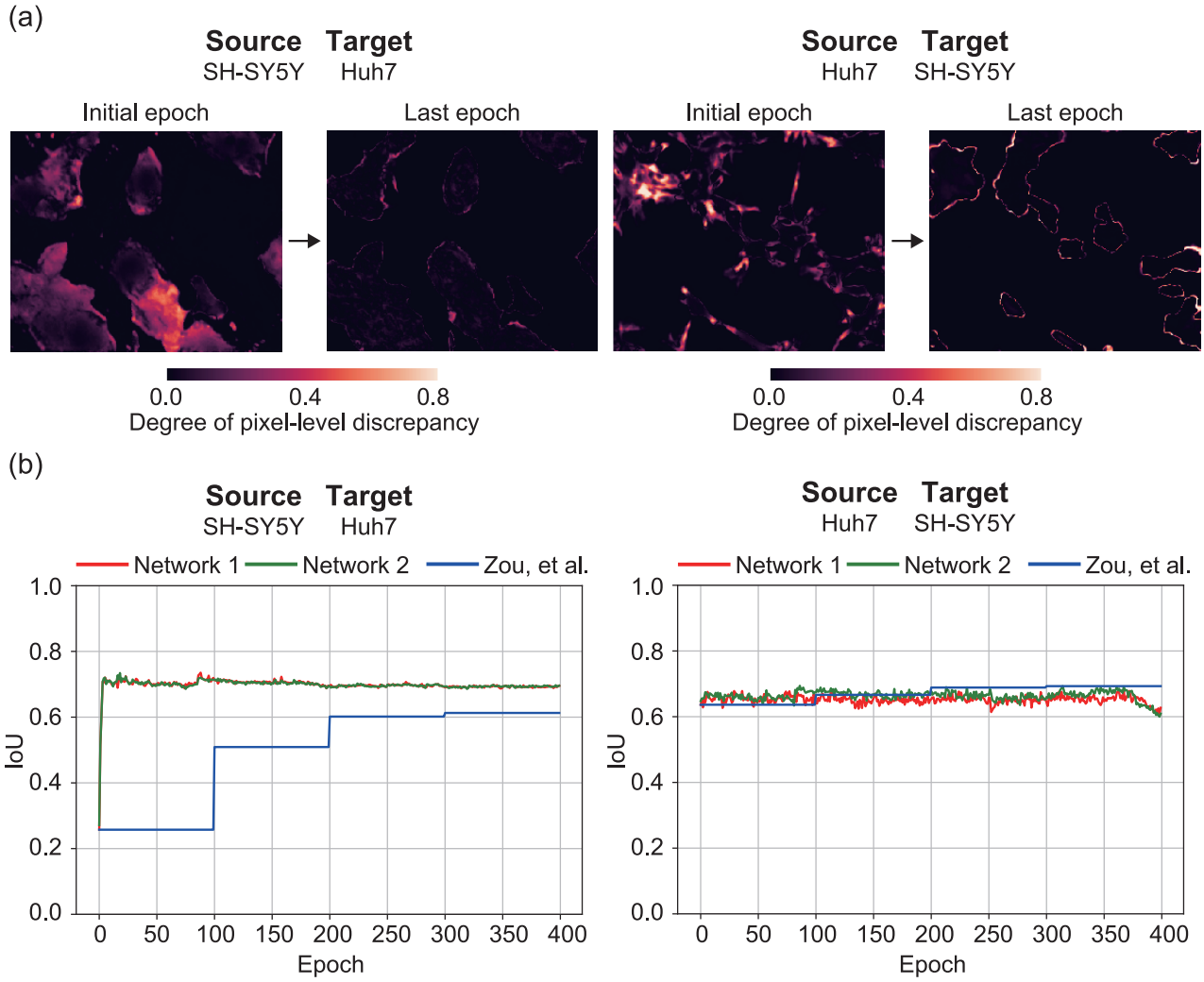

Figure 5: Performance transition in the learning process
(a) Transition of the pixel-level discrepancy. Black represents regions with low discrepancy, and red represents regions with high discrepancy. (b) Transition of the IoU for the pseudo-labels. The red and green lines represent the IoUs of the pseudo-labels created from the two networks in the proposed method. The blue line represents the IoUs of the pseudo-labels created using the Zou et al. method[14]. The IoUs in the proposed method are 0.808 $\pm$ 0.061 and 0.672 $\pm$ 0.074 when the source and target cell type combinations are SH-SY5Y $\rightarrow$ Huh7 and Huh7 $\rightarrow$ SH-SY5Y, respectively.