Sensors and Materials, Vol. 35, No. 2 (2023) 399–409
MYU Tokyo

S & M 3176

Cell Nucleus Detection in Oral Cytology Using Artificial Intelligence

Yoichi Shimomoto,[1*] Kirin Inoue,[1] Ikuo Yamamoto,[2]
Seigo Ohba,[3] Kinuko Ogata,[3] and Hideyuki Yamamoto[3]

[1]Graduate School of Engineering, Nagasaki University,
1-14, Bunkyo-machi, Nagasaki City, Nagasaki 852-8521, Japan
[2]Organization for Marine Science and Technology, Nagasaki University,
1-14, Bunkyo-machi, Nagasaki City, Nagasaki 852-8521, Japan
[3]Graduate School of Medical Science, Nagasaki University,
1-12-4, Sakamoto, Nagasaki City, Nagasaki 852-8523, Japan

(Received December 30, 2022; accepted February 1, 2023)

Keywords: cell nucleus detection, artificial intelligence, sliding window method, mask-RCNN method, oral cancer

We describe the detection of cell nuclei in oral cytology using artificial intelligence (AI). We focused on the detection of cell nuclei because the ratio of cell nuclei to cytoplasm increases with increasing cell malignancy. As an initial step in the development of AI-assisted cytology, we investigated two methods for the automatic detection of cell nuclei in blue-stained cells in cytopreparation images. We evaluated the usefulness of the sliding window method (SWM) and the mask region-based convolutional neural network (Mask-RCNN) method in identifying cell nuclei in oral cytopreparation images. Thirty cases of liquid-based oral cytology were analyzed. First, we performed the SWM by dividing each image into 96 × 96 pixels. Overall, 591 images with or without blue-stained cell nuclei were prepared as the training data and 197 as the test data (from among 1576 images in total). Next, we performed the Mask-RCNN method by preparing 130 images of Class II and III lesions and creating mask images showing cell regions based on these images. By the SWM method, the highest detection rate for blue-stained cells in the evaluation group was found to be 0.9314. For Mask-RCNN, 37 cell nuclei were identified, and 1 cell nucleus was identified as a non-nucleus after 40 epochs (error rate: 0.027). Mask-RCNN was shown to be more accurate than SWM in identifying the cell nuclei. If the blue-stained cell nuclei can be correctly identified automatically, the entire cell morphology can be grasped faster, and the diagnostic performance of cytology can be improved.

1. Introduction

In recent years, remarkable progress has been made in the utilization of artificial intelligence (AI) in medicine. The use of AI for diagnosis using images(1) and pathological diagnosis(2) is

*Corresponding author: e-mail: goma@nagasaki-u.ac.jp
https://doi.org/10.18494/SAM4293

ISSN 0914-4935
https://myukk.org/

increasing. The mortality rate of oral cancer remains high, with 15000 people affected and approximately 7000 deaths reported per year in Japan.[3] Although the oral cavity can be easily observed and palpated, treatment is often delayed because oral cancer is not well recognized owing to its rarity, and early-stage oral cancer is treated as stomatitis, whose appearance is similar to that of oral cancer.

Oral cytology is a diagnostic technique in which lesions are rubbed in order to collect cells, and the atypicality of the stained cells is assessed, allowing the diagnosis of Class I to V lesions (Papanicolaou classification). Class I and II lesions are usually considered inflammatory reactions, while Class III and above lesions require biopsies to obtain a definitive diagnosis of dysplasia or malignancy. Oral cytology is one of the most useful modalities for the early detection of oral cancer. Although cytology is a simple and reproducible technique, the accurate diagnosis of cellular atypia is based on the experience of the pathologist and may vary among experts. These factors are barriers to performing cytology among general practitioners. The development of a system using AI to assist in detecting abnormalities in digital cytological images could solve this problem. During Papanicolaou staining for cytological screening for oral cancer, the blue-stained cells were located on the basal side and the red-stained cells on the epidermal side. The detection and recognition of blue-stained cells are required for the early detection of dysplastic changes. As the atypia of the cells increases, the nucleus-to-cytoplasm (N/C) ratio increases. In addition, a significant variation is observed in the cytoplasmic staining, while only a slight variation is observed in the nuclear staining. Therefore, it is necessary to identify the nuclei of blue-stained cells in cytopreparation images. Because Class III-V lesions should be considered for biopsy, Class I and II lesions should be distinguished from Class III, IV, and V lesions. In this study, we sought to detect the nuclei in blue-stained cells and to classify them.

During image recognition using AI, objects are detected and identified. Two methods are used to propose object region candidates: the sliding window method (SWM)[4–10] and the mask region-based convolutional neural network (Mask-RCNN) method.[11] The SWM extracts candidate regions by shifting the region to a fixed size at a fixed pixel interval. The extracted regions are applied to an image discriminator to determine the presence of important objects in the window. This is a slow but reliable method. Mask-RCNN is a method used for object detection and pixel-by-pixel segmentation of images, and a study detailing this method was selected as the best study at the 16th International Conference on Computer Vision.[10] Mask-RCNN can detect object-like regions in an image and classify them. Object-like regions were detected in large numbers; they were obtained by dividing the image into specific regions and evaluating the regions thoroughly. By narrowing down the image to those regions where the "nucleusiness" is higher than the threshold value or those regions with the highest "cell nucleusiness" from a group of regions where the overlap between regions is higher than a threshold value, highly accurate results can be obtained. Therefore, we studied the possibility of classifying the presence or absence of cell nuclei using Mask-RCNN. We compared the performance of SWM and Mask-RCNN in detecting cell nuclei in oral cytological images and investigated whether AI is effective in identifying cell nuclei in oral cytological images.

2. Methods

2.1 Dataset

Specimens for oral cytology for the detection of oral mucosal diseases were collected at the Department of Oral and Maxillo-facial Surgery, Nagasaki University Hospital. Abrasion (conventional) cytology and liquid cytology were performed simultaneously at the time of specimen collection. For conventional cytology, only a small number of cells were collected, and a significant overlapping of cells was observed. Many foreign bodies including debris were also detected. However, a large number of cells were retrieved, and the debris was removed from the samples for liquid cytology, which made it easier to identify the cells. For this study, liquid cytology was performed. The microscopic images of cytopreparations were taken using a Nikon Eclipse Ti-S inverted microscope with a DS-Ri1 digital camera (Nikon Corp. Tokyo, Japan) equipped with a 40 × objective lens. The images were saved in tagged image file format at 1024 × 1280 pixels.

2.2 Papanicolaou classification of oral cytology

An oral pathologist evaluated all images. The Papanicolaou classification system consists of five classes. Class 1 tumors do not contain abnormal or atypical cells. Class 2 tumors have atypical cells but are not malignant. Class 3 tumors are suspected of being malignant and cannot be ruled out. Class 4 tumors are highly suspicious for malignancy. Class 5 tumors are almost certainly malignant. In addition, evaluation is made on the basis of cytoplasmic staining, cytoplasmic luminosity, cytoplasmic thickness and structure, cell shape, cell size difference, N/C ratio, nuclear shape, nuclear size difference, nuclear limbus appearance, nuclear number, chromatin amount, distribution and pattern, and nucleolus appearance.

2.3 Applications of SWM

First, an oral cytology image was divided into 96 pixels [Figs. 1(a) and 1(b)]. The images were manually sorted into those that contained cell nuclei and those that did not, and a dataset for classification was created using a convolutional neural network (CNN). Only images containing more than half of the cell nuclei in a sample were extracted to determine whether they actually were cell nuclei. During cytological examination (Papanicolaou staining), only the blue-line-age cell nuclei were extracted, because they are important cells for the early detection of dysmorphic cells (basal cells).

We only extracted the nuclei of the blue-stained cells. A total of 788 images with [Fig. 1(c)] and without [Fig. 1 (d)] nuclei were prepared. The data for cells with and without cell nuclei were then randomly divided into training and test data. Overall, 591 images of cells with or without nuclei were prepared for the training group, while 197 images were prepared for the evaluation group (from among 1576 images in total).

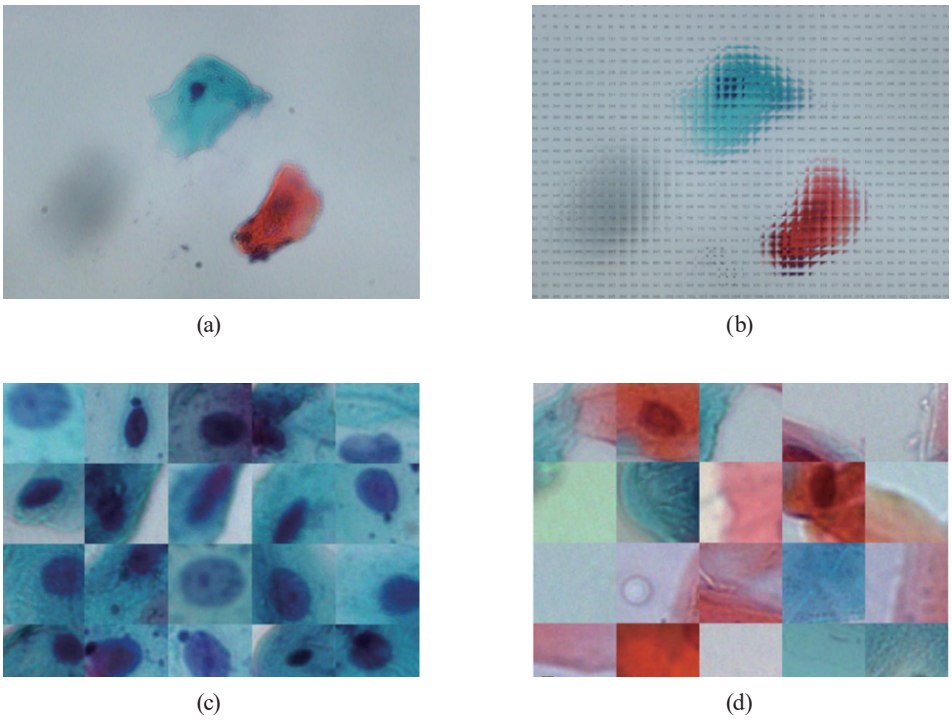


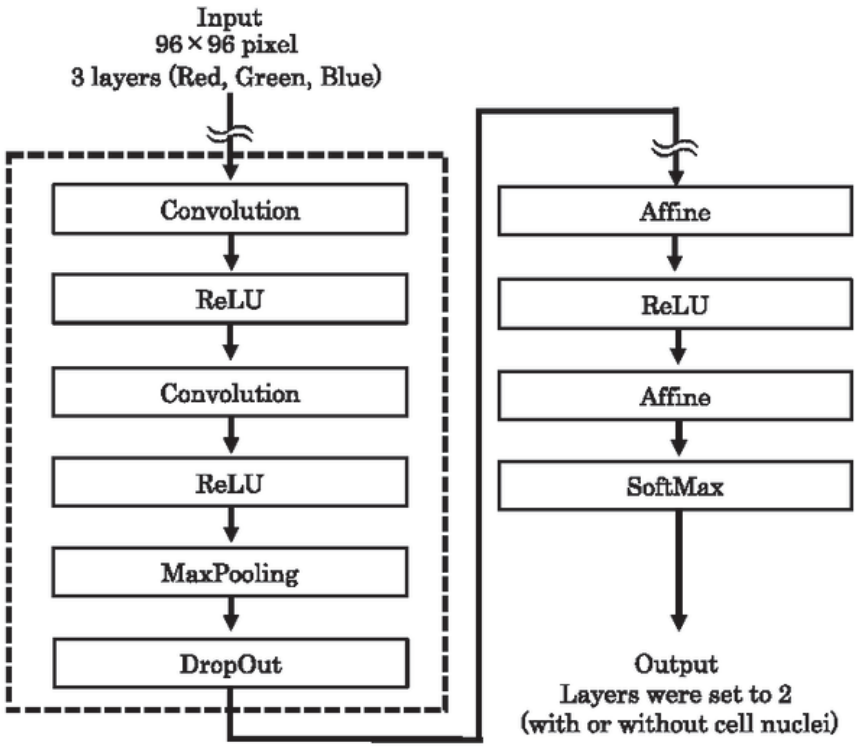
Fig. 1. (Color) Dataset preparation for SWM. (a) Original image from liquid-based oral cytology (1280 × 1024 pixels). (b) SWM applied; the size of the sliding window is 96 × 96 pixels. (c) Images of windows containing only cell nuclei. (d) Images of windows containing no cell nuclei.

Second, we constructed an image classifier, the CNN [Fig. 2 (a)]. The input image was set to 6 × 96 pixels with three channels (red, green, and blue), while the output layer was set to two channels (with or without cell nuclei). Categorical cross-entropy was used as a loss function to calculate the label and output errors, and optimization was performed using the ADAM software to update the network parameters. [12, 13] To obtain more accurate loss function values and results, we increased the number of epochs and dropped out of the system to avoid overlearning. We deepen the network (convolution, pooling, and increasing the set of activation functions) and varied the activation functions accordingly.

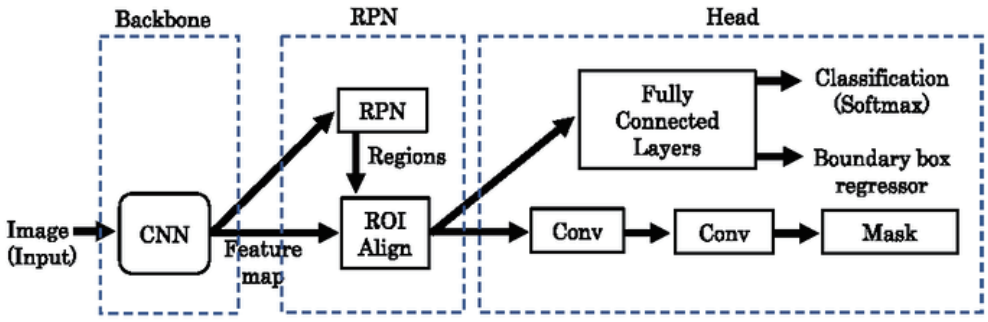
2.4 Mask-RCNN method

To train the Mask-RCNN, mask images were required as teaching images. A dataset of 130 images (original and mask images) of each class was constructed.

The structure of the image identifier with Mask-RCNN is shown in Fig. 2(b). Mask-RCNN was largely divided into three layers: backbone, region proposal network (RPN), and head. The backbone extracts features of the input image. The RPN determines whether each fixed region is correct and whether the overlap of regions is correct. The head layer pools the candidate RPN regions into the same image size and then calculates the probability for each class.



(a)



(b)

Fig. 2. Constructed image classifier. (a) Image classifier using the sliding window method. (b) Image classifier with Mask-RCNN. For the meaning of the abbreviations in Fig. 2, see Refs. 10–12.

2.5 Diagnostic performance of AI system

The AI systems were implemented on an NVIDIA GeForce RTX 2080Ti (NVIDIA Corp., Santa Clara, CA, USA) with an Intel Core i9-9900 K processor (Intel Corp., Santa Clara, CA, USA) and 64 GB of memory.

3. Results

3.1 Results of SWM

The CNN training results are shown in Table 1. In No. 6 in Table 1, which has the highest accuracy rate for training data, the value of the loss function for the training data is sufficiently small (0.0628). However, for the test data, it is not sufficiently small (0.4614). This indicates that result No. 6 in Table 1 is due to overlearning. In addition, although the accuracy rate for the training data is sufficiently high (99.8%), the accuracy rate for the test data is 89.1%, which is not sufficiently high. This confirmed that overlearning occurred. The CNN with the highest accuracy without overlearning was No. 2, which increased the proportion of correct answers to approximately 93% of the test data.

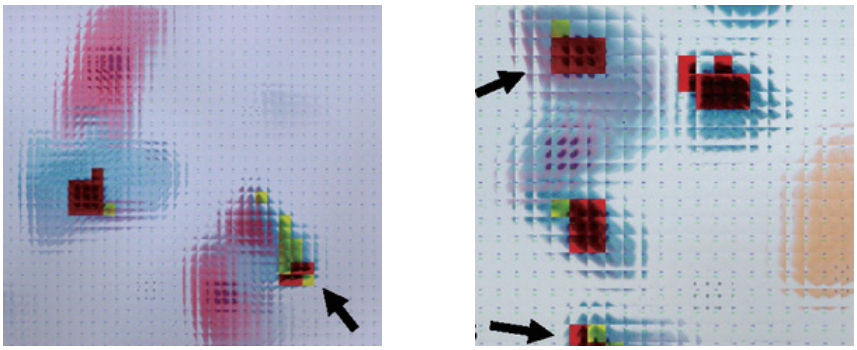
Figures 3(a) and 3(b) show the colored images indicated by CNN as cell nuclei. Red indicates that the probability of being a cell nucleus is more than 90%, while yellow indicates that the probability is more than 50%. As shown in Fig. 3(a), this finding was not sufficiently accurate because it reacted to the background image and did not react to the nuclei in the blue cell, as shown in Fig. 3(b). We also considered the possibility of overlearning owing to the presence of multiple cell nuclei in a single image.

3.2 Results of Mask-RCNN

At the beginning of the course (epoch 1), the value of the loss function was 0.8904, and the value of the validation loss function was 0.7262. At the end of the course (epoch 40), the value of the loss function was 0.4547, and the value of the validation loss function was 0.4820. The value of the loss function was lower at the end (40 epochs) compared with that at the beginning (one epoch). When the prepared images were examined using the model, we succeeded in detecting the cell nuclei, as shown in Fig. 4. In the other images, the accuracy of this method was higher than that of the SWM model [Fig. 2(a)]. This result shows that Mask-RCNN can be applied to the detection of Papanicolaou-stained cell nuclei in the oral cytology of liquid-based samples.

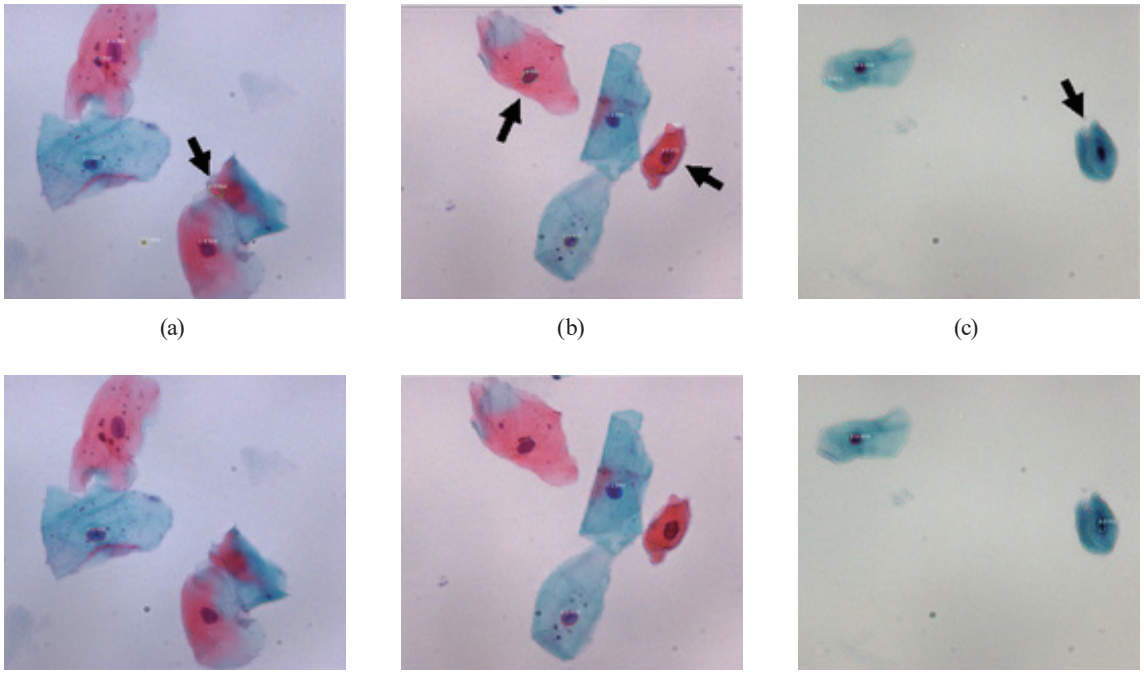
Table 1
Results of the SWM.

No.	Activation function	Epoch	No. of layers	Dropout	Value of the training loss function	Percentage of correct answers for the training group	Value of the loss function of the test group	Percentage of correct answers in the test group
1	ReLU	5	18	0.25	0.3356	0.8668	0.3266	0.8654
2	ReLU	20	18	0.25	0.0891	0.9562	0.2104	0.9314
3	ReLU	50	18	0.25	0.0526	0.9851	0.6082	0.9111
4	ReLU	5	24	0.25	0.351	0.8524	0.3188	0.8705
5	ReLU	20	24	0.25	0.0798	0.9251	0.1958	0.9314
6	ReLU	50	24	0.25	0.0628	0.9982	0.4614	0.8908
7	ReLU	5	18	0.5	0.3752	0.8226	0.3225	0.8705
8	ReLU	20	18	0.5	0.2151	0.9198	0.2514	0.9213
9	Sigmoid	20	18	0.25	0.7456	0.4659	0.7285	0.4785



(a) (b)

Fig. 3. (Color) Example of images which the CNN identified as cell nuclei. Red indicates cell nuclei with a probability greater than 90%, while yellow indicates cell nuclei with a probability greater than 50%. (a) An example of detection in the background (arrow) and (b) an example where a nucleus was not detected (arrows).



(a) (b) (c)

(d) (e) (f)

Fig. 4. (Color) Detection of cell nuclei using the region-based CNN method. (a) Red cell nuclei, Example 1 (arrow), Epoch 1, (b) blue absent cell nuclei, Example 2, Epoch 1 (arrows), (c) learning allowed the discrimination of the blue-tinted nuclei alone, Example 3, Epoch 1 (arrow), (d) Example 1, Epoch 40, (e) Example 2, Epoch 40, and (f) Example 3, Epoch 40.

3.2 Results of Mask-RCNN

At the beginning of the course (epoch 1), the value of the loss function was 0.8904, and the value of the validation loss function was 0.7262. At the end of the course (epoch 40), the value of the loss function was 0.4547, and the value of the validation loss function was 0.4820. The value

of the loss function was lower at the end (40 epochs) than that at the beginning (one epoch). When the prepared images were examined using the model, we succeeded in detecting the cell nuclei, as shown in Fig. 4. In the other images, the accuracy of the Mask-RCNN method was higher than that of the SWM method [Fig. 2(a)].

4. Discussion

4.1 Avoiding gradient loss function in CNN

The basic structure of the CNN used includes convolutional, pooling, dropout, and affine layers.[12] The generalization capability was improved so that it was possible to detect the features in other images as well. After applying the convolutional and pooling layers, the affine layer performed the compilation (all joins), while the softmax function was used to report the results.

The more convolutional and pooling layers that were used, the higher the expected accuracy; however, the gradient loss function occurred during training, which caused the accuracy to deteriorate. Because the error backpropagation method was used to train the neural network, the gradient was lost, learning was not possible, and the accuracy was not improved. The dropout layer prevented overlearning. Overlearning of the training data could be avoided by deliberately setting the output data to zero during the layer-to-layer transfer. Because the output data that dropped out changed randomly as the training was repeated, a new layer was learned each time, and overlearning of the training data was avoided. On the basis of these findings, a CNN was constructed.

4.2 Comparison of the results

In the detection of cell nuclei, the SWM and Mask-RCNN methods have different criteria for the loss function. Therefore, we will make comparisons without loss functions in this section. A visual comparison was performed. We randomly prepared 10 whole images of Class II and III tumors and assessed them using the SWM and Mask-RCNN methods on the basis of the following criteria: (1) number of detections, (2) percentage incorrectly detected from background, (3) average number of undetectable cell nuclear regions per cell nucleus, and (4) average number of regions detected for each cell nucleus (Table 2). The Mask-RCNN method showed a lower percentage of background detection and a lower average number of undetected regions per cell

Table 2
Comparison of the results of the SWM Mask-RCNN method.

	Sliding window	Mask R-CNN
Number of detections	320	37
Percentage of incorrectly detected from backgrounds	0.3031 (97/320)	0.027 (1/37)
Average number of undetectable cell nuclear regions per cell nucleus	1.8611 (67/36)	0 (0/36)
Average number of regions detected for each cell nucleus	6.1944 (223/36)	1 (36/36)

nucleus. These results suggest that the Mask-RCNN method is more accurate because it can detect the cell nuclei and does not react to extra regions. The average number of detected regions for each cell nucleus in the Mask-RCNN method was one (36/36). This finding indicates that there was no overlap in the number of detected cell nuclei, thus indicating that the data are suitable to use for the next stage of training. Figure 5 shows detection results for the visual evaluation of three images: Example 1, Example 2, and Example 3. Figures 5(a) and 5(d) are SWM and Mask-RCNN detection results for Example 1, respectively. Figures 5(b) and 5(c) are SWM and Mask-RCNN detection results for Example 2, respectively. Figures 5(c) and 5(f) are the detection results from SWM and Mask-RCNN for Example 3, respectively. Comparison of the detection results of SWM and Mask-RCNN for each example showed that the Mask-RCNN method could detect the outline of the cell nucleus. In other words, the regions other than the cell nuclei were omitted. This makes it easier to extract the features of the cell nucleus when the training is performed in the next stage. Therefore, the Mask-RCNN method appeared to be more accurate than the SWM in detecting the cell nuclei and could be applied to the next stage of training. However, the cell nuclei were more accurately detected using the Mask-RCNN method because the Class II and III images used in this study were relatively clean, with little debris in the dataset. The results in cases with large amounts of debris or overlapping cells still need to be determined. If the cell nuclei can be automatically detected, the accuracy of cytological diagnosis can be improved by detecting the cells using an attention mechanism.[14]

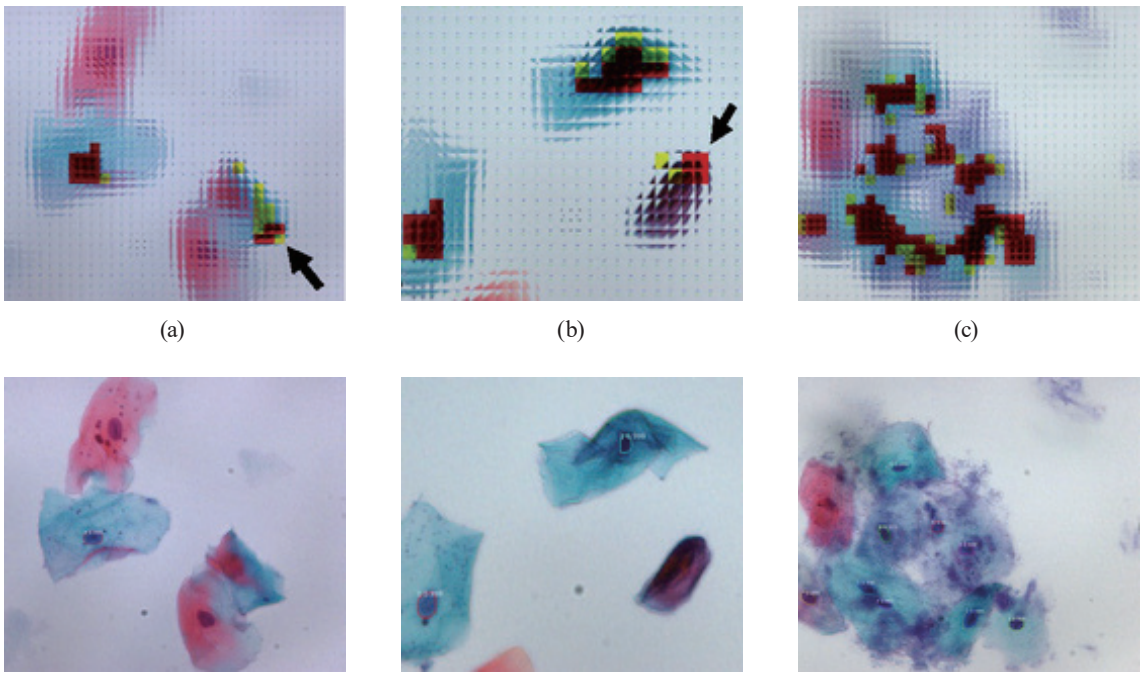


Fig. 5. (Color) Comparison of the detection accuracy between the SWM method and the Mask-RCNN) method using the same image from the end of epoch 40. (a) SWM Example 1, (b) SWM Example 2, (c) SWM Example 3, (d) Mask-RCNN Example 1, (e) Mask-RCNN Example 2, and (f) Mask-RCNN Example 3.

Diagnosis through cytological examination is based on cell morphology and the N/C ratio. However, in this study, the positivity rate was relatively high despite focusing only on the recognition of cell nuclei. This clarified the possibility of classifying cytological diagnoses by focusing on the morphology of cell nuclei. In the medical field, AI has been investigated for its potential use in diagnostic imaging and pathology. It has been suggested that AI may provide more accurate results than the manual reading of CT and chest X-ray images by radiologists.[15] The cytology of cervical cancer has been well studied,[2] and the number of cervical cancer cases is higher than that of oral cancer cases, making it easier to construct datasets. In contrast, oral cancer is rare (less than 6 people per 100000 population in Japan), making it difficult to collect cases and build datasets. We applied Mask-RCNN to the cytological screening of oral cancer. If the results of this study can be applied to other rare types of cancer, the early detection of cancer may be improved through cytological screening.

5. Conclusion

In this paper, we described the detection of cell nuclei in oral cytology using AI. We focused on the detection of cell nuclei because the ratio of cell nuclei to cytoplasm increases with increasing cell malignancy. As an initial step in the development of AI-assisted cytology, we investigated two methods for the automatic detection of cell nuclei in blue-stained cells in cytopreparation images.

We evaluated the usefulness of the SWM method and Mask-RCNN method in identifying the cell nuclei in oral cytopreparation images. Thirty cases of liquid-based oral cytology were analyzed. First, we carried out the SWM method by dividing each image into 96 × 96 pixels. Overall, 591 images with or without blue-stained cell nuclei were prepared as the training data and 197 as the test data (from among 1576 images in total). Next, we carried out the Mask-RCNN method by preparing 130 images of Class II and III lesions and creating mask images showing cell regions based on these images. Mask-RCNN is more accurate than SWM in identifying the cell nuclei.

Pathologists have found that if blue-stained cell nuclei can be automatically and accurately identified, the morphology of the entire cell can be understood more quickly, and the diagnostic performance of cytology can be expected to improve. A future task is to develop an AI system that can automatically and accurately identify blue-stained cell nuclei on the basis of the results of this study.

References

1 A. A. Borkowski, N. A. Viswanadhan, L. B. Thomas, R. D. Guzman, L. A. Deland, and S. M. Mastorides: Federal Practitioner **37** (2020) 398. https://doi.org/10.12788/fp.0045
2 H. Bao, H. Bi, X. Zhang, Y. Zhao, Y. Dong, X. Luo, D. Zhou, Z. You, Y. Wu, Z. Liu, Y. Zhang, J. Liu, L. Fang, and L. Wang: Gynecol. Oncol. **159** (2020) 171. https://doi.org/10.1016/j.ygyno.2020.07.099
3 CANCER STATISTICS IN JAPAN 2021: https://ganjoho.jp/public/qa_links/report/statistics/2021_en.html (accessed August 2022).
4 M. M. Cheng, Y. Liu, W. Y. Lin, Z. Zhang, P. L. Rosin, and P. H. S. Torr: Comput. Visual Media **5** (2019) 3. https://doi.org/10.1007/s41095-018-0120-1
5 P. Viola and M. Jones: Proc. 8th IEEE Int. Conf. Computer Vision (ICCV, 2001) 747.

6 H. A. Rowley, S. Baluja, and T. Kanade: IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998) 23. https://doi.org/10.1109/34.655647
7 N. Dalal and B. Triggs: Proc. Int. Conf. Computer Vision and Pattern Recognition (IEEE, 2005) 886.
8 G. Papandreou, I. Kokkinos, and P. A. Savalle: arXiv:1412.0296:Ar.Xiv (2014) https://doi.org/10.48550/arXiv.1412.0296
9 A. Arasu and G. Manku: Proc. Symp. Principles of Database Systems (PODS, 2004) 286.
10 H. Shen and Y. Zhang: J. Comput. Sci. Technol. **23** (2008) 973. https://doi.org/10.1007/s11390-008-9192-1
11 B. Kedra, M. Chomczyk, M. Zlotkowski, W. Stokowska, A. Borsuk, M. Bicz, M. Pietruska, G. Tokajuk, R. Charkiewicz, P. Czajka, L. Chyczewski, L. Zimnoch, and B. Kedra: J. Comput. Sci. Technol. **50** (2012) 375. https://doi.org/10.5603/19746
12 K. He, G. Gkioxari, P. Dolla, and R. Girshick: Mask-RCNN, https://arxiv.org/pdf/1703.06870.pdf (accessed August 2022).
13 D. P. Kingma and J. L. Ba: ADAM: A Method for Stochastic Optimization, https://arxiv.org/pdf/1412.6980.pdf (accessed August 2022).
14 H. Robinson, A. Rasheed, and O. San: Dissecting Deep Neural Networks, https://arxiv.org/pdf/1910.03879.pdf (accessed August 2022).
15 M. Lin, Q. Chen, and S. Yan: Network in Network, https://arxiv.org/pdf/1312.4400.pdf (accessed August 2022).