技術論文

A Novel Two-Stage Deep Learning Method for Enhanced Cell Nuclei Segmentation

Chetan Gupta　Rupesh Kumar　Amit Shakya　Shruti Phutke　Lalit Sharma

This paper, titled “A Novel Two-Stage Deep Learning Method for Enhanced Cell Nuclei Segmentation,” was presented at CVR-2025 (5th International Conference on Computer Vision and Robotics), held in hybrid mode at the National Institute of Technology Goa, India, from April 25–26, 2025.

Reprinted with permission. Copyright © 2025 CVR and SCRS.
Further use or distribution is not permitted without permission from CVR and SCRS.

Abstract

Cell nuclei segmentation is essential for microscopic image analysis. It facilitates the detailed micro-environmental insights for clinical studies. Automated nuclei segmentation can simplify the work of pathologists and address the variability and subjectivity among them, thereby improving diagnostic consistency and accuracy. Although deep learning (DL) techniques usually offer better performance than traditional methods for nuclei segmentation, they still struggle with challenges especially when the nuclei are clustered and overlapped with each other. To address these challenges, many effective deep learning techniques have been developed. Nonetheless, these approaches still exhibit limitations, such as the tendency to overlook certain nuclei. To mitigate this issue, we propose a two-stage network to boost accuracy by incorporating an enhancement network (second stage) on top of the widely used encoder-decoder architectures (first stage). The enhancement network refines the results by utilizing the decoder’s output and original image. The primary goal of this proposed method is to segment those regions that might have been overlooked by the base model, by re-considering the original image within the network. The study employs popular models such as U-Net, Micro-Net, and U-Net++ as base models. The results illustrate improvement of 0.8±1.8% in Precision, 1.5±3.0% in Recall, 1.4±0.5% in Dice Score, and 1.6±0.7% in IoU across different datasets through the proposed method.

1 INTRODUCTION

Cancer is the second leading cause of death globally which make the early detection crucial for better treatment[1]. Morphological changes in nuclei served as the biomarker for cancer diagnosis[2]. Currently, pathologists use microscopes to examine tissue samples, a method that is slow, labor-intensive and dependent on human skills[3]. Recent Research shows that pathologists often disagree on diagnoses, with a 24.7% disagreement rate[4]. It highlights the critical necessity for computer-assisted pathology to enhance diagnostic consistency and precision. Cell nuclei segmentation is the fundamental task in computer-assisted pathology, involving the identification and outlining of cell nuclei in microscopic images[5].

Traditional methods like Watershed segmentation[6], K-means clustering[7], and fuzzy C-means[8] have been used for this purpose but often struggle with blurry images[9]. As a result, researchers are seeking more effective and accurate solutions. Deep learning has proven to be a highly effective approach for segmenting cell nuclei, as it can automatically learn and extract features from images, resulting in more precise and reliable segmentation outcomes[10]. Unlike conventional methods, deep learning models excel in handling nuclei across a variety of image conditions. This flexibility is essential in biomedical imaging, where nuclei can differ significantly in size, shape, and intensity. Although single-stage training methods[11]-[15] are commonly used for medical image segmentation, some researchers have found the two-stage approach to be highly effective and promising. Existing methods often rely on specific

architectures, for instance, Jiang et al.[16] utilized a modified U-Net in a two-step process, where an initial simplified U-Net provides a preliminary result, followed by an enhanced model with additional decoding layers for refinement. In contrast, our method is flexible and independent of any particular architecture. It can integrate with any model in the first stage, offering adaptable and improved cell nuclei segmentation without being restricted to a specific design. In medical diagnostics, the structure and count of cell nuclei are crucial factors for decision-making[17]. For a cell nuclei segmentation model, one aspect is to accurately delineate the region of interest (cell nuclei) and another is to precisely segment the correct number of nuclei. However, there remains a gap between the quality and quantity of segmented nuclei. To bridge this gap, our paper introduces a two-stage cell nuclei segmentation approach falling under the category of representation enhancement. The proposed method uses existing encoder-decoder segmentation networks as the base model (first stage), followed by an enhancement network (second stage) to refine the initial prediction mask produced by the base network. This enhancement network is a lightweight model with approximately 77k trainable weights that can be easily integrated with the base model and produce better quantitative and qualitative results, as shown in the results section later. Notably, the proposed method not only captures initially missed cell nuclei but also refines the pixels along the boundaries of segmented nuclei. Our main contributions are as follows:

- We propose a novel two stage deep learning approach for cell nuclei segmentation which effectively enhances the segmentation results of existing methods.
- The proposed pipeline is architecture independent, allowing it to integrate various advanced segmentation methods in the first stage, followed by an enhancement network in the second stage. To illustrate its effectiveness, we evaluated it using three state-of-the-art segmentation models: U-Net[12], Micro-Net[13], and U-Net++[14].
- The comparison of proposed method is carried out on two publicly available histopathogical datasets.

The structure of the paper is as follows: Section 2 explains the proposed method, Section 3 details about the dataset used, evaluation metrics, and implementa-tion of proposed approach, Section 4 discusses the results, and finally, Section 5 concludes the paper.

2 PROPOSED METHOD

The proposed method, as shown in Figure 1, consists of two stages that aims to enhance the performance of the cell segmentation. The method uses an encoder- decoder network, followed by an enhancement network that improves the quality of initial prediction mask produced by the first network. In this section, deep convolutional encoder-decoder network has been explained that

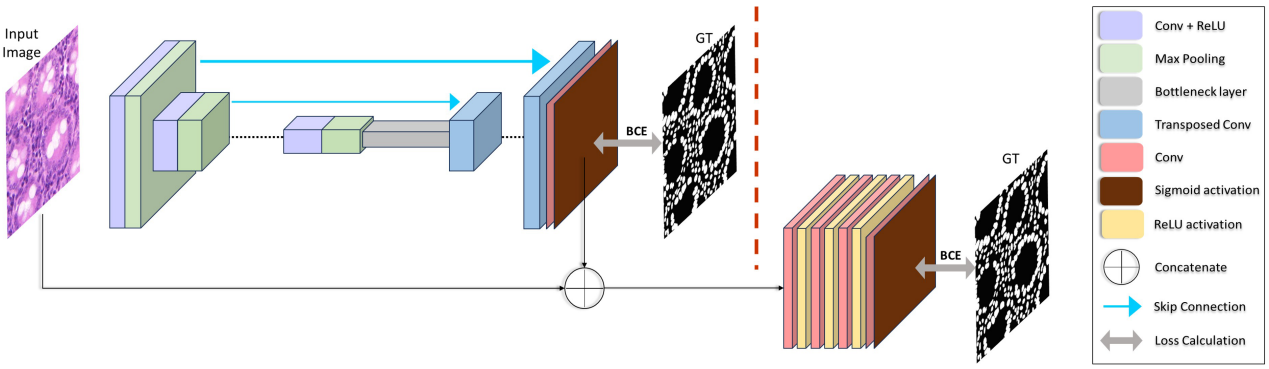

Fig. 1 Proposed method architecture. (a) first stage (architecture on the left side of red dashed line): an encoder-decoder based segmentation network. (b) second stage (architecture on the right side of red dashed line): enhancement network

constitutes the first stage of the proposed method. These networks take the original RGB image as input and produce a raw prediction mask. Subsequently, in the second stage, the enhancement network has been proposed, which refines the initial prediction mask, resulting in a more accurate and sharper edge mask.

2-1. Encoder-Decoder Stage

The first stage of our method comprises of a general segmentation network, as depicted in Figure 1(a). To comprehensively illustrate the effectiveness of the proposed method, we utilize three different popular segmentation networks U-Net, Micro-Net and U-Net++. U-Net[12] is a pioneering U-shaped encoder-decoder network specifically developed for efficient and accurate processing of biological images, even when trained on small datasets. It features a structured architecture with four encoder blocks that progressively down-sample the input image to capture feature representations, and four decoder blocks that up-sample these representations to produce the final segmentation. The skip connections between the encoder and decoder blocks help retain spatial details and enhance segmentation accuracy. Building on the foundational design of U-Net, Micro-Net[13] introduces additional convolutional layers that bypass the traditional max-pooling layers. This modification allows Micro-Net to more effectively capture and learn both weak and strong features in fluorescence and H&E stained images, improving its ability to segment cells, nuclei, and glands. The inclusion of intermediate connections between layers helps to preserve contextual information and localization, further enhancing segmentation performance. U-Net++[14] takes the advancements of U-Net a step further by integrating deeply supervised learning. It employs a series of nested, dense skip pathways that connect the encoder and decoder sub-networks. These redesigned skip pathways aim to bridge the semantic gap between the feature maps generated by the encoder and those used in the decoder, resulting in more accurate and refined segmentation. Each of these models builds upon the previous innovations, contributing to the development of increasingly sophisticated and precise segmentation techniques.

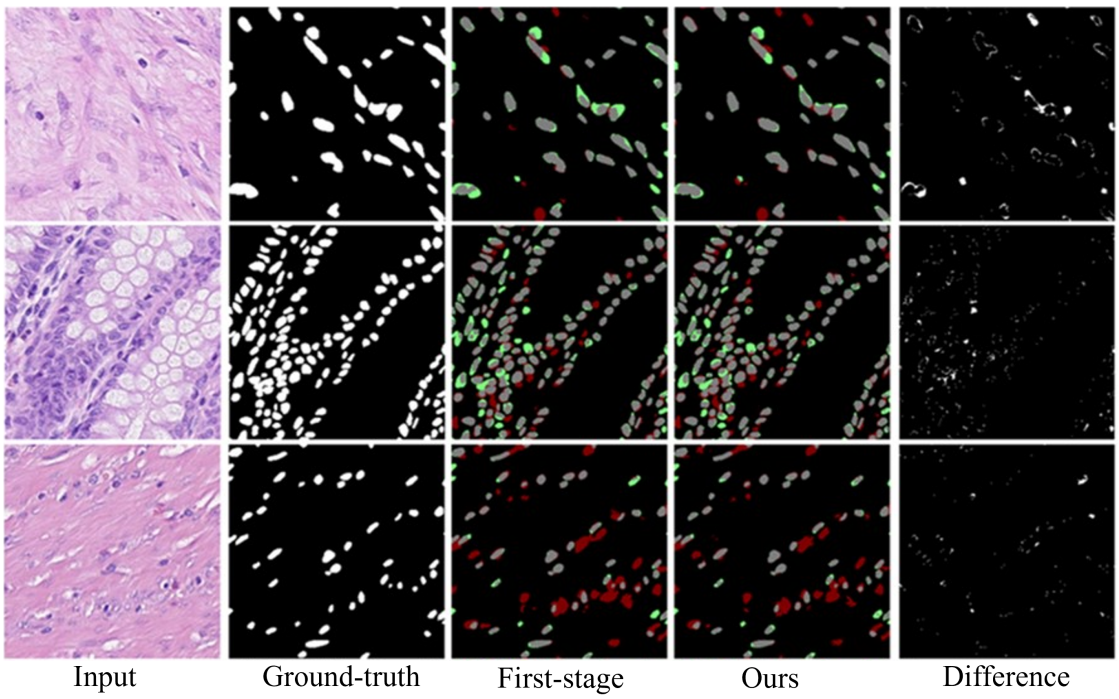

Fig. 2 Visual results of proposed refinement approach on Lizard dataset. **Red** indicates false positive pixels, **green** indicates false negative pixels and **gray** indicates true positive pixels. Fifth column - image obtained by difference between the refinement (ours) output and First-stage output. Images in lines 1, 2 and 3 are results from U-Net, Micro-Net, and U-Net++, respectively

2-2. Enhancement Stage

Enhancement stage is the second part of the proposed method, focusing on refining the initial prediction mask.

Network structure: The input image is concatenated channel-wise with the prediction from the first stage, creating a 4-channel input for the enhancement network. As illustrated in Figure 1(b), the network contains 4 convolutional layers, where each of the first three layers contains 64 filters of size 3×3 followed by 'ReLU' activation layer. The final layer is a 1×1 convolutional layer with 'sigmoid' activation. Notably, downsampling is deliberately omitted in each layer to preserve subtle structures that may have been missed in the first stage. As this enhancement stage is provided with the histopathological image and initial segmentation mask as input, the task here is to refine the segmentation results. This is done via extracting the features from the concatenated 4 channel input which already has localization of nuclei present in it. Processing this localized/referenced histopathological image with the proposed enhancement network helps to refine the earlier segmented nuclei by considering maximum receptive fields with the help of stack of convolution layers. The loss is calculated between the network's output and the corresponding ground truth.

Table 1 Comparison of evaluation metrics of segmentation networks with and without the proposed method, trained on Lizard dataset[18] († is proposed two-stage approach)

	Precision	Recall	Dice	IoU
U-Net	0.810	0.768	0.786	0.665
U-Net†	0.798	0.818	0.806	0.689
Micro-Net	0.790	0.774	0.780	0.655
Micro-Net†	0.804	0.780	0.790	0.664
U-Net++	0.692	0.854	0.757	0.627
U-Net++†	0.715	0.844	0.769	0.642

3 EXPERIMENTATION

3-1. Dataset

We utilize two datasets to demonstrate the generalization potential of our approach.

Lizard We use Lizard dataset[18], which is a large-scale dataset for colonic nuclear instance segmentation and classification. In total, there are 4981 image patches of size 256×256 which contain nearly half a million labeled nuclei in H&E stained colon tissue. The dataset has nuclear class labels for epithelial cells, connective tissue cells, lymphocytes, plasma cells, neutrophils and eosinophils. However, for our binary segmentation task, we pooled these different cell types into a single class. We split the dataset into training set, validation set and test set, each consisting of 3842, 499, and 640 image-mask pairs, respectively.

TNBC Triple Negative Breast Cancer (TNBC) dataset[19], consists of 50 images each of size 512×512 with a total of 4022 annotated cells that include normal epithelial and myoepithelial breast cells (localized in ducts and lobules), invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages and inflammatory cells (lymphocytes and plasmocytes). In our binary segmentation study, we group these diverse cell types together as a single foreground class. We divided dataset into three subsets: training set, validation set and test set, each containing 30, 10, and 10 image-mask pairs, respectively. After splitting the data, we addressed limitations of small dataset by applying data augmentation techniques like random horizontal flip, rotation and zoom, individually to each set.

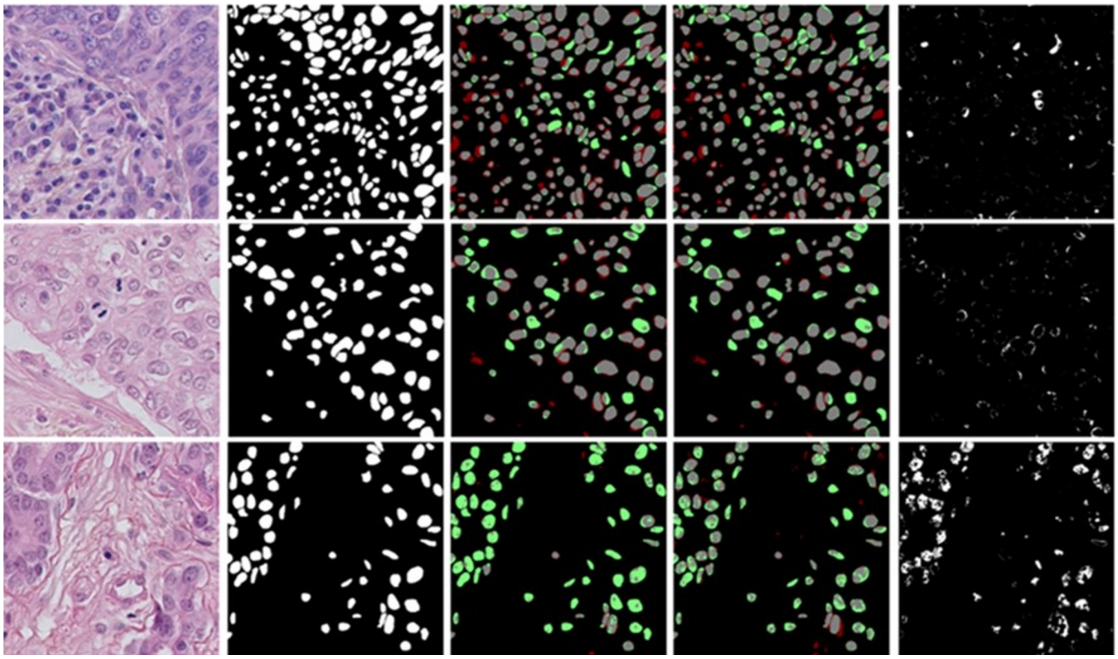

Fig. 3 Visual results of proposed refinement approach on TNBC dataset. **Red** indicates false positive pixels, **green** indicates false negative pixels and **gray** indicates true positive pixels. Fifth column - image obtained by difference between the refinement (ours) output and First-stage output. Images in lines 1, 2 and 3 are results from U-Net, Micro-Net, and U-Net++, respectively

3-2. Loss Function

During the training of both stages in the proposed method, we employ binary cross-entropy (BCE) loss.

$$\mathrm{BCE}(y_i, p(y_i)) = -\frac{1}{N}\sum_i (y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i))) \quad (1)$$

We calculate pixel-wise binary cross-entropy loss between actual pixel value y_i and predicted pixel value $p(y_i)$.

Table 2 Comparison of evaluation metrics of segmentation networks with and with- out the proposed method, trained on TNBC dataset[19] († is proposed two-stage approach)

	Precision	Recall	Dice	IoU
U-Net	0.800	0.857	0.827	0.705
U-Net†	0.840	0.826	0.832	0.708
Micro-Net	0.759	0.828	0.789	0.653
Micro-Net†	0.807	0.805	0.804	0.673
U-Net++	0.782	0.665	0.649	0.519
U-Net++†	0.801	0.703	0.721	0.577

3-3. Evaluation Metrics

To evaluate the performance of the proposed method, we employ the commonly used metrics, including Precision, Recall, Intersection over Union and Dice coefficient. Precision reflects the quality of results, while recall indicates the quantity of relevant items identified. Intersection over Union (IOU) measures the ratio of overlapping pixels to the total pixels in two compared images, with a value ranging from 0 to 1, 0 indicates no overlap, while 1 signifies complete overlap while the Dice score shows how similar two binary images are, like comparing a predicted mask to the actual mask. It is calculated by taking twice the amount of overlap between them and dividing by the total number of pixels in both images. A score of 1 means they match perfectly, while 0 means no overlap.

$$\mathrm{Precision} = \frac{tp}{tp+fp} \quad (2)$$

$$\mathrm{Recall} = \frac{tp}{tp+fn} \quad (3)$$

$$\mathrm{Dice\ Score} = \frac{2\cdot tp}{2\cdot tp+fp+fn} \quad (4)$$

$$\mathrm{IoU} = \frac{tp}{tp+fp+fn} \quad (5)$$

Here, *tp* and *tn* denote the total counts of accurately identified positive and negative cases. On the other hand, *fn* and *fp* indicate the counts of incorrectly identified negative and positive cases, respectively.

3-4. Implementation Details

As mentioned in the section 3.1, our proposed method is trained on two different datasets. The RGB images were first normalized to a scale of 0 to 1 before being fed to the model. The entire network undergoes a two-stage training process. After both parts of the network converge, the entire network is fine-tuned in a single-stage training session.

During first stage training, we first trained the encoder-decoder network without the enhancement network. After the encoder-decoder part converges, we freeze it and then update the enhancement network. After the enhancement part also converges, finally we fine-tune the whole network together.

We use the Adam optimizer with a learning rate of 0.0001 and apply binary cross-entropy loss, mentioned in section 3.2, during each training session. Additionally, each training is facilitated by early stopping technique which halts the training once the validation loss stops improving.

4 RESULTS

Table 1 and Table 2 showcase the evaluation scores for precision, recall, dice coefficient and IoU on Lizard and TNBC dataset, respectively. The symbol † on a segmentation network represents the combination of the encoder-decoder network followed by the enhancement network. In the tables, scores highlighted in blue indicate improvements in metrics with the proposed method. In each architecture, U-Net, Micro-Net and U-Net++, the integration of the enhancement network leads to improvements in at least three evaluation metrics. It can be noticed that the Dice coefficient improved in each case.

Figures 2 and 3 provide a clear visual analysis of the difference in the predictions of segmentation network with and without the enhancement network for Lizard and TNBC dataset respectively. The enhancements in the evaluation scores presented in the tables are visually reflected in Figure 2 and 3. The 'Difference' column is derived by subtracting the images in the 'First Stage Prediction' column from those in the 'Ours' column. White pixels in the images of the 'Difference' column indicate instances where the proposed method successfully identifies complete cells or parts of cells that were overlooked by the first stage network. It is important to note that some white pixels in these images may also result from false positive pixels generated by the proposed method, although they are relatively few in number.

5 CONCLUSION

Histopathological Images are complex in nature which makes segmentation of cell nuclei a challenging task. In this paper, we have proposed a method to integrate an enhancement network atop widely used existing segmentation networks. To showcase the generalizability of the proposed method, we apply it to diverse datasets, including Lizard and TNBC, across popular segmentation architectures like U-Net, Micro-Net, and U-Net++. The quantitative and qualitative results indicate that the proposed method not only refines pixels along the cell boundary but also segments cells that were initially missed by the first stage network. In the future, we would like to extend our experiments to encompass additional datasets across other popular segmentation architectures with the aim to improve the quality of output generated by the enhancement network.

REFERENCES

[1] Mortality, G., of Death Collaborators, C.: Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the global burden of disease study 2015. The Lancet 388, 1459-1544 (2016),

https://doi.org/10.1016/S0140-6736(16)31012-1
[2] Singh, I., Lele, T. P.: Nuclear morphological abnormalities in cancer: a search for unifying mechanisms. In: Nuclear, chromosomal, and genomic architecture in biol-ogy and medicine, pp. 443-467. Springer (2022)
[3] Onega, T., Barnhill, R. L., Piepkorn, M. W., Longton, G. M., Elder, D. E., Weinstock, M. A., Knezevich, S. R., Reisch, L. M., Carney, P. A., Nelson, H. D., et al.: Accuracy of digital pathologic analysis vs traditional microscopy in the interpretation of melanocytic lesions. JAMA dermatology 154(10), 1159-1166 (2018)
[4] Hsu, W., Han, S. X., Arnold, C. W., Bui, A. A., Enzmann, D. R.: A data-driven ap- proach for quality assessment of radiologic interpretations. Journal of the American Medical Informatics Association 23(e1), e152-e156 (2016)
[5] Alom, Z., Asari, V. K., Parwani, A., Taha, T. M.: Microscopic nuclei classification, segmentation, and detection with improved deep convolutional neural networks (dcnn). Diagnostic Pathology 17(1), 38 (2022)
[6] Beucher, S.: The watershed transformation applied to image segmentation. Scan- ning microscopy 1992(6), 28 (1992)
[7] Ahmed, M., Seraj, R., Islam, S. M. S.: The k-means algorithm: A comprehensive survey and performance evaluation. Electronics 9(8), 1295 (2020)
[8] Bezdek, J. C., Ehrlich, R., Full, W.: Fcm: The fuzzy c-means clustering algorithm. Computers & geosciences 10(2-3), 191-203 (1984)
[9] Thomas, R. M., John, J.: Detection and segmentation of mitotic cell nuclei in breast histopathology images. In: Proceedings of the 2017 International Conference on Networks & Advances in Computational Technologies (NetACT). pp. 246-250. IEEE (2017)
[10] Gudhe, N. R., Kosma, V. M., Behravan, H., Mannermaa, A.: Nuclei instance seg- mentation from histopathology images using bayesian dropout based deep learning. BMC Medical Imaging 23(1), 162 (2023)
[11] Xing, F., Yang, L.: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Reviews in Biomedical Engineering 9, 234-263 (2016)
[12] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomed-ical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234-241. Springer (2015)
[13] Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., Khan, M., Rajpoot, N. M.: Micro-net: A unified model for segmentation of various objects in microscopy images. Medical image analysis 52, 160-173 (2019)
[14] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th In- ternational Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. pp. 3-11. Springer (2018)
[15] Ibtehaz, N., Rahman, M. S.: Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. Neural networks 121, 74-87 (2020)
[16] Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In: Proceedings of the International MICCAI Brainlesion Workshop. pp. 231-241 (2019)
[17] Xing, F., Yang, L.: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE reviews in biomedical engineering 9, 234-263 (2016)
[18] Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y. W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al.: Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684-693 (2021)
[19] Naylor, P., Láe, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE transactions on medical imaging 38(2), 448-459 (2018)

■著者

Chetan Gupta
Emerging Technology and Innovation Lab, Yamaha Motor Solutions India

Rupesh Kumar
Emerging Technology and Innovation Lab, Yamaha Motor Solutions India

Amit Shakya
Emerging Technology and Innovation Lab, Yamaha Motor Solutions India

Shruti Phutke
Emerging Technology and Innovation Lab, Yamaha Motor Solutions India

Lalit Sharma
Emerging Technology and Innovation Lab, Yamaha Motor Solutions India