https://doi.org/10.2352/EI.2024.36.5.MLSI-310

Segmentation of Starch Granules in Microscopic Images Using a U-Net Model

Ye Jin, Pierce Cui and Jinshan Tang
Department of Health Administration and Policy, College of Public Health, George Mason University; Fairfax, VA 22310, USA

Abstract

Starch plays a pivotal role in human society, serving as a vital component of our food sources and finding widespread applications in various industries. Microscopic imaging offers a straightforward, efficient, and precise approach to examine the distribution, morphology, and dimensions of starch granules. Quantitative analysis through the segmentation of starch granules from the background aids researchers in exploring their physicochemical properties. This article presents a novel approach utilizing a modified U-Net model in deep learning to achieve the segmentation of starch granule microscope images with remarkable accuracy. The method yields impressive results, with mean values for several evaluation metrics including JS, Dice, Accuracy, Precision, Sensitivity and Specificity—reaching 89.67%, 94.55%, 99.40%, 94.89%, 94.23% and 99.70%, respectively.

1. Background

Natural starch mainly exists in plants, playing a critical role in plant growth and metabolism. It serves as a vital source for various human products, including food, clothing, and paper, contributing significantly to our bioeconomic value. Starch granules are composed of two glucan polymers, namely amylose and amylopectin [11]. These granules generally exhibit circular, elliptical, oval, or polyhedral shapes, with most having random overlapping distributions within the 1-100μm range [6]. However, the morphology, size, and distribution of starch granules vary among different plant species, resulting in distinct physicochemical properties. These variations have implications for diverse applications in industrial products and human health, such as resistant starch that can improve colon health and regulate blood glucose levels. This natural starch is mainly derived from potato tubers due to the characteristics of large round granules with a smooth surface [2]. Researchers may sometimes need to modify starch granules according to product requirements. Therefore, the exploration of starch is a significant task for researchers, ensuring its adaptability to various applications.

To deepen our comprehension of starch granule properties, conducting quantitative analysis and evaluation involving distribution, morphology, and size is essential. Image segmentation algorithms can help segment starch granules from the background. Guo et al. [4] developed an improved watershed segmentation algorithm, they used GVF field analysis and Fuzzy c-means method to reduce over-segmentation. However, various deep learning methods based on convolutional neural network (CNN) improvements are insensitive to image noise and contrast during feature learning, which is superior to earlier machine learning and traditional imaging techniques [13]. Therefore, deep learning methods have made great achievements in the field of image segmentation.

Ronneberger et al. [12] proposed a U-Net model based on CNN, capable of not only classifying and locating each pixel for segmentation but also achieving relatively high accuracy with only a small amount of data. Gu et al. [3] applied a CE-Net model which added the inception structure to U-Net to prevent more information loss during image segmentation and proved to be superior to U-Net. Alom et al. [8] combined U-Net, Residual Network and Recurrent Neural Network (RNN) to develop RU-Net and R2U-Net, enhancing the segmentation performance. Jonathan et al. [7] applied a Fully Convolutional Network (FCN) that can accept inputs of any size and combine shallow and deep feature information to achieve more accurate segmentation results. Badrinarayanan et al. [1] employed a Seg-Net Model, significantly reducing the number of parameters while outperforming the FCN model. Wang et al. [14] proposed the integrated Zoom-In-Net to highlight suspicious regions for segmentation. Moeskops et al. [10] presented an automatic segmentation of anatomical MR brain images into multiple classes method using multi-scale CNN, achieving accurate segmentation details as well as spatial consistency. Fausto et al. [9] showed a V-Net implementation based on FCN for 3D image segmentation. Zhou et al. [15] proposed a U-Net++ model that allows the network to automatically learn different layer features and reduce parameters through pruning.

Therefore, this article will present the application of the U-Net model to accomplish the segmentation of starch granules through deep learning methods.

2. Methods

2.1. Existing Algorithm of The Deep Learning-Based Segmentation

We employed the U-Net architecture as described in [12], comprising three components: down-sampling (encoding process), up-sampling (decoding process), and skip connections. The down-sampling allows the model to compress input image features, preserving key information and serving as an encoder. Up-sampling transforms low-resolution feature maps into high-resolution ones and outputs a segmented image of the original size. The skip connection fuses the pixel-level features of the image with the semantic-level features to achieve pixel-level semantic segmentation. Additionally, we utilized the same convolution to keep the image size unchanged after convolution process. Figure 1 illustrates our 4-layer U-Net model designed for starch granules segmentation.

2.2. Improvements

In starch granule images, the low variability in contrast between the foreground part (starch granules) and the background is beneficial for segmentation accuracy. However, challenges arise due to the blurred edges of starch granules, particularly when dealing with overlapping granules. To address this, we apply Gaussian blur with a kernel size of (7,7) and sigmax set to 0 to eliminate noise from the image and utilize the Canny edge detection algorithm to get the

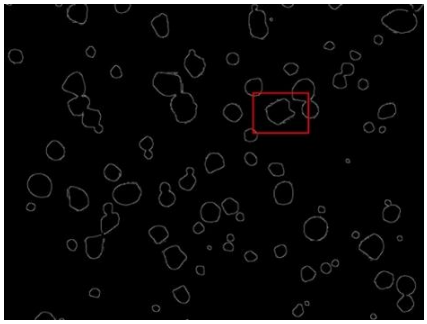

Figure 2a. *Before removing noise*

Figure 2b. *Before removing noise*

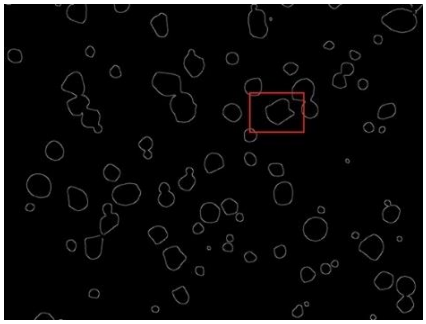

Figure 2c. *After removing noise*

Figure 2d. *After removing noise*

edges of the starch granules. The obtained edge channel is then added to the original RGB image to get a 4-channel image. Figure 2a and Figure 2b depicts the image before noise removal and Figure 2c and Figure 2d show the image after noise removal. Figure 3 shows 4-channle image after employing edge detection.

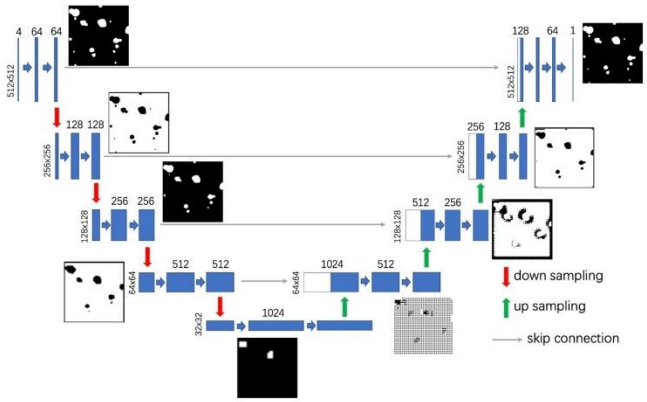

Figure 1. *An illustration of 4-layer U-Net model for starch granules segmentation*

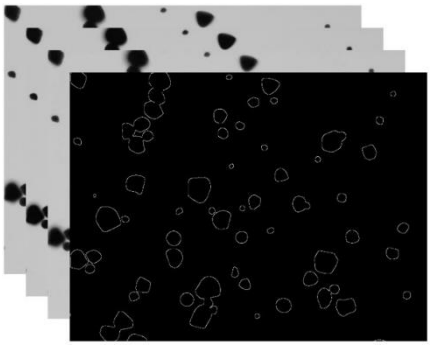

Figure 3. *RGB-edge (4-channels image)*

2.3. Implementation Details

The U-Net segmentation model was implemented using the PyTorch framework and tested on a server equipped with an Intel® Xeon® Silver 4314 CPU (2.40 GHz; 503 GB RAM) and two NVIDIA A100-PCIE GPUs (40GB GPU RAM). The initial learning rate was set to 0.01, gradually reduced through supervised decay with a coefficient of 0.1. An Adam optimizer was selected to train the proposed model with a batch size of 5.

3. Results

3.1. Datasets

We performed in situ analysis by observing iodine-stained tissue sample under a light microscope and capturing images with a high-resolution camera. This method is straightforward, efficient, and relatively accurate.

After acquiring the starch granule images, manual labeling was performed to create the ground truth dataset. Figure 4a and 4b depicts the microscopic image and corresponding ground truths.

Each of 20 (1360x1024 pixel) starch granule images was cropped into 6 (512x512 pixel) images, resulting in a total of 120 images,

constituting a 3-channel dataset. Manual labeling was then conducted on these 120 images to generate a ground truth dataset. Subsequently, noise reduction and edge detection techniques were applied to obtain a 4-channel dataset. The dataset was split, with 100 images allocated for the training set, 10 for the validation set, and another 10 for the test set. Figure 5 illustrates the data processing flow.

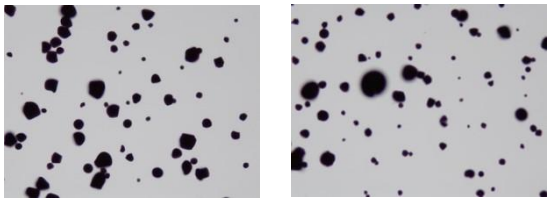

Figure 4a. *Microscopic Images*

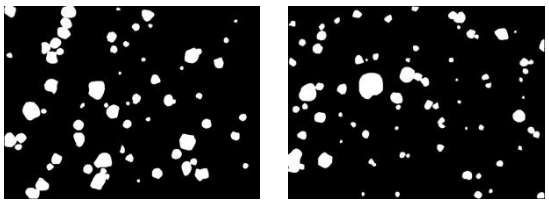

Figure 4b. *Ground Truths*

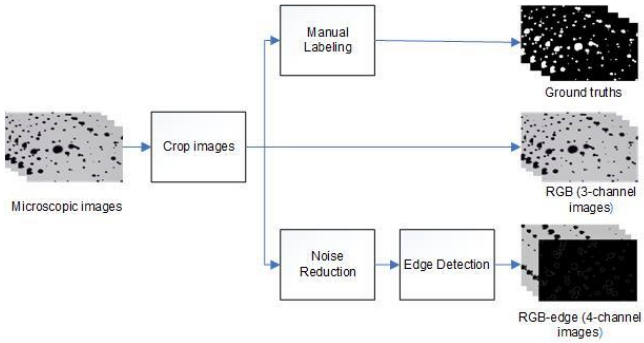

Figure 5. *Data Processing*

3.2. Performance Measurement

We chose Accuracy (ACC), Precision (PREC), Sensitivity (SE), Specificity (SP), Dice similarity coefficient (DSC), and Jaccard similarity (JS) as our evaluation metrics. The first four metrics are frequently employed for classification objectives, whereas the latter pair, DSC and JS, rely on the ground truth (GT) and are utilized for assessing the regional likeness between the segmentation outcomes and GT annotations. ACC, PREC, SE, SP, DSC, and JS are defined as follows:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$\text{PREC} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{SE} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{SP} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{DSC} = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{5}$$

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

In the performance comparisons metric table, the upward arrow ↑ indicates that the larger the value, the better the performance. The downward arrow ↓ means the opposite. The best results are highlighted in bold.

3.3. Quantitative Evaluation

We explored the optimal number of layers for the U-Net model structure in starch granule image segmentation. Our results reveal marginal differences in metrics across models with varying numbers of layers. Table 1 demonstrates that the 4-layer model exhibits superior performance in the JS, DSC, and SE metrics, with values of 89.23%, 94.30%, and 93.83%, respectively. Additionally, the 5-layer model achieves higher scores in ACC, PREC, and SP, with values of 99.47%, 95.35% and 99.73%, respectively.

Table 1: Quantitative performance comparisons (mean ± 95% confidence interval) for various layer models using six metrics.

Model	JS(%)↑	DSC(%)↑	ACC(%)↑	PREC(%)↑	SE(%)↑	SP(%)↑
3-layer	89.17 ±1.59	94.26 ±0.89	99.38 ±0.12	95.04 ±1.30	93.54 ±1.58	99.71 ±0.09
4-layer	**89.23 ±1.56**	**94.30 ±0.87**	99.38 ±0.13	94.81 ±1.42	**93.83 ±1.24**	99.70 ±0.11
5-layer	88.93 ±1.60	94.13 ±0.9	**99.47 ±0.12**	**95.35 ±1.18**	92.98 ±1.57	**99.73 ±0.09**

Subsequently, we discovered that utilizing 4-channel images (incorporating noise reduction and edge detection channel) will expedite the training process. Figure 6 presents a comparison of the time taken, as well as the validation dice and loss metrics, between 3-channel and 4-channel datasets.

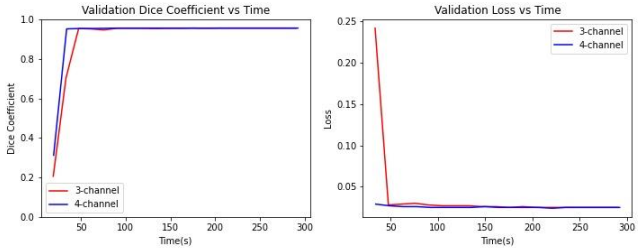

Figure 6. *Validation dice and loss metrics comparison between 3-channel images (without edge channel) and 4-channel images*

Table 2 compares the segmentation performance of 3-channel and 4-channel models across six metrics. The 4-layer model using 4-channel images demonstrates superior performance in the JS, DSC, ACC, SE and SP, achieving values of 89.23%, 94.30%, 99.38%, 93.83%, and 99.70%, respectively. The 4-layer model using 3-channel images achieves a higher score in PREC, with a value of 94.89%. Figure 7d illustrates the enhanced ability of the 4-channel model in detecting the edges of starch granules. Figure 8 presents

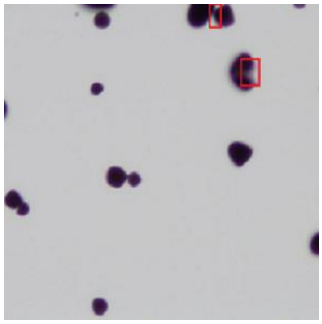

Figure 7a. *Starch Microscopic Image*

Figure 7b. *Ground Truth Image*

Figure 7c. *Segmentation Result Generated by 3-channel Model*

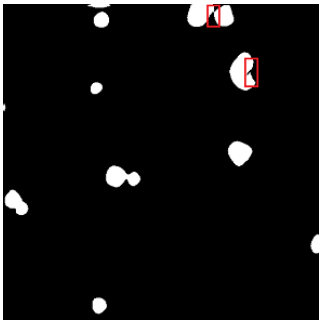

Figure 7d. *Segmentation Result Generated by 4-channel Model*

the Dice and loss metrics throughout the training and validation phases for a 4-layer model trained on a 4-channel image dataset.

Table 2: Quantitative performance comparisons (mean ± 95% confidence interval) for 3-channel and 4-channel utilizing six metrics.

Model	JS(%)↑	DSC(%)↑	ACC(%)↑	PREC(%)↑	SE(%)↑	SP(%)↑
3-channel	88.85 ±1.71	94.08 ±0.97	99.36 ±0.12	**94.89 ±1.30**	93.32 ±1.35	99.70 ±0.09
4-channel	**89.23 ±1.56**	**94.30 ±0.87**	**99.38 ±0.13**	94.81 ±1.42	**93.83 ±1.24**	**99.70 ±0.11**

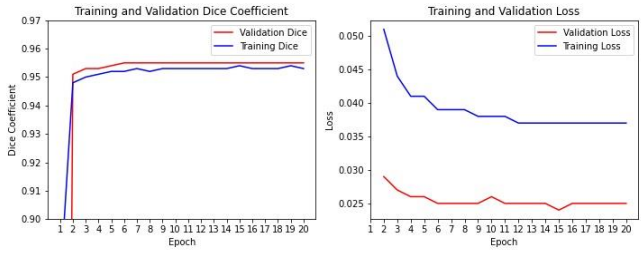

Figure 8. *Dice and loss metrics during training and validation for a 4-layer model using 4-channel images*

Loss Functions

We further explored the performance of applying various loss functions and compared their performance across six matrix using a 4-layer model trained on 4-channel image dataset. Mean Squared Error (MSE) loss and Binary Cross-Entropy(BCE) loss emerged as top contenders. MSE loss calculates the average of squared differences between the actual and predicted values:

$$\text{MSE Loss} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (7)$$

BCE loss represents cross-entropy loss, commonly used in binary classification task:

$$\text{BCE loss} = -\frac{1}{n}\sum_{i=1}^{n}(y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)) \quad (8)$$

where y_i is true target value for the i-th sample, $\hat{y}_i$ is the predicted value for the i-th sample.

Table 3 presents a quantitative comparison of the segmentation performance between BCELoss and MSELoss using a 4-layer model trained on 4-channel dataset across six metrics. The MSELoss function demonstrates superior performance in all six matrix: JS, DSC, ACC, PREC, SE and SP, achieving values of 89.67%, 94.55%, 99.40%, 94.89%, 94.23% and 99.70%, respectively.

Table 3: Quantitative performance comparisons (mean ± 95% confidence interval) for various loss function using six metrics.

Loss Function	JS(%)↑	DSC(%)↑	ACC(%)↑	PREC(%)↑	SE(%)↑	SP(%)↑
BCELoss	89.23 ±1.56	94.30 ±0.87	99.38 ±0.13	94.81 ±1.42	93.83 ±1.24	**99.70 ±0.11**
MSELoss	**89.67 ±1.32**	**94.55 ±0.74**	**99.40 ±0.12**	**94.89 ±1.34**	**94.23 ±1.01**	**99.70 ±0.1**

Comparative Models Our U-Net model was also compared with nnU-Net [5]. To be fair, out U-Net implementation and nnU-Net were trained and tested on the same data with same environment configuration.

Quantitative Results The segmentation performance of two models on starch dataset is detailed in Table 4. The nnU-Net model yields the best performance in the JS, DSC and SE metrics, achieving values of 89.84% , 94.64% and 95.01%, respectively. Moreover, our 4-layer model utilizing MSELoss function achieves the higher score in the ACC, PREC and SP, with values of 99.40%, 94.89% and 99.70%, respectively.

Table 4: Quantitative performance comparisons (mean ± 95% confidence interval) between out U-Net and nnU-Net segmentation models using six metrics.

Model	JS(%)↑	DSC(%)↑	ACC(%)↑	PREC(%)↑	SE(%)↑	SP(%)↑
4-layer	89.67 ±1.32	94.55 ±0.74	**99.40 ±0.12**	**94.89 ±1.34**	94.23 ±1.01	**99.70 ±0.1**
nnU-Net	**89.84 ±1.01**	**94.64 ±0.56**	**99.40 ±0.10**	94.31 ±1.34	**95.01 ±0.83**	99.67 ±0.08

4. Discussion

Our image-augmentation based U-Net model obtained good segmentation results, which demonstrate the effectiveness of deep learning in segmenting starch granules. This is also due to the strong contrast and low variability between the foreground and background

in starch granule images. However, on the one hand our model still needs improvement. Nevertheless, there is room for improvement. The evaluation metrics, including mean JS and mean Dice, fall below 90% and 95%, respectively, potentially due to the common occurrence of overlapping or sticking starch granules. Addressing this challenge requires a distinct boundary to facilitate the separation of starch granules with different sizes and shapes. Furthermore, in our starch granule images, the background area surpasses the foreground area, creating a data imbalance that may hinder the extraction of small foreground regions. The use of Binary Cross Entropy Loss and Mean Squared Error Loss in this context are not conducive to effectively segmenting unbalanced samples, necessitating the exploration of a loss function that emphasizes the foreground. Additionally, in deep learning methods, manual annotation is time-consuming, posing challenges for image preprocessing. Finding more efficient annotation strategies could alleviate this issue.

5. Conclusion

Deep learning has found wide applications in different fields, including image enhancement [16], image retrieval [17], image segmentation [18], and speckle reduction [19]. Compared with traditional machine learning models]20][21], deep learning can obtain high accuracy in different applications. This paper employed a deep learning approach to segment microscopic images of starch granules. And we successfully extracted starch granules from the background using the classification and localization features of the U-Net model, and obtained high accuracy, illustrating the effectiveness of the deep learning method.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.

[2] G. E. Bednar, A. R. Patil, S. M. Murray, C. M. Grieshop, N. R. Merchen, and G. C. Fahey, “Starch and Fiber Fractions in Selected Food and Feed Ingredients Affect Their Small Intestinal Digestibility and Fermentability and Their Large Bowel Fermentability In Vitro in a Canine Mode,” The Journal of nutrition, vol. 131, no. 2, pp. 276–286, 2001.

[3] Z. Gu et al., “CE-Net: Context Encoder Network for 2D Medical Image Segmentation,” IEEE Transactions on Medical Imaging, vol. 38, no. 10, pp. 2281–2292, 2019, doi: 10.1109/TMI.2019.2903562.

[4] S. Guo, J. Tang, Y. Deng, and Q. Xia, “An improved approach for the segmentation of starch granules in microscopic images,” BMC genomics, vol. 11 Suppl 2, no. Suppl 2, pp. S13–S13, 2010.

[5] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” Nature methods, vol. 18, no. 2, pp. 203–211, 2021.

[6] J.-L. Jane, T. Kasemsuwan, S. Leas, H. Zobel, and J. F. Robyt, “Anthology of Starch Granule Morphology by Scanning Electron Microscopy,” Starch - Stärke, vol. 46, no. 4, pp. 121–129, 1994, doi: https://doi.org/10.1002/star.19940460402.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440. doi: 10.1109/CVPR.2015.7298965.

[8] Md Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” CoRR, vol. abs/1802.06955, 2018, [Online]. Available: http://arxiv.org/abs/1802.06955.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571. doi: 10.1109/3DV.2016.79.

[10] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, “Automatic Segmentation of MR Brain Images With a Convolutional Neural Network,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1252–1261, 2016, doi: 10.1109/TMI.2016.2548501.

[11] S. Pérez and E. Bertoft, “The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review,” Starch - Stärke, vol. 62, no. 8, pp. 389–420, 2010, doi: https://doi.org/10.1002/star.201000013.

[12] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer Assisted Intervention − MICCAI 2015, Cham: Springer International Publishing, 2015, pp. 234–241.

[13] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical image segmentation using deep learning: A survey,” IET image processing, vol. 16, no. 5, pp. 1243–1267, 2022.

[14] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, “Zoom-in-Net: Deep Mining Lesions for Diabetic Retinopathy Detection,” in Medical Image Computing and Computer Assisted Intervention − MICCAI 2017, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds., Cham: Springer International Publishing, 2017, pp. 267–275.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation,” IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1856–1867, 2020, doi: 10.1109/TMI.2019.2959609.

[16] J. Tang, Q. Sun, and K. Agyepong, "An Image Enhancement Algorithm Based on a Contrast Measure in the Wavelet Domain for Screening Mammograms," *2007 IEEE International Conference on Image Processing*, San Antonio, TX, USA, 2007, pp. V - 29-V – 32.

[17] J. Tang and S. Acton, "An image retrieval algorithm using multiple query images," *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.*, Paris, France, 2003, pp. 193-196.

[18] Juanjuan He, Qi Zhu, Kai Zhang, Piaoyao Yu, Jinshan Tang, An evolvable adversarial network with gradient penalty for COVID-19

infection segmentation, Applied Soft Computing, Volume 113, Part B,2021,107947, ISSN 1568-4946.

[19] Liu, X., Liu, J., Xu, X. *et al.* A robust detail preserving anisotropic diffusion for speckle reduction in ultrasound images. *BMC Genomics* 12 (Suppl 5), S14 (2011). https://doi.org/10.1186/1471-2164-12-S5-S14

[20] J. Tang, X. Liu, H. Cheng, and K. M. Robinette, "Gender Recognition Using 3-D Human Body Shapes," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 898-908, Nov. 2011.

[21] J. Xu, Y. -Y. Cao, Y. Sun and J. Tang, "Absolute Exponential Stability of Recurrent Neural Networks with Generalized Activation Function," in *IEEE Transactions on Neural Networks*, vol. 19, no. 6, pp. 1075-1089, June 2008.

Author Biography

Ye Jin received her BS and MS in computer science from Peking University, Beijing, China. She is currently pursuing the Ph.D. degree in health informatics in the Department of Health Administration and Policy, College of Public Health at George Mason University, Fairfax, VA, USA. Her research interest includes artificial intelligence in medicine (e.g., computer-aided cancer detection and diagnosis), biomedical image analysis, and natural language understanding and processing.

Pengyuan Cui received his BS degree in College of Pharmacy from China Pharmaceutical University, Nanjing, China. He also received an MS degree in health informatics in the Department of Health Administration and Policy, College of Public Health at George Mason University, Fairfax, VA, USA. His research interest includes clinical statistical analysis and artificial intelligence in medicine. Since then, he has worked on clinical trial data analysis and management in Shanghai, China.

Jinshan Tang is a professor of Health Informatics in the Department of Health Administration and Policy at George Mason University. Before joining George Mason University, Dr. Tang was a full professor in the College of Computing at Michigan Technological University and a Founding Director of the Joint Center for Biocomputing and Digital Health. Dr. Tang's research covers broad areas related to image processing and artificial intelligence. His specific research interests include biomedical image analysis, biomedical imaging, artificial intelligence in medicine (e.g., computer-aided cancer detection, AI for COVID-19 detection). He has obtained over three million dollars grants as a PI or Co-PI and has published more than 130+ refereed journal and conference papers. He has also served as a committee member at various international conferences. He is a senior member of IEEE and a Co-chair of the Technical Committee on Information Assurance and Intelligent Multimedia-Mobile Communications, IEEE SMC society.