

# NUCLEAR DENSITY DISTRIBUTION FEATURE FOR IMPROVING CERVICAL HISTOPATHOLOGICAL IMAGES RECOGNITION

Zhuangzhuang Wang<sup>1</sup>, Mengning Yang<sup>1</sup>, Yangfan Lyu<sup>2</sup>, Kairun Chen<sup>1</sup>, Qicheng Tang<sup>1</sup>

<sup>1</sup>School of Big Data and Software Engineering, Chongqing University, Chongqing, China

<sup>2</sup>Department of Pathology, Second Affiliated Hospital, Army Medical University, Chongqing, China

## ABSTRACT

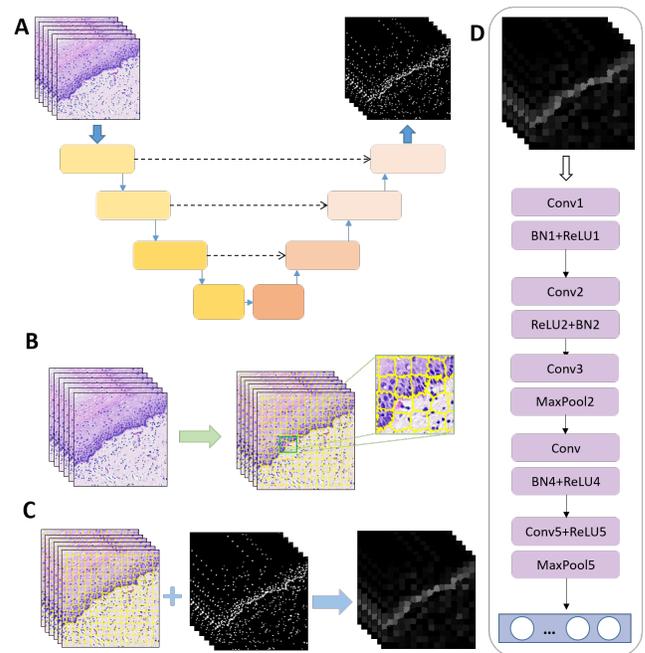
Cervical carcinoma is a common type of cancer in the female reproductive system. Early detection and diagnosis can facilitate immediate treatment and prevent progression of the disease. However, in order to achieve better performance, DL-based algorithms just stack various layers with low interpretability. In this paper, a robust and reliable Nuclear Density Distribution Feature (NDDF) based on priors of the pathologists to promote the Cervical Histopathological Image Classification (CHIC) is proposed. Our proposed method combines the nucleus mask segmented by U-Net with the segmentation grid-lines generated from pathology images utilizing SLIC to obtain the NDDF map, which contains information about the morphology, size, number, and spatial distribution of nuclei. The result shows that the proposed model trained with NDDF maps has better performance and accuracy than that trained on RGB images (patch-level histopathological images). More significantly, the accuracy of the two-stream network trained with RGB images and NDDF maps is steadily improved over the corresponding baselines of different complexity.

**Index Terms**— Nuclear Density Distribution, Superpixel, Deep Learning, Histopathological Image Classification

## 1. INTRODUCTION

In the training of traditional fully supervised deep learning methods, digital pathology images are fed directly into the model for efficient training, and then the trained model is leveraged on the test set to rapidly diagnose[1]. Neural networks can learn color, texture, and other information well from RGB images. Nevertheless, owing to inconsistent degrees of staining, messy nuclei, and complex fields of view, the network could learn invalid information, resulting in the need for more complex networks and larger training datasets to get good results. Meanwhile, a crucial problem facing deep learning is the lack of general theory in that interpretation for

the network design which is essential in the medical field. Therefore, it makes sense to extract features based on the prior knowledge of pathologists to improve the performance of the model. Some methods have explored cervical cancer diagnosis using digital image processing techniques to facilitate histopathological image classification by extracting features such as morphology and number of cell nuclei[2, 3, 4]. But there has never been a comprehensive and integrated feature that does not affect the performances of the model to represent histopathological images.



**Fig. 1.** Processing steps of the proposed approach. A: The U-Net structure-like neural network model ANet segments the nucleus mask. B: SLIC is utilized to generate segmentation grid-lines from histopathological images. C: The nucleus density is calculated in each region. D: The DNet model contains a 5-layer convolutional neural network for training utilizing NDDF maps.

The nuclei distribution feature of the epithelium region

This work is supported by the Fundamental Research Funds for the Central Universities under Grant 2020CDCGRJ013 and the Science and Technology Innovation Capacity Enhancement Special Project of Army Medical University under Grant 2019XQY14.

is completely different in normal and cancerous images: the nuclei in normal epithelial tissues are densely to sparsely distributed, with a clear direction of distribution; in cancerous epithelial tissues, there is no such clear direction of growth, and sometimes even the base cannot be distinguished. Moreover, nuclear atypia is characterized by the enlargement of the nucleus, resulting in the different shape and size of the nucleus in the epithelium region, which is another important feature of cervical cancer in addition to the gradual increase in the number of atypical cells [5]. Therefore, we propose a method to calculate the nucleus density in each region, utilizing segmentation grid-lines generated based on the SLIC algorithm which can combine information on the morphology, size, number and spatial distribution of nuclei.

Our method is inspired by how pathologists observe histopathological images to make diagnoses. In order to simulate pathologists' behaviour, we adopt the idea of a deep learning framework for extracting features and identifying diseases. Our framework consists of three stages. The first stage contains the U-Net structure-like neural network to segment nucleus from RGB images [6]. The labels are obtained by the threshold-based segmentation method and then processed by image denoising. The second stage is the extraction of NDDF maps using the SLIC algorithm. The third stage is composed of a two-stream network trained with the NDDF map and the RGB image. Pathologists classify the histopathological images of the epithelium region into well, moderately, and poorly differentiated stages, and then use the diagnosis as a label for the deep learning model. For the task of training on a small percentage of samples and testing on a large percentage of samples, training the two-stream neural network model over both RGB images and NDDF maps achieves a higher accuracy with little impact on the speed of classification.

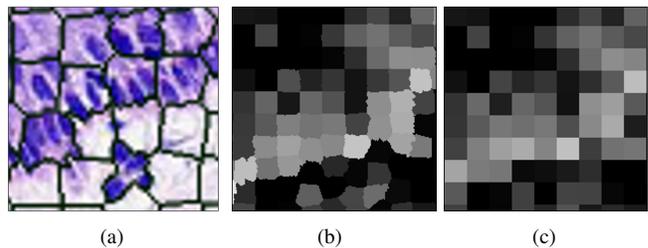
## 2. METHODOLOGY

The current mainstream approach is to train deep learning models directly on RGB images. But these works could learn invalid information from RGB images, and they require larger data sets or more complex neural networks to achieve better results [7]. To tackle this issue, we propose the nuclear density distribution feature (NDDF) to further improve the performance of the deep learning model. In this section, we will describe the process of segmenting the nucleus mask from the pathology image, and then extracting the NDDF from mask maps utilizing the superpixel algorithm, finally training the deep learning model on NDDF maps. The overall framework is shown in Fig. 1.

### 2.1. The nucleus segmentation

Histopathology images are labeled by the pathologists and divided into three types of data at a magnification of  $20\times$ . No

additional labeling is done by pathologists except for category labeling. In our experiments, the size of each patch is  $512 \times 512$ . Since we end up with a  $32 \times 32$  sized feature map, as shown in Fig. 1C, the nucleus segmentation performance has little effect on the final classification results. U-Net could perform well in the field of medical image segmentation, learning a robust edge extraction model with a small amount of data [8]. We train the U-Net structure-like neural network model ANet (Fig. 1A) with the publicly available datasets MoNuSeg [9] and TNBC [10] to segment nuclei masks on histopathological images [11]. The right half of the ANet contains five blocks, all of which are up-sampled. The feature maps from each up-sampling are concatenated with the feature maps output from the corresponding layer on the left. The concatenated feature map is then fed into the next decode block [12]. Nucleus masks of the same size as the input image are finally detected.



**Fig. 2.** (a) Superpixel grid-lines. (b) NDDF based on the SLIC. (c) NDDF based on the square grids.

### 2.2. Extraction of density distribution feature

After acquiring nucleus masks, the next step is to generate the NDDF matrix, which is implemented in our approach utilizing the simple linear iterative clustering (SLIC) algorithm, as shown in Fig. 1B. In the CIELAB color space, each pixel is considered as a five-dimensional vector  $(a, b, c, x, y)$ , where the first three dimensions contain LAB color space information and the last two dimensions contain pixel position information. The SLIC algorithm first converts the RGB image to CIELAB color space, then initializes  $K$  seed pixels on the image, and calculates the superpixel distance  $S = \sqrt{(N/K)}$  ( $N$  is the pixel number). Next, under the metric distance, center each seed pixel to find the center of the nearest superpixel region within the surrounding  $2S \times 2S$  range [13] and determine which superpixel each pixel belongs to, with the following details:

$$\begin{aligned}
 d_c &= \sqrt{(a_j - a_i)^2 + (b_j - b_i)^2 + (c_j - c_i)^2} \\
 d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\
 D &= \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}
 \end{aligned} \tag{1}$$

Where  $i$  and  $j$  denote two pixels for distance comparison, and  $d_c$  and  $d_s$  are the color distance and pixel distance, respectively.  $N_s = S = \sqrt{(N/K)}$ , and  $N_c$  is directly defined as a constant 5, which is used to standardize  $d_s$  and  $d_c$ .

After determining the pixel class based on the calculated metric distance, the five-dimensional average of each category of pixels is taken as the center of the new superpixel and is used to calculate the residuals between the current epoch of classification and the previous epoch. Finally, the above process is repeated continuously until the residuals converge to zero. From this, we can obtain a 32x32 matrix on the pathological image by setting the value of  $K$  to 1024. In order to calculate the nuclear density  $T_i$  within each superpixel block, we combine the seg-grid-lines (superpixel segmentation grid-lines) generated by SLIC with the nucleus mask (Fig. 1C), as follows:

$$T_i = \frac{\sum_{j=1}^N n_{ij}}{\sum_{j=1}^M s_{ij}} \quad i \in K \quad (2)$$

Where  $n_i$  is the area of the nucleus in each region, which is the number of white pixels, and  $s_i$  is the area of each region, which is the number of region pixels.

The brightness value of each superpixel block in the feature map corresponds to the density value of nuclei in each region, as shown in Fig. 2. SLIC generates the superpixels along the nucleus boundaries, and the clustered nuclei are contained intactly in the same superpixel. Comparing Fig. 2b and Fig. 2c, in the superpixel-processed digital histopathology image, the basal edge is clear and intact, while the square grid method of directly dividing the image area disrupts the boundary both of the nucleus and the basal area in the local area, making it difficult to discern the direction of nucleus distribution[14].

### 2.3. The NDDF classification model

The number, location and size of nuclei in epithelium region are meaningful indicators for pathologists to diagnose cervical cancer. Therefore, the method proposed by us extracts the NDDF in patches of epithelial histopathology images and then feeds them into a deep learning model for training. We design a model DNet (Fig. 1D) containing a 5-layer convolutional neural network for training utilizing NDDF maps. Last layer is a fully connected layer. The experimental results show that training DNet classification utilizing a small amount of NDDF maps is fast and accurate. However, only the information in the NDDF is not enough; color and texture information are also important, so we also leverage the histopathological images to train the neural network model. In order to verify the validity of the density distribution feature for cervical cancer diagnosis, we select models of different complexity as the baseline for training RGB images, including DNet, VggNet16 [15], ResNet50 [16], and SeNeXt50 [17, 18]. We then fuse the baseline and DNet models into the

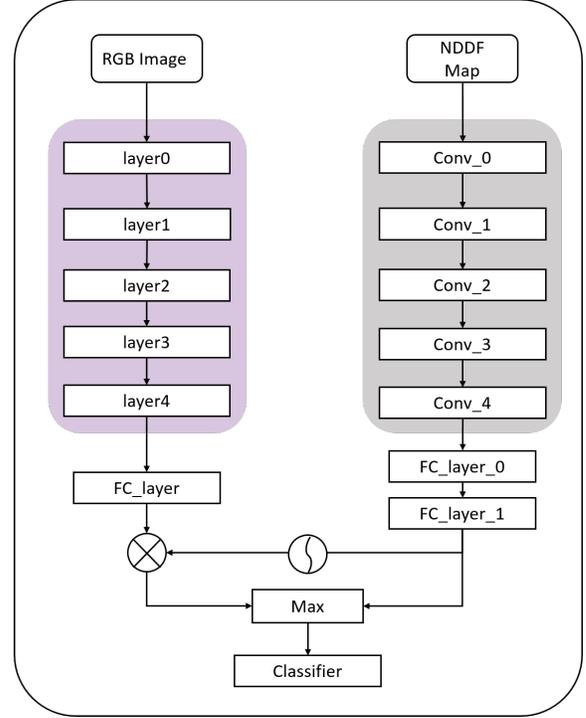


Fig. 3. The structure of the two-stream network

two-stream network, inputting RGB images and NDDF maps for training, respectively. As shown in Fig. 3, in the structure of the two-stream network, the outputs of baseline and DNet are fused together. In addition, a new FC layer is added to the classifier layer of DNet before. The FC layer is combined with the Sigmoid function to produce a weight vector, which is then dot-multiplied by the baseline output features. And then we concatenate the dot-multiply result with the output features of DNet. Finally, the results of the max function output are classified as follows:

$$Predict = \max\{\sigma(out_D) \cdot out_B, out_D\} \quad (3)$$

where  $out_D$  is the output features of the DNet,  $out_B$  is the output features of the baseline model, and  $\sigma(\cdot)$  is the Sigmoid activation function. The maximum value of each channel feature is obtained by the max operation, the fusion feature  $Predict$  is input into the classification layer.

Table 1. The performance of DNet for classification of the RGB image, the Grid map and the NDDF map respectively

Input	Model	Recall(%)	Precision(%)	F1-score(%)	Accuracy(%)
RGB image	DNet	83.70	77.27	80.30	76.79
Grid map	DNet	77.90	76.37	77.14	73.81
NDDF map	DNet	<b>88.79</b>	<b>83.95</b>	<b>86.30</b>	<b>84.04</b>

**Table 2.** The performance of four baselines with their corresponding two-stream networks

Model	Dataset	FPS	2-class				3-class			
			Recall(%)	Precision(%)	F1-score(%)	Accuracy(%)	Recall(%)	Precision(%)	F1-score(%)	Accuracy(%)
DNet	RGB	196	83.70	77.27	80.36	76.79	60.34	61.82	57.63	70.71
VggNet16	RGB	68	72.86	76.81	72.66	79.16	70.47	75.00	71.33	75.37
ResNet50	RGB	113	82.68	89.36	85.89	84.58	77.05	78.25	77.42	82.12
SeNeXt50	RGB	63	90.32	92.20	91.25	90.19	79.09	82.85	80.00	84.90
DNet+DNet	RGB+NDDF	179	84.55	90.22	87.29	86.06	69.98	74.73	70.11	78.03
VggNet16+DNet	RGB+NDDF	64	88.04	87.14	87.52	88.26	77.44	78.29	77.68	81.43
ResNet50+DNet	RGB+NDDF	109	<b>93.20</b>	88.83	90.97	89.52	80.05	81.85	80.69	84.80
SeNeXt50+DNet	RGB+NDDF	59	92.53	<b>93.32</b>	<b>92.92</b>	<b>92.01</b>	<b>81.90</b>	<b>84.03</b>	<b>82.60</b>	<b>86.06</b>

The dot product operation weights the output features of the baseline as an attention mechanism leveraging the DNet implementation to improve the performance of the baseline model [19]. The max operation can combine the valid information learned from the two networks to improve the classification performance. It was found that the two-stream model significantly improves the performance compared to the corresponding baseline, while the speed barely decreases for inferring.

### 3. EXPERIMENT

The dataset used for the experiment is a dataset of cervical histopathology images collected from a collaborating hospital. It contains multiple data sources, all of which are annotated by pathologists. The experimental data consists of patches containing varying proportions of cervical epithelium intercepted from 84 digitized histopathology images, which are labeled by the experts, of which 34 are well differentiation stage, 25 are moderately differentiation stage, and 25 are poorly differentiation stage. The size of the original image is  $512 \times 512$ , and the size of the NDDF map is  $32 \times 32$ . All evaluations and comparisons reported in this section are done on the test set.

In our experiments, we train the DNet model separately utilizing three types of data: digital histopathological images, nucleus density distribution maps generated with square grid, and SLIC-based NDDF maps. We randomly split the dataset, by patients, into training, and test sets, with ratio 2:8. Samples are randomly selected from each category to obtain the same number of samples from each class for training and testing. We set the same hyperparameter during model training to ensure that the variation is due only to the density distribution feature. In order to provide a comprehensive comparison, we report the quantitative evaluation scores obtained by training the DNet model with each data, including the accuracy, recall, precision and F1 scores.

The density distribution feature is extracted based on the prior knowledge of pathologists in diagnosing cervical cancer on pathological images and the SLIC algorithm is able

to take good account of morphological and size features of the nuclei. As shown in Table 1, the NDDF maps performed the best among the three data results. This indicates that the information contained in the NDDF is discriminative. We achieve an accuracy of up to 84.04% by training a 6-layer neural network leveraging NDDF maps. To verify the robustness of this method, we utilize four different baselines, including DNet, VggNet16, ResNet50 and SeNeXt50, with the same hyperparameters, to do the 2-classification task and the 3-classification task, respectively, and the final performance is shown in Table 2.

As can be seen in Table 2, the NDDF extracted brings clear benefits to all the metrics evaluated, and our two-stream model achieves better results in all metrics. More importantly, the effect of our proposed method of two-stream networks is robust and its classification performance is improved under different baselines. This indicates that the two-stream model can learn valid and complementary information from RGB images and NDDF maps to improve the classification performance of histopathology images. Also from Table 2, it can be seen that the classification speed of the two-stream network using RGB images and NDDF maps did not decrease significantly. The NDDF based on the superpixel algorithm is robust for the classification of cervical pathology images.

### 4. CONCLUSION

In this paper, we present an approach to extract the robust NDDF based on U-Net and SLIC algorithm for improving the performance of digital pathology image classification. Due to the complexity and the heterogeneity of digital histopathological images, the previous works are not valid to learn the discriminative features and the prior knowledge of pathologists in diagnosing cervical cancer is discovered to promote the model performance. Therefore, both RGB images and NDDF maps are as the inputs to train the two-stream network for histopathological images. Comprehensive experiments demonstrate that the classification accuracy of cervical tissue pathology images are improved by the proposed approach comparing with several methods.

## 5. REFERENCES

- [1] Y. Xue, Q. Zhou, J. Ye, L. R. Long, S. Antani, C. Cornwell, Z. Xue, and X. Huang, "Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 387–396.
- [2] F. Sheikhzadeh, R. K. Ward, A. Carraro, Z. Chen, D. van Niekerk, C. MacAulay, M. Follen, P. Lane, and M. Guillaud, "Confocal fluorescence microscopy for detection of cervical preneoplastic lesions," in *Medical Imaging 2015: Digital Pathology*, vol. 9420. International Society for Optics and Photonics, 2015, p. 942009.
- [3] A. J. Schaumberg, W. Juarez, S. J. Choudhury, L. G. Pastroján, B. S. Pritt, M. P. Pozuelo, R. S. Sánchez, K. Ho, N. Zahra, B. D. Sener *et al.*, "Large-scale annotation of histopathology images from social media," *BioRxiv*, p. 396663, 2018.
- [4] C. Li, Z. Hu, H. Chen, D. Xue, N. Xu, Y. Zhang, X. Li, Q. Wang, and H. Ma, "Cervical histopathology image clustering using graph based unsupervised learning," in *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*. Springer, 2020, pp. 141–152.
- [5] C. Li, H. Chen, X. Li, N. Xu, Z. Hu, D. Xue, S. Qi, H. Ma, L. Zhang, and H. Sun, "A review for cervical histopathology image analysis using machine vision approaches," *Artificial Intelligence Review*, pp. 1–42, 2020.
- [6] X. Xie, J. Chen, Y. Li, L. Shen, K. Ma, and Y. Zheng, "Instance-aware self-supervised learning for nuclei segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 341–350.
- [7] H. Le, D. Samaras, T. Kurc, R. Gupta, K. Shroyer, and J. Saltz, "Pancreatic cancer detection in whole slide images using noisy label annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 541–549.
- [8] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.
- [9] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.
- [10] P. Naylor, M. Laé, F. Reyat, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.
- [11] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," *arXiv preprint arXiv:1801.05746*, 2018.
- [12] Y. Tu, L. Niu, J. Chen, D. Cheng, and L. Zhang, "Learning from web data with memory module," 2019, pp. arXiv–1906.
- [13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," vol. 40, no. 4. IEEE, 2017, pp. 834–848.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [18] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.
- [19] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.