ORIGINAL ARTICLE

A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images

Yuxin Cui[1] · Guiying Zhang[2] · Zhonghao Liu[1] · Zheng Xiong[1] · Jianjun Hu[1] (ORCID)

Abstract

This paper addresses the task of nuclei segmentation in high-resolution histopathology images. We propose an automatic end-to-end deep neural network algorithm for segmentation of individual nuclei. A nucleus-boundary model is introduced to predict nuclei and their boundaries simultaneously using a fully convolutional neural network. Given a color-normalized image, the model directly outputs an estimated nuclei map and a boundary map. A simple, fast, and parameter-free post-processing procedure is performed on the estimated nuclei map to produce the final segmented nuclei. An overlapped patch extraction and assembling method is also designed for seamless prediction of nuclei in large whole-slide images. We also show the effectiveness of data augmentation methods for nuclei segmentation task. Our experiments showed our method outperforms prior state-of-the-art methods. Moreover, it is efficient that one 1000×1000 image can be segmented in less than 5 s. This makes it possible to precisely segment the whole-slide image in acceptable time. The source code is available at https://github.com/easycui/nuclei_segmentation.

Keywords Deep learning · Nuclei segmentation · Fully convolutional neural network · Data augmentation

1 Introduction

With the progress of image processing and pattern recognition techniques, computer-assisted diagnosis (CAD) has been widely utilized to assist medical professionals in interpreting medical images. Digital pathology is earning more and more attention from both image analysis researchers and pathologists due to the advent of whole-slide imaging. The potential applications of digital pathology span a wide range such as segmentation of desired regions or objects, counting normal or cancer cells, recognizing tissue structures, classifying cancer grades, and prognosis of cancers [5, 33].

Yuxin Cui and Guiying Zhang are equally contributed.

✉ Jianjun Hu
jianjunh@cse.sc.edu

Yuxin Cui
ycui@email.sc.edu

[1] Department of Computer Science and Technology, University of South Carolina, Columbia, SC 29208, USA

[2] Department of Medical Information Engineering, Zunyi Medical University, Zunyi, China

As an essential part of digital pathology, histopathology image analysis is playing an increasingly important role in cancer diagnosis, which can provide direct and reliable evidence to diagnose the grade and type of cancer. This paper deals with nuclei segmentation, an important step in histopathology image analysis. The purpose of nuclei segmentation is not only counting the number of nuclei but also obtaining the detailed information of each nucleus. Hence, we can exactly extract each nucleus from the image and make it available for further analysis. For example, the features of the individual nucleus and the distribution of nuclei clusters can be used to grade and classify status of breast cancers [2, 19]. Because of appearance variations such as color, shape, and texture, nuclei segmentation from histopathology images could be very challenging, as illustrated in Fig. 1, in which it is very difficult even for humans to recognize and segment all nuclei within the images. Figure 1a and b illustrate two histopathology images from different organs. Figure 1c and d are two histopathology images from the the same organ (breast) but have different cancer grades.

Current deep learning methods for nuclei segmentation usually need a complex post-processing procedure to obtain the final nuclei boundaries [15, 20, 35]. Here, we proposed an end-to-end approach for nuclei segmentation based on

Fig. 1 **a** Colon cancer. **b** Prostate cancer. **c** Breast cancer (grade I). **d** Breast cancer (grade III)

(a)   (b)

(c)   (d)
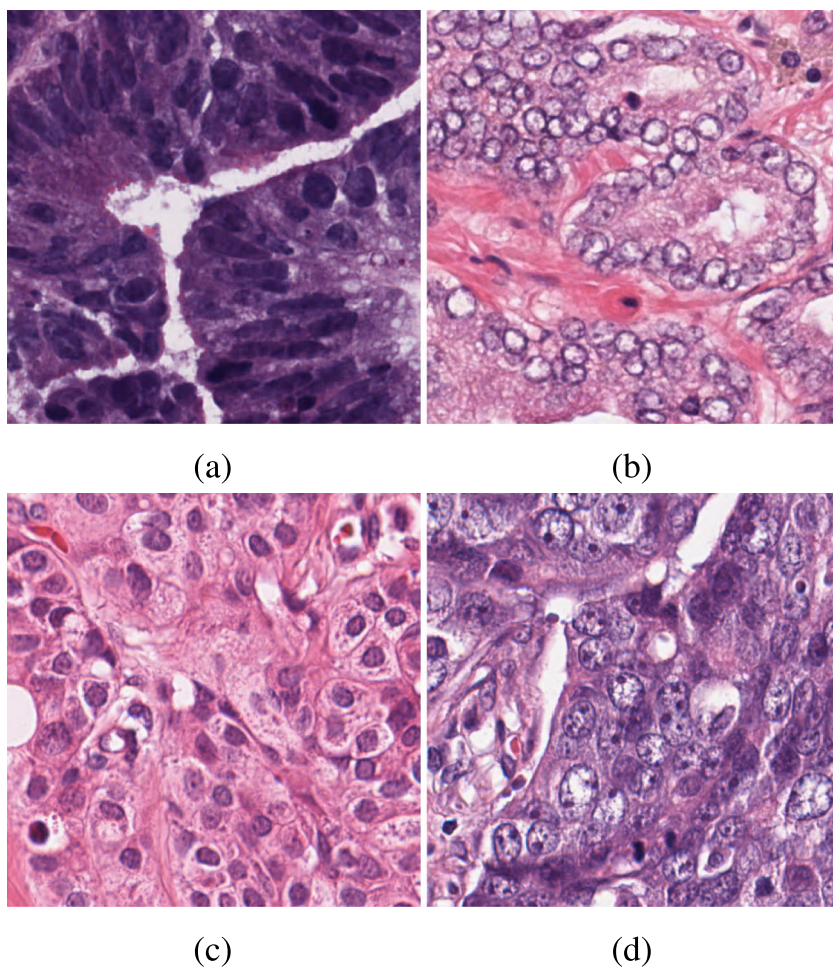
U-net [26]. Unlike prior binary classifiers [17, 29, 36], which only discriminate nuclei against the background, our nuclei-boundary segmentation model predicts the nuclei and their contours at the same time. Due to the accurate prediction of nucleus and boundary in our approach, the final segmentation can be generated by a simple and fast post-processing procedure. To segment the whole-slide image, a pixel-wise segmentation strategy is necessary. However, the border area of each patch cannot be predicted accurately because of a lack of contextual information. A seamless patch extraction and assembling method is proposed to handle this problem. The main contributions of this paper are as follows:

– We propose a nuclei-boundary model to explicitly detect nuclei and their boundaries simultaneously from histopathology images. Detecting boundary is able to improve the accuracy of nuclei detection and help split the touched and overlapped nuclei. Given the raw segmentation results by our nuclei-boundary model, only a simple dilation operation and noise-removing steps are needed to produce the final segmentation results.

– We develop an effective approach to segment extra-large high-resolution images that U-net cannot handle due to limited GPU memory using a seamless patch-wise segmentation. A weighted loss map is utilized to train the model and a vote mechanism is used to assemble the patches.

– Extensive studies on the effects of a variety of data augmentation methods for nuclei segmentation are provided.

– We introduce four evaluation criteria for more accurate nuclei segmentation performance evaluation: missing detection rate, false detection rate, under-segmentation rate, and over-segmentation rate. They are designed to help the pathologist obtain more in-depth understanding of the performance of automatic segmentation methods and choose the right one for their specific application.

2 Related work

Nuclei segmentation methods can be largely divided into two categories: unsupervised or supervised approaches.

Among unsupervised methods, the most popular method to detect nuclei is intensity thresholding such as Otsu's method [22]. Another popular approach for nuclei detection is clustering including K-mean clustering [4], graph cut–based methods [27], etc. Furthermore, a few filtering-based methods have been proposed that utilize the features of the nuclei, such as Laplacian-of-Gaussian (LoG) filters [1] and fast radial symmetry transformation [32]. The above unsupervised methods have one common weakness: they are only effective for one or a few specific types of nuclei or images, since the appearances of nuclei are so diverse that one can hardly develop a single model suitable for all these different images. In recent years, supervised learning-based approaches are becoming more and more attractive including multilayer neural networks [17], stacked sparse autoencoder [36], and spatially constrained convolutional neural networks (CNNs) [29]. In these methods, each pixel of the image is usually classified into one of two categories: nuclei or background. After the nuclei area or the nucleus seed is predicted by the nuclei detection stage, the next step would be obtaining the contours of all nuclei. If the nuclei area is predicted in the nuclei detection stage, this could be achieved by methods such as bottleneck detection [14] and ellipse fitting [9, 30]. If the seed of a nucleus is generated, its contour could be obtained by using marker-controlled watershed [24, 32] or region growing [35].

Deep learning–based methods are becoming increasingly popular in image segmentation due to their dominating performance in many tasks of computer vision such as object classification, object detection, and image segmentation. Since 2014, numerous convolutional neural network–based image segmentation methods have been proposed. Long et al. proposed the fully convolutional neural network (FCN) [15] for semantic segmentation. Compared with prior models, it is demonstrated that the FCN algorithm is much more efficient and accurate. Converting fully connected layers into convolutional neural networks makes it possible to predict the heatmap of the objects in the image that needs to be segmented, which unifies the detection and segmentation steps in traditional approaches. The skip architecture of FCN as first introduced in residual networks [7] helps boost its performance by fusing different levels of semantic information.

A major progress in biomedical segmentation was made by U-net [26], an FCN-based network architecture proposed in 2015, which won the Grand Challenge for Computer-Automated Detection of Caries in Bitewing Radiography at ISBI 2015. Naylor [20] employs FCN to discriminate nuclei from the background and then applies the watershed method to split the nuclei. However, the resulting boundaries are not accurate. Xing [35] proposed a sophisticated shape deformation method to generate each nucleus's boundary. Kumar [12] designed a CNN3 model based on a CNN network to detect the nuclei from the image and a region growing method to obtain the contours. But both of these have high running time complexity.

3 Method

3.1 Overview

Our nuclei segmentation method adopts an end-to-end deep learning framework. As shown in Fig. 2, the procedures to segment nuclei from H&E stain normalized images are as follows. First, the image is processed by H&E stain normalization. In the training phase, we randomly extract thousands of patches (samples) from training images.

Fig. 2 The overview of our method
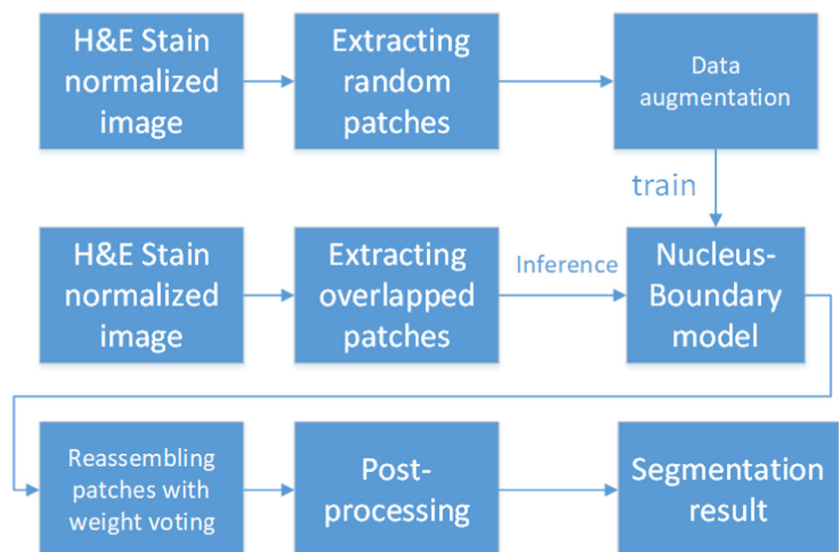
During the training, each minibatch is processed by data augmentation before it is fed into the deep neural network to train the nucleus-boundary model. During the testing phase, We extract overlapped patches from test images based on sliding windows. The prediction result of these overlapped patches yielded by the nucleus-boundary detector shows inside nuclei area and the boundaries. At last, the area of each nucleus is obtained via a simple, fast, and parameterless post-processing procedure.

3.2 Data preprocessing

H&E stain is the most widely used stain protocol in medical diagnosis. Typically, the nuclei of cells are stained to blue by hematoxylin while the cytoplasm is colored to pink by eosin. However, in practice, the colors of H&E-stained images could vary a lot as shown in Fig. 1 due to variation in the H&E reagents, the staining process, the scanner, and the specialist who performs the staining. A few H&E stain normalization methods [8, 16, 31] have been proposed to eliminate the negative interference caused by color variation. We tried two of them [16, 31] to normalize the raw H&E-stained images. However, we did not find any considerable difference between these two normalization methods in terms of prediction performance of our segmentation algorithm. In particular, the result shown in the experiment Section 4 was generated based on the images normalized by the method proposed in [31]. Given a target image, this method is able to convert a source image's color into the target image's color space based on sparse non-negative matrix factorization (SNMF) [31]. Compared with the nonnegative matrix factorization (NMF) [13], a technique that has been used for stain separation [25], SNMF introduces L1 sparseness regularization to preserve the biological structure. We chose one well-stained H&E image as the target and convert other images into its color space. The hyperparameter λ, which controls the trade-off between sparseness and reconstruction accuracy, is set to 0.1 according to [12].

Intuitively, it would be much easier to distinguish the foreground (nuclei) from the background (cytoplasm) in a pure hematoxylin-channel grayscale image compared with a RGB image. A large number of nuclei segmentation methods [3, 24, 34] employ some deconvolution algorithms to extract the H-channel from H&E-stained images for better segmentation performance. However, based on our experiments, we noticed that our deep fully convolutional neural network works better in extracting the nuclei from raw RGB images than from H-channel grayscale images. A visual comparison between H-channel image and original RGB image is shown in Fig. 3. The reason would be that the H-channel misses some information that might be helpful for distinguishing nuclei from the cytoplasm. Given well-labeled training images, our deep neural networks can then learn the optimal way to extract the features that discriminate samples of different categories. Based on the above analysis, we skipped the step of H-channel extraction and directly took the RGB color images as the input to our deep neural networks.

3.3 Nucleus-boundary model

Traditional supervised nuclei segmentation methods usually apply a binary classifier to segment the nuclei areas by classifying all pixels into nuclei or background type. These methods usually predict the category of the central pixel given a small patch. To segment the whole image, it needs to extract all the sliding windows (patches) with a stride of 1 pixel and predict the central pixel category for each of these patches. A major limitation of this procedure is its high computational complexity. Given an image of size 1000×1000, this method needs to process one million sliding windows in order to segment it. To make it worse, a typical whole-slide histopathology image may have billions

Fig. 3 A comparison between H-channel and RGB images. **a** An original histopathology image; **b** corresponding H-channel image

(a)　　　　　　　　　　　(b)
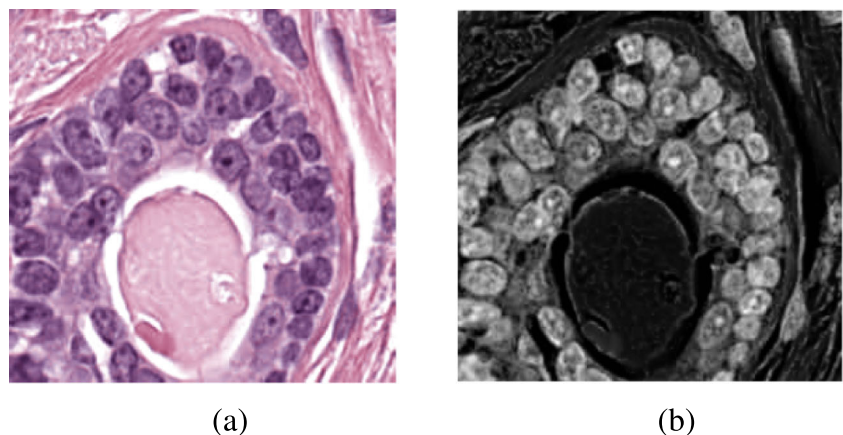
of pixels, making it impossible to process it in an acceptable time using this strategy. Instead, our method is based on a fully convolutional network (FCN) framework, which allows predicting the category of all the pixels of an image with only one pass. The input of the network is one image; the output is the estimated class map.

The task of nuclei segmentation can be roughly divided into two stages: the first stage is extracting the foreground (nuclei); the second stage is segmenting the connected foreground area into separated nuclei and finding out the boundary of each nucleus. Our method intends to merge these two steps by extracting the nuclei and their boundaries at the same time. So, it is named "nuclei-boundary (NB) model." As shown in Fig. 4, the output of the NB model has three channels, and each has the same height and width with the input image. Its values represent the probabilities of each pixel being *background*, *boundary*, or *inside* class, respectively. The manual annotation for our segmentation problem is the boundary of each nucleus. A pixel belonging to the *boundary* class means that it is on or inside an annotated boundary and within 2 pixels from the boundary. Pixels of the *inside* class are those that are inside the annotated boundary but are not *boundary* pixels. Correspondingly, the output can be regarded as an RGB image and the estimated maps of the *background*, *boundaries*, and *nuclei* are represented by red, green, and blue, respectively, as shown in Fig. 4. To generate the ternary mask for training, we apply a morphology operator to each nucleus to obtain the *inside* pixels, and then subtract *inside* pixels from the nucleus to get *boundary* pixels.

### 3.3.1 The architecture of our NB network

Figure 4 shows the network architecture of our algorithm, which consists of a couple of encoding and decoding layers. The encoding layers are used to extract different levels of contextual feature maps. The decoding layers are designed to combine these feature maps produced by the encoding layers to generate the desired segmentation maps. Due to the memory limitation of our GPU, the size of the input layer is set to 128×128 in our experiments. The weight of each convolutional layer is initialized by glorot uniform [6] and bias is initialized to 0. The glorot uniform is defined as:

$$W \sim U \left[\frac{-\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \qquad (1)$$

where W means the initialized weight; n_j means the size of the convolutional layer j.

The scaled exponential linear units (SELUs) [18] activation function is used in all convolutional layers. SELUs is designed to make the forward neural network (FNN) have self-normalizing capability [11]. The FNN using SELUs is shown to be able to outperform the ones using explicit normalization methods, such as batch normalization, layer normalization, and weight normalization. Hence, our network does not have any normalization layers.

The padding property of each convolutional layer is the "same" in order to ensure it keeps the same size with its previous layer. The size of all convolutional filters is 3×3. Each convolutional layer is followed by a dropout layer with 0.2 drop rate. The network is trained by Adam optimizer
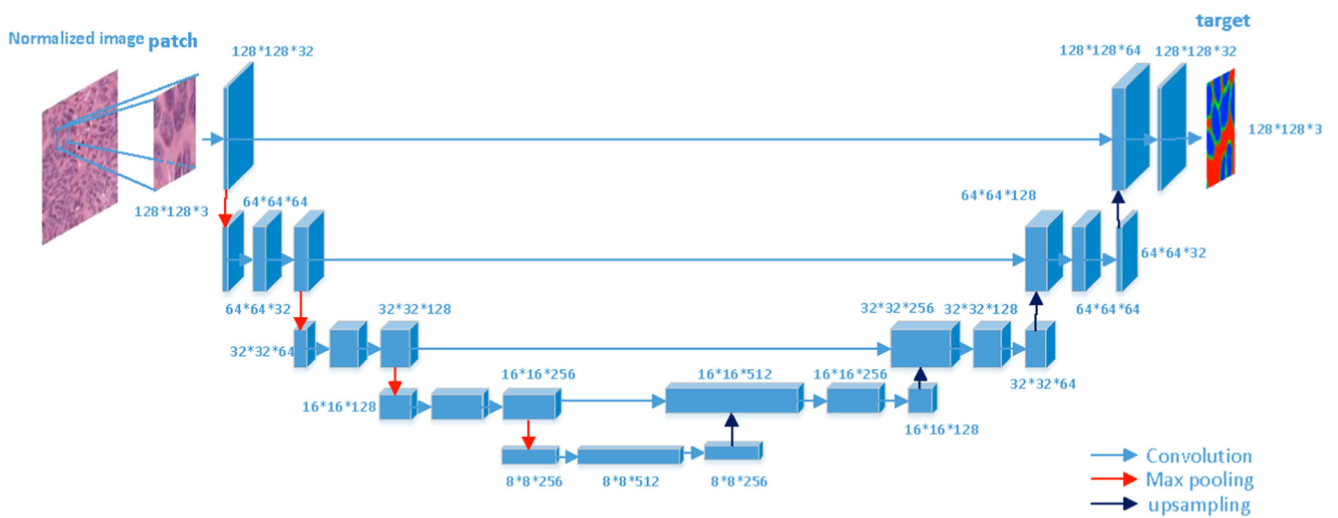
Fig. 4 The structure of our network. The size of each layer is shown in $height \times width \times channels$. The height and width of each layer are not fixed, which are determined by the size of input images. Here, we assume the input image is of size 128×128

[10]. This stochastic optimization method is able to compute adaptive learning rate for each parameter. It automatically controls the learning rate along the training, so it is not necessary to manually set the momentum and decay.
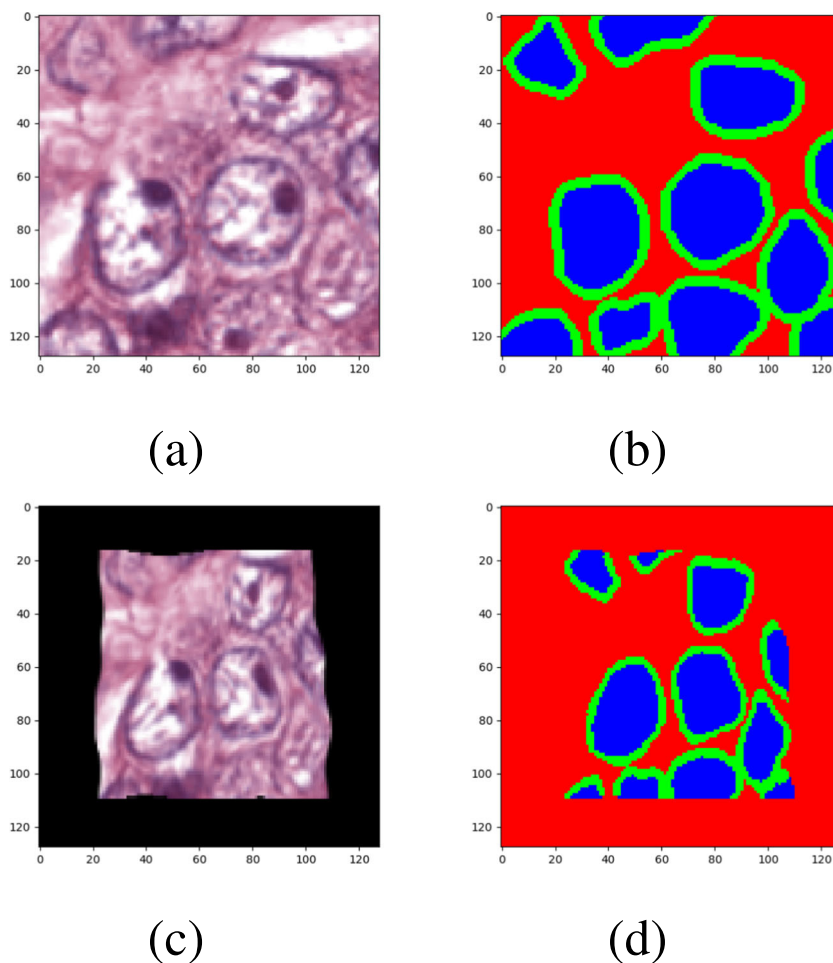
3.3.2 Data augmentation

Deep learning models often have millions of parameters so that it needs a large-scale sample dataset to avoid the overfitting problem. In fact, the datasets of our nuclei segmentation task often contain only tens of images. Moreover, labeling an 1000×1000 image which contains hundreds of nuclei usually costs a specialist at least 5 h. Hence, it is impossible to manually label sufficient nuclei boundaries accurately for training deep learning models. Data augmentation is an essential approach to overcome the overfitting problem caused by a lack of samples. The training samples, i.e., the patches, are randomly extracted from the H&E-stained images in the training datasets. Five augmentation techniques are used together in our

method including random elastic transformation, rescale, affine transformation, shift, flip, and rotate. Each training sample (one patch extracted from a whole image) and the corresponding target are processed by the data augmentation procedure. Given a training sample, which is a RGB image I with its corresponding ground truth I_{gt}, we transform I to I' and I_{gt} to I'_{gt}. I' and I'_{gt} are the input and target of the neural network. The rescaling factors are set as a random number between 0.5 and 1.5. We employ Simard's method [28] to do elastic transforming. Two hyperparameters α and σ need to be manually set to control how dramatic the original image is transformed. In our experiment, α is set to a random number between 100 and 200, σ is set to 12.

Besides transforming the input sample, it is necessary to do the same transformation on targets to maintain consistency. The one-hot encoding target consists of only binary values. However, the transformed target has some float-point numbers caused by bilinear interpolation we used for data augmentation. They need to be binarized by the following rules:

Fig. 5 Example of data augmentation: **a** one patch extracted from a normalized image; **b** corresponding ground truth of (**a**). **c** A training sample generated by data augmentation procedure based on patch (**a**). **d** The corresponding ground truth of (**c**)

(a)

(b)

(c)

(d)

Let the value of one pixel be (t_i, t_b, t_o), where t_i, t_b, and t_o represent the labels for *inside*, *boundary*, and *background*, respectively.

1. If $t_b > 0.5$, $t_b = 1$, else $t_b = 0$
2. If $t_i > 0$ and $t_b == 0$, $t_i = 1$, else $t_i = 0$
3. If $t_i == 1$ or $t_b == 1$, $t_o = 0$, else $t_o = 1$

An example of data augmentation is illustrated in Fig. 5.

3.3.3 Weighted loss

The U-net [26] model tends to predict the pixels with full context in the input image, which leads to generation of a smaller segmentation map than the input image. The border area of the input image is not predicted because of a lack of enough context information. This strategy can solve the problem that the prediction of the border area is not accurate to some extent. One issue of this is that this U-net defines a border area whose size is immutable without modifying the network structure. However, in practice, the border area size could vary in different histopathology images and it mainly depends on the size of the nuclei. Another limitation is that we have to do some cropping operation in neural network training to make the size of layers match each other, which might lose useful surrounding information.

As a trade-off of these issues, we designed a weighted loss and a scheme for patch extraction and assembling to allow the neural network to predict a segmentation map of equal size without concerning the lack of context issue in the border area.

The model is trained by minimizing the categorical softmax cross-entropy loss between predictions and targets, which is described in Eq. 2:

$$L = \sum_i \sum_j W_{i,j} log(p_{t(i,j)}(i, j)) \tag{2}$$

where $t(i, j)$ denotes the true label of the pixel at (i, j) position; $p_{t(i,j)}(i, j)$ is the output of soft-max activation layer which indicates the probability of the pixel at (i, j) being $t(i, j)$. W is the proposed weight map, which is defined as:

$$W_{i,j} = \alpha \frac{D_{i,j}^e}{(D_{i,j}^c + D_{i,j}^e)}$$
$$\alpha = \frac{h \cdot w}{\sum_{i=1}^{h} \sum_{j=1}^{w} \frac{D_{i,j}^e}{D_{i,j}^c + D_{i,j}^e}} \tag{3}$$

where $W_{i,j}$ is the weight of position i, j; $D_{i,j}^e$ is the distance from border; $D_{i,j}^c$ denotes the distance from center. h and w are the height and width of the map, respectively (Fig. 6).
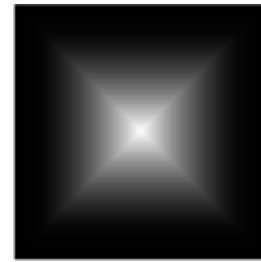

Fig. 6 The weighted loss map generated by Eq. 3

3.3.4 Extra-large image segmentation using overlapped patch extraction and assembling

Current medical image segmentation algorithms based on U-net and its derivatives have an unsolved problem for segmenting extra-large high-resolution histopathology images: due to the limited memory of the GPU, it is possible to feed the whole-slide image into the deep neural network. It has to be cut into patches and perform patch-wise training and prediction. However, there is no reported solution to deal with this issue.

With close examination, we found the main issue of U-net algorithm on patch-based segmentation is that the prediction at the border area is not accurate as demonstrated in Fig. 12. Here, we propose an overlapped patch extraction and assembling method. The patches are extracted by sliding window with a stride. For assembling, a vote mechanism is applied to predict each pixel using:

$$P(i, j) = \frac{\sum_k W_{k(i,j)} p(k(i, j))}{\sum_k W_{k(i,j)}}$$

where $P(i, j)$ is the final prediction of the pixel at position (i, j) in an image. $k(i, j)$ means the position of it in the kth patch.

3.3.5 Post-processing

From Fig. 7, we can see that the raw prediction results already show clear *inside* nucleus areas and boundaries. Due to this reliable prediction results, we no longer need the complex region-growing algorithms [12, 35] and splitting algorithms [34] to extract the final segmented areas. These methods usually strongly rely on manual parameter tuning to get good performance and are computationally demanding. Instead, we use a parameter-free postprocessing procedure that runs in a negligibly short time. Since our NB model detects both *inside* and *boundary* classes, all we need is the *inside* class map. Then, the *inside* class map is transformed to a binary map using a constant threshold 0.5. In this way, each connected component in the binary image indicates the *inside* area of one nucleus. At the end, in order to recover the shape, based on the way *inside*
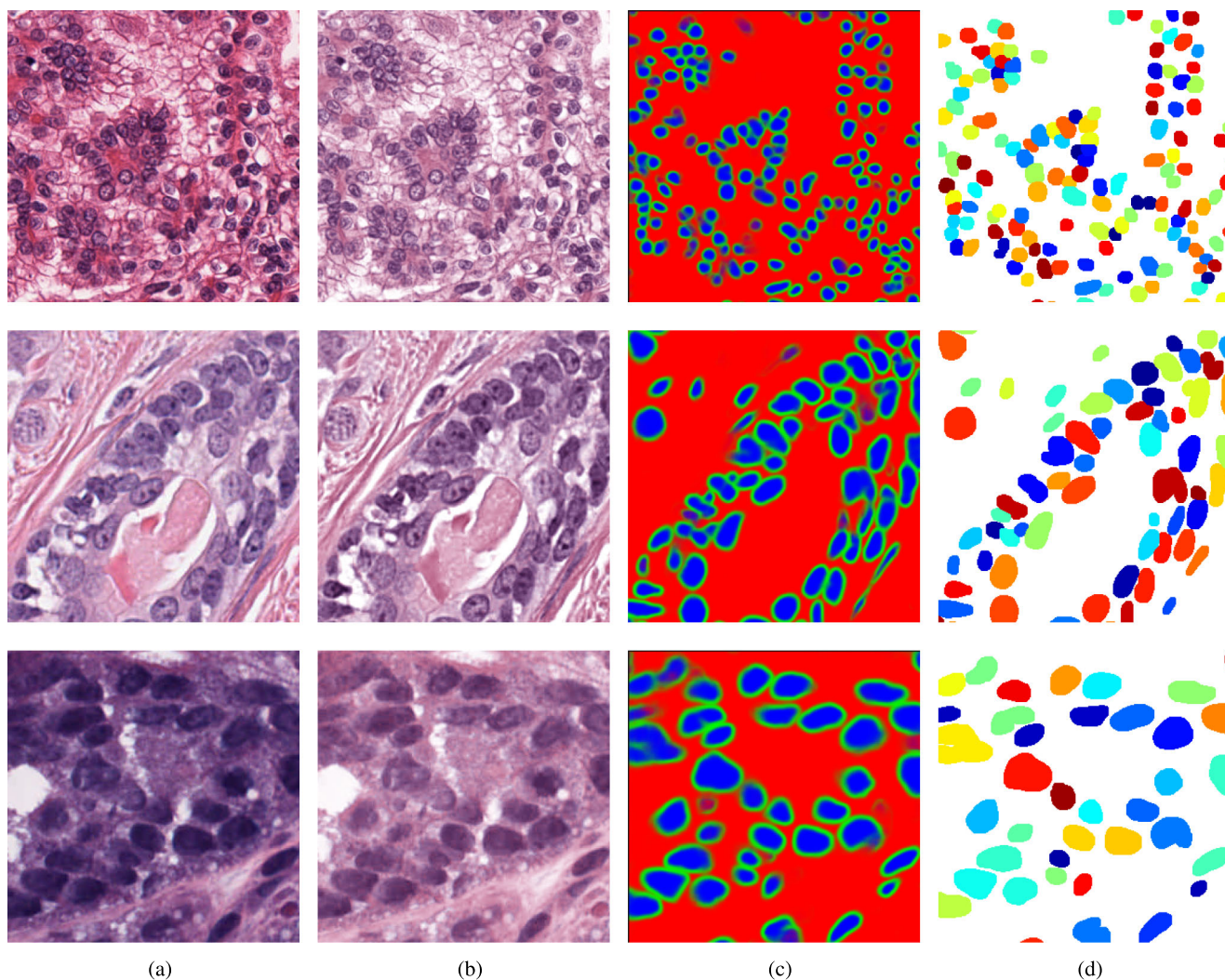
Fig. 7 **a** Examples of original histopathology images; **b** corresponding images after color normalization. **c** Raw segmentation results by our algorithm. **d** Final segmentation result

class is generated (3.3), we can simply dilate each connected component by a radius 3 of disk structuring element.

4 Experiment

4.1 Evaluation criteria

Two levels of criteria are usually used to measure the performance of nuclei segmentation methods: one is object-level criteria, another is pixel-level criteria. The most common object-level criteria for object detection tasks include *precision*, *recall*, and *F1score*. *Precision* is defined as:

$$precision = \frac{TP}{TP + FP}$$

recall is defined as:

$$recall = \frac{TP}{FN + TP}$$

F1score considers both of the *precision* and *recall*, as shown in following equation.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

where the TP is true positive, FP means false positive, and FN means false negative. Given a manually labeled ground truth nucleus T_i, if there is one nucleus P_j in automatic segmentation result that matches T_i, P_j can be counted as one TP.

F1 score is the harmonic average of *precision* and *recall* and its value is in the range of [0,1].

We noticed that FN can be caused by two different types of errors: one is miss-detection (nuclei is predicted as

Table 1 Quantitative comparison results of segmentation performance on MOD dataset

Methods	Precision	Recall	F1	Dice's coefficient	MDR	FDR	USR	OSR
CNN3 [12]	-	-	0.827	0.762	-	-	-	-
NB model 1 (our method)	0.813	0.914	0.854	0.812	0.09	0.09	0.09	0.01
NB model 2	0.861	0.856	0.846	0.808	0.05	0.13	0.08	0.03
NB model 3	0.880	0.864	0.854	0.818	0.07	0.11	0.05	0.03
NB model 4	0.812	0.925	0.861	0.805	0.09	0.07	0.09	0.01
NB model 5	0.814	0.910	0.846	0.803	0.10	0.08	0.10	0.01
NB model *	0.845	0.892	0.850	0.81	0.06	0.11	0.02	0.08

cytoplasm); another is under-segmentation (multiple ground truth nuclei are detected as one nucleus, hence only one of these nuclei ground truth nucleus has corresponding detected nucleus.). Similarly, FP consists two types of errors: one is false detection (cytoplasm is detected as nuclei); another is over-segmentation (one ground-truth nucleus is segmented into several nuclei; each of them is a part of the ground truth nucleus and at most only one among them can be considered as the corresponding detected nucleus). Let us think about this situation: one segmentation method is weak on discriminating the nuclei and cytoplasm while another one is weak on splitting the nuclei area. But they may have similar *precision* and *recall*, even *F1score*. Apparently, *precision*, *recall*, and *F1score*,
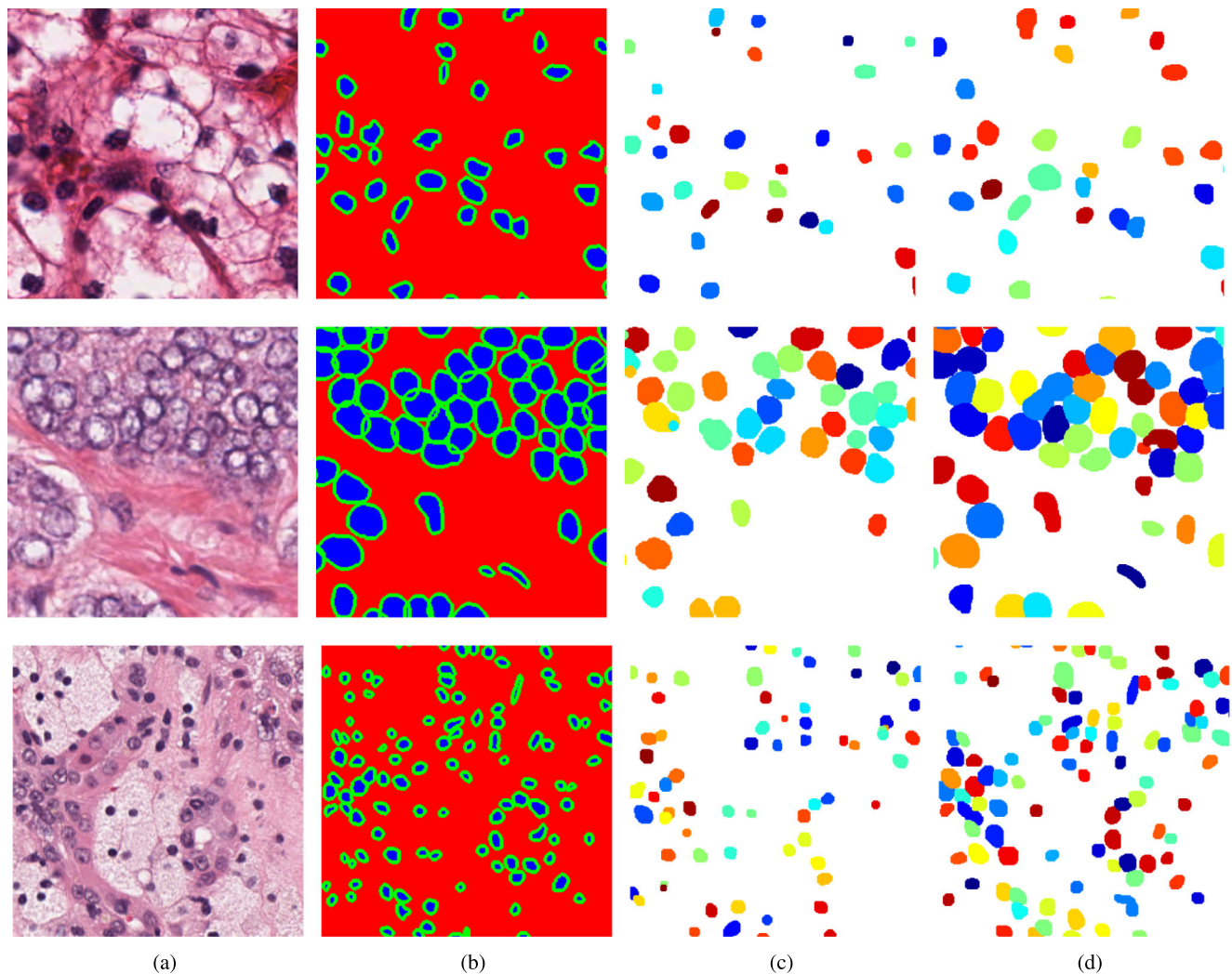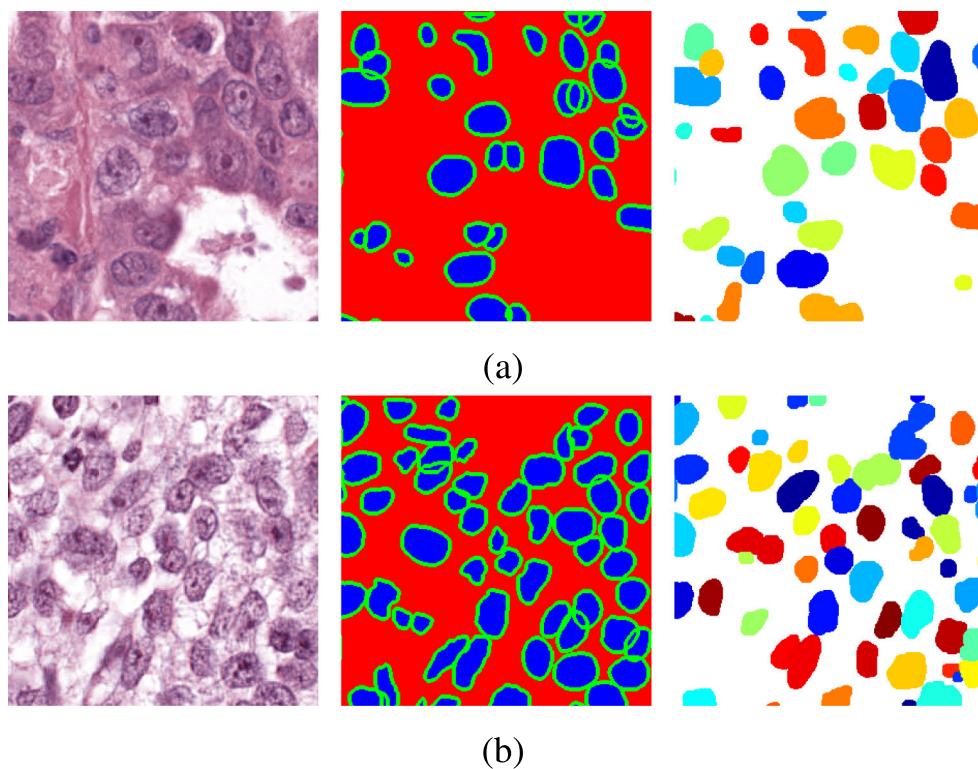

(a) (b) (c) (d)

Fig. 8 The comparison between our method and CNN3 [12]: **a** raw images; **b** ground truth; **c** CNN3 results; **d** our results

Fig. 9 Cropped portions of two images. **a** precision = 0.76, recall = 0.83, OSR = 0.05, USR = 0.15, MDR = 0.02, FDR = 0.20; **b** precision = 0.78, recall = 0.83, OSR = 0.13, USR = 0.05, MDR = 0.12, FDR = 0.10

(a)

(b)

and their combination, fail to differentiate the performance of these two segmentation methods. To handle this issue, we introduce four new criteria to evaluate automatic nuclei segmentation methods: missing detection rate (MDR), false detection rate (FDR), under-segmentation rate (USR), over-segmentation rate (OSR), as shown in Eq. 4.

$$
\begin{aligned}
MDR &= \frac{MD}{FN + TP} \\
FDR &= \frac{FD}{TP + FP} \\
USR &= \frac{US}{P} \\
OSR &= \frac{OS}{S}
\end{aligned}
\tag{4}
$$

where MD is the number of missing detections; FD indicates the number of false detections; US means the number of nuclei which are not detected caused by under-segmentation. P is the number of ground truth nuclei in the region of TP, which can be defined as $FN + TP - MD$. OS means the number of false positives caused by over-segmentation and S means the number of segmented nuclei in the region of TP's corresponding ground truth nuclei, which can be defined as $FP + TP + FD$. The combination of MDR and FDR measures the capacity

of discriminating the nuclei and cytoplasm while the combination of USR and OSR measures the performance of handling overlapped nuclei area. On the other hand, *recall* value is negatively correlated with MDR and USR while *precision* is negatively correlated with FDR and OSR. These four criteria are able to help pathologists select proper automatic segmentation methods for specific tasks.

The pixel-level criteria are used to measure the accuracy of segmentation algorithms in predicting the shape and size of the detected nuclei. The most essential one is Dice's coefficient, which is defined as:

$$
D(X, Y) = 2 \frac{|X \cap Y|}{|X| + |Y|}
\tag{5}
$$

where X indicates a manual segmentation and Y means its corresponding automatic segmentation. That is, a manual segmentation is considered as a FP if there is no corresponding automatic segmentation with a Dice coefficient of at least 0.2.

4.1.1 Datasets

We evaluate the performance of our method on three public available nuclei segmentation datasets. One is a multiple-organ H&E-stained image dataset [12] (MOD). It consists
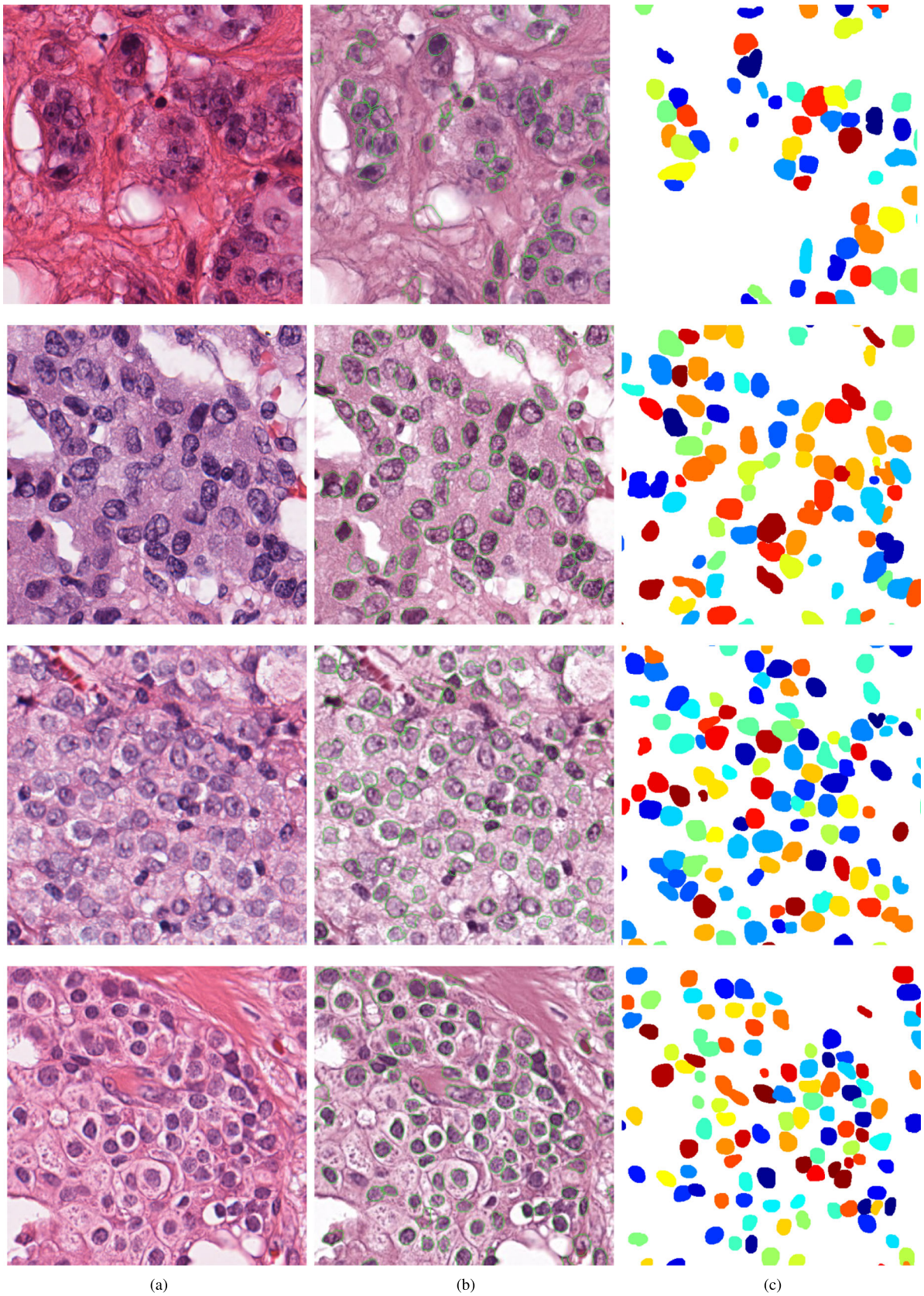
Fig. 10 Nuclei segmentation result over the BCD dataset. **a** Two breast cancer image samples. **b** Automatic segmentation result of [32]. **c** Result of our method

Table 2 Quantitative comparison of segmentation performance on the BCD dataset

Dataset	Method	Precision	Recall	F1	DC
BCD	Veta's method [32]	0.863	0.886	0.874	0.88
	TV-MRF-BP [23]	0.801	0.823	0.811	0.84
	NB model	0.942	0.915	0.923	0.862
BNS	FCN [15]	0.823	0.752	0.763	-
	DeconvNet [21]	0.864	0.773	0.805	-
	Ensemble [20]	0.741	0.9	0.802	-
	NB model	0.920	0.7835	0.84	0.83

of 30 images which were captured from 7 organs: the breast, liver, kidney, prostate, bladder, colon, and stomach. The resolution of each image is 1000×1000. In total, about 21,000 nuclei boundaries are manually annotated. These 30 images are split into two subsets: the training set with 16 images composed of 4 from the breast, 4 from the liver, 4 from the kidney, and 4 from the prostate; and the test set with 14 images composed of 2 images from each organ.

The second dataset is the breast cancer histopathology image dataset (BCD). It contains two subsets: subset A and subset B. Subset A includes 21 images and subset B has 18 images. In [32], Subset A is used to tune the parameters. In a similar way, we utilize subset A as the training set and subset B as the test set. Since one image may contain thousands of nuclei, it is impractical to manually label all the training images. We randomly select five images from subset A and crop a 1000 × 1000 subimage from each of them to build the training set. It is manually annotated under the supervision of a specialist.

The third one is also a breast cancer image dataset (BNS) [20]. It is composed of 33 H&E-stained images of size 512 × 512 from 7 triple negative breast cancer patients. There are a total of 2754 manually annotated nuclei.

4.2 Experiment result

Figure 7 shows how our method segments the nuclei step by step. The color variety is well controlled by the color normalization procedure. The prediction result shows clear nuclei areas and boundaries. In the final segmentation result and ground truth image, each nucleus is represented by a different color.

First, we test our method on the MOD dataset. Unfortunately, the dataset publicly provided online does not explicitly divide the whole dataset into the training set and test set. We do not know which image belongs to the training set exactly as introduced in their paper [12]. To make a fair comparison, we randomly select 16 images from the breast,

liver, kidney, and prostate. Then, we combine the remaining 8 images of these four types and the 6 images from the bladder, colon, and stomach to build the test images. A total of 12,000 patches are randomly extracted from 12 training images to train our model. To eliminate the bias caused by random selection, 5 different training sets and the corresponding test sets are randomly generated. Then, the model is trained and tested on the 5 pairs of training set and test set separately. All of the models are trained for 300 epoch in 7.5 h. For testing, the stride of overlapped patch extraction is set to 64. The quantitative comparison is listed in Table 1, which demonstrates that our method outperforms the state-of-the-art method CNN3 as reported in [12] in terms of both F1 score and Dice's coefficient. Moreover, it shows that the under-segmentation error is much more significant than over-segmentation error and it achieves a balance between the false detection error and missing detection error. Figure 8 shows a visual comparison between our method and [12]. As shown in the sample image, our segmentation result has fewer false negatives and higher accuracy in terms of nuclei boundaries than [12]. Our method is not only more accurate but also much faster. It takes about 5 s to predict a 1000 × 1000 image by one Nvidia Titan X GPU and the time used for post-processing is less than 0.1 s. Given the same hardware environment and test images, [12] takes about 4 min to predict one image and 80 s to do the post-processing. Additionally, a 10-folder cross-validation is performed to validate our method. The result is shown in Table 1 NB model*.

To show the benefit of our proposed evaluation metrics for nuclei segmentation, we compared two images with similar precision and recall, but different segmentation quality. As shown in Fig. 9, these two images have similar precision and recall. From our proposed criteria, we can find that the segmentation error on the first image is mainly caused by under-segmentation and false detections while that it is mainly caused by over-segmentation, missing detection and false detection in the second image.
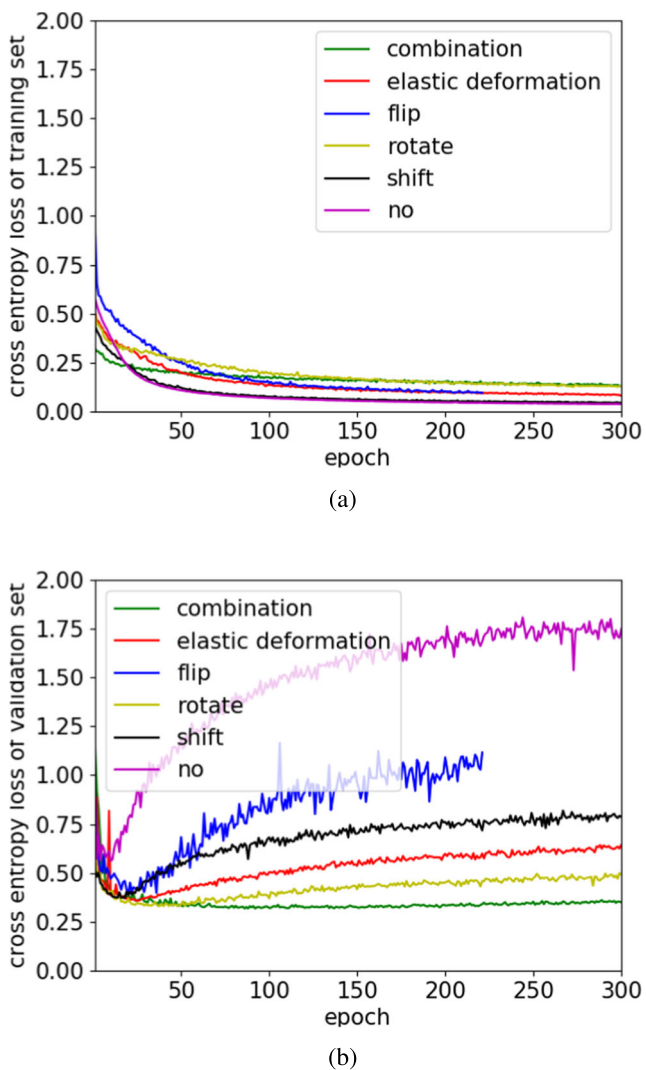
(a)

(b)

Fig. 11 a Shows how the training loss changes during training. **b** Indicates the validation loss

Second, we test our method on the BCD dataset. The manually labeled training set consists of five 1000 × 1000 images. Instead of training the models from random initialization, we use the training data to fine-tune the network model trained on the MOD dataset. Thus, the model would adjust to a new dataset with much shorter time by training on a limited training set for a small number of epochs. In this experiment, only 2000 patches are extracted to fine-tune the pretrained model. It takes about 10 s to train one epoch and the training is terminated after 70 epochs. The visual comparison between our algorithm and algorithm in [32] can be seen in Fig. 10.

At last, we follow the same strategy in [20] to validate our method. The strategy is called leave-one-patient-out cross-validation. That is, every time we train the model on 6 patients and use the rest for validation. Table 2 shows that our method outperforms the state-of-the-art breast cancer nuclei segmentation method by a large margin in terms of precision, recall, and F1 score.

4.3 Discussion

4.3.1 Data augmentation for fully convolutional networks

Data augmentation is a widely used technique to handle the overfitting issue caused by limited training samples. In image segmentation tasks, one can generate more images from one image using image transformation methods. The most common methods include rotation, flipping, shifting, and rescaling. Elastic deformation transform, a higher level transformation method, is also employed in some image segmentation works. Ronneberger et al. [26] claim that elastic deformation is the key method to do data augmentation for a segmentation network with very limited annotated images.

However, to the best of our knowledge, there is no systematic study of the effectiveness of these image transformation methods for nuclei segmentation using a fully convolutional network. We compare different training processes using rotation, flipping, shifting, rescaling, and elastic deformation transform to augment the training data. To make fair comparisons, we let the training set and validation set have similar appearances by splitting each whole image into two sub-images and placing one in the training set and another one in the validation set. We randomly extract 6000 patches from the training set to train our neural networks and 6000 patches from the validation set for validation. The setting of these transformation methods is same with those reported in Section 3.3.2. The comparison is shown in Fig. 11: "no" means do not apply data augmentation; "combination" means data augmentation is performed by combining elastic deformation, flip, rotate, shift, and rescale. It is very clear that without data augmentation, the network has severe overfitting issue, and validation loss starts to increase rapidly from epoch 5. Unexpectedly, rotating rather than elastic deformation has achieved the best performance in performance improvement. But only rotating operation still cannot prevent the overfitting. One has to combine all of these transform methods together to do data augmentation to get good performance as done in this paper.

Fig. 12 **a** An H&E-stained image. **b** The raw segmentation results of our method. **c** The prediction result without applying weight map and overlapped patch extraction and assembling. **d** The prediction results using overlapped patch extraction and assembling but without weight map

(a)

(b)

(c)

(d)
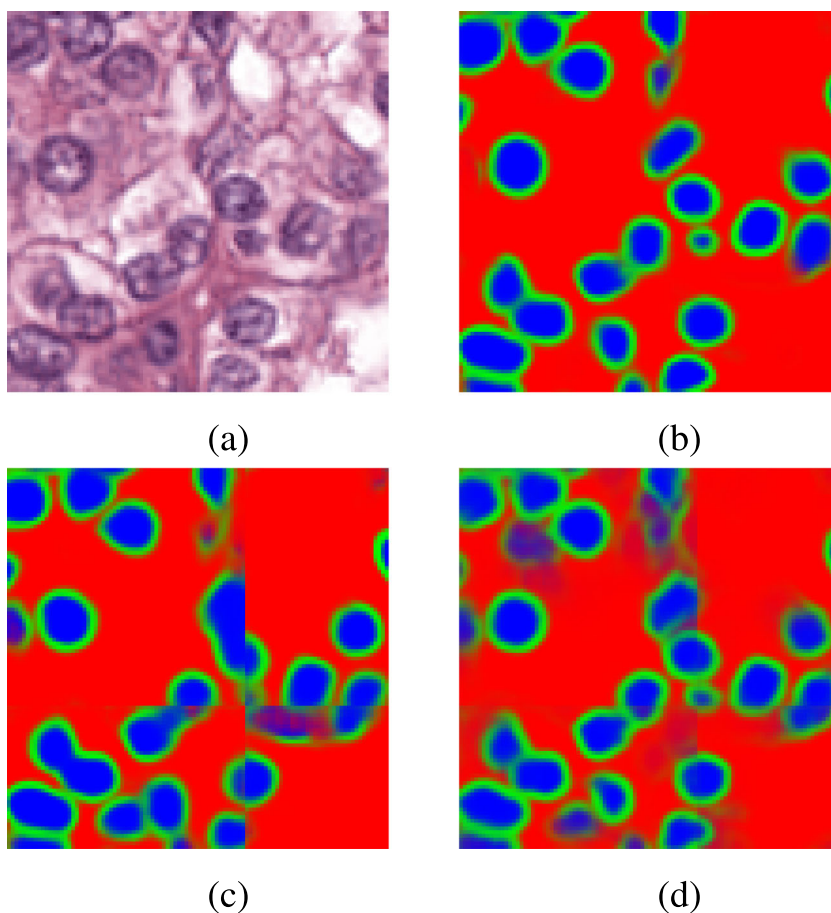
4.3.2 Nuclei segmentation on extra-large images

To evaluate the effectiveness of the proposed weight map and overlapped patch extraction and assembling method for extra-large image segmentation, we compared the segmentation results with and without the proposed method in Fig. 12. We can see that the raw segmentation results without using those two techniques contain obvious seams between the patches. It also demonstrates that the predictions in the border area are not accurate. As shown in Fig. 12d, if we employ the overlapped patch extraction and assembling but without the weight map (which means all the pixels in a patch have the same weight), the segmentation

Table 3 Quantitative comparison of prediction performance on whole images

	Cross entropy of assembled image
Uniform loss	0.85
Overlapped	0.59
Overlapped + weight map	0.52

result still shows noticeable seams. Figure 12b and d have the same stride, which is 64. Table 3 shows the quantitative comparison of prediction performance on whole MOD test images.

4.3.3 NB model versus the mixed nucleus model + boundary model

An alternative way to detect nuclei and their boundaries is training two binary classifiers to detect *inside* and *boundary* separately and then merge the detections together. We apply the same method with our NB model to train the nucleus model and boundary model except that the three-class classification is replaced by binary classification. Figure 13 depicts why the NB model outperforms the mixed nucleus model + boundary model. The NB model is able to learn the latent relationships between *inside*, *boundary*, and *background*. That is, there should be no gaps between *inside* and *boundary* classes and *inside* should not cross the *boundary* class. From the samples shown in Fig. 13, we can easily find out that NB model predicts the *inside* class and *boundary* classes more precisely.

Fig. 13 The comparison of NB model against the mixed nucleus model and boundary model. First column shows the histopathology images. Second column shows estimated nuclei and boundaries using our NB model. Third column shows the estimated result generated by the mixed nucleus model and boundary model. Fourth column represents the ground truth
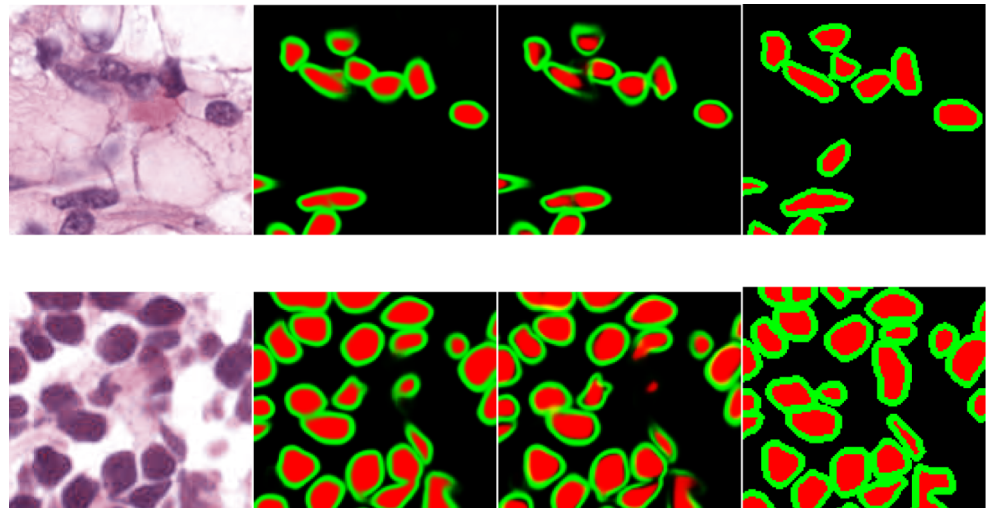
5 Conclusion

In this paper, we have presented a state-of-the-art supervised fully convolutional neural network–based method for nuclei segmentation in histopathology images. First, the images are normalized into the same color space. To handle the extra-large image issue, one whole image is split into overlapping patches for succeeding processing. Next, we propose a novel nucleus-boundary model to detect nuclei and boundaries on each patch. Then, the predictions of all the patches are seamlessly reassembled to build the raw prediction result of the whole image. At the end, we apply a fast and non-parameter post-processing to generate the final nuclei segmentation results. The nucleus-boundary model is trained on a very limited number of images and has been tested on the images that may have different appearances. Comparison with the state-of-the-art algorithm shows that our proposed method is accurate, robust, and fast. It is also found that our idea of simultaneous nucleus-boundary identification model can be applied to other biomedical image segmentation tasks such as gland segmentation and bacteria segmentation.

Data Availability The source code is public available in https://github.com/easycui/nuclei_segmentation and licensed under MIT license https://github.com/easycui/nuclei_segmentation/blob/master/LICENSE.

References

1. Al-Kofahi Y, Wiem LW (2010) Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 57(4):841–852

2. Chen JM, Qu AP, Wang LW, Yuan JP, Yang F, Xiang QM, Maskey N, Yang GF, Liu J, Li Y (2015) New breast cancer prognostic factors identified by computer-aided image analysis of he stained histopathology images. Scientific reports 5

3. Cui Y, Hu J (2016) Self-adjusting nuclei segmentation (sans) of hematoxylin-eosin stained histopathological breast cancer images. In: 2016 IEEE International Conference on Bioinformatics and biomedicine (BIBM). IEEE, pp 956–963

4. Filipczuk P, Kowal M, Obuchowicz A (2011) Automatic breast cancer diagnosis based on k-means clustering and adaptive thresholding hybrid segmentation. In: Image processing and communications challenges 3. Springer, pp 295–302

5. Gandomkar Z, Brennan PC, Mello-Thoms C (2016) Computer-based image analysis in breast pathology, vol 7

6. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Artificial Intelligence and Statistics, pp 249–256

7. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, pp 770–778

8. Khan AM, Rajpoot N, Treanor D, Magee D (2014) A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Trans Biomed Eng 61(6):1729–1738

9. Kharma N, Moghnieh H, Yao J, Guo YP, Abu-Baker A, Laganiere J, Rouleau G, Cheriet M (2007) Automatic segmentation of cells from microscopic imagery using ellipse detection. IET Image Process 1(1):39–47

10. Kingma D, Ba J (2014) Adam: A method for stochastic optimization. arXiv:14126980

11. Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. arXiv:170602515

12. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A (2017) A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Transactions on Medical Imaging

13. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788

14. Liao M, Yq Z, Li Xh, Ps D, Xu Xw, Jk Z, Bj Z (2016) Automatic segmentation for cell images based on bottleneck detection and ellipse fitting. Neurocomputing 173:615–622

15. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 3431–3440

16. Macenko M, Niethammer M, Marron J, Borland D, Woosley JT, Guan X, Schmitt C, Thomas NE (2009) A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical imaging. IEEE, pp 1107–1110

17. Mouelhi A, Sayadi M, Fnaiech F, Mrad K, Romdhane KB (2013) Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. Biomed Signal Process Control 8(5):421–436

18. Herrera de la Muela M, Garcia Lopez E, Frias Aldeguer L, Gomez-Campelo P (2017) Protocol for the BRECAR study: a prospective cohort follow-up on the impact of breast reconstruction timing on health-related quality of life in women with breast cancer. BMJ Open 7(12):e018108

19. Nawaz S, Yuan Y (2015) Computational pathology: Exploring the spatial dimension of tumor ecology. Cancer letters

20. Naylor P, Laé M, Reyal F, Walter T (2017) Nuclei segmentation in histopathology images using deep neural networks. In: 2017 IEEE International Symposium on Biomedical imaging. IEEE, pp 933–936

21. Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: IEEE International Conference on Computer Vision, pp 1520–1528

22. Otsu N (1975) A threshold selection method from gray-level histograms. Automatica 11(285-296):23–27

23. Paramanandam M, O'Byrne M, Ghosh B, Mammen JJ, Manipadam MT, Thamburaj R, Pakrashi V (2016) Automated segmentation of nuclei in breast cancer histopathology images. PloS one 11(9):e0162053

24. Qu A, Chen J, Wang L, Yuan J, Yang F, Xiang Q, Maskey N, Yang G, Liu J, Li Y (2014) Two-step segmentation of hematoxylin-eosin stained histopathological images for prognosis of breast cancer. In: 2014 IEEE International Conference on Bioinformatics and biomedicine (BIBM). IEEE, pp 218–223

25. Rabinovich A, Agarwal S, Laris C, Price JH, Belongie SJ (2004) Unsupervised color decomposition of histologically stained tissue samples. In: Advances in neural information processing systems, pp 667–674

26. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 234–241

27. Rother C, Kolmogorov V, Blake A (2004) Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM Transactions on graphics ACM, vol 23, pp 309–314

28. Simard PY, Steinkraus D, Platt JC et al (2003) Best practices for convolutional neural networks applied to visual document analysis. In: ICDAR, vol 3, pp 958–962

29. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DR, Cree IA, Rajpoot NM (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging 35(5):1196–1206

30. Su H, Xing F, Lee JD, Peterson CA, Yang L (2014) Automatic myonuclear detection in isolated single muscle fibers using robust ellipse fitting and sparse representation. IEEE/ACM Trans Comput Biol Bioinform 11(4):714–726

31. Vahadane A, Peng T, Albarqouni S, Baust M, Steiger K, Schlitter AM, Sethi A, Esposito I, Navab N (2015) Structure-preserved color normalization for histological images. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, pp 1012–1015

32. Veta M, van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JP (2013) Automatic nuclei segmentation in h&e stained breast cancer histopathology images. PloS one 8(7):e70221

33. Veta M, Pluim JP, van Diest PJ, Viergever MA (2014) Breast cancer histopathology image analysis: a review. IEEE Trans Biomed Eng 61(5):1400–1411

34. Wang P, Hu X, Li Y, Liu Q, Zhu X (2016) Automatic cell nuclei segmentation and classification of breast cancer histopathology images. Signal Process 122:1–13

35. Xing F, Xie Y, Yang L (2016) An automatic learning-based framework for robust nucleus segmentation. IEEE Trans Med Imaging 35(2):550–566

36. Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, Madabhushi A (2016) Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. IEEE Trans Med Imaging 35(1):119–130

Yuxin Cui received the B.E. degree in Bioinformatics in 2010 and M.E. degree in Computer Science and Technology in 2013 from Harbin Institute of Technology, China. He is currently pursuing the Ph.D. degree in Computer Science at University of South Carolina, Columbia, USA. His research interests include machine learning and computer vision.

Guiying Zhang is an associate professor at Department of Medical Information Engineering, Zunyi Medical University, Zunyi, China. She received the B.E. degree of Computer Science and Technology in 2005 from Northwest Normal University. Dr. Zhang received M.E. degree of Computer Application in 2010 and the Ph.D. in College of Computer Science and Technology in 2017 respectively from Guizhou University, Guiyang, China. Her research interests include machine learning, pattern recognition, and computer vision.

Zhonghao Liu is currently a Ph.D. student at Department of Computer Science, University of South Carolina. He obtained his bachelor's degree from Sichuan University in China. His research interests include machine learning and bioinformatics.

Jianjun Hu is an associate professor at the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA. He is also guest professor at the School of Mechanical Engineering, Guizhou University, Guiyang, China. Dr. Hu received the B.S. degree and M.S. degree of Mechanical Engineering in 1995 and 1998 respectively from Wuhan University of Technology, China. He received the Ph.D. in Computer Science and Engineering in 2004 from Michigan State University in the area of machine learning. He was a postdoctoral fellow at both Purdue University and University of Southern California from 2004 to 2007. His research interests include machine learning, data mining, evolutionary computation, and bioinformatics. Dr. Hu is a member of ACM and IEEE.

Zheng Xiong received the B.E. and M.Sc degree in Electrical Engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011 and 2014, respectively. Since then, he began his pursuit of Ph.D. degree in Computer Science at the University of South Carolina. His research interests include machine learning, data mining, material informatics, and bioinformatics.