HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2019 February 01.

Published in final edited form as:

Med Image Anal. 2018 February ; 44: 245–254. doi:10.1016/j.media.2017.07.003.

Efficient and Robust Cell Detection: A Structured Regression Approach

Yuanpu Xie[1], **Fuyong Xing**[2], **Xiaoshuang Shi**[1], **Xiangfei Kong**[3], **Hai Su**[1], and **Lin Yang**[1,2,*]

[1]Department of Biomedical Engineering, University of Florida, FL 32611 USA

[2]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, USA

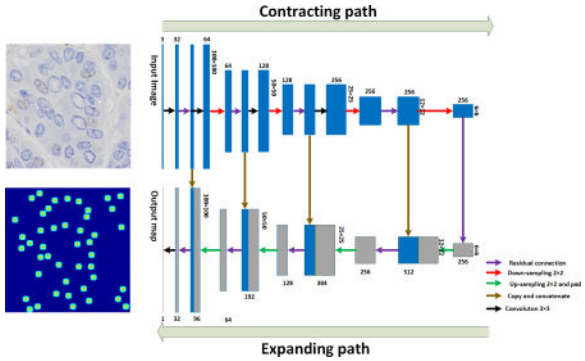
[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Drive 637553 Singapore

Abstract

Efficient and robust cell detection serves as a critical prerequisite for many subsequent biomedical image analysis methods and computer-aided diagnosis (CAD). It remains a challenging task due to touching cells, inhomogeneous background noise, and large variations in cell sizes and shapes. In addition, the ever-increasing amount of available datasets and the high resolution of whole-slice scanned images pose a further demand for efficient processing algorithms. In this paper, we present a novel structured regression model based on a proposed fully residual convolutional neural network for efficient cell detection. For each testing image, our model learns to produce a dense proximity map that exhibits higher responses at locations near cell centers. Our method only requires a few training images with weak annotations (just one dot indicating the cell centroids). We have extensively evaluated our method using four different datasets, covering different microscopy staining methods (e.g., H & E or Ki-67 staining) or image acquisition techniques (e.g., bright-filed image or phase contrast). Experimental results demonstrate the superiority of our method over existing state of the art methods in terms of both detection accuracy and running time.

Graphical abstract

*Corresponding author, lin.yang@bme.ufl.edu.



Keywords

Biomedical image analysis; structured regression; deep learning; cell detection

1. Introduction

Manual analysis of microscopy images is not only laborious and expensive but is also inclined to suffer from inter-observer variabilities. Recent progress Gurcan et al. (2009) shows that digitized specimen analysis can significantly improve the objectivity and reproducibility of computer-aided diagnosis (CAD). In the context of microscopy image analysis-based CAD, automatic and robust cell detection are highly desirable and serve as an essential prerequisite for a wide variety of subsequent tasks, such as cell segmentation, tracking and morphological measurements Veta et al. (2014). In addition, the combination of cell detection and a successive stage of cell classification can provide clinically useful information about objects of interest, such as the presence (or quantity) of cancer cells in a microscopy image.

Many research efforts have been devoted to cell detection during the past decades Gurcan et al. (2009); Xing and Yang (2016). Unfortunately, the success of cell detection is hindered by the complex characteristics of microscopic images such as touching cells, background clutters, large variations in the shape and size of cells, poor contrast, and differences between image acquisition techniques (see Figure 1). In addition, microscopy images often have very high resolution, which further pose a challenge on the computational resources.

A general approach towards object detection or localization can be reduced to finding local maxima on a response (e.g. probability) map. The essential assumption here is that the object centers should have larger responses than their surrounding pixels. A typical way to obtain the probability map is sliding window based classification. In this paper, instead of working on the hard class labels with classification, we take a further step to directly regress the proximity value for each pixel, whose definition is based on the Euclidean distance between each pixel and the closest nucleus. This approach has been shown to be more robust and accurate than hard label classification in (Xie et al., 2015c; Kainz et al., 2015). In addition, training with structured output is more effective than using a single value for each training image (Xie et al., 2015c). One of the reasons is that training with structured labels

makes use of the structured information exhibited in the labels, thereby the supervision information is much richer than a single value.

In this paper, we propose a novel structured regression model based on a newly introduced neural network architecture for cell detection in microscopy images with both accuracy and efficiency in mind. Instead of providing a single class label like many traditional methods, our algorithm generates structured predictions (referred to as proximity map), which exhibit higher values for pixels near cell centers. The final cell centroids are then localized by identifying all the local maximum positions.

We conduct experiments using four datasets (Ki-67 stained Neuroendocrine Tumor (NET) microscopy images, phase contrast HeLa cervical cancer microscopy images, and H&E stained Breast cancer and Bone marrow microscopy images, for detailed description of the datasets, please refer to Section 3.) The experimental results demonstrate that the proposed method can achieve very promising detection performance despite the background noise, dense cell overlapping and large variations of cell morphology.

This work is an extended version of our recent conference paper (Xie et al., 2015c). To the best of our knowledge, this paper as well as our previous work (Xie et al., 2015c) is the first study to report the application of structured regression using deep architectures for efficient cell detection. In comparison to the conference version, this work has the following extensions:

1. We discuss the original convolutional neural network based structured regression framework used in the conference paper (Xie et al., 2015c), and provide further insight into why it is hard to train the model when the output proximity patch is very large. Inspired by the fully convolutional neural network Long et al. (2015); Ronneberger et al. (2015), and deep residual learning, we introduce a novel fully convolutional residual network which directly outputs a dense proximity prediction that has the same size as the input image. We demonstrate that our method is robust and general on various microscopy image datasets.
2. We provide a generalized version of the original weighted square error loss function. This new loss function allows the model to adjust weights of the loss coming from different areas of the output based on the actual distribution of each training batch, rather than a predefined fixed value (more details are presented in Section 4).
3. Inspired by the work of (Kainz et al., 2015; Sironi et al., 2014), we adopt a more sophisticated proximity definition that leads to better distinctive peaks at cell centroids.

The rest of the paper is organized as follows. In Section 2, we present a literature review on the related works with emphasis on cell detection in microscopic images. We briefly introduce the datasets used in this study in Section 3. We present the fully residual convolutional neural network and our general structured regression model in Section 4, and describe the detailed experimental settings, results and the experiment analysis in Section 5. Finally, we conclude the paper and discuss the potential future work in Section 6.

2. Related works

Numerous works can be found in the literature for cell detection in microscopy images. Laplacian-of-Gaussian filter based on Euclidean distance map (Al-Kofahi et al., 2010) has been applied to automatic nuclei localization, and graph partition methods are reported in (Bernardis and Yu, 2010; Zhang et al., 2014) to detect cells automatically. Several other cell localization and segmentation methods using concave point based touching cell splitting are reported in (Yang et al., 2008; Kong et al., 2011), where the performances heavily rely on concave point detection. Assuming that cells are approximately circular or elliptical, Parvin et al. (2007) introduce a kernel based radial voting method to iteratively localize cells, which is relatively insensitive to image noise. Several other radial voting-based methods are presented in (Qi et al., 2012; Xing et al., 2013) for automatic cell detection on pathology images. Cosatto et al. (2008) employ difference of Gaussian (DoG) together with Hough transformation to detect radially symmetric nuclei. However, many of the aforementioned unsupervised methods are based on heuristics, and can not generalize well to different microscopy image modalities. In addition, the inhomogeneous background, irregular cell morphology and touching cell further challenge those unsupervised methods. To address those problems, supervised learning based methods have also attracted considerable attentions due to their promising performance.

Ali and Madabhushi (2012) apply an active contour-based shape model to detect and split overlapping cells in histological images and obtain encouraging results on their datasets. Arteta et al. (2012) propose a method based on maximally stable extremal region (MSER) selection, where each MSER is scored by a structured SVM (Bertelli et al., 2011), and the final detection results are obtained by finding the optimal configuration of the selected regions using dynamic programming to explore the tree structure of MSER. However, for images with strong inhomogeneous background or/and low intensity contrast, the MSER detector can not generate feasible region candidates and thus its usage is limited. Recently, Vink JP (2012) apply two Adaboost classifiers with Haar-like features and pixel-based features to nucleus detection in histological images, and the outputs of two detectors are merged using an active contour algorithm to obtain the final results. Irshad (2013) propose a framework that employs the morphological and statistics features in selected image channels to detect mitosis in histopathology for breast cancer grading. Kainz et al. (2015) use random forests to regress the proximity value for each pixel with a bag of pre-defined features to detect cell, and a similar architecture is presented in (Sironi et al., 2014) to extract linear structures (e.g. neurons) in medical images.

Despite their success of the methods above, the performance of aforementioned works heavily relies on various hand-crafted features such as shapes, gradients, colors, etc.

Recent literature has demonstrated that deep learning based methods have a remarkable ability to learn task specific feature representations, which are generally superior (Cruz-Roa et al., 2013) to hand-crafted feature. These learned features have been adopted to achieve state-of-the-art performance on many biomedical image analysis tasks (Cruz-Roa et al., 2013; Liao et al., 2013; Li et al., 2014; Xing et al., 2015). Xu et al. (2015) explore a stacked sparse auto-encoder to learn high-level features of sliding window patches, which are fed

into a softmax classifier to categorize as nuclear or non-nuclear samples, and obtain promising nuclei detection results on breast cancer histopathology images. Another stacked auto-encoder based method is reported in (Su et al., 2015) for cell detection and segmentation in microscopy images. Ciresan et al. (2013) employ a convolutional neural network (CNN) (LeCun et al., 1998) as a sliding window classifier to detect mitosis in breast cancer histological images. A similar architecture has also been utilized to segment membrane neuronal (Ciresan et al., 2012) in electron microscopy images. Recent work (Xie et al., 2015b) also shows that CNN is capable of learning the geometric information exhibited in the training images.

Recently, (Sirinukunwattana et al., 2016) propose a novel spatially constrained convolutional neural network and achieve promising performance for cell detection in routine colon cancer histology images. Another similar architecture that deploys the idea of spatial constrained CNN is reported in (Sirinukunwattana et al., 2015).

However, almost all of the aforementioned methods involve expensive sliding window classification or regression, which is not efficient enough to be applicable to large microscopy whole-slices. Xie et al. (2015c) propose a CNN based structured regression method, which, for every testing image patch, produces a proximity patch encoding every pixel's proximity to it's closest nucleus. This allows the use of a stride strategy which skips a large portion of the testing pixels. Meanwhile, a large proximity patch is preferred, since it allows us to use a large testing stride which can greatly improve the efficiency. However, the sub-sampling (e.g, max-pooling) operations used in Xie et al. (2015c) results in significant information loss, which makes it challenging to train the structured regression model with a large-size proximity patch as the model's output.

Recently, residual learning He et al. (2015) has been a revolutionizing technique in designing deep convolutional neural networks. It has allowed effective training of very deep networks and achieved the 1st place in the ILSVRC 2015 classification task. The following-up work in He et al. (2016) gives a complete theoretical and experimental analysis of the importance of identity mapping used in deep residual learning.

Fully convolutonal neural networks (FCNs) Long et al. (2015) are an end-to-end trainable architecture which is greatly efficient and can scale well to large image size. In addition, since it does not contain fully connected layers which require fixed input image size, the learned model can be applied to arbitrary images of various sizes. However, the conventional FCN Long et al. (2015) usually fails to produce highly accurate pixel level prediction. The works presented in Noh et al. (2015); Yang et al. (2016); Xie et al. (2015a); Laina et al. (2016) utilize an up-pooling to help the subsequent convolutional layers to generate results with higher localization accuracy. Most of the FCN and it's variations Xie and Tu (2015); Chen et al. (2016); Zheng et al. (2015) are initialized with weights transfered from pretrained models like VGG Simonyan and Zisserman (2014), Inception Szegedy et al. (2016), etc.

FCN and it's variants have a large number of successful applications in semantic image labeling Chen et al. (2016); Zheng et al. (2015); Long et al. (2015), edge detection Xie and

Tu (2015), depth estimation Laina et al. (2016), etc. In the area of biomedical image analysis, Xie et al. (2015a) applies a variant of FCN to compute the cell density distribution for cell counting. U-Net Ronneberger et al. (2015), which extends the FCN by providing expanding layers with information from corresponding contracting layers, is capable of producing precise segmentation results on various biomedical images. Fakhry et al. (2016) propose a residual deconvolutional network for brain electron microscopy image segmentation. Some other works that utilize FCN are reported in Wang et al. (2016); Christ et al. (2016).

3. Datasets

In this section, we introduce four microscopy image datasets used in this paper, each dataset represents one distinct stain preparation, modality, or image acquisition technique. One sample image for each dataset is shown in Fig 1.

3.1. Neuroendocrine tumor (Fig. 1a)

Neuroendocrine Tumor (NET) is considered as one of the most common leading cause of cancer deaths worldwide. Early diagnosis and treatment are crucial for the survival of NET patients. Ki-67 proliferation index, defined as the ratio of the count of immunopositive tumor cells to all the tumor cells, is an important prognostic and grading cue for NET (Dhall et al., 2012). Accurate detection of all cells in the entire image can serve as the starting point for the subsequent cell classification and counting, which can support the assessment of Ki-67 proliferation index.

This dataset contains 59 cropped Ki-67 stained bright-field NET images patches, the size is rough 400 × 400 × 3 and all the images have human annotations as ground truth. All the images are obtained at 20× magnification. This dataset is randomly divided into two halves for training and testing. The cell detection on this dataset is challenging due to touching cells, blurred (or weak) cell boundaries and inhomogeneous background noise.

3.2. HeLa cervical cancer (Fig. 1b)

This dataset contains 22 phase contrast microscopy images of HeLa cervical cancer cells. It is collected to monitor the detailed colony growth in radiation experiments (Arteta et al., 2012). Half of the images are chosen for training and the other half for testing. This image dataset exhibits background similarity and large variations in the cell shape and size.

3.3. Breast cancer (Fig. 1c)

The H&E stained bright-field breast cancer dataset is obtained from The Cancer Genome Atlas (TCGA) (National Cancer Institute, 2013). In total 70 image patches are randomly cropped and manually annotated, the size of each roughly ranges from 300 × 300 × 3 to 500 × 500 × 3. We use half of this dataset for training and the other half for testing. As is shown in Fig. 1c, this dataset contains a large portion of highly inhomogeneous background noise, and a significant variation in cell size and shape.

3.4. Bone marrow (Fig. 1d)

This dataset is first introduced in (Kainz et al., 2015) and contains 11 1200 × 1200 pixel H&E stained bright-field microscopy images of human bone marrow tissue from eight different patients. The dataset is split into two sets: 8 for training and 3 for testing. Each image is cropped from whole-slice scanned images at 40× magnification. Similar to the breast cancer image data, this dataset exhibits inhomogeneous background noise and large variations of cell sizes.

4. Methodology

Since microscopy images usually contain a large portion of cell clusters, it is critical to accumulate large context information for separating between cell clusters from background noise. Meanwhile, it is also important to capture fine, local information for accurately splitting touched cells. In the previous conference paper (Xie et al., 2015c), we use the convolutional neural network to conduct structured regression. However, we observed that it is difficult to train an ideal model that is capable of producing a large-size target proximity mask (e.g. the same size as the input image patch). Although it is reasonable to apply max-pooling operations to reduce the feature dimension and increase the receptive filed for image classification and recognition tasks, this operation actually results in massive loss of high resolution information contained in the input image, which is crucial for dense prediction problem.

In this paper, we present a fully residual convolutional neural network that is capable of achieving both of those two objectives (large receptive filed and high resolution information). Instead of doing patch-wise classification, our method produces dense proximity mask that is of the same size to the input image. Our model encodes the topological structured information exhibited in the training data and explicitly forces the pixels near cell centers to get higher values than their neighbor pixels.

4.0.1. Fully residual convolutional neural network

The detailed network architecture is illustrated in Fig. 2. It consists of one contracting path (upper side) which encodes the input to high-level features and one expanding path (lower side) that decodes those features to the output mask. The contracting path consists of a repeated stack of 3×3 convolution followed by a residual block (Fig. 3) and a 2×2 down-sampling layer. Before we down-sample the feature maps, we double the feature map channels using 1×1 convolution until it reaches 256. All the down-sampling operations used in our method are mean-pooling. The biggest difference between the expanding and contracting paths is that the down-sampling in the contracting path is replaced with up-sampling. In this paper, we use bilinear interpolation to up-sample the feature maps. To compensate the massive loss of high-resolution information during down-sampling layers and improve the localization accuracy, we concatenate the higher resolution feature maps from the contracting path to the corresponding up-sampled feature maps in the expanding path.

Residual connection—An identity mapping in residual blocks can be expressed as:

$$x_{l+1} = x_l + \mathcal{F}(x_l, \mathcal{W}_l) \quad (1)$$

where, $\mathcal{F}$ denotes the residual function, $\mathcal{W}_l$ the parameters of the l–th residual block. This type of residual identity mapping has a nice back-propagation property that the gradient does not vanish even when the weights are arbitrarily small.

In He et al. (2015), the $\mathcal{F}$ that achieves highest classification accuracy consists of a sequence of layers: BN-ReLu-Conv-BN-ReLu-Conv, where BN, ReLu, Conv stands for Batch Normalization, rectified linear activation, and convolution respectively. During the model training, we find that adding batch normalization in the residual blocks makes the training slower and does not have positive effects on the performance, especially on breast cancer dataset. One of the possible reasons is that we already use extensively dropout in the model. However, as is observed in Shah et al. (2016), without batch normalization, the gradient tends to explode for large network. To stabilize the training Szegedy et al. (2016), we scale down the activation of the last convolution layer in $\mathcal{F}$ before adding them to the input, we pick 0.3 as the scaling factor in all of our experiments. All the convolution kernels are 3×3. To prevent overfiting, we add dropout between two convolutional layers. We also replace the ReLu activation function with the more recent ELU Clevert et al. (2015), and thus the resulting residual function can be summarized as: **ELU-Conv-Dropout-ELU-Conv-Scaling**. This type of residual block is illustrated in Fig 3b. The residual connection require that the input should have the identical dimension to the output. In this paper, we use 1×1 convolution to map the feature channels to the desired output's dimension when the dimensions of input and the output mismatch with each other.

Feature map concatenation—During the expanding path, up-sampling layer is used to expand the feature map size, however, the up-sampling layer and down-sampling layer are not exactly 'symmetric'. For instance, the size of 25×25 feature maps after down-sampling operation becomes 12×12. However, when we up-sample 12×12 feature maps back, we can only get 24×24 feature maps. In order to preserve the size during this process, we pad the up-sampled feature maps to make match the size of the corresponding feature maps in the contracting path. Those two set of feature maps are then concatenated and passed to the following convolutional layers to inference the output. This process is illustrated using brown arrows in Fig 2.

4.1. Structured regression for cell detection

4.1.1. Data preprocessing—Denote $\boldsymbol{x} \in R^{d \times d \times c}$ as one local image patch extracted from image I at location (u, v), in which d and c represent the patch size and image channel, respectively. For simplicity, we only use square local image patches, and $\boldsymbol{x}$ can be identified by one quintuple $\{u, v, d, c, I\}$. Please note that image patches have the same image channel with the original image For each image I with human annotations, we compute the corresponding proximity map M using the following function:

$$\mathcal{M}(u,v)=\begin{cases}\dfrac{e^{\alpha\left(1-\frac{D(u,v)}{d}\right)}-1}{e^{\alpha}-1} & \text{if } D(u,v)\le d,\\ 0 & \text{otherwise,}\end{cases} \quad (2)$$

in which $D(u, v)$ is the Euclidean distance from pixel (u, v) to the closest human annotation. d is a distance threshold and α is the decay ration, and both of them are used to control the shape of this exponential function. This function is a normalized version of the one used in (Kainz et al., 2015; Sironi et al., 2014), in practice, we can choose a scaling factor to scale up this proximity value.

After obtaining the proximity map M for image I, we can define the proximity patch $\boldsymbol{s} \in R^{d\times d}$ for $\boldsymbol{x}$. $\boldsymbol{s}$ can be viewed as the structured label of $\boldsymbol{x}$, and can be identified as a quintuple $\{u, v, d, \mathcal{M}\}$. This data generation process is illustrated in Fig. 3a.

4.1.2. Inference in structured regression—We define $\{f_l\}_{l=1}^{L}$ as the transformation of each of the L layers parameterized by $\{\boldsymbol{\theta}_l\}_{l=1}^{L}$, respectively. Our goal is to learn a mapping function $\psi = f_L \circ f_{L-1} \circ \cdots f_1$, which maps the input local image patch to a proximity patch. Please note that, f_i is a general notation for the transformation of i-th layer, the corresponding $\boldsymbol{\theta}_i$ has distinct forms for different types of f_i. For example, $\boldsymbol{\theta}_i$ is given as $[\boldsymbol{W_i}, \boldsymbol{b_i}]$ if f_i denotes a conventional fully connected layer. Given one input $\boldsymbol{x}^i$, the network computes the output $\boldsymbol{o}^i$ as $\psi\ (\boldsymbol{x}^i; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$.

In order to evaluate the model's parameters, we formulate the structured regression as the following optimization problem:

$$\arg\min_{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_L}\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\mathcal{L}(\psi(\boldsymbol{x}^i;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_L),\boldsymbol{y}^i), \quad (3)$$

in which $\mathcal{L}$ is the loss function. A widely adopted loss function is mean square error, but in our case, a dominant portion of the value in the proximity patch is zeros and only a small portion of pixels has positive response. This might bias our model to produce trivial outputs containing all zeros. To solve this problem, we adopt a weighting strategy to allow the model to assign different weights to the loss coming from different regions of the proximity patches. More specifically, it is defined as:

$$\mathcal{L}(\psi(\boldsymbol{x}^i;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_L),\boldsymbol{y}^i)=\frac{1}{2}\sum_{j=1}^{p}(\beta y^i{}_j+\lambda\bar{\boldsymbol{y}}^i)(y^i{}_j-o^i{}_j)^2, \quad (4)$$

in which o^i_j denotes the j-th element of $\boldsymbol{o}^i$, $\bar{\boldsymbol{y}}^i$ represents the mean value of $\boldsymbol{y}^i$, β, b and λ are predefined constants and used to tune the weights of the losses coming from different parts

of the model's output. This loss function does not use a fixed weight for every training sample; instead, it allows the model to determine based on the mean value of the training proximity patch.

Denote $\boldsymbol{a}^i$ as the inputs to the last layer for training sample $\boldsymbol{x}^i$. We can obtain $\boldsymbol{a}^i = \psi(\boldsymbol{x}^i; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{L-1})$. We denote the j-th element of $\boldsymbol{y}^i$, and $\boldsymbol{a}^i$ as $y^i{}_j$ and $a^i{}_j$, respectively.

In order to back propagate the gradients from the last layer (structured regression layer) to the lower layers, we need to calculate the partial derivative of (4) with respect to the input to the last layer. More specifically, if the activation function is chosen to be sigmoid in the last layer, it can be given by

$$\frac{\partial \mathscr{L}(\boldsymbol{o}^i, \boldsymbol{y}^i)}{\partial a^i{}_j} = \frac{\partial \mathscr{L}(\boldsymbol{o}^i, \boldsymbol{y}^i)}{\partial o^i{}_j} \frac{\partial o^i{}_j}{\partial a^i{}_j} = (\beta y^i{}_j + \lambda \bar{\boldsymbol{y}}^i)(o^i{}_j - y^i{}_j) a^i{}_j (1 - a^i{}_j). \quad (5)$$

After computing the value of (5), we can calculate the gradients of (3) with respect to $\{\boldsymbol{\theta}_l\}_{l=1}^{L}$ recursively using the chain-rule based back-propagation algorithm.

Our algorithm uses a more complicated output layer since the regression is performed on proximity patches $\boldsymbol{y}^i \in \mathscr{Y}$ that encode topological information. The output of the proposed model is explicitly computed as quantitative predictions related to the locations of cell centers and thus produce much more precise and robust local maxima for locating cell centers than conventional sliding-window based classification methods. After obtaining the proximity prediction map (denoted as $\mathscr{P}$), a small threshold $\xi \in [0, 1]$ is applied to remove the values smaller than $\xi \cdot \max(\mathscr{P})$. The final procedure for nucleus localization is to find all the local maximum locations in $\mathscr{P}$.

5. Experiment

5.1. Evaluation metrics

Before we describe the metrics used to evaluate the performance of different methods and their variations in this paper, we first define the ground-truth regions as circular regions with radius r centered at all the human annotations. In our experiment, r is roughly chosen to be half of the average radius of all nucleus for each data set. For each testing image, we match all the detected cell centroids with the corresponding human annotations using Hungarian algorithm. The matching is performed with the constraint that any matched detection results must lie within the ground-truth region. All the matched detected cell centroids are considered as true positive (*TP*). False positive (*FP*) refers to detected cell centroids that are not matched with any human annotations. The ground-truth human annotations that are not matched with any detected cell centroids are considered to be false negative (*FN*).

Based on the above definitions, we define the following evaluation metrics: (1) The mean (μ_c) and standard deviation (σ_c) of the counting error. Specifically, given N testing images, we can have $\mu_c = \frac{1}{N}\sum_{i=1}^{N} \hat{c}_i$, and $\sigma_c = \frac{1}{N}\sqrt{\sum_{i=1}^{N} (\hat{c}_i - \mu_c)^2}$, where $\hat{c}_i$ represents the absolute

difference between the total number of detected cells and the ground-truth annotations for the i-th image. (2) The mean (μ_d) and standard deviation (σ_d) of the detection distance error. For N testing images, we can have $\mu_d = \frac{1}{N}\sum_{i=1}^{N}\hat{d}_i$, and $\sigma_d = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\hat{d}_i - \mu_d)^2}$, where $\hat{d}_i$ refers to the average Euclidean distance between human-annotated dots and the corresponding matched true positive detections for the i-th image. This metric is used to measure the localization accuracy. (3) The precision (P), recall (R), and F_1 score, which can be calculated as $R = \frac{TP}{TP+FN}$, $P = \frac{TP}{TP+FP}$ and $F_1 = \frac{2\times P\times R}{P+R}$, respectively. We also report the average running time (T) over all testing images in each dataset.

5.2. Implementation details

We implement our model in python using Theano Bergstra et al. (2010); Bastien et al. (2012) and Keras Chollet (2015). We train and evaluate our method on a machine with an Intel Xeon E5-1650 CPU, and an NVIDIA Quadro K4000 GPU. Our model is trained with adadelta Zeiler (2012), and the learning rate is set as 0.0001. The hyper-parameters are set as $d = 15$, $\alpha = 3$ in Equation. 2, we scale the proximity value by 3. We set $\beta = 0.2$, $\lambda = 1$ in Equation. 4. The hyper parameters are chosen based on heuristic and trade-off between model complexity and running time. All the images are processed to have 3 RGB channels. The ground truth is given as a set of coordinates of dot annotations (one dot near cell centroid). We random cropped 135 × 135 × 3 image patches as the training data. Data augmentation (random rotation, shifting and mirroring) are used to prevent over-fitting.

Please note that, for all the four different datasets, we use exactly the same network architecture. We compare with the cell detection results of Non-overlapping Extremal Regions Selection (*NERS*) Arteta et al. (2012), and our original conference work, convolutional neural network (CNN) based structured regression model Xie et al. (2015c), represented as (*CNN-SR*). Pixel-wise classification based methods are shown to be inferior to structured regression based methods Xie et al. (2015c), so we exclude them from the comparison. We also compare the our method with two FCN based cell counting architectures proposed in Xie et al. (2015a), denoted as (*FCRN-A*) and (*FCRN-B*), respectively. Meanwhile, we also include the following unsupervised methods: Iterative Radial Voting (*IRV*) Parvin et al. (2007), Image-based Tool for Counting Nuclei (*ITCN*) Byun et al. (2006), and Laplacian-of-Gaussian filtering (*LoG*) Al-Kofahi et al. (2010).

5.3. Neuroendocrine tumor

With significant challenges including cell overlapping, background noise, blurred cell boundaries and weak staining, it is not surprised that supervised methods win over unsupervised methods in this dataset. To eliminate the effects of incomplete cells, we exclude detection results lie on the image border. The detailed model configuration for *CNN-SR* used in this dataset is: Input(39 × 39 × 3) − C(34 × 34 × 32) − M(17 × 17 × 32) − C(14 × 14 × 32) − M(7 × 7 × 32)−Dense(1024) − Dense(1024) −Dense(289), in which C, M and Dense respectively represents the convolutional layer, max pooling layer, and fully connected layer. The sizes of C and M layers are represented as *width* × *height* × *depth*, where *width* × *height* defines the size of each feature map and *depth* represents the number of feature maps. The max pooling layer uses a window of size 2 × 2 with a stride of 2.

The detailed comparison results are shown in Table 1. It can be seen that *Ours* produces the highest detection precision and F_1 score, and exhibits strong robustness with the lowest mean and standard deviation of the counting error. Two example images marked with detection results of *Ours* are shown in Fig. 5.

It's worth noting that our conference work *CNN-SR* achieves the comparable recall and same overall performance in terms of F_1 score. It also obtains promising detection accuracy evidenced by the lowest mean and standard deviation of the detection distance error. On the other hand, the dense cell overlapping and low intensity contrast hinder the non-overlapping extremal regions detection algorithm used in *NERS* Bertelli et al. (2011) from producing high-quality region candidates pool, thereby leading to a relatively low recall value and a large mean μ_c and deviation σ_c of the counting error. *FCRN-A* also achieves very competitive results on this dataset.

Since cells in this data set mostly exhibit round or elliptical shapes, Iterative Radial Voting (*IRV*) Parvin et al. (2007) and Laplacian-of-Gaussian blob detector based methods (*LoG* Al-Kofahi et al. (2010) and *ITCN* Byun et al. (2006)) are also capable of achieving reasonable results.

5.4. HeLa cervical cancer

Since phase contrast microscopy images exhibit very low intensity contrast and contain a large number of cells with irregular shapes, supervised methods outperform the unsupervised methods by a large margin.

Following the similar paradigm of the one used in Section 5.3. The detailed model configuration for *CNN-SR* used in this dataset can be represented as: Input($35 \times 35 \times 3$) − C($30 \times 30 \times 32$) − M($15 \times 15 \times 32$) − C($12 \times 12 \times 32$) − M($6 \times 6 \times 32$)−Dense(1024) − Dense(1024) −Dense(289). The ground truth region's radius r is set as 8, and the detailed quantitative comparison results are presented in Table 2. It can be seen that *Ours* achieves the best overall performance with the lowest counting error and the highest precision and F_1 score. It also demonstrates promising detection accuracy with a low mean μ_d and standard deviation σ_d of the detection distance error. Our conference work *CNN-SR*, which utilizes a dense sliding window based structured regression, also achieves similar performance in this dataset. *NERS* can also achieve promising results. This is because there are few overlapping cells, and the cell boundaries, although weak, are mostly well defined such that it can lead to a high quality extremal region candidate pool.

Two example images marked with detection results of *Ours* are shown in Fig. 5.

5.5. Breast Cancer

The H&E stained breast cancer microscopy images usually exhibit a large portion of inhomogeneous background noise and the cell morphology including size and shape, also vary to a large degree. Since the highly complex nature of the H&E breast cancer images significantly challenges the unsupervised methods, which are unable to produce comparable results, we exclude them in this experiment. Meanwhile, we also find it difficult to make *NERS* work on this dataset, so it is excluded from the comparison as well.

Similar to the one used in Section 5.3. The detailed model configuration for *CNN-SR* is summarized as: Input(55 × 55 × 3) − C(50 × 50 × 32) − M(25 × 25 × 32) − C(22 × 22 × 32) − M(11 × 11 × 32)−Dense(1024) − Dense(1024) −Dense(289).

All the annotations that lie within 7 pixels of the image borders are removed from evaluation to eliminate the effects of incomplete cells, and the radius r of the ground truth circle is set as 15. The detailed comparison results are presented in Table 3. Overall, the comparison results are in favor of *Ours* in terms of precision, recall, and F_1 score. *Ours* also demonstrates strong robustness evidenced by the lowest counting error. *CNN-SR* obtains a slightly smaller detection distance error and a comparable recall. *FCRN-B* from Xie et al. (2015a) are not be able to achieve competitive results on this challenging dataset, partially because of the massive information loss during the pooling operations make it difficult for the network to handle complex background noise and large shape variation. On the other hand, our method utilizes the high resolution image features in the expanding path and achieve much better results. Two example images and the detection results of *Ours* are shown in Fig. 6.

5.6. Bone marrow microscopy

Cell detection and localization in bone marrow microscopy images is also challenging given the fact that images in this dataset contain exhibit significant variations in their sizes and intensities. In this experiment, we include the result reported in Kainz et al. (2015), representing as *Regr. Forest*. Unsupervised methods (*IRV* Parvin et al. (2007), *LoG* Al-Kofahi et al. (2010) and *ITCN* Byun et al. (2006)) are excluded from comparison due to incomparable performance.

The detailed model configuration for *CNN-SR* can be summarized as: Input(41 × 41 × 3) − C(36 × 36 × 32) − M(18 × 18 × 32) − C(14 × 14 × 32) − M(7 × 7 × 32)−Dense(1024) − Dense(512) −Dense(100).

The ground truth region radius r is set as 16, and the comparative cell localization results are detailed in Table 4. As is shown in the table, our method obtain the best overall performance and the lowest running time. *FCRN-A* Xie et al. (2015a), *Regr. Forest* and *CNN-SR* achieve similar performance in terms of F_1 score. Both of *CNN-SR* and *Ours* show strong localization accuracy with low mean (μ_d) and standard deviation (σ_d) of the detection distance error. Nevertheless, *CNN-SR* produces a larger portion of false positive compared to *Ours* evidenced by the relatively smaller precision and higher counting error. The detection results of *Ours* on three sample images are shown in Fig. 7.

6. Conclusion

In this paper, we present a structured regression model using our proposed fully residual convolutional neural network for robust and efficient cell detection in microscopy images. Our method is particularly suitable to process large-size images containing dense cell clusters, large variations of cell morphology and inhomogeneous background noise. We conduct extensive experiments using four datasets representing different image modalities and image acquisition techniques, and the experiments demonstrate the effectiveness,

efficiency and generality of our proposed method. For future work, we will exploit the potential of our method on other applications, such as detection of instances (other than cells) in crowded scenes and semantic image labeling; We will also consider extending our network to 3D models and apply it on volumetric medical datasets.

Acknowledgments

We acknowledge the authors of Kainz et al. (2015); Arteta et al. (2012) for sharing their dataset. This research is supported by NIH grant R01 AR065479-02.

References

Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. Biomedical Engineering, IEEE Transactions on. Apr; 2010 57(4): 841–852.

Ali S, Madabhushi A. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. Medical Imaging, IEEE Transactions On. 2012; 31(7): 1448–1460.

Arteta C, Lempitsky V, Noble JA, Zisserman A. Learning to detect cells using non-overlapping extremal regions. Medical Image Computing and Computer-Assisted Intervention MICCAI 2012. 2012; 7510:348–356.

Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow IJ, Bergeron A, Bouchard N, Bengio Y. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop. 2012

Bergstra J, , Breuleux O, , Bastien F, , Lamblin P, , Pascanu R, , Desjardins G, , Turian J, , Warde-Farley D, , Bengio Y. Theano: a CPU and GPU math expression compiler; Proceedings of the Python for Scientific Computing Conference (SciPy); Jun, 2010

Bernardis E, Yu SX. Finding dots: segmentation as popping out regions from boundaries. CVPR. 2010:199–206.

Bertelli L, , Yu T, , Vu D, , Gokturk B. Kernelized structural SVM learning for supervised object segmentation; The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011; 2011 21532160

Byun J, Verardo MR, Sumengen B, Lewis G, Manjunath BS, Fisher SK. Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. Mol. Vis. 2006; 12:949–960. [PubMed: 16943767]

Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. 2016 Vol. abs/1606.00915.

Chollet F. keras. GitHub. 2015 https://github.com/fchollet/keras

Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D'Anastasi M, Sommer WH, Ahmadi S-A, Menze BH. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. 2016:415–423.

Ciresan D, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. Medical Image Computing and Computer-Assisted Intervention MICCAI 2013. 2013; 8150:411–418.

Ciresan DC, , Giusti A, , Gambardella LM, , Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images; 26th Annual Conference on Neural Information Processing Systems NIPS 2012; 2012 28522860

Clevert D-A, , Unterthiner T, , Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus); International Conference on Learning Representations; 2015

Cosatto E, , Miller M, , Graf H, , Meyer J. Grading nuclear pleomorphism on histological micrographs; Pattern Recognition, 2008. ICPR 2008. 19th International Conference on; 2008 14

Cruz-Roa A, Arevalo-Ovalle J, Madabhushi A, Gonzalez-Osorio F. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Medical Image Computing and Computer-Assisted Intervention MICCAI 2013. 2013; 8150:403–410.

Cruz-Roa AA, Ovalle JEA, Madabhushi A, Osorio FAG. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013. 2013:403–410.

Dhall D, Mertens R, Bresee C, Parakh R, Wang HL, Li M, Dhall G, Colquhoun SD, Ines D, Chung F, Yu R, Nissen NN, Wolin E. Ki-67 proliferative index predicts progression-free survival of patients with well-differentiated ileal neuroendocrine tumors. Human Pathology. 2012; 43:489–495. [PubMed: 21937080]

Fakhry A, Zeng T, Ji S. Residual deconvolutional networks for brain electron microscopy image segmentation. IEEE Transactions on Medical Imaging. 2016; (99):1–1. [PubMed: 26151933]

Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. Biomedical Engineering, IEEE Reviews in. 2009; 2:147–171.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv prepring arXiv: 1506.01497. 2015

He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Computer Vision–ECCV 2016. 2016

Irshad H. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. Journal of Pathology Informatics. 2013; 4:10. [PubMed: 23858385]

Kainz P, Urschler M, Schulter S, Wohlhart P, Lepetit V. You should use regression to detect cells. Medical Image Computing and Computer- Assisted Intervention -MICCAI 2015. 2015; 9351:276–283.

Kong H, Gurcan MN, Belkacem-Boussaid K. Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting. IEEE Trans. Med. Imaging. 2011; 30(9):1661–1677. [PubMed: 21486712]

Laina I, , Rupprecht C, , Belagiannis V, , Tombari F, , Navab N. Deeper depth prediction with fully convolutional residual networks; 2016 International Conference on 3D Vision, 3DV 2016; Stanford, California, USA. 2016

LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient- Based Learning Applied to Document Recognition. Proceedings of the IEEE. Nov; 1998 86(11):2278–2324.

Li R, Zhang W, Suk H-I, Wang L, Li J, Shen D, Ji S. Deep learning based imaging data completion for improved brain disease diagnosis. Medical Image Computing and Computer-Assisted Intervention MICCAI 2014. 2014; 8675:305–312.

Liao S, Gao Y, Oto A, Shen D. Representation learning: A unified deep learning framework for automatic prostate mr segmentation. Medical Image Computing and Computer-Assisted Intervention MICCAI 2013. 2013; 8150:254–261.

Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. CVPR (to appear). Nov.2015

National Cancer Institute. The cancer genome atlas. 2013 retrieved from https://tcga-data.nci.nih.gov

Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. arXiv preprint arXiv:1505.04366. 2015

Parvin B, Yang Q, Han J, Chang H, Rydberg B, Barcellos-Hoff MH. Iterative voting for inference of structural saliency and characterization of subcellular events. IEEE Transactions on Image Processing. 2007; 16(3):615–623. [PubMed: 17357723]

Qi X, Xing F, Foran DJ, Yang L. Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. IEEE Trans. Biomed. Engineering. 2012; 59(3):754–765.

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015:234–241.

Shah A, Kadam E, Shah H, Shinde S. Deep residual networks with exponential linear unit. 2016 CoRR abs/1604.04112.

Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014 CoRR abs/1409.1556.

Sirinukunwattana K, , Ahmed Raza SE, , Tsang Y-W, , Snead D, , Cree I, , Rajpoot N. Patch-Based Techniques in Medical Imaging: First International Workshop, Patch-MI 2015; Held in Conjunction with MICCAI 2015; Munich, Germany. October 9, 2015; 2015 154162Revised Selected Papers

Sirinukunwattana K, Raza S, Tsang YW, Snead D, Cree I, Rajpoot N. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Transactions on Medical Imaging PP. 2016; (99)

Sironi A, , Lepetit V, , Fua P. Multiscale centerline detection by learning a scale-space distance transform; 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014; Columbus, OH, USA. June 23–28, 2014; 2014 26972704

Su H, Xing F, Kong X, Xie Y, Zhang S, Yang L. Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. 2015; 9351:383–390.

Szegedy C, , Ioffe S, , Vanhoucke V, , Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. ICLR 2016 Workshop 2016 URL https://arxiv.org/abs/1602.07261

Veta M, Pluim J, van Diest P, Viergever M. Breast cancer histopathology image analysis: A review. IEEE Transactions on Biomedical Engineering. May; 2014 61(5):1400–1411. [PubMed: 24759275]

Vink JP, Van Leeuwen MB. V. D. C. D. H. G. Efficient nucleus detector in histopathology images. Journal of Microscopy. 2012; 249:124–135. [PubMed: 23252774]

Wang Y, , Sun Z, , Liu C, , Peng W, , Zhang J. Mri image segmentation by fully convolutional networks; 2016 IEEE International Conference on Mechatronics and Automation; 2016 16971702

Xie S, , Tu Z. Holistically-nested edge detection; Proceedings of the IEEE International Conference on Computer Vision; 2015 13951403

Xie W, Noble JA, Zisserman A. Microscopy cell counting with fully convolutional regression networks. MICCAI 1st Workshop on Deep Learning in Medical Image Analysis. 2015a

Xie Y, Kong X, Xing F, Liu F, Su H, Yang L. Deep voting: A robust approach toward nucleus localization in microscopy images. Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. 2015b; 9351:374–382.

Xie Y, Xing F, Kong X, Su H, Yang L. Beyond classification: Structured regression for robust cell detection using convolutional neural network. Medical Image Computing and Computer-Assisted Intervention -MICCAI 2015. 2015c; 9351:358–365.

Xing F, Su H, Yang L. An integrated framework for automatic ki-67 scoring in pancreatic neuroendocrine tumor. Medical Image Computing and Computer-Assisted Intervention -MICCAI 2013. 2013:436443.

Xing F, Xie Y, Yang L. An automatic learning-based framework for robust nucleus segmentation. Medical Imaging, IEEE Transactions on PP. 2015; (99):1–1.

Xing F, Yang L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. accepted to Biomedical Engineering, IEEE Reviews in PP. 2016; (99):1–1.

Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, Madabhushi A. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging. 2015; 35(1):119–130. [PubMed: 26208307]

Yang J, , Price B, , Cohen S, , Lee H, , Yang M-H. Object contour detection with a fully convolutional encoder-decoder network; Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on; 2016

Yang L, Tuzel O, Meer P, Foran D. Automatic image analysis of histopathology specimens using concave vertex graph. Medical Image Computing and Computer-Assisted Intervention MICCAI 2008. 2008; 5241:833–841.

Zeiler MD. ADADELTA: an adaptive learning rate method. 2012 CoRR abs/1212.5701.

Zhang C, Yarkony J, Hamprecht FA. Cell detection and segmentation using correlation clustering. Medical Image Computing and Computer-Assisted Intervention -MICCAI 2014. 2014; 8673:9–16.

Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z, Du D, Huang C, Torr PHS. Conditional random fields as recurrent neural networks. ICCV. 2015

Highllights

- A highly efficient and effective fully residual convolutional neural network is proposed for cell detection.
- We validate the superiority of structured regression over the conventional pixel wise classification method for cell detection.
- We prove the robustness and generalization capability of our model using four datasets, each corresponding to a distinct staining method or image modality.

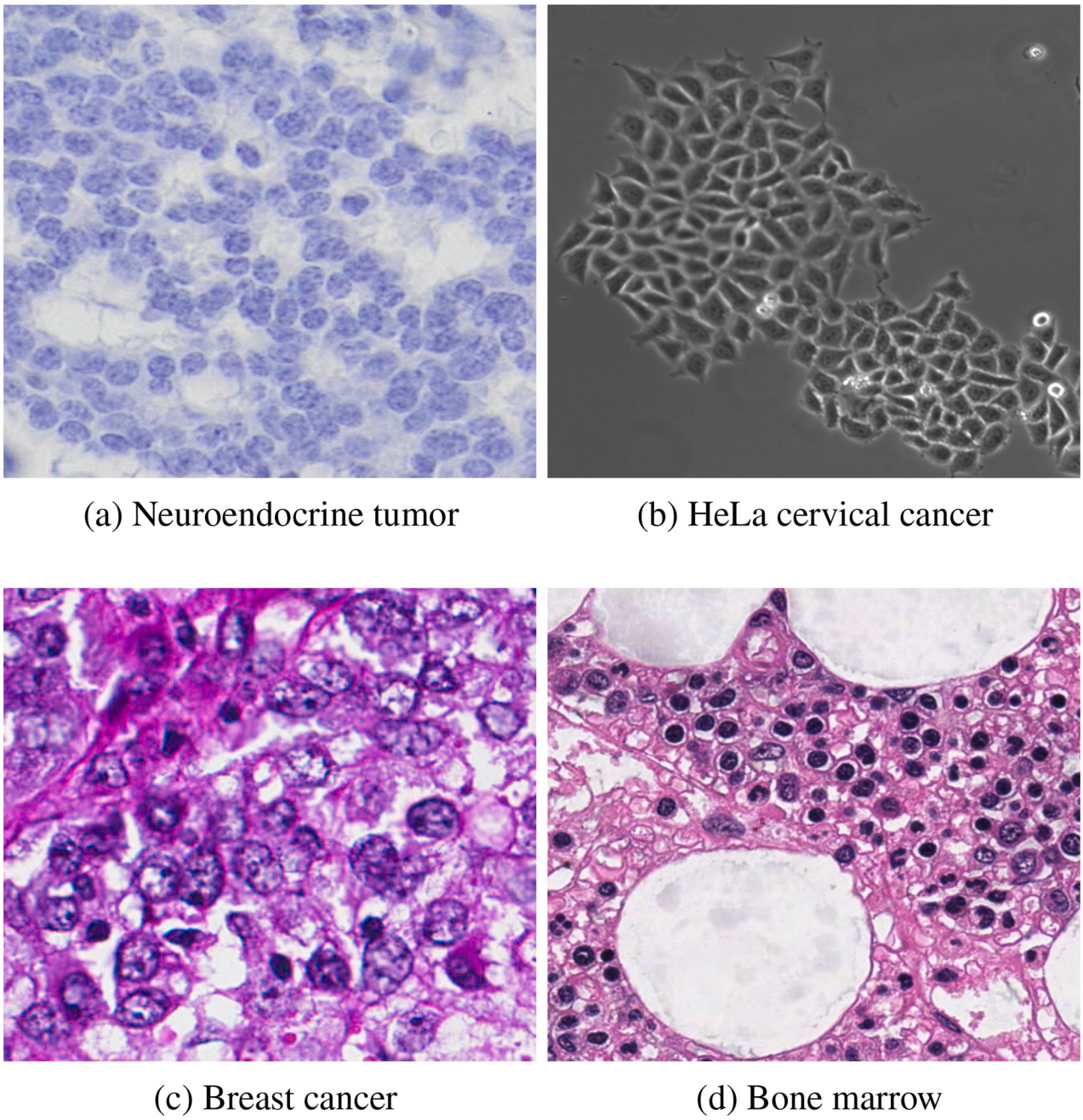


Fig 1.
Sample images of the four datasets used in this paper.

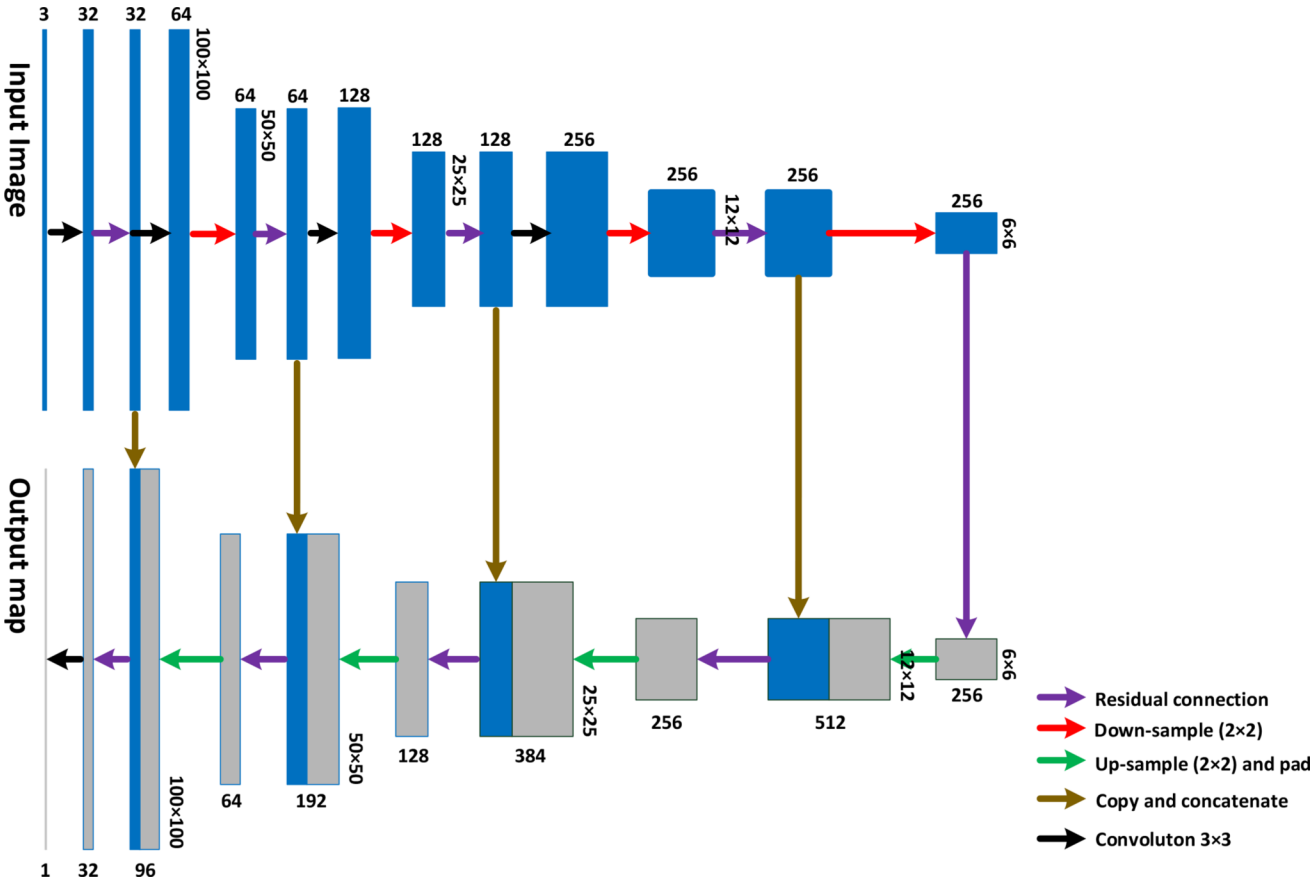
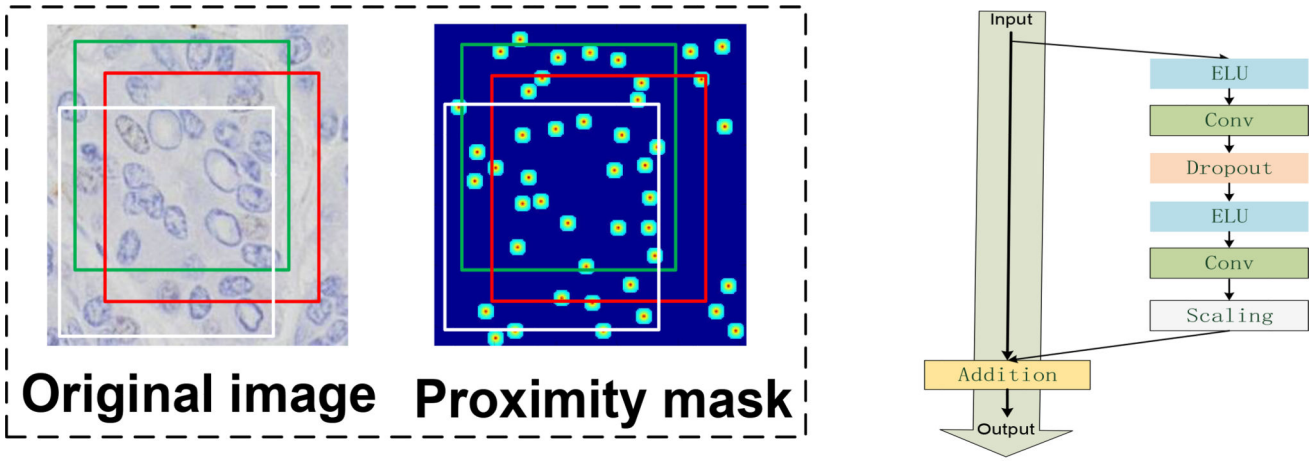


Fig 2.

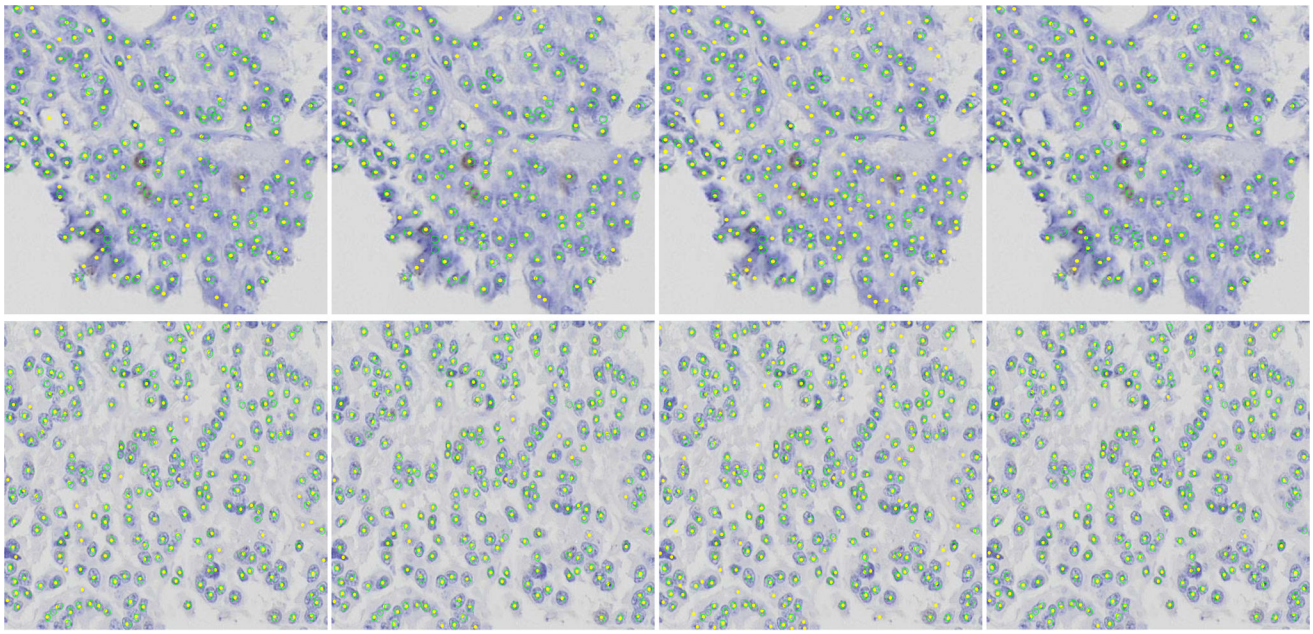
Fully residual convolutional neural network architecture (Please note that the input image size does not need to be fixed value). The blue or gray box denote the feature maps, the number of feature map channel is marked on top or bottom of each box. The feature maps size (along row and column dimension) is denoted on the right-hand side of each box. Different operations are denoted using arrows with different colors. For the details of each type of connection, please refer to Section 4



(a) Training data generation (b) Residual block

Fig 3.

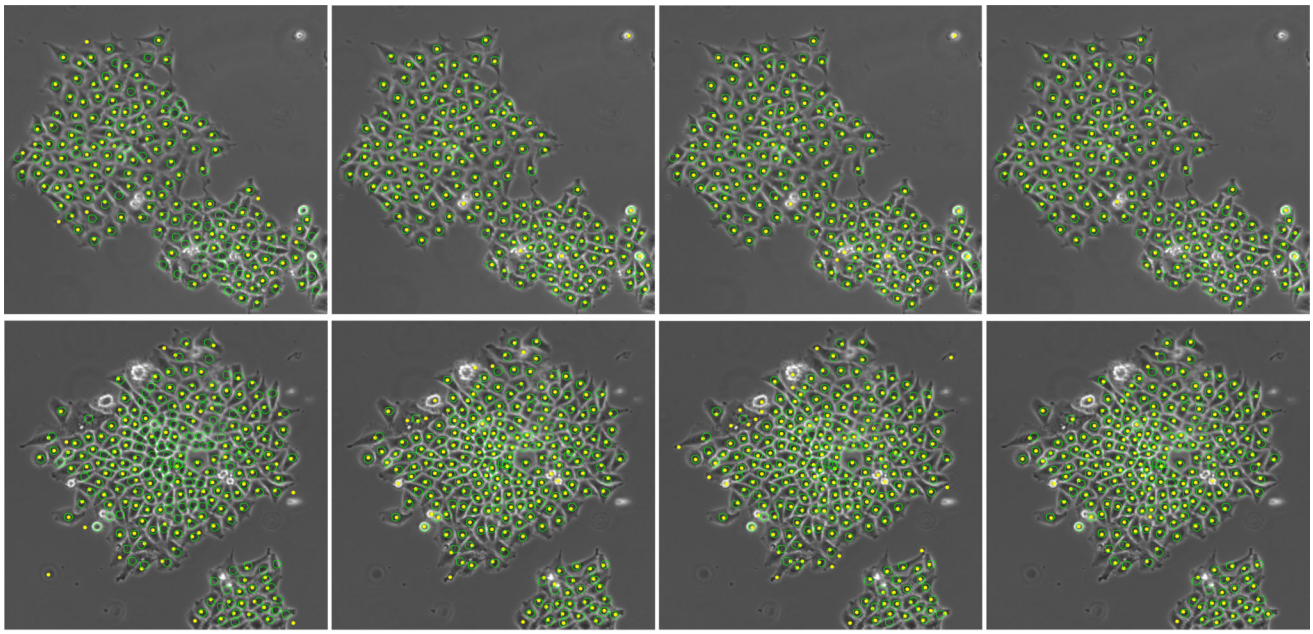
(a): The training data generation process. Each original image has a corresponding proximity mask that has the same size and each cropped local image patch (illustrated as colorful rectangle) has a proximity patch serving as structured *label*. Please note that the human annotations only contain one dot near the cell centers. (b) The residual blocks used in our network, where conv and ELU denote convolution and exponential rectified units Clevert et al. (2015), respectively. The residual output is scaled down before being added to the input.



(a) FCRN-A results (b) FCRN-B results (c) CNN-SR results (d) Ours results

Fig 4.

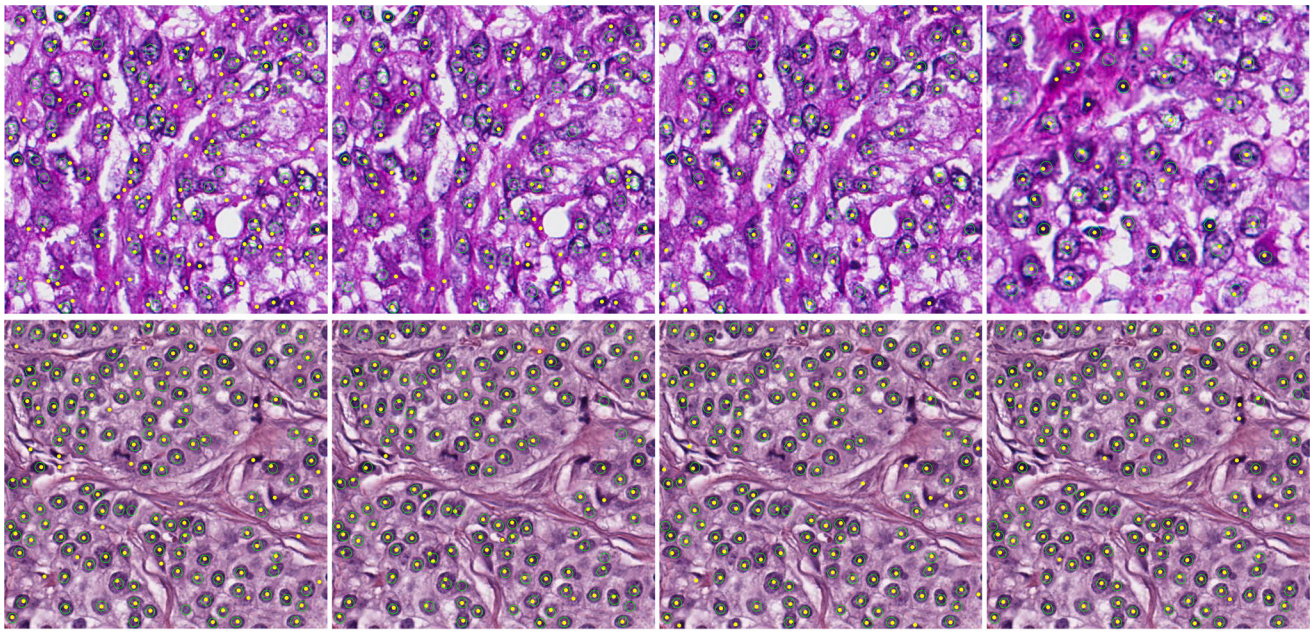
Example cell detection results in neuroendocrine tumor microscopy dataset. Each row represents one testing image. The detected cells are marked by yellow dots, while the ground truth are represented as green circles for better visualization.



(a) FCRN-A results (b) FCRN-B results (c) CNN-SR results (d) Ours results

Fig 5.

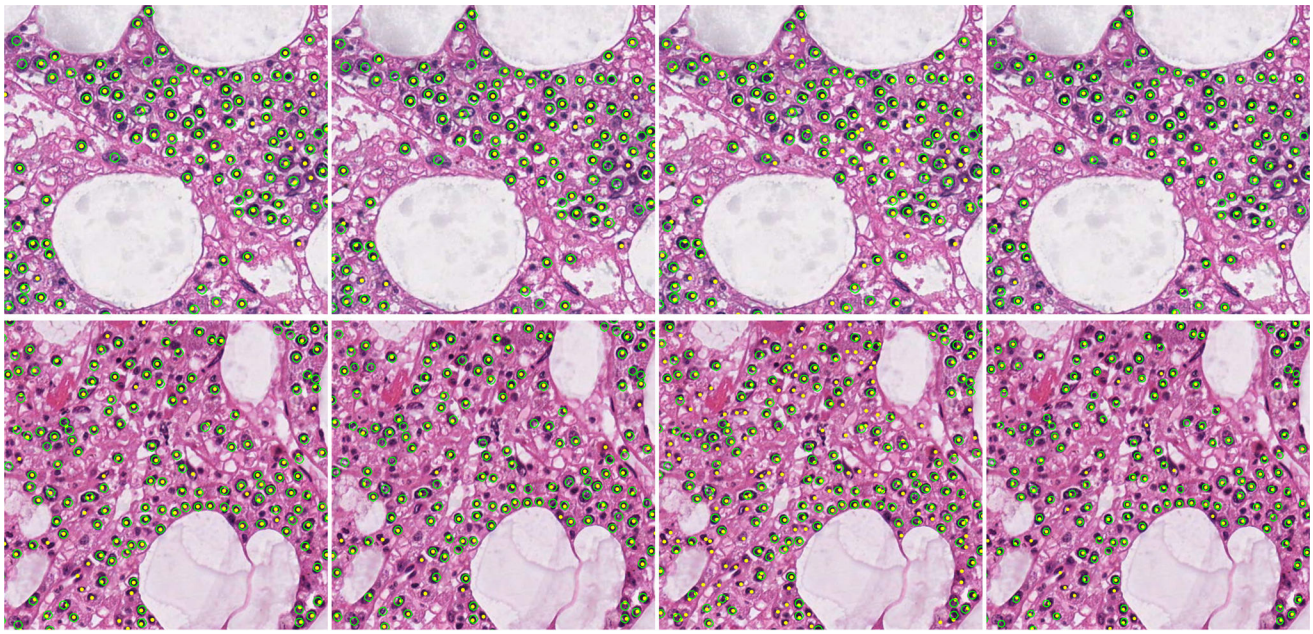
Example HeLa cervical cancer cells detection results in phase contrast microscopy images. Each row represents one testing image. The detected cells are marked by yellow dots, while the ground truth are represented as green circles for better visualization.



(a) FCRN-A results (b) FCRN-B results (c) CNN-SR results (d) Ours results

Fig 6.

Example breast cancer cell detection results in H&E stained microscopy images. Each row represents one testing image. The detected cells are marked by yellow dots, while the ground truth are represented as green circles for better visualization.



(a) FCRN-A results (b) FCRN-B results (c) CNN-SR results (d) Ours results

Fig 7.

Example cell detection results in one patch of the bone marrow microscopy images. Each row represents one testing image. The detected cells are marked by yellow dots, while the ground truth are represented as green circles for better visualization.

Table 1

The comparative cell detection results in neuroendocrine tumor (NET) microscopy image dataset. μ_d, σ_d represent the mean and standard deviation of detection accuracy (d), and μ_c, σ_c represent the mean and standard deviation of counting error (c). T represents the average running time over all testing images (measured in second).

	P	R	F_1	$\mu_d \pm \sigma_d$	$\mu_c \pm \sigma_c$	T
Ours	0.94	0.92	0.93	2.43 ± 1.46	4.87 ± 6.82	0.19
CNN-SR	0.87	0.96	0.91	1.88 ± 1.39	8.23 ± 10.81	54.66
FCRN-A	0.88	0.94	0.91	2.38 ± 1.35	5.01 ± 5.03	0.18
FCRN-B	0.91	0.93	0.92	2.35 ± 1.36	6.13 ± 7.21	0.36
NERS	0.89	0.67	0.75	3.39 ± 3.25	28.47 ± 46.45	-
IRV	0.86	0.75	0.78	3.54 ± 2.79	13.1 ± 14.24	-
LoG	0.78	0.88	0.82	3.49 ± 2.54	12.47 ± 22.88	-
ITCN	0.85	0.75	0.78	4.34 ± 2.98	21.9 ± 36.76	-

Table 2

The comparative cell detection results in HeLa cervical cancer phase contrast microscopy image dataset. μ_d, σ_d represent the mean and standard deviation of detection accuracy, and μ_c, σ_c represent the mean and standard deviation of the counting error. T represents the average running time over all testing images (measured in second).

	P	R	F_1	$\mu_d \pm \sigma_d$	$\mu_c \pm \sigma_c$	T
Ours	0.98	0.98	0.98	2.63 ± 1.43	1.36 ± 1.67	0.31
CNN-SR	0.96	0.98	0.97	2.13 ± 1.32	2.18 ± 3.82	28.82
FCRN-A	0.91	0.61	0.72	3.24 ± 1.7	32.9 ± 21.9	0.16
FCRN-B	0.96	0.98	0.97	2.54 ± 1.42	1.91 ± 1.38	1.1
NERS	0.96	0.92	0.94	2.28 ± 1.46	9.10 ± 9.83	-
IRV	0.85	0.49	0.61	3.07 ± 1.81	54.36 ± 40.06	-
LoG	0.69	0.77	0.72	3.65 ± 1.78	20.82 ± 13.91	-
ITCN	0.66	0.30	0.39	2.81 ± 1.7	71.72 ± 41.63	-

Table 3

The comparative cell detection results on breast cancer microscopy data set. μ_d, σ_d represent the mean and standard deviation of detection accuracy, and μ_c, σ_c represent the mean and standard deviation of the counting error. T represents the average running time over all testing images (measured in second).

	P	R	F_1	$\mu_d \pm \sigma_d$	$\mu_c \pm \sigma_c$	T
Ours	0.89	0.91	0.90	2.8 ± 1.44	6.06 ± 4.96	0.38
CNN-SR	0.86	0.91	0.88	2.36 ± 1.56	8.4 ± 5.78	67.52
FCRN-A	0.71	0.8	0.73	3.06 ± 1.55	21.8 ± 21.8	0.21
FCRN-B	0.69	0.82	0.73	3.42 ± 1.98	24.2 ± 28.9	0.75

Table 4

The comparative cell detection results in bone marrow microscopy image dataset. μ_d, σ_d represent the mean and standard deviation of detection accuracy, and μ_c, σ_c represent the mean and standard deviation of the counting error. T represents the average running time over all testing images (measured in second).

	P	R	F_1	$\mu_d \pm \sigma_d$	$\mu_c \pm \sigma_c$	T
Ours	0.86	0.94	0.90	2.69 ± 1.59	36.3 ± 19.4	2.18
CNN-SR	0.82	0.95	0.88	2.34 ± 1.5	58.67 ± 25.37	499.7
FCRN-A	0.84	0.93	0.88	2.79 ± 1.67	44.3 ± 28.6	2.89
FCRN-B	0.84	0.82	0.83	2.92 ± 1.8	16 ± 1.63	8.27
Regr. Forest	-	-	0.88	-	-	15