

Received February 8, 2020, accepted March 24, 2020, date of publication April 1, 2020, date of current version May 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984996

Point Supervised Extended Scenario Nuclear Analysis Framework Based on LSTM-CFCN

DONG SUI¹, MAOZU GUO¹, FEI YANG², AND LEI ZHANG³

¹School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

²School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 276000, China

³Diagnostic Radiology and Nuclear Medicine, University of Maryland, Baltimore, MD 21201, USA

Corresponding author: Dong Sui (suidongcs2016@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702026, in part by the Natural Science Foundation of Shandong Province under Grant ZR2019MF011, and in part by the China Postdoctoral Science Foundation Project under Grant 2017M622210.

ABSTRACT Cells and cell like particles detection and segmentation are of significant interest to many biological and clinical studies. Traditionally, these tasks are usually performed by visual inspection, which is time consuming, labor intensity and prone to induce subjective bias between different people. These make automatic cell analysis protocols essential for large-scale and objective studies. In recent years, imaging technical has been significantly advanced following the great success by computer vision. In addition, these technologies enable the cross module microscopy analysis, and make the task of cell analysis extremely challenging. Over these decades, computer aided cell detection, counting and segmentation have evolved from earlier filter based methods to the state-of-art deep learning protocols. However, there are still few suitable frameworks that can process multiple source cell images at the same time. In this paper, we seek a different route and propose a novel efficient framework for robust cell analysis based on Long Short Term Memory Channeled Fully Convolution Neural Networks (LSTM-CFCN). The results demonstrates that our framework is able to perform most of cell detection, counting and segmentation tasks from different cell type, and it can also cover most kinds of microscopy images scenarios including dark field, bright field, pathological and electron images. We have perform substantial experiments on several benchmark datasets, the LSTM-CFCN achieves the highest or at least top-2 performance in terms of F1-score, compared with other state-of-the-art methods.

INDEX TERMS Cell analysis, LSTM-CFCN, multi-task learning, multi-scale image processing.

I. INTRODUCTION

Crowded scene object detection and segmentation is an extremely time consuming and tedious task (especially in videos and images), and has attracted great attentions during the past decades. For biological scientists and clinical researchers, cells' and cell like particles' analysis are highest priority tasks for many studies, such as cancer region and cells detection in pathological images, high content and high throughput gene and protein expression analysis, cell morphology researches, molecular markers and phenotype quantification [1]. A efficient cell analysis protocol can reveal critical information that related with the cell physiological status and progress of disease infectious. Such as density

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Afzal¹.

estimation for $s/9$ cells can monitor the cells' state of health. Especially in the progress of cancer detection, for invasion and immune response analyzing, cells migration and deformation in the level of quantity and quality are also play very important role [2].

Traditionally, the cells counting and segmentation tasks are usually performed manually, this leads to extremely tedious and prone to errors with intra- or inter-individual variability. Automated cell analysis has the benefit of reducing time consuming and labor cost, minimizing errors caused by different persons, furthermore it can improving consistency of results between individuals and clinics. To simplify the task and improve robustness, many scientists stain the cells prior to automatic counting [3], [4], however, this is not feasible for many applications (such as red cells and $s/9$ cell in bright field). Thus, there is a significant interest and great demand

to analysis the position and area distribution of cell nuclei or cell like particles with high accuracy and performance.

With the great progress in object detection and recognition in computer vision community, researchers have developed many methods to perform the same tasks [5], [6]. However, these methods still cannot deal with irregular shape cells and only become effective on single cell category [4]. Moreover, cell analysis task is not simply a general object detection and segmentation in computer vision, which typically deals with extended objects, such as humans and vehicles assistant. In the past decades, developments of machine learning technology promote amounts of supervised learning methods for solving this problem, for example, R-CNN, Fast R-CNN, Faster R-CNN and FCN have become the state of the art algorithms for the extended object detection problem, but most methods can only deal with specific kinds of cells, and can not easy to transfer to small object localization [7]–[11].

In this paper, we consider 3 challenges in the application of deep learning method to cell image analysis: (i) cell detection with point labeled images, (ii) detection of cells or cell like particles from different scales, (iii) jointly of cell detection and segmentation. Particularly, based on these challenges, we introduce a novel multi-task cell analysis framework based on the fully convolution networks. The major contributions are as follows:

- We propose a novel framework for jointing detection, counting and segmentation of extended types cells or cell like particles by using global to local information and their fusion results.
- We proposed an LSTM-CFCN network structural to process the multi-scale images combined with adaptive Gaussian filling mapping protocol to establish a point labeled training process, followed by a NMS processing, we achieve the cells detection and counting.
- We propose a Conditional Random Field (CRF) smoothing protocol for the output of LSTM-CFCN, after several iteration steps on these datasets, it finally accomplish the image level segmentation for the multiple scale cells.

The remainder of this paper is organized as follows. Section II describes related works and Section III describes the materials and proposed method with experiments designing. Experimental results and discussions about the advantages and limitations are given in Section IV, as well as future research directions. Conclusions are drawn in Section VI.

II. RELATED WORKS

Nucleus analysis has attracted many research interests. Most existing methods for this task can be classified into the following categories: non-learning methods including thresholding, morphological operations, region growing, level sets, and graph-cuts. Learning methods give priority to supervised learning include SVM, random forests and deep learning-based methods. Unfortunately, the aforementioned methods commonly focused on specific images or cell types. Furthermore, there is a substantial interest in development of a method that is not dependent on specific images and capable

of detection across a wide range of cell/particle types. These methods are reviewed as follows:

A. LOW LEVEL FEATURE BASED CELL ANALYSIS METHODS

Before deep learning stage, many cell analysis methods have been proposed, and most of them are based on segmentation strategy, a comprehensive review can be found in [12]. Traditionally, the basic image process protocol such as image thresholding, feature selection, image morphological, deformable model and region growth method have been used for the cell analysis. A popular choice is Laplacian of Gaussian filter, it treat the cell as a blob to detect the center, besides this, some texture properties have also been used for the same task [13], [14]. All of the aforementioned methods are following the “hand crafted feature”+“classifier” baseline, such as Sobel operator, SIFT, HOG LBP, and etc [15]–[17].

Besides these, one of the common used to detect markers is DT based methods, which assigns each pixel with the distance to the nearest feature point [3]. Park proposed an improved UE operation to exploit a noise-robust measurement of convexity as the stopping criterion for erosion [18]. HIT is also a popular method based on the morphology operation [19]. Compared with DT using all minima as markers, HIT can remove spurious local minima caused by uneven object shapes or noises and generates correct markers. The MSER detector is also used to locate blob objects. Lu have applied this strategy to nucleus detection in Pap smear microscopy images [20], [21]. However, these criteria require empirical selection of the parameter values, which might limit the algorithms applications.

Filter-based methods are usually correspond to the central regions of nuclei or cells in microscopy images. Cosatto proposed detection of cell nuclei using Difference of Gaussian (DoG) followed by Hough transform to find radially symmetrical shapes [13], [22]. Wang proposed a novel filter based on nonlinear transformed sliding band filter which performs well in inset cells images [23]. However, these methods can not be applied to clumping and overlapping cells.

B. LEARNING BASED CELL ANALYSIS METHODS

Supervised learning based methods have attracted much attention for depicting the complex nature of pathological processes from images, especially on histopathology. These methods are mainly classified into two categories, the shallow learning methods and deep learning methods.

For the shallow learning methods, Su applied a binary SVM classifier to automatic mononuclear detection in isolated single muscle fiber fluorescence images [24]. Mao proposed a supervised learning method for nucleus detection and segmentation in bladder cancer images [25]. Random forest is another popular method due to its fast training and testing, and fair error tolerance in training data [26]. Mualla applied a random forest to cell detection in microscopy images [27]. Gall have introduced a general object detection algorithm,

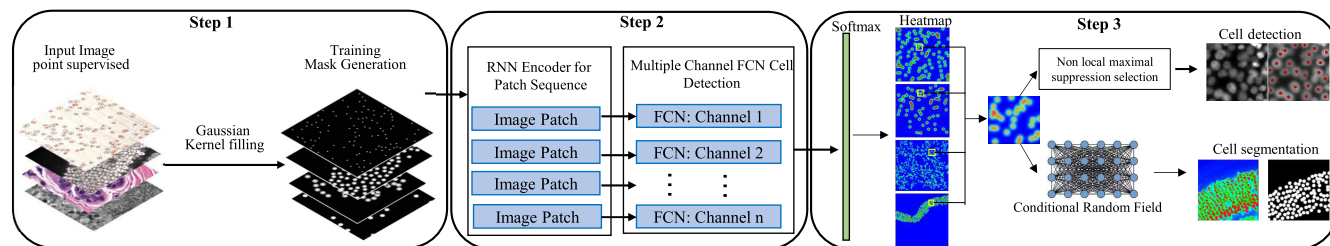


FIGURE 1. The workflow of the LSTM-CFCN multi-task cell analysis framework.

namely Hough forests, which constructs discriminative class-specific part appearance codebooks based on random forests that are able to cast probabilistic votes within the HT framework [28].

For deep learning-based models, especially CNN [29], have attracted particular attention for applying to the problem of cell counting and detection recently [30]–[32]. Different from SVM and random forests that rely on hand crafted features for object classification, CNN can automatically learn multi-level hierarchies of features that are invariant to irrelevant variations of samples while preserving relevant information [31], [33]–[35]. However, the networks were trained to perform in well-controlled environments, with clean background and little cell overlap on synthetic images. Shelhamer proposed an end-to-end FCN [36], [37] method to classify the pixels at image level so that to solve the semantic segmentation problem at pixel level, and can accept any size of the input image by using the deconvolution layer to the last convolution layer. The feature map is up-sampled so that it returns to the same size as the input image. A prediction can be made for each pixel while preserving the spatial information in the original input image, and finally classify the pixels. Xie proposed a FCN-based method to predict counting numbers in RPE cells, and get an ideal result [31].

III. METHODS

A. OVERVIEW

In the proposed framework, we take the multi-task scenario into consideration. we combine cell detection and segmentation into one scheme at pixel level to scale the whole images, and choose a recurrent neuron networks encoder followed by fully convolution networks structure combined with devolution part to keep the patches' sequence and integration of the whole image. Each channel of the FCN can extract the total features of a sequential patches separately and concat them into a whole image. Finally, we evaluate the results on multiple source image data sets. Details are illustrated in Figure 1, The total scheme can be divide into 3 parts as follows:

In step1, we mainly focus on the training data preparation. We designed a Gaussian kernel to fill the dot label cell images, in this way the cell mask was then generated for the networks training. In step2, we design a recurrent neuron network(RNN) patches sequence encoder and a stack

channel fully convolution network (FCN) model to perform the cell feature extraction. The aim of RNN model is to encode the patch sequence into an integrated image. For each image patch, a parallel deep neural networks which share the weights of layers is employed for cell heat map generation, in this part, Deconvnet model is employed to learn the target heat maps. According to the RNN sequence encoder, we integrate the whole images in this step for the following detection and segmentation tasks. In the third part, our aims is to detect the target cells and perform the cell body segmentation jointly. For the cell detection part, after heat map generation, a non maximal suppression method is employed for local maximal searching, which is assumed the center of each cell. For the cell body segmentation, we choose conditional random field model to smooth the heat map area to generate the accurate results.

To verify the proposed pipeline, we choose four types of cells image data, including pathologically stained whole slide images, confocal microscopy retinal cell images, insect cells under bright field microscopy, and pancreatic vesicle cells under projection electron microscopy. For each dataset, the accuracy of detection and segmentation will be listed and compared with the traditional and deep learning methods.

B. DATA SETS

In this paper, we choose four types of cell datasets: (1)breast cancer pathological cells image, (2)insect cells image, (3)vesicles images, (4)retinal cell images. A multiple scale form of these images are listed in Figure 2. To cover the source scope of cell data, breast cancer pathological cell image dataset is sampled from the public Camlyon 17 dataset, retinal cell images is download form the biological image center of UCSB. Others are prepared by ourselves in biology lab. All of these datasets are labeled manually by biological scientists, and more descriptions are as follows:

Breast pathological cells image dataset

The hematoxylin and eosin (H&E) stain is commonly used to screen abnormal cells in pathological researches and clinic diagnosis. It induces sharp blue/pink contrasts across various (sub)cellular structures and it is applied across many different tissue types [38], [39]. In this part, Camlyon 17 dataset is choose as the template, the region contains normal and cancer cells are selected manually, and then the testing and training sets are labeled manually by biological scientist.

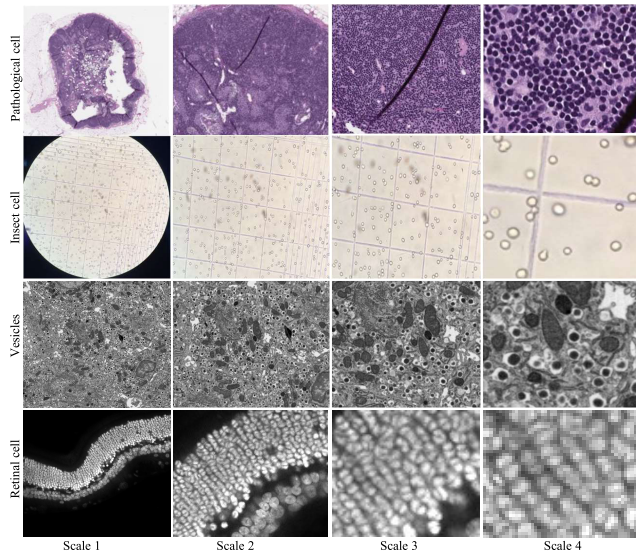


FIGURE 2. Examples of data sets. From the top to the bottom in row, are breast pathological cells image, Insect cell images, Vesicles image and Retinal image. Column indicates 4 scales from each image set.

Insect cells image dataset

There are four type insect cell lines commonly used for application, which support various levels of expression and differential glycosylation with the same recombinant protein [40]. We chose the most widely used cell type *Sf9*, a clonal isolated from the *Spodoptera frugiperda* cell line *IPLB-Sf21-AE*, as the host cells. In this paper, all of the training data is collected from bright field microscopy and labeled.

Vesicles image dataset

Vesicles Slices are generated from a human pancreas cell and stained for electronic microscope (EM) imaging. Focused ion beam scanning electron microscopy imaging of our samples are performed. Briefly, the block face was imaged using an electron beam with an acceleration voltage of 3 keV, a current of 2.75 nA, and a dwell time of 15 μs/pixel. We used to remove the 50-nm-thick for the next slice imaging. As a Z-stack image dataset, collector manually labeled as our training and testing sets.

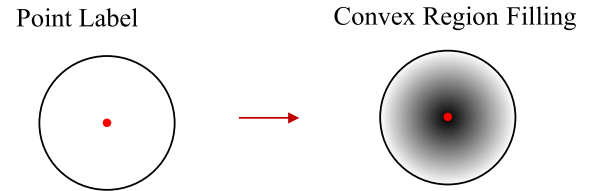
Retinal image dataset

The retinal images datasets are collected by confocal imaging from a 100-μm-thick retinal tissue section for the purpose of understanding the mechanism of the loss and recovery of vision caused by retinal detachment and reattachment [41]. The cells density of photoreceptor has been prove to be close relation with retinal detachment and reattachment [41]. Images dataset used in this paper are from datasets of Retinal Cell Biology Laboratory in Neuroscience Research Institute of University of California [41].

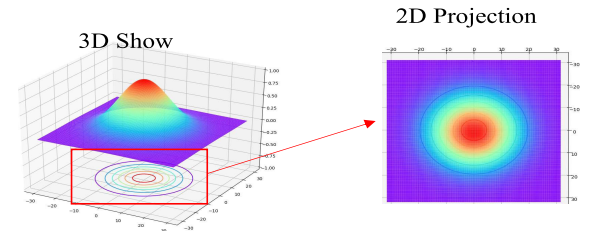
C. TRAINING DATA PREPARATION

1) CELL IMAGE PREPROCESSING

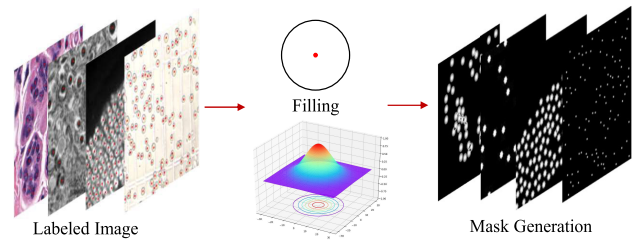
Biomedical images collected from microscopy are with multi-scale and large size, such as the pathological images



(a) Rounded convex region filling scheme around the labeled center



(b) General kernel distribution shown in 2D and 3D



(c) Mask filling process by using the Gaussian kernel

FIGURE 3. Mask generation using Gaussian kernel rounded convex region filling.

from Camlyon17 is up to 2Gb per image. Most of these data cannot be fed into the network directly. We need to split images for several steps' processing. Firstly, active sample mining is performed by the OSTU algorithm to extracted the region of interest(ROI) in the whole images, which can remove the background in the original data. We select the high availability area and took the foreground area as large as possible. Then, based on these ROIs, biological scientists label the cell center with dot manner for preparing the training data.

2) MASK GENERATION

In practice scenario, biological scientists usually label the large number of cells use the dots to mark for each cell nucleus, this seems possible for making image-level annotations. Whereas, this manner cannot cover all of the nucleus' area accurately and efficiently, such as in cell segmentation tasks. With the case of less labeling or incomplete information, we propose a cell area filling process to automatically expand the point label to achieve full cell region mark with weakly supervision. In this situation, we assume that the morphology of the cells is circular, and extend the marker to a circular area based on the size of cell. The cells is filled by a rounded convex region whose radius is near to the cells [4]. Finally, a Gaussian pixel filling protocol is employed to cover the circular area. See figure 3, and the formula 1 describes the

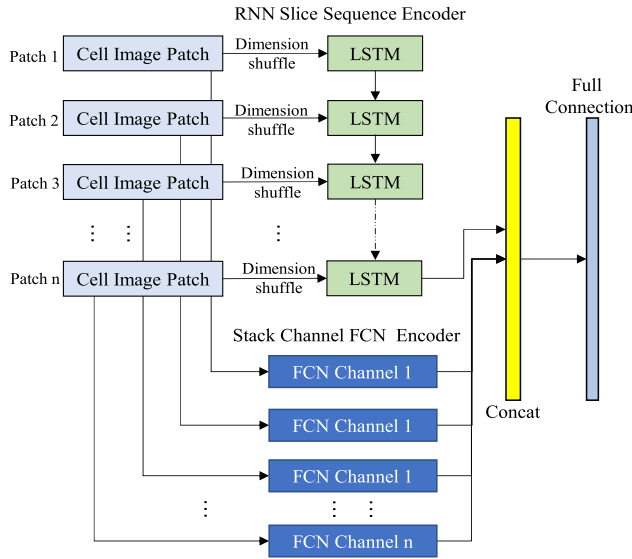


FIGURE 4. Long Short Term Memory enhanced Channeled Fully Convolution Network pipeline with Dimension Shuffle layer.

filling function.

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \left\{ \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\mu_X^2} + \frac{(y-\mu_Y)^2}{\mu_Y^2} - \frac{2\mu(x-\mu_X)(y-\mu_Y)}{\mu_X\mu_Y} \right] \right) \right\}. \quad (1)$$

where, x, y donate the location of labeled dot in the cells center, $\sigma_1, \sigma_2, \mu_X, \mu_Y$ is the variance and mean value respectively, and ρ is a constant located in the interval $(-1, 1)$. The two-dimensional Gaussian kernel function f is used to filling the dot labeled images with centered by the label coordinates, and the final aim is to generate a target density map by filling all-zero image. In this way, the average diameter of the cells obtained by analyzing the image to estimate the average cell diameter. Firstly, we transform all values and scale them into 0-1 to make sure the probability of the center point is 1. In addition, the relatively small value of the core is compressed to the same level as the background by a threshold. The marker thus constructed can well optimize the confidence probability of the kernel, and the further away from the kernel with lower probability. Under the situation of bright field cell images, the cells exhibit rounded circle with a blank, we used a reverse Gaussian filter to filling and adjust these area.

D. NETWORK ARCHITECTURE DESIGN

Generally, the content of a high resolution biomedical image can up to 2Gb, and cells in these images usually locates at a small part of area. As an multiple tasks frame that consisted of object detection and segmentation, most of the model can not cover cell and cell like particles under multi-scale sub-sampling process. This means all of the model have to be trained on the ground scale to perform these tasks. This make

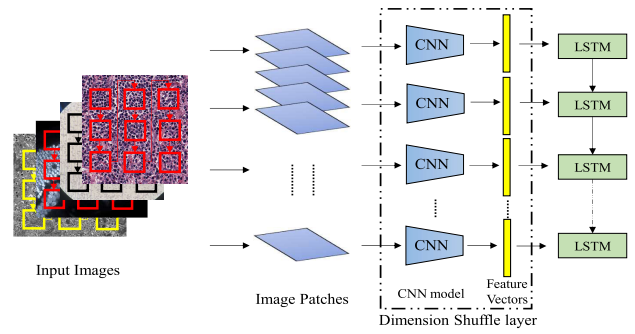


FIGURE 5. Details of Dimension Shuffle layer, functional block is in the dotted box, and CNN model is stacked for the feature extraction.

a biological image usually have to be spliced into thousands of patches, and cells are separated in each of them.

In order to merge these patches into a large integrated image, we proposed a LSTM labeled stack channel FCN scheme to accelerate the processing speed and keeping the integrity of result generation, see figure 4.

In this scheme, we use LSTM network to get a fixed-length feature representation for the whole image feature sequence [42], [43]. The channel FCN block and the LSTM block perceive the same patch input in two different views. The channel FCN block views the patch sequence as a univariate time series with multiple time steps. If the it receive N patches, the channel FCN block will receive the data in N time steps. Contrarily, the LSTM block in our framework receives the input patch sequences as a multivariate time series with a single time step that accomplished by the dimension shuffle layer.

At the same time, the image path sequence are then fed into a stack channel FCN pipeline to perform the cell feature extraction process, details are shown in figure 6. We choose the same structural for each FCN channel to output the probability map with the same size to input image, and indicating the class probability of each pixel. In the convolution part, it has 13 convolution layers, rectification and max pooling are performed between convolutions, and for imposing the class-specific projection, 2 fully connection layers are employed for augmenting. In the deconvolution part, it is a mirror edition of the convolution side, and has several unpooling, deconvolution and rectification layers, which enlarge the activation size with the combination of these layers. Simultaneously, the input image and the target density maps can be connected in the form of end to end. See figure 6.

E. LSTM-CFCN MULTI-TASK LEARNING

As it is stated in the section C, the center of each cell is labeled as dot, the ground truth of cell count for sequent patch i is the total number of labeled dots. In part 2 of section B, we define the counting ground truth density $P_i^0(p)$ for each pixel p in the i th sequent patch as a 2D Gaussian kernel covering area centered at the dotted point, as follows:

$$P_i^0(p) = \sum_{d \in \mathcal{A}_i} N(p; x, \delta), \quad (2)$$

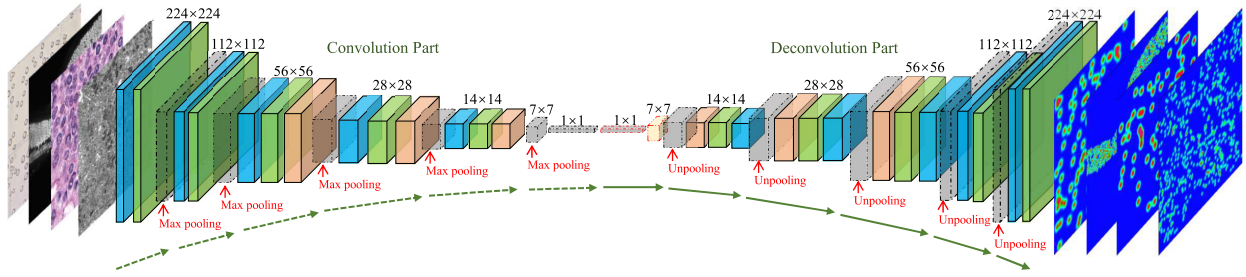


FIGURE 6. The architecture of FCN with deconvolution part in one channel of CFCN.

where \mathcal{A}_i is the set of annotations, x is the pixel of annotation and δ is the parameter of Gaussian kernel decided by the heat map. The Channel FCN aims to estimate the density map of each cell and the LSTM block is to estimate the relative global cell count for each sequent patch. These two blocks a jointly achieved by end to end training protocol of whole LSTM-CFCN network. Finally the cell density is predicted form the feature map generated by the deconvolution part, here we employ euclidean distance to measure the difference between the detected density map and the ground truth. So our loss function is defined as follows:

$$L_{\mathcal{A}} = \frac{1}{2N} \sum_{i=1}^N \sum_{p=1}^P \|F_i(p; \Theta_{FCN}) - F_i^0(p)\|_2^2, \quad (3)$$

where, N is the batch size, $F_i(p)$ is the estimated cell density map at pixel p in the i -th sequential patch, and Θ_{FCN} is the parameter of CFCN.

For the second task, global cell detection from reality whole images, which is learned from the LSTM block. Its total loss function including two parts:(i) basic cell counting: the integration of the heat map for the relative whole image; (ii) residual cell counting: learned by the LSTM layers. To achieve the final aims, a sum of these two part to get the estimated cell counting results:

$$R_i = C(F_i; \Gamma, \Phi) + \sum_{p=1}^P F_i(p), \quad (4)$$

where, $C(F_i; \Gamma, \Phi)$ represents the residual count, F_i is the estimated heat map for the sequential patch i , Γ is the LSTM parameter to be learned, and Φ is the parameters of the fully connected layers. Here we choose the same hypothesize on the optimization protocol with [44]. The loss function of the global cell count is:

$$L_R = \frac{1}{2N} \sum_{i=1}^N (R_i - R_i^0)^2, \quad (5)$$

where, R_i^0 is the ground truth in the i -th image patch, R_i is the estimated cell number of the same patch. Then the total loss function for the LSTM-CFCN is defined as:

$$L = L_{\mathcal{A}} + \xi L_R, \quad (6)$$

where ξ is a weight of global cell counting loss, and to be tuned for the best accuracy. Simultaneously, each task can be trained with fewer parameter to achieve a better training process. In the network, we use batch-based Adam optimizer and backpropagation to optimize the total loss function. Finally, our proposed LSTM-CFCN can process different input cell images with multiple scale and achieve a robustness for different field scenes.

Algorithm 1 Fine-Tuning Algorithm

- 1: **for** $i < T$ **do**
 - 2: $new_model_weights \leftarrow initial_model_weights$;
 - 3: $Train_{phase}(model, initial_{lr}, batch_{size})$;
 - 4: $initial_model_weights \leftarrow new_model_weights$;
 - 5: $i \leftarrow i + 1$;
 - 6: $initial_{lr} \leftarrow updateLR(initial_{lr}, i)$;
 - 7: $batch_{size} \leftarrow updateBatchsize(batch_{size}, i)$;
 - 8: **end for**
-

F. MODEL FINE-TUNING

To make sure our framework suitable to various kind of cell image data set, a transfer learning strategy is employed for the knowledge learned from one large dataset can be reused when training another dataset, such that the domain of these four kind of data have some similarity with each other [45]. For the multiple tasks framework, we split the training procedure into two distinct phases. In the *initial phase*, the optimal hyper parameters are trained on the pathological data set for the initial model, and then we apply the *fine-tuning* to this initial model. we iterate the transfer learning procedure in the *fine-tuning* phase using the original dataset step by step based on the previous model weight, and the learning rate is halved at each iteration step. Furthermore, we also halved the batch size once every alternate iteration with the initial learning rate $1e-4$ and batch size is 32. This procedure is repeated T times, where T is a constant, generally set as 5 in our framework.

G. POST PROCESSING

In step 3, we construct 2 blocks for cell detection and segmentation separately. In the cell detection block, the density map is fed into a Non-Maximum Suppression (NMS) module to screen the cell location. In this paper, the NMS is used to detect the local maximal value, the local maximal is assumed

Algorithm 2 $2D(n+1) \times (n+1)$ -Block NMS

```

1: for  $(k, l) \in \{n, 2n+1, \dots\}^2 \cap [0, W-n] \times [0, H-n]$ 
   do
2:    $(max\ k, max\ l) \leftarrow (k, l)$ ;
3:   for  $(k_1, l_1) \in [k, k+n] \times [l, l+n]$  do
4:     if  $I(k_1, l_1) I(maxk, maxl)$  then
5:        $(max\ k, maxl) \leftarrow (k_1, l_1)$ ;
6:     end if
7:   end for
8:   for  $(k_1, l_1) \in [max\ k - n, max\ k + n] \times [max\ l - n, max\ l + n] - [k, k+n] \times [l, l+n]$  do
9:     if  $I(k_1, l_1) I(maxk, maxl)$  then
10:      goto failed;
11:    end if
12:     $MaxlocAt(max\ k, max\ l)$ 
13:    failed;
14:  end for
15: end for

```

the cell center, and details of the NMS is listed in Algorithm 2. In this block, the cell detection task is performed directly from the heat map of LSTM-CFCN by using NMS on the ground scale of the input images that can make sure all of the original location of each cell being located.

In the cell segmentation block, we intend to achieve a coarse to fine segmentation task by integrating into our framework the fully connected CRF model to overcome the limitations of short-range CRFs [46]. We adopt the following energy function

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \psi_{ij}(x_i, x_j), \quad (7)$$

where,

$$\begin{aligned} \theta_i(x_i) &= -\log P(x_i), \\ \psi_{ij}(x_i, x_j) &= \mu(x_i, x_j) \sum_{m=1}^K \omega_m \cdot k^m(f_i, f_j), \\ \mu(x_i, x_j) &= \begin{cases} 1 & x_i \neq x_j; \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

where x represents the pixels assigned label, $\theta_i(x_i)$ is unary potential, and $P(x_i)$ is the output of DCNN represents the probability of label assignment at the position i , $\psi_{ij}(x_i, x_j)$ is pairwise potential. In this part, the pair term is omitted for the model's factor graph is fully connected. The feature f is extracted for pixel i and j , and ω_m weighted Gaussian kernels k^m , details are as follow:

$$\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right), \quad (8)$$

in which the first kernel depends on pixel color intensities I and pixel positions p , whereas the second kernel depends on

pixel position only and the parameters σ_α , σ_β and σ_γ control the scale of these kernels.

H. EVALUATION METHODS**1) EVALUATION OF CELL DETECTION AND COUNTING**

We evaluate the cell detection and counting results by using the traditional protocol employed in [4], [23], [47], [48]. We compare the results from the proposed LSTM-CFCN framework and other state of the art methods. The performance of detector is evaluated by error criterion according to the following:

$$ER_1 = \frac{1}{N} \frac{\|CD - GT\|}{GT}, \quad (9)$$

$$ER_2 = \frac{1}{C} \sum \frac{1}{N} \frac{\|MD - AV\|}{AV}, \quad (10)$$

$$AV = \frac{1}{C} \sum MD. \quad (11)$$

where, ER_1 represents the absolute relative Error Rate of the cell detection block and ER_2 represents the absolute relative Error Rate of manual counting. N is the image number of the testing set and C represents the number of lab collaborators who count cells. We manually labeled the cell number as the Ground Truth (GT), and the results detected by the cell detection block is defined as Cell Detected (CD). Average number (AV) is the average number of manual counting according to all of the validating dataset. To test the counting effectiveness, we choose some lab collaborators manually detected cells in the images to compare with our cell detection blocks, and the Manual Detected results is defined as MD .

For the comparison, we defined that only the detected centroids lie inside the cells are considered as correct one. The results of our framework and other methods are reported in terms of recall, precision and f1-measure values as follows:

$$Recall = \frac{TP}{TP + FN}, \quad (12)$$

$$Precision = \frac{TP}{TP + FP}, \quad (13)$$

$$F1 = \frac{2PR}{P + R}, \quad (14)$$

2) EVALUATION OF CELL SEGMENTATION

For the cell segmentation part, we evaluated the segmentation performance by using Quelhas' methods in [47]. In this part, to get a hand craft ground truth, we use the Kc167 cell data [47] to evaluate the performance of cell segmentation. We map each detected cell body C_i to fit the ground truth (GT_m). F_β score is used as the coverage measure according to the following:

$$F_\beta(C_i, GT_m) = \frac{(1 + \beta^2)p(C_i, GT_m)r(C_i, GT_m)}{p(C_i, GT_m) + \beta^2r(C_i, GT_m)}, \quad (15)$$

where, p and r represents precision and recall respectively as follows:

$$\begin{aligned} p(C_i, GT_m) &= \frac{C_i \cap GT_m}{C_i}, \\ r(C_i, GT_m) &= \frac{C_i \cap GT_m}{GT_m}, \end{aligned} \quad (16)$$

where β is a parameter for F_β score. In our experiments, we choose F_1 as it is in cell detection part.

IV. RESULTS AND DISCUSSIONS

A. NETWORK TRAINING

All experiments and test bed are carried out on a PC (Intel Core(TM) i9, 12 cores, 2.90GHz processor with 32 GB of RAM) and a NVIDIA GeForce GTX 1080 Graphics Processor Unit. The software implementation is performed using Tensorflow+Keras. During implementation, as Figure 4 and Figure 6.

The proposed models have been tested on all the 4 kinds of data sets, in this part we divided each data set into two parts, 80% and 20% for training and testing, and data augmentations is applied on these data sets. To make sure the stability of the network parameters, the CFCN block is kept constant throughout all of these experiments. We choose the optimal number of LSTM cells by using hyperparameter search within the range from 8 cells to 128 cells. Concerning about the scale of input pathological images and processing efficiency, we finally choose the N with 8 cells. Training epochs' number is generally kept constant at 2000, but the model requires a longer converge time when it confront the pathological dataset for the image is too large. Whereas, some of the feature sequence can be very long and lead to the training become a frustrated process for the reason that there are no enough memory. When the feature sequence is longer than 5000, a random sampling is employed to summarize the sequence. Furthermore, a truncated back propagation through time technique is also applied to the LSTM model for training [49]. Initial batch size is 128, and it is halved in each successive *fine-tuning* iteration. All of the models are trained using the Adam optimizer with the starting learning rate $1e-3$ and the final learning $1e-4$, and it is reduced by a factor of $1/\sqrt[3]{2}$ every 200 epochs of no improvement in the validation score, until the final learning rate is reached.

After the CFCN is fine-tuned, we cross-validate the CRF parameters by using default values of $\omega_2 = 3$, $\sigma_\gamma = 3$ and we search for the best values of ω_1 , σ_α and σ_β by cross-validation on 100 images from the testing dataset. An coarse to fine searching protocol is employed for find a suitable CRF parameters, we choose the initial search range are ω_1 and σ_β from 3 to 6 with step 1, σ_α from 30 to 100 with step 10.

B. RESULTS OF CELL DETECTION

The qualitative cell detection of LSTM-CFCN is shown in figure 7, we apply the framework to carry out cell counting task on these four kinds of data sets. For the cell detection task, our aim is to find the center of each cell and calculate

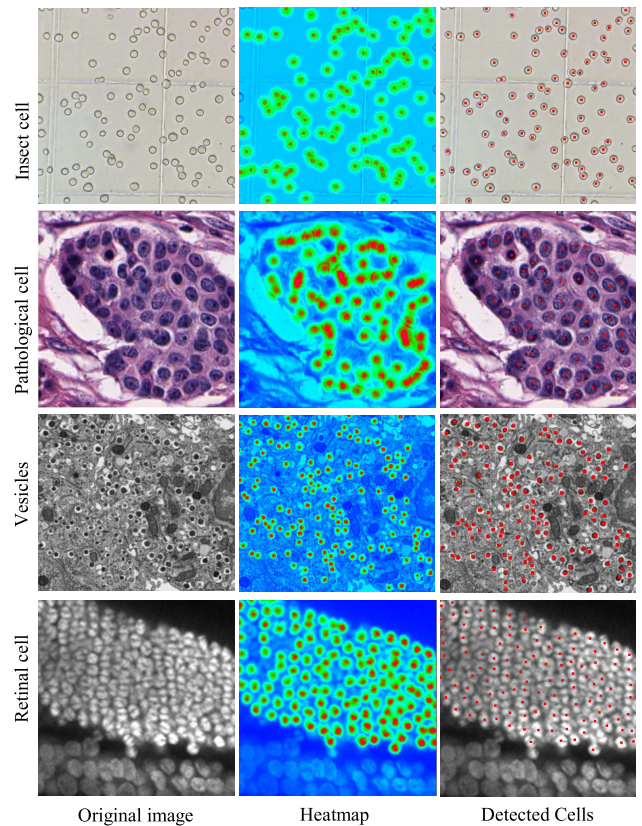


FIGURE 7. Cell detection results for the four kind of data set.

the number of these locations. Here we select several results as an indication about the cell detection process and results,. As shown in figure 7, we firstly use the LSTM-CFCN generate a suitable heat-map of the cell image, this is indicated in the second column of figure 7, and then a 2D non-maximal suppression method is employed for the final cell center detection, as shown in the third column, most of the nuclei are labeled accurately by our method. In this part, for a better comparison, we change the CFCN part with CNN based method such as ResNet50, VGG 16 and U-Net, and these methods are state-of-art in object detection and image segmentation, a traditional LoG methods and SSAE are also used for the comparison [50]–[55]. The prediction recall and F_1 -measure are used to compare proposed detection methods with the others. Some experimental results are listed in 1.

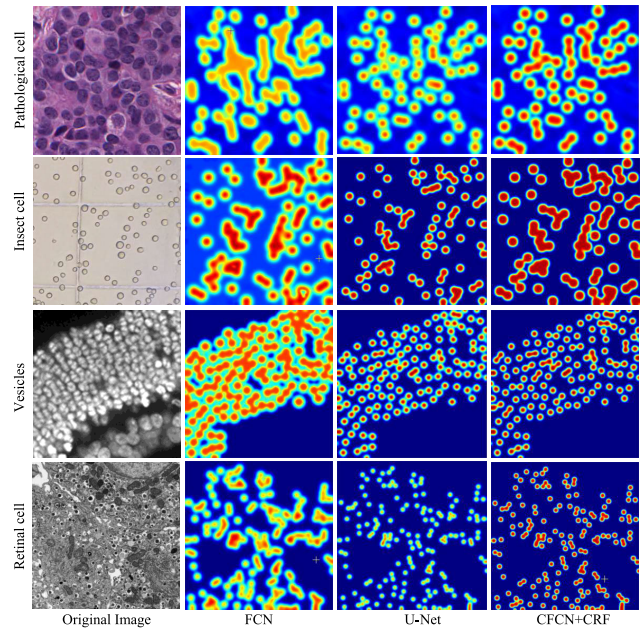
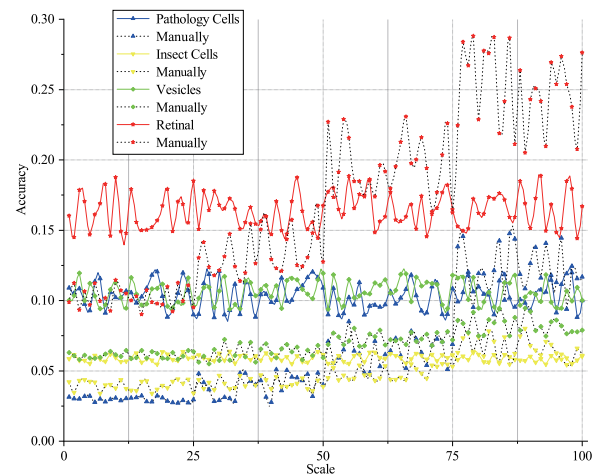
We chose SSAE, ResNet50, VGG 16 and U-Net as the final test bed, each of these model is performed based on the same training and testing data. Totally speaking, see Table 1, our method and U-Net get better results for these 4 kinds data, and the proposed LSTM-CFCN method get the top-3 precision and F1 score. For pathology image, it contains several cell types, and SSAE get a higher result, because this method is designed for pathological images [51]. VGG16, ResNet50 and U-Net get relative lower in precision and F1 score for the limitation of training set. Insect cells, Retinal and Vesicles have relative pure cell component, for these data set deep learning based method exhibited a better

TABLE 1. Results and comparison of cell detection.

Dataset	Methods	Precision(%)	Recall(%)	F1
Pathology	LoG	13.7	42.7	20.7
	SSAE	80.2	83.7	81.9
	VGG 16	67.2	66.1	66.7
	ResNet 50	73.6	68.7	71.1
	U-Net	77.5	69.4	73.2
	LSTM-CFCN	90	90.9	90.4
Insect cells	LoG	18.7	12.6	15.1
	SSAE	70.6	73.1	71.8
	VGG 16	83.4	87.9	85.6
	ResNet 50	85.2	77.3	81.1
	U-Net	87.4	85.1	86.2
	LSTM-CFCN	90.2	94.4	92.3
Vesicles	LoG	11.2	92.4	20.0
	SSAE	69.5	71.9	70.7
	VGG 16	84.7	90.5	87.5
	ResNet 50	88.3	82.1	85.1
	U-Net	90.6	87.9	89.2
	LSTM-CFCN	91.7	90.2	90.9
Retinal	LoG	9.1	32.8	14.2
	SSAE	53.6	60.1	56.7
	VGG 16	76.2	71.8	73.9
	ResNet 50	75.4	78.1	76.8
	U-Net	82.5	81.8	82.1
	LSTM-CFCN	82.6	83.3	82.9

performances than SSAE. ResNet, VGG and U-Net are the popular architecture for the cell detection tasks, and get a relative higher accuracy on these dataset. Our proposed framework can get a better results on these cells, but for the retinal cells data, which contain lots of density packed cell, and these cells usually crowded together and cannot be discriminated clearly, the total precision of our method is lower than that in the other data.

We have proposed a manually counting comparison protocol in [4], and this method can highlight the performance of automatic cell counting task from a scientists point of view. Take the scale and labor intensity together into consideration, we have also compare the Error Rate (ER) which is used as an evaluation indicator to compare human counting with our proposed framework at all of these data sets. The results is shown in figure 9. Because of the biomedical image from these four dataset contains multi-scales information, we hence randomly chose 4 scales from these dataset, and 25 patches for each scale to evaluate the error rate of our proposed method and manually counting results. The scale are chose by the discrimination of single cells, scale 1 is from 0 to 25th patches with the fewest cells in each patches, and scale 4 is from the 75th to 100th patches and contains most amount of cells. For pathological images, our lab scientist' error rate is about lower than 5% when the cell number is less than 80 per patch, the manually error rate exhibits a rise up trends with the images amount increase, this is caused by the scientist's personal problems with increased labor intensity and lack of concentration. When the cell number increase such as in the fourth scale, the manually error rate is higher than 10%, on this situation, our proposed framework still keep it around 10%. Here we see a error rate fluctuation of our framework in these four data set, this is caused by

**FIGURE 8. Cell detection and segmentation results of backbone substitution with FCN, U-Net and CFCN+CRF.****FIGURE 9. Cell detection results comparison between manually counting and proposed method.**

the image's noise on staining differences, uneven lighting intensity and blurred cell boundaries. The lowest and stable changes error rate is from insect cells for the reason that cell density is diluted into a degree of identification for light field microscopy. For retinal cell counting, manually counting lead to a most unstable exhibition from scale 1 to scale 4, for the reason that the cell is with density packed and even can not distinguished with vague edges. Under this situation, our lab scientist can not get a clearly counting results, and our framework can keep a stable error rate at 15% 18%. This results demonstrated that when the training set and cell number is large, our proposed methods can keep a stable error rate, and it can handle a large scale cell detection task when it confront the multiple scale whole scan slice pathology images.

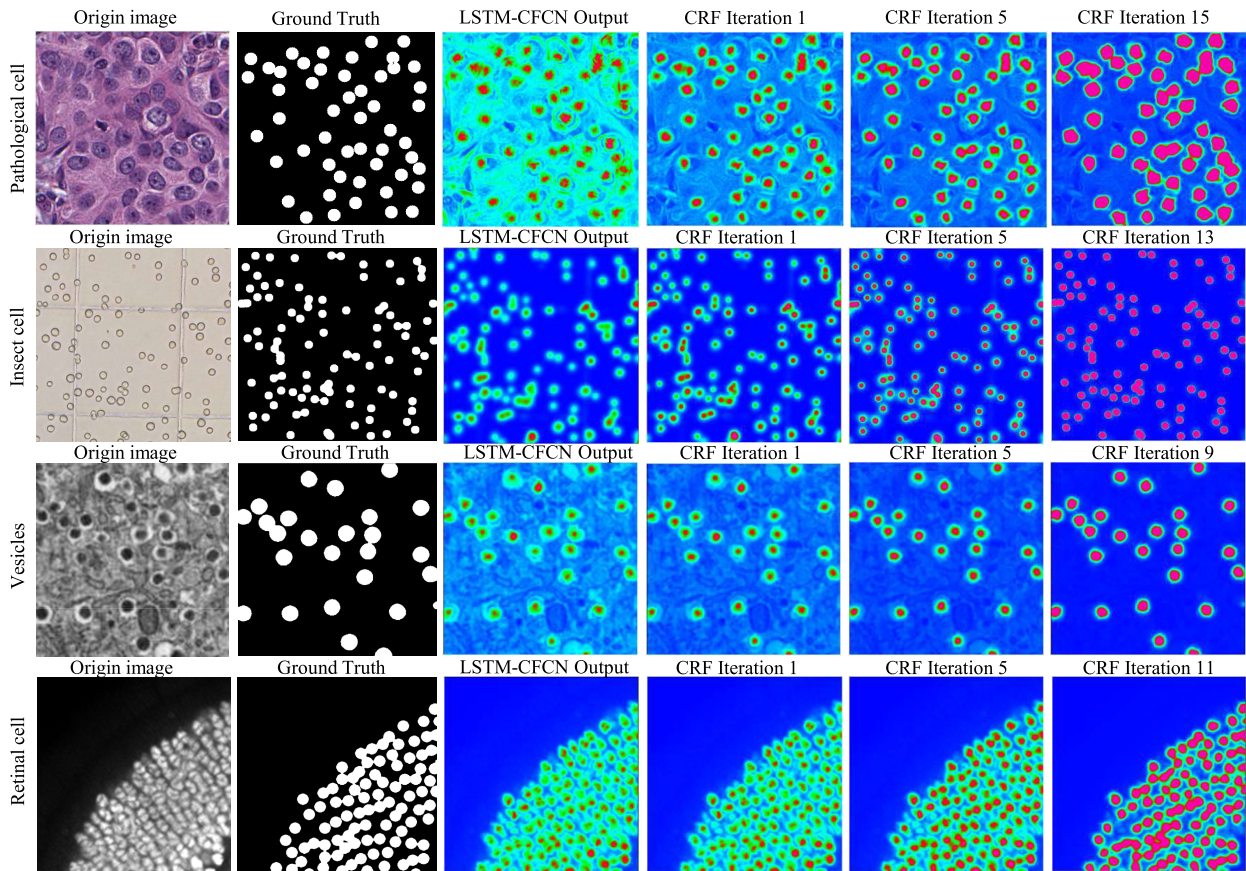


FIGURE 10. LSTM-CFCN segmentation results with CRF smoothed iterations at 1 time, 5 time and 15 times.

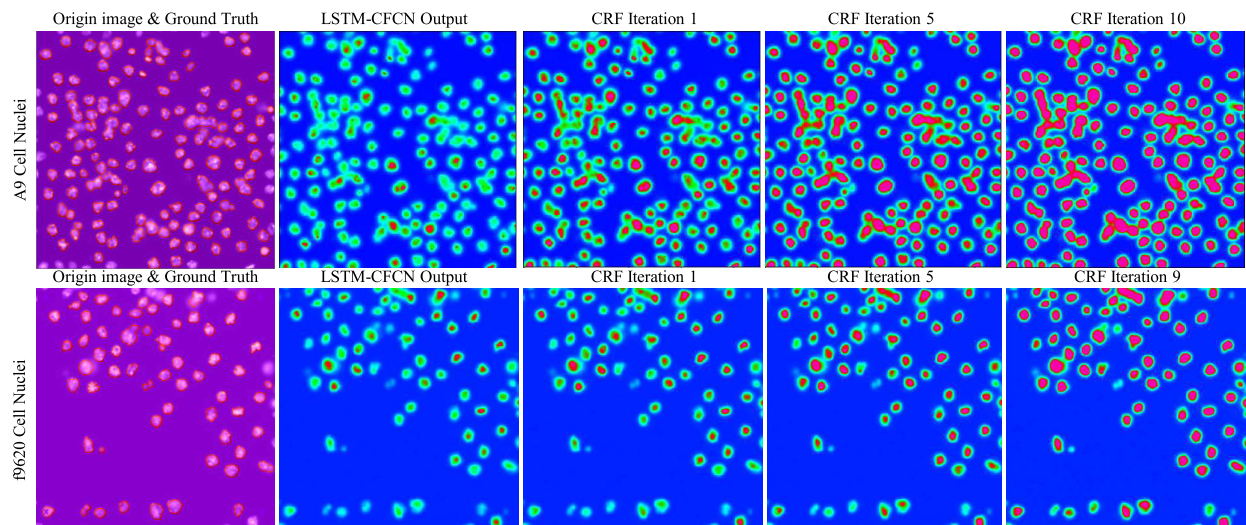


FIGURE 11. Segmentation results comparison on the Kc167 dataset after CRF smoothing.

C. RESULTS OF CELL SEGMENTATION

For the cell segmentation problems, most of the state-of-art methods are suitable for only specially kind of cell type and will get a low accuracy when it confront a distinct data set. We proposed a jointly segmentation method based on a CRF smooth manipulation after the detection task follows the heat

map of the network output. We have listed the comparison result on heat map generation with FCN and U-Net, see figure 8, Column 1 to column 3, represents origin image, ground truth, LSTM-CFCN outputs without CRF. Column 4 to column 6 represents LSTM-CFCN+CRF with 1 times iteration to 15 times iteration.. After CRF smooth on CFCN,

TABLE 2. Segmentation Results of Kc167 cell line images dataset.

Cell Type	Methods	Precision(%)	Recall(%)	F1
j9620 Cell Nuclei	VGG 16	72.1	64.3	67.9
	ResNet 50	71.6	66.7	69.1
	U-Net	77.8	71.3	74.4
	LSTM-CFCN	88.4	74.7	81.0
j9620 Cell Cytoplasm	VGG 16	54.3	28.7	37.6
	ResNet 50	49.7	32.1	39.0
	U-Net	55.7	31.6	40.3
	LSTM-CFCN	46.9	36.4	40.9
A9 Cell Nuclei	VGG 16	69.1	63.4	66.1
	ResNet 50	66.7	67.2	66.9
	U-Net	74.4	73.3	73.8
	LSTM-CFCN	77.1	72.4	74.6
A9 Cell Cytoplasm	VGG 16	43.2	36.7	39.7
	ResNet 50	45.6	41.2	43.3
	U-Net	53.6	49.8	51.6
	LSTM-CFCN	61.2	52.8	56.7

we can get a better results than U-Net and FCN on the same data set, see the fourth column in figure 8. FCN and U-Net need pixel label and end to end model training, here we choose the same training protocol with our framework, the results shows CFCN+CRF can better adapt the edge of the cells.

Segmentation results is shown in figure 10. Here we listed the source image, LSTM-CFCN outputs and 3 CRF smoothed results, which indicates that the proposed framework can get a proper cell segmentation by the CRF block after sever iterations. In this part, we make a separated channel to test the boosting effect of CRF block, the results shows that this block can get a better segmentation.

For a benchmark level comparison, we choose the Kc167 cell line images dataset from the Broad Institute's Imaging Platform, the ground truth and source image is listed in the first column of figure 11. This dataset is designed for cell detection and segmentation tasks, and divided into cell nuclei and cell cytoplasm separately. We use VGG16, ResNet50, U-Net and our proposed LSTM-CFCN as comparison test bed. During the training phase, we make a data augmentation such as rotation and reversal, a fine tuned protocol of the pre-trained network is also employed to get a better result as much as we can. Table 2 is the segmentation results on this dataset, our proposed method can reach a better result for cell nuclei segmentation, but for cell cytoplasm data, our method get top-3 precision, this is because the uneven illumination and data with irregular morphology of cell cytoplasm.

In summary, by using our proposed framework, we can get a better results than traditional Log method and state-of-art deep learning method. For different images data sets, the LSTM block is to make sure keeping the integration of a large images, this is effective for pathology cell images. Taking the computing time into consideration, the processing time for a 100000×200000 WSI image is about 7 minutes on our computing platform, and a 1024×1024 microscopic image is from 0.34 ms to 1 second that depending on the cell numbers.

V. CONCLUSION

In this paper, we proposed a LSTM-CFCN framework to establish a multi-task cell analysis platform, and jointly achieve the cell counting, detection and segmentation tasks in variety of medical images. We tested the proposed method on four kinds of cell datasets, and results shows the method performs well. A validation of segmentation on Kc 167 cell line demonstrates that the proposed framework can get a relative higher result than the traditional method and CNN based method. Therefore, our experimental result show that the proposed framework can cover most scenario of microscopic cell datasets from light field, dark field, multi-scale and electrical microscopy. For the disadvantages, our proposed framework can be restricted by the morphology of cells, and we can not handle the cell lines with irregular morphology, this is because of the label manner is only by the way of Gaussian filling protocol. Furthermore, the use of the Gaussian kernel is a significant limitation of the model for many other tasks such as irregular cell types. In the future, our aim will focus on multi-scale feature extraction and fine-tune the structure of the proposed networks that would allow improved performance on prediction.

ACKNOWLEDGMENT

The authors would like to thank GPU donations from NVIDIA.

REFERENCES

- [1] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.
- [2] V. A. Sindagi and V. M. Patel, "A survey of recent advances in CNN-based single image crowd counting and density estimation," *Pattern Recognit. Lett.*, vol. 107, pp. 3–16, May 2018.
- [3] B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. A. W. M. van der Laak, "Automated detection of DCIS in whole-slide H&E stained breast histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2141–2150, Sep. 2016.
- [4] D. Sui and K. Wang, "A counting method for density packed cells based on sliding band filter image enhancement," *J. Microsc.*, vol. 250, no. 1, pp. 42–49, Apr. 2013.
- [5] R. M. Thomas and J. John, "A review on cell detection and segmentation in microscopic images," in *Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT)*, Apr. 2017, pp. 1–5.
- [6] J. Wang, J. Liu, J. Liu, H. Yan, and R. Mao, "Automatic cell segmentation and signal detection in fluorescent in situ hybridization," in *Proc. Chin. Intell. Syst. Conf.* Singapore: Springer, 2019, pp. 285–293.
- [7] P. Naylor, M. Lae, F. Reyat, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.
- [8] Y. Tian, A. Dehghan, and M. Shah, "On detection, data association and segmentation for multi-target tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2146–2160, Sep. 2018.
- [9] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [11] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379–387.
- [12] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 140–145, Sep. 2012.

- [13] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.
- [14] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1719–1733, Dec. 2013.
- [15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.
- [16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [17] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
- [18] C. Park, J. Z. Huang, J. X. Ji, and Y. Ding, "Segmentation, inference and classification of partially overlapping nanoparticles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, p. 1, Mar. 2013.
- [19] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer-Verlag, 2013, doi: 10.1007/978-3-662-05088-0.
- [20] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.
- [21] Z. Lu, G. Carneiro, and A. P. Bradley, "Automated nucleus and cytoplasm segmentation of overlapping cervical cells," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2013, pp. 452–460.
- [22] E. Cosatto, M. Miller, H. P. Graf, and J. S. Meyer, "Grading nuclear pleomorphism on histological micrographs," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
- [23] K. Wang, D. Sui, W. Wang, Y. Yuan, and W. Zuo, "A cell counting method for BEVS based on nonlinear transformed sliding band filter," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 118–121.
- [24] H. Su, F. Xing, J. D. Lee, C. A. Peterson, and L. Yang, "Automatic myonuclear detection in isolated single muscle fibers using robust ellipse fitting and sparse representation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 4, pp. 714–726, Jul. 2014.
- [25] K. Z. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for P53 immunohistochemistry," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1153–1163, Jun. 2006.
- [26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [27] F. Mualla, S. Schöll, B. Sommerfeldt, S. Steidl, R. Buchholz, and J. Hornegger, "Improving joint learning of suspended and adherent cell detection using low-pass monogenic phase and transport of intensity equation," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 927–930.
- [28] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2188–2202, Nov. 2011.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [30] S. Seguí, O. Pujol, and J. Vitria, "Learning to count with deep object features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 90–96.
- [31] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 6, no. 3, pp. 283–292, May 2018.
- [32] A. Khan, S. Gould, and M. Salzmann, "Deep convolutional neural networks for human embryonic cell counting," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 339–348.
- [33] F. Liu and L. Yang, "A novel cell detection method using deep convolutional neural network and maximum-weight independent set," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Cham, Switzerland: Springer, 2017, pp. 63–72.
- [34] T. Chen and C. Chef'd'Hotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2014, pp. 17–24.
- [35] S. Wang, J. Yao, Z. Xu, and J. Huang, "Subtype cell detection with an accelerated deep convolution neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 640–648.
- [36] W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 474–478, Mar. 2018.
- [37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [38] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," 2017, *arXiv:1703.02442*. [Online]. Available: <http://arxiv.org/abs/1703.02442>
- [39] B. Lee and K. Paeng, "A robust and effective approach towards accurate metastasis detection and pN-stage classification in breast cancer," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 841–850.
- [40] W. F. Hink, D. R. Thomsen, D. J. Davidson, A. L. Meyer, and F. J. Castellino, "Expression of three recombinant proteins using baculovirus vectors in 23 insect cell lines," *Biotechnol. Prog.*, vol. 7, no. 1, pp. 9–14, Jan. 1991.
- [41] G. P. Lewis, D. G. Charteris, C. S. Sethi, and S. K. Fisher, "Animal models of retinal detachment and reattachment: Identifying cellular events that may affect visual recovery," *Eye*, vol. 16, no. 4, pp. 375–387, Jul. 2002.
- [42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [43] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3168–3176, May 2020.
- [44] S. Zhang, G. Wu, J. P. Costeira, and J. M. F. Moura, "FCN-rLSTM: Deep spatio-temporal neural networks for vehicle counting in city cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3667–3676.
- [45] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
- [46] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
- [47] P. Quelhas, M. Marcuzzo, A. M. Mendonca, and A. Campilho, "Cell nuclei and cytoplasm joint segmentation using the sliding band filter," *IEEE Trans. Med. Imag.*, vol. 29, no. 8, pp. 1463–1473, Aug. 2010.
- [48] D. Sui, K. Wang, H. Park, and J. Chae, "Bright field microscopic cells counting method for BEVS using nonlinear convergence index sliding band filter," *Biomed. Eng. OnLine*, vol. 13, no. 1, p. 147, 2014.
- [49] I. Sutskever, *Training Recurrent Neural Networks*. Toronto, ON, Canada: Univ. of Toronto, 2013.
- [50] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, May 2016.
- [51] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.
- [52] W.-P. Jan and X. Liu, "Deep residual learning for image recognition," in *Proc. CVPR*, 2015, pp. 1–26.
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [55] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.



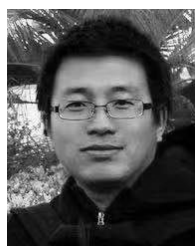
DONG SUI received the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2015. From January 2015 to August 2016, he was a Researcher with Aerospace Changfeng Company Ltd., Beijing, China. From September 2016 to September 2018, he was a Postdoctoral Research Fellow with the College of Computer Science, Beihang University, Beijing. He is currently a Lecture with the College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing. His current research interests include machine learning, computer vision, virtual reality, and medical image analysis.



FEI YANG received the Ph.D. degree in computer science from the Harbin Institute of Technology, China, in 2012. From April 2010 to April 2011, he worked under the joint Ph.D. Program with the School of Computer Science and Technology, Harbin Institute of Technology, and the School of Physics and Astronomy, University of Manchester. He is currently a Lecture with the School of Mechanical, Electrical, and Information Engineering, Shandong University, Weihai. His current research interests include scientific data visualization, and virtual organ modeling.



MAOZU GUO received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology. He is currently a Professor with the College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China. His research interests include machine learning, bioinformatics, and urban computing.



LEI ZHANG received the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2013. He was a Lecturer with the School of Art and Design, Harbin University, Harbin, China. From September 2011 to August 2012, he was a Research Intern with Siemens Corporate Research, Inc., Princeton, NJ, USA. From July 2015 to September 2017, he was a Postdoctoral Research Fellow with the College of Engineering, Temple University, PA, USA, and a Postdoctoral Research Fellow with the Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, USA. From September 2017 to September 2019, he was a Postdoctoral Associate with the School of Medicine, Department of Radiology, University of Pittsburgh. He is currently a Faculty of diagnostic radiology and nuclear medicine with the University of Maryland, Baltimore, MD, USA. His current research interests include machine learning, computer vision, visualization, and medical image analysis.

...