








Deep Learning-Based Microscopic Cell Detection Using Inverse Distance Transform and Auxiliary Counting

Rui Liu , Wei Dai , *Graduate Student Member, IEEE*, Cong Wu , Tianyi Wu ,
Min Wang , *Graduate Student Member, IEEE*, Junxian Zhou, Xiaozhen Zhang,
Wen Jung Li , *Fellow, IEEE*, and Jun Liu , *Senior Member, IEEE*

Abstract—Microscopic cell detection is a challenging task due to significant inter-cell occlusions in dense clusters and diverse cell morphologies. This paper introduces a novel framework designed to enhance automated cell detection. The proposed approach integrates a deep learning model that produces an inverse distance transform-based detection map from the given image, accompanied by a secondary network designed to regress a cell density map from the same input. The inverse distance transform-based map effectively highlights each cell instance in the densely populated areas, while the density map accurately estimates the total cell count in the image. Then, a custom counting-aided cell center extraction strategy leverages the cell count obtained by integrating over the density map to refine the detection process, significantly reducing false responses and thereby boosting overall accuracy. The proposed framework demonstrated superior performance with F-scores of 96.93%, 91.21%, and 92.00% on the VGG, MBM, and ADI datasets, respectively, surpassing existing state-of-the-art methods. It also achieved the lowest distance error, further validating the effectiveness of the proposed approach. These results demonstrate significant potential for automated cell analysis in biomedical applications.

Index Terms—Cell detection, healthcare automation, deep learning, inverse distance transform, cell counting.

I. INTRODUCTION

CELL detection, which involves identifying the positions of individual cells, is a common yet challenging task

Received 24 January 2024; revised 15 May 2024; accepted 14 June 2024. Date of publication 20 June 2024; date of current version 4 October 2024. This work was supported in part by the Research Grant Council (RGC) of Hong Kong under Grant 11212321, Grant 11217922, and Grant ECS-21212720 and in part by the Science and Technology Innovation Committee of Shenzhen under Grant Type-C SGD20210823104001011. (*Corresponding author: Jun Liu.*)

Rui Liu, Wei Dai, Tianyi Wu, Min Wang, Junxian Zhou, Xiaozhen Zhang, and Jun Liu are with the Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, SAR, China, and also with the Centre for Robotics and Automation, City University of Hong Kong, Hong Kong, SAR, China (e-mail: rui.liu@my.cityu.edu.hk; jun.liu@cityu.edu.hk).

Cong Wu and Wen Jung Li are with the Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE), Hong Kong, SAR, China, and also with the Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, SAR, China.

Digital Object Identifier 10.1109/JBHI.2024.3417229

in biology and medicine. Manual detection under microscopy is labor-intensive and prone to subjective errors, especially in densely populated regions where cells are closely clustered. Therefore, there is a strong demand to automate the cell detection procedure with higher accuracy and enhanced objectivity.

Traditionally, automated cell detection strategies are carefully tailored based on the given microscopic images [1], [2], [3]. The approach for one particular scenario may vary significantly from another, contingent upon factors such as cellular characteristics, image acquisition techniques, and background texture. Moreover, the detection accuracy is usually hindered by inherent defects present in the images and the limitations associated with handcrafted features. These challenges have spurred researchers to develop more robust, universally applicable methods for fully automatic cell detection.

Deep learning has recently achieved unprecedented success across diverse domains [4], [5], [6], [7], [8]. In biomedicine, this revolutionary technology has also demonstrated its powerful capabilities in various applications [9], [10], [11], [12], [13]. Consequently, a bunch of systems based on deep learning have been developed, aiming to enhance the intelligence and generalizability of the cell detection process [14], [15], [16].

One widely employed approach for cell detection involves predicting individual bounding boxes for each cell instance. However, this strategy often struggles in scenarios where cells are densely clustered or exhibit considerable variation in morphology. Additionally, the process of annotating objects with bounding boxes is time-consuming and expensive [17], [18]. In this annotation paradigm, annotators need to meticulously focus on the contours of objects of various sizes and carefully draw bounding boxes, which is labor-intensive. An alternate approach for automatic cell detection is based on map regression, which can effectively alleviate the aforementioned issue. This strategy involves regressing a spatial representation map for a given microscopic image, where the value of each pixel indicates the likelihood of cell presence. The problem is subsequently reformulated as the task of identifying a local maximum for each cell, as the center of the cell is expected to generate a more potent response compared to the surrounding background. The map regression pipeline only requires weak annotations, typically in the form of a dot near the center of each object. Compared to

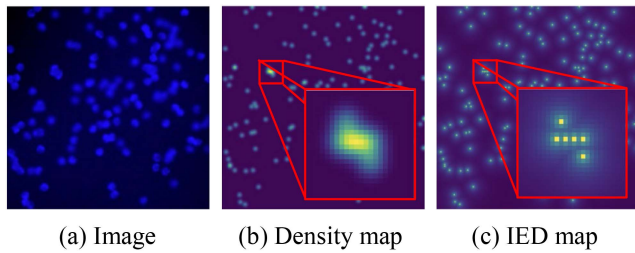


Fig. 1. The differences between the density map and the IED map. (a) The input image. (b) The cell instances become indistinguishable in occlusion regions in the density map. (c) The IED map accurately pinpoints cell locations, even in highly dense areas.

drawing a polygon tightly enclosing an object, clicking on the center of that object is much simpler and less time-consuming.

The spatial representation map, functioning as the supervisory signal for training the detector, is an intermediate representation created from the dot map. One common way to obtain the ground truth spatial representation map is by applying a Gaussian kernel to each dot, resulting in a blurred representation, also referred to as the density map. The density map is effective in accurately determining the number of objects by integrating over the entire map. However, it is difficult for Gaussian blobs to provide individual cell locations in highly dense regions with substantial cell occlusions. In such regions, as shown in Fig. 1(b), the overlapping of blobs can lead to a false detection where the local maxima may not converge to the cell centers. To address this issue, the inverse Euclidean distance-based (IED) map is employed to act as the intermediate ground truth. As shown in Fig. 1(c), the individual cell centers are highlighted in the IED map, making it work effectively even in highly dense areas. While this method successfully localizes individual instances, it performs poorly in acquiring the number of cells, potentially leading to numerous false responses.

In previous studies [15], [32], density maps were used as the intermediate representation for supervised training of detection networks. However, this approach resulted in suboptimal performance in the dense regions. To address this, the IED map is improved to serve as the intermediate representation. The proposed pipeline begins with the microscopic image fed into the detection network (detector) to generate the corresponding IED map. This map facilitates the detection of cell centers by identifying local maxima. Despite this improvement, another challenge emerges due to the unknown number of cells, which can cause the vanilla local maxima determination approach to yield numerous false responses. To address this issue, a counting-aided strategy is introduced for cell center extraction. In this strategy, another network functions as a counter to estimate a density map by fusing the IED map and the input image. The cell count, obtained by integrating the predicted density map, aids the determination of the cell center by eliminating false detections.

The new method involves a synergistic interaction between the localization and counting components. The detector boosts the accuracy of the density map regression through feature fusion, and conversely, the counter improves the precision of the local maximum extraction algorithm for determining cell

TABLE I
AN OVERVIEW OF THE RELATED WORK

Techniques	Supervisory signals	References
Traditional Methods	NA	[1] [2] [3] [19] [20] [21]
Deep Learning-based Methods	Bounding box	[23] [27] [26] [29]
	Spatial representation map	[15] [31] [32] [34] [35] [36] [37]

centers. Moreover, the structural information within the IED map is crucial for accurate cell localization. Consequently, a component has been incorporated into the proposed pipeline specifically to refine the texture pattern of the IED map during training.

In summary, this study makes the following primary contributions:

- A novel framework that utilizes the localization capabilities of the IED map along with the counting proficiency provided by the density map is introduced to enhance the effectiveness of deep learning-based cell detection.
- Development of a U-shaped network with embedded residual blocks specifically targets the regression of high-quality IED maps, ensuring robust feature extraction and processing.
- A counting-aided strategy for cell center extraction is implemented, which significantly improves accuracy in locating individual cell instances.
- A specialized module is designed to refine the IED map, thereby boosting the overall performance of the detection framework.
- The effectiveness of the proposed framework is evaluated across three public databases. The experimental results demonstrate that the proposed method outperforms existing state-of-the-art approaches, highlighting its superiority.

The remainder of the article is organized as follows. Section II reviews the recent advancements in cell detection approaches. Section III then introduces the proposed cell detection framework. The experimental setup, including datasets and implementation details, is described in Section IV. Section V reports the experimental results, followed by a discussion in VI. Finally, Section VII concludes this work.

II. RELATED WORK

This section reviews recent advances in automatic cell detection, discussing a variety of techniques and methodologies that have been presented in the literature to tackle the challenges associated with cell detection. The related work is summarized in Table I.

A. Traditional Methods

Traditional image processing techniques have been foundational in the development of cell detection systems. For instance, Mualla et al. [1] and Lu et al. [2] applied techniques such as

scale-invariant feature transform (SIFT) key points extraction and Bayesian modeling, respectively, to enhance cell detection for specific applications. Similarly, Xing et al. [19] recognized seeds with hierarchical voting-based detection to provide priori information for Ki-67 scoring. However, the effectiveness of these methods in scenarios involving significant cell overlap remains to be validated. To tackle this challenge, Arteta et al. [20] introduced a tree-structured graphical model for detecting overlapping objects. Additionally, Ge et al. [3] employed distance transform and spectral clustering to differentiate objects within overlapping regions to develop a system for counting blood cells. Aiming to improve both accuracy and efficiency in cell detection, Kainz et al. [21] proposed a simple system that predicted a score map for a given image using Random Forests [22] and then found the local maximums as the cell centers. Although traditional methods have demonstrated decent performance in certain situations, their robustness and generalizability are constrained by the need to adjust algorithms manually depending on image texture features.

B. Deep Learning-Based Methods

The emergence of deep learning has transformed the field of cell detection, enabling end-to-end learning directly from raw imaging data. This advancement significantly surpasses the constraints of traditional image processing methods, providing more robust and versatile solutions for cell identification challenges.

1) *Bounding Box-Based Approaches*: Deep learning-based object detection algorithms typically utilize bounding-box annotations to provide information for supervised training. Techniques like YOLO [24] and the advanced version [28] were modified by Aham and Islam [23], and Shakarami et al. [27], respectively, to enhance the accuracy of cell identification. Similarly, Faster R-CNN [25] has been re-implemented by Hung et al. [26] to develop an easy-to-use tool that facilitates cell detection in both brightfield and fluorescence images. To address the issue of low recognition rates caused by dense cell distributions, Gu and Sun [29] enhanced the feature processing capability of the network in high-density feature areas by integrating an attention module into YOLOv5 [30]. These deep learning-based detectors can be trained to recognize various cells without meticulous customization specific to particular image texture features, thus improving generalizability. However, these methods typically rely on the prediction of bounding boxes for individual cells, which can be computationally intensive and particularly challenging when cells are tightly clustered.

2) *Spatial Representation Map-Based Approaches*: To address the limitations associated with bounding box predictions and streamline the annotation process, research efforts have shifted towards developing deep learning networks to produce spatial representation maps for cell detection. Xie et al. [15] constructed fully convolutional models to regress Gaussian-based density maps from the microscopic images to enhance cell detection and counting. Similarly, Lu et al. [31] employed the density map as an intermediate representation for object

identification in their class-agnostic counting system. Subsequently, Guo et al. [32] extended this paradigm to 3D cell recognition by modifying the infant U-Net [33] with 3D operations. While these methods effectively circumvent the drawbacks of bounding boxes by using Gaussian transform maps, challenges still remain in densely populated cell areas where Gaussian peaks may become indistinguishable due to occlusion, thereby diminishing detection accuracy.

Fortunately, the introduction of inverse distance maps, where pixel intensity is inversely proportional to the Euclidean distance from the nearest cell center, offers a potential solution to this issue by distinctly highlighting each cell center. Xie et al. [34] employed a custom convolutional neural network to produce spatial proximity patches, which were then fused to yield the final prediction map for the cell image. Building on this foundation, the same team [35] improved this method by using a well-designed hierarchical encoder-decoder network to directly regress a proximity map that matches the size of the input. Furthermore, the effectiveness of the inverse distance kernel for cell localization was also validated by Guo et al. [36].

Other efforts to tackle the challenges in cell identification have also been made. For instance, to address the challenges of localizing cells in densely overlapping and low-contrast scenes, Chen et al. [37] introduced an innovative concept of the direction field map. In this map, each foreground pixel is represented as a vector pointing towards the center of the nearest object, while background pixels are set to zero. This novel representation has demonstrated significant potential in enhancing cell localization accuracy across various complex scenarios.

III. METHODOLOGY

This section introduces to the overall structure of the proposed pipeline, followed by a detailed description of each component within the framework.

A. Framework Overview

The basic process of spatial representation map-based cell detection can be described formally as follows. Given a microscopic image $X \in \mathbb{R}^{H \times W \times C}$, the function $F_d(\cdot)$ will estimate a spatial representation map $\hat{Y} \in \mathbb{R}^{H \times W \times 1}$ corresponding to it. Then the local maxima in \hat{Y} can be found as the cell centroids. The map serves as an intermediate representation in this process because directly mapping the cells to sparse dots is challenging. To derive the locations of each cell, the proposed framework is built with three components, including ground truth generation, basic detection, and auxiliary counting, as shown in Fig. 2.

In the basic detection part, a customized network (detector) acts as the mapping function $F_d(\cdot)$. The deep learning-based detector can be trained to generate IED maps for the cell images by leveraging its superior nonlinear representation ability. The resulting maps are then processed by a local maximum identification algorithm for the final outputs. However, due to the lack of knowledge about the number of cells in the input image, local maximum identification process may generate a significant number of false responses, resulting in a decrease in detection accuracy. To address this issue, an auxiliary counting module is

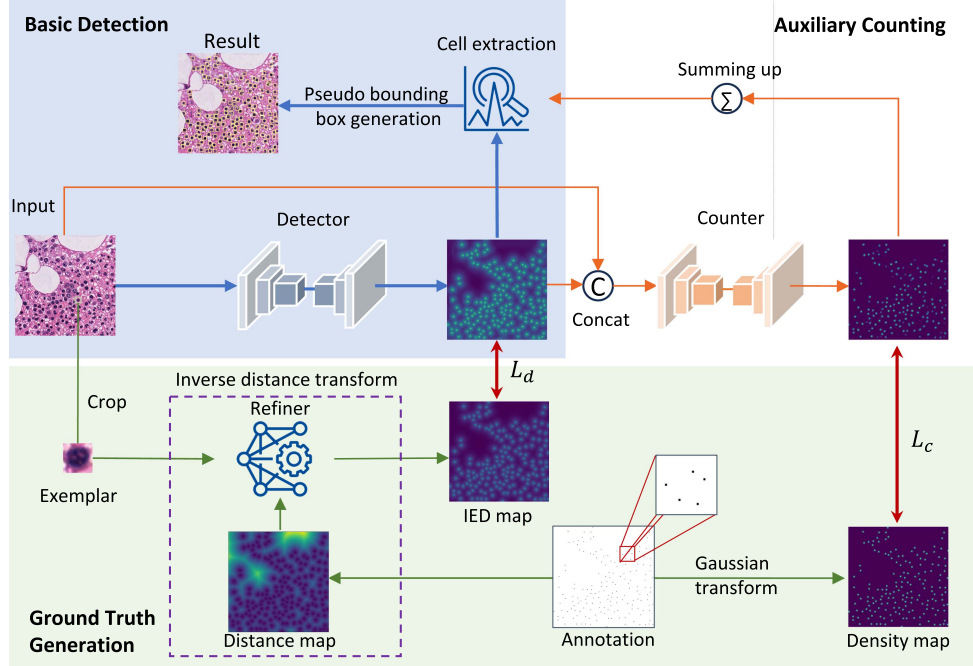


Fig. 2. The proposed framework for cell detection in microscopic image analysis. During the training stage, the cell sample is input into the detection network to regress the IED map, and the auxiliary counter estimates the density map based on the concatenation of the input image and the obtained IED map. The ground truth generation branch generate the ground truths for training the detector and counter. During the testing phase, the trained detector and counter regress an IED map and a density map for the unseen image, respectively. Finally, the cell extraction algorithm extracts cell instances from the predicted IED map with the assistance of the estimated cell count. In the detection result, pseudo bounding boxes are generated from the predicted map for better visualization.

employed to assist the detection module in accurately identifying the cell centroids in a given image by providing the count of objects presented in that image. In the auxiliary counting branch, a deep learning model (counter) is trained to regress a density map based on the input image and its corresponding IED map. The cell counting is obtained by summing up the density map. These two components are complementary to each other.

The ground truth generation component is responsible for providing supervisory information for the training of the detector and counter. In addition, a network (refiner) is utilized to polish the pattern of the IED map. The refiner dynamically adjusts the ground truth during the training process. It is worth noting that during the testing phase, the ground truth generation part is not involved.

B. Ground Truth Generation

Inspired by [38] and [39], a modified inverse distance transform is proposed to produce ground truth IED maps for training the detector. Given a human-annotated dot map $A \in \mathbb{R}^{H \times W \times 1}$ with a set of cell centers C , the corresponding IED map $Y \in \mathbb{R}^{H \times W \times 1}$ can be expressed as:

$$Y(h, w) = \frac{1}{D(h, w)^{(\lambda B_1 \times D(h, w) + B_2)} + \alpha} \quad (1)$$

where $D(h, w)$ is the distance between the position (h, w) and its nearest cell center (c_h, c_w) , α is set to 1, B_1 and B_2 are modulated by the refiner to dynamically adjust the ground truth features during training. To avoid potential arithmetic overflow

TABLE II
THE ARCHITECTURE DETAILS OF THE REFINER

Operations	Input channels/nodes	Output channels/nodes
Conv-ReLU-MaxPool	3	32
Conv-ReLU-MaxPool	32	64
Conv-ReLU-MaxPool	64	128
FC1	*	128
FC2-Sigmoid	128	2

* Is determine by the feature maps from the preceding layer.

caused by the exponential operation, B_1 is scaled by a small factor λ , which is set to 0.01. $D(h, w)$ can be computed as:

$$D(h, w) = \sqrt{(h - c_h)^2 + (w - c_w)^2}, (c_h, c_w) \in C \quad (2)$$

The refining network consists of three convolutional layers and two fully connected layers. During the training process, the refiner is optimized along with the detector through back propagation. To reduce the computational burden, one single cell instance is cropped from each training image to serve as the representative input for the refining network. B_1 and B_2 are the first and second elements in the output of the refiner, respectively. Explicitly, the structure details of the refiner are summarized in Table II. The Conv-ReLU-MaxPool denotes a 2D convolution operation followed by ReLU activation and max pooling with stride of 2. The kernel size and stride of the convolution operation are configured as 3×3 and 1, respectively. FC represents a fully connected layer.

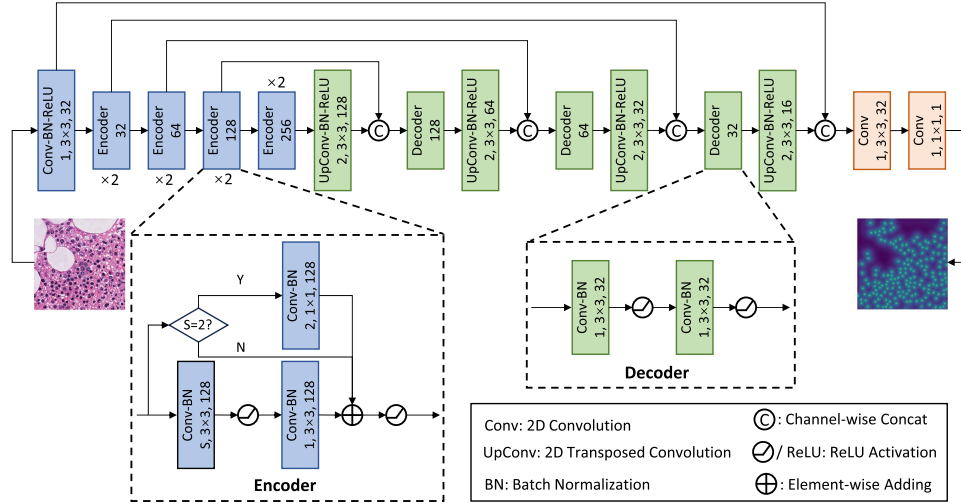


Fig. 3. The architecture of the detector. [Conv(-BN-ReLU), $S, k \times k, N$] represents a 2D convolution operation with the stride of S , the output channel number of N and the kernel size of $k \times k$, followed by batch normalization (if applicable) and ReLU activation (if applicable). The transposed convolution operation also has a similar description. The [Encoder M] or [Decoder M] indicates that the number of output channels in this block is M . The padding operation is used to adjust the feature maps to the desired size in each layer.

The Gaussian transform map is adopted to supervise the learning of the counter. The ground truth density map is obtained by convolving the expert-annotated dot map A with a Gaussian kernel. In the experiments, the standard deviation and size of the bivariate Gaussian kernel are set to (2,2) and 17×17 , respectively.

C. The IED Map Regression

A hierarchically concatenated network, based on modifications to the U-Net architecture [33], is designed to produce IED map, as illustrated in Fig. 3. The input layer has three channels, corresponding to the number of channels of the cell image. Compared to the original U-Net, the detector increases the depth of the network by integrating two residual encoders into each stage of the encoding process, thereby enhancing feature extraction capability. In each encoder, a residual connection [40] is employed to improve training stability. Instead of utilizing max pooling, the first residual encoder in each stage performs downsampling through a convolutional operation with a stride of 2. This alternative not only better preserves spatial information but also enhances the learning capacity of the network. To reduce the complexity of the architecture, the number of channels of feature maps is halved compared to the original. Additionally, appropriate padding is applied to each layer to make the output size equal to the input size, thus ensuring accurate localization of cell instances.

To supervise the training of the detector, a combination of mean squared error (MSE) and structural similarity index (SSIM) [41] is employed. This combination has been demonstrated to be effective in obtaining high-quality regression maps in previous studies [39], [42]. The MSE loss L_d^{mse} and the SSIM loss L_d^{ssim} for training the detector can be calculated as follows:

$$L_d^{mse} = \left\| \hat{Y} - Y \right\|^2 \quad (3)$$

$$L_d^{ssim} = 1 - \frac{(2\mu_{\hat{Y}}\mu_Y + C_1)(2\sigma_{\hat{Y}Y} + C_2)}{(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1)(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2)} \quad (4)$$

where Y and \hat{Y} are the ground truth IED map and its corresponding prediction, respectively. The μ and σ are the mean and variance. Following [41], the C_1 and C_2 are set to $0.0001l^2$ and $0.0009l^2$ to avoid unstable results. The l is the range of the value in the estimated image. Finally, the objective of the joint training of the detector and the refiner is obtained by adding the MSE and SSIM losses:

$$L_d = L_d^{mse} + L_d^{ssim} \quad (5)$$

D. Auxiliary Counting

The auxiliary counting module aims to estimate the number of cells in the input image. Previous research has demonstrated the effectiveness of density maps in cell counting [32], [52]. In the proposed system, the counter is created by modifying the detector. Since the input of the counter is the channel-wise concatenation of cell images and the obtained IED map, the number of channels in the input layer of the counter has been adjusted to four.

The training process of the counter is similar to that of the detector, with the main difference being the supervisory information. To reduce the computational complexity of the training process, the counter is trained using the MSE loss, which is calculated as follows:

$$L_c = \left\| \hat{Z} - Z \right\|^2 \quad (6)$$

where $Z \in \mathbb{R}^{H \times W \times 1}$ and $\hat{Z} \in \mathbb{R}^{H \times W \times 1}$ are the ground truth density map and its corresponding prediction, respectively.

E. Cell Extraction

The cell centers can be obtained by identifying the local maxima in the inferred IED map. Following the strategy in [39],

Algorithm 1: The Cell Center Extraction Process.

Input: The predicted IED map $\hat{Y}(h, w)$ and the corresponding estimated number of cells \hat{c} .

Output: The predicted cell centers $\{p_i\}_1^{\hat{c}}$.

Obtain the candidates

- 1: $\tilde{Y}(h, w) = \text{MaxPool}(\hat{Y}(h, w), \text{size} = (3, 3), \text{stride} = 1)$
- 2: **if** $\tilde{Y}(h, w) == \hat{Y}(h, w)$ **then**
- 3: $\tilde{Y}(h, w) = 1$
- 4: **else**
- 5: $\tilde{Y}(h, w) = 0$
- 6: **end if**
- 7: $P(h, w) = \tilde{Y}(h, w) \times \hat{Y}(h, w)$
- # Get the top \hat{c} values in $P(h, w)$*
- 8: $P_{\text{reshape}} = P(h, w).\text{clone.view}(-1)$
- 9: $V_{\text{top-}\hat{c}}, \text{indices} = \text{get_topk}(P_{\text{reshape}}, \hat{c})$
- # Obtain the predicted cell centers*
- 10: $V_{\text{threshold}} = \min\{V_{\text{top-}\hat{c}}\}$
- 11: **if** $P(h, w) \geq V_{\text{threshold}}$ **then**
- 12: $P(h, w) = 1$
- 13: **else**
- 14: $P(h, w) = 0$
- 15: **end if**
- 16: $\{p_i\}_1^{\hat{c}} = \text{nonzero}(P(h, w))$
- 17: **return** $\{p_i\}_1^{\hat{c}}$.

the map is first filtered using a 3×3 max pooling operation, and then locations where the values remain unchanged are retained as candidate cell centers. Unsurprisingly, there are many false responses contained in these candidates. In previous research, the local maxima were recognized in individual sliding window patch under the guidance of the corresponding region count [43]. However, this strategy yielded suboptimal results for cell detection, as the number of objects in the microscopy field of view is typically much lower compared to dense crowd scenarios. The intuitive reason behind this observation is that the kernel of a certain cell instance can be split into different regions, resulting in significant region-counting errors for counting a small number of objects.

To obtain a high accuracy for cell extraction, a different approach is adopted by extracting the cell centers from the entire predicted IED map at once, constrained by the estimated number of cells in the image. This allows for consideration of the global context and mitigates the errors caused by region counting limitations. Specifically, the process begins with the summation of the corresponding density map of the predicted IED map to obtain the estimated cell count, denoted as \hat{c} . The top- \hat{c} candidates, based on their probabilities, are then retrieved to serve as the inferred cell centers. This selection strategy primarily focuses on identifying and pinpointing the most probable locations of cells in the image, thereby enhancing the accuracy of cell detection.

The cell instance extraction process is described in Algorithm 1.

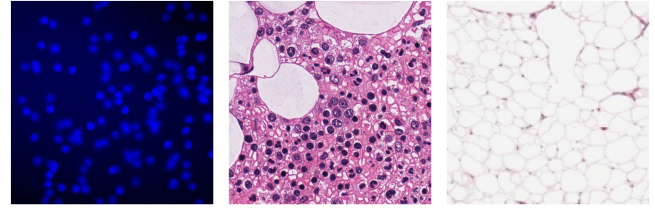


Fig. 4. Example images of the three datasets. From left to right: VGG, MBM, ADI.

IV. EXPERIMENTAL SETUP

This section describes the dataset and implementation details in the experiments.

A. Dataset Description

In this study, three public datasets are employed to evaluate the proposed method, as shown in Fig. 4. One dataset consists of synthetically generated suspension cells, while the other two originate from biological tissues. In one of the tissue image sets, cells are sparsely distributed, while the other dataset features densely packed cells. These three datasets represent three distinct scenarios, enabling an evaluation of the algorithm's performance across diverse situations.

1) *VGG Dataset:* The VGG dataset, introduced by Lempitky et al. [45] from the Visual Geometry Group (VGG) at the University of Oxford, comprises 200 synthetic images. The images were generated using the system presented by Lehmussola et al. [46]. Each image in this dataset has a dimension of $256 \times 256 \times 3$ pixels, and the dataset contains a total of 35,192 bacterial cells. The system was trained using 100 images, with the remaining images reserved for testing. The cell detection in this dataset is challenging due to variations in focal distances and occlusions among the bacteria.

2) *MBM Dataset:* The second dataset is the modified bone marrow (MBM) dataset, created by Paul et al. [47]. This dataset consists of 44 images with a dimension of $600 \times 600 \times 3$ pixels and contains 5,553 cells in total. These samples were obtained from eight healthy subjects and stained with Hematoxylin and Eosin. Thirty images were used to train the network and the remaining for testing. This dataset has special challenges in cell detection due to the inhomogeneous background texture and variations in cell morphologies.

3) *ADI Dataset:* The Adipocyte (ADI) dataset was collected from the Genotype-Tissue Expression (GTEx) project [48] by Paul et al. [47]. This dataset consists of 200 microscopic images of human subcutaneous adipose tissue with a size of $150 \times 150 \times 3$ pixels. The total number of cells in these data is 29,684. This dataset was equally split into training and test sets. These adipose cells vary significantly in shape and size and are packed tightly together, making the cell detection difficult.

B. Implementation Details

The proposed method was implemented using PyTorch. During the training phase, the batch size was set to 16 for the VGG

and ADI datasets, and to 8 for the MBM dataset. random cropping, flipping, and rotating techniques were applied to augment the training data. The new size after cropping was obtained by multiplying the original size by 0.875 and rounding it down to the nearest value divisible by 16. This calculation was performed due to the model's architecture, which includes four cascaded downsampling operations with a stride of 2. The cropped images were then randomly flipped horizontally and rotated with an angle of $k \times 90^\circ$ ($k = 1, 2, 3, 4$) before being input into the system. At the testing stage, the sizes of the MBM and ADI images were padded to 608×608 and 160×160 , respectively. This ensured that the test samples could be processed compatibly by the framework while retaining all the cell instances. The aforementioned operations were applied consistently to both the images and their corresponding labels.

The proposed framework was optimized using the Adam algorithm [49] with a weight decay of 0.0001 for 2000 epochs. The initial learning rate was set to 0.0005 for both the detector and counter components, while the refiner had an initial learning rate of $1e-6$. Inspired by [50], the learning rate was increased linearly to the initial value during the first ten epochs and then followed a cosine annealing schedule for the remaining iterations. The experiments were conducted on a server equipped with four Nvidia RTX3090 GPUs and an Intel Xeon Platinum 8375 C CPU.

V. EXPERIMENTAL RESULTS

In this section, the evaluation metrics are presented and the detection results of the framework are compared with state-of-the-art methods. Additionally, ablation studies are conducted to validate the effectiveness of the design.

A. Evaluation Protocol

The performance of detection methods is evaluated using precision, recall, F-score, and distance error. Prior to computing these evaluation metrics, criterion for successful detection of a given cell instance is established. A cell is deemed successfully detected if the detection occurs within a radius of R_d around it. Here, the tolerance R_d is determined based on the average cell size in each dataset. A correct detection is classified as a true positive (TP). A detection that does not match any annotation is considered a false positive (FP). A false negative (FN) refers to a ground truth cell that has not been assigned a detection. In the experiment, the Hungarian algorithm was employed to match the annotated dots with the predicted points to achieve maximum bipartite matching.

The precision (P) and recall (R) can be calculated by $TP/(TP + FP)$ and $TP/(TP + FN)$, respectively. The F-score is the harmonic mean of the recall and precision, given by $2 \times P \times R / (P + R)$. The distance error is quantified by computing the normalized mean (μ_d) of the Euclidean distance (\hat{d}) between each ground truth location and its correctly matched detection. For a test image set containing a total of T cell

TABLE III
COMPARATIVE RESULTS FOR CELL DETECTION IN VGG DATASET

Methods	Precision (%)	Recall (%)	F-score (%)
Singletons [20] [†]	98.87	72.07	83.37
Full system w/o surface [20] [†]	95.00	91.97	93.46
Direction Field Map [37] [†]	98.95	93.90	96.36
GeoNet [44] [†]	93.50	91.90	92.70
Class-Agnostic [31] [†]	99.43	82.50	90.18
SAU-Net [32] [†]	99.94	89.65	94.51
FIDT [39]	99.53	91.98	95.61
C-FCRN-Aux [52]	99.80	88.75	93.96
IDN [51]	40.43	47.10	43.51
FCRN-A [15]	99.91	84.76	91.72
FCRN-B [15]	99.97	88.02	93.61
The proposed	99.16	94.80	96.93

[†]: results obtained from the original articles.

Best results are marked in bold.

instances, the μ_d can be calculated as follows:

$$\mu_d = \frac{\sum_{i=1}^T \hat{d}}{T \cdot R_d} \quad (7)$$

B. Comparison With the State-of-The-Art Methods

In this part, a comparative analysis between the proposed method and the state-of-the-art cell detection techniques is conducted. Additionally, the evaluation is extended to encompass advanced cell counting algorithms such as Interact Dual Network (IDN) [51] and C-FCRN-Aux [52], which can be used for cell detection with post-processing techniques.

1) *VGG Dataset*: Following [37], the radius of the ground truth region was set to 4 for the highly realistic synthetic bacterial cells in the VGG dataset. The experiment results are summarized in Table III. It is noteworthy that for the re-implemented methods, the hyperparameters were fine-tuned to achieve their optimal performance. All methods except IDN achieved high detection accuracy with an F-score exceeding 80%. The proposed approach surpassed state-of-the-art methods in terms of F-score, a comprehensive metric measuring the detection capability of an algorithm. Thanks to the IED map and the tailored counting-aided cell center extraction strategy, the new method achieved an optimal balance between precision and recall. Specifically, this system attained the highest recall rate while also delivering competitive performance in terms of precision.

The location capabilities of individual approaches are further revealed in the distance error. As depicted in Fig. 5, the proposed method achieves a marginally lower distance error compared to its competitors.

2) *MBM Dataset*: In accordance with [35] and [37], a distance threshold of 16 pixels was set between the actual cell center and its correctly matched detection for the bone marrow samples. The results of the experiments conducted on the MBM database are reported in Table IV. The hyperparameters for the re-implemented methods were fine-tuned in order to attain their optimal performance. The proposed method achieved the highest precision and F-score on this dataset. The GeoNet [44] and structured regression model [35] obtained relatively high recall rate. However, they perform poorly in precision. In contrast, the proposed framework yielded highly competitive results across

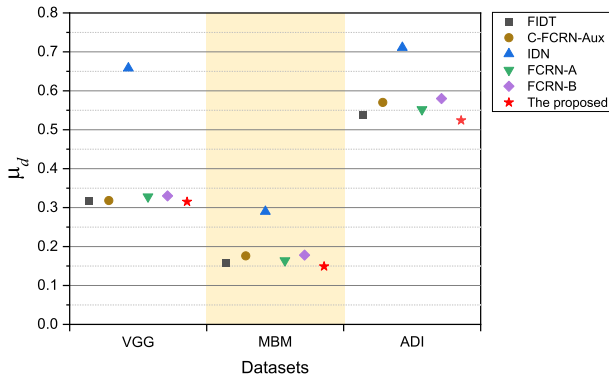


Fig. 5. The distance errors of the proposed method on the three datasets: VGG, MBM, ADI.

TABLE IV

COMPARATIVE RESULTS FOR CELL DETECTION IN MBM DATASET

Methods	Precision (%)	Recall (%)	F-score (%)
Direction Field Map [37] [†]	87.04	90.44	88.71
GeoNet [44] [†]	86.8	95.0	90.7
Structured Regression [35] [†]	86	94	90
SAU-Net [32] [†]	88.76	92.83	90.73
FIDT [39]	89.97	91.55	90.75
C-FCRN-Aux [52]	88.18	90.47	89.31
IDN [51]	48.76	81.35	60.96
FCRN-A [15]	88.79	89.71	89.25
FCRN-B [15]	86.94	90.19	88.53
The proposed	91.93	90.50	91.21

[†]: results obtained from the original articles.
Best results are marked in bold.

TABLE V

COMPARATIVE RESULTS FOR CELL DETECTION IN ADI DATASET

Methods	Precision (%)	Recall (%)	F-score (%)
SAU-Net [32] [†]	88.57	93.13	90.78
FIDT [39]	92.20	90.06	91.12
C-FCRN-Aux [52]	88.57	92.70	90.59
IDN [51]	38.46	85.05	52.97
FCRN-A [15]	90.99	72.86	80.92
FCRN-B [15]	91.53	88.58	90.03
The proposed	92.52	91.48	92.00

[†]: results obtained from the original articles.
Best results are marked in bold.

all evaluation metrics, thanks to the counting-aided cell center extraction strategy.

Regarding the localization performance observed in the bone marrow images (Fig. 5), all of these methods demonstrated distance errors significantly smaller than the radius of the ground truth region. This can be attributed to the inherent distinctiveness of cell features within the bone marrow sample, as shown in Fig. 4. Consequently, the networks trained with ground truth maps emphasizing distinct features near the object centroids could effectively capture individual cell instances. Additionally, the proposed method exhibited the lowest average distance error, further validating its robustness and superiority.

3) ADI Dataset: The radius of the ground truth region for the human subcutaneous adipose tissue sample was set to 12 pixels, based on the size of the cell. The experimental results are presented in Table V. Hyperparameter fine-tuning was conducted for the re-implemented methods, aiming to attain their best

TABLE VI
COMPARISON OF CELL COUNTING RESULTS

Methods	VGG (MAE ↓)	MBM (MAE ↓)	ADI (MAE ↓)
Singletons [20] [†]	51.2	-	-
Full system w/o surface [20] [†]	5.1	-	-
Direction Field Map [37] [†]	8.8	9.2	-
Structured Regression [35] [†]	-	36.3	-
Class-Agnostic [31] [†]	3.6	-	-
SAU-Net [32] [†]	2.6	5.7	14.2
FIDT [39]	7.6	7.0	13.1
C-FCRN-Aux [52]	8.9	6.8	12.3
IDN [51]	1.9	4.0	11.1
FCRN-A [15]	15.3	9.3	26.4
FCRN-B [15]	11.2	8.1	13.7
The proposed	5.3	5.5	10.9

↓: lower MAE values indicate better performance.

[†]: results obtained from the original articles.

Best results are marked in bold.

performance. The proposed method surpassed its competitors in terms of F-score and precision on this dataset. Due to the high sensitivity of the IED map in localization and the counting-aided cell extraction strategy, this approach neatly trade off precision and recall. In contrast, the SAU-Net [32] and C-FCRN-Aux [52] achieved high recall rates but performed poorly in terms of precision.

The cells in the human subcutaneous adipose tissue lack distinctive features, posing a challenge for detectors to accurately capture individual objects, as indicated in Fig. 4. As a result, all methods demonstrated relatively poor performance on this dataset in terms of distance error, as observed in Fig. 5. However, primarily owing to the inverse distance transform map that prominently highlighted the center of each cell, the proposed framework and the FIDT approach achieved slightly lower distance errors compared to other competitors.

C. Counting Performance

The proposed framework is also compared with other approaches in terms of counting performance. The mean absolute error (MAE) is utilized as a metric to quantify the counting performance. The MAE is defined as the average absolute difference between the estimated cell number and the real count, and it can be expressed as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{c}_i - c_i| \quad (8)$$

where N refers to the number of test images. \hat{c}_i and c_i represent the estimated cell count and real cell count of the i -th sample, respectively.

As shown in Table VI, the proposed method achieved the lowest and second-lowest counting errors on the ADI and MBM datasets, respectively, and its performance on the VGG dataset was also above average. Therefore, compared to other competitors, the proposed framework demonstrates a very competitive counting capability. Interestingly, some methods exhibit contradictory behavior between cell counting and cell detection. For example, IDN performed poorly in detection but achieved

TABLE VII
THE EXPERIMENT RESULTS OF THE FRAMEWORKS WITH DIFFERENT DESIGNS ON THE VGG DATASET

Trials	Ground Truth			Cell Center Extraction Strategies		Precision (%)	Recall (%)	F-score (%)
	Density map	Vanilla IED map	Refined IED map	Threshold-based	counting-aided			
1	✓	-	-	✓	-	99.57	89.35	94.18
2	✓	-	-	-	✓	99.03	90.21	94.42
3	-	✓	-	✓	-	99.34	92.17	95.62
4	-	✓	-	-	✓	99.03	93.18	96.01
5	-	-	✓	✓	-	98.96	93.54	96.17
6	-	-	✓	-	✓	99.16	94.80	96.93

Best results are marked in bold.

impressive results in cell counting. This is because the post-processing techniques employed for cell counting pipeline are completely different from those used in cell detection. Cell counting treats the microscopic image as a whole, whereas detection requires focusing on each individual cell instance.

D. Ablation Analysis

Ablation studies were conducted to justify the rationality of the proposed approach.

1) *Analysis of the Framework Design*: The proposed framework was investigated with varying configurations on the VGG dataset. In the first trial, the detector was trained using the density map, and then extracted the cell centers from the predicted map based on the empirical threshold. The experimental results are summarized in Table VII. With this configuration, the system can still yield acceptable results, validating the effectiveness of the detecting network. The counting-aided cell extraction strategy was then implemented to retrieve cells from the density-based predicted map. This new trial achieved a certain improvement in terms of F-score. Specifically, the first configuration had a slightly higher precision, while the second one performed better in terms of recall rate.

In the third and fourth experiments, the detector was trained under the supervision of the vanilla IED map without refinement. As illustrated in Table VII, the F-scores attained by both the threshold-based and counting-aided configurations surpassed those of their respective counterparts trained with the density map. This underscores the efficacy of the inverse distance kernel in cell detection.

In the subsequent trial, the detector was trained using the refinable IED map, and the inferred map was processed by the threshold-based local maximum determination strategy to yield the final detections. As can be seen from Table VII, the F-score increases significantly compared to the previous designs. The final experiment was the proposed design, where the detector was trained under the supervision of the refinable IED map, and cell instances were extracted using the counting-aided strategy. The proposed method achieved the highest F-score and recall compared to its counterparts, while performing competitively in terms of precision.

The prediction maps and final detection results of the second trial and the proposed configuration are visually presented in Fig. 6. It can be seen that in high-density regions where cells overlap and occlude each other (as indicated within the red ellipses), it is challenging to distinguish individual cells in

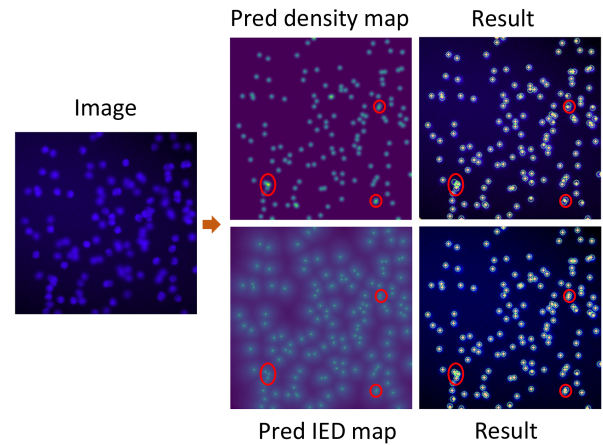


Fig. 6. Predicted results of the detector trained under the supervision of density map and IED map, respectively. The green circles denote the ground truth cells, while the yellow dots represent the detected cells. The red ellipses highlight some exemplary areas where different configurations yield varying outcomes.

the predicted density map. In contrast, the cell instances are distinguishable in the predicted IED map.

2) *Analysis of the Hyper-Parameters in IED Map Generation*: The characteristics of the IED map play a pivotal role in cell detection. A well-designed ground truth map should be able to clearly distinguish objects within dense regions and effectively supervise the training of the detector. In the ground truth generator of the proposed framework, the coefficients λ and α are used to shape the fundamental attributes of the IED map. Specifically, a higher λ value results in a more pronounced decay of pixel values from the cell centers to the background, whereas a lower value has the opposite effect. α is utilized to modulate the pixel values at the cell centers. It is critical to avoid setting λ too high, as the excessive exponential calculations could lead to computational overflow. The experimental outcomes from various hyper-parameter combinations on the VGG dataset are detailed in Table VIII. In the proposed method, α and λ are set to 1 and 0.01, respectively.

3) *Analysis of the Detector*: An additional ablation study was conducted to validate the modifications applied to the U-Net architecture in the proposed detector. As summarized in Table IX, the proposed detector outperformed its original counterpart across all indicators. This improvement can be primarily attributed to the enhanced feature extraction capabilities of the revised architecture. Firstly, the deepening of the network is essential for capturing subtle variations in cell morphology and

TABLE VIII
ABLATION STUDY OF THE HYPER-PARAMETERS IN THE IED MAP
GENERATION ON THE VGG DATASET

α	λ	Precision (%)	Recall (%)	F-score (%)
0.5	0.01	98.57	94.36	96.43
1	0.01	99.16	94.80	96.93
1.5	0.01	98.87	94.83	96.79
0.5	0.02	98.36	93.57	95.91
1	0.02	99.04	94.22	96.57
1.5	0.02	98.34	93.93	96.08

Best results are marked in bold.

TABLE IX
ABLATION STUDY OF THE DETECTOR ON THE VGG DATASET

Methods	Precision (%)	Recall (%)	F-score (%)
U-Net [33]	98.79	93.32	95.98
The proposed	99.16	94.80	96.93

Best results are marked in bold.

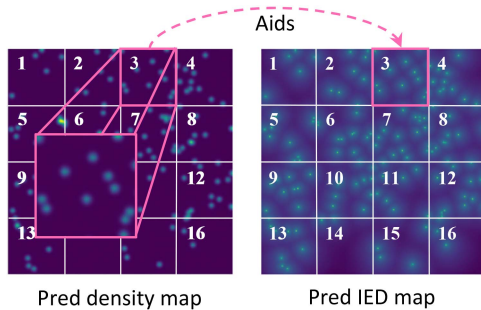


Fig. 7. The schematic diagram of block-based counting-aided cell extraction strategy. In this setting, the maps are divided into 8×8 blocks. The count within each density map block aids the cell extraction process in the corresponding block of the IED map. For instance, the 3rd region in the density map provides count information for the cell extraction in the 3rd region of the IED map.

texture, which are critical for accurate cell detection. Moreover, the introduction of convolutional downsampling in the proposed detector helps in preserving more spatial information than traditional max pooling, which may result in the loss of features during the pooling process. Additionally, appropriate padding was also incorporated into the vanilla U-Net to ensure that the position of the cell was obtained on the full-sized image.

4) Analysis of the Counting-Aided Pattern in the Cell Extraction Strategy: The counting pattern in the counting-aided cell extraction strategy plays a pivotal role in obtaining accurate detection results. Intuitively, region-based counting can provide valuable guidance, as depicted in Fig. 7. To further investigate this, experiments were conducted using various block-based counting-aided extraction strategies. Surprisingly, the experimental results, as presented in Fig. 8, revealed an unexpected trend that the F-score decreased as the number of blocks increased. This decline in the F-score can be attributed to a decrease in precision, while the recall rate remained relatively consistent regardless of the number of blocks. The observation can be explained by the fact that the Gaussian kernels near the regional boundary are divided into different blocks, as illustrated

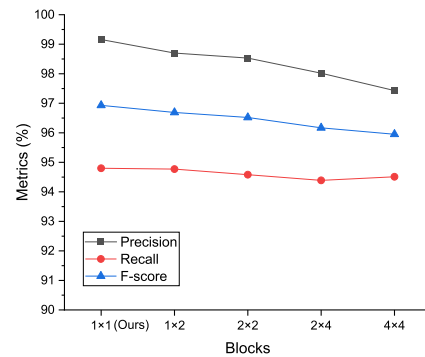


Fig. 8. The experimental results of cell extraction strategies based on different fine-grained regions on the VGG dataset.

in the partial enlargement of Fig. 7. Furthermore, as the number of blocks increases, this situation becomes more pronounced. The division of the map into more blocks exacerbates the aforementioned issue, resulting in a corresponding decrease in precision and ultimately leading to a lower F-score.

VI. DISCUSSION

To further illustrate the detection capabilities of various methods, the cell detection outcomes across different datasets are delineated in Fig. 9. As depicted in the first row, the proposed framework exhibits superior accuracy in the VGG dataset, significantly reducing the incidence of both false negatives and false positives. This is particularly evident in regions with occluded cells, as highlighted by the red ellipse, where each cell instance is effectively identified. This is attributable to the enhanced cell center focus provided by the IED map within the framework. In contrast, the IDN pipeline tends to overpredict, frequently resulting in multiple detections for single cells. This issue arises because the IDN framework lacks a prominent focus near the cell center in the intermediate representation map, making the system prone to generating multiple responses.

In the second row of Fig. 9, which showcases results from the MBM dataset, the presented method consistently outperforms other approaches in object identification. The predicted cells are predominantly centered within the ground truth markers, indicating a smaller positional error. Other competitors, however, frequently generate spurious detections. This is exemplified by the IDN method's tendency to generate multiple detections for a same real cell, further illustrating the challenges other techniques encounter in achieving precise localization.

The third row of Fig. 9 reveals the detection outcomes on the ADI image. Here again, the proposed system demonstrates superior cell detection accuracy. Notably, in cell-free areas (as indicated by red ellipses), the reported system accurately reflects the absence of cells, sharply contrasting with other methods that mistakenly identify multiple cells in these regions. This underscores the robustness and reliability of the proposed approach across different and challenging imaging conditions, emphasizing its practical utility in diverse biological imaging contexts.

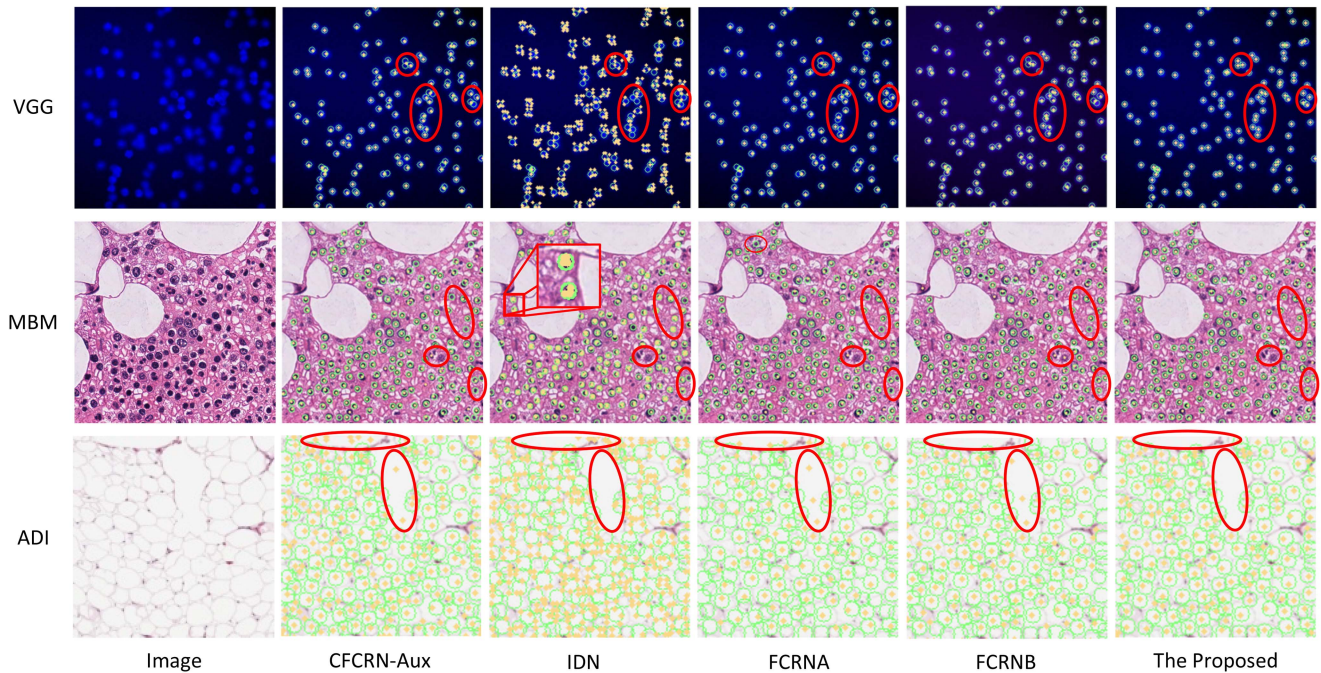


Fig. 9. Example cell detection results of different methods on the synthetic bacterial cell image (first row), bone marrow sample (second row), and the human subcutaneous adipose tissue (third row). In each image, green circles indicate the ground truth cells, while yellow dots represent the detected cells. The red ellipses highlight specific areas that demonstrate the difference in prediction results among different methods.

To further demonstrate the improvements of the proposed method in cell detection, an additional non-parametric statistical test was conducted. Specifically, a Wilcoxon rank-sum test was performed to compare the proposed method with the FIDT map, which served as the baseline. The analysis was based on F-score values obtained from five independent trials on the VGG dataset for each method. The results yielded a p-value of 0.008, below the significance threshold of 0.01. The statistical analysis indicates a significant enhancement in detecting cells, thereby validating the effectiveness of the proposed method.

In the development of an artificial intelligence system, time complexity is a critical consideration, especially in practical applications where the system is required to deliver rapid responses [53], [54], [55], [56]. Compared to machine learning [57], [58], [59], [60] and other intelligent computing approaches [61], [62], deep learning methods [63], [64] are generally more computationally intensive. As a result, the reliance on substantial computing resources and the resulting long inference time are limitations of the proposed method. This is primarily due to the fact that the system employs an additional auxiliary counter to enhance the cell detection accuracy, which increases the number of parameters and computational complexity. Therefore, the design of lightweight models while maintaining detection accuracy could be a direction for future improvements.

In practical applications, robustness against adversarial attacks is a critical capability for deep learning models [65], [66]. The training of the proposed cell detector utilized a dataset free from interference signals, leading to impressive performance in controlled laboratory settings. However, this trained detector may experience performance degradation in more complex real-world scenarios, especially when confronted with sophisticated

adversarial attacks. This vulnerability is of critical concern in intelligent medical diagnostic systems, where inaccuracies in output could compromise the reliability of diagnostic results and subsequent treatment decisions. Given these challenges, a potential research focus for the future should be enhancing the robustness of the detector against various disturbed inputs, ensuring the effectiveness and reliability in real-world applications.

VII. CONCLUSION

This paper presents a novel framework for cell detection in microscopic images by leveraging the refined IED map as an intermediate representation. The IED map effectively retains the precise location information from the original annotations and discriminates every individual cell instance within densely clustered areas. The framework is enhanced by a counting-aided cell center extraction strategy to extract the individual instances from the predicted IED map accurately. In this pipeline, an auxiliary counting network estimates the number of cells in the input image, which then guides the elimination of false responses during the cell instance identification process. The proposed framework was evaluated across three public datasets, namely VGG, MBM, and ADI, achieving F-scores of 96.93%, 91.21%, and 92.00%, respectively. These results set new records for state-of-the-art cell detection performance, demonstrating the universality and efficacy of the proposed system. In summary, the foremost contribution of this research is the use of the improved inverse distance transform as an intermediate representation for cell detection, coupled with the development of a bespoke deep learning-based system to implement this

concept. The study has made significant progress in the field of microscopic cell detection and provides a promising solution to overcome the challenges in cell biology and histology.

In the future, efforts can be directed toward extending the proposed framework to 3D cell detection scenarios. Additionally, the training data used in this study is assumed to be without defects. However, in real-world applications, imperfect data are commonly encountered. Therefore, exploring techniques for effectively training the detector under conditions of imperfect data is critical. Moreover, enhancing the robustness of the proposed detector against adversarial attacks will significantly improve diagnostic reliability. Lastly, there is a demand to design lightweight detectors, especially for mobile and point-of-care applications.

REFERENCES

- [1] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, and J. Hornegger, "Automatic cell detection in bright-field microscope images using sift, random forests, and hierarchical clustering," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2274–2286, Dec. 2013.
- [2] C. Lu and M. Mandal, "Toward automatic mitotic cell detection and segmentation in multispectral histopathological images," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 2, pp. 594–605, Mar. 2014.
- [3] J. Ge et al., "A system for counting fetal and maternal red blood cells," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2823–2829, Dec. 2014.
- [4] J. Jeong, T. Kim, B. J. Lee, and J. Lee, "Predicting AFM topography from optical microscope images using deep-learning," *Adv. Intell. Syst.*, vol. 5, no. 1, 2023, Art. no. 2200317.
- [5] Z. Wu et al., "Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 3072.
- [6] K. Shaikat, S. Luo, and V. Varadharajan, "A novel deep learning-based approach for malware detection," *Eng. Appl. Artif. Intell.*, vol. 122, 2023, Art. no. 106030.
- [7] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, "A survey of deep learning applications to autonomous vehicle control," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 712–733, Feb. 2021.
- [8] K. Shaikat, S. Luo, and V. Varadharajan, "A novel machine learning approach for detecting first-time-appeared malware," *Eng. Appl. Artif. Intell.*, vol. 131, 2024, Art. no. 107801.
- [9] X. Zhang et al., "CXR-NET: A multitask deep learning network for explainable and accurate diagnosis of COVID-19 pneumonia from Chest X-ray images," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 980–991, Feb. 2023.
- [10] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.
- [11] Z. Liu et al., "Automatic diagnosis of significant liver fibrosis from ultrasound b-mode images using a handcrafted-feature-assisted deep convolutional neural network," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 10, pp. 4938–4949, Oct. 2023.
- [12] L. Devnath et al., "Deep ensemble learning for the automatic detection of pneumoconiosis in coal worker's chest X-ray radiography," *J. Clin. Med.*, vol. 11, no. 18, 2022, Art. no. 5342.
- [13] T. M. Alam et al., "An efficient deep learning-based skin cancer classifier for an imbalanced dataset," *Diagnostics*, vol. 12, no. 9, 2022, Art. no. 2115.
- [14] T. Falk et al., "U-Net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.
- [15] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," *Comput. Methods Biomech. Biomed. Eng.: Imag. Visualization*, vol. 6, no. 3, pp. 283–292, 2018.
- [16] C.-W. Wang, S.-C. Huang, Y.-C. Lee, Y.-J. Shen, S.-I. Meng, and J. L. Gaol, "Deep learning for bone marrow cell detection and classification on whole-slide images," *Med. Image Anal.*, vol. 75, 2022, Art. no. 102270.
- [17] D. P. Papadopoulos, J. R. Uijlings, F. Keller, and V. Ferrari, "Training object class detectors with click supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6374–6383.
- [18] D. P. Papadopoulos, J. R. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 854–863.
- [19] F. Xing, H. Su, J. Neltner, and L. Yang, "Automatic Ki-67 counting using robust cell detection and online dictionary learning," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 859–870, Mar. 2014.
- [20] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Detecting overlapping instances in microscopy images using extremal region trees," *Med. Image Anal.*, vol. 27, pp. 3–16, 2016.
- [21] P. Kainz, M. Urschler, S. Schuler, P. Wohlhart, and V. Lepetit, "You should use regression to detect cells," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 276–283.
- [22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [23] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells," *Healthcare Technol. Lett.*, vol. 6, no. 4, pp. 103–108, 2019.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [25] 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. 1–9.
- [26] J. Hung et al., "Keras R-CNN: Library for cell detection in biological images using deep neural networks," *BMC Bioinf.*, vol. 21, pp. 1–7, 2020.
- [27] A. Shakarami, M. B. Menhaj, A. Mahdavi-Hormat, and H. Tarrah, "A fast and yet efficient Yolov3 for blood cell detection," *Biomed. Signal Process. Control*, vol. 66, 2021, Art. no. 102495.
- [28] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [29] W. Gu and K. Sun, "AYOLOv5: Improved YOLOv5 based on attention mechanism for blood cell detection," *Biomed. Signal Process. Control*, vol. 88, 2024, Art. no. 105034.
- [30] Ultralytics, "YOLOv5," 2020. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [31] E. Lu, W. Xie, and A. Zisserman, "Class-agnostic counting," in *Proc. 14th Asian Conf. Comput. Vis., Comput. Vis.*, 2018, pp. 669–684.
- [32] Y. Guo, O. Krupa, J. Stein, G. Wu, and A. Krishnamurthy, "SAU-Net: A unified network for cell counting in 2D and 3D microscopy images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 4, pp. 1920–1932, Jul./Aug. 2022.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [34] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 358–365.
- [35] Y. Xie, F. Xing, X. Shi, X. Kong, H. Su, and L. Yang, "Efficient and robust cell detection: A structured regression approach," *Med. Image Anal.*, vol. 44, pp. 245–254, 2018.
- [36] Y. Guo, D. Borland, C. McCormick, J. Stein, G. Wu, and A. Krishnamurthy, "Cell counting with inverse distance Kernel and self-supervised learning," in *Proc. Int. Workshop Med. Opt. Imag. Virtual Microsc. Image Anal.*, 2022, pp. 1–10.
- [37] Y. Chen, D. Liang, X. Bai, Y. Xu, and X. Yang, "Cell localization and counting using direction field map," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 359–368, Jan. 2022.
- [38] J. Wan, Q. Wang, and A. B. Chan, "Kernel-based density map generation for dense object counting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1357–1370, Mar. 2022.
- [39] D. Liang, W. Xu, Y. Zhu, and Y. Zhou, "Focal inverse distance transform maps for crowd localization," *IEEE Trans. Multimedia*, vol. 25, pp. 6040–6052, 2023.
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [42] X. Cao, Z. Wang, Y. Zhao, and F. Su, "Scale aggregation network for accurate and efficient crowd counting," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 734–750.

- [43] J. Cheng, H. Xiong, Z. Cao, and H. Lu, "Decoupled two-stage crowd counting and beyond," *IEEE Trans. Image Process.*, vol. 30, pp. 2862–2875, 2021.
- [44] J. Chen, Y. Zhu, and Z. Chen, "Graph-embedded online learning for cell detection and tumour proportion score estimation," *Electronics*, vol. 11, no. 10, 2022, Art. no. 1642.
- [45] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1–9.
- [46] A. Lehmussola, P. Ruusuvuori, J. Selinummi, H. Huttunen, and O. Yli-Harja, "Computational framework for simulating fluorescence microscope images with cell populations," *IEEE Trans. Med. Imag.*, vol. 26, no. 7, pp. 1010–1016, Jul. 2007.
- [47] J. Paul Cohen, G. Boucher, C. A. Glastonbury, H. Z. Lo, and Y. Bengio, "Count-ception: Counting by fully convolutional redundant counting," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 18–26.
- [48] J. Lonsdale et al., "The genotype-tissue expression (GTEx) project," *Nature Genet.*, vol. 45, no. 6, pp. 580–585, 2013.
- [49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [50] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.
- [51] R. Liu et al., "Interactive dual network with adaptive density map for automatic cell counting," *IEEE Trans. Automat. Sci. Eng.*, early access, Nov. 09, 2023, doi: [10.1109/TASE.2023.3329973](https://doi.org/10.1109/TASE.2023.3329973).
- [52] S. He, K. T. Minn, L. Solnica-Krezel, M. A. Anastasio, and H. Li, "Deeply-supervised density regression for automatic cell counting in microscopy images," *Med. Image Anal.*, vol. 68, 2021, Art. no. 101892.
- [53] K. Shaukat et al., "Performance comparison and current challenges of using machine learning techniques in cybersecurity," *Energies*, vol. 13, no. 10, 2020, Art. no. 2509.
- [54] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A survey on machine learning techniques for cyber security in the last decade," *IEEE Access*, vol. 8, pp. 222310–222354, 2020.
- [55] K. Shaukat, S. Luo, S. Chen, and D. Liu, "Cyber threat detection using machine learning techniques: A performance evaluation perspective," in *Proc. IEEE Int. Conf. Cyber Warfare Secur.*, 2020, pp. 1–6.
- [56] M. R. Kumar et al., "Dementia detection from speech using machine learning and deep learning architectures," *Sensors*, vol. 22, no. 23, 2022, Art. no. 9311.
- [57] T. M. Alam et al., "A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining," *Biomed. Signal Process. Control*, vol. 68, 2021, Art. no. 102726.
- [58] Z. Ali et al., "A proposed framework for early prediction of schistosomiasis," *Diagnostics*, vol. 12, no. 12, Art. no. 3138, 2022.
- [59] X. Yang, M. Khushi, and K. Shaukat, "Biomarker CA125 feature engineering and class imbalance learning improves ovarian cancer prediction," in *Proc. IEEE Asia-Pacific Conf. Comput. Sci. Data Eng.*, 2020, pp. 1–6.
- [60] T. M. Alam et al., "A machine learning approach for identification of malignant mesothelioma etiological factors in an imbalanced dataset," *Comput. J.*, vol. 65, no. 7, pp. 1740–1751, 2022.
- [61] M. Khushi et al., "A comparative performance analysis of data resampling methods on imbalance medical data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021.
- [62] T. M. Alam et al., "A fuzzy inference-based decision support system for disease diagnosis," *Comput. J.*, vol. 66, no. 9, pp. 2169–2180, 2023.
- [63] S. Luo and K. Shaukat, *Computational Methods for Medical and Cyber Security*. Basel, Switzerland: MDPI Books, 2022.
- [64] C. Srinivas et al., "Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images," *J. Healthcare Eng.*, vol. 2022, 2022, Art. no. 3264367.
- [65] K. Shaukat, S. Luo, and V. Varadharajan, "A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks," *Eng. Appl. Artif. Intell.*, vol. 116, 2022, Art. no. 105461.
- [66] M. Xu, T. Zhang, and D. Zhang, "Medrdf: A robust and retrain-less diagnostic framework for medical pretrained models against adversarial attack," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 2130–2143, Aug. 2022.