

Real-Time Cell Counting in Unlabeled Microscopy Images

Yuang Zhu

School of Computer Science and Technology, Donghua University

zhuyuang1@gmail.com

Zhao Chen*

School of Computer Science and Technology, Donghua University

Department of Computer Science, University of Warwick

chenzhao@dhu.edu.cn, Zhao.Chen.1@warwick.ac.uk

Yuxin Zheng, Qinghua Zhang and Xuan Wang

School of Computer Science and Technology, Donghua University

Abstract

Deep learning is largely applied to cell counting in microscopy images. However, most of the existing cell counting models are fully supervised and trained off-line. They adopt the usual training-testing framework, whereas the models are trained in advance to infer numbers of cells in test images. They require large amounts of manually labeled data for training but lack the ability to adapt to newly-collected unlabeled images that are fed to processing systems dynamically. To solve these problems, we propose a novel framework for real-time (RT) cell counting with density maps (DM). It is a semisupervised system which enables training with upcoming unlabeled images and predicting their cell counts simultaneously. It is also flexible enough to allow almost any cell counting model to be embedded within it. With a reliable and automatic training set renewing mechanism, it ensures counting accuracy while optimizing the models by both historical data and new images. To deal with cell variability and image complexity, we propose a Semisupervised Graph-Based Network (SGN) for within the RT counting framework. It leverages a count-sensitive measurement to construct dynamic graphs of DM patches. With the graph constraint, it regularizes an encoder-decoder to represent underlying data structures and gain robustness for cell counting. We have realized SGN along with several baseline networks and state-of-the-art methods within the RT counting framework. Experimental results validate the effectiveness and robustness of SGN. They also demonstrate the feasibility, efficacy and generalizability of the proposed framework for cell counting in unlabeled images.

*Corresponding author.

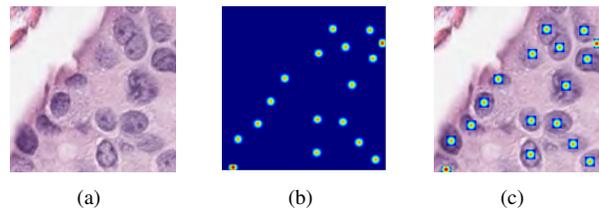


Figure 1. (a) A microscopy image from the PanNuke data set[7], (b) its DM generated from cell annotations and (c) the visualized effect whereas the image is overlaid with the DM (erased of its background) to illustrate that DM can reflect cell count and nuclear spatial details.

1. Introduction

Cell counting is a crucial step in many biomedical studies and applications. The number of cells in a microscopy image can indicate the presence of diseases, help differentiate tumor types and assist in understanding cellular and molecular genetic mechanisms[10]. As manual counting is tedious, time-consuming and prone to subjective errors, computer-aided counting methods are developed. Specifically, image analysis and machine learning techniques have successfully increased accuracy in cell counting. Advanced deep learning models with their feature representation power and generalizability have further reduced counting errors. Recently, it is discovered that density maps (DMs) generated from dot annotations of nuclei can represent cell counts and local spatial patterns as illustrated in

This work was partly supported by the Dawn Project of Shanghai Municipal Education Commission (No. 18CG38), the National Natural Science Foundation of China (No. 61702094) and the Young Scientists' Sailing Project of Science and Technology Commission of Shanghai Municipal (No. 17YF1427400).

Fig. 1, thus feeding more information to the deep networks than the numbers of cells. Therefore, many counting methods using DMs as labels are proposed[10][25][24]. However, there is still room for improvements. Most deep learning methods for cell counting are fully supervised, which require large amount of training data manually annotated *a priori*. They adopt the usual training-testing framework, training a model beforehand and testing it on unlabeled data afterwards. Meanwhile, unknown images may exhibit substantial new patterns that are unseen by the trained models, as there are large variations in image acquisition techniques, tissue backgrounds and cell sizes, shapes and counts. If there are new images to input, as there always are in real applications, the generalizability of the trained models and the information capacity of the historical training set may no longer support the application to analyze them. Systems with “human in the loop” such as the active/interactive learning methods[14] may alleviate these problems, but require constant inputs from the users to keep them working.

Motivated by the concerns above, we propose a novel DM regression framework for real-time (RT) cell counting. The strengths of this method are as follows. 1) Timeliness. It can learn from unlabeled images in a semisupervised manner and predict their DMs/cell counts simultaneously as the data are fed to the system. Moreover, it updates and refines the predicted cell counts iteratively during the semisupervised training. 2) Reliability. It optimizes the counting models by both historical data and new images, thus exploiting the prior information provided by manual labels and the variable patterns in unlabeled images. Specifically, it involves a training set renewing mechanism. Being fully automatic, it selects the images with predicted DMs of high confidence and adds them to the training set for the next round of optimization. In this way, it can avoid introducing erroneous information and ensure counting accuracy. 3) Flexibility. It allows almost any cell counting model to be embedded within it. Thus, it can leverage the existing powerful supervised models, turn them into semisupervised versions and further upgrade their performances in cell counting.

The contributions of this work lie in the following aspects. 1) It proposes perhaps the first semisupervised regression framework for RT cell counting, to the best of our knowledge. It allows for timely adaptation to dynamic data and saves manual efforts. It utilizes the powerful counting models and achieves state-of-the-art performances. 2) Within the framework, a new cell counting model, Semisupervised Graph-Based Network (SGN) is designed to address cell variability and image complexity. It constructs dynamic graphs of DM patches with a count-sensitive measurement[16]. With the graph constraint, it regularizes an encoder-decoder to represent underlying data structures and gain robustness for cell counting. 3) Thor-

ough experiments have been conducted on four sets of microscopy images with large variations. Two of them are benchmark data sets for cell counting, the simulated Bacterial Cells in Fluorescence-light Microscopy images (BCFM)[16] and the Bone Marrow (BM) images[25]. The other two are recently published Kaggle[2] and PanNuke data sets [7], including cells of various tissues from different patients. The experimental results validate the merits of the RT framework embedded with different models as well as the edges of SGN in cell counting.

2. Related Works

Most cell counting methods fall into three categories, counting after cell segmentation, counting after cell detection and direct counting[29]. The first approach counts cells/nuclei in segmentation masks[1][5][8][9][14][15][20][22][26]. There are unsupervised segmentation methods applied to blood cell counting[5] and fully supervised convolutional neural networks (CNN) that have largely increased counting accuracy[9][15][20][22][26]. While the segmentation networks usually require contour or area masks as labels, the detection-counting methods can work with simpler annotations, such as dots on nuclei or boxes around cells[10][23][25][27]. The counting accuracy of these two approaches depends on segmentation or detection results, which are heavily affected by cell occlusions and morphology variations. These problems can be solved by the regression models that directly predict cell counts or count-related DMs[10][24][27]. DMs are generated from point annotations of cells to utilize local spatial information and improve counting accuracy[4][10][16][24][25]. Xie et al. [24] proposed a fully convolutional regression network (FCRN), estimating DMs for arbitrary-sized input images. He et al. [10] used auxiliary CNNs (AuxCNNs) to assist in the training of a concatenated FCRN (C-FCRN), thus boosting DM regression performance and counting accuracy.

However, these fully supervised models achieve their success at the cost of large amounts of labeled data [17]. While the manual annotations are hard to come by, there are ample unlabeled data. Thus, it is natural to resort to semisupervised methods that can leverage the unlabeled data in training. It can even adapt to dynamic data sets such as videos for object tracking [3][11]. There are two typical ways to realize semisupervised learning, label propagation and graph embedding. The first approach uses a powerful model trained on annotated data to infer high-quality pseudo-labels for unlabeled data, which are fed back to the model to train it [11][12][17]. It opens the door to learning from real unlabeled, large scale data. The second strategy constructs graphs to exploit connectivity patterns between labeled and unlabeled samples to improve learning

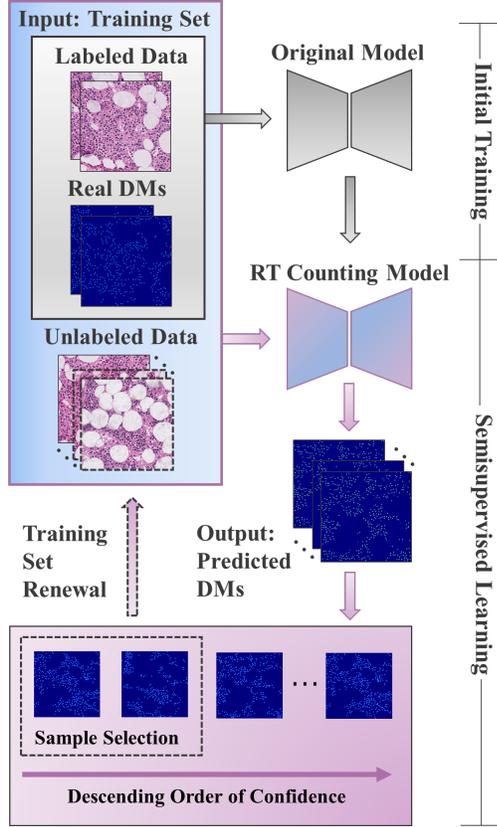


Figure 2. Illustration of the proposed real-time counting framework.

performance[3][6][18][19]. Though being proven flexible and powerful, semisupervised learning has not been well applied to cell counting based on DM regression.

3. Method

Inspired by the success of semisupervised learning in computer vision [12][17], we adopt the notion and propose a generic DM regression framework for real-time cell counting. To specifically address cell variability and image complexity, we design a SGN within the framework. Before beginning describing the method, we define basic definitions and notations. Denote a microscopy image by $\mathbf{X}_i \in R^{B \times J}$ with B variables and J pixels, where $i = 1, 2, \dots, N^{(l)} + N^{(u)}$, $N^{(l)}$ and $N^{(u)}$ represent numbers of historically labeled images and newly collected unlabeled images, respectively. For RGB images, $B = 3$. Suppose the number of cells in \mathbf{X}_i is C_i . If \mathbf{X}_i is annotated with dots ($\mathbf{y}_i = \{y_{i,k}\}_{k=1}^J$) at nuclear centers, the groundtruth density map $\mathbf{m}_i = \{m_{i,j}\}_{j=1}^J$ of \mathbf{X}_i is composed of

$$m_{i,j} = \sum_k N(j; y_{i,k}, \sigma^2 \mathbf{1}_{2 \times 2}), \quad (1)$$

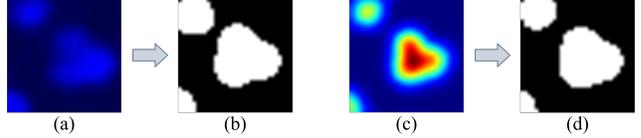


Figure 3. Illustration of the morphology masks for confidence measurement. (a) An image patch from BCFM (b) and its binary morphology mask. (c) The predicted DM of the image patch and (d) is the binary morphology mask. The white pixels indicate nuclear areas and black for background. The masks are obtained via unsupervised thresholding and dilation.

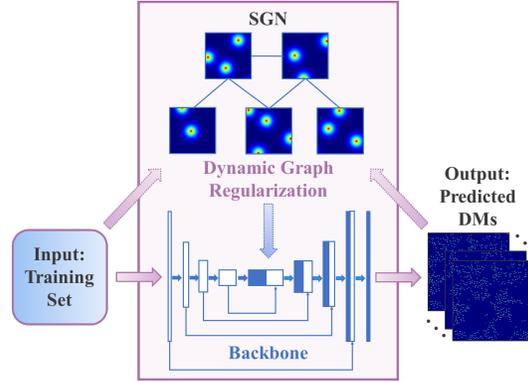


Figure 4. Illustration of the proposed specific model, SGN.

where $N(\bullet)$ is the normalized 2D Gaussian kernel. The relation between DM and cell count is defined by

$$C_i \approx \sum_j m_{i,j}, \quad (2)$$

whereas the tiny approximation errors can be omitted.

3.1. Real-Time Cell Counting Framework

As shown in Fig. 2, the RT framework comprises two phases: initial training and semisupervised learning. At first, we train a computational model $M(\bullet)$ for cell counting with historically labeled images $\{\mathbf{X}_i\}_{i=1}^{N^{(l)}}$ in a fully supervised manner. This way, the original model (coloured in grey in Fig. 2) acquires basic feature representation ability. It should be noted that $M(\bullet)$ can be almost any supervised counting model, given the flexibility of the framework.

As unlabeled new images $\{\mathbf{X}_i\}_{i=1}^{N^{(u)}}$ flow in, we switch to the semisupervised mode and let both labeled and unlabeled data train the network. In this way, $M(\bullet)$ iteratively adapts to the dynamic data and enhances representation ability. In an arbitrary semisupervised iteration t , while being optimized, the model $M(\bullet)$ predicts a DM $\mathbf{m}'_{i,t}$ for each unlabeled image $\mathbf{X}_i \in \{\mathbf{X}_i\}_{i=1}^{N^{(u)}}$ and estimates the number of cells by Eq. (2). As the process continues, the real-time counting errors are gradually reduced.

Algorithm 1 The RT Counting Framework

Input: the historically labeled data $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(1)}}$ and the upcoming unlabeled images $\{\mathbf{X}_i\}_{i=1}^{N_t^{(n)}}$;

Step 1 Initial Training:

1a) Pretrain the model $M(\cdot)$ on $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(1)}}$;

1b) Let the semisupervised iteration number $t = 0$ and initialize the training set $\{\mathbf{X}_i, \mathbf{a}_i\}_{i=1}^{N_t}$ as $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(1)}}$.

Step 2 Semisupervised Learning:

2a) Feed $\{\mathbf{X}_i\}_{i=1}^{N_t^{(n)}}$ to the pretrained $M(\cdot)$, infer their initial DMs and collect them in set $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(n)}}$;

// Training set renewal;

2b) Sort $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(n)}}$ by the morphology confidence described in Subsection 3.1 in descending order;

2c) Select the top $N_t^{(u)}$ samples $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(u)}}$ and let $\{\mathbf{X}_i, \mathbf{a}_i\}_{i=1}^{N_{t+1}} \leftarrow \{\mathbf{X}_i, \mathbf{a}_i\}_{i=1}^{N_t} \cup \{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(u)}}$, $N_{t+1} = N_t + N_t^{(u)}$, $\{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_{t+1}} \leftarrow \{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t} \cup \{\mathbf{X}_i, \mathbf{m}'_i\}_{i=1}^{N_t^{(u)}}$, $N_{t+1} = N_t + N_t^{(u)}$ and $t \leftarrow t + 1$;

2d) Optimize $M(\cdot)$ with the renewed set $\{\mathbf{X}_i, \mathbf{a}_i\}_{i=1}^{N_{t+1}}$;

2e) Predict DM \mathbf{m}'_i and repeat Steps **2a)**-**2e)** until $N_t^{(n)} = 0$;

Output: the predicted DM \mathbf{m}'_i and the cell count estimated by Eq. (2) for each new image.

Layer No.	Layer type	Filter dimension	Filter number
1	Conv.+ReLU	3×3×3	8
2	Conv.+ReLU	3×3×8	16
3	Conv.+ReLU	3×3×16	32
4	Conv.+ReLU	3×3×32	64
5	Conv.+ReLU	3×3×64	64
6	Upsample+Conv.+ReLU	3×3×128	32
7	Upsample+Conv.+ReLU	3×3×64	16
8	Upsample+Conv.+ReLU	3×3×32	8
9	Upsample+Conv.+ReLU	3×3×16	8
10	Conv.+ReLU	1×1×8	1

Table 1. Architecture of the backbone in SGN.

In the semisupervised stage, a training set renewing mechanism is activated. It is designed to ensure the reliability of the system. As shown in Fig. 2, it selects the otherwise unlabeled samples associated with highly confident predicted DMs and adds them to the training set to optimize $M(\cdot)$. Intuitively, we prefer DMs that can faithfully mirror the cell patterns in the original images, meaning that the network has indeed learned some features from the training data. Therefore, we measure the confidence of by the Euclidean Distance (ED) between a binary mask obtained from the original image \mathbf{X}_i (Fig. 3 (a), (b)) and that

of $\mathbf{m}'_{i,t}$ (Fig. 3 (c), (d)). The smaller the distance is, the higher the confidence is. The masks can reflect cell morphology patterns such as shapes and areas that are crucial to recognition of occluded or overlapping cells. If the masks are similar to each other, it means that the prediction is consistent with the original image, thus being highly confident. Theoretically, the morphology masks can be produced by an arbitrary unsupervised segmentation method. Empirically, we use thresholding and dilation operations [29], which require little efforts on parameter tuning. For double insurance in reliability, each time the system only selects a very small amount ($N_t^{(u)}$) of samples from the new data and adds them to the training set. With erroneous data largely avoided, the updated training set can positively help model optimization. Since the renewing process does not require any annotations or priors, it enables automatic counting without increasing manual workloads.

The implementation details are summarized in **Algorithm 1**. Throughout the semisupervised iterations, the groundtruth DMs of historically annotated data are involved since they provide the most reliable prior knowledge. The difference between the proposed method and the usual label propagation is that the former does not take the highly confident DMs immediately as the final output for the unlabeled images. Instead, it adopts the DMs predicted at the end of semisupervised iterations. It should also be noted that once the semisupervised learning finishes, $M(\cdot)$ can be seen as a fine-tuned model with extended generalizability. It can be used for inferring cell counts in totally unknown images that are not involved in semisupervision. The inference mode is preferred over the real-time mode when the need for efficiency rather than accuracy is in dominant.

3.2. Semisupervised Graph-based Network

Within the proposed framework, we design a specific model, SGN for accurate cell counting. As illustrated by Fig. 4, SGN can be substituted for $M(\cdot)$ in **Algorithm 1**. The backbone of SGN is an encoder-decoder using DMs as regression labels. Here we use UNet[21] as it is designed to capture context and enables precise localization, which is helpful in spotting cells against complex background tissues. The structure of the backbone is provided in Table 1. As SGN is trained on both labeled and unlabeled data, the empirical loss is

$$\begin{aligned} \ell_{emp}^t = & \arg \min_{\mathbf{m}'_{i,t}} \sum_{i=1}^{N_t^{(1)}} \|\mathbf{m}_i - \mathbf{m}'_{i,t}\|^2 \\ & + \alpha \sum_{i=N_t^{(1)}+1}^{N_t^{(1)}+N_t^{(u)}} \|\mathbf{m}'_{i,t-1} - \mathbf{m}'_{i,t}\|^2, \end{aligned} \quad (8)$$

where α is the trade-off coefficient, set as zero in Step 1 and reset at a non-zero value in **Step 2 of Algorithm 1**.

To exploit underlying structures of DMs, we embed graph Laplacian within SGN as in Fig. 4. Depict the graph

of DMs by $G = \{V, E, \mathbf{W}\}$, where $V = \{\mathbf{a}_i\}$ (\mathbf{a}_i can be the groundtruth DMs for the manually annotated images or the predicted DMs for the unlabeled data), $E = \{e_{i,j}\}$ and $\mathbf{W} = \{w_{i,j}\}$ denote vertices set, edge set and edge weights, respectively. Link vertices \mathbf{a}_i and \mathbf{a}_j by edge $e_{i,j} = \{\mathbf{a}_i, \mathbf{a}_j | \mathbf{a}_i \in S(\mathbf{a}_j) \vee \mathbf{a}_j \in S(\mathbf{a}_i)\}$, where $S(\cdot)$ obtains the neighborhood of the argument. Instead of ED or other usual measures, we utilize Maximum-Excess-over-SubArrays (MESA)[16] to define $S(\cdot)$ because it is robust to local modifications but sensitive to change in number of objects, which is a pleasant characteristic for cell counting. Therefore, we split each \mathbf{a}_i into sub-patches \mathbf{f}_p^i and compute the MESA measurement between \mathbf{a}_i and \mathbf{a}_j by

$$d_{MESA}(\mathbf{a}_i, \mathbf{a}_j) = \max_{p \in P} \left| \sum_q^Q f_{p,q}^i - \sum_q^Q f_{p,q}^j \right|, \quad (4)$$

where $f_{p,q}^i$ stands for pixels in sub-patch \mathbf{f}_p^i and P is the index set of the sub-patches. Then find k vertices with the smallest MESA associated with \mathbf{a}_i to construct the neighborhood $S(\mathbf{a}_i)$. Further, we design the edge weights as

$$w_{i,j} = \begin{cases} \exp(-\|\mathbf{a}_i - \mathbf{a}_j\|^2 / \|\mathbf{a}_i - \mathbf{a}_j\|_1), & e_{i,j} \neq \emptyset \\ 0, & e_{i,j} = \emptyset \end{cases}, \quad (5)$$

where $\|\cdot\|_1$ is ℓ_1 -norm. Storing all the predicted DMs in matrix \mathbf{V} , we construct the graph regularization

$$\begin{aligned} \ell_{reg}^t &= \min \sum_{i,j} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 w_{i,j} \\ &= \min \text{Tr}(\mathbf{V}^T \mathbf{L}_i \mathbf{V}), \end{aligned} \quad (6)$$

where $\text{Tr}(\cdot)$ is the trace function and \mathbf{L} denotes the normalized graph Laplacian matrix[21], which is computed by $\mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$ and \mathbf{D} is a diagonal matrix whose entries are the row-wise or column-wise sums of \mathbf{W} . Finally, we obtain the overall loss

$$\ell^t = \ell_{emp}^t + \lambda \ell_{reg}^t, \quad (7)$$

where λ is a trade-off coefficient. Different from the manifold regularized regression network (MRRN)[29], SGN is designed for semisupervised learning as the loss function and the graph regularizer consider both annotated images and unknown samples. Further, SGN is more robust to variation in cell distribution than MRRN as the former establishes graph edges based on MESA while the latter only uses simple geometrics. When implemented, SGN is optimized by Adam optimizer [13] in **Step 2d**, **Algorithm 1**.

4. Experiments

4.1. Data Sets

To evaluate performance of the proposed method, four sets of microscopy images are used, namely, BCFM images[16], BM images[25], Kaggle 2018 Data Science

Bowl[2] and the PanNuke data set released in 2020[7]. The first two are benchmark data sets for cell counting with point annotations of nuclei. They can give a general evaluation of different counting methods. The last two are recently published with nuclear segmentation masks. We turn these masks into point annotations by locating the center of each cell area to generate DMs. These images are acquired through different means and exhibit great variability in cell type and morphology. They can demonstrate generalizability of the RT framework and effectiveness of SGN. In experiments, we split each data set into three sub-sets: a labeled set with groundtruth DMs for initial training, an unlabeled set for semisupervised learning which simulates the new data flowing in dynamically without any prior annotations, and a test set simulating the unknown images to be analyzed off-line. The performances on the unlabeled set and the test set mainly reflect real-time counting ability and generalizability, respectively. Details of each data set are provided as follows. 1) BCFM is a synthetic data set designed by Lempitsky et al.[16]. Each image contains 256×256 pixels. We randomly select 32 images from the first 50 images as the labeled set, use the 50-100 images as the unlabeled set and the last 100 samples as the test set. 2) BM contains 11 HE stained bright-field microscopy images, cropped from whole-slide images (40× magnification) of human bone marrow from eight different patients, each with 1200×1200 pixels[25]. We slice them into 600×600 sub-images as in the references[25]. We randomly select 15 of the 44 sub-images as the labeled set, 18 images as the unlabeled set and the rest 11 images as the test set. 3) Kaggle[2] contains 670 HE stained microscopy images. The images were obtained under different conditions, varying in size, cell type, magnification, and imaging method (bright field and fluorescence). The data set is designed to challenge an algorithm’s ability to generalize across these variations. We use 335 samples as the labeled set, 135 as the unlabeled set, and the remaining 200 as the test set. 4) PanNuke[7] consists of 7870 images of 19 different types of tissues. Each image has 256×256 pixels. As the publisher has already been split it into three subsets, we use the first one (2625 images) as the labeled set, the second one (2523 images) as the unlabeled set and the third one (2722 images) for as the test set.

4.2. Implementation Details

To demonstrate the flexibility of the proposed framework, we embed various supervised cell counting models, including three baseline networks, UNet[21] and FCN-A/B[24], and three advanced methods, Structured Regression (SR)[25], Count-Caption[4] and C-FCN+Aux[10]. Substituting a competing model “X” for $M(\cdot)$ in **Algorithm 1**, we create its counterpart for RT counting, i.e., RT-X. We also compare them with a self-training (ST) framework[28].

Despite the name, it is little different from semisupervised learning. After cell segmentation by ST, we count the cells in the predicted masks. Moreover, we implement MRRN[29] with the graph regularizer defined by Eq. (6) to serve as a supervised version of SGN. Among the competing methods, SR (2018), C-FCRN+Aux (2021) and ST (2020) can be viewed as state-of-the-art. Although some methods are initially proposed for cell detection or segmentation, they are adapted to cell counting with great care.

To be fair, we use the same configurations for the RT framework while different models are embedded. As input, each image of BCFM, Kaggle and PanNuke is sliced into 32x32 patches. Since the BM images are large, they are sliced into 50x50 patches. The batchsize is 32 during the initial training and semisupervised learning. We have carefully tuned the hyperparameters of each model until they reach their optimal performances as in the reference papers. Instead of the original distance map, density map is used for SR for better counting results. The initial learning rate is 0.001 and decayed with a factor 0.9 by 30 steps. For the proposed SGN, λ is 0.2 and the number (k) of nearest neighbors is 8. During its semisupervised learning phase, α is 1 and the number ($N_t^{(u)}$) of selected samples is 1 for BM, 2 for BCFM and Kaggle, and 25 for PanNuke.

The evaluation metrics are Mean Absolute Error (MAE) and Mean Relative Error (MRE)[10]. MAE directly reflects the number of miss-counted cells while MRE provides a relative assessment as it is independent of the total cell number in an image. The mean result and standard deviations of 5 repetitions of each experiment are reported. All the algorithms are run on Intel Xeon E3, 11G GPU and 64G RAM. The codes and optimized models will be available at <https://github.com/Yihouyihou/SGN>.

4.3. Experimental Analysis

1) RT counting

i) Feasibility and flexibility. Table 2 gives a comparison between real-time counting and off-line prediction. The error indexes of the RT methods are low enough to indicate that the framework is feasible and the real-time counting is a success. The framework learns from the unlabeled images and their cells simultaneously, thus reducing counting errors. The results also show that the RT framework is flexible enough to extend various supervised models to their semisupervised versions.

ii) Efficacy and generalizability. Fig. 5 presents some typical results by SGN. It visualizes four microscopy images overlaid with their groundtruth or predicted DMs. It shows that SGN can accurately count various types of cells in very different settings. Table 3 shows that SGN outperforms the ST method on all the data sets, demonstrating the superiority of the proposed RT framework. Moreover, each semisupervised model outperforms its supervised counter-

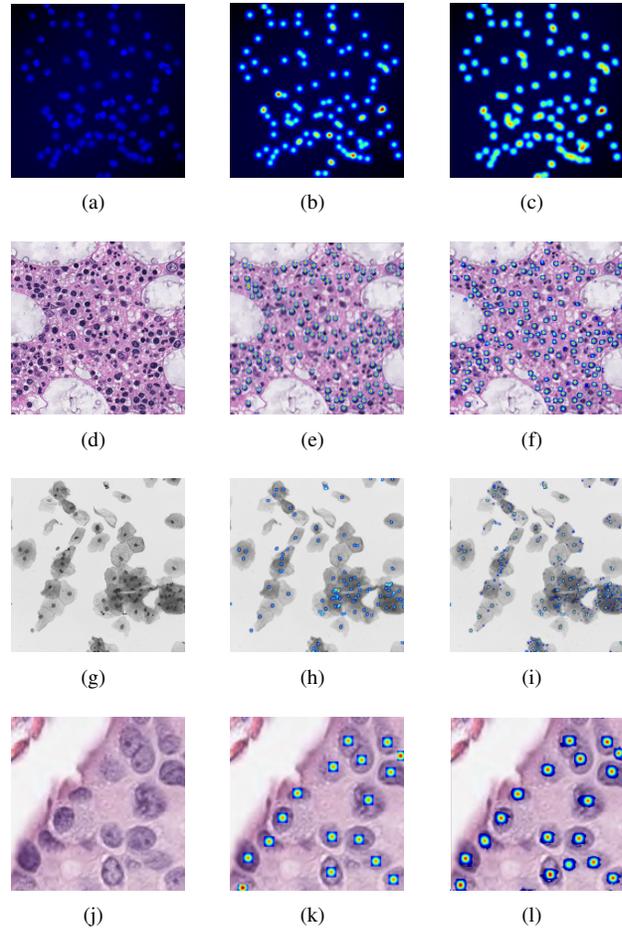


Figure 5. Typical results of SGN on the test sets. (a) A BCFM image and (b) its true and (c) predicted DMs (MAE: 1.75, MRE: 1.63%). (d) A BM image and (e) its true and (f) predicted DMs (MAE: 7.52, MRE: 8.35%). (g) A Kaggle image and (h) its true and (i) predicted DMs (MAE: 10.32, MRE: 15.53%). (j) A PanNuke image and (k) its true and (l) predicted DMs (MAE: 1.87, MRE: 10.64%). Overlaying the DMs with the original images in the same way as in Fig. 1 is only for the purpose of illustration.

part on the test sets. It indicates that the semisupervised learning mechanism is effective in reducing counting errors. By engaging ample amount of unlabeled data in optimization without increasing manual workload, the RT framework let the counting models gain representation power automatically. With enhanced generalizability, the semisupervised models yield better results for unseen data than the supervised ones. Meanwhile, SGN gives less counting errors than the competing models except for RT-C-FCRN+Aux on the BM data set. The cells in the BM images are overcrowded and difficult to recognize, as illustrated by Fig. 5(d). Fortunately, the proposed framework can leverage the powerful models such as C-FCRN+Aux and keep its edge in cell counting.

Data Set Method	BCFM		BM		Kaggle		PanNuke	
	MAE	MRE	MAE	MRE	MAE	MRE	MAE	MRE
UNet[21]	4.86±1.51	3.88	13.45±1.78	8.75	8.52±0.45	18.93	5.20±0.73	19.30
FCRN-A[24]	3.00±0.52	2.40	22.52±6.22	14.71	9.78±0.72	21.73	5.38±0.64	19.93
FCRN-B[24]	3.33±0.78	2.66	28.75±4.88	18.79	12.24±0.72	27.20	5.37±1.02	19.89
SR[25]	3.02±0.65	2.41	9.58±1.68	6.26	9.02±0.60	20.04	5.55±0.89	20.56
Count-Ception[4]	2.75±0.31	2.20	9.02±2.00	5.83	9.37±0.60	20.82	4.21±0.25	15.59
C-FCRN+Aux[10]	2.78±0.77	2.24	8.53±2.10	5.57	8.99±0.79	19.97	5.75±0.30	21.30
MRRN[29]	2.62±0.14	2.09	8.96±1.12	5.85	8.42±0.50	18.71	3.46±0.21	12.81
ST[28]	4.89±0.58	3.91	8.68±1.87	5.68	9.51±0.55	21.13	9.00±0.57	33.33
RT-UNet	2.98±0.74	2.38	8.45±1.58	5.52	6.72±0.45	14.93	4.31±0.34	15.96
RT-FCRN-A	2.56±0.51	2.05	19.45±7.70	12.70	7.77±0.56	17.26	4.71±0.42	17.44
RT-FCRN-B	2.67±0.52	2.14	21.29±7.20	13.91	7.31±0.66	16.24	4.89±0.56	18.11
RT-SR	2.87±0.41	2.29	6.40±1.85	4.18	7.98±0.38	17.73	4.87±0.42	18.04
RT-Count-Ception	2.24±0.20	1.79	6.22±1.57	4.06	7.65±0.40	17.00	3.45±0.29	12.79
RT-C-FCRN+Aux	2.23±0.65	1.78	5.65±1.35	3.69	6.79±0.52	15.08	4.90±0.31	18.15
SGN	1.98±0.12	1.58	5.72±1.03	3.73	5.38±0.30	11.95	2.82±0.29	10.44

Table 2. Evaluation of cell counting performances on the unlabeled sets in terms of MAE and MRE(%).

Data Set Method	BCFM		BM		Kaggle		PanNuke	
	MAE	MRE	MAE	MRE	MAE	MRE	MAE	MRE
UNet[21]	5.75±1.25	3.46	12.32±2.30	9.78	8.15±0.55	19.40	5.10±0.69	20.64
FCRN-A[24]	3.15±0.69	1.74	28.22±8.62	22.39	9.22±0.69	21.96	5.58±0.64	23.05
FCRN-B[24]	3.84±0.84	2.12	27.87±9.78	22.06	9.85±0.70	23.45	5.69±1.02	22.51
SR[25]	3.22±0.54	1.89	8.88±1.57	7.04	8.68±0.60	20.67	5.59±0.89	23.10
Count-Ception[4]	2.85±0.20	1.67	8.65±1.80	6.86	9.03±0.60	21.50	4.14±0.25	18.48
C-FCRN+Aux[10]	2.89±0.87	1.69	7.13±1.24	5.67	8.74±0.79	20.93	5.86±0.30	23.91
MRRN[29]	2.70±0.14	1.58	7.98±1.05	6.03	7.90±0.50	18.99	3.46±0.21	14.29
ST[28]	5.19±0.79	3.19	7.86±1.42	6.23	9.51±0.67	22.64	8.92±0.48	34.03
RT-UNet	4.33±0.85	2.67	10.61±2.11	8.42	7.86±0.45	18.71	4.45±0.33	18.39
RT-FCRN-A	2.97±0.69	1.64	25.52±8.50	20.25	9.00±0.58	21.42	4.96±0.48	20.50
RT-FCRN-B	3.03±0.84	1.67	25.33±8.45	20.13	9.21±0.60	21.92	5.23±0.64	21.61
RT-SR	3.00±0.42	1.78	8.02±1.95	6.36	7.43±0.35	17.69	4.88±0.39	20.16
RT-Count-Ception	2.78±0.21	1.63	7.95±1.56	6.31	8.78±0.42	20.70	3.87±0.27	15.99
RT-C-FCRN+Aux	2.70±0.34	1.57	6.99±1.35	5.54	8.01±0.47	18.67	5.20±0.29	21.49
SGN	2.68±0.12	1.57	7.64±0.89	6.06	6.57±0.30	15.64	3.38±0.18	14.08

Table 3. Evaluation of cell counting performances on the test sets in terms of MAE and MRE (%).

iii) Reliability and efficiency. Figs. 6(a)-(b) show that the RT counting error, MAE decreases as the semisupervised iterations move on. This is strong evidence for the reliability of the training set renewing mechanism. By selecting a small amount of unlabeled images with top morphology confidence values, the RT framework successfully manage the risks of introducing erroneous information and guarantees counting accuracy. Figs. 6(c)-(d) show that the learning errors are also minimized as all the loss curves of **Step 2d**), **Algorithm 1** in the semisupervised iterations con-

verges nicely. It indicates that SGN is a stable and reliable system. As most iterations end within 20 epochs, it is efficient to perform RT counting by SGN.

2) SGN

i) Robustness and count sensitivity. It is already shown by Tables 2-3 that SGN performs accurate cell counting. Here the effects of the MESA-based graph regularization, which plays a major role in SGN are analyzed. Fig. 7 gives two examples of DM neighborhood (sub-graph) used in the graph construction. It shows that the normalized MESA dis-

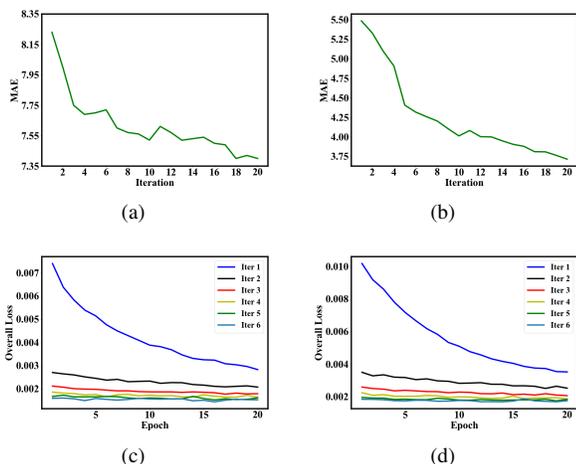


Figure 6. Analysis on reliability and efficiency of RT counting by SGN. Count MAE against semisupervised iterations for the unlabeled sets of (a) BM and (b) PanNuke. Overall loss against optimization epochs in the first six semisupervised iterations for the (c) BM and (d) PanNuke data sets.

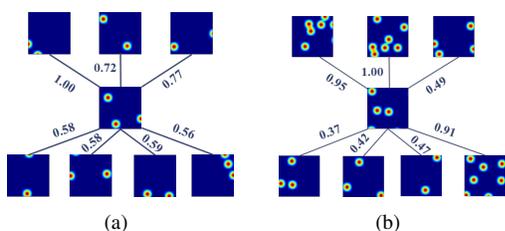


Figure 7. Sub-graphs of predicted DMs for test images of (a) Kaggle and (b) PanNuke. The numerical values indicate the normalized MESA.

tance between two DM patches is small as long as their cell counts are close, no matter how the cells vary in positions. Even when the cells are cut-off by the edges of DM patches, the edges and weights of the sub-graphs can capture the patterns of invariant cell counts. Thus, it is reasonable to employ MESA to build graph regularizers with robustness to spatial variations and sensitivity to cell counts.

ii) Ablation study. MRRN and UNet can be regarded as two ablated models of SGN since MRRN is the supervised version of SGN while UNet is the backbone. As shown in Fig. 8, SGN outperforms UNet and MRRN in different training conditions. It justifies the contributions of the graph regularization and the semisupervised mechanism to counting accuracy. Moreover, as the number of labeled images decreases, MAE of SGN increases the slowest among the competing models. It should be attributed to the graph embedding and the semisupervised framework of SGN.

3) Discussion. The performance analysis above has validated the merits of the RT framework and SGN in cell counting. Interestingly, as illustrated by Fig. 9, SGN is open to possibilities of cell detection and classification. With

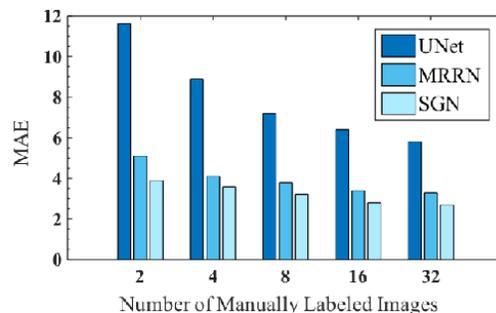


Figure 8. Count MAE of SGN, MRRN and UNet for the test set of BCFM.

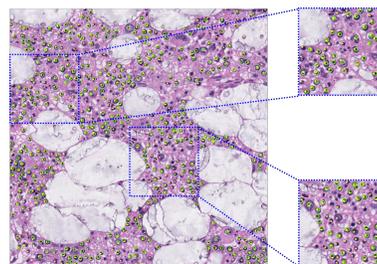


Figure 9. Visualization cell localization results of SGN for a BM image, whereas green rings stands for real cell locations and yellow dots indicate the predicted ones. Best viewed when zoomed in.

proper postprocessing of the predicted DMs, we can obtain cell locations as detection results as well as nuclear regions for classification. We could also use different feature maps as regression labels[25][9]. While the topic is not within the scope of this paper, we are not going to dwell on it but will include it in our future work.

5. Conclusion

To count cells simultaneously as new microscopy images are captured and fed to application systems, we propose a flexible and reliable RT counting framework. It is a regression system that uses density maps as labels. It could almost extend any supervised counting model to the semisupervised version and improve its performance in cell counting. The framework can automatically adapt to unlabeled new data and perform real time counting. With the delicate design for training set renewal, this process does not require extra manual efforts while the system gains representation power and refines the counting results. Moreover, we propose a robust counting model, SGN to address the specific issues such as cell variability and image complexity. Experimental results on four different data sets have validated the generalizability of the RT framework as well as the efficacy of SGN.

References

- [1] Salim Arslan, Tulin Ersahin, Rengül Çetin-Atalay, and Cigdem Gunduz Demir. Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images. *IEEE Trans. Medical Imaging*, 32(6):1121–1131, 2013.
- [2] Kaggle 2018 Data Science Bowl. <https://www.kaggle.com/c/data-science-bowl-2018/discussion/54426>.
- [3] Bingzhi Chen, Jinxing Li, Guangming Lu, Hongbing Yu, and David Zhang. Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray image classification. *IEEE J. Biomed. Health Informatics*, 24(8):2292–2302, 2020.
- [4] Joseph Paul Cohen, Geneviève Boucher, Craig A. Glastonbury, Henry Z. Lo, and Yoshua Bengio. Count-ception: Counting by fully convolutional redundant counting. In *2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017, Venice, Italy, October 22-29, 2017*, pages 18–26. IEEE Computer Society, 2017.
- [5] B. J. Ferdosi, S. Nowshin, F. A. Sabera, and Habiba. White blood cell detection and segmentation from fluorescent images with an improved algorithm using k-means clustering and morphological operators. In *International Conference on Electrical Engineering and Information Communication Technology*.
- [6] Cristina Gallego-Ortiz and Anne L. Martel. A graph-based lesion characterization and deep embedding approach for improved computer-aided diagnosis of nonmass breast MRI lesions. *Medical Image Anal.*, 51:116–124, 2019.
- [7] Jevgenij Gampfer, Navid Alemi Koohbanani, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir M. Rajpoot. Pannuke dataset extension, insights and baselines. *CoRR*, abs/2003.10778, 2020.
- [8] Ji Ge, Zheng Gong, Jun Chen, Jun Liu, John Nguyen, Zongyi Yang, Chen Wang, and Yu Sun. A system for counting fetal and maternal red blood cells. *IEEE Trans. Biomed. Eng.*, 61(12):2823–2829, 2014.
- [9] Simon Graham, Quoc Dang Vu, Shan e Ahmed Raza, Ayesha Azam, Yee-Wah Tsang, Jin Tae Kwak, and Nasir M. Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Anal.*, 58, 2019.
- [10] Shenghua He, Kyaw Thu Minn, Lilianna Solnica-Krezel, Mark A. Anastasio, and Hua Li. Deeply-supervised density regression for automatic cell counting in microscopy images. *Medical Image Anal.*, 68:101892, 2021.
- [11] Hongwei Hu, Bo Ma, Jianbing Shen, Hanqiu Sun, Ling Shao, and Fatih Porikli. Robust object tracking using manifold regularized convolutional neural networks. *IEEE Trans. Multim.*, 21(2):510–521, 2019.
- [12] Ahmet Iscen, Giorgos Toliass, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 5070–5079. Computer Vision Foundation / IEEE, 2019.
- [13] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [14] Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajadin, and Nasir M. Rajpoot. Nuclick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Anal.*, 65:101771, 2020.
- [15] Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, and Amit Sethi. A multi-organ nucleus segmentation challenge. *IEEE Trans. Medical Imaging*, 39(5):1380–1391, 2020.
- [16] Victor S. Lempitsky and Andrew Zisserman. Learning to count objects in images. In John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta, editors, *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, pages 1324–1332. Curran Associates, Inc., 2010.
- [17] Qimai Li, Xiao-Ming Wu, Han Liu, Xiaotong Zhang, and Zhichao Guan. Label efficient semi-supervised learning via graph filtering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 9582–9591. Computer Vision Foundation / IEEE, 2019.
- [18] Yawei Luo, Rongrong Ji, Tao Guan, Junqing Yu, Ping Liu, and Yi Yang. Every node counts: Self-ensembling graph convolutional networks for semi-supervised learning. *Pattern Recognit.*, 106:107451, 2020.
- [19] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5425–5434. IEEE Computer Society, 2017.
- [20] Peter Naylor, Marick Lae, Fabien Rey, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans. Medical Imaging*, 38(2):448–459, 2019.
- [21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells III, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.
- [22] Monjoy Saha and Chandan Chakraborty. Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. *IEEE Trans. Image Process.*, 27(5):2189–2200, 2018.
- [23] Mohammad Tofiqi, Tiantong Guo, Jairam K. P. Vanamala, and Vishal Monga. Prior information guided regularized deep learning for cell nucleus detection. *IEEE Trans. Medical Imaging*, 38(9):2047–2058, 2019.

- [24] Weidi Xie, J. Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Comput. methods Biomech. Biomed. Eng. Imaging Vis.*, 6(3):283–292, 2018.
- [25] Yuanpu Xie, Fuyong Xing, Xiaoshuang Shi, Xiangfei Kong, Hai Su, and Lin Yang. Efficient and robust cell detection: A structured regression approach. *Medical Image Anal.*, 44:245–254, 2018.
- [26] Fuyong Xing, Yuanpu Xie, and Lin Yang. An automatic learning-based framework for robust nucleus segmentation. *IEEE Trans. Medical Imaging*, 35(2):550–566, 2016.
- [27] 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 Transactions on Medical Imaging*, 35(1):119–130, 2016.
- [28] T. Zhao and Z. Yin. Weakly supervised cell segmentation by point annotation. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2020.
- [29] Yuxin Zheng, Zhao Chen, Yanfei Zuo, Xiaosong Guan, Zhaoyu Wang, and Xiao Mu. Manifold-regularized regression network: A novel end-to-end method for cell counting and localization. In *ICIAI 2020: The 4th International Conference on Innovation in Artificial Intelligence, Xiamen, China, May 8-11, 2020*, pages 121–124. ACM, 2020.