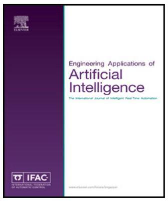
Engineering Applications of Artificial Intelligence 132 (2024) 107948

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Exponential distance transform maps for cell localization

Bo Li [a,1], Jie Chen [a,1], Hang Yi [a], Min Feng [b], Yongquan Yang [a], Qikui Zhu [c], Hong Bu [a,*]

[a] Department of Pathology and Institute of Clinical Pathology, West China Hospital, Sichuan University, Chengdu, China
[b] Department of Pathology, West China Second University Hospital, Sichuan University & Key Laboratory of Birth Defect and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China
[c] Department of Biomedical Engineering, Case Western Reserve University, OH, USA

ARTICLE INFO

Dataset link: https://github.com/Boli-trainee/MHFAN

Keywords:
Cell localization
Location map
Exponential distance transform map
Difference convolution

ABSTRACT

Cell localization in medical image analysis aims for precise identification of cell positions. Existing methods involve predicting density maps from images, followed by post-processing to extract cell location and number details. The quality of generated density maps significantly impacts the model's localization and counting performance. However, density maps produced with Gaussian kernels exhibit stacking in dense regions, resulting in inaccurate cell location information and suboptimal localization performance. In this study, we propose an exponential distance transform map that ensures accurate location information and provides well-defined gradient details for effective model learning, setting a new benchmark for high performance. Additionally, to address the challenge of substantial variations in cell color within images, we introduce a multi-scale gradient aggregation module that enhances the model's color recognition robustness through gradient information utilization. Experimental results across diverse datasets showcase notable improvements, establishing a novel benchmark for cell localization.

1. Introduction

The cell localization task is dedicated to accurately predicting the specific location and interpretable number of cells in an image, offering valuable insights for physicians in their diagnostic process. This task in the medical field encompasses a diverse range of applications, broadly classified into two types. The first involves directly deriving results from localization and counting information, as seen in calculating the Ki-67 index. This index plays a critical role in elucidating the molecular staging of breast cancer, assessing the administration of cytotoxic therapy, and predicting prognosis. The second type utilizes acquired information as a foundation to offer references for subsequent tasks. For instance, in clinical settings, this task provides cell information with robust interpretability for different departments, establishing a scientifically credible theoretical basis for the individualized treatment of tumor patients.

In recent years, the remarkable advancements in deep learning have facilitated the utilization of Convolutional Neural Networks (CNN) by numerous researchers for predicting cell location and number. For instance, some studies have employed a detection or segmentation based paradigm to localize individual cells (Shakarami et al., 2021; Alam and Islam, 2019; Kutlu et al., 2020; Stringer et al., 2021; Pachitariu and Stringer, 2022; Zhu et al., 2021b). However, the annotation of bounding boxes and precise edge information in this paradigm is both costly and labor-intensive, and many medical scenarios necessitate solely the cell's location information, without the need for size details. As a result, existing datasets in the field of cell localization typically offer only point-level annotations (Sirinukunwattana et al., 2016; Tofighi et al., 2019; Huang et al., 2020). To take advantage of point-level supervision, many researchers have adopted a location map-based cell localization paradigm (Huang et al., 2020; Lempitsky and Zisserman, 2010; Guo et al., 2021; Morelli et al., 2021a; Raza et al., 2019; Xie et al., 2018), which has emerged as a prevalent approach in this field.

1.1. Existing works for cell localization

The current cell localization paradigm is primarily based on density maps, as shown in Fig. 1(a). Initially, the cell images are inputted into a CNN model based on vanilla convolution to predict the corresponding density maps. Subsequently, a post-processing strategy is applied to obtain the location and count information. The existing density maps primarily capture the density information of cells within the image (Huang et al., 2020; Sirinukunwattana et al., 2016). The generation process can be roughly summarized as follows: Firstly, an activation

[*] Corresponding author.
E-mail address: hongbu@scu.edu.cn (H. Bu).
[1] Bo Li and Jie Chen contributed equally to this manuscript.

https://doi.org/10.1016/j.engappai.2024.107948
Received 6 July 2023; Received in revised form 22 December 2023; Accepted 18 January 2024
Available online 7 February 2024

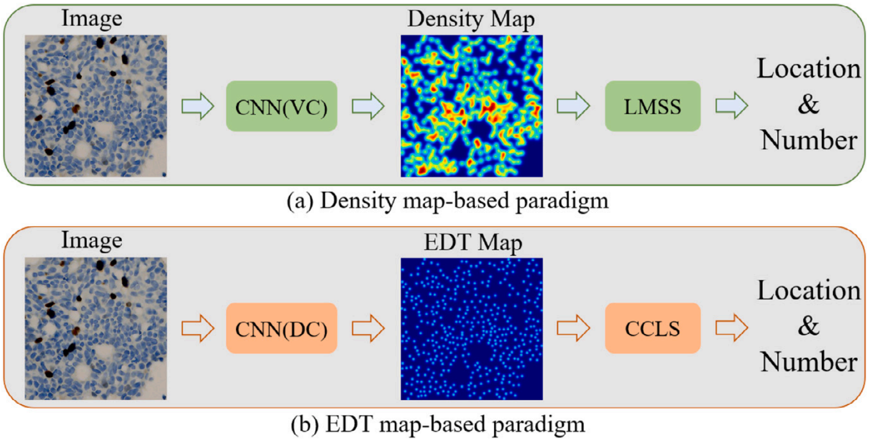


Fig. 1. Comparison of localization paradigms: (a) The existing paradigm, which relies on density maps, involves feeding the image into a Vanilla Convolution (VC)-based CNN model to generate a density map. Subsequently, a Local Maximum Search Strategy (LMSS) is employed to extract location and number information. However, this paradigm is plagued by several problems. Firstly, the VC-based model struggles to effectively handle images with substantial variations in cell color. Second, the density map does not provide accurate cell location information and ideal gradient details. Lastly, the LMSS encounters difficulties in handling location map noise. (b) In contrast, our proposed paradigm, based on EDT maps, operates by feeding the image into a Difference Convolution (DC)-based CNN model to obtain an EDT map. This EDT map is then utilized in conjunction with the Cell Center Localization Strategy (CCLS) to derive location and number information. Our paradigm addresses these challenges and offers notable improvements. Firstly, the DC-based model effectively mitigates the issues posed by large variations in cell color, thereby enhancing the model's color robustness. Secondly, the EDT map accurately provides crucial cell location information and ideal gradient details. Finally, the CCLS significantly enhances localization performance by effectively reducing background noise.

function $\delta(x - x_i)$ is placed in the center of each cell. Assuming that there are N cells in an image, which can be represented as

$$H(x) = \sum_{i=1}^{N} \delta(x - x_i). \tag{1}$$

To obtain a continuous density map, researchers (Huang et al., 2020; Zhang et al., 2019) commonly convolve the entire image using a Gaussian kernel $G_\sigma(x)$, where the size of the Gaussian kernel is determined either by the local density or by a fixed size, denoted as

$$F(x) = H(x) * G_\sigma(x). \tag{2}$$

To finally obtain cell location and number information based on the density map, Huang et al. (2020) employ separate predictions for negative and positive cells and subsequently process the output density maps with Local Maximum Search Strategy (LMSS). Specifically, the LMSS is employed to identify local maximum points within the density map, which are considered as candidate points. These candidate points are subsequently filtered to yield the final output.

1.2. Challenges

However, the existing localization paradigm described above still has several unresolved issues, which can be categorized into three points.

1.2.1. Density maps

The density map generated by Eqs. (1), (2) exhibits three notable drawbacks. Firstly, when the Gaussian kernel is small, the density map centers around a single pixel. As a result, there is too little supervisory information in the model and it is difficult to learn effective information. Secondly, if the Gaussian kernel is large, distinguishing dense cell regions on the density map becomes difficult, and the gradient information becomes less prominent. This hinders the model from effectively learning location information, as depicted in Fig. 2. Lastly, the gradient of the cell center, obtained from the Gaussian kernel, decreases at a slow rate, making accurate center localization difficult. These challenges persist to varying degrees, regardless of the Gaussian kernel value (Liang et al., 2022).

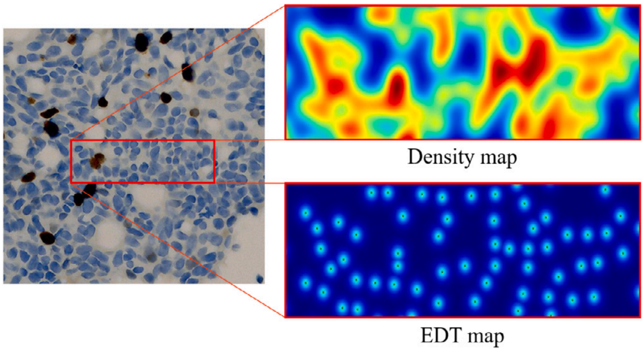


Fig. 2. Comparison of the responses between density map (Huang et al., 2020) and our EDT map. In areas of high cell density, the density map is difficult to discriminate and the location information of the cells is lost.

1.2.2. Post-processing strategy

Current post-processing strategies suffer from susceptibility to background noise points as they directly employ a LMSS on the output density map. Moreover, the presence of negative and positive cells leads to interference between them. For instance, when predicting lighter-colored negative cells, the threshold is vulnerable to darker-colored positive cells, which subsequently leads to the neglect of negative cells.

1.2.3. Large variations in cell color

Cell staining results exhibit significant variation between different laboratories due to subjective factors such as the staining technique, scoring method, and choice of scoring area. Even with fully automated staining procedures, it remains challenging to completely eliminate this discrepancy. Consequently, the presence of dramatic variations in cell color poses a considerable difficulty in all cell-related tasks. In Fig. 3, it is evident that cells with lighter colors are consistently disregarded in the EDT maps generated using conventional vanilla convolution. Notably, Huang et al. (2020) cleverly address this challenge by separately predicting positive (dark-colored) and negative (light-colored) cells.

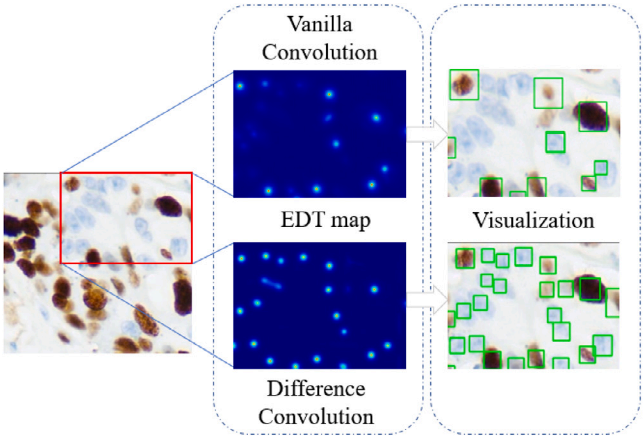


Fig. 3. The comparison of responses from various convolutions in the EDT map for cells exhibiting significant color variations reveals that difference convolution can effectively enhance the model's robustness against such variations in cell color.

However, this approach increases the complexity of dataset labeling and limits the model's applicability in realistic scenarios. Vanilla convolution, on the other hand, tends to ignore lighter-colored cells, resulting in an unreasonable EDT map and a significant decline in localization and counting performance.

1.3. Our method

To address the aforementioned challenges, we present a comprehensive overhaul of the cell localization paradigm. The revised paradigm, depicted in Fig. 1(b), incorporates three key components: a novel Exponential Distance Transform (EDT) map, an enhanced post-processing strategy for accurate cell location retrieval, and a CNN model based on difference convolution. Firstly, the EDT map offers precise location information and superior gradient details for each cell by leveraging point supervision. Subsequently, we introduce a Cell Center Localization Strategy (CCLS) that facilitates the simultaneous prediction of negative and positive cells, effectively mitigating background noise interference. Furthermore, to mitigate the challenges posed by dramatic variations in cell color, we propose a multi-scale gradient aggregation module based on difference convolution. This module generates a more rational EDT map. Our extensive experimental evaluation demonstrates that our approach yields substantial improvements in both cell localization performance, thereby establishing a new performance baseline for cell localization tasks.

In summary, the contributions of this paper are summarized as follows:

- We introduce a comprehensive update to the cell localization paradigm, encompassing a novel exponential distance transform map that enables precise cell localization, as well as a post-processing strategy called cell center localization strategy for accurate retrieval of cell location information.
- An innovative multi-scale gradient aggregation module based on difference convolution is proposed, which offers a novel solution to address the challenges posed by dramatic variations in cell color.
- Extensive experiments demonstrate that our method enables multiple models to achieve highly competitive cell localization performance, providing the new benchmark for future research.

The remaining sections of this paper are organized as follows: Related works of cell localization and counting and difference convolution are reviewed in Section 2. The proposed method including EDT map, CCLS and MGA module are detailed in Section 3. Extensive experiments and analysis are provided in Section 4, and Section 5 is the conclusion.

2. Related works

In this section, we briefly describe the current state of research in the field of CNN-based cell localization, mainly including detection-based and map-based approaches. In addition, related works on difference convolution are reviewed.

2.1. Detection-based methods

Detection-based methods for cell localization and counting aim to predict the location and number of cells in an image by detecting individual cell instances (Shakarami et al., 2021; Alam and Islam, 2019; Kutlu et al., 2020). To achieve rapid and efficient detection of blood cells in microscopic images, Shakarami et al. (2021) propose a detector based on YOLOV3 (Redmon and Farhadi, 2018). They enhance the receptive field by utilizing dilated convolutions and reduce model complexity through the use of depthwise separable convolutions, resulting in remarkable detection performance on the BCCD dataset. Similarly, Alam and Islam (2019) devise a cell detector based on YOLO for automatic identification and counting of red blood cells, blood cells, and platelets. To enhance detection accuracy, they employ a K-Nearest Neighbors and Intersection over Union based method to handle cases of multiple counts for the same cell. Furthermore, Kutlu et al. (2020) propose a deep learning and migration learning-based approach for automatic leukocyte detection in smear images.

The detection-based approach demonstrates outstanding detection performance in scenarios where cells are sparsely distributed. However, its effectiveness diminishes significantly as cell density increases. Furthermore, the process of annotating cell datasets with bounding boxes is intricate and costly, imposing limitations on its broader applicability. As a result, researchers employ point-based annotation methods (Tofighi et al., 2019; Sirinukunwattana et al., 2016; Huang et al., 2020) as a more common alternative.

2.2. Map-based methods

In order to leverage the available point-based supervision effectively, researchers have introduced probability maps (Sirinukunwattana et al., 2016) and density maps (Tofighi et al., 2019; Huang et al., 2020; Zhang et al., 2022; Liu et al., 2022). These maps provide insight into cell density variations across different regions of an image, collectively referred to as density maps. Most existing studies on cell localization utilize density maps as a fundamental component.

Typically, Xue et al. (2016) approach the cell counting task as a regression density map problem, training a residual convolutional neural network for this purpose. To alleviate the scarcity of datasets in the field of cell counting, Sirinukunwattana et al. (2016) propose a spatially constrained convolutional neural network and introduce a dataset named UW. Subsequently, Tofighi et al. (2019) propose a shape priors convolutional neural network and create the PSU dataset. Their approach involves predicting the probability of patches becoming the center of cells and aggregating these results to generate a probability map. In light of the relatively small size of these datasets, Huang et al. (2020) release the largest dataset named BCData in cell localization and counting. They design the U-CSRNet based on CSRNet (Li et al., 2018) to regress density maps for localization and counting. Recognizing the laborious nature of dataset-related efforts, Zhu et al. (2021a) propose a semi-supervised density map-based cell counting framework that can be trained using unlabeled images. In addition, some researchers are committed to applying cell localization and counting techniques to practical medical scenarios (Falk et al., 2019; Morelli et al., 2021b; Mandracchia et al., 2017; Hosseini et al., 2020; Zhang et al., 2023; Kleinovink et al., 2019; Huang et al., 2017). For example, to enable non-specialists to analyze their data remotely, Falk et al. (2019) offer an ImageJ plug-in specifically for U-Net-based cell localization counting and detection. Furthermore, some researchers (Guo et al., 2021, 2019; Li et al., 2022) focus on cell localization and counting in 2D and 3D scenarios. Recently, Chen et al. (2021a) train a CNN model to predict a two-dimensional direction field map, which is subsequently utilized for individual cell localization and counting. However, the reliance on the direction field map introduces challenges such as mutual overlap between cells, leading to a decrease in localization performance.

The aforementioned studies have made significant contributions to the advancement and implementation of cell localization and counting. Nevertheless, the current density map encounters challenges when differentiating densely populated cell areas, thereby impeding precise cell localization and counting. Additionally, the gradient information provided by the density map generated through Gaussian algorithms lacks clarity, posing difficulties in obtaining accurate cell location information.

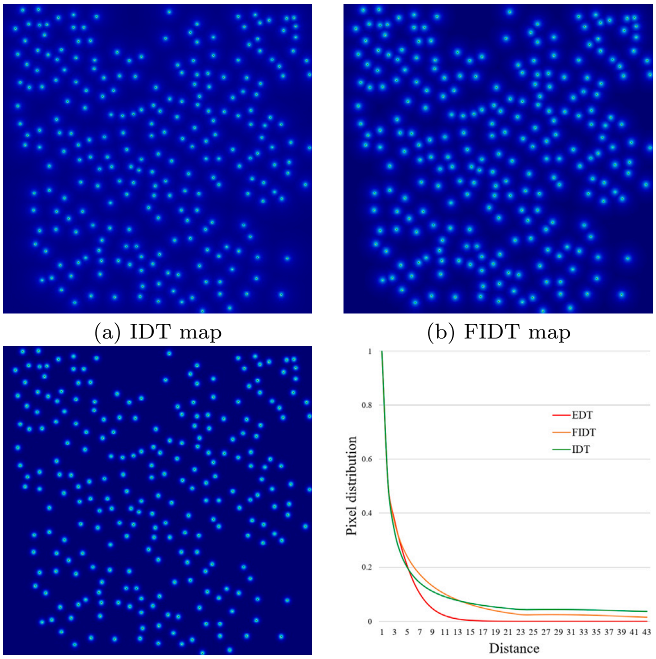
2.3. Works on difference convolution

To make full use of the local texture information of the image, Ojala et al. (2002) adopt the local relationships between adjacent features as effective discriminative features to improve the recognition performance. Inspired by this idea, Yu et al. (2020) propose the central difference convolution, which aggregates intensity and gradient information to capture local gradient details, thereby enhancing the model's resilience to environmental variations. Subsequently, Yu et al. (2021b) argue that the central difference convolution involves redundant difference operations on all neighborhood features, leading to computational inefficiencies. Therefore, they propose cross-centered difference convolution, which decouples into two symmetrically crossed suboperators horizontally vertically, and diagonally to reduce the computational cost. Later, Su et al. (2021) introduce difference convolution to the edge detection, using the difference information to enhance the model's ability to characterize the abrupt and detailed features of the edge background. Recently, Wang et al. (2023a) propose an adaptive multiscale differential graph convolutional network based on differential convolution for capturing implicit associations between joints and handling actions across different time intervals. In addition, to reduce feature redundancy while obtaining distinctive features, Zhou et al. (2021) and Bi et al. (2022) design a multi-grain perception module based on differential dilated convolutions. This module captures gradient features within different ranges to achieve a discriminative multigrain representation.

The utilization of difference convolution has significantly improved the characterization of image texture information by leveraging the local gradient relationships among adjacent features. In this paper, we pioneer the application of difference convolution in the domain of cell localization and counting, aiming to address the challenge posed by the considerable variations in cell color. By enhancing the gradient information within the images, our approach offers a promising solution to this challenge.

3. Method

The methodology's overview is illustrated in Fig. 1b. During the training phase, a CNN model establishes the mapping relationship between a cell image and a corresponding location map. Subsequently, the test phase processes the location map to obtain the exact location and number of cells. The performance of the cell localization and counting is mainly determined by three main components: the CNN-based model, the quality of the location map, and the post-processing strategy. Given that the CNN model's evaluation hinges on the location map, we sequentially introduce the Exponential Distance Transform (EDT) map proposed in this study, followed by the Cell Center Localization Strategy (CCLS) applied in the post-processing step, and conclude with the CNN model grounded in the Multi-scale Gradient Aggregation (MGA) module.



(a) IDT map (b) FIDT map (c) EDT map (d) Distance distribution

Fig. 4. The visualization comparison includes the IDT map, FIDT map, and EDT map. The IDT map (a) exhibits faster gradient decay in the foreground region, while decaying slower in the background, leading to a sustained high response. The FIDT map (b) demonstrates slower gradient decay in the background while maintaining a high response. In contrast, the EDT map (c) exhibits slower gradient decay in the foreground region and faster decay to 0 in the background region. Comparing the distance distributions of the IDT map, FIDT map, and EDT map (d), it is evident that the EDT map closely approximates the ideal distribution.

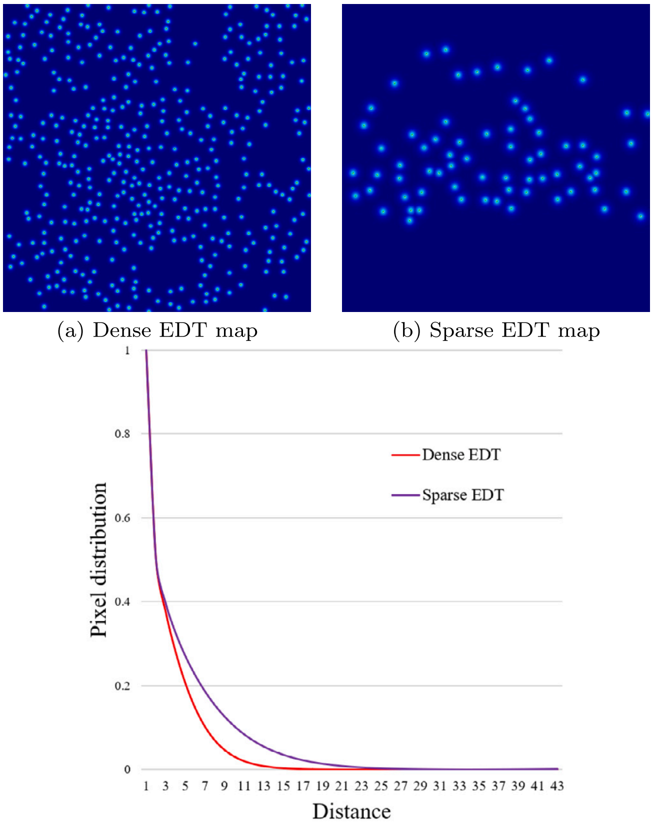
3.1. Exponential distance transform map

Building on prior works in crowd localization tasks (Olmschenk et al., 2019; Liang et al., 2022), we present the Exponential Distance Transform (EDT) map as a fresh approach that encompasses two pivotal stages: the Inverse Distance Transform (IDT) map and an adaptable scaling exponential optimization strategy.

To address the challenge of distinguishing dense regions in the density map, Olmschenk et al. (2019) propose the IDT map (Fig. 4(a)), which ensures independence among the targets in each region through distance inversion. However, IDT maps suffer from rapid decay of pixel values in target foreground features and slow decay in background areas. In response, Liang et al. (2022) introduce the FIDT map (Fig. 4(b)) that employs a primary term as an exponent of the distance function, enabling slower decay in the foreground and rapid decay in the background. Nonetheless, the linear relationship between the principal term and distance leads to slow decay in the background and non-zero response (Fig. 4(d)). To overcome this limitation, we propose an EDT map based on an adaptive scaling exponential optimization strategy (Fig. 4(c)). Compared to IDT and FIDT map, our EDT map demonstrates a more reasonable pixel distribution, with slower decay in the target foreground region and rapid decay to zero in the background region (Fig. 4(d)).

Initially, an array of identical dimensions to the original image is generated. The cells in the image are mapped to zero-valued pixel points in the array, while the remaining points are assigned a pixel value of 255. The nearest distance between the zero-valued pixel points in the array and each pixel point is then determined, denoted as

$$DT(x, y) = \min \sqrt{(x - x_i)^2 + (y - y_i)^2}, \text{ where } i \in I, \quad (3)$$



(a) Dense EDT map

(b) Sparse EDT map

(c) Distance distribution

Fig. 5. The visualization of EDT maps at different cell densities is presented for comparison. Figure (a) illustrates a rapid decay to zero in a dense cell scenario, while Figure (b) shows a slower decay to zero in a sparse cell scenario. The decay of pixel values in the EDT maps for these two scenarios is further compared in Figure (c).

where I denotes all zero-valued points in the image, i.e., the cell centroids. The Inverse Distance Transform (IDT) map is generated as

$$IDT = \frac{1}{DT(x,y) + C}, \tag{4}$$

where C is a constant introduced to prevent division by zero. The IDT map is constructed following the 1KNN method (Olmschenk et al., 2019) (Fig. 4(d)). The IDT map exhibits a steep gradient in the foreground region, enabling more precise localization. However, a drawback of the IDT maps is the rapid decay of gradients in the foreground area, while the gradients in the background region decay slowly, leading to a sustained high response.

To address this issue, Liang et al. (2022) proposed the FIDT map for optimization using the equation

$$FIDT = \frac{1}{DT(x,y)^{\alpha \cdot DT(x,y)+\beta} + C}, \tag{5}$$

where the primary term is exponentiated by the distance function in the IDT. This formulation ensures a slow decay of the target in the foreground region and rapid decay in the background region. However, the growth rate of the principal term $\alpha \cdot DT(x,y) + \beta$ tends to be linear with respect to the distance function $DT(x,y)$. Consequently, the FIDT map still exhibits slow decay in the background region, and the response is not reduced to zero.

Ideally, the location map exhibits a slow gradient decay in the foreground region and a fast decay to zero in the background region. To this end, we propose an adaptive scaling exponential optimization strategy to make the location map decay to zero rapidly in the background region, and propose different decay rates for different location maps, as shown in Eq. (6).

$$EDT = \frac{1}{DT(x,y)^{\frac{C_1 \cdot DT(x,y)}{Max(DT(x,y))+C_2}} + C_3}, \tag{6}$$

where C_1, C_2, and C_3 are hyperparameters that control the gradient of the distance map. $Max(DT(x,y))$ represents the maximum distance to the nearest zero-valued pixel for each pixel in the map. As the distance between the background and foreground points (cell centroids) approaches its maximum value, the pixel value in the background region rapidly decays to zero. Moreover, considering the diverse cell density distribution in different pathology images, the EDT map exhibits distinct decay rates for various densities, as depicted in Fig. 5. In comparison to the IDT and FIDT maps, our EDT map demonstrates a more reasonable distribution of pixel values: the foreground region associated with the target cells decays slowly, while the background region decays rapidly to zero (Fig. 4(d)). Subsequent ablation experiments involve further analysis and comparison of the EDT maps with the IDT and FIDT maps.

3.2. *Cell center localization strategy*

In the preceding subsection, we performed the transformation of the original cell image into an EDT map. The subsequent step involves a post-processing procedure to acquire the final cell locations and counts. Motivated by the Local Maxima Detection Strategy (LMDS Liang et al., 2022), we have devised a Cell Center Localization Strategy (CCLS) outlined in Algorithm 1. Our CCLS is specifically tailored for the task of cell localization, considering our optimized downsampling approach and adjusted filtering thresholds, distinguishing it from the LMDS method.

Algorithm 1 Cell center localization strategy

Input: The predicted EDT map.
Function: Get the coordinates and the total number of cell centers according to the EDT map.

```
EDT = max_pool2d(EDT, size=(11,11))
Cand_Spots = max_pool2d(EDT, size=(11,11))
Loc_map = Cand_Spots $\times$ EDT
if max(Loc_map) $\le 0.01$ then
    Coordinates = None
    number = 0
else
    Loc_map[Loc_map $\le$ threshold] = 0
    Loc_map[Loc_map $\ge$ 0] = 1
Number = sum(Loc_map)
Locations = nonzeros(Loc_map)
```

Output: The corresponding locations and number of cells.

Given an EDT map, we initially apply the maximum pooling technique to identify local maxima, followed by employing a threshold to eliminate background noise. In the context of pathological sections, where the thickness is typically between 3–5 μm and image acquisition is performed with the microscope's camera focused on the tissue area through the objective lens, extreme situations involving objects that are extremely small or excessively large, often encountered in photography, are absent. Nonetheless, the presence of tumor cell heterogeneity leads to variations in size among the targets. Therefore, to capture multiple candidate cell centers, we utilize a large pooling layer of 11 × 11 and perform two pooling operations. These resulting local maxima serve as potential cell center points. Furthermore, as illustrated in Fig. 2, given the comparatively weaker response of light-colored cells in the predicted EDT maps, we employ a lower threshold to ensure the inclusion of these light-colored cells.

Table 1
Comparison of existing publicly available datasets in the field of cell localization and counting.

Dataset	Venue	Images	Annotated objects	Resolution
UW (Sirinukunwattana et al., 2016)	TMI 16	100	29,756	500 × 500
PSU (Tofighi et al., 2019)	TMI 19	120	25,462	612 × 452
BCData (Huang et al., 2020)	MICCAI 20	1338	181,074	640 × 640
ccRCC Grading (Gao et al., 2021)	MICCAI 21	1000	70,945	512 × 512
CoNIC (Graham et al., 2021b,a)	ICCV 21	4981	495,179	256 × 256

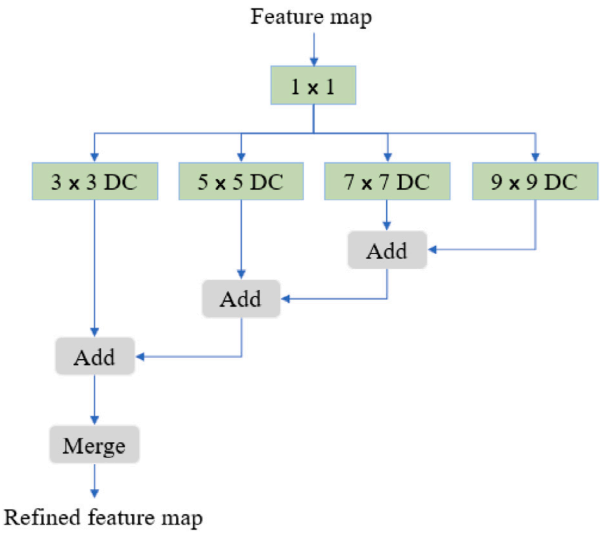


Fig. 6. Multi-scale gradient aggregation module. Capture multi-scale gradient information using branches with different dilated rates, and output after continuous superposition.

3.3. Multi-scale gradient aggregation module

In the previous subsection, the candidate points showed minimal distinction between the background noise and light-colored cells on the EDT map. To address this issue and mitigate the impact of significant color variations in cell localization and counting tasks, we present a novel Multi-scale Gradient Aggregation (MGA) module based on difference convolution. The MGA module aims to enhance the response of light-colored cells on EDT maps. To provide context, we first introduce the principle of difference convolution and subsequently present our MGA module.

The conventional vanilla convolution can be mathematically expressed as follows:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n), \tag{7}$$

where p_0 denotes the central position of the local receptive field R, p_n denotes the relative position of each value in the R to p_0, and $w(p_n)$ is the learnable parameter. On the other hand, the difference convolution (Yu et al., 2020) can be expressed as

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot (x(p_0 + p_n) - x(p_0)). \tag{8}$$

That is, each value $x(p_0 + p_n)$ in the local receptive field R is subtracted from its centroid $x(p_0)$ to form the local gradient information. Additionally, to incorporate the stronger semantic information offered by the conventional vanilla convolution, the final form of difference convolution is given by:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) + \theta(-x(p_0) \cdot \sum_{p_n \in R} w(p_n)), \tag{9}$$

where θ is a hyperparameter with a default value of 0.7, which we will discuss further in the ablation experiment.

We propose the novel MGA module based on difference convolution, with the aim of improving the robustness of the model to cell color variations. The MGA module uses difference convolution with different dilated rates to extend the receptive field of gradients, enabling the model to obtain rich gradient information. Specifically, as shown in Fig. 6, the input feature map is first convolved by 1 × 1 convolution to adjust the channel dimension, and then fed into the difference convolution branches with different dilated rates, respectively. There are 4 branches with corresponding field sizes of 3 × 3, 5 × 5, 7 × 7, and 9 × 9 respectively. Each branch is designed based on a 3 × 3 convolutional kernel with corresponding dilated rates of 1, 2, 3, and 4, respectively. For example, 5 × 5 DC indicates that this branch has a receptive field size of 5 × 5, that is, a convolution kernel size of 3 × 3 and a dilated rate of 2. Finally, the optimized feature map is obtained by stitching the outputs of different branches.

4. Experiments and analysis

4.1. Datasets and experimental details

In the field of cell localization and counting, widely used publicly available datasets include BCData (Huang et al., 2020), ccRCC Grading (Gao et al., 2021), CoNIC (Graham et al., 2021b,a), PSU (Tofighi et al., 2019), and UW (Sirinukunwattana et al., 2016), as summarized in Table 1. We provide a brief description of each dataset along with the experimental details.

The **BCData** (Huang et al., 2020) dataset is currently the largest dataset for cell localization and counting in the field. It comprises 1338 images of breast tumor cells, all with a uniform resolution of 640 × 640. The dataset includes a total of 181,074 annotated cells and is specifically designed for Ki-67 index assessment tasks. Notably, the BCData dataset exhibits three key characteristics: (1) uneven distribution of tumor cell density, (2) varying positive rates of cells, and (3) diverse cell traits. The dataset is divided into a training set (803 images), validation set (133 images), and test set (402 images). We would like to highlight that our experiments were conducted based on the U-CSRNet$^{\propto}$ implementation available at https://openi.pcl.ac.cn/xuf01/ki67, as indicated in Table 2.

The **ccRCC Grading** (Gao et al., 2021) dataset consists of 1000 H&E stained images that contain a total of 70,945 labeled cell nuclei. Each image has a resolution of 512 × 512, and each cell nucleus has an instance segmentation mask and a classification mask. In order to use this dataset for cell localization, we derived the centroids of the cells by a connected domain algorithm and used them for cell localization tasks.

The **CoNIC** (Graham et al., 2021b,a) dataset includes histology images stained with Haematoxylin and Eosin, captured at a 20x objective magnification, sourced from six distinct data repositories. It consists of 4981 images, each with a resolution of 256 × 256 pixels, and provides annotations for 495,179 labeled cell nuclei. Each image is accompanied by both instance segmentation and classification masks, offering detailed insights into the spatial distribution of cell nuclei and their corresponding classifications. This comprehensive dataset is tailored to support extensive research and analysis in the field of cell image recognition. Adopting the same processing approach as the ccRCC Grading dataset, the processed dataset can be accessed **here**.

The **PSU** (Tofighi et al., 2019) dataset consists of 120 images of porcine colon tissue, each with a resolution of 612 × 452. It encompasses 25,462 annotated cells and represents cross-sections of colonic

Table 2
Comparison of localization performance with different location maps and post-processing strategies on the BCData validation dataset. Models default to a categorical prediction method that distinguishes between negative and positive cells, while models marked with an asterisk * represent uniform predictions. We have indicated the best-performing records in each group with bold font, as done for all the subsequent tables.

Methods	Map	Post-process	Positive F1/Pre/Rec	Negative F1/Pre/Rec	Average F1/Pre/Rec(%) ↑
SC-CNN (Sirinukunwattana et al., 2016)	Density	LMSS	79.8/77.0/82.8	77.8/73.4/82.9	75.2/82.9/78.8
CSRNet (Li et al., 2018)	Density	LMSS	82.9/82.4/83.4	81.4/80.9/81.9	81.7/82.6/82.2
U-CSRNet (Huang et al., 2020)	Density	LMSS	86.3/86.9/**85.7**	**85.2**/84.4/**86.0**	**85.7**/85.6/**85.9**
U-CSRNet$^{\propto}$	Density	LMSS	84.7/84.4/85.0	84.6/84.7/84.5	84.7/84.5/84.8
U-CSRNet	Density	CCLS	86.1/87.2/85.0	83.9/83.6/84.3	85.0/85.4/84.7
U-CSRNet	EDT	LMSS	85.8/88.3/83.4	83.9/84.8/83.0	84.8/86.5/83.2
U-CSRNet	EDT	CCLS	**86.5**/**89.5**/83.8	84.8/**85.2**/84.5	**85.7**/**87.4**/84.1
U-CSRNet*	Density	LMSS	–	–	85.2/85.4/84.9
U-CSRNet*	Density	CCLS	–	–	85.5/**86.7**/84.3
U-CSRNet*	EDT	LMSS	–	–	85.9/85.4/86.5
U-CSRNet*	EDT	CCLS	–	–	**86.9**/86.6/**87.2**

epithelial cells. The dataset deliberately includes areas with artifacts, over-coloring, and autofocus failures to capture real-scene outliers. We employed the first 90 images as the training set and the remaining 30 images as the validation set.

The **UW** (Sirinukunwattana et al., 2016) dataset encompasses 100 H&E-stained histological images of colorectal adenocarcinoma, each with a resolution of 500 × 500. This dataset comprises non-overlapping regions extracted from 10 full cross-section images of 9 patients, resulting in a total of 29,756 annotated cells. To construct the dataset, we randomly crop non-overlapping regions from the 10 whole images. We allocate 70 images to the training set and 30 images to the validation set, maintaining a training set to validation set ratio of 7:3.

Considering the similar scale of the aforementioned datasets, we opt to uniformly resize them to 512 × 512 before generating the corresponding EDT maps. Our experimental setup is as follows: we utilize the MSE loss function for optimization, set the learning rate to 1e–4, apply a decay rate of 1e–5 after 200 epochs, employ the Adam optimizer, and select a minimum batch size of 4. The experiments are conducted on a NVIDIA GeForce RTX 3090 GPU with approximately 24 GB of memory, and the project code will be publicly available at https://github.com/Boli-trainee/MHFAN

4.2. Evaluation criteria

This paper focuses on the task of cell localization and counting, with evaluation criteria primarily focused on the performance of localization and counting.

Localization criteria: In order to assess the accuracy of the model's localization, we employ F1 score, precision, and recall as evaluation metrics.

$$Precision = \frac{True\ \ Positive}{True\ \ Positive + False\ \ Positive} \tag{10}$$

$$Recall = \frac{True\ \ Positive}{True\ \ Positive + False\ \ Negative} \tag{11}$$

$$F1 = \frac{Precision \cdot Recall}{Precision + Recall}, \tag{12}$$

where $True\ \ Positive$ indicates a successful match when the distance between a given predicted point and the true value point is less than a threshold σ. The selection of thresholds is closely tied to the characteristics of real cell images. For this study, we choose two fixed threshold levels ($\sigma = 5, 10$) to evaluate the model's performance. A smaller threshold value corresponds to a higher level of precision in positioning accuracy.

Counting criteria: Instead of directly regressing the number of cells, this paper relies on the results obtained from cell localization. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are utilized to evaluate the counting performance of the model. The equations for MAE and RMSE are given as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i|, \tag{13}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}, \tag{14}$$

where m is the number of cells, y_i is the number of ground truth, and $\hat{y}_i$ is the predicted number for ith cell.

4.3. Experiments and analysis

In this section, we commence by first validating the effectiveness of the EDT map and the CCLS strategy. To align with the ultimate clinical objectives of cell localization tasks, this paper employs two approaches to validate the effectiveness of the proposed algorithm. One involves separate predictions of positive and negative cells, specifically designed for calculating the Ki67 index. The other approach is the more widely applicable unified prediction for all cells. Therefore, we initially implemented separate predictions using BCData, which provides information about cell positivity and negativity. Subsequently, we confirmed unified predictions on all datasets, as shown in Tables 2 and 3. Building on this foundation, we proceed to confirm the efficacy of the MGA module through module replacements. As demonstrated in Tables 4–9, replacing several existing popular models with the MGA module resulted in performance improvements, signifying the effectiveness of the MGA module.

4.3.1. EDT map and CCLS

To calculate the Ki67 index directly, BCData (Huang et al., 2020) opted for the direct prediction of two density maps corresponding to negative and positive cells. To ensure a fair comparison, we also employed a method that separately predicts negative and positive cells to validate the Ki67 predictive performance based on EDT maps and CCLS, as outlined in the upper section of Table 2. Subsequently, in order to extend the applicability of cell localization to a broader spectrum of clinical tasks, we adopted a uniform prediction approach to validate our method, as demonstrated in the lower section of Tables 2 and 3.

Based on the results from Tables 2 and 3, we can deduce four conclusions. Firstly, the CCLS enhances the post-processing procedure, leading to improved localization performance while maintaining the same location map. Secondly, the combination of Peak_local_max algorithms and EDT maps yields subpar results. Our observation suggests that this outcome may be attributed to the vulnerability of light-colored negative cells to interference from dark-colored cells, thereby diminishing localization accuracy. This factor contributes to the rationale behind the uniform prediction approach. Thirdly, our EDT maps demonstrate significant advantages when employed within the

Table 3
Comparing the localization performance using various location maps and post-processing strategies across four datasets: CoNIC, ccRCC, PSU, and UW dataset.

Methods	Map	Post-process	CoNIC F1/Pre/Rec	ccRCC F1/Pre/Rec	PSU F1/Pre/Rec	UW F1/Pre/Rec (%) ↑
U-CSRNet	Density	LMSS	76.0/77.2/75.9	85.3/84.1/85.5	78.1/77.6/78.7	79.5/77.8/81.2
U-CSRNet	Density	CCLS	76.4/78.0/75.8	85.6/84.9/86.3	78.5/77.8/79.1	79.5/79.0/80.1
U-CSRNet	EDT	LMSS	77.6/77.4/**77.9**	86.2/**85.0**/87.5	79.6/**78.9**/80.3	80.8/80.1/81.5
U-CSRNet	EDT	CCLS	**78.7**/**80.0**/77.5	**88.2**/84.6/**92.1**	**80.4**/78.2/**82.7**	**81.8**/**80.5**/**82.2**
U-Net	Density	LMSS	80.6/83.4/77.8	87.5/88.0/87.1	79.6/78.9/80.3	80.2/80.0/80.3
U-Net	Density	CCLS	81.2/84.6/77.9	87.9/88.8/86.9	80.1/79.8/80.5	80.8/80.1/81.6
U-Net	EDT	LMSS	82.0/85.6/78.4	89.2/91.0/**88.4**	80.5/80.1/80.9	81.6/80.8/82.4
U-Net	EDT	CCLS	**82.9**/**87.4**/**78.7**	**89.4**/**92.6**/86.4	**81.0**/**80.3**/**81.7**	**82.9**/81.5/**84.3**

Table 4
Quantitative comparison of localization and counting performance of different models on the BCData validation dataset.

Methods	Counting MAE/RMSE↓	Localization(5) F1/Pre/Rec(%) ↑	Localization(10) F1/Pre/Rec(%) ↑
UNext	22.8/28.8	67.5/66.4/68.5	81.9/80.6/83.2
U_CSRNet	**19.8**/**25.3**	75.7/74.3/77.0	**86.9**/86.6/87.2
MPViT	21.5/27.7	74.4/76.5/72.4	85.8/88.3/83.5
U-Net	22.3/28.3	75.6/73.1/78.2	86.3/83.4/89.3
Attention U-Net	23.4/28.5	75.1/71.9/78.5	86.1/82.5/**90.0**
TransUNet	23.6/29.6	76.2/**79.1**/73.5	86.3/**89.6**/83.2
Swin Transformer	19.9/25.8	76.1/74.7/77.5	86.8/85.2/88.3
VGG16	19.9/25.8	**77.7**/76.0/79.3	86.6/84.8/88.5
Hover-Net	20.9/27.2	77.6/75.6/**79.7**	86.7/84.4/89.0
HRNet	22.2/27.8	77.0/76.5/77.5	86.5/86.0/87.1
M_U_CSRNet	**18.9**/**24.8**	76.8/75.6/78.0	**88.0**/87.5/89.6
M_U-Net	20.7/25.6	77.1/76.3/78.1	87.6/85.5/**89.7**
M_HRNet	20.7/26.5	**79.6**/**79.9**/**79.4**	**88.0**/**88.3**/87.7

Table 5
Quantitative comparison of localization and counting performance of different models on the BCData test dataset.

Methods	Counting MAE/RMSE↓	Localization(5) F1/Pre/Rec(%) ↑	Localization(10) F1/Pre/Rec(%) ↑
UNext	20.4/27.0	68.9/68.8/69.0	82.4/82.3/82.5
U_CSRNet	18.1/23.8	73.8/73.7/74.0	85.6/85.4/85.7
MPViT	22.3/29.2	75.3/79.3/71.6	85.5/**90.1**/81.4
U-Net	24.9/33.4	76.7/**81.7**/72.1	85.7/80.7/**91.4**
Lite-UNet	18.1/24.3	76.5/77.0/76.1	86.3/86.3/86.4
Attention U-Net	19.6/24.9	77.1/74.9/79.5	86.5/84.0/89.1
TransUNet	17.7/23.3	77.3/76.5/78.1	86.9/86.1/87.8
Swin Transformer	**17.1**/**23.1**	78.1/77.8/78.4	87.1/86.7/87.4
VGG16	17.7/23.2	**79.2**/78.6/79.8	86.9/86.3/87.5
Hover-Net	18.3/23.8	**79.2**/78.3/**80.1**	87.0/86.1/88.0
HRNet	18.5/24.7	**79.2**/80.1/78.3	**87.3**/88.4/86.3
M_U_CSRNet	**17.4**/**22.5**	74.9/74.1/75.7	87.1/87.5/86.7
M_U-Net	22.8/31.0	77.8/78.5/76.9	86.9/82.1/**91.8**
M_HRNet	19.9/25.6	**79.3**/**80.2**/**78.3**	**87.2**/**88.3**/86.2

Table 6
Quantitative comparison of localization and counting performance of different models on the ccRCC Grading test dataset.

Methods	Counting MAE/RMSE↓	Localization(5) F1/Pre/Rec(%) ↑	Localization(10) F1/Pre/Rec(%) ↑
UNext	6.1/7.5	80.7/78.9/82.4	88.4/86.5/90.5
U_CSRNet	7.5/8.8	81.8/78.5/85.5	88.2/84.6/**92.1**
MPViT	6.2/7.8	82.0/82.5/81.6	88.5/88.4/88.7
U-Net	5.8/7.9	83.7/86.7/80.8	89.4/**92.6**/86.4
Lite-UNet	5.2/7.4	84.4/85.8/83.1	89.6/91.3/86.9
Attention U-Net	**4.6**/**6.0**	83.4/83.6/83.2	89.7/89.8/89.5
TransUNet	5.0/7.0	83.6/84.0/83.3	90.0/90.3/89.6
Swin Transformer	6.0/7.9	82.4/80.2/84.6	89.4/87.0/91.9
VGG16	5.3/6.8	83.9/82.8/84.9	89.4/87.8/90.9
W-Net	–/–	85.0/83.0/**88.0**	–/–/–
Hover-Net	4.8/6.3	85.4/84.3/86.5	90.2/89.7/90.7
HRNet	4.9/6.9	**85.6**/**87.3**/83.9	**90.4**/92.3/88.7
M_U_CSRNet	7.0/8.5	82.5/81.2/83.8	89.2/86.6/**91.9**
M_U-Net	5.3/7.0	84.6/85.8/83.5	90.6/**92.5**/88.7
M_HRNet	**4.7**/**6.2**	**86.2**/**87.3**/**85.0**	**91.3**/91.8/90.7

unified prediction paradigm (indicated by the asterisk in Table 2), surpassing the performance of commonly used density maps. Finally, the combination of EDT maps and CCLS, as proposed in this paper, achieves the highest localization performance in both separate and uniform predictions of cells.

To further validate the performance of our proposed EDT map and CCLS across different datasets and models, we conducted experiments on four additional datasets using a consistent prediction approach. Specifically, we employed the U-CSRNet (Huang et al., 2020) and the widely recognized U-Net (Ronneberger et al., 2015) in the field of medical imaging. The performance of the models was evaluated with a threshold value of $\sigma = 10$. As evident from Tables 2 and 3, our introduced EDT map exhibits superior performance compared to the existing density map, achieving the optimal localization performance in conjunction with the CCLS.

4.3.2. MGA module

The above comparison clearly demonstrates the superior performance of our EDT map and CCLS over existing density map and LMSS in the task of cell localization. In the subsequent experiments, we exclusively rely on EDT maps and CCLS for validation. Firstly, we reevaluated the localization performance of several widely used models across multiple datasets, including UNext (Valanarasu and Patel, 2022), U-CSRNet (Huang et al., 2020), MPViT (Lee et al., 2022), U-Net (Ronneberger et al., 2015), Lite-UNet (Li et al., 2024), Attention U-Net (Oktay et al., 2018), TransUNet (Chen et al., 2021b), Swin Transformer (Liu et al., 2021), hover (Graham et al., 2019), W-Net (Mao et al., 2021), and HRNet (Sun et al., 2019). It is worth noting that, in order to align the input–output structure of the aforementioned models with the cell localization task, we employed the same strategy as in HRNet. This involves overlaying feature maps from different levels and performing deconvolution to obtain the final output. Secondly, we validated the effectiveness of the proposed MGA module using several models. Specifically, these models encompass the U-CSRNet, the widely employed U-Net in medical image analysis, and the high-performing HRNet across multiple datasets in this field. We replaced the initial set of coding layers in these models with the MGA module to enhance the model's resilience to variations in cell color. For instance, in the case of HRNet (Sun et al., 2019), we substituted the first encoder layer with the MGA module, which is responsible for transforming the feature channel dimension from 3 to 64.

Tables 4 and 5 present the results for counting performance and localization performance on the BCData validation and test datasets. Tables 6, 7, 8 and 9 illustrate the performance of the models on the ccRCC Grading, CoNIC, PSU and UW datasets. In these tables, the prefix "M_" indicates the replacement of a portion of the encoder layer in the aforementioned model with our MGA module. From the analysis of these tables, the following insights can be derived: (1) Most of the classical models achieve remarkable counting and localization performance when utilizing EDT maps; (2) The incorporation of the

Table 7
Quantitative comparison of localization and counting performance of different models on the CoNIC test dataset.

Methods	Counting MAE/RMSE↓	Localization(5) F1/Pre/Rec(%) ↑	Localization(10) F1/Pre/Rec(%) ↑
UNext	25.1/33.0	71.6/76.9/66.9	77.6/83.5/72.6
U_CSRNet	17.8/24.7	72.6/73.7/71.5	78.7/80.0/77.5
MPViT	19.5/25.5	74.0/74.7/73.3	79.7/80.4/78.9
U-Net	20.0/24.9	77.7/81.9/73.9	82.9/87.4/78.8
Attention U-Net	**14.3/19.7**	79.0/79.9/78.1	**83.9**/84.9/82.9
TransUNet	20.0/27.5	74.1/76.8/71.7	80.5/83.3/77.8
Swin Transformer	24.1/30.8	76.3/81.4/71.8	80.8/86.2/76.0
HRNet	18.0/28.4	77.8/76.9/**78.7**	82.4/81.5/**83.4**
Hover-Net	20.4/25.1	**80.3/85.2**/75.4	**83.9/89.4**/79.1
M_U_CSRNet	16.7/22.5	74.1/76.2/72.0	80.2/83.6/76.8
M_U-Net	18.3/23.1	**78.9/82.0**/75.9	83.7/85.1/**82.3**
M_HRNet	**14.5/21.0**	78.6/80.5/**76.8**	**83.9/85.9**/82.0

Table 8
Quantitative comparison of localization and counting performance of different models on the PSU dataset.

Methods	Counting		Localization	
	MAE↓	RMSE↓	F1↑(σ = 5)	F1↑(σ = 10)
UNext	43.2	52.1	59.2	78.1
U_CSRNet	32.1	38.2	54.5	80.4
MPViT	36.6	44.7	61.1	81.0
U-Net	32.6	38.5	61.0	81.0
Lite-UNet	33.7	38.4	60.2	79.6
Attention U-Net	31.4	36.8	63.5	82.1
TransUNet	40.1	49.4	58.9	80.1
Swin Transformer	**26.6**	**32.1**	62.2	82.1
Hover-Net	26.9	33.1	64.5	82.6
HRNet	27.6	32.4	**66.1**	**83.5**
M_U_CSRNet	29.5	36.1	56.2	81.2
M_U-Net	30.4	34.7	63.1	82.3
M_HRNet	**24.8**	**30.3**	**69.6**	**85.3**

Table 9
Quantitative comparison of localization and counting performance of different models on the UW dataset.

Methods	Counting		Localization	
	MAE↓	RMSE↓	F1↑(σ = 5)	F1↑(σ = 10)
UNext	24.5	32.3	68.3	79.6
U_CSRNet	**21.7**	**29.9**	69.3	81.8
VGG16	22.4	30.6	71.7	83.3
U-Net	29.5	38.4	70.5	82.9
TransUNet	24.5	32.0	71.4	83.4
Swin Transformer	22.9	33.1	**72.0**	83.6
Hover-Net	22.3	31.2	71.8	**85.1**
HRNet	26.2	36.1	71.5	84.9
M_U_CSRNet	**20.1**	**26.5**	70.5	82.5
M_U-Net	29.1	37.5	71.6	84.0
M_HRNet	23.6	28.2	**72.2**	**85.2**

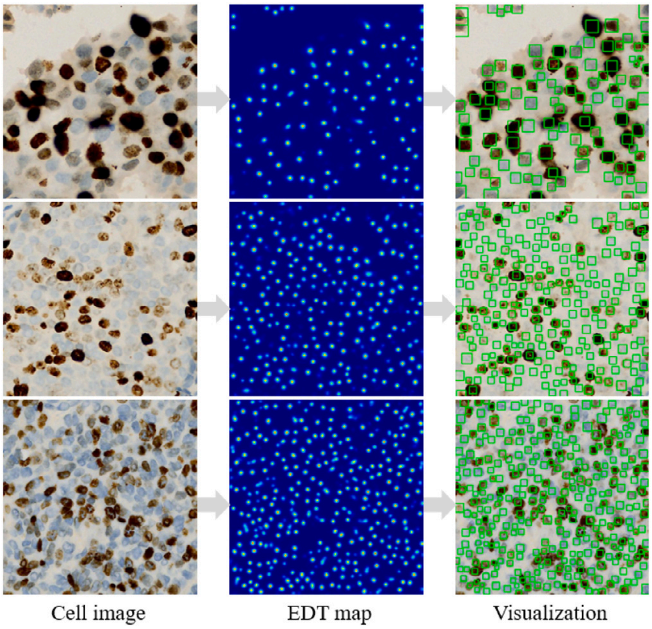


Fig. 7. Visualization results of model localization performance. The first column displays the original cell image, the second column illustrates the cell location map, and the third column presents the visualization of the localization results, where detected cell centers are highlighted with bounding boxes.

Table 10
Ablation experiments of hyperparameters C_1 and C_2 in EDT maps, comparison of counting and localization performance of models on EDT maps under different hyperparameter combinations.

Methods	C_1	C_2	Localization			Counting	
			Pre (%)	Rec (%)	F1↑	MAE↓	RMSE↓
IDT	–	–	84.9	83.5	84.2	21.7	27.3
FIDT	–	–	85.8	84.9	86.6	20.1	25.0
EDT	5	0.5	84.5	84.1	84.3	20.3	25.7
EDT	5	1	86.4	84.6	85.5	20.6	26.4
EDT	10	1	86.5	85.7	86.1	19.4	24.5
EDT(Ours)	10	0.5	**86.6**	**87.2**	**86.9**	**18.4**	**23.9**

MGA module significantly enhances both the counting and localization performance of the aforementioned models; (3) Among these models, the combination of HRNet and the MGA module yields the highest counting and localization performance. To provide a more visual representation of the localization performance, we display the localization visualization results for cell images using M_HRNet model in Fig. 7.

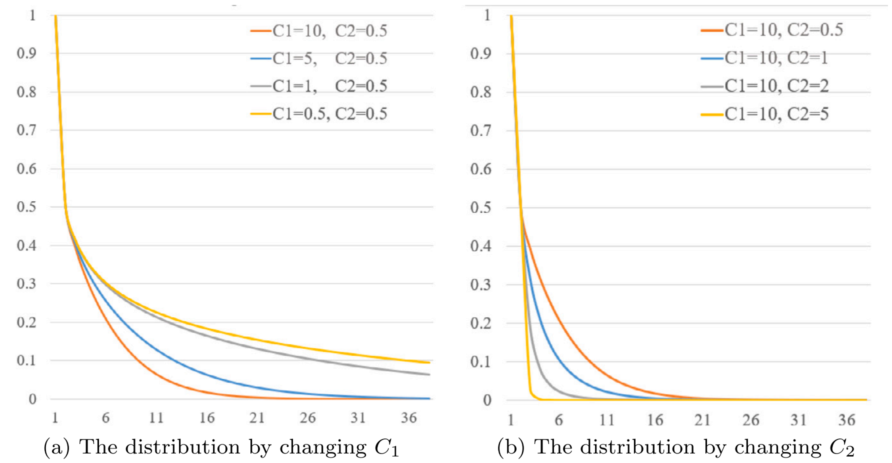
4.4. Ablation experiments

4.4.1. EDT map

As described in the previous Section 3.1, our goal is to achieve rapid decay of the EDT map in the center of the cell for better localization of the center point. The decay should then be slower throughout the foreground area and finally decay quickly in the background area, with most images decaying to zero beyond 15 pixels. To gain a deeper understanding of the decay process of the EDT map, we conducted ablation experiments on the hyperparameters C_1 and C_2 in Eq. (6). By changing only one of the hyperparameters, we observed the distance change of the EDT map, as shown in Fig. 8. The results indicate that the C_1 parameter primarily affects the decay rate of the EDT map outside the central region, while the C_2 parameter mainly affects the decay rate in the central region. Furthermore, we conducted ablation experiments based on U-CSRNet using various combinations of hyperparameters, as presented in Table 10. The table reveals two key findings: (1) Different combinations of hyperparameters significantly impact the final localization performance; (2) The table also includes a comparison of IDT and FIDT maps, demonstrating that our EDT map outperforms both in terms of localization performance. For this paper, we adopt the combination that achieves the best localization performance, with C_1 and C_2 set to 10 and 0.5, respectively. Additionally, we visualize the inferred results of several cell images using different location maps, as illustrated in Fig. 9.

4.4.2. Difference convolution

From Eq. (9), it is evident that the advantage of difference convolution lies in its gradient information. When the hyperparameter θ in Eq. (9) is set to 0, the difference convolution reduces to the conventional vanilla convolution, resulting in the loss of relative gradient information. Hence, we conducted ablation experiments on the



(a) The distribution by changing C_1

(b) The distribution by changing C_2

Fig. 8. The effect of changing C_1 and C_2 on the distribution of EDT map.

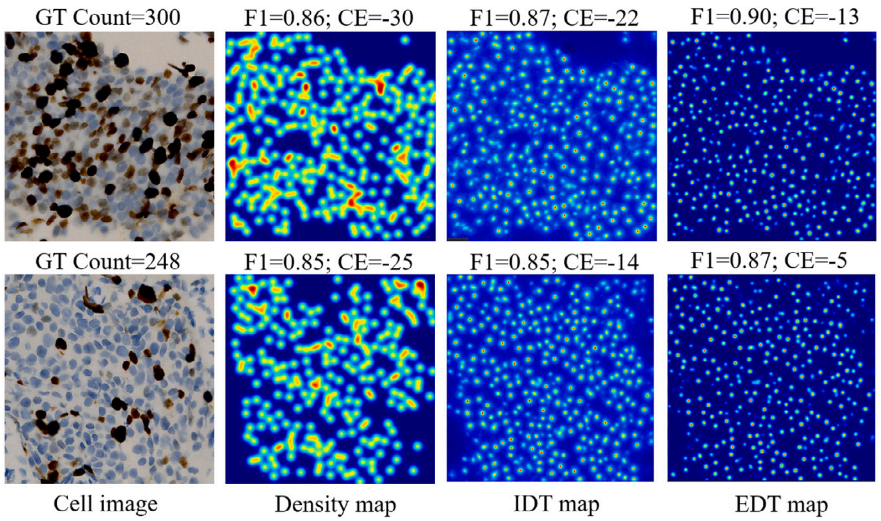
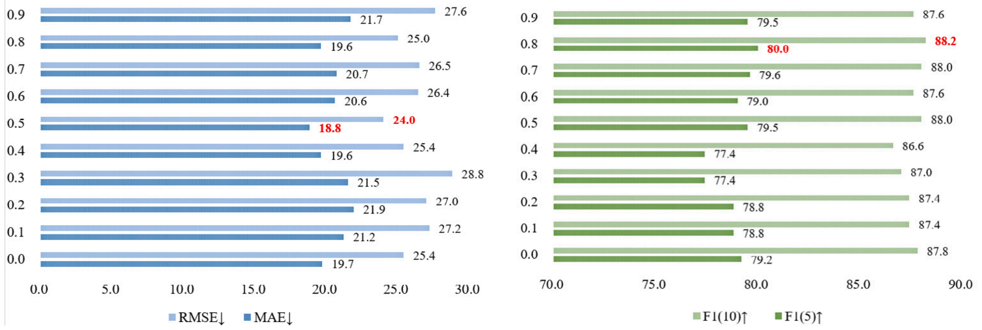


Fig. 9. Visual comparison of inferred results based on different location maps for several cell images, where CE indicates count errors.



(a) Counting

(b) Localization

Fig. 10. Ablation experiments of difference convolution hyperparameters in the MGA module. It is worth noting that when $\theta = 0$, the difference convolution degenerates into the conventional vanilla convolution.

hyperparameter θ to investigate its impact on cell localization and counting performance across a range from 0 to 1.

In this experiment, we employed the M_HRNet model on the BCData validation dataset, as illustrated in Fig. 10. The following findings were obtained: (1) The gradient information plays a significant role in the model's localization and counting performance. The hyperparameter θ directly influences the weight balance between gradient information and semantic information obtained by the model, thereby directly affecting performance outcomes. (2) Effective utilization of gradient information greatly enhances cell localization and counting performance. When the hyperparameter θ is set to 0, resulting in the degeneration of the difference convolution to the vanilla convolution, the model's performance notably declines. (3) The optimal localization and counting performance of the model is achieved with hyperparameters $\theta = 0.8$ and $\theta = 0.5$. It is worth noting that setting the hyperparameter $\theta = 0$ can lead to training instability. The primary reason for this lies in the near elimination of semantic information, making it challenging for the model to capture feature distributions. Therefore, we recommend avoiding the setting of $\theta = 0$ whenever possible.

5. Conclusion and outlook

Cell localization plays a crucial role in medical image analysis. This paper presents a substantial advancement in the field of cell localization, introducing several key innovations. Firstly, we propose an exponential distance transform map that accurately determines the location of cells, while maintaining a reasonable gradient. Additionally, we develop a corresponding cell center localization strategy that provides precise information regarding the final location and number of cells. Moreover, we introduce a novel multi-scale gradient aggregation module based on difference convolution, which enhances the model's ability to handle color variations. Extensive experimental evaluations demonstrate the remarkable performance improvement of our method in cell localization and counting tasks, establishing a new benchmark in the field.

Our future research will focus on advancing the cell localization task using the EDT map as a foundation. Our observations indicate that the ground truth representation of cells within the EDT map showcases a consistent distribution in all directions, with the cell center as the central point. Based on this finding, we propose to reframe the cell localization task as a feature alignment problem. Specifically, our objective is to align the complex and diverse cell distributions observed in pathological images with the uniformly distributed hillsides. This strategic approach allows us to leverage the concept of feature alignment to effectively address the inherent challenges. Consequently, our forthcoming investigations will focus on conducting a comprehensive exploration of this feature alignment methodology.

CRediT authorship contribution statement

Bo Li: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jie Chen:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision. **Hang Yi:** Investigation, Methodology, Validation, Visualization, Writing – original draft. **Min Feng:** Data curation, Formal analysis, Resources, Software. **Yongquan Yang:** Methodology, Software, Validation, Visualization. **Qikui Zhu:** Methodology, Software, Validation, Writing – review & editing. **Hong Bu:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

None Declared

Data availability

The experimental code and processed datasets will be publicly available at https://github.com/Boli-trainee/MHFAN.

Acknowledgment

This work was supported by the 1·3·5 project for disciplines of excellence (ZYGD18012); the Technological Innovation Project of Chengdu New Industrial Technology Research Institute (2017-CY02–00026-GX).

References

Alam, M.M., Islam, M.T., 2019. Machine learning approach of automatic identification and counting of blood cells. Healthc. Technol. Lett. 6 (4), 103–108.

Bi, Q., Zhou, B., Qin, K., Ye, Q., Xia, G.-S., 2022. All grains, one scheme (AGOS): Learning multigrain instance representation for aerial scene classification. IEEE Trans. Geosci. Remote Sens. 60, 1–17.

Chen, Y., Liang, D., Bai, X., Xu, Y., Yang, X., 2021a. Cell localization and counting using direction field map. IEEE J. Biomed. Health Inf. 26 (1), 359–368.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021b. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al., 2019. U-Net: deep learning for cell counting, detection, and morphometry. Nature Methods 16 (1), 67–70.

Gao, Z., Shi, J., Zhang, X., Li, Y., Zhang, H., Wu, J., Wang, C., Meng, D., Li, C., 2021. Nuclei grading of clear cell renal cell carcinoma in histopathological image by composite high-resolution network. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 132–142.

Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.-W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al., 2021a. Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684–693.

Graham, S., Jahanifar, M., Vu, Q.D., Hadjigeorghiou, G., Leech, T., Snead, D., Raza, S.E.A., Minhas, F., Rajpoot, N., 2021b. Conic: Colon nuclei identification and counting challenge 2022. arXiv preprint arXiv:2111.14485.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.

Guo, Y., Krupa, O., Stein, J., Wu, G., Krishnamurthy, A., 2021. Sau-net: A unified network for cell counting in 2d and 3d microscopy images. IEEE/ACM Trans. Comput. Biol. Bioinform. 19 (4), 1920–1932.

Guo, Y., Stein, J., Wu, G., Krishnamurthy, A., 2019. Sau-net: A universal deep network for cell counting. In: Proceedings of the ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. pp. 299–306.

Hosseini, S.H., Chen, H., Jablonski, M.M., 2020. Automatic detection and counting of retina cell nuclei using deep learning. In: Proceedings of the Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging. Vol. 11317, pp. 634–646.

Huang, Z., Ding, Y., Song, G., Wang, L., Geng, R., He, H., Du, S., Liu, X., Tian, Y., Liang, Y., et al., 2020. Bcdata: A large-scale dataset and benchmark for cell detection and counting. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 289–298.

Huang, K., Doyle, F., Wurz, Z.E., Tenenbaum, S.A., Hammond, R.K., Caplan, J.L., Meyers, B.C., 2017. FASTmiR: an RNA-based sensor for in vitro quantification and live-cell localization of small RNAs. Nucleic Acids Res. 45 (14), e130.

Kleinovink, J.W., Mezzanotte, L., Zambito, G., Fransen, M.F., Cruz, L.J., Verbeek, J.S., Chan, A., Ossendorp, F., Löwik, C., 2019. A dual-color bioluminescence reporter mouse for simultaneous in vivo imaging of T cell localization and function. Front. Immunol. 9, 3097.

Kutlu, H., Avci, E., Özyurt, F., 2020. White blood cells detection and classification based on regional convolutional neural networks. Med. Hypotheses 135, 109472.

Lee, Y., Kim, J., Willette, J., Hwang, S.J., 2022. Mpvit: Multi-path vision transformer for dense prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7287–7296.

Lempitsky, V., Zisserman, A., 2010. Learning to count objects in images. In: Proceedings of the Advances in Neural Information Processing Systems. p. 23.

Li, S., Ach, T., Gerig, G., 2022. Improved counting and localization from density maps for object detection in 2D and 3D microscopy imaging. arXiv preprint arXiv:2203.15691.

Li, Y., Zhang, X., Chen, D., 2018. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1091–1100.

Li, B., Zhang, Y., Ren, Y., Zhang, C., Yin, B., 2024. Lite-UNet: A lightweight and efficient network for cell localization. Eng. Appl. Artif. Intell. 129, 107634.

Liang, D., Xu, W., Zhu, Y., Zhou, Y., 2022. Focal inverse distance transform maps for crowd localization. IEEE Trans. Multimed..

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 10012–10022.

Liu, Z., Wang, Q., Meng, F., 2022. A benchmark for multi-class object counting and size estimation using deep convolutional neural networks. Eng. Appl. Artif. Intell. 116, 105449.

Mandracchia, B., Bianco, V., Wang, Z., Mugnano, M., Bramanti, A., Paturzo, M., Ferraro, P., 2017. Holographic microscope slide in a spatio-temporal imaging modality for reliable 3D cell counting. Lab Chip 17 (16), 2831–2838.

Mao, A., Wu, J., Bao, X., Gao, Z., Gong, T., Li, C., 2021. W-net: A two-stage convolutional network for nucleus detection in histopathology image. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. pp. 2051–2058.

Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., Zoccoli, A., 2021a. Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Sci. Rep. 11 (1), 1–11.

Morelli, R., Clissa, L., Dalla, M., Luppi, M., Rinaldi, L., Zoccoli, A., 2021b. Automatic cell counting in flourescent microscopy using deep learning. arXiv preprint arXiv:2103.01141.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Olmschenk, G., Tang, H., Zhu, Z., 2019. Improving dense crowd counting convolutional neural networks using inverse k-nearest neighbor maps and multiscale upsampling. arXiv preprint arXiv:1902.05379.

Pachitariu, M., Stringer, C., 2022. Cellpose 2.0: how to train your own model. Nature Methods 19 (12), 1634–1641.

Raza, S.E.A., AbdulJabbar, K., Jamal-Hanjani, M., Veeriah, S., Le Quesne, J., Swanton, C., Yuan, Y., 2019. Deconvolving convolutional neural network for cell detection. In: Proceedings of the International Symposium on Biomedical Imaging. pp. 891–894.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement.

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

Shakarami, A., Menhaj, M.B., Mahdavi-Hormat, A., Tarrah, H., 2021. A fast and yet efficient YOLOv3 for blood cell detection. Biomed. Signal Process. Control 66, 102495.

Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.-W., Snead, D.R., Cree, I.A., Rajpoot, N.M., 2016. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging 35 (5), 1196–1206.

Stringer, C., Wang, T., Michaelos, M., Pachitariu, M., 2021. Cellpose: a generalist algorithm for cellular segmentation. Nature Methods 18 (1), 100–106.

Su, Z., Liu, W., Yu, Z., Hu, D., Liao, Q., Tian, Q., Pietikäinen, M., Liu, L., 2021. Pixel difference networks for efficient edge detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5117–5127.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5693–5703.

Tofighi, M., Guo, T., Vanamala, J.K., Monga, V., 2019. Prior information guided regularized deep learning for cell nucleus detection. IEEE Trans. Med. Imaging 38 (9), 2047–2058.

Valanarasu, J.M.J., Patel, V.M., 2022. Unext: Mlp-based rapid medical image segmentation network. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 23–33.

Wang, X., Gan, Z., Jin, L., Xiao, Y., He, M., 2023a. Adaptive multi-scale difference graph convolution network for skeleton-based action recognition. Electronics 12 (13), 2852.

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., Yang, L., 2018. Efficient and robust cell detection: A structured regression approach. Med. Image Anal. 44, 245–254.

Xue, Y., Ray, N., Hugh, J., Bigras, G., 2016. Cell counting by regression using convolutional neural network. In: Proceedings of the European Conference on Computer Vision. pp. 274–290.

Yu, Z., Qin, Y., Zhao, H., Li, X., Zhao, G., 2021b. Dual-cross central difference network for face anti-spoofing. arXiv preprint arXiv:2105.01290.

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., Zhao, G., 2020. Searching central difference convolutional networks for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5295–5305.

Zhang, C., Chen, J., Li, B., Feng, M., Yang, Y., Zhu, Q., Bu, H., 2023. Difference-deformable convolution with pseudo scale instance map for cell localization.

Zhang, A., Shen, J., Xiao, Z., Zhu, F., Zhen, X., Cao, X., Shao, L., 2019. Relational attention network for crowd counting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6788–6797.

Zhang, S., Zhang, X., Li, H., He, H., Song, D., Wang, L., 2022. Hierarchical pyramid attentive network with spatial separable convolution for crowd counting. Eng. Appl. Artif. Intell. 108, 104563.

Zhou, B., Yi, J., Bi, Q., 2021. Differential convolution feature guided deep multi-scale multiple instance learning for aerial scene classification. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing. pp. 4595–4599.

Zhu, Y., Chen, Z., Zheng, Y., Zhang, Q., Wang, X., 2021a. Real-time cell counting in unlabeled microscopy images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 694–703.

Zhu, Y., Huang, R., Wu, Z., Song, S., Cheng, L., Zhu, R., 2021b. Deep learning-based predictive identification of neural stem cell differentiation. Nat. Commun. 12 (1), 2614.

Further reading

Lin, Q., Xiongbo, G., Zhang, W., Cai, L., Yang, R., Chen, H., Cai, K., 2023. A novel approach of surface texture mapping for cone-beam computed tomography in image-guided surgical navigation. IEEE J. Biomed. Health Inf..

Liu, H., Xu, Y., Chen, F., 2023. Sketch2Photo: Synthesizing photo-realistic images from sketches via global contexts. Eng. Appl. Artif. Intell. 117, 105608.

Lu, S., Yang, J., Yang, B., Yin, Z., Liu, M., Yin, L., Zheng, W., 2023. Analysis and design of surgical instrument localization algorithm. CMES-Comput. Model. Eng. Sci. 137 (1).

Wang, W., Qi, F., Wipf, D., Cai, C., Yu, T., Li, Y., Yu, Z., Wu, W., 2023b. Sparse Bayesian learning for end-to-end EEG decoding. IEEE Trans. Pattern Anal. Mach. Intell..