



# Review of research on the instance segmentation of cell images

Tingxi Wen, Binbin Tong, Yu Liu, Ting Pan, Yu Du, Yuping Chen\*, Shanshan Zhang\*

College of Engineering, Huaqiao University, Quanzhou 362021, China

## ARTICLE INFO

### Article history:

Received 2 May 2022

Revised 27 October 2022

Accepted 30 October 2022

### Keywords:

Cell segmentation

Target detection

Deep learning

## ABSTRACT

The instance segmentation of cell images is the basis for conducting cell research and is of great importance for the study and diagnosis of pathologies. To analyze current situations and future developments in the field of cell image instance segmentation, this paper first systematically reviews image segmentation methods based on traditional and deep learning methods. Then, from the three aspects of cell image weak label extraction, cell image instance segmentation, and cell internal structure segmentation, deep-learning-based cell image segmentation methods are analyzed and summarized. Finally, cell image instance segmentation is summarized, and challenges and future developments are discussed.

© 2022 Published by Elsevier B.V.

## 1. Introduction

As the basic unit of the human body, a cell consists of a nucleus, cytoplasm, and cell membrane. The detection and study of cells is a vital topic, facilitating the accurate assessment of health conditions. The diagnosis and classification of many diseases depend on the number, shapes, and sizes of cells. Obtaining information about the status of some cells quickly through cell image analysis is vital for the study and diagnosis of pathologies. Rapid developments in microscopic imaging and computer technology have enabled medical personnel to acquire precise cell image data. This task would be labor intensive if the human eye alone is used. Therefore, using computer vision processing techniques in segmenting and counting cell images can reduce the burden imposed on medical personnel and improve efficiency. In recent years, developments in biomedical science have facilitated the use of computer vision processing technologies in medical imaging, and a large number of image analysis techniques have been used in automatic cell segmentation, cell counting, tumor cell identification, and pathology image detection. Computer vision processing technologies facilitate the characterization of morphology and number of cells in microscopic images through cell segmentation, not only providing a basis for medical diagnosis but also having essential significance for the reconstruction of cellular ultrastructure and promoting research related to quantitative cellular information analysis.

Cell segmentation is an essential basis for disease diagnosis and classification, and the detection of many diseases cur-

rently depends on the number and morphology of cells. Cell segmentation divides cellular images into disjoint regions according to features, such as grayscale, geometry, color, and texture. These features show similarity or consistency in the same areas while varying widely in different regions [1]. Cell segmentation, an essential prerequisite for cell image recognition and counting, is a fundamental and vital topic in medical imaging [2]. In addition, cell segmentation is a classical problem, and a general and efficient segmentation method is needed. Currently, cell images are segmented manually by pathologists, and this method is tedious and time consuming. An automatic segmentation method is urgently needed to reduce manpower and resource consumption.

Computer vision processing technologies play an increasingly important role in medical image processing, and deep learning algorithms related to convolutional neural networks (CNNs) have attracted considerable interest. These technologies allow the batch analysis of cell images and effectively prevent human errors. Developments in CNN technologies have led to renewed interest in the study of cell images for biological image research. Compared with traditional segmentation methods, deep learning algorithms can automatically generate optimal features for cell images without requiring researchers to design features, and deep learning algorithms have high accuracy. However, these algorithms suffer from disadvantages, and current deep learning-based cell segmentation algorithms need large amounts of labeled data to provide models for training. The accuracy of an algorithm depends on the quality of labeled and high-quality marked data, such as medical expert's manually labeled data. Thus, the cost is high. A high-precision model requires a large number of model parameters for support. However, inference efficiency of a model decreases with the number of model parameters.

\* Corresponding authors.

E-mail addresses: [chenyup@hqu.edu.cn](mailto:chenyup@hqu.edu.cn) (Y. Chen), [zshanshan@hqu.edu.cn](mailto:zshanshan@hqu.edu.cn) (S. Zhang).

## 2. Image Segmentation

Image segmentation technology is an essential part of image analysis because subsequent tasks, such as target recognition, depend on image segmentation, and its accuracy depends on image segmentation quality [3]. Image segmentation has attracted interest since the 1970s. Researchers have proposed different solutions for different fields and tasks. However, owing to the breadth and depth of a subject, no universal algorithm has been established. A large number of studies related to automatic image segmentation have been conducted, and many methods have been proposed, including the threshold segmentation method, region growth method, watershed algorithm, active contour segmentation method, and deep-learning-based algorithms. The following is a brief description of the most commonly used segmentation algorithms.

### 2.1. Traditional methods for image segmentation

#### 2.1.1. Threshold segmentation method

The basic principle of the threshold segmentation method is to classify pixel features by setting different thresholds. This method can be divided into single-threshold and multithreshold according to the number of selected thresholds. In single threshold segmentation, an image is divided into foreground and background regions according to a selected grayscale threshold compared with the grayscale in the image, and its transformation formula is shown in Equation (1).

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq T, \\ 0 & f(x, y) < T. \end{cases} \quad (1)$$

where  $f(x, y)$  is the input image,  $g(x, y)$  is the output image, and  $T$  is the threshold value. Therefore, the selection of threshold is the key technique of the method. Otsu [4] proposed the OTSU method, which is based on the grayscale histograms of images and selects a suitable threshold by calculating the maximum interclass variance of the foreground and background. It is a classical single-threshold segmentation method. However, in scenes where an image is not uniformly illuminated, a slightly blurred edge target image becomes the background. To prevent this scenario, Wei et al. [5] proposed the adaptive thresholding method, which divides a picture into  $N$  parts and uses the moving average method to calculate a threshold value. The method can calculate a local threshold value according to the brightness distribution of the different regions of an image. Methods for selecting threshold values include Mode, entropy, P-tile, and minimum error.

As a traditional image segmentation method, the OTSU method has a simple implementation process and low computational requirement and applies to many types of fields. A threshold segmentation method is limited because it only focuses on grayscale features of pixels and does not consider spatial features. Moreover, noise has substantial impacts on threshold segmentation methods, which are not robust. The segmentation effect of a threshold segmentation method is poor in images with little difference between foreground and background gray values.

#### 2.1.2. Region growing method

A neighborhood-based image segmentation method connects neighboring pixels with the same or similar properties to obtain a segmented image. This method requires some a priori knowledge, such as seed pixels and criteria for target boundaries. One of its advantages is that it uses spatial information and inter-pixel correlation. The basic idea of the method is to continuously add neighboring pixels with similarity to a seed region and finally achieve the detection and image segmentation of a target object. First, suitable seed points are selected as the starting point of growth, and

according to specific growth rules, the points that meet growth properties are merged into a growth region to form a new growth region. This process is repeated until no growth is possible, thus segmenting the target from the background [6]. Therefore, the selection of seed points, formulation of growth rules, and setting of termination conditions are the factors determining the segmentation performance of the region growing method.

The method has the advantages of being computationally simple and have the ability to partition out connected regions with the same characteristics. The disadvantages are that it requires artificial selection of growth points, is sensitive to noise, and is easy to produce voids and over-segmentation.

#### 2.1.3. Watershed algorithm

The watershed segmentation method is based on the mathematical morphology of topological theory, and its basic idea is to consider an image as a topological landform in geodesy. The gray value of each pixel in the image represents the elevation of the point, and each minimal local value and its influence area constitute a catchment basin. The basin's boundary is the watershed [7,8], as shown in Fig. 1. The watershed algorithm has a good segmentation effect on the segmentation problem of adherent objects. However, the algorithm easily causes the over-segmentation of images during image segmentation. It is sensitive to noise, so images require preprocessing [9].

Mitko et al. [10] proposed an improved watershed segmentation method in which watershed segmentation is conducted on preprocessed images before segmentation error regions are eliminated. The subsequent merging of segmentation results at different scales allows accurate image detection and segmentation. Chen et al. [11] proposed a method for the automatic segmentation and identification of cone photoreceptor cells through morphological processing and by using a watershed algorithm; The method was able to accurately segment and identify cone photoreceptor cells, being comparable to manual identification. Marker control watershed is a way to reduce over-segmentation by marking difference between foreground and background objects in an image during watershed segmentation. Nurçin et al. [12] improved the segmentation efficiency of marker-controlled watershed segmentation by introducing a local minimum histogram background segmentation method and a selective hole-filling algorithm. Jung et al. [13] proposed an improved watershed stacked cell segmentation method with adaptive H-minima, which uses the average roundness of stacked cells as an evaluation function and adaptively calculates the h-value as a way to suppress over- and under-segmentation situations in the segmentation process and improve segmentation performance. Koyuncu et al. [14] proposed an adaptive H-minima marker-controlled watershed cell nucleus segmentation method by iterating h-values with a preset cell area and radius as thresholds that can be used by a marker in selecting an appropriate h-value for subsequent watershed segmentation. In addition, the fuzzy clustering segmentation algorithm can improve the over-segmentation problem generated by the watershed algorithm. Fuzzy clustering classifies things according to their similarity. The fuzzy C-mean clustering algorithm is the most commonly used fuzzy theory algorithm because of its simple operation, fast convergence, and suitability for processing images with large data volumes.

#### 2.1.4. Active contour segmentation method

In addition to watershed algorithms, the active contour segmentation method can resolve image overlaps effectively. The principle of the method is to calculate the minimum value of an energy function by constructing the energy function and using the differential equation in iteration continuously. In the calculation, a contour curve gradually approaches an object's edge to be detected

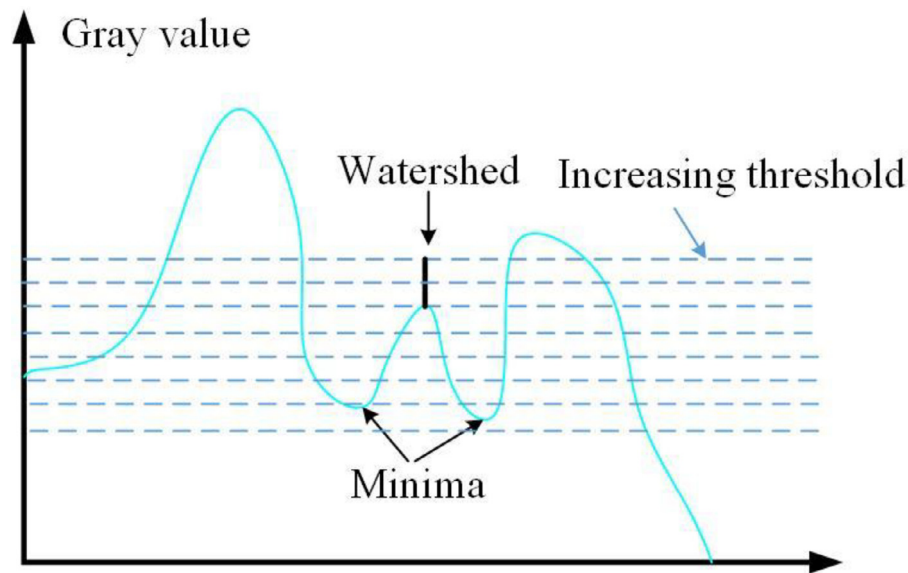


Fig. 1. Watershed algorithm schematic.

and segments a target. This model integrates the initial contour selection, target contour features, and related constraints in a single feature extraction process, which is more suitable for cellular images. The original active contour model was the Snake model proposed by Kass et al. [15]. Researchers have continued to design many active contour segmentation models that are based on the Snake model but have better performance, including the Balloon Snake [16] and edge-based geometric active contour models. To solve the limitation that the initial contour line of the Snake model must be similar to the actual contour to segment the image correctly, the Balloon Snake model adds a constant balloon force when setting an initial contour. This feature expands the curve continuously until it approaches the contour to be segmented and allows the segmentation of any shape of a region. Ali et al. [17] proposed an active contour line model based on boundary and region information; the model combines the prior knowledge of a shape according to the level set model with the watershed algorithm to prevent the overlapping of multiple objects. However, the model is easily affected by the initial position and owing to the nonconvexity of the model, it may fall into local extrema during the training process.

## 2.2. Deep learning methods for image segmentation

Deep neural networks have become the leading methods for computer vision processing. They are widely used in processing pathological images. The most considerable challenge in analyzing pathological images is analyzing individual cells because the diagnosis and classification of most diseases depend on the cells' morphological and quantitative information. Deep neural networks can detect and segment cellular images accurately, so it is of great significance for cancer diagnosis at the cellular level [18]. The essence of deep learning is to build a network model with multiple hidden layers and learn its features by training a large amount of sample data to improve the accuracy of classification or prediction, as shown in Fig. 2.

CNNs [19] constitute an early type of network that can recognize and classify images but cannot segment them. Alex et al. [20] built AlexNet using CNN with 60 million parameters and 65,000 neurons and constructed a network with five layers of convolution and three layers of full connectivity; the final output layer was a 1000-channel softmax. AlexNet is split into two parts and

can use two GPUs to train a model and use Dropout to prevent overfitting. Karen et al. [21] designed a VGG network using multiple small  $3 \times 3$  convolutional kernels instead of the large convolutional kernels of AlexNet, reducing the model's parameters while increasing the depth of a network and enhancing classification of a model. However, as the network continues to deepen, the performance of the network decreases. To solve this problem, He et al. [22] proposed ResNet with constant mapping using Residual Building Block to prevent the deterioration of a network's performance as the depth of the network increases. The fully convolutional network proposed in 2014 as the pioneer of CNNs in image segmentation tasks has laid the foundation for CNNs in image segmentation [23]. In 2016, Lin et al. [24] proposed a multiscale fusion network called feature pyramid networks (FPNs). FPNs change the connections of networks and upsample the deep layers of the networks. Each network layer obtained by upsampling is fused with the corresponding lower layer of the network for features. Each fused network layer performs prediction, effectively solving the multiscale object detection problem. Similar to the FCN, U-Net solves the segmentation problem of medical images and contains coding-decoding and hopping layer structures to effectively address the limitation caused by few medical image data [25]. Zhou et al. [26] proposed the U-Net++ model based on U-Net, redesigned the jump connection, changed the connectivity of the encoding and decoding layers, and pruned the model through deep supervision.

Deep learning methods based on CNNs are widely used for image segmentation. Naylor et al. [27] used and compared three network structures: FCN, U-Net, and Mask R-CNN, which were used to segment H&E-stained histological slice images; the segmentation problem was found to be a regression task, which predicts the distance transformation of a binarized image; finally, the segmentation results were obtained with the watershed algorithm. Chen et al. [28] proposed a deep contour-aware network with a multi-task learning framework, in which contextual features at different levels were used in the supervision of accurate gland segmentation, and regularization was added to the network training process to further improve the discriminative power of the network; the network was able to output not only accurate probability maps for gland segmentation but also depicted clear contour lines to separate different gland objects, thus further improving the performance of gland segmentation. Simon et al. [29] designed a CNN

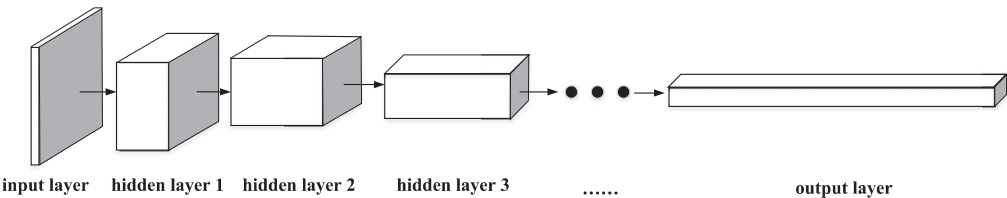


Fig. 2. Convolutional neural network concept demonstration diagram.

Table 1  
Typical methods for image segmentation.

Common image segmentation methods	Representative literature
Threshold segmentation	[4]-1979
Watershed algorithm	[5]-2011 [7]-2015 [8]-1979
Active Contour Segmentation	[10]-2018 [15]-1988 [16]-2019
Segmentation method based on deep neural network	[17]-2012 [22]-2016 [23]-2015 [24]-2016 [25]-2015 [26]-2018 [46]-2017

to segment cell nuclei; the distance of each cell nucleus boundary from the center was calculated, the computed distance was used for the accurate segmentation of overlapping cell nuclei, and the method was used for four different pathological image datasets, which achieved good segmentation results. Raza et al. [30] proposed a CNN for segmenting cells, nuclei, and glands in pathological images; the network was trained by scaling an input image to different scales for image enhancement and connecting intermediate layers with contextual information to obtain accurate location information. Yi et al. [31] combined a fully convolutional network with a marker-controlled watershed segmentation algorithm to use a fully convolutional network for predicting red cells, use predicted results as markers, and use the marker-controlled watershed to segment the cells further; these approaches improved red cell separation.

A comparison of various image segmentation algorithms is shown in Table 1. In summary, CNN-based deep learning methods exhibit excellent performance compared with traditional image segmentation methods and have facilitated medical research and medical diagnosis. In recent years, the CNNs' advantages have been used in cell image segmentation. Many deep learning methods with higher segmentation performance have been proposed and applied cell image segmentation to provide a reliable basis for subsequent medical diagnosis.

3. Cell light field image instance segmentation

The following section will focus on three aspects: cell image weak label extraction, cell image instance segmentation, and cell internal structure segmentation to analyze and summarize deep learning-based cell image segmentation methods shown in Fig. 3.

3.1. Cell image weak label extraction method

The expensive and time-consuming manual annotation has become an essential bottleneck in the research and application of

high-quality machine learning models. Employing physicians (experts) to annotate datasets in medical image research is difficult. In machine learning, strongly supervised learning uses high-confidence labeled data (e.g., manual labeling) for model training, which requires a high cost. Weakly supervised learning uses low-confidence labeled data (weak labeling) for model training and covers a wide range. Weakly supervised learning is considered when labeled information is incomplete, inexact, or imprecise. Supervised learning of models by the weak label in instance segmentation of cells is a critical way to reduce costs.

The commonly used annotation methods for weakly supervised learning in semantic segmentation are Scribes [32], Box [33], and Point [34] (Fig. 4). These annotation methods only roughly identify a target object's location in an image without clear edge information. Training methods based on weakly supervised segmentation models can be divided into two types: literature [32,33], in which an iterative optimization-like approach is used to obtain an initial segmentation model on the basis of original annotations (Scribes [32] or Box [33]) and then use the results obtained from the segmentation of the initial segmentation model as input for the second iteration, and literature [35–39], in which target location and edge-related cues are obtained according to annotation information and pixel-level weak labels are computed and used in training a segmentation model.

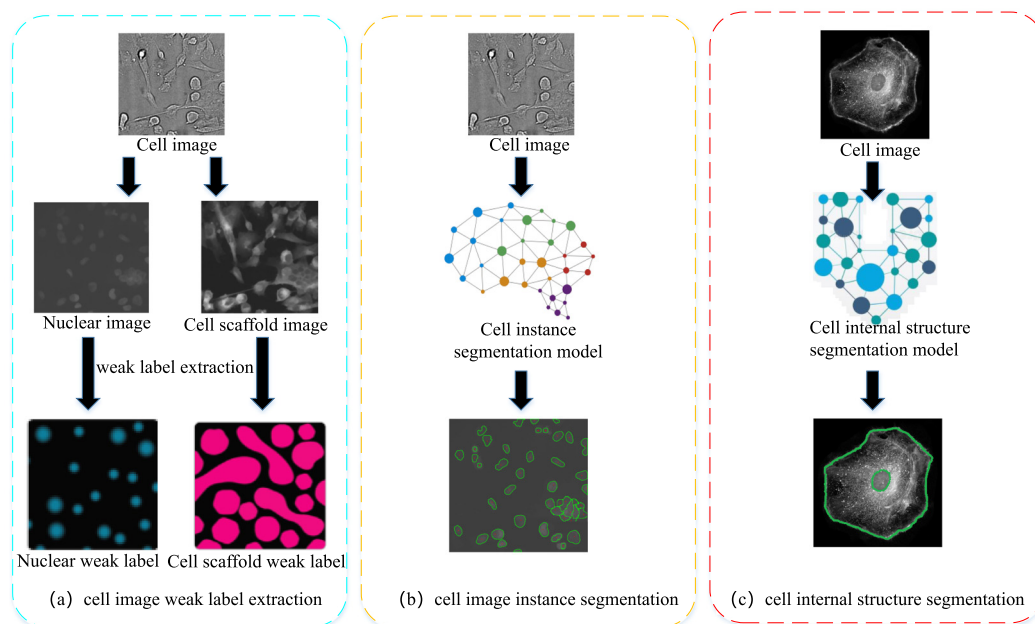
The quality of generating pixel-level weak labels affects the performance of a segmentation model. Wei et al. [37] used a trained classification model to classify images and obtain regions associated with underlying annotation information; they were able to achieve the automatic generation of pixel-level annotations of images. Huang et al. used a dynamic supervision method and obtained pixel-level annotation data by generating a seed region (the initial seed is generally the position information of a segmented object) in each iteration process. Xu Shukui et al. used GrabCut [38] algorithm to train the parameters of Gaussian mixture model by Box annotation to distinguish foreground and background and obtain pixel-level annotation data. Belharbi S et al. [39] proposed an Active Learning framework that progressively generates pixel-level annotations during the training process. Shi et al. [40] proposed a gated graph convolutional network to refine the prediction of weak and coarse semantics and generate clear boundaries and fine-grained pixel-level classification.

How to segment stacked cells in cell images is an essential factor affecting the quality of weak labeling of cell images. Given that instance segmentation is performed for cell images, individual and pixel-level annotations for each cell need to be obtained. Therefore, cell images should be segmented at different scales (e.g., stacked cell regions vs. nonstacked cell regions), and existing cell structure (cytoplasm and nucleus) images should be fused for the generation of individual and pixel-level annotations for each cell.

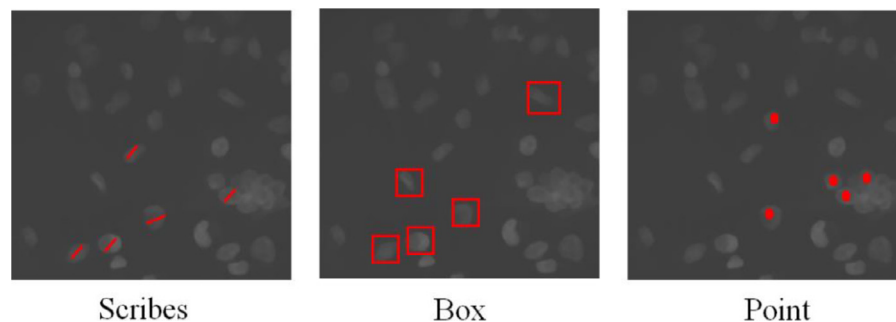
3.2. Instance Segmentation Methods for Cell Images

Cell instance segmentation is the basis of cell tracking and cell division detection and occupies an important position in medical image processing and analysis. Unlike cell segmentation, cell instance segmentation not only requires the identification of the





**Fig. 3.** Cell image instance segmentation. (a) Cell image weak label extraction: The more readily available weak label information of an image (e.g., weak label information of fluorescence-stained images) is extracted at the pixel level for weak label extraction. Weak label extraction can reduce work load for subsequent deep learning. (b) Cell image instance segmentation: instance segmentation of cell images on the basis of convolutional neural networks (e.g., U-Net) or instance segmentation framework based on target recognition algorithms. (c) Internal cell structure segmentation: comprehensive joint learning of cells (cytoplasmic contours) and nuclei for the construction of a joint model of cell instance segmentation and structure segmentation.



**Fig. 4.** Schematic diagram of Scribes, Box and Point labeling in the nuclear staining image.

class of each pixel point in an image but also requires the absence of overlaps between different cells.

Ronneberger et al. [25] proposed U-Net. First, U-Net's U-shaped network can effectively combine information from the deep and low-level networks to improve the segmentation performance of the networks. Second, U-Net uses an overlap-tile strategy for the seamless segmentation of input images with various sizes. Finally, U-Net uses elastic deformation to enhance images. The elastic deformation method is well-suited for cellular images. Overall, U-Net can be used to solve biomedical image segmentation problems with small data sets and achieved excellent performance in the segmentation of biomedical images. Li et al. [41] combined neural ordinary differential equations (NODEs) with a U-Net for blood smear image segmentation. The NODEs-UNet network was constructed by adding the ODE-block module to the convolutional layer of the U-Net network, determining the optimal position of the ODE-block in the network through experiments, and select the appropriate error tolerance. The experimental results showed that the network model improved the accuracy of blood cell image segmentation and reduces the computational cost of the network. Long [42] designed improved coded branching and densely connected convolutional blocks to improve the segmentation

performance of U-Net (U-Net+). The network performed accurate segmentation with few data. Huang et al. [43] proposed an improved U-Net algorithm based on the mixed convolution block for cell image segmentation. First, different resolution modes were obtained by using convolution kernels of various sizes, two hyperparameters (width multiplier and resolution multiplier) were used in adjusting a model, and a residual path was used to improve training efficiency. Zamora-Cárdenas et al. [44] proposed a cell instance segmentation method based on the U-Net architecture, using a depth distance transformer (DDT) to enhance a model's learning of morphological information in images. The output of DDT was fed to the U-Net model to improve the performance of the model in cell instance segmentation.

Instance segmentation detects targets in an image and marks different target instances one by one and the pixels in the corresponding regions of different instances [45]. In simple terms, cell instance segmentation classifies each pixel on a cell image and distinguishes among different objects of the same class. Current mainstream methods for instance segmentation are divided into two types. The first type is based on candidate regions, where target candidate frames are obtained through target detection, and target pixel point classification is performed according to the

candidate boxes. The second type is based on segmentation and directly enables a model to learn a specific design or target boundary.

He et al. [46] proposed the Mask R-CNN instance segmentation framework, which implements the classical “detection before segmentation” two-stage instance segmentation framework, and adds semantic segmentation branches to existing classification and border regression branches in the Faster-RCNN framework. Yi et al. [47,48] proposed a bounding box method based on the two-stage instance segmentation framework; the method produced a key-point map and its subsequences and quickly generated candidate boxes based on keypoints as cell detectors. This method is suitable for cell border regression at different scales in cell images. In subsequent studies, target features were used as guides for the separation of target objects from neighboring objects in candidate boxes. An auxiliary feature refinement module was proposed to further refine the boundary features of a network and improve segmentation quality. Zhou et al. [49] constructed an instance-relationship network model based on Mask R-CNN and used it to store and exchange the feature information of each cell instance with a cell association matrix; this information was used in segmenting stacked regions in cell images. Kromp et al. [50] compared and evaluated the segmentation performance of four deep learning frameworks (U-Net, U-Net with ResNet34 backbone network, Deepcell, and Mask-RCNN) and trained and tested them on the fluorescent nuclear images of different samples, sample preparation types, annotation quality, image scale, and segmentation complexity; the final results showed that three deep learning frameworks (U-Net, U-Net with ResNet34 backbone network, and Mask-RCNN) can segment fluorescent cell nuclear images on most sample preparation types and tissue sources. Prangemeier et al. [51] applied a target recognition algorithm, detection transformer (DETR), to CELL instance segmentation and proposed an attention-based cell detection transformer (Cell-DETR) to achieve direct end-to-end cell instance segmentation. DETR directly predicts the final detection result by combining the common CNN and transformer architecture. CELL-DETR applies the basic idea of DETR to cell segmentation and successfully solves the bottleneck of automatic cell instance segmentation of yeast cells in a microstructure environment. Up-schulte et al. [52] proposed contour proposal network (CPN) and used a Fourier descriptor to establish a CPN model. Meanwhile, the most advanced object detection architecture can be used as a backbone network to select an instance segmentation model.

In object detection, the single-shot multibox detector (SSD) [53] is an essential representative of single-stage object detection. Yi et al. [54] used the SSD framework to add a target detection particular layer (ROIs Crop) in order to access information in the network model input layer, ResNet101 intermediate correlation layer, and border detection layer; the output information was used in the prediction of cell pixel points in target regions in cell images. However, when classifying and recognizing cell images, the number of classifications is often small, or even only the two categories of foreground and background, so the use of ResNet101 as the backbone network is relatively bloated. Xie et al. [55] proposed the PolarMask polar contour modeling method to determine the contour of an object for modeling. This method is a single-stage instance segmentation network framework, simplifies the instance segmentation process in the past two stages (detection before segmentation), and transforms an instance segmentation problem into instance center point classification and dense distance regression problems. However, instance segmentation methods based on candidate regions usually detect a target by using a pretrained cell image classification model. These classification models are basically large models, and most of the labeled cell images on the Internet are small datasets, which are difficult to train these models effectively.

In summary, the existing instance segmentation methods based on candidate regions require a large amount of identification data to be provided to the model for training. Thus, they incur considerable cost, and two problems have been identified: (1) if the target detection border does not frame the entire cell, mask segmentation yields only a portion of a cell within the detection border; (2) mask segmentation uses the average value of the cross entropy of all pixels in the target detection frame as the loss function to perform background and foreground segmentation for each pixel. If the pixel-level labeling of confidence is low, it causes the cells to be divided into multiple parts, incomplete cell segments, and jagged outlines. Given that cells have certain geometric shapes (most cells in cell images are circular, oval, and kidney shaped), instance segmentation that directly enables a model to learn the boundaries of specific target objects is a feasible solution. Therefore, the cell instance segmentation model needs to build a method for modeling cell contours, and contour integrity and smooth constraints need to be designed to enable the model to learn the geometric features of cells.

### 3.3. Methods of segmenting the internal structures of cells

The segmentation of key structures, such as tumors, glands, and cells, is an important step in automated microscopic image analysis systems. Qaiser et al. [56] extracted features by using topological features and convolutional networks for stained tissue images and then used random forest classification to locate tumor regions. Awan R et al. [57] used the morphology of glandular cells to establish two rating classification models: binary classification and multiclassification; the shape measurement mechanism proposed by the author and the convolutional network cell edge detection were used in extracting features, and then an SVM classifier was used to classify features. Raza et al. [58] proposed a network (Micro-Net) and used different training data to train multiple segmentation models based on the network (Micro-Net) for different segmentation objects (cells, nuclei, and glands); each segmentation model segmented the corresponding target. Vicar et al. [59] reviewed the method of cell image segmentation without labeling. In the nuclear image segmentation in fluorescent stained images, the performance of the relevant cell nucleus detection algorithm based on the Laplace Gaussian filter was improved. The performance related algorithms of watershed segmentation based on marker control enhanced. Thi Le P et al. [60] proposed an automatic cell nucleus segmentation method based on the convolutional blur attention network, which can automatically generate pixel-level labels from the features of input images. The network performs upsampling and downsampling. In the downsampling process, the blur attention module and blur pooling operation are used to improve segmentation accuracy. In the upsampling process, the pyramid blur pooling module obtains multiscale information. Khamael et al. [61] used edge-based geometric active contours combined with level set forces (curvature-based, normal direction, and vector field-based) to extract the complex detailed features of cell images; the method effectively segmented leukocytes and their nuclei. These segmentation algorithms for cell content structures (e.g., nucleus and cytoplasm) are basically performed for stained images.

In research related to the segmentation of internal cell structures in brightfield images, given that fluorescent staining can selectively stain some components inside cells with dyes, Christiansen et al. [62] used a deep learning model to predict the location of fluorescent staining in cell brightfield images. In their study, a deep learning model (ISL) was proposed to learn stained images directly as label samples of brightfield images, and ISL predicted the stained images of the brightfield images. Dai et al. [63] constructed a recognition model to identify endothelial cells from

induced pluripotent stem cell images through deep learning and used traditional immunofluorescence technology to evaluate recognition results. The recognition model exhibited high classification accuracy.

In summary, in the current segmentation research of cell images based on fluorescence microscopy, most algorithms are based on the respective research algorithms of each channel image, and the information of multiple channels is not fused. The segmentation of internal structure of cells is difficult in brightfield images, but it will be an effective and feasible way to guide the segmentation of brightfield images through fluorescence-stained images. Therefore, the fusion of multichannel cell image information and joint learning of individual cell segmentation and internal structure segmentation can further improve the segmentation performance of a model.

#### 4. Discussion

In the research of cell image segmentation, low contrast and blurred edges have always caused problems affecting cell segmentation performance. How to effectively segment overlapping and adherent cells has been an important research direction in the field of cell segmentation nowadays. Current trends in the field of image segmentation can be seen from the cell instance segmentation algorithms in recent years, which mostly use deep learning algorithms such as CNNs, such as U-Net to segment cells or improve them to improve the segmentation performance. These deep learning algorithms can effectively segment cell images by instances. However, a large number of annotated datasets are required to provide models for training, and this requirement needs a lot of manpower and demands substantial material resources. How to use a small amount of manually annotated data or weak labels in improving training effect (weakly supervised learning) is an important research direction in cell segmentation, so how to extract high-quality weak labels is an issue that cannot be ignored in future studies of cell segmentation.

Current cell instance segmentation studies are mainly performed on stained cell images, which are easy to manually annotate. However, stained images do not clearly represent cell edges. **Instance segmentation models only predict possible stained pixel points rather than segmented edges without learning information about cell morphological features. Among segmentation methods, region-based methods have some problems. Specifically, a target detection border cannot frame an entire cell, cell segmentation is incomplete, and segmented contours are jagged.** Cells have distinct geometry, and thus an instance segmentation approach that directly makes a model learn specific target object boundaries is a feasible solution. **Therefore, methods for modeling cell contours and contour-integrity-smoothing constraints are needed to enable models to learn the geometric morphological features of cells.**

In the segmentation of intracellular structures, different segmentation models are used for different structures within cells, and interconnections among different structures within the cells are ignored. Hence, multiple instances of segmentation model training are conducted, thus requiring a high computational load. Most algorithms study algorithms for segmenting each channel image, without fusing information of multiple channels. Therefore, fusing the cell image information of multiple channels and jointly learning the segmentation of individual cells with the segmentation of internal structures can further improve the segmentation performance of a model.

#### 5. Conclusion

Cellular segmentation is an essential basis for disease diagnosis and classification, and the detection of many diseases de-

pends on the number and morphology of cells. Cell segmentation is a cumbersome task that consumes considerable resources in terms of manpower and material resources and requires experienced pathologists to perform assessments. Early cell segmentation tasks used traditional segmentation methods, such as threshold, region growth, watershed, and active contour line, which are computationally less expensive but have limited segmentation effects. Rapid developments in computer performance in recent years have fostered developments in CNN research. CNNs are effective in various tasks in computer vision processing, where learning is achieved by simulating neurons in the human brain. Compared with the traditional segmentation method, a segmentation method based on deep CNN greatly improves segmentation accuracy but imposes high requirements on computer hardware and is computationally expensive.

In this paper, traditional segmentation and deep learning methods for image segmentation are systematically reviewed. Considering cell image segmentation algorithms proposed in recent years, we found that the current trend in the field of cell segmentation is to combine deep learning algorithms with traditional segmentation methods or optimize CNNs by changing the connectivity of neural networks, adding attention mechanisms to the network, and combining multiple network models. These approaches allow the identification of a robust general algorithm with enhanced segmentation performance. Although the deep learning algorithm has achieved ideal results for cell instance segmentation, the success of deep learning depends on the support of a large amount of data. Hence, research of weakly supervised or semisupervised learning for cell segmentation becomes critical.

#### Declaration of Competing Interest

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

#### Acknowledgments

This research was funded by [Natural Science Foundation of Fujian Province](#), with grant numbers [2020J01086](#) and [2022J01309](#).

#### References

- [1] J Cao, X Lin, S Guo, et al., Bipartite graph embedding via mutual information maximization[C], in: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 635–643.
- [2] F Li, X Zhou, J Ma, et al., Multiple Nuclei Tracking Using Integer Programming for Quantitative Cancer Cell Cycle Analysis[J], *IEEE Trans. Med. Imaging* 29 (1) (2009) 96–105.
- [3] Xiaowei Han, Junsheng Li, Yanping Li, Xinhe Xu, An Approach of Color Object Searching for Vision System of Soccer Robot, in: *2004 IEEE International Conference on Robotics and Biomimetics*, 2004, pp. 535–539, doi:[10.1109/ROBIO.2004.1521836](#).
- [4] N Otsu, Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems Man and Cybernetics*[J], *IEEE Trans. Syst. Man. Cybern.* (1979) 9.
- [5] S G Wei, Y Lei, C Zhen, et al., Motion Detection Based on Optical Flow and Self-adaptive Threshold Segmentation[J], *Procedia Engineering* 15 (2011) 3471–3476.
- [6] F Y Shih, S Cheng, Automatic seeded region growing for color image segmentation[M], *Butterworth-Heinemann*, 2005.
- [7] A Seal, A Das, P. Sen, Watershed: an image segmentation approach[J], *Int. J. Comp. Sci. I.T.* 6 (3) (2015) 2295–2297.
- [8] S. Beucher, Use of watersheds in contour detection[C], in: *Proceedings of the International Workshop on Image Processing, CCETT*, 1979.



- [9] X Zhang, W Ye, An adaptive fourth-order partial differential equation for image denoising[J], *Comput. Math. Appl.* 74 (10) (2017) 2529–2545; 12. Mitko V, Van DPJ, Robert K, André H, Viergever MA, Pluim JPW, Konradin MJPO. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. 2013, 8(7): e70221..
- [10] V Mitko, DPJ Van, K Robert, H André, MA Viergever, JPW Pluim, MJPO Konradin, Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images 8 (7) (2013) e70221.
- [11] Y Chen, Y He, J Wang, et al., Automated Cone Photoreceptor Cell Segmentation and Identification in Adaptive Optics Scanning Laser Ophthalmoscope Images Using Morphological Processing and Watershed Algorithm[J], *IEEE Access* (99) (2020) 1 PP-1.
- [12] FV Nurçin, E. Imanov, Selective Hole Filling of Red Blood Cells for Improved Marker-Controlled Watershed Segmentation, *Scanning* 2021 (2021) 5678117 Dec 6PMID: 34950282; PMCID: PMC8668327, doi:10.1155/2021/5678117.
- [13] C Jung, C Kim, Segmenting Clustered Nuclei Using H-minima Transform-Based Marker Extraction and Contour Parameterization[J], *IEEE Trans. Biomed. Eng.* 57 (10) (2010) 2600–2604.
- [14] CF Koyuncu, E Akhan, T Ersahin, R Cetin-Atalay, C Gunduz-Demir, Iterative h-minima-based marker-controlled watershed for cell nucleus segmentation, *Cytometry A* 89 (4) (2018) 338–349.
- [15] M Kass, A Witkin, D. Terzopoulos, Snakes: Active contour models[J], *Int. J. Comput. Vision* 1 (4) (1988) 321–331.
- [16] H J Chien, C J Chang, Application of the Balloon Snake in the Volume Measurement of Subretinal Fluid in Central Serous Chorioretinopathy[C], *Seminars in Ophthalmology*, Taylor & Francis 34 (6) (2019) 403–408.
- [17] Ali S, Madabhushi AJIToMI. An Integrated Region-, Boundary-, Shape-Based Active Contour for Multiple Object Overlap Resolution in Histological Imagery. 2012.
- [18] Z Huang, Z Qi, S Jiang, et al., Medical Image Segmentation Based on the Watersheds and Regions Merging[C], in: 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), IEEE, 2016.
- [19] A S Eltrass, M B Tayel, Ammar A I. A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary Gabor transform[J], *Biomed. Signal Process. Control* 65 (2021) 102326.
- [20] A Krizhevsky, I Sutskever, G E Hinton, Imagenet classification with deep convolutional neural networks[J], in: *Advances in neural information processing systems*, 2012, p. 25.
- [21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [22] K He, X Zhang, S Ren, et al., Deep residual learning for image recognition[C], in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [23] J Long, E Shelhamer, T Darrell, Fully Convolutional Networks for Semantic Segmentation[J], *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (4) (2015) 640–651.
- [24] Lin T Y, P Dollár, Girshick R, et al. Feature Pyramid Networks for Object Detection[J]. *arXiv e-prints*, 2016.
- [25] Ronneberger O, Fischer P, Brox TJS, Cham. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
- [26] Z Zhou, MM Rahman Siddiquee, N Tajbakhsh, et al., in: Unet++: A nested u-net architecture for medical image segmentation[M]//Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, Cham, 2018, pp. 3–11.
- [27] P Naylor, M Laé, F Reyat, L. Walter, Nuclei segmentation in histopathology images using deep neural networks, *IEEE Int. Symp. Biomed. Imag.* (2017) 2017.
- [28] H Chen, X Qi, L Yu, et al., DCAN: Deep contour-aware networks for object instance segmentation from histology images[J], *Med. Image Anal.* 36 (2017) 135–146.
- [29] Graham S, Vu QD, Raza S, Azam A, Tsang YW, Kwak JT, Rajpoot N. HoVer-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images. 2018.
- [30] S E A Raza, L Cheung, M Shaban, et al., Micro-Net: A unified model for segmentation of various objects in microscopy images[J], *Med. Image Anal.* 52 (2019) 160–173.
- [31] F Yi, I Moon, B Javidi, Automated red blood cells extraction from holographic images using fully convolutional neural networks[J], *Biomed. Opt. Express* (2017).
- [32] D Lin, J Dai, J Jia, K He, Sun J. ScribbleSup, Scribble-Supervised Convolutional Networks for Semantic Segmentation, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, 2016, pp. 3159–3167.
- [33] Khoreva A BR, Hosang J, et al., Simple Does It: Weakly Supervised Instance and Semantic Segmentation. *IEEE Computer Society* 2016.
- [34] O Russakovsky, AL Bearman, V Ferrari, F-F Li, What's the point: Semantic segmentation with point supervision, *ECCV*, 2016.
- [35] A Kolesnikov, Lampert CH. Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation, *European Conference on Computer Vision*, 2016.
- [36] SJ Oh, R Benenson, A Khoreva, Z Akata, M Fritz, B. Schiele, Exploiting saliency for object segmentation from image level labels, *IEEE Conference on Computer Vision*, 2017.
- [37] Wei Y, Feng J, Liang X, Cheng MM, Zhao Y, Yan S. Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. 2017.
- [38] C Rother, GrabCut: Interactive foreground extraction using iterated graph cut[J], *ACM Trans. Graph.* (2004) 23.
- [39] Belharbi S, Ayed I B, McCaffrey L, et al. Deep Active Learning for Joint Classification & Segmentation with Weak Annotator[C] 2020.
- [40] Y Shi, Q Li, X X Zhu, Building segmentation through a gated graph convolutional neural network with deep structured feature embedding[J], *ISPRS J. Photogramm. Remote Sens.* 159 (2020) 184–197.
- [41] D Li, P Tang, R Zhang, C Sun, Y Li, J Qian, Y Liang, J Yang, L Zhang, Robust Blood Cell Image Segmentation Method Based on Neural Ordinary Differential Equations, *Comput. Math. Methods Med.* (2021) 5590180 PMID: 34413897, doi:10.1155/2021/5590180.
- [42] F Long, Microscopy cell nuclei segmentation with enhanced U-Net[J], *BMC Bioinf.* (2020) 21.
- [43] C Huang, H Ding, C Liu, in: Segmentation of Cell Images Based on Improved Deep Learning Approach[J], *IEEE Access*, 2020, p. 1. PP-1.
- [44] W Zamora-Cárdenas, Mendez M, Calderon-Ramirez S, et al. Enforcing Morphological Information in Fully Convolutional Networks to Improve Cell Instance Segmentation in Fluorescence Microscopy Images[C] *International Work-Conference on Artificial Neural*.
- [45] B Romera-Paredes, PHS Torr, Recurrent Instance Segmentation, *ECCV 2016 14th European Conference on Computer Vision*, 2016.
- [46] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV) 2017.
- [47] J Yi, P Wu, Q Huang, et al., in: Multi-scale cell instance segmentation with keypoint graph based bounding boxes[C]// International conference on medical image computing and computer-assisted intervention, Springer, Cham, 2019, pp. 369–377.
- [48] J Yi, P Wu, H Tang, et al., Object-Guided Instance Segmentation With Auxiliary Feature Refinement for Biological Images[J], *IEEE Trans. Med. Imaging* 40 (9) (2021) 2403–2414.
- [49] Y Zhou, H Chen, J Xu, Q Dou, Heng P-A. IRNet, Instance Relation Network for Overlapping Cervical Cell Segmentation (2019) 640–648.
- [50] F Kromp, L Fischer, E Bozsaky, I Ambros, W Doerr, S Taschner-Mandl, P Ambros, A Hanbury, Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation, *Computer Vision and Pattern Recognition*, 2019.
- [51] T Prangemeier, C Reich, H Koepl, Attention-Based Transformers for Instance Segmentation of Cells in Microstructures[C]//, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2020.
- [52] Upschulte E, Harmeling S, Amunts K, et al. Contour Proposal Networks for Biomedical Instance Segmentation[J]. 2021.
- [53] W AD Liu, D Erhan, et al., SSD: Single Shot MultiBox Detector[J]. 2015. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016 ECCV 2016 Lecture Notes in Computer Science*, vol 9905, Springer, Cham, 2015.
- [54] J Yi, P Wu, M Jiang, DJ Hoepfner, DN. Metaxas, Instance Segmentation of Neural Cells, *European Conference on Computer Vision Workshops*, 2018.
- [55] Xie E, Sun P, Song X, Wang W, Liang D, Shen C, Luo P. PolarMask: Single Shot Instance Segmentation with Polar Representation, 2019.
- [56] Kaiser T, Tsang Y, Epstein D, Rajpoot N. Tumor Segmentation in Whole Slide Images Using Persistent Homology and Deep Convolutional Features, 2017.
- [57] R Awan, K Sirinukunwattana, D Epstein, S Jefferyes, U Qidwai, Z Aftab, I Mujeed, D Snead, N. Rajpoot, Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images, *Sci. Rep.* 7 (1) (2017) 16852.
- [58] SEA Raza, L Cheung, M Shaban, S Graham, D Epstein, S Pelengaris, M Khan, NM. Rajpoot, Micro-Net: A unified model for segmentation of various objects in microscopy images, *Med. Image Anal.* (2018).
- [59] T Vicar, J Gumulec, J Balvan, M Hrach, R. Kolar, in: Label-Free Nuclear Staining Reconstruction in Quantitative Phase Images Using Deep Learning. 2019, Springer Singapore, Singapore, 2019, pp. 239–242.
- [60] P Thi Le, T Pham, YC Hsu, JC Wang, Convolutional Blur Attention Network for Cell Nuclei Segmentation, *Sensors (Basel)* 22 (4) (2022) 1586 Feb 18PMID: 35214488; PMCID: PMC8878074, doi:10.3390/s22041586.
- [61] A L D Khamael, I Tomeo-Reyes, J Banks, et al., Evaluation and benchmarking of level set-based three forces via geometric active contours for segmentation of white blood cell nuclei shape[J], *Comput. Biol. Med.* 116 (2020) 103568.
- [62] Christiansen EM, SJ Yang, DM Ando, A Javaherian, G Skibinski, S Lipnick, E Mount, A O'Neil, K Shah, AK Lee, In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images, *Cell* (2018) S0092867418303647.
- [63] K Dai, M Lachmann, T Kunihiro, S Yuasa, K Fukuda, Automated Deep Learning-Based System to Identify Endothelial Cells Derived from Induced Pluripotent Stem Cells, *Stem. Cell Rep.* 10 (6) (2018).