

Supplementary Information

Dynamic image recognition for cell encapsulated microfluidic droplet

Yuanhang Mao ^{1*}, Xiao Zhou ^{1*}, Weiguo Hu ¹, Weiyang Yang ¹, and Zhen Cheng ^{1†}

¹ Department of Automation, Tsinghua University, Beijing, 100084, China

* These authors contributed equally: Yuanhang Mao, Xiao Zhou.

† Corresponding authors: Department of Automation, Tsinghua University, Haidian District, Beijing, 100084, China. Phone: +86-10-62799220,

email: zhouxiao@pjlabor.org.cn; zcheng@mails.tsinghua.edu.cn

Table of contents

Supplementary Note 1. Details for application software of weakly supervised cell-counting network (WSCApp) with graphical user interface (GUI).

Supplementary Note 2. Illustration of two segmentation algorithms for dark-field and bright-field scenarios.

Supplementary Note 3. Details for the design and implementation of CNN-based classifiers and WSCNet models.

Figure S0. A graphical user interface (GUI) preview of the proposed WSCApp software.

Figure S1. The encapsulated probability restricted by the Poisson's distribution.

Figure S2. Representative images of bright-field and dark-field droplets under different illumination conditions.

Figure S3. Comparison of two segmentation algorithms on bright-field and dark-field droplets.

Figure S4. Illustration for evaluation metric calculation of segmented droplets under different IoU thresholds.

Figure S5. Distribution diagram of recognition rate and average IoU value of different datasets.

Figure S6. Representative results of droplet classification, cell counting and location predicted by ResNet18 classifier and WSCNet model.

Figure S7. Representative results of detected droplets with encapsulated mammalian cells by CEHT algorithm and LeNet-5 classifier.

Figure S8. Representative results of classification failure by WSCNet model.

Figure S9. Multiclass confusion matrixes showing classification performance on different datasets.

Figure S10. Comparative result of cell amount distribution encapsulated in microfluidic droplets.

Table S1. Summary of different approaches in cell encapsulated microfluidic droplets.

Table S2. Annotation summary and inference time of cell encapsulated droplets on different datasets.

Table S3. Detailed parameters of droplet diameter distribution of three independent experiments.

Table S4. Comparative performance of different algorithms for the segmentation of dark-field droplets.

Supplementary Note 1. Details for application software of weakly supervised cell-counting network (WSCApp) with graphical user interface (GUI)

WSCApp is a practical and effective tool for dynamic video and image recognition for cell-encapsulated microfluidic droplets, including the droplet information of the generation process (e.g., diameter, and distribution), and the cell information of encapsulated droplets (e.g., cell amounts and position). In addition, WSCApp can not only distinguish droplet encapsulations but also locate each cell without any supervised location information. Therefore, empty, multicell, and single-cell encapsulated droplets could be recognized and quantitatively counted. Please perform the following steps to install and run WSCApp.exe. (Currently, this software only supports the window system).

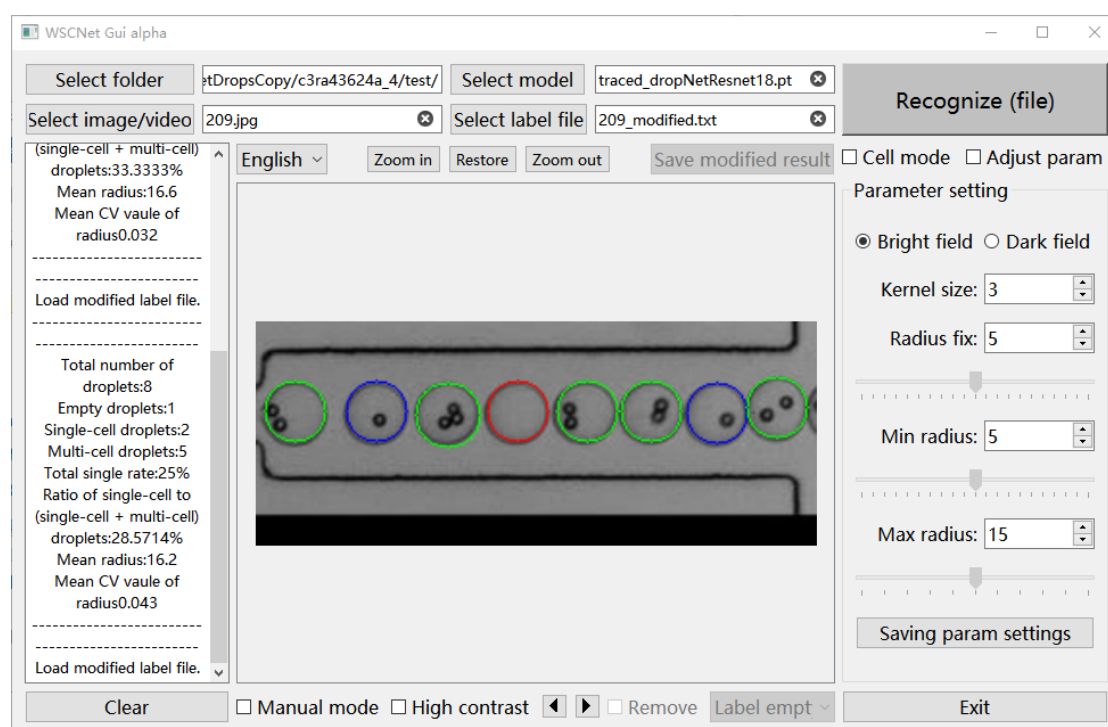


Figure S0. A graphical user interface (GUI) preview of the proposed WSCApp software.

Step 1. File preparation

Before configuring the WSCApp.exe operating environment, some packaged files should be unpacked to their respective paths. In detail, WSCApp_x86_64.zip should be downloaded and unzipped, the same as examples_dataset.zip if you need to test the software on our sample dataset.

Step 2. Training environment configuration

If you use the model trained in the paper, you can skip this step.

If you want to achieve great performance on your own datasets, a perfect development environment is essential. Although the software provides packaged recognition functions, the training process is highly customizable, considering the debuggability and ease of use of the training process, we recommend that users install anaconda to set up a model training environment.

Anaconda environment

1) Install Anaconda

<https://www.anaconda.com/>

2) Create conda environment, then activate it

```
conda create -n wscnetTrain python=3.8
```

```
conda activate wscnetTrain
```

3) Install Pytorch

<https://pytorch.org/>

```
pip3 install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118
```

4) Install opencv-python

```
pip install opencv-python
```

Step 3. An overview of the WSCApp

As shown in Figure S0, there are 4 primary components in WSCApp software (Multi language support), encompassing *Inputs*, *Recognition*, *Modification*, *Window and Details*:

Inputs consists of modifiable and default file paths and file names used for video/figure recognition.

Recognition lists all the modifiable parameters of the dark-field as well as the bright-field recognition algorithms, and automatically process all files or individual files for image recognition.

Modification support semi-automatic or manual modification of the image label of droplet proposals and droplet categories. To avoid tedious and manual cell-level annotation, only three droplet-level labels, including empty, single-cell, and multicell encapsulation are adopted.

Window demonstrates the detailed recognized figure that before and after the modification. Users can also annotate and modify the recognized images in this window.

Details provides some helpful prompt messages through all the periods of WSCApp software, including the recognition period, the modification period, and the evaluation period.

3.1 Inputs

Inputs could set the workspace file path, select the single figure/video or a folder that contains multiple videos or figures, and select their corresponding annotation files.

Work space, select the work path, and specify the storage path for the input object and output files;

Select image/video, specify an image file or a single video to be processed by this program. If the path is empty, the program will automatically process all files that meet the requirements in the *work space*.

Select label, specify an annotation file for the selected image. When image has been chosen, the program will automatically search for an annotation file or recognition result file sharing a similar name with the *selected image*.

When running in recognition mode, WSCApp software can perform folder recognition or single image/video recognition by selecting a folder (without specifying a single internal file) or selecting a single image/video, respectively.

When running in modification mode, WSCApp software can annotate and modify recognized images by selecting the figure and the corresponding annotation file (can be loaded automatically or specified manually).

User can click the clear button in the right corner to clear chosen image/video/label file.

3.2 Recognition

Recognition can automatically load recognition configuration files that occur after selecting a folder. After a single click, automatically process all files or individual files for image recognition according to the input settings of the file/folder path, and display the recognition results in the window position.

There are several parameters that can be adjusted according to the application scenarios, including the following parameters:

Field type Boolean button, including *Bright field* and *Dark field*, this button is set for different segmentation algorithms that are developed for dark or light backgrounds. If the solution background inside the droplet is a dark/light background (low/high grayscale value), it is suitable for calling dark/light background recognition algorithms. Two different sets of feature engineering algorithms are designed and integrated into the software.

Kernel size, this value relates to the size of the convolutional kernel, and can be adjusted based on the image resolution.

Minimum radius, set the minimum droplet radius in pixels.

Maximum radius, set the maximum droplet radius in pixels.

Foreground ratio, set the ratio of the foreground area to the background area in the range 0 to 1.

You can save all these parameters to param.json in the current folder, so that the next time you open this folder, the software will automatically load them.

The program performs real-time video/image processing by calling CNN-based classifiers or WSCNet model through multiple back-end processes. After images being automatic recognized, a new 'imgResult' folder will be created in the work path to store the recognized images in .png formats, and a new 'textResult' folder will be created to store the recognized droplet information in .txt format. For videos, this software will firstly divide it into frames and store them into a new folder created by the video's name, then apply the previous steps for images, finally regenerate the recognized video result.

3.3 Modification

Modification can display the processing results based on automatic recognition or loading the modified files, mark and modify the identified key parameters, including droplet size, position, and categories of cells with different contents (empty, single-cell, multi-cell droplet).

Users can click on "Editing mode" to enter semi-automatic modification mode and modify the following parameters:

Radius correction, uniformly adjusting the size of the identified droplet proposals;

Minimum radius, set a threshold to filter out droplets smaller than this radius;

Maximum radius, set a threshold to filter out droplets larger than this radius;

Add new circle, left click on droplet image and drag to all a new circle;

Drag circle, right click on droplet proposal, and drag to move the droplet to the specified position.

In the modification mode, users can click on "Manual editing" to enter manual modification mode and modify the following parameters:

Delete the bounding box, right-click on droplet proposal to remove the label with the negative mark, right-click and drag to generated new droplet proposals at the specified location.

Label the empty/single/multi droplets, select the corresponding category from the dropdown box, then set it as the corresponding category by left-clicking on the target droplet proposal to complete the modification of the droplet category.

To avoid tedious and manual cell-level annotation, only three droplet-level labels, including empty, single-cell, and multicell encapsulation are adopted. After the modification of recognized images, user can left-click the “save the results” button to save the modified results with a new “_modified.text” file adding to the ‘textResult’ folder.

3.4 Details

Details prints out all the necessary prompt information (Figure SX) during the whole cycle of video/image recognition. During the parameter input period, the input value to *Inputs*, *Recognition*, and *Modification* parameters will be displayed in *Details* if Enter key is pressed. When recognition push-button is pressed, all the pre-set parameters are printed to Input portion in *Details*. During or after the video/figure recognition period, some useful information will be printed out to Output portion in *Details*, such as the total number of the empty, multicell, and single-cell encapsulated droplet. During the modification period, the statical information of the modified image will be printed out to the Output portion in *Details*.

Moreover, *Clear* button can be used to clear the prompt box.

3.5 Window

Window is located in the middle of the program and displays the labeled results before and after image processing and modification. Users can gradually correct the visualized image in *Window* and save the final results with the following function:

High contrast image, open the high contrast image of the original one by automatic contrast adjustment;

Zoom-in, enlarge the image for easy annotation;

Zoom-out, reduce the image to facilitate overall viewing.

As mentioned above, if one candidate figure displayed in *Window* is recognized or modified, more detailed information will be printed out to the *Details*.

3.6 Other tips

Multi-language support: Currently supports both Chinese and English.

Software automatically selects the appropriate annotation files, prioritizes selecting manual annotation files, and then automatic annotation files.

Recognition function supports both folder recognition (automatically detects all files under the path) and individual image recognition.

Step 4 Getting started with WSCApp.exe

Several dynamic video examples will be given, available on GitHub at the following link <https://github.com/Loyage/WSCNet2>, and recognized to cast a glance at the application of WSCApp.

Supplementary Note 2. Illustration of two segmentation algorithms for dark-field and bright-field scenarios.

We designed both adaptive scale template matching (ASTM) and contour extraction & Hough transform (CEHT) for dark-field and bright-field scenarios, respectively.

Firstly, the Otsu threshold segmentation algorithm is employed by ASTM algorithm to extract the droplet foreground in dark-field scenario. Specifically, assuming the binary template is noted by \mathcal{T} and its initial size is $r \times r$, the matching response of each pixel in the foreground can be computed by:

$$\mathcal{D}(x, y) = \sum_{a, b \in [0, r)} \mathcal{T}(a, b) \mathcal{F}\left(x + a - \frac{r}{2}, y + b - \frac{r}{2}\right) \quad \text{Equation (S.1)}$$

Where \mathcal{F} denotes the binary foreground image, the bounding circle of the i th droplet is denoted by $\mathcal{C}_i = [x_i, y_i, r_i]$ and $[x_i, y_i]$ represents the center of i th droplet and r_i suggests its radius.

The maximum of matching responses \mathcal{D}_{\max} corresponds to the largest foreground area covered by the template, which can be computed by $\pi r_{\text{cur}}^2 \mathcal{D}_{\max}$. Intuitively, the shrinking template radius for the next time matching can be provided by:

$$r_{\text{next}} = \sqrt{\mathcal{D}_{\max}} r_{\text{cur}} \quad \text{Equation (S.2)}$$

In the implementation, we simply need to set the upper and lower bound of the template radius to perform a limited number of searches. Since overlapping bounding circles might occur in one true droplet, non-maximum suppression (NMS)^[1] is employed to remove redundant circles.

The CEHT algorithm was designed for bright-field scenario. In the contour extraction (CE) algorithm, we first perform adaptive binarization on the droplet image, followed by an "open operation" (corrosion followed by dilation) to reduce burrs in the image. We then use the findContours method to extract the contours in the image and select them based on the number of 0 and 1 values in the contours. Next, we transform the circumferential circles on the selected contours and filter these circles using a series of conditions, including but not limited to maintaining suitable areas, centers within the image, not overlapping with other circumferential circles or sharing a large intersection to union ratio (IoU). After the above steps, most of the circular droplets will be extracted by the algorithm.

However, in the experiments, we found that the CE algorithm occasionally fails to

successfully extract droplet proposals. Therefore, we added the Hough transform (HT) algorithm to complement the CE algorithm. HT is a very common algorithm for extracting circular objects. However, due to HT's poor ellipse recognition performance and the difficulty in maintaining standard circular contours during the dynamic droplet generation process, the center position of droplets directly recognized using HT is often far from the true position and the radius deviation is also significant. Therefore, we designed the CE algorithm supplemented by HT (CEHT) as the main algorithm for droplet extraction in a bright-field scenario. In short, we first applied the CE algorithm to obtain a batch of droplet proposals, then applied the HT algorithm to the same image to obtain a new batch of results, then searched for undiscovered droplet proposals from the new results (with a distance greater than the sum of the radii of any existing proposals), and finally added them to the previous droplet proposals.

The final segmentation of droplet proposals was produced by CEHT combining the results of two parallel methods: 1. After filtering and thresholding, edge operators are used to detect the droplet edge with significant changes in pixel values; 2. Hough transforms 2D feature extraction into searching for a point in the high-dimension space determined by the radius and center coordinates of the circle.

Supplementary Note 3. Details for the design and implementation of CNN-based classifiers and WSCNet models

The second stage is the recognition of different cells encapsulated droplets from the droplet proposals provided by the droplet segmentation algorithms. Most studies train deep convolutional neural networks to simplify this issue by classifying the droplet proposals into four categories: false positive, empty droplet, single-cell, and multicell encapsulations. Classification-based algorithms, including Lenet-5^[2], ResNet18^[3], ResNet50^[3] and Mobilenet-v3-small^[4], were applied as benchmarks to simplify this task by classifying the proposals. All these networks are implemented directly by calling the basic network structure in torchvision.models, with finetuning the output size of the last fully connected layer of the network to fit the four categories.

During the training process, we used cross entropy loss and Adam optimizer^[5] with an initial learning rate of 0. The batch size is typically set to 64. However, when training with small batches of dataset to test the performance of different models under insufficient training samples, the batch size is appropriately adjusted according to the sample number of the dataset. During the training process, we used data augmentation, including random flipping (e.g. RandomHorizontalFlip and RandomVerticalFlip functions) and rotation of images (e.g. RandomAffine and RandomRotation functions), to improve the generalization performance of the model and reduce the impact of overfitting phenomena.

Unfortunately, their training time and resource consumption will greatly increase as the network deepens. Moreover, lacking interpretability, the classification task can neither count the specific number of cells inside each droplet proposal nor predict the location of each cell. We previously design the WSCNet model to estimate the number of cells inside each droplet proposal and predict the location of each cell. To avoid tedious manual cell-level annotation, only three droplet-level labels, including empty, single-cell, and multicell capsulation, are adopted in this work.

The WSCNet consists of two branches: classification and counting. The classification branch serves as a filter to remove false positives in droplet proposals, and the counting branch counts and locates cells inside each droplet. The output of the counting branch is a single channel density map of cells. The loss function of the counting branch can be given by:

$$L_{count} = \|f(\mathcal{D}) - y\|^2 + \psi(\mathcal{D}) \quad \text{Equation (S.3)}$$

Where, $f(\mathcal{D})$ and $\psi(\mathcal{D})$ are shown in Eq. (4) and Eq. (5), respectively.

$$f(\mathcal{D}) = \begin{cases} \mathcal{D}_{sum} & \mathcal{D}_{sum} < 2 \\ 2 + \gamma \mathcal{D}_{sum} & \mathcal{D}_{sum} \geq 2 \end{cases} \quad \text{Equation (S.4)}$$

$$\psi(\mathcal{D}) = \begin{cases} \mathcal{D}_{max} - 1 & \mathcal{D}_{max} > 1 \\ 0 & \mathcal{D}_{max} \leq 1 \end{cases} \quad \text{Equation (S.5)}$$

Where, \mathcal{D}_{sum} indicates the integral of the density map, which can be obtained by global sum pooling:

$$\mathcal{D}_{sum} = \sum_{i,j} \mathcal{D}(i,j) \quad \text{Equation (S.6)}$$

\mathcal{D}_{max} represents the max value in the density map, which can be obtained by global max pooling.

The loss function of the whole network containing the classification and the counting branches is given with a weight of λ :

$$L = \lambda L_{count} + (1 - \lambda) L_{class} \quad \text{Equation (S.7)}$$

Where, L_{class} denotes the loss function of cross entropy for the classification branch. The integral and the local maxima of the density map indicate the number and the location of cells, respectively.

In experiments, we have found that, due to the limited characteristics of some objects (e.g. beads), a shallow CNN-based classifier can achieve high recognition accuracy, such as Lenet-5. Another advantage of small models is that the time spent on model training and inference is further reduced, which provides great help in processing large-scale videos or images. Overall, however, Resnet18 is the best choice (of the four models we tested) for classifiers. First, its structure is not complex and the training and inference time is relatively low. Second, when the difficulty of the dataset increases, such as cells being difficult to identify due to high transparency, the accuracy of Resnet18 is much better than other classifiers. Third, our WSCNet model can not only categorize droplets by encapsulating cell populations, but also locate cells without any supervised location knowledge, showing an advantage over CNN-based classifiers. In summary, the selection of models needs to vary according to the encapsulated cells of the droplet experiment.

Supplementary Figure

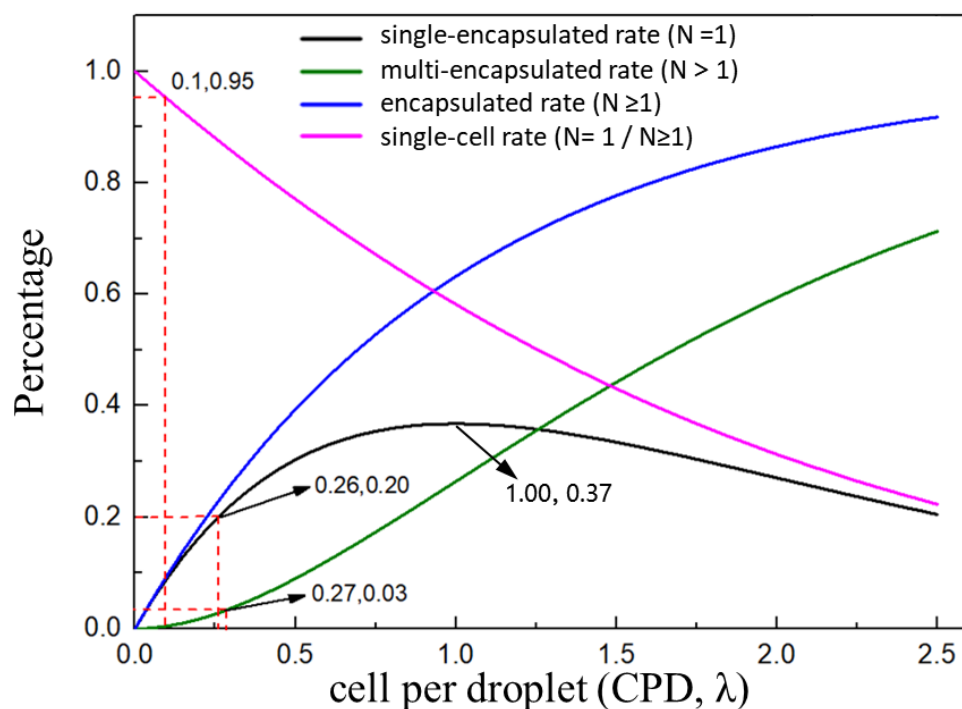


Figure S1. The encapsulated probability restricted by the Poisson's distribution. The number of cells encapsulated per droplet in passive methods is randomly restricted by Poisson statistics^[6]. The probability distribution of single-encapsulated ($N = 1$), multi-encapsulated ($N > 1$), encapsulated ($N \geq 1$), and single-cell rate ($N=1 / N \geq 1$, the proportion of droplets containing exactly one cell to that containing cells) is plotted with different colors. Supposing the number of cells encapsulated per droplet is X , X follows Poisson distribution $X \sim P(\lambda)$.

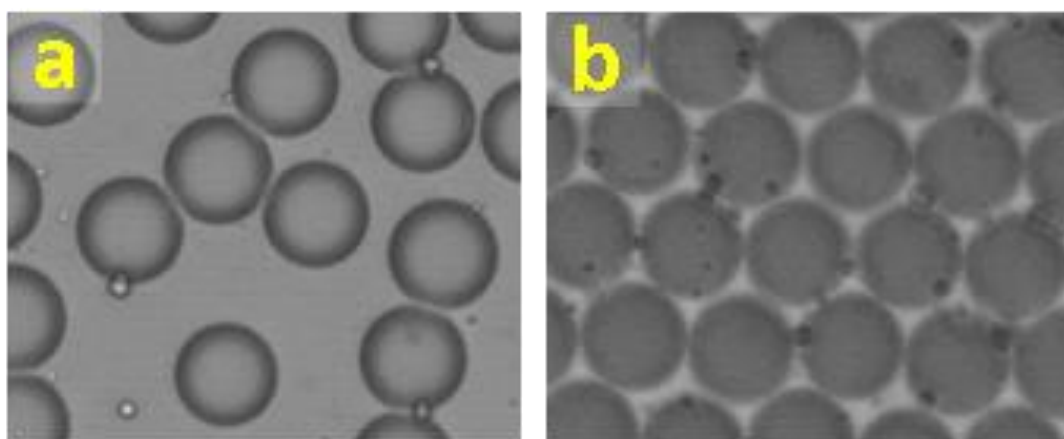


Figure S2. Representative images of bright-field and dark-field droplets under different illumination conditions. (a) Bright-field images of discrete and transparent droplets with clear edges. (b) Dark-field images of highly adherent and opaque droplets. Both transparent and separate droplets or opaque and highly adherent droplets are observed in this study. Mineral oil-based droplets usually have a rough surface and were more difficult as shown in S2 (b).

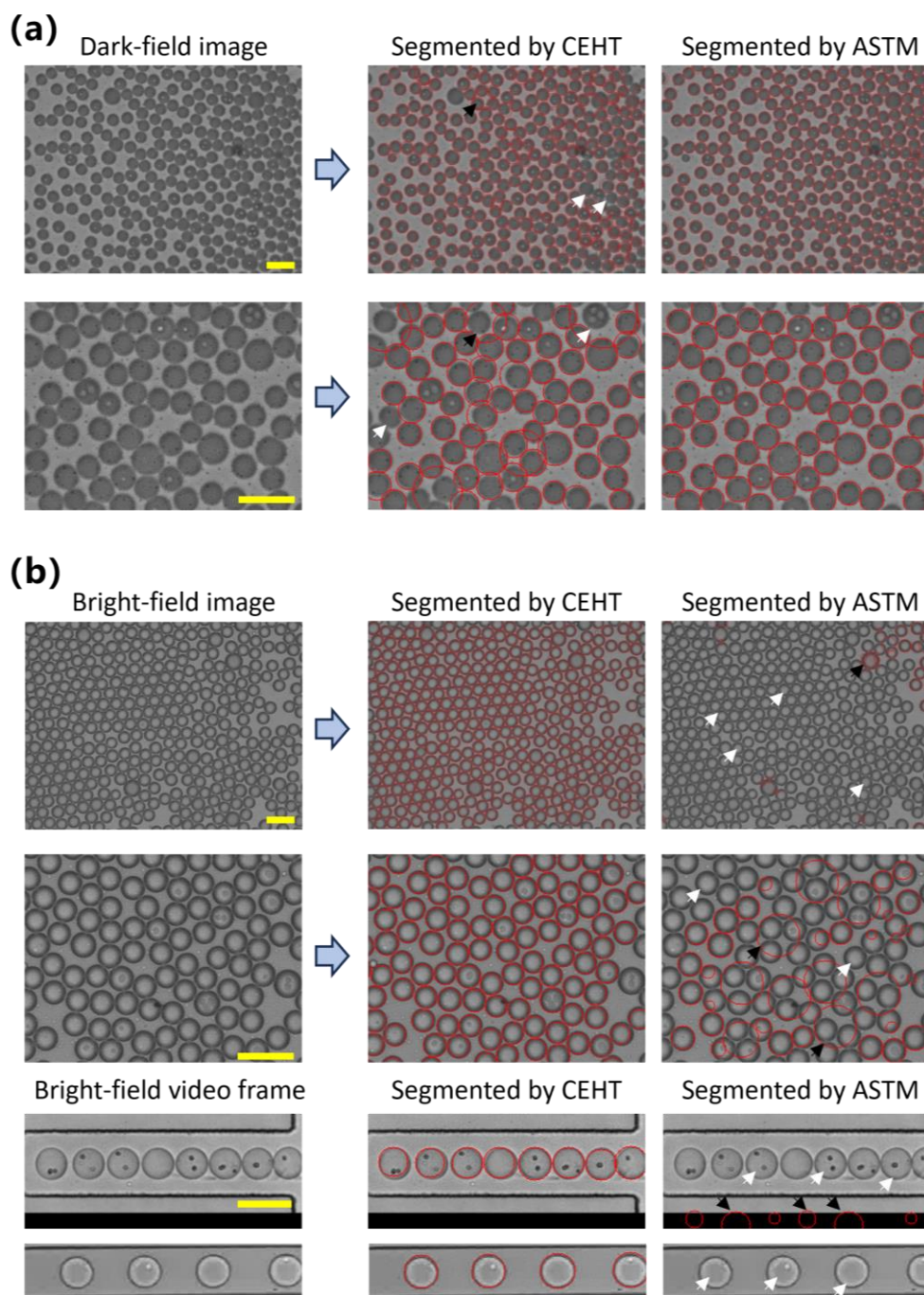


Figure S3. Comparison of two segmentation algorithms on bright-field and dark-field droplets. (a) Representative results of two segmentation algorithms on bright-field images with low-grayscale droplets. (b) Representative results of two segmentation algorithms on light-field images or video frames with high-grayscale droplets. We designed adaptive scale template matching (ASTM) and contour extraction followed by Hough transformation (CEHT) segmentation algorithms for dark-field and bright-field imaging conditions, respectively. White and black arrows denote false negative and false positive, respectively.

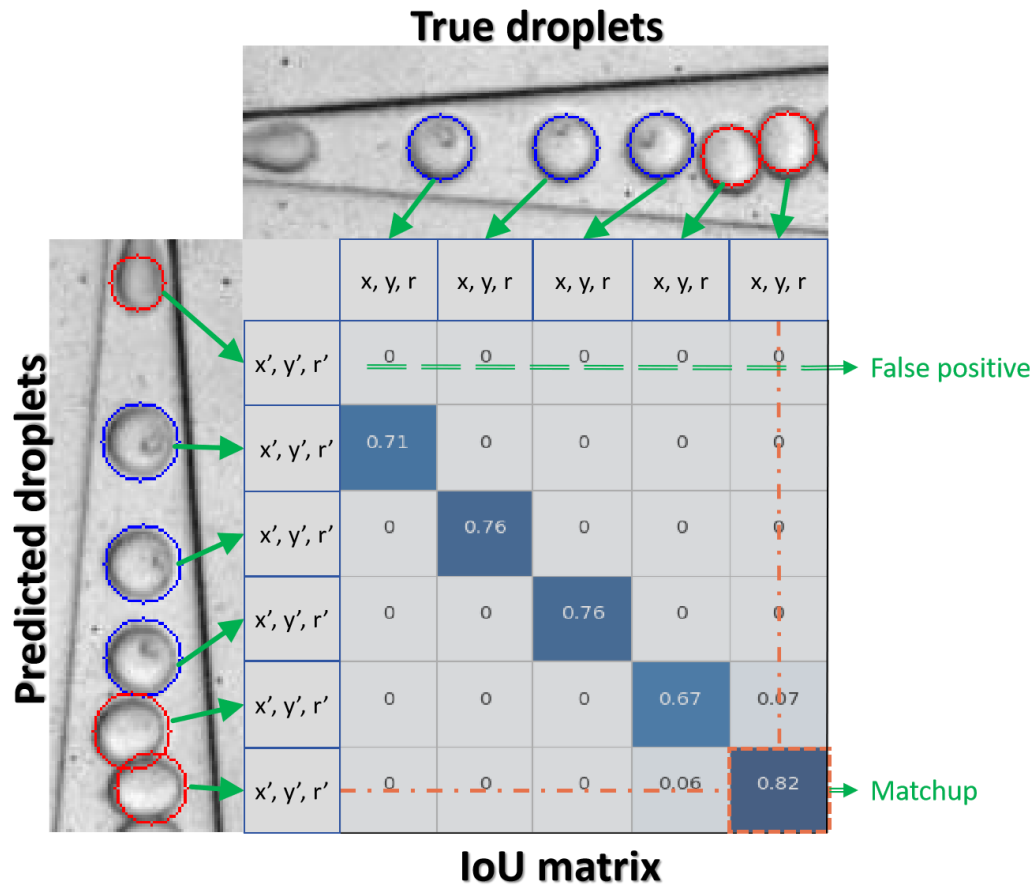


Figure S4. Illustration for evaluation metric calculation of segmented droplets under different IoU thresholds. The number of droplets that has an IoU result greater than a threshold θ is illustrated. In the evaluation metric, the algorithm will compare the identified droplet positions with their true positions one by one to form an IoU matrix, as shown above. The maximum value will be selected from the IoU matrix to determine the matching relationship between droplets. In addition, if the IoU values of a certain row/column in the matrix are all below the IoU threshold (e.g. 0.5), it is determined that the droplet is a false positive/false negative target. This metric is used to measure the degree of correlation between predicted and actual droplets, and a higher IoU represent a more stringent requirement on the correlation.

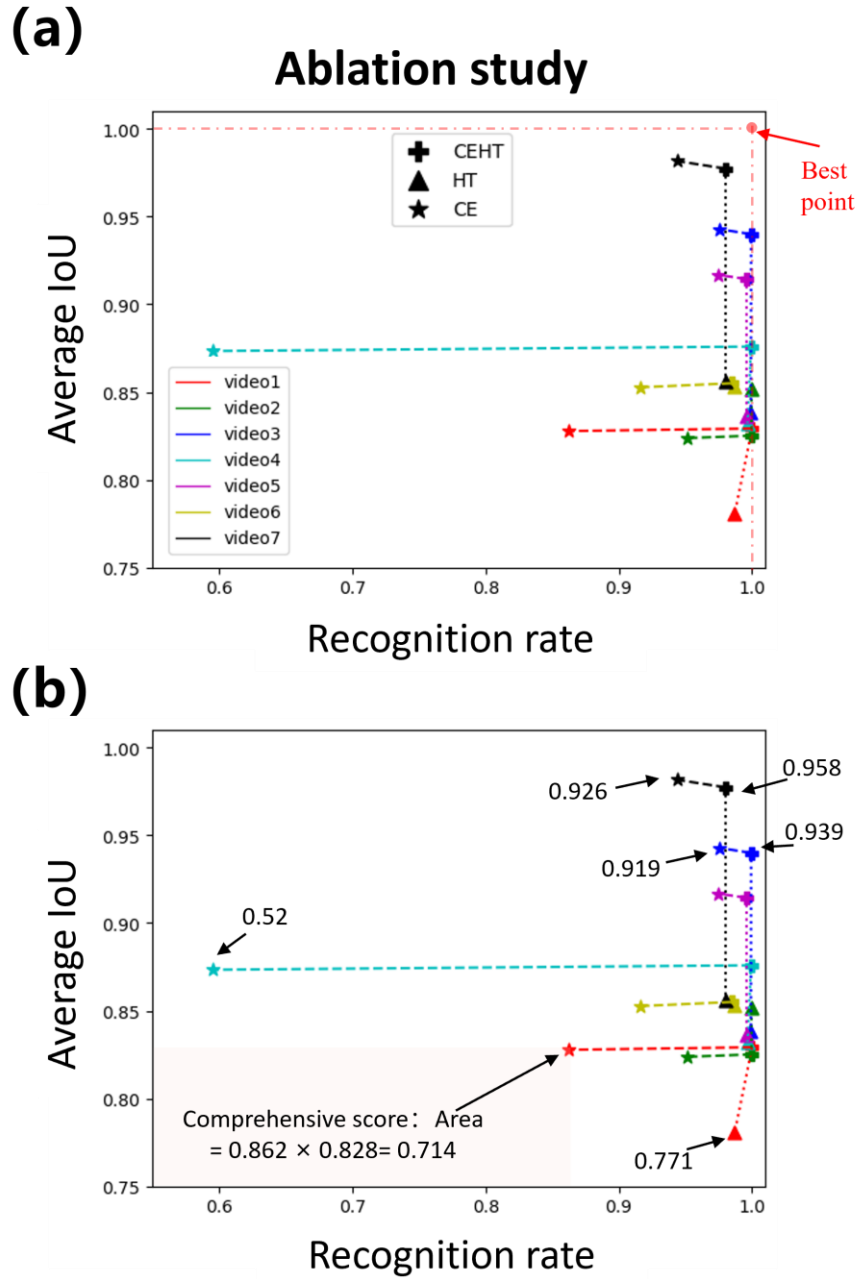
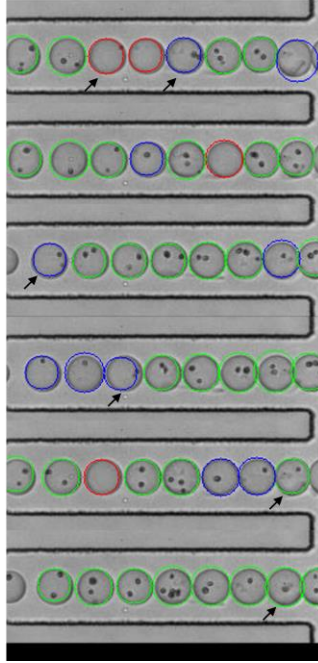


Figure S5. Distribution diagram of recognition rate and average IoU value of different datasets. (a) Ablation study of CEHT with contour extraction (CE) alone or Hough transform (HT) alone. The average Interaction over Union (IoU) value represents the average value of all bounding circles for droplet proposals, including leak detection (IoU = 0), which demonstrates the accuracy of droplet positioning. The best point denotes an ideal result with a 100% recognition rate at an average IoU of 1 (completely overlapping). (b) Comparative results of the ablation study for the segmentation of darkfield droplets. The comprehensive score is calculated as the product of the recognition rate (recall) and average IoU and indicated as the area. Results from different datasets are represented by colors, crosses, triangles, and pentagrams representing CEHT, CE, and HT methods, respectively.

(a)

**Droplet classification
by ResNet18 classifier**



(b)

**Droplet classification, cell counting
and location by WSCNet model**

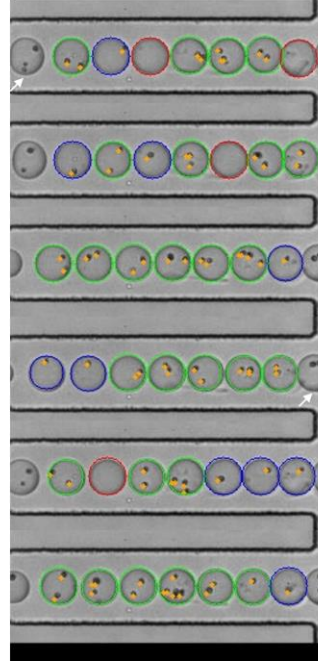


Figure S6. Representative results of droplet classification, cell counting and location predicted by ResNet18 classifier and WSCNet model. (a) Representative frames of droplets segmented and classified by ResNet18 classifier. **(b)** Representative results of droplet classification, cell counting and location using our WSCNet model. The predicted cell location is indicated as a yellow dot. All droplets were divided into empty, single-cell, and multicell encapsulations (red, blue, and green circular masks) according to the number of encapsulated cells. White and black arrows denote false negative and false positive, respectively.

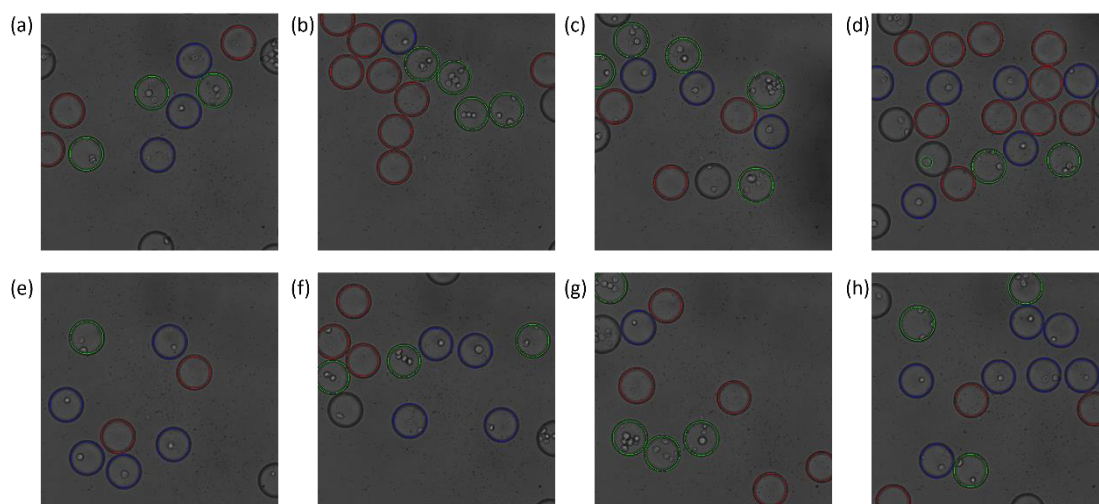


Figure S7. Representative results of detected droplets with encapsulated mammalian cells by CEHT algorithm and LeNet-5 classifier. The droplets from video 8 were divided into empty, single-cell and multicell encapsulations (red, blue, and green circular masks) and shown in (a)-(h) according to the encapsulated cell count. Image recognition of validation droplets (2129 droplets in 103 images) with encapsulated PC3 cancer cells was segmented and classified by WSCNet model, which is retrained on the training data (10452 droplets in 412 images) from scratch.

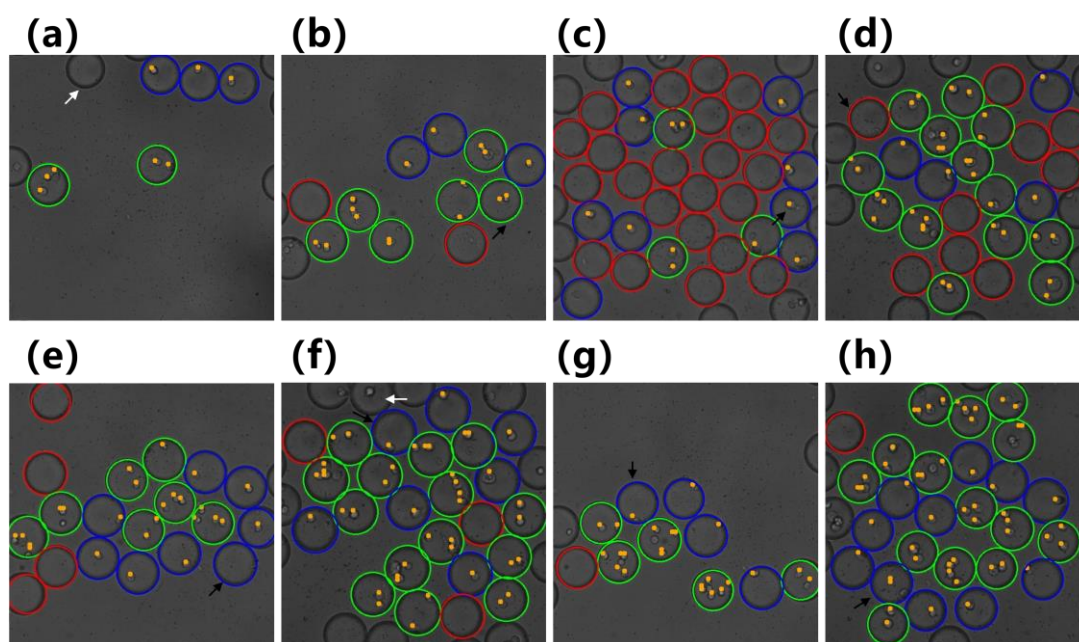


Figure S8. Representative results of classification failure by WSCNet model. White and black arrows denote false negative and false positive, respectively. Only a small number of droplet labels, e.g. locating at the edge of the field of view or cell doublet, are predicted as the wrong category. The predicted cell location is indicated as a yellow dot. All droplets were divided into empty, single-cell, and multicell encapsulations (red, blue, and green circular masks) according to the number of encapsulated cells.

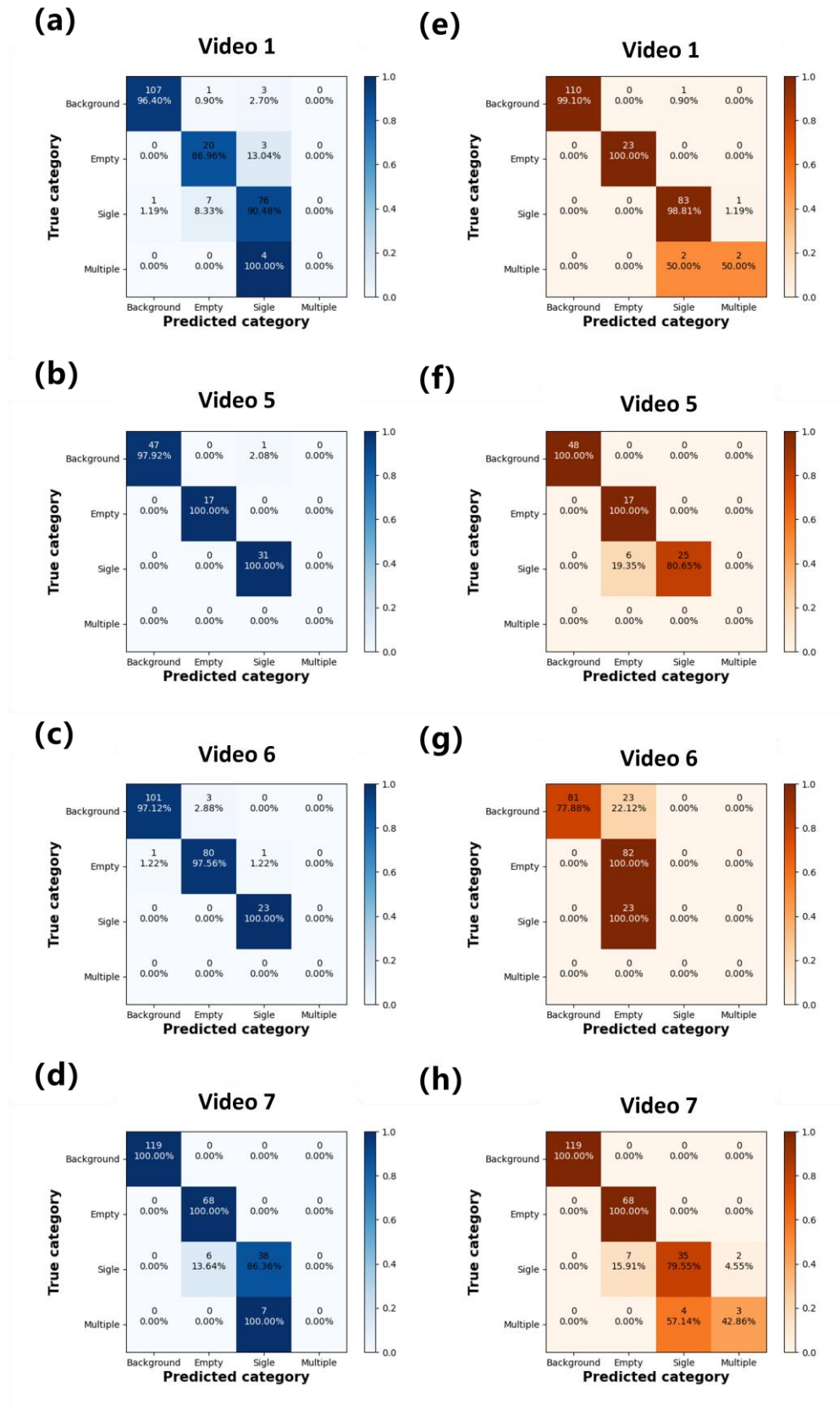
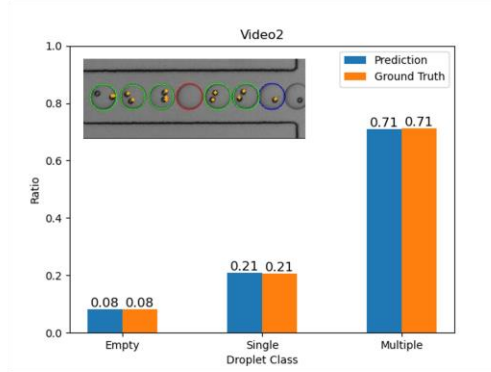


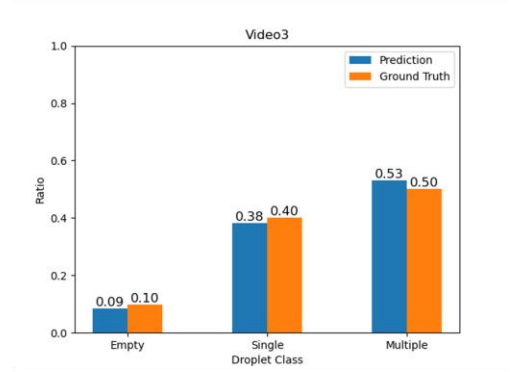
Figure S9. Multiclass confusion matrixes showing classification performance on different datasets. Confusion matrixes of independent test of video 1, video 5, video 6, and video 7 is plotted using ResNet-18 classifier (a-d) and WSCNet model (e-f) for image recognition,

respectively. Both classifier and WSCNet model were retrained from scratch. The vertical axis represents the true category of droplets while the horizontal axis represents the predicted category of droplets. The color bar on the right shows the colormap with percentage predictions where the darkest color indicates the maximum predictions.

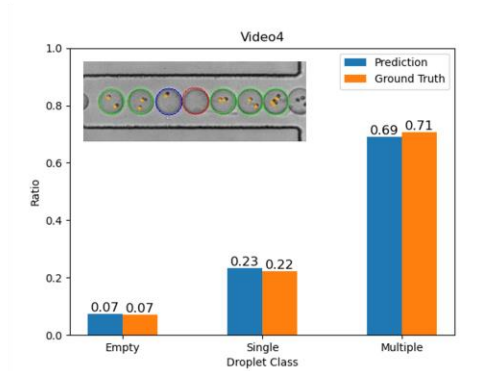
(a)



(b)



(c)



(d)

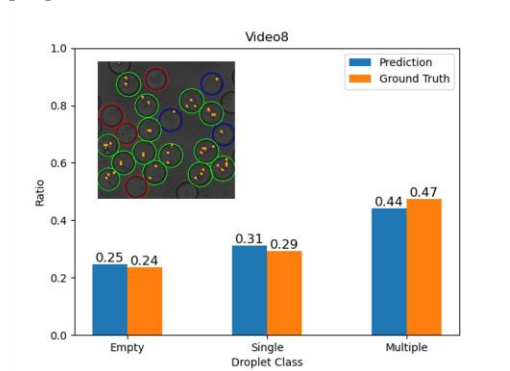


Figure S10. Comparative result of cell amount distribution encapsulated in microfluidic droplets. Cells encapsulated in droplets were counted and localized by our proposed WSCNet model while number of encapsulated cells is obtained (marked in prediction) and compared with the manual annotation (marked in ground truth). Cells in droplets are counted from the video frame by hand and marked as ground truth. The insert diagram shows the results of representative cell counting and location results.

Supplementary Table

Table S1. Summary of different approaches in cell encapsulated microfluidic droplets.

Method	Function	Encapsulated content	Speed	Precision	Annotation	Reference
Handcrafted CNN	Classification & sorting	Spheres, beads and MCF-7 cells	≈ 200 drop/s	90%	Droplet-level 100% sample	Lab on a Chip, 2020, 20(5): 889-900.
Morphological processing	Location & classification & sorting	K562 cells	≈ 50 drop/s	85%	Non	Scientific reports, 2020, 10(1): 8736
YOLO, SSD300	Location & classification	Dyes	≈ 78 drop/s	85%	Pixel-level 100% sample	Biomicrofluidics, 2020, 14(3)
YOLO v3, YOLO v5	Location & classification & cell counting	Cancer cells	≈ 1080 drop/s	91%	Pixel-level 100% sample	Lab on a Chip, 2022, 22(21): 4067-4080
YOLO v4-tiny	Location & classification & cell counting	PA beads, SH-SY5Y cells, and cell aggregates	≈ 60 drop/s	92%	Pixel-level 100% sample	Advanced Materials Technologies, 2022, 7(5): 2101053.
Image gray discrimination	Volume measurement & flow rate calculation	Dyes and Particles	≈ 100 drop/s	/	Non	Micromachines, 2023, 15(1): 27
ASTM + WSCNet	Location & classification	Yeast cells	≈ 100 drop/s	90%	Droplet-level 100% sample	Biosensors, 2023, 13(8): 821. Our previous study.
ASTM or CEHT WSCNet or classifiers (optional)	Location & classification & cell counting	Microspheres, yeast, algae, PC3 cancer cells, HL60 and K562 cells	≈ 350 drop/s	94%	Droplet-level 20% sample	This study.

* The definitions of different evaluation metrics are as follows:

1. Speed = [processed droplets] / [unit time]




2. Precision = [Correctly classified cells] / [All droplets]

** Since some papers use multi-class methods to deal with multi-class cells, the precision data cited in this table is roughly calculated from the average of the original results on each dataset based on the best performing methods.

*** Encapsulated content refers to the droplet encapsulated cells, beads, dyes or particles.

ASTM: Adaptive scale template matching; CEHT: Contour extraction followed by Hough transform; CE: Contour extraction; HT: Hough transform.

Table S2. Annotation summary and inference time of cell encapsulated droplets on different datasets. The processing frames per second (FPS) is defined as inverse of the average inference time when using Resnet-18 as the classification model.

all Datasets	Rec	Frames	Droplets	Empty	Single	Multiple	Processing FPS	Reference
 Video1.avi		108	537	118	410	9	50	Kemna E. W. M., Schoeman R. M., Wolbers F., et al. High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel. Lab on a Chip, 2012, 12(16): p. 2881-2887.
 Video2.avi		201	1425	115	295	1015	47	Lagus T. P., Edd J. F. High-throughput co-encapsulation of self-ordered cell trains: cell pair interactions in microdroplets. Rsc Advances, 2013, 3(43): p. 20512-20522.
 Video3.avi		733	5332	524	2137	2671	47	





	351	2540	182	563	1795	46	
Video4.avi							
	80	236	93	143	0	61	Edd J. F., Di Carlo D., Humphry K. J., et al. Controlled encapsulation of single-cells into monodisperse picolitre drops. Lab on a Chip, 2008, 8(8):1262.
Video5.mov							
	112	514	389	125	0	52	
Video6.mov							
	300	1203	749	398	56	54	https://www.lbc.espci.fr/home/gallery/microfluidic-movies/
Video7.mp4							

Table S3. Detailed parameters of droplet diameter distribution of three independent experiments, including 4 135, 2 286 and 3 981 droplets, respectively.

Statistical Parameters	Experiment 1	Experiment 2	Experiment 3
Number of static images	20	22	20
Number of statistic droplets	4135	2286	3981
Minimum radius (μm)	13.8417	13.7067	14.2379
Maximum radius (μm)	27.6848	28.1934	29.9137
Mean radius (μm)	21.8777	23.1193	22.2542
Median radius (μm)	20.0861	23.1389	22.2595
Standard deviation in radius (μm)	1.0005	0.8092	0.9983
Coefficient of variation (CV, %)	4.5733	3.5002	4.4861

Table S4. Comparative performance of different algorithms for the segmentation of dark-field droplets. **Bold** and **blue** fonts suggest the best and second-best performance, respectively. Comprehensive score is calculated as the product of recognition rate (recall) and average IoU.

Evaluation metrics	Recognition rate			Average IoU			Comprehensive score		
Algorithm	HT	CE	CEHT	HT	CE	CEHT	HT	CE	CEHT
Video 1	0.987	0.862	1.000	0.781	0.828	0.829	0.771	0.714	0.829
Video 2	1.000	0.951	1.000	0.852	0.824	0.825	0.852	0.783	0.825
Video 3	0.999	0.975	0.999	0.838	0.943	0.940	0.837	0.919	0.939
Video 4	0.996	0.595	0.999	0.832	0.873	0.876	0.829	0.520	0.875
Video 5	0.996	0.975	0.996	0.836	0.917	0.914	0.833	0.893	0.911
Video 6	0.986	0.916	0.984	0.853	0.853	0.855	0.842	0.781	0.842
Video 7	0.980	0.943	0.980	0.856	0.982	0.977	0.839	0.926	0.958

CEHT: Contour extraction followed by Hough transform; CE: Contour extraction ; HT: Hough transform.

References:

- [1] Bodla, N., B. Singh, R. Chellappa, *et al.* *Soft-NMS - Improving Object Detection With One Line of Code*. in *16th IEEE International Conference on Computer Vision (ICCV)*. **2017**. Venice, ITALY: IEEE.
- [2] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. *Gradient-based learning applied to document recognition*. *Proceedings of the IEEE*, **1998**, 86(11), 2278-2324.
- [3] He, K.M., X.Y. Zhang, S.Q. Ren, *et al.* *Deep Residual Learning for Image Recognition*. in **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. 2016. Seattle, WA: IEEE.
- [4] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. *Searching for mobilenetv3*. In *Proceedings of the IEEE/CVF international conference on computer vision*, **2019** (pp. 1314-1324).
- [5] Kingma, D. P., & Ba, J. *Adam: A method for stochastic optimization*. arXiv preprint arXiv: **2014**, 1412.6980.
- [6] Collins, D.J., A. Neild, A. deMello, *et al.*, *The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation*. *Lab on a Chip*, **2015**. 15(17): p. 3439-3459.
- [7] Anagnostidis V., Sherlock B., Metz J., *et al.*, *Deep learning guided image-based droplet sorting for on-demand selection and analysis of single cells and 3D cell cultures*. *Lab on a Chip*, **2020**. **20**(5): p. 889-900.
- [8] Sesen, M., Whyte, G. *Image-Based Single Cell Sorting Automation in Droplet Microfluidics*. *Scientific Reports*, **2020**. 10: p. 8736.
- [9] Cantwell, C., McGrath, J.S., Smith, C.A. *et al.* *Image-Based Feedback of Multi-Component Microdroplets for Ultra-Monodispersed Library Preparation*. *Micromachines*, **2023**. **15**(1): p. 27.
- [10] Gardner K., Uddin M.M., Tran L., *et al.* *Deep learning detector for high precision monitoring of cell encapsulation statistics in microfluidic droplets*. *Lab on a chip*, **2022**, 22(21): p. 4067-4080.
- [11] Arjun, A., R. R. Ajith, and S. Kumar Ranjith. *Mixing characterization of binary-coalesced droplets in microchannels using deep neural network*. *Biomicrofluidics*, **2020**, 14(3).
- [12] Howell L., Anagnostidis V., Gielen F. *Multi-object detector YOLOv4-tiny enables high-throughput combinatorial and spatially-resolved sorting of cells in microdroplets*. *Advanced Materials Technologies*, **2022**, 7: 2101053
- [13] Zhou, X., Mao, Y., Gu, M., *et al.*, *WSCNet: Biomedical Image Recognition for Cell Encapsulated Microfluidic Droplets*. *Biosensors*. **2023**; 13(8): p.821.
- [14] Kemna E. W. M., Schoeman R. M., Wolbers F., *et al.* *High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel*. *Lab on a Chip*, **2012**, 12(16): p. 2881-2887.
- [15] Lagus T. P., Edd J. F. *High-throughput co-encapsulation of self-ordered cell trains: cell pair interactions in microdroplets*. *Rsc Advances*, **2013**, 3(43): p. 20512-20522.
- [16] Edd J. F., Di Carlo D., Humphry K. J., *et al.* *Controlled encapsulation of single-cells into monodisperse picolitre drops*. *Lab on a Chip*, **2008**, 8(8): p.1262.