An Efficient Approach for Droplet Coalescence Videos Processing based on Instance Segmentation and Multi-Object Tracking Algorithms

Wenle Xu,a,b Shuyuan Zhang,a,b Kai Wang,a,c Tong Qiua,b\*

aDepartment of Chemical Engineering, Tsinghua University, Beijing, 100084, CHINA

bBeijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing, 100084, CHINA

cThe State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, CHINA

qiutong@tsinghua.edu.cn

Abstract

Controlled coalescence of droplets is a crucial method of performing reactions and synthesises within droplets. Among all methods employed for droplet characterization within microchannels, microscopic imaging stands out for its capacity to capture ample information. However, the processing of images and videos still predominantly relied on massive manual works, which falls short of meeting the demands for high-throughput analysis. To address this problem, this paper proposes an efficient approach based on instance segmentation and multi-object tracking algorithms to analyse the droplet coalescence videos in microchannels. This approach initially segments droplets in microscopic images and consequently associate the identical droplets and recognize the coalescence processes across consecutive frames. Finally, further analysis of these data can yield critical statistics of the droplet coalescence process, such as coalescence probability and coalescence time. This approach enables automated and efficient analysis of videos to decipher the droplet coalescence process, thereby accelerating the discovery and exploration of droplet coalescence patterns in microfluidics.

**Keywords**: Microfluidics, Droplet Coalescence, Instance Segmentation, Multi-Object Tracking.

* 1. Introduction

Droplet coalescence is considered as a pivotal technique in microfluidics, with significant potential applications in versatile chemical such as the formation of particles, kinetics studies, and chemical synthesis (Teh et al., 2008). Conducting droplet coalescence experiments on microfluidic platforms equipped with microscope and high-speed cameras offers distinct advantages, including high throughput, precise control, and minimal resource consumption. However, researchers often grapple with the manual processing of large volumes of microscopic photos and videos, a task that consumes several weeks to accumulate sufficient data and obtain related statistics. This manual approach results in a substantial gap between the speed of data generation and analysis. Although several commercial tools can assist in automatically identifying droplets in photos, these tools are struggled to handle droplets with irregular shapes, not to mention that the extraction of coalescence time and coalescence probability requires analyzing videos frame by frame.

Some researchers have made attempts to apply deep learning based computer vision algorithms to process microchannel microscopic photos and videos (Rutkowski et al., 2022; S. Zhang et al., 2022). Notably, up to this point, no deep learning model has been proposed to process videos recorded coalescence process, the models aforementioned lack the ability to identify droplets before and after the coalescence occurs.

This paper proposes an efficient approach based on instance segmentation and multi-object tracking algorithms. This approach begins by employing a convolutional neural network-based instance segmentation model to detect and segment droplets in microscopic images. Subsequently, the droplet tracking and coalescence judging algorithm can associate identical droplets and recognize coalescence processes across consecutive frames. Finally, further analysis of these data can yield critical statistics of the droplet coalescence processes such as coalescence probability, the distribution of coalescence numbers and coalescence time. Compared to manual analysis, this approach can automatically and intelligently analyze the droplet coalescence videos, achieving human-level accuracy and significantly faster video processing speed.

* 1. Method

The approach to process droplet coalescence videos involves three main stages: data pre-processing, instance segmentation, and droplet tracking and coalescence judging.

* + 1. Data pre-processing

The videos analyzed in our paper are derived from a droplet coalescence experiment in microchannels with a sudden expansion chamber (Wang et al., 2016).

The data pre-processing stage involves two main steps: video splitting and image annotation. Since the subsequent instance segmentation and droplet tracking and coalescence judging algorithms operate on images as the basic unit rather than videos, it is necessary to split the video into individual images. To capture the complete coalescence process, the videos are split at a frequency of once per frame. Following this, image annotation is performed. Once considered a laborious task, it requires the outlining of thousands of droplets to train an instance segmentation model. Each droplet necessitated more than 30 precise clicks around its contour. Fortunately, leveraging state-of-the-art pretrained models such as Segment Anything Model (Kirillov et al., 2023), the outlining task can now be achieved with a single click anywhere on the droplets’ bodies. It significantly accelerates the annotation by more than 30x.

A total of approximately 1500 droplets in 473 images are annotated. Subsequent to annotation, the images are divided into training and testing sets, with detailed information shown in Table 1.

Table 1. The characteristics of training set and testing set.

|  |  |  |
| --- | --- | --- |
|  | Training Set | Testing Set |
| Total number of images | 372 | 101 |
| Average number of droplets per image | 3.2 | 3.3 |
| Average projection area (pixels) | 11450 | 12190 |

* + 1. Instance segmentation

Instance segmentation models based on convolution neural networks (CNNs) can be categorized into 2 types: mask-based models and contour-based models. While both types leverage CNNs to extract features, the former processes the mask of each instance, and the latter focuses on contours. In common objects dataset, mask-based models usually outperform in accuracy but tend to be slower compared to contour-based models. However, the most notable distinction of the droplets and background lies in their contours, contour-based models may achieve a higher accuracy. Consequently, unlike the mask-based model Mask R-CNN utilized in previous work, a contour-based model, E2EC, is employed to extract the contours of droplets in our paper, potentially achieving better accuracy and faster inference speed (He et al., 2017; T. Zhang et al., 2022).

Transfer learning is particularly useful in downstream applications where there is a lack of sufficient and high-quality data. In this paper, transfer learning is implemented by transferring the weights of E2EC pretrained on COCO dataset (Lin et al., 2014). Furthermore, a series of reductions are applied to the training set to simulate the common dilemma of lacking sufficient annotated images, the sizes of training sets are shown in Table 2. The performances of the models (E2EC, E2EC with transferred weights, and Mask R-CNN) trained on these reduced training sets are compared.

Table 2. Total number of the training sets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Training Set 1 | Training Set 2 | Training Set 3 | Training Set 4 | Training Set 5 |
| Total number of images | 372 | 186 | 93 | 37 | 18 |

The confusion matrix depending on Intersection over Union (IoU) is depicted in Figure 1. The mean Average Precision (mAP) across a series of IoU is used to evaluate the instance segmentation results, as defined in Eq. (1).

|  |  |
| --- | --- |
|  | (1) |

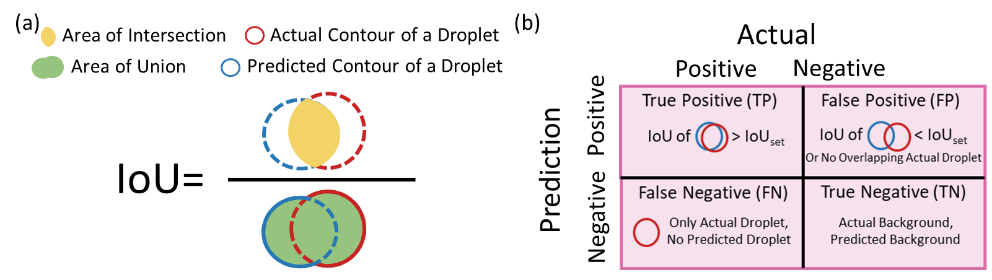


Figure 1. (a) A computation example of intersection over union. (b) Confusion matrix of droplet instance segmentation.

* + 1. Droplet tracking and coalescence judging

This paper proposes a droplet tracking and coalescence judging algorithm (shown in Figure 2.), which consists of a droplet tracking part and a coalescence part. The droplet tracking part adopts the idea of performing assignments based on a distance matrix in SORT (Bewley et al., 2016), and the coalescence judging part is designed based on three fundamental assumptions that (1). No droplet will disappear unless it flows out of the microchannel, (2). The droplet’s area does not increase suddenly, and (3). Coalescence is based on contacting. The workflow of the algorithm can be summarized as follows:

I. Estimate: Current positions of tracks (referring to the historical droplets in frame T-1) are estimated based on their historical speeds and positions.

II. Associate: An association refers to link detections (meaning the instance segmentation results in frame T) to their tracks. An association cost matrix between all tracks and detections is computed by Eq. (2). Then, the association problem is converted to an assignment problem which is solved by using the Hungarian algorithm. Additionally, any association with an association cost exceeding a predetermined threshold is rejected.

|  |  |
| --- | --- |
|  | (2) |

After assignment, the tracks and detections are divided into three sets: the unassociated tracks set (UT), the unassociated detections set (UD), and the associations set (AS).

III. Coalescence judging: Figure 2(b) illustrates two different coalescence scenarios. (1). Two droplets (A and B) with similar area coalesce and produce a droplet (C) in Frame T. The association cost of (A, C) and (B, C) are high, and consequently A and B are distributed to UT and C is distributed to UD. (2). A small droplet and a large droplet (A and B) coalesce and produce a droplet. (B, C) has a low enough cost to be associated.

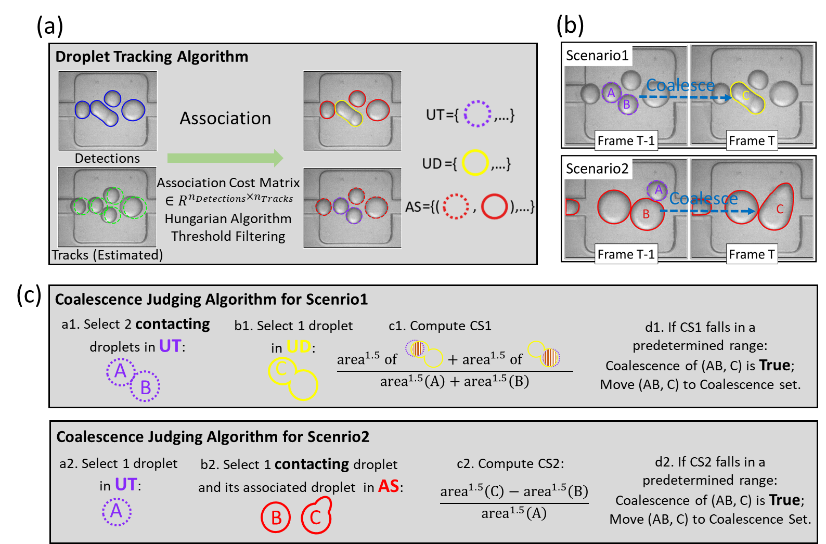


Figure 2. (a) The pipeline of droplet tracking algorithm. (b) Two different coalescence scenarios. (c) Coalescence judging algorithm for two different coalescence scenarios.

IV. Two coalescence judging algorithms are designed for these two scenarios. For the first coalescence scenario, the algorithm work as bellows:

a1. Select two contacting droplets from UT, denoted as A and B. b1. Select a droplet from UD, denoted as C. c1. Compute the coalescence score 1 (CS1) of A, B, and C by Eq. (3). d1. If CS1(AB, C) falls in a predetermined range, the coalescence event (AB, C) is considered to have occurred and (AB, C) is moved to coalescence set.

|  |  |
| --- | --- |
|  | (3) |

For the second coalescence scenario, the algorithm work as bellows:

a2. Select a droplet from UT, denoted as A. b2. Select one contacting droplet of A and its associated droplet from AS, denoted as B and C. c2. Compute the coalescence score 2 (CS2) of A, B, and C by Eq. (4). d2. If CS2(AB, C) falls in a predetermined range, the coalescence event (AB, C) is considered to have occurred and (AB, C) is moved to coalescence set.

|  |  |
| --- | --- |
|  | (4) |

V. After coalescence judging, the aforementioned three sets are further subdivided into four sets: individual droplets set, new detections set, leaving tracks set, and coalescence droplets set. The coalescence time, coalescence probability, and other critical statistics can be figured out directly based on this subdivision.

Multiple Object Tracking Accuracy (MOTA) serves as an evaluation metric for assessing the performance of tracking algorithms. It is calculated using Eq (5), which incorporates false positives (FP), false negatives (FN), and identity switches (IDSW), normalized over ground-truth (GT) tracks. FP represents the number of incorrect assignments of detections to tracks when the droplets have flowed out or coalesced. FN represents to the number of missed assignments of detections to tracks. IDSW represents the number of wrong assignments of detections to track. GT represents the total number of droplets. Precision rate and recall rate are employed to evaluate the results of coalescence judging.

|  |  |
| --- | --- |
|  | (5) |

* 1. Results and discussions
     1. Instance segmentation

The mAP of three models (E2EC, E2EC with transferred weights, and Mask R-CNN) trained on the reduced training sets are shown in Figure 3(a). Notably, E2EC outperforms than Mask R-CNN in the droplet segmentation task, and this superiority becomes more pronounced as the training set size decreases. Figure 3(b) presents a comparative example between E2EC and Mask R-CNN, both trained with a dataset consisting of 37 annotated images. E2EC demonstrates precise segmentation even for droplets with irregular shapes, while Mask R-CNN has a rough segmentation at the droplets’ contours, with large areas of missed detection. Additionally, E2EC achieves an inference speed of 22.24 images per second, while Mask R-CNN only infers 11.91 images per second, tested on a RTX 3070Ti GPU. Figure 3(c) compares E2EC and E2EC with transferred weights trained with 18 annotated images. The E2EC trained from scratch struggles to maintain its performance under the limitation of a small training set. Conversely, E2EC with transferred weights performs well, showcasing the efficacy of transfer learning for downstream tasks with sparse data.

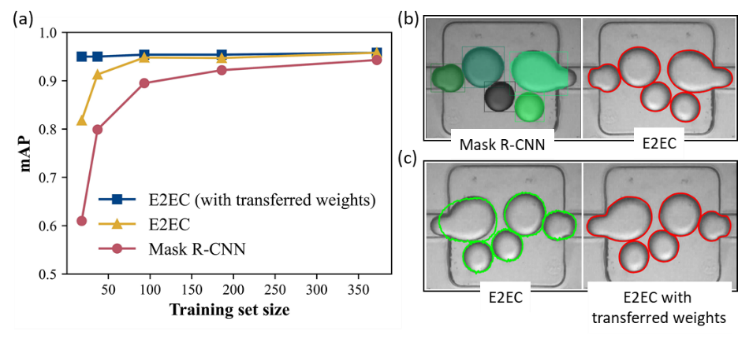


Figure 3. (a) The mAP of E2EC, E2EC with transferred weights, and Mask R-CNN trained on the training sets. (b) A comparison of E2EC and Mask R-CNN trained with 37 annotated images. (c) A comparison of E2EC and E2EC with transferred weights trained with 18 annotated images.

* + 1. Droplet tracking and coalescence judging

We select 12 videos to evaluate the algorithm. Manual analysis shows the presence of 415 droplets totally, of which 158 coalesce. The testing results for droplet tracking shows no FN or FP, and IDSW is 1, resulting in a high MOTA of 0.998. For coalescence judging, both FN and FP are zero, resulting in high precision and recall rate of coalescence judging of 1. Based on it, coalescence time and coalescence number can be precisely determined. Take coalescence time as an example, it can be directly figured out by counting the frames from when two droplets come into contact to the moment of coalescence.

* + 1. Analysis speed comparison

The only stage requiring researchers' participation is image annotation. However, this step demands only a few minutes to annotate sufficient images for training an instance segmentation model, with the assistance of pretrained models and transfer learning technology. In comparison to manual analysis, our approach demonstrates a significantly superior processing speed. For instance, the analysis of a video consisting of 1000 frames, involving the determination of droplet sizes, coalescence numbers, and coalescence time, could consume researchers dozens of hours. In contrast, our approach completes this task in approximately 50 seconds, showcasing a huge superiority.

* 1. Conclusions

In general, this paper proposes an approach based on instance segmentation and multi-object tracking algorithms to analyze droplet coalescence videos automatically and intelligently. This approach is able to precisely identify droplets and coalescence events in the video. Compared to manual analysis, this approach presents a human-level accuracy and a much faster video processing speed. Our work exhibits a possibility of applying deep learning-based computer vision technology for analyzing the critical interphase of multiphase systems in chemical engineering.

Acknowledgements

The authors gratefully acknowledge the National Natural Science Foundation of China for its financial support (Grant No. 21991100, 21991104).

References

Teh, S.-Y., Lin, R., Hung, L.-H., Lee, A.P., 2008. Droplet microfluidics. Lab Chip 8(2), 198-220.

Rutkowski, G.P., Azizov, I., Unmann, E., Dudek, M., Grimes, B.A., 2022. Microfluidic droplet detection via region-based and single-pass convolutional neural networks with comparison to conventional image analysis methodologies. Machine Learning with Applications 7, 100222.

Zhang, S., Liang, X., Huang, X., Wang, K., Qiu, T., 2022. Precise and fast microdroplet size distribution measurement using deep learning. Chem Eng Sci 247, 116926.

Wang, K., Yi, S., Zhou Q., Luo G., 2016. Effect of nanoparticles on droplet coalescence in microchannel. Journal of Chemical Industry and Engineering 67, 469-475.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN.

Zhang, T., Wei, S., Ji, S., 2022. E2EC: An End-to-End Contour-based Method for High-Quality High-Speed Instance Segmentation.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L., 2014. Microsoft COCO: Common Objects in Context.

Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B., 2016. Simple Online and Realtime Tracking.