Petroleum Science 22 (2025) 5212–5227

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Automated labeling and segmentation based on segment anything model: Quantitative analysis of bubbles in gas–liquid flow

Jia-Bin Shi [a,c], Li-Jun You [b], Jia-Chen Dang [a,c], Yi-Jun Wang [b], Wei Gong [b], Bo Peng [a,c,*]

[a] School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[b] State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[c] Intelligent Oil and Gas Laboratory, Southwest Petroleum University, Chengdu, 610500, Sichuan, China

ARTICLE INFO

Article history:
Received 11 March 2025
Received in revised form
5 August 2025
Accepted 8 August 2025
Available online 11 August 2025

Edited by Yan-Hua Sun

Keywords:
Dispersed phases
Bubble segmentation
Microfluidic system
Segment anything model
Gas–liquid flow
Artificial intelligence

ABSTRACT

The quantitative analysis of dispersed phases (bubbles, droplets, and particles) in multiphase flow systems represents a persistent technological challenge in petroleum engineering applications, including CO_2-enhanced oil recovery, foam flooding, and unconventional reservoir development. Current characterization methods remain constrained by labor-intensive manual workflows and limited dynamic analysis capabilities, particularly for processing large-scale microscopy data and video sequences that capture critical transient behavior like gas cluster migration and droplet coalescence. These limitations hinder the establishment of robust correlations between pore-scale flow patterns and reservoir-scale production performance. This study introduces a novel computer vision framework that integrates foundation models with lightweight neural networks to address these industry challenges. Leveraging the segment anything model's zero-shot learning capability, we developed an automated workflow that achieves an efficiency improvement of approximately 29 times in bubble labeling compared to manual methods while maintaining less than 2% deviation from expert annotations. Engineering-oriented optimization ensures lightweight deployment with 94% segmentation accuracy, while the integrated quantification system precisely resolves gas saturation, shape factors, and interfacial dynamics, parameters critical for optimizing gas injection strategies and predicting phase redistribution patterns. Validated through microfluidic gas–liquid displacement experiments for discontinuous phase segmentation accuracy, this methodology enables precise bubble morphology quantification with broad application potential in multiphase systems, including emulsion droplet dynamics characterization and particle transport behavior analysis. This work bridges the critical gap between pore-scale dynamics characterization and reservoir-scale simulation requirements, providing a foundational framework for intelligent flow diagnostics and predictive modeling in next-generation digital oilfield systems.

1. Introduction

As global energy demand continuously grows and shallow oil and gas resources are gradually being depleted, the extraction of deep and unconventional resources (e.g., shale oil and gas, tight oil and gas, etc.) has increasingly become a research focus (He et al., 2023). In the development of unconventional oil and gas, gas flooding technologies (such as CO_2 flooding and N_2 flooding) have been widely applied due to their ability to significantly enhance oil and gas recovery rates (Hamza et al., 2021; Tan et al., 2025). However, the microscopic mechanisms of gas–liquid two-phase flow during gas flooding processes are highly complex, especially within porous media, where the migration, aggregation, and distribution of bubbles directly affect gas flooding efficiency and energy transfer effectiveness (Liu et al., 2019a). Therefore, accurately observing and analyzing the state and behavior of bubbles within porous media is critical for understanding the gas flooding mechanisms in unconventional oil and gas development (Liu and Zhang, 2019).

* Corresponding author.
E-mail address: bopeng@swpu.edu.cn (B. Peng).

Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2025.08.018

To address this issue, microfluidic technology has been increasingly applied in energy engineering in recent years (Wang Y.-J. et al., 2025). Microfluidic systems, by designing micron-scale channels and pore structures, can simulate fluid behavior that more closely resemble those within real porous media, providing a visual platform for studying gas–liquid two-phase flow (Gavoille et al., 2019; Jahanbakhsh et al., 2020; Song et al., 2022). By etching microfluidic chips with specific pore structures, the flow processes of gas and liquid in complex porous media can be simulated, enabling the visual observation of dynamic behavior such as the migration and distribution of discontinuous phase bubbles, thereby revealing the microscopic mechanisms and dynamic characteristics of gas–liquid two-phase flow (Jia et al., 2023; Li et al., 2024; Ouyang et al., 2024).

Discontinuous phase bubbles exhibit significant instability and complex dynamic behavior. Their morphology, size, and spatial distribution are influenced not only by the structure of the porous medium but also by fluid viscosity, surface tension, and interfacial dynamics. These factors significantly affect gas–liquid distribution, interphase mass transfer, and energy exchange (Liu et al., 2019b). Therefore, precise segmentation and analysis of discontinuous phase bubbles in microfluidic systems are crucial methods for clarifying microscopic flow mechanisms and optimizing gas injection strategies.

Recent advancements in deep learning-based algorithms have further enhanced the precision and efficiency of bubble tracking and segmentation. For instance, Fang et al. (2024) developed a framework combining YOLOv5 with DeepSORT for rapid tracking of gas liquid two-phase bubbly flow, achieving real-time monitoring with improved accuracy by incorporating semantic segmentation and multi-object tracking techniques. Similarly, Yang et al. (2025) proposed an enhanced segmentation method based on YOLOv8 for handling multi-scale and overlapping bubbles, using omni-dimensional dynamic convolution and high-resolution feature layers to address challenges under high void fraction conditions. These approaches demonstrate the potential of deep learning in overcoming the limitations of traditional methods for complex bubble dynamics analysis.

Nguyen et al. (2015) employed fluorescence microscopy to observe the mixing process of CO_2 bubbles with crude oil in microchannels, developing a platform for measuring the minimum miscible pressure parameter during CO_2 injection. Their study used the changes in fluorescence intensity to reflect variations in bubble morphology, interfacial tension, and miscibility. However, due to limitations in fluorescent dye concentration and coloration effects, the bubble edges were not sufficiently distinct. To address the shortcomings of the previous study, Tian et al. (2021) designed a two-dimensional pore model based on glass chips and utilized high-definition cameras to visualize the evolution of the gas–water interface and two-phase flow behavior at the pore scale. In their study, manual segmentation was employed to characterize discontinuous phase bubbles, and gas–water saturation was measured every 3 h, without achieving continuous dynamic observation of two-phase flow. Notably, Durve et al. (2022) applied a combination of YOLOv5 and DeepSORT for tracking microfluidic droplets, showing that hybrid datasets composed of synthetic and real images can reduce manual annotation efforts by 60% while maintaining high accuracy, offering a promising solution for large-scale bubble analysis in porous media.

Gas–liquid flow in porous media exhibits bubble characteristics that differ from those in conventional multiphase systems. In traditional multiphase flow systems, bubbles tend to exhibit a relatively uniform size distribution, with frequent overlapping observed. However, the challenge in our study does not lie in the overlap of bubbles, as the experiments are conducted in a quasi-two-dimensional microfluidic chip where the imaging plane limits vertical stacking. Instead, the challenge lies in the significant variation in bubble sizes and the complexity of texture patterns. Within a single experimental image, dispersed phase bubbles with approximately 1% saturation typically show near-spherical morphology, while bubbles forming a continuous phase can reach around 10% saturation and form larger structures spanning multiple pore throats. This substantial size disparity within the same image poses a considerable challenge for multi-scale segmentation. Furthermore, the bubble surfaces exhibit intricate grayscale textures due to light scattering at irregular pore interfaces, increasing the difficulty of accurate mask extraction.

With the rapid development of artificial intelligence technology, numerous key fields have utilized image segmentation techniques to process and analyze image data and achieved significant results (He et al., 2024; Sakshi and Kukreja, 2023). Research on image segmentation technology can be traced back to traditional methods based on edge detection, region growing, and threshold segmentation (Pavlidis and Liow, 1990). These methods perform segmentation based on changes in pixel intensity or similarity, but they perform poorly in complex backgrounds (Marmanis et al., 2018). With the advancement of CNN (convolutional neural network) and transformer architectures in the field of deep learning, variants based on these two models have emerged successively and have been widely applied in fields such as medicine, remote sensing, autonomous driving, industrial robotics, and oil and gas extraction (Lai-Dang et al., 2023; Pan et al., 2023; Zhang et al., 2022). The CNN architecture possesses advantages such as a compact structure and high computational efficiency in image feature extraction, enabling excellent segmentation and detection performance even in scenarios with limited computational resources. This is critically important for handling tasks that involve high-resolution image processing (Alzubaidi et al., 2021). The YOLO (you only look once) method, proposed by Redmon et al., is a real-time object detection algorithm that transforms the object detection task into a regression problem, requiring only a single forward pass from input image to detection results, thereby greatly enhancing detection speed (Redmon et al., 2016). To date, YOLO has been iteratively updated to version 11, achieving remarkable results in image segmentation in general domains (Jegham et al., 2024).

Considering the significant size variations, diverse morphologies, and uneven distribution of discontinuous phase bubbles in porous media, image segmentation networks face more stringent requirements. Concurrent with the ongoing advancements in image segmentation technology, researchers are also investigating data augmentation and automated annotation techniques to minimize dependence on manual labeling (Koeshidayatullah, 2023; You et al., 2024; Zhang et al., 2021). Data annotation is typically the initial step in supervised deep learning methodologies, and both the volume and quality of annotated data critically influence model training outcomes (Wang et al., 2021). The segment anything model (SAM), as an image segmentation model trained on the largest dataset to date, has exhibited exceptional performance in the field of image segmentation and has been widely adopted by numerous researchers for automated data annotation purposes (Chen et al., 2024; Luo et al., 2024; Osco et al., 2023). However, SAM has yet to be applied in the domain of oil and gas extraction.

To address the challenges of data annotation and the need for a high-precision segmentation model for discontinuous bubbles, this study proposes an image segmentation method for discontinuous phase bubbles based on the SAM. This method first employs the SAM2 model to achieve high-precision automatic segmentation and measurement of bubbles. Subsequently, we further develop a lightweight image segmentation model, MSDD-YOLO (multi space to depth down YOLO), to address the challenges of the large size and slow inference speed of the SAM2 model. By modifying the feature extraction module of the YOLOv11 image segmentation model to accommodate the segmentation of bubbles of varying sizes, the segmentation accuracy of the model is enhanced. Additionally, lightweight modifications are implemented in the detection head and backbone network, thereby achieving a balance between segmentation accuracy and inference speed.

2. Methodology

This study employs the SAM2 segmentation model and the lightweight MSDD-YOLO model to segment discontinuous phase bubbles in a microfluidic system. Gas–water flow videos are obtained through the microfluidic system, and after image preprocessing, the SAM2 segmentation model is utilized for automated bubble segmentation. The effectiveness of SAM2's automated bubble segmentation is evaluated using the intersection over union (IoU) metric. Considering the characteristics of discontinuous phase bubbles, we designed a lightweight MSDD-YOLO bubble segmentation model. First, a DWCC module is incorporated into the YOLOv11 backbone network, significantly reducing computational load and the number of parameters while maintaining segmentation performance. Next, we designed an MSDD downsampling module that employs multi-dimensional space-to-depth convolution to enhance the network's ability to extract multi-scale features and integrates the CBAM lightweight attention module for channel and spatial attention downsampling. Finally, the Pixels-IoU loss function replaces the original loss function in the detection head to more effectively assess the differences between predicted and target bounding boxes. The segmented bubbles then undergo quantitative analysis to obtain their area, shape factor, and positional information. The workflow is illustrated in Fig. 1.

2.1. Segment anything model

Segment anything model (SAM), a foundation model trained on the largest image segmentation dataset to date, has been widely applied in various computer vision applications since its introduction (Wang et al., 2024). The SAM's dataset comprises 11 million images and 1.1 billion mask images. It employs a strategy that combines an image encoder, a prompt encoder, and a mask decoder to achieve the ability to segment anything (Kirillov et al., 2023). This segmentation capability comes from the model's ability to detect changes in pixel brightness, texture, and image structure, which helps it distinguish between different objects. The vision transformer-based image encoder captures not only local image details but also the broader context, allowing SAM to accurately locate object edges where the visual characteristics change noticeably, such as between a bubble and the surrounding fluid or particles (Shi et al., 2025). The prompt encoder accepts various input types (such as points, boxes, text, and pre-existing masks), while the mask decoder processes inputs from both the image encoder and the prompt encoder to ultimately generate segmentation masks.

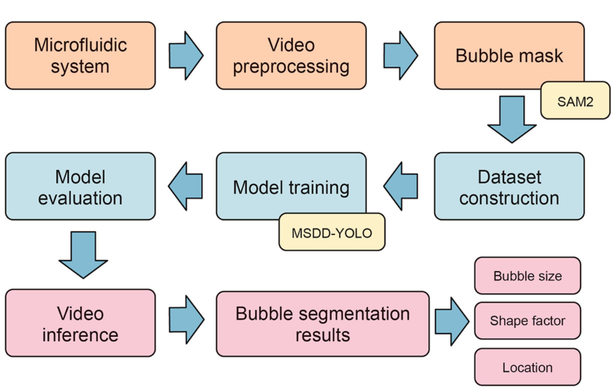

Fig. 1. Schematic workflow of the proposed methodology.

With the rapid growth of video stream data, a large portion of current image data is stored in video format. To address the limitations of the first-generation SAM, Ravi et al. (2024) developed the segment anything model 2 (SAM2). SAM2 employs the hierarchical image encoder, which enables the model to capture multi-scale features through a multi-level feature extraction mechanism, significantly improving computational efficiency and inference speed compared to traditional vision transformer (ViT) (Zhou et al., 2024). Additionally, it's streaming memory mechanism stores key information from the previous frame to provide contextual references for the segmentation tasks of subsequent frames. This mechanism not only enhances the consistency of video frame segmentation but also significantly reduces the model's reliance on independent frame features, demonstrating higher robustness and accuracy in handling dynamic scenes. The SAM2 network architecture is illustrated in Fig. 2.

Additionally, SAM2 has developed an interactive data engine that employs a collaborative strategy between humans and models to construct SA-V, the largest and most diverse video segmentation dataset to date, comprising 50.9 K videos and 642.6 K mask images. Compared to the first-generation model, SAM2 achieves faster processing speeds and higher accuracy with a smaller model size and fewer parameters. Specific parameters are presented in Table 1.

Beyond its architectural and dataset enhancements, SAM2 offers flexible integration into domain-specific segmentation workflows. A novel pipeline is introduced for automated bubble annotation in microfluidic images, integrating SAM2 with an interactive verification mechanism. The approach addresses the limitations of manual labeling by combining global automatic segmentation with user-guided selection:

(1) Global automatic segmentation. The process begins with SAM2 performing global segmentation on the input image. All visible objects, including target bubbles and non-target structures such as rock skeletons or particulate impurities, are detected and segmented. Each identified object is assigned a precise mask and its spatial coordinates.
(2) Interactive bubble mask selection and validation. A visualization interface is developed to facilitate bubble region selection. Users click on bubble areas in the displayed image. A coordinate mapping algorithm links the clicked position with the corresponding pre-segmented mask. The matched mask is immediately displayed for user confirmation and saved as part of the final training dataset. This verification process removes irrelevant or erroneous masks and ensures that only accurate bubble regions are retained.

The two-stage pipeline, combining global automatic segmentation with interactive bubble mask selection and validation, offers an effective solution for automating bubble annotation in microfluidic images. By leveraging SAM2's powerful segmentation capabilities and integrating user input for mask verification, the pipeline ensures both high segmentation accuracy and efficiency. The process not only accelerates the annotation workflow but also

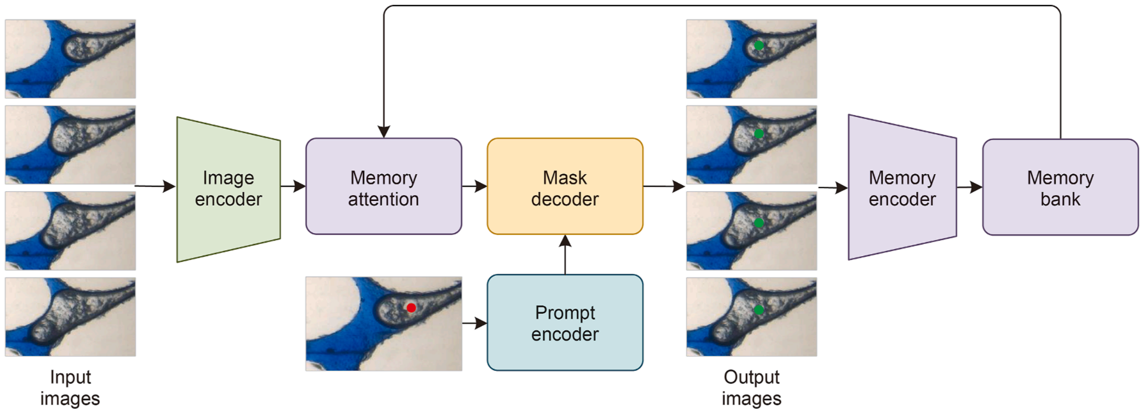

Fig. 2. Overview of the SAM2 network architecture.

Table 1
Comparison of model parameters and size for different versions of SAM and SAM2.

Name	Parameters, M	Model size
SAM-B	91	358 MB
SAM-L	308	1.2 GB
SAM-H	636	2.4 GB
SAM2-tiny	38.9	149 MB
SAM2-small	46	176 MB
SAM2-base+	80.8	309 MB
SAM2-large	224.4	856 MB

guarantees the precision of the final dataset. The implementation details, including the code for the two stages, are provided in Appendix A.

2.2. Bubble segmentation model MSDD-YOLO

In the field of image segmentation, the YOLO series models are widely applied to various computer vision tasks due to their high computational efficiency and real-time performance. YOLOv11, the latest version in this series, has undergone multiple optimizations and achieved excellent results in detecting objects of various scales and types (Alif, 2024). In microfluidic gas flooding experiment videos, YOLOv11 demonstrates a certain level of effectiveness in segmenting bubbles when the size variation is relatively small. However, its segmentation accuracy decreases when dealing with bubbles that exhibit significant size variations. Especially when the areas of bubbles differ by several tens of times within a single image, the fixed receptive field and uniform feature resolution of conventional YOLO architectures become insufficient. Such a large disparity in object size within the same image leads to challenges in accurately capturing and representing features across different scales. As noted by Chen et al. (2018b), significant image scale variation can substantially degrade segmentation performance due to the inconsistency in feature representation.

This work proposes the MSDD-YOLO high-precision bubble segmentation algorithm. Firstly, a DWCC (depthwise convolution with channel attention) module is integrated into the YOLOv11 backbone network to reduce computational load and model parameters while preserving performance. Given that shallow layers incur higher computational overhead when processing high-resolution images, the original bottleneck modules are replaced with lightweight DWCC modules. By utilizing depthwise convolutions in place of standard convolutions, the computational burden in the shallow layers is significantly reduced, without compromising performance. The channel attention mechanism further enhances the model's ability to focus on critical features, thereby ensuring minimal impact on accuracy. Secondly, the MSDD (multi space to depth down) module is introduced, which incorporates the lightweight CBAM (convolutional block attention module) to perform channel and spatial attention during the downsampling process. Additionally, multi-dimensional SPD (space-to-depth down) convolution is utilized to enhance the network's ability to extract multi-scale features, catering to the segmentation needs of bubbles of different sizes. In the detection head, the Pixels-IoU loss function replaces the original loss function to more effectively assess the differences between predicted and target bounding boxes. The MSDD-YOLO network architecture is illustrated in Fig. 3.

2.2.1. Lightweight structure DWCC

In the task of bubble segmentation within gas–water flooding experiments, the objective is to accurately identify and delineate the shapes and edges of bubbles. However, the morphological changes of bubbles in the fluid, complex backgrounds, varying lighting conditions, and blurred bubble boundaries pose significant challenges, making it difficult for the original YOLOv11 method to maintain stable performance. To enhance segmentation accuracy and speed, a network architecture capable of adaptively handling complex backgrounds and blurred boundaries is required to improve the model's generalization capability.

To address the aforementioned issues, this paper proposes the DWCC (depthwise convolution with channel attention) module. Compared to the original bottleneck structure of YOLOv11, the DWCC module replaces standard convolution with depthwise convolution (DWConv) in the convolutional layers, thereby reducing computational complexity and enhancing the network's inference speed. Furthermore, a channel attention mechanism is introduced to assign adaptive weights to features along the channel dimension, enabling the network to focus on bubble regions while suppressing interfering information in the background. The DWCC network architecture is illustrated in Fig. 4.

In the DWCC module, the input bubble features are first processed through a depthwise convolution layer to extract spatial features and learn the morphological characteristics of the bubbles. Subsequently, the channel attention mechanism adaptively enhances these features, enabling the model to focus more on the bubble regions while suppressing background interference. The channel attention mechanism analyzes information across different channels by separately extracting global information from each channel using global average pooling and max pooling.

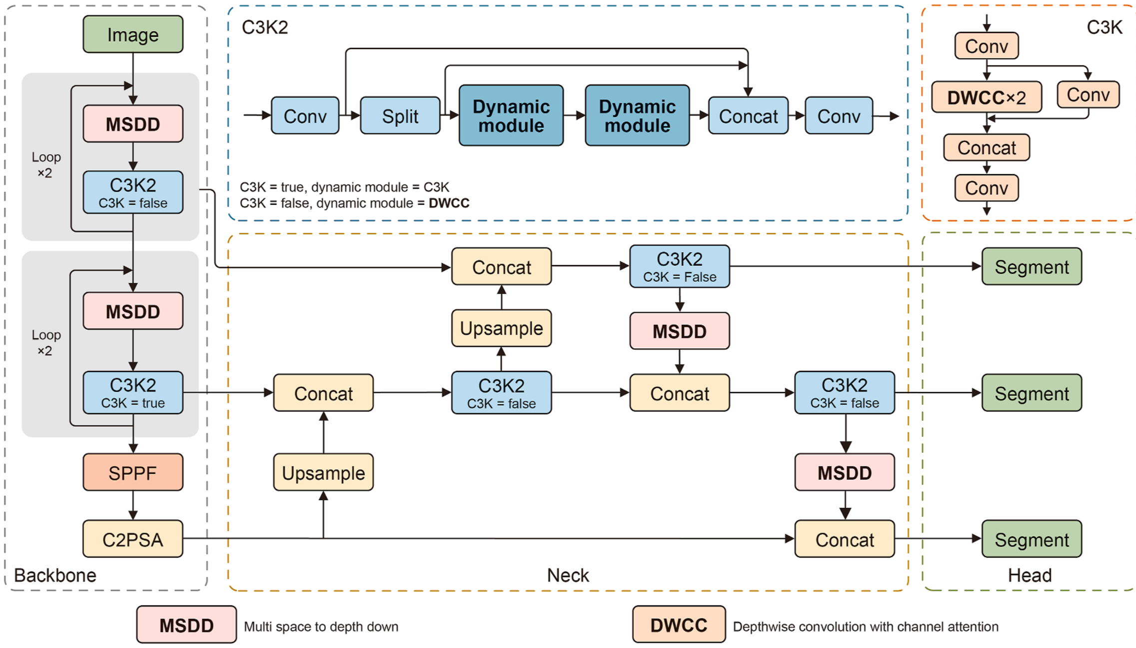

Fig. 3. The network architecture of the proposed MSDD-YOLO model.

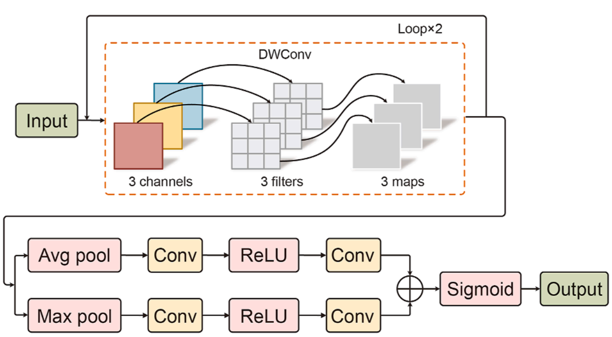

Fig. 4. Diagram of the DWCC module.

It then generates weights through sequential convolution layers, ReLU activation functions, and sigmoid functions, which are applied to each channel to adjust the feature responses. This module can adaptively enhance features related to the bubble regions while reducing background interference, thereby improving the model's segmentation accuracy in complex backgrounds.

2.2.2. Feature enhancement network MSDD

In the original YOLOv11 network, downsampling is implemented using standard convolutional layers, which include convolutional layers, batch normalization, and SiLU activation functions. After multiple downsampling operations, global semantic information is retained, but the extraction of features for small bubbles is insufficient. The downsampling module in YOLOv9 adopts a lightweight ADown branch structure (Wang C.-Y. et al., 2025). The feature map first undergoes average pooling, then is split into two parts along the channel dimension, forming a dual-branch structure, $x1$ and $x2$. The $x1$ branch performs downsampling through a 3 × 3 convolution operation to extract local spatial features; the $x2$ branch uses 3 × 3 max pooling and 1 × 1 convolution to further retain global feature information and adjust the number of channels to suit subsequent network layers.

Inspired by this strategy, this paper proposes the MSDD downsampling module. The MSDD employs a multi-branch convolution strategy to enhance feature extraction by decomposing the input feature map into four sub-feature maps along the channel dimension and applying the SPD-Conv downsampling method to each sub-feature map (Sunkara and Luo, 2022). The SPD-Conv consists of a space-to-depth layer and a non-strided convolution layer. The SPD layer partitions the feature map into blocks based on predefined rules, resulting in several sub-feature maps, each containing pixels from specific spatial locations. The blocking rules are defined as follows. Given an intermediate feature map $\boldsymbol{X}$, its structure is illustrated below:

$$\{S,\ S,\ C\} \tag{1}$$

The sub-feature maps are computed according to the following equations:

$$\begin{aligned}\boldsymbol{f}_{0,0} &= \boldsymbol{X}[0:S:stride,\ 0:S:stride],\ \boldsymbol{f}_{1,0}\\ &= \boldsymbol{X}[1:S:stride,\ 0:S:stride],\ \ldots,\ \boldsymbol{f}_{stride-1,0}\\ &= \boldsymbol{X}[stride-1:S:stride,\ 0:S:stride];\end{aligned} \tag{2}$$

$$\begin{aligned}\boldsymbol{f}_{0,1} &= \boldsymbol{X}[0:S:stride,\ 1:S:stride],\ \boldsymbol{f}_{1,1},\ \ldots,\ \boldsymbol{f}_{stride-1,1}\\ &= \boldsymbol{X}[stride-1:S:stride,\ 1:S:stride];\end{aligned} \tag{3}$$

$$\vdots$$

$$\begin{aligned}\boldsymbol{f}_{0,stride-1} &= \boldsymbol{X}[0:S:stride,\ stride-1:S:stride],\\ &\quad \boldsymbol{f}_{1,stride-1},\ \ldots,\ \boldsymbol{f}_{stride-1,stride-1}\\ &= \boldsymbol{X}[stride-1:S:stride,\ stride-1:S:stride]\end{aligned} \tag{4}$$

Under the constraints of the aforementioned equations, given any feature map $\boldsymbol{X}$, the sub-feature maps $f_{x,y}$, where x, $y \in \{0,\ 1,\ \ldots,\ stride-1\}$, satisfy the following:

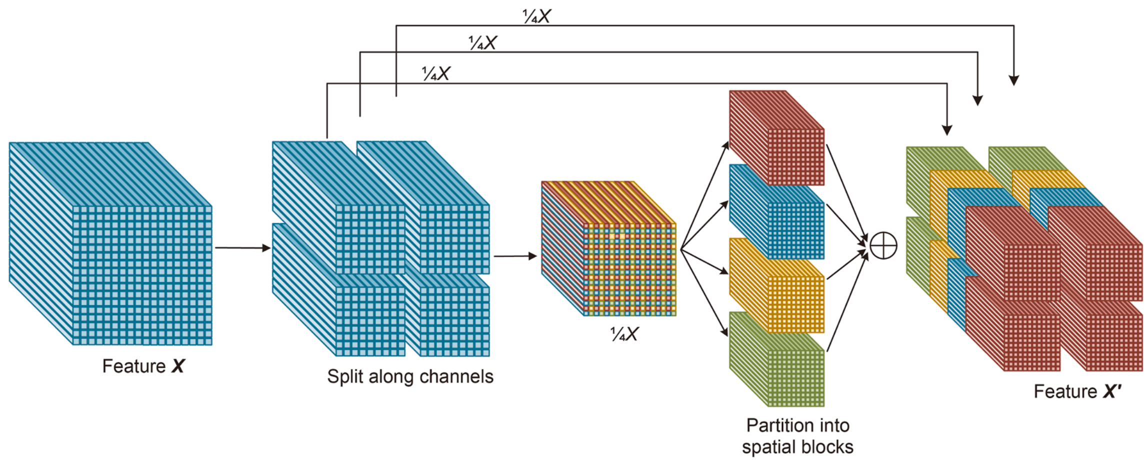

Fig. 5. The MSDD downsampling process.

$$\begin{cases} (i+x) \bmod stride = 0 \\ (j+y) \bmod stride = 0 \end{cases} \tag{5}$$

Through the feature map block downsampling operation, each sub-feature map $f_{x,y}$ has the shape:

$$\left\{\frac{S}{stride}, \frac{S}{stride}, C\right\} \tag{6}$$

Finally, the sub-feature maps are concatenated along the channel dimension to form a new feature map, which has the shape:

$$\left\{\frac{S}{stride}, \frac{S}{stride}, stride^2 C\right\} \tag{7}$$

Fig. 5 illustrates the MSDD downsampling process with a stride of 2.

In the downsampling process, the MSDD method first performs adaptive weighting of the channel and spatial dimensions through the CBAM module, thereby adaptively enhancing features in different regions. The resulting feature map is then decomposed into four sub-feature maps along the channel dimension. Subsequently, the sub-feature maps undergo SPD operations, which partition the feature map spatially and expand it into the channel dimension to increase the local representation capacity of the feature map. Compared to traditional downsampling methods, the MSDD module retains more spatial information and avoids excessive information loss, ensuring that features of smaller bubble targets are not overlooked during the downsampling process. We conducted ablation studies comparing the dual-branch ADown and four-branch MSDD structures. The results are summarized in Appendix B.

2.2.3. Pixels-IoU loss function

Considering the variable shapes of bubbles in porous media and the wide range of shape factors (width/height), relying on the loss function that was solely based on the overlap between standard bounding box makes it difficult to achieve rapid model convergence. Therefore, this paper introduces the pixel-level loss function Pixels-IoU (PIoU) (Chen et al., 2020). Compared to the original IoU, PIoU employs oriented bounding boxes (OBB) with rotational parameters, adding angular information to measure the degree of overlap between the ground truth and predicted boxes. Additionally, it precisely calculates the overlap area at the pixel level, thereby more effectively evaluating the differences between the predicted and ground truth boxes. The calculation method of PIoU is shown in Fig. 6.

As shown in Fig. 6, the detection box b is defined as follows:

$$(c_x,\ c_y,\ w,\ h,\ \theta) \tag{8}$$

where $\{c_x, c_y\}$ represents the coordinates of the center point c of the detection box; w and h denote the width and height; and θ indicates the inclination angle of the detection box. The green point p is the pixel to be calculated, and $t_{i,j}$ is the intersection point of the perpendicular from point p along the perpendicular line of the detection box. Points c, p, and $t_{i,j}$ form a right-angled triangle S; β is the angle between the triangle and the perpendicular bisector. The lengths of the sides of the triangle are $d^w_{i,j}$, $d^h_{i,j}$, and $d_{i,j}$. The calculation method is as follows:

$$d_{i,j} = \sqrt{(c_x - i)^2 + (c_y - j)^2} \tag{9}$$

$$d^w_{i,j} = |d_{i,j}\cos\beta| \tag{10}$$

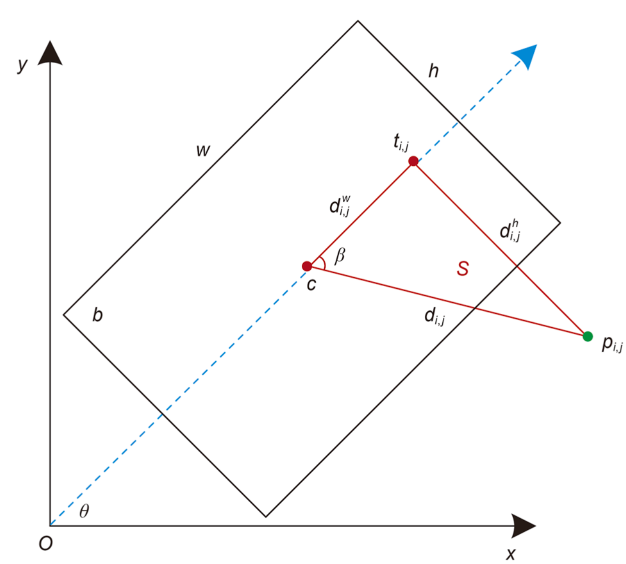

Fig. 6. Schematic diagram illustrating the calculation components of PIoU.

$$d_{i,j}^{h} = |d_{i,j}\sin\beta| \tag{11}$$

$$\beta = \begin{cases} \theta + \arccos\dfrac{c_x - i}{d_{i,j}}, & c_y - j \geq 0 \\ \theta - \arccos\dfrac{c_x - i}{d_{i,j}}, & c_y - j < 0 \end{cases} \tag{12}$$

The core of PIoU lies in performing calculations for each pixel individually, determining the distance from the pixel to the center of the bounding box, and calculating its contribution based on its position using the following kernel function:

$$F\left(p_{i,j}\middle|b\right) = K\left(d_{i,j}^{w},\ w\right)\cdot K\left(d_{i,j}^{h},\ h\right) \tag{13}$$

where $K(d, s)$ is the coefficient that adjusts the distance sensitivity, and its calculation is as follows:

$$K(d,\ s) = \frac{1}{1 + \mathrm{e}^{-k(d-s)}} \tag{14}$$

Define the detection box and target box as b and b'. The intersection area $S_{b\cap b'}$ is defined as the total contribution of the pixels that belong to both the bounding boxes b and b'. The formula is as follows:

$$S_{b\cap b'} = \sum_{p_{i,j}\in B_{b,b'}} F\left(p_{i,j}\middle|b\right)\cdot F\left(p_{i,j}\middle|b'\right) \tag{15}$$

where $B_{b,b'}$ is the smallest horizontal bounding box that contains both the bounding boxes b and b', used to define the statistical range of the pixels. The union area $S_{b\cup b'}$ is defined as the total contribution of the pixels that belong to either bounding box b or b'. The formula is as follows:

$$S_{b\cup b'} = \sum_{p_{i,j}\in B_{b,b'}} \left[F\left(p_{i,j}\middle|b\right) + F\left(p_{i,j}\middle|b'\right) - F\left(p_{i,j}\middle|b\right)\cdot F\left(p_{i,j}\middle|b'\right)\right] \tag{16}$$

This formula avoids redundant calculations in the intersection area by weighting the pixel contributions. To reduce computational complexity, the union area $S_{b\cup b'}$ can be simplified as follows:

$$S_{b\cup b'} = w\cdot h + w'\cdot h' - S_{b\cap b'} \tag{17}$$

where w, h and w', h' represent the width and height of bounding boxes b and b' respectively. Based on the definitions of the intersection and union areas, the Pixels-IoU formula for the two bounding boxes is as follows:

$$\mathrm{PIoU}(b,\ b') = \frac{S_{b\cap b'}}{S_{b\cup b'}} \tag{18}$$

The objective of the loss function is to make the value of PIoU as close to 1 as possible, thereby maximizing the overlapping area between the predicted and ground truth boxes. Its calculation formula is as follows:

$$L_{\mathrm{PIoU}} = -\frac{1}{|M|}\sum_{(b,b')\in M} \ln(\mathrm{PIoU}(b,\ b')) \tag{19}$$

where M represents the set of all positive sample pairs, which includes predicted boxes b and ground truth boxes b' whose IoU is greater than the threshold of 0.5; $|M|$ represents the number of positive sample pairs. By maximizing PIoU, the model can more accurately predict bounding boxes, especially when dealing with high aspect ratios, dense targets, and complex backgrounds.

2.3. Data acquisition

The dataset used is derived from gas–water flow experiments conducted using the microfluidic system, as illustrated in Fig. 7. The system consists of the microfluidic chip, a gas generator, a peristaltic pump, a control device, a high-definition camera, and a reflux device. Due to the slow nature of the gas–liquid flow process, images were extracted at 5-s intervals, resulting in a total of 50 raw experimental images with a resolution of 3264 × 2448. After removing artifacts within the camera's field of view, the image size was reduced to 2810 × 1850.

2.4. Model performance evaluation

This study uses IoU (intersection over union) to validate the segmentation effectiveness of the SAM2 model. For the MSDD-YOLO model, the evaluation metrics include the number of floating-point operations (FLOPs) and model parameters to assess the model's lightweight performance. The segmentation performance of the model is evaluated using precision, recall, and mean average precision (mAP), with the following calculation formulas:

$$\mathrm{IoU} = \frac{area(C)\cap area(G)}{area(C)\cup area(G)} \tag{20}$$

$$P = \frac{TP}{TP + FP} \tag{21}$$

$$R = \frac{TP}{TP + FN} \tag{22}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{23}$$

where $area(C)$ represents the area of the bubble region segmented by the SAM2 model, and $area(G)$ represents the area of the bubble region manually annotated; TP (true positive) denotes the number of pixels correctly predicted as bubble regions, while FP (false positive) represents the number of pixels incorrectly predicted as bubble regions; FN (false negative) refers to the number of pixels in the bubble regions that were missed; AP is the average precision, and n represents the number of classes.

2.5. Model training environment

This study used the Pytorch 2.0.1 deep learning framework to build the SAM2 and MSDD-YOLO networks. The operating system was Ubuntu 20.04 with 32 GB of system memory, and the GPU

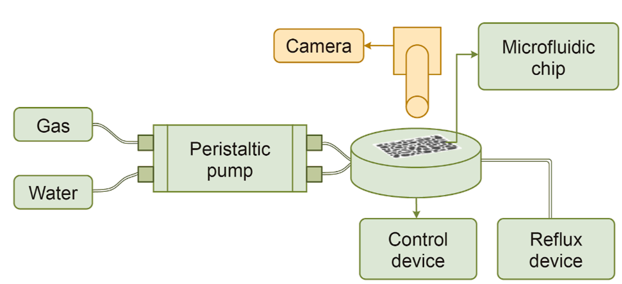

Fig. 7. Schematic diagram of the microfluidic system.

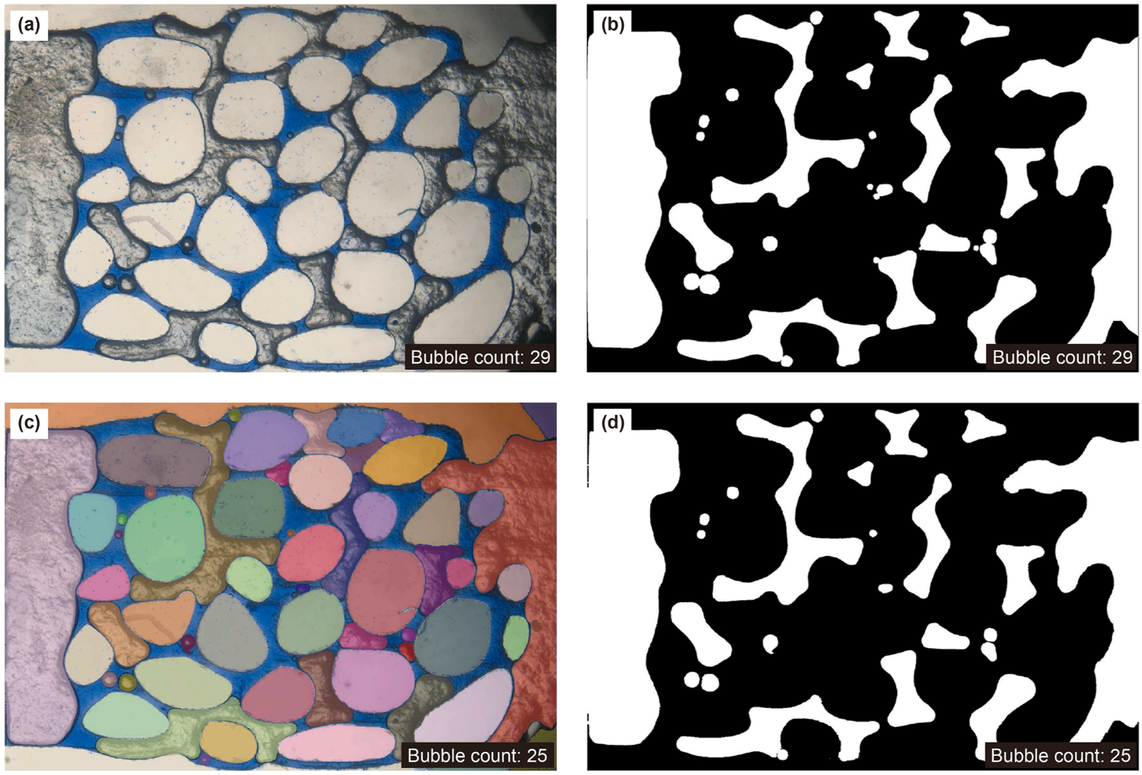

Fig. 8. Comparison of SAM2 segmentation and manual segmentation. (**a**) Original image; (**b**) manually annotated bubble image; (**c**) all target masks segmented by SAM2; (**d**) filtered bubble mask image from (c).

device was an NVIDIA GeForce RTX 4090 24 GB graphics card, running CUDA version 11.8. The SAM2 model utilized default hyperparameters. The training process for the MSDD-YOLO model employed a cosine annealing strategy and the SGD (stochastic gradient descent) optimization method, with a learning rate set to 0.01. The batch size and number of epochs were set to 16 and 1500, respectively, with an input image size of 640 × 640. The progressive training strategy involves a three-stage training process, with input image sizes of 160, 320, and 640, each trained for 500 epochs. The dataset was split with an 8:2 ratio, resulting in the training set consisting of 40 images and the test set containing 10 images.

3. Results and discussion

3.1. SAM2 segmentation results

To evaluate the performance of the SAM2 method in the bubble segmentation task, this study selected 10 manually annotated experimental bubble images as references for analyzing the SAM2 segmentation results. After the images are input into the SAM2 model, the model covers all targets (including bubbles and skeletons) in the image with masks of different colors. These masks are then manually filtered to obtain the bubble mask images. Fig. 8 shows the segmentation results of SAM2 for bubbles across the entire field of view. Fig. 9 displays the segmentation results for localized bubbles, including the original bubble image, the manually segmented bubble mask, and the SAM2 segmentation result. SAM2 can accurately segment most of the bubble contours, but some smaller bubbles remain unsegmented. This is due to the indistinct boundaries and the similarity in color between the bubbles and the background, which prevents their capture by the SAM2 model.

To quantify the segmentation performance, the IoU value for each image was calculated, as well as the ratio of the count of bubbles segmented by SAM2 to those manually segmented. Fig. 10 shows the comparison between the bubble counts obtained from manual segmentation and those detected by SAM2. Fig. 11 shows the distribution of IoU values for the test images (The corresponding experimental images used for Figs. 10 and 11 are provided in Appendix C). The calculated IoU values demonstrate a strong correspondence between the SAM2 segmentation results and the manual annotations, indicating that the SAM2 model is highly effective in capturing the true contours of the bubbles.

To evaluate the efficiency of the SAM2 model in the bubble segmentation task, we compared the time consumption of the SAM2 automatic segmentation method with the traditional manual segmentation method. The experimental results are shown in Fig. 12. Specifically, after receiving 10 images, the SAM2 model can complete the segmentation of all targets in the images and generate the corresponding mask information within 2 min. Subsequently, manually selecting the bubble masks takes approximately 10 min, resulting in a total time of about 12 min. The manual segmentation time was measured using the LabelMe annotation tool, where annotators delineated bubble contours by manually clicking along their boundaries in each of the 10 images, requiring approximately 350 min in total. This process is significantly more time-consuming compared to the SAM2-assisted method. To illustrate the manual workflow and its complexity, representative examples are provided in Appendix D. Compared to the manual method, the SAM2 model is approximately 29 times more efficient.

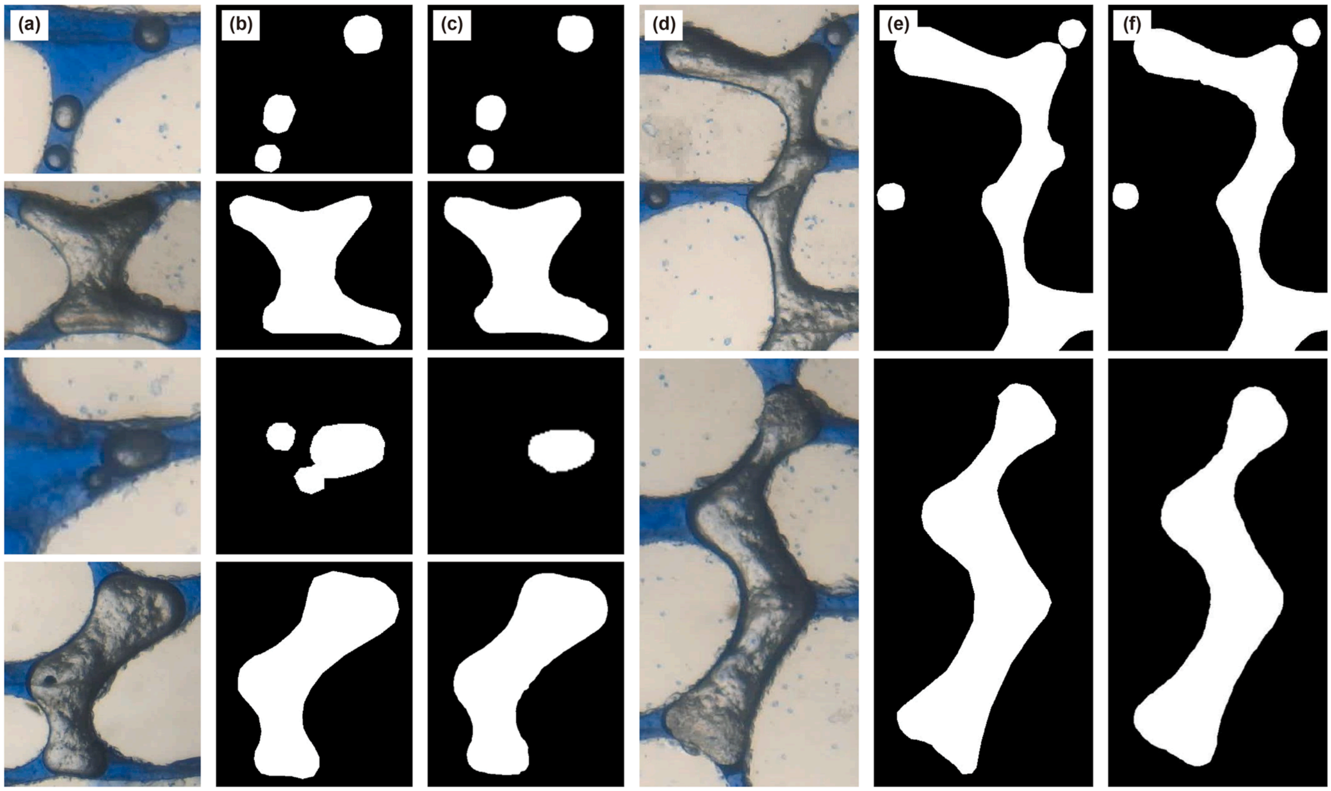

Fig. 9. Local comparison of SAM2 segmentation and manual segmentation. (**a**), (**d**) Original bubble images; (**b**), (**e**) manual segmentation results; (**c**), (**f**) SAM2 segmentation results.

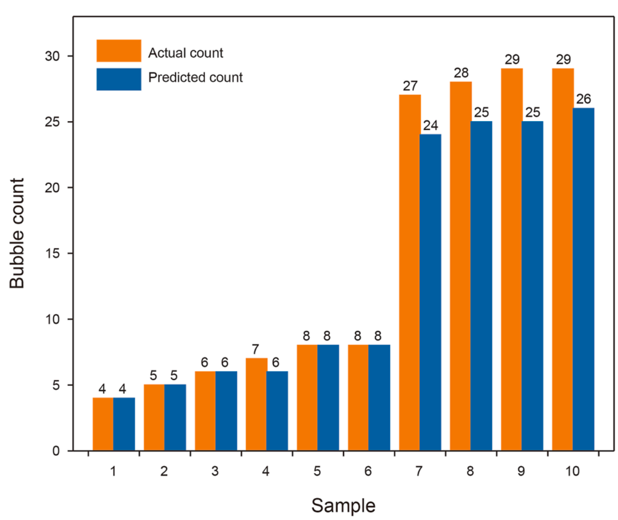

Fig. 10. The number of bubbles from manual segmentation and the number predicted by the SAM2 model.

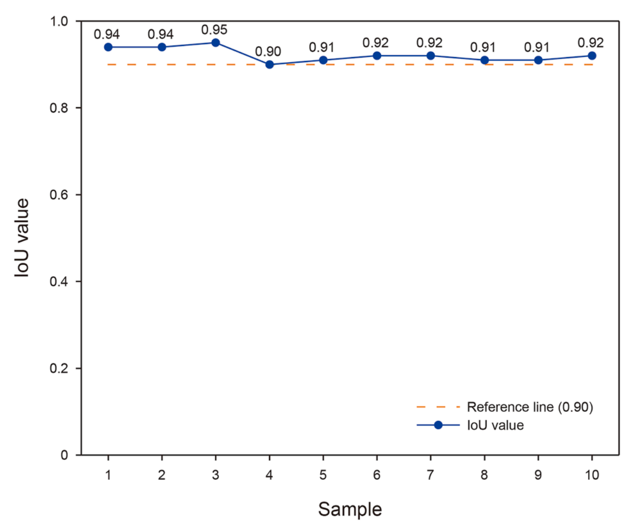

Fig. 11. Distribution of IoU values.

This study also validated the quantitative analysis capability of the SAM2 model in bubble segmentation. We analyzed segmentation accuracy by comparing the bubble areas segmented by the SAM2 model with those obtained through manual annotation. Fig. 13 illustrates the area calculation comparison between SAM2 segmentation and manual segmentation for some bubble images, with the quantitative results shown in Table 2. The analysis is based on pixel count to assess the area differences between the two methods. The experimental results indicate that the discrepancy between the SAM2 model's segmentation and manual segmentation is less than 2%, demonstrating a high degree of consistency in segmentation accuracy. This demonstrates the reliability of the SAM2 model for assisting bubble annotation tasks.

3.2. Ablation study

To validate the effectiveness of the proposed DWCC method, MSDD method, PIoU loss function, and the progressive training strategy, this study designed ablation experiments to analyze the

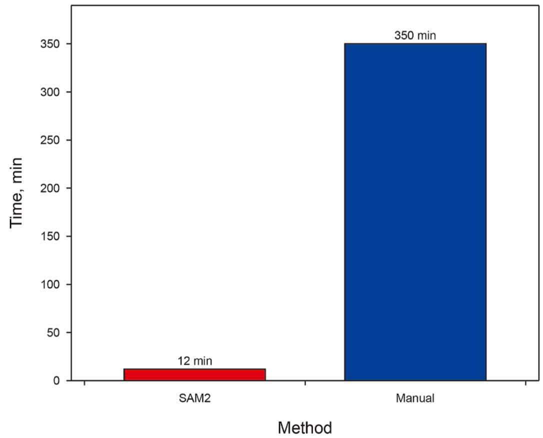

Fig. 12. Comparison of segmentation time.

impact of each module and their coupling performance. The experimental results are shown in Table 3. After introducing the DWCC and MSDD modules, the model's parameters and computational complexity decreased, while precision improved. However, there was a slight decrease in the recall rate. Replacing the loss function with PIoU improved the model's mAP and recall without adding computational cost.

Notably, the progressive training strategy substantially enhanced model performance, increasing mAP@0.5 from 0.77 to 0.93 and recall from 0.65 to 0.87. This strategy adopts a gradually increasing input resolution, allowing the model to initially capture coarse features of large-scale bubbles and subsequently refine its representation of smaller-scale structures. Based on the microfluidic experimental images used in this study, where bubble sizes vary significantly from very small to relatively large, the proposed coarse-to-fine training paradigm effectively facilitates multi-scale feature learning without introducing additional computational cost (Zhu et al., 2023).

Finally, by combining all the improvements, the MSDD-YOLO model outperforms the original YOLOv11-n algorithm, with bubble segmentation, precision increasing by 3%, mAP@0.5 rising by 15%, and recall improving by 18%. Meanwhile, the parameters and FLOPs are reduced by 25% and approximately 18%, respectively. The ablation study results demonstrate that the modules proposed in this paper complement each other, contributing to the overall improvement of the model's performance.

Table 2
Quantitative results of the SAM2 method and manual method.

Sample	Pixel count		Discrepancy, %
	SAM2	Manual	
Sample 1	55238	54856	0.6
Sample 2	47429	46782	1.4
Sample 3	20268	19936	1.7
Sample 4	11023	10859	1.5

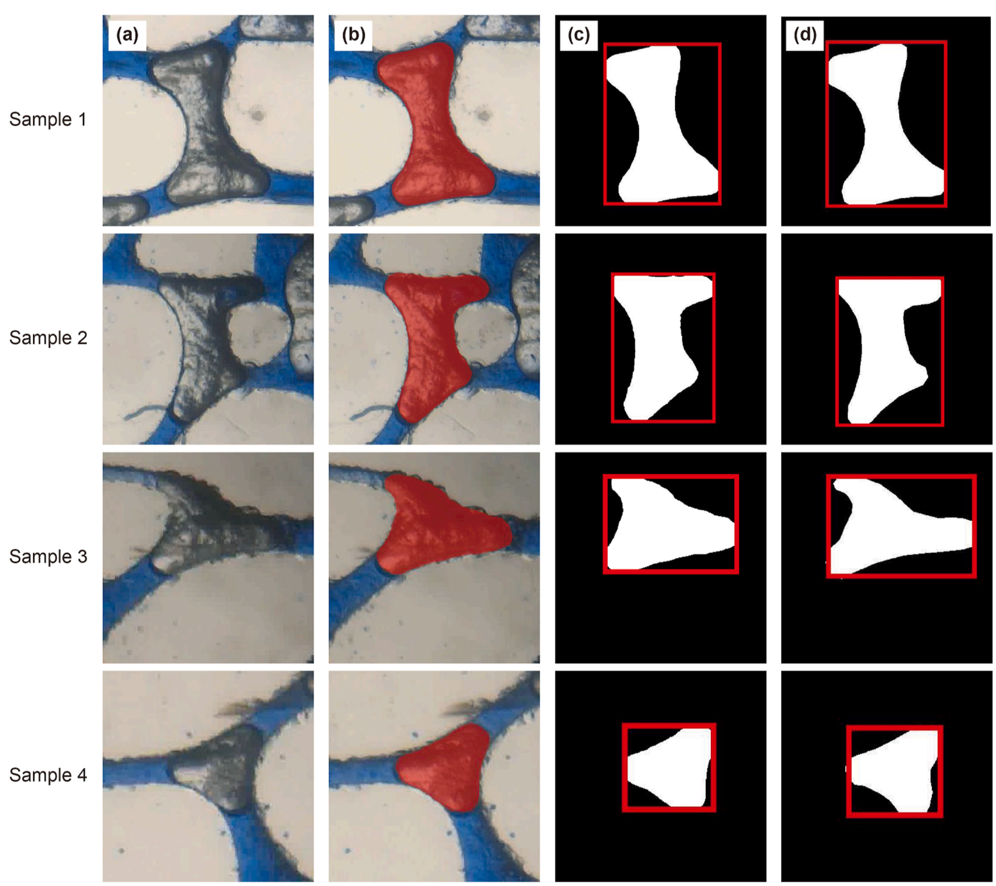

Fig. 13. Diagram of SAM2 quantitative analysis. (**a**) Original bubble image; (**b**) mask automatically generated by SAM2; (**c**) SAM2 segmentation result; (**d**) manual segmentation result.

Table 3
Ablation study of model efficiency and segmentation performance. ↑ indicates that a higher value is better for segmentation accuracy. ↓ indicates that a lower value is better for model efficiency. The best results for each metric are highlighted in bold.

No.	YOLOv11	DWCC	MSDD	PIoU loss	Progressive training	Parameters, M↓	FLOPs, G↓	Mask metrics			
								mAP@0.5↑	mAP@0.50:0.95↑	Precision↑	Recall↑
1	√					2.8	10.2	0.79	0.56	0.89	0.72
2	√	√				2.4	9.3	0.80	0.53	0.91	0.67
3	√		√			2.4	9.0	0.80	0.57	0.92	0.73
4	√	√	√			2.1	8.4	0.77	0.52	0.92	0.65
5	√	√	√	√		2.1	8.4	0.80	0.55	0.91	0.71
6	√	√	√		√	2.1	8.4	0.93	0.63	0.90	0.87
7	√	√	√	√	√	**2.1**	**8.4**	**0.94**	**0.66**	**0.92**	**0.90**

3.3. Comparative experiment

To validate the segmentation performance of the MSDD-YOLO algorithm, this study compared it with other image segmentation algorithms, including DeepLabv3+ (Chen et al., 2018a), YOLOv8 (Jocher et al., 2023), improved-YOLOv8 (Shi et al., 2024), Mamba-YOLO (Wang Z. et al., 2025), and YOLOv11-n (Jocher and Qiu, 2024), under the same experimental conditions. The proposed method was also compared with other gas–liquid image segmentation approaches (Durve et al., 2022; Fang et al., 2024). A total of 20 images (containing 460 bubbles) were selected for testing. As shown in Table 4, the proposed algorithm achieved the highest IoU of 0.91, while also attaining the lowest parameter count (2.1 M) and model size (4.6 MB) among all compared methods, with a highly competitive inference time. Fig. 14 shows the local bubble segmentation results for each algorithm. By analyzing the segmentation results of different models, it can be seen that the proposed model produces results with shapes that more closely resemble the actual bubble shapes, indicating that the MSDD-YOLO method is more effective in accurately extracting bubble edge features, thereby achieving higher precision in bubble segmentation tasks.

3.4. MSDD-YOLO segmentation results

Fig. 15 demonstrates the segmentation results of the proposed method on bubble images, where the bubble regions are covered with red masks, and each bubble is annotated with a rectangular

Table 4
Comparison of results of different models.

Method	MIoU↑	Parameters, M↓	Inference time, ms↓	Model size, MB↓
DeepLabv3+	0.87	5.8	17.0	22.4
YOLOv8	0.85	2.9	7.4	6.1
Improved-YOLOv8	0.89	2.7	6.5	6.7
Mamba-YOLO	0.81	5.9	11.2	11.8
YOLOv11	0.82	2.8	4.7	7.8
Durve et al. (2022)	0.73	7.4	6.1	15.2
Fang et al. (2024)	0.63	7.0	5.9	14.4
Ours	**0.91**	**2.1**	**4.1**	**4.6↓**

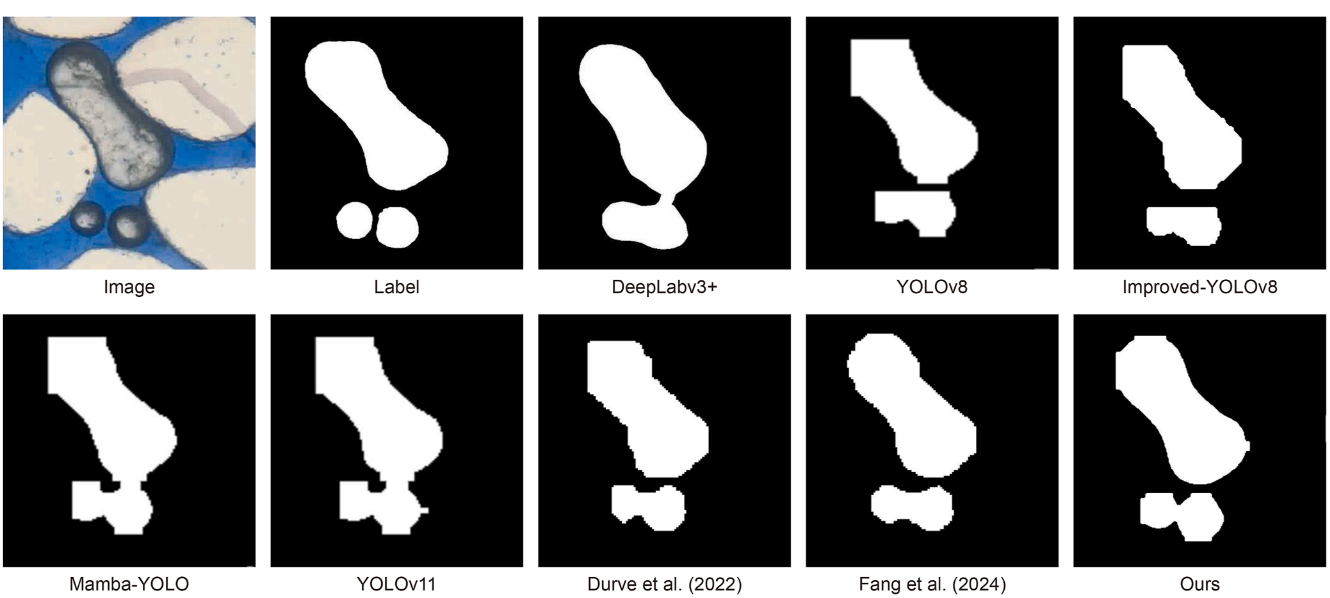

Fig. 14. Comparison of local bubble segmentation.

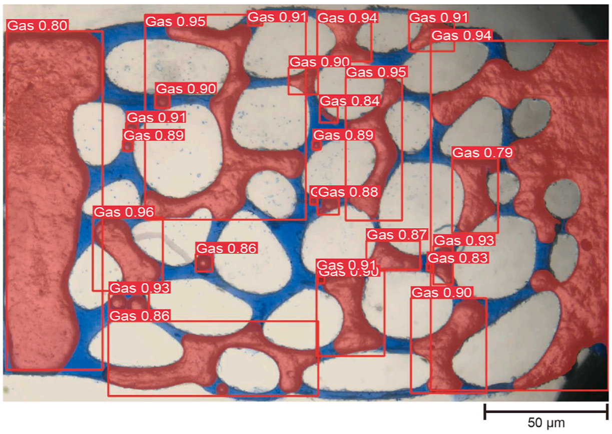

Fig. 15. Segmentation results of the proposed method.

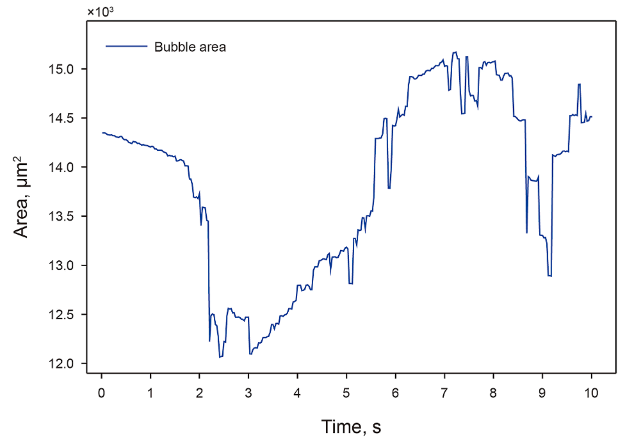

Fig. 17. Variation of bubble area over time.

box. The value in the top-left corner represents the bubble's confidence score. Visually, the proposed method accurately segments the contours of the bubbles. To quantitatively characterize the bubble size, shape factor, and spatial location, Fig. 16 presents a schematic of the quantitative analysis for selecting bubble images, with the results shown in Table 5. Additionally, a 10-s video segment was analyzed, and the results are shown in Fig. 17, which displays the curve of total bubble area (the sum of areas of all detected bubbles in each frame) over time, clearly illustrating the dynamic variation in total bubble area. After obtaining the bubble masks, the proposed method can calculate the bubble area, shape factor, and the central position of the bubbles, providing tools for the visualization and quantitative analysis of discontinuous phase bubbles.

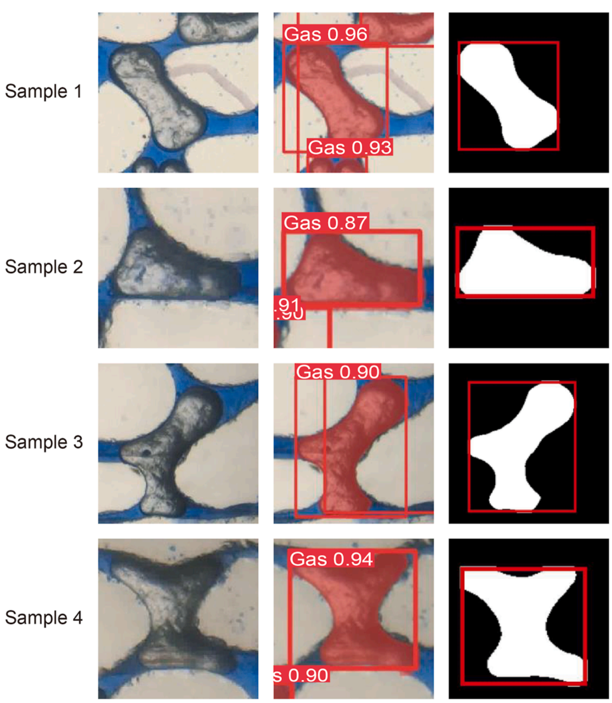

Fig. 16. Schematic diagram of bubble quantification analysis.

Table 5
Results of bubble quantification analysis.

Sample	Area, μm^2	Shape factor	Center location
1	478	0.93	(576, 1160)
2	215	1.95	(1813, 1169)
3	597	0.83	(2072, 1576)
4	295	1.07	(1589, 175)

4. Conclusions

This study successfully addressed the challenge of discontinuous phase bubble segmentation by developing a novel framework centered on the segment anything model. Additionally, leveraging the segment-anything capability of the SAM2 model, this method assists in data annotation and reduces the workload of manual labeling. The main contributions are summarized as follows:

(1) We introduced and validated the SAM2 model for segmenting discontinuous phase bubbles, achieving a segmentation discrepancy of less than 2% compared to manual annotation.
(2) By combining SAM2 segmentation with manual selection strategies, the method assists in bubble data annotation, improving annotation efficiency by approximately 29 times while reducing the associated manual labor cost.
(3) We developed the MSDD-YOLO bubble segmentation model, which integrates the DWCC and MSDD modules into the YOLOv11 network. This approach significantly reduces the model's computational load and parameters, achieving a mAP@0.5 of 94% and a recall rate of 90%. The model offers researchers visualization and quantitative tools for analyzing discontinuous phase bubbles.

Validated through gas–liquid displacement experiments in microfluidic pore networks, this methodology demonstrates transformative potential for CO_2-EOR efficiency evaluation, foam flooding optimization, and tight reservoir gas–water distribution analysis. The proposed method enables high-throughput analysis of multiphase flow dynamics, offering a transformative approach for studying fundamental flow physics and enabling field-scale digital twin construction.

Despite its advantages, the model has some limitations, such as high memory consumption during high-resolution video processing and inefficient video stream input/output performance. Future studies will explore the potential of this methodology in different application scenarios and optimize the model performance to adapt to a wider range of discontinuous phase image segmentation needs.

CRediT authorship contribution statement

Jia-Bin Shi: Writing – original draft, Visualization, Methodology, Investigation. **Li-Jun You:** Supervision, Resources, Funding acquisition. **Jia-Chen Dang:** Writing – review & editing, Supervision, Methodology. **Yi-Jun Wang:** Software, Funding acquisition, Formal analysis, Data curation. **Wei Gong:** Software, Investigation. **Bo Peng:** Writing – review & editing, Supervision, Software, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Sichuan Province Outstanding Young Scientist Fund (Grant No. 2025NSFJQ0009), Sichuan Regional Innovation Cooperation Fund (Grant No. 2025YFHZ0270), and the numerical calculations in this paper have been done on the supercomputing system in the High Performance Computing Center of Southwest Petroleum University. We would like to give thanks to the editors and anonymous reviewers for their suggestions on the manuscript improvement.

Appendix A

The code for the two-stage pipeline.

The code for the two-stage pipeline, including the visualization interface and coordinate mapping algorithm, is available on GitHub: https://github.com/shidizai-swpu/SAM2-MSDD-YOLO.

Appendix B

Ablation study of MSDD branch decomposition.

To justify the design choice of decomposing the input feature map into four sub-feature maps in the MSDD module, we conducted an ablation study comparing different configurations. All experiments were performed under identical training settings and on the same dataset to ensure fairness. The results are summarized in Table B1.

Table B1
Performance comparison of different MSDD branch decomposition strategies.

Model	Branch number	mAP@0.5	mAP@0.5:0.95	Parameters, M
YOLOv11	–	0.733	0.521	2.83
MSDD-2B	2	0.721	0.475	2.16
MSDD-4B (ours)	4	**0.791↑**	**0.542↑**	**2.11↓**

The best results for each metric are highlighted in bold.
↑ indicates that a higher value is better for segmentation accuracy. ↓ indicates that a lower value is better for model efficiency.

Appendix C

The images of the SAM2 model segmentation and manual segmentation.

Column 1 of Fig. C1 illustrates the original bubble images, Column 2 displays the global segmentation results of SAM2, Column 3 presents the SAM2-segmented bubble masks, and Column 4 shows the manually segmented bubble masks.

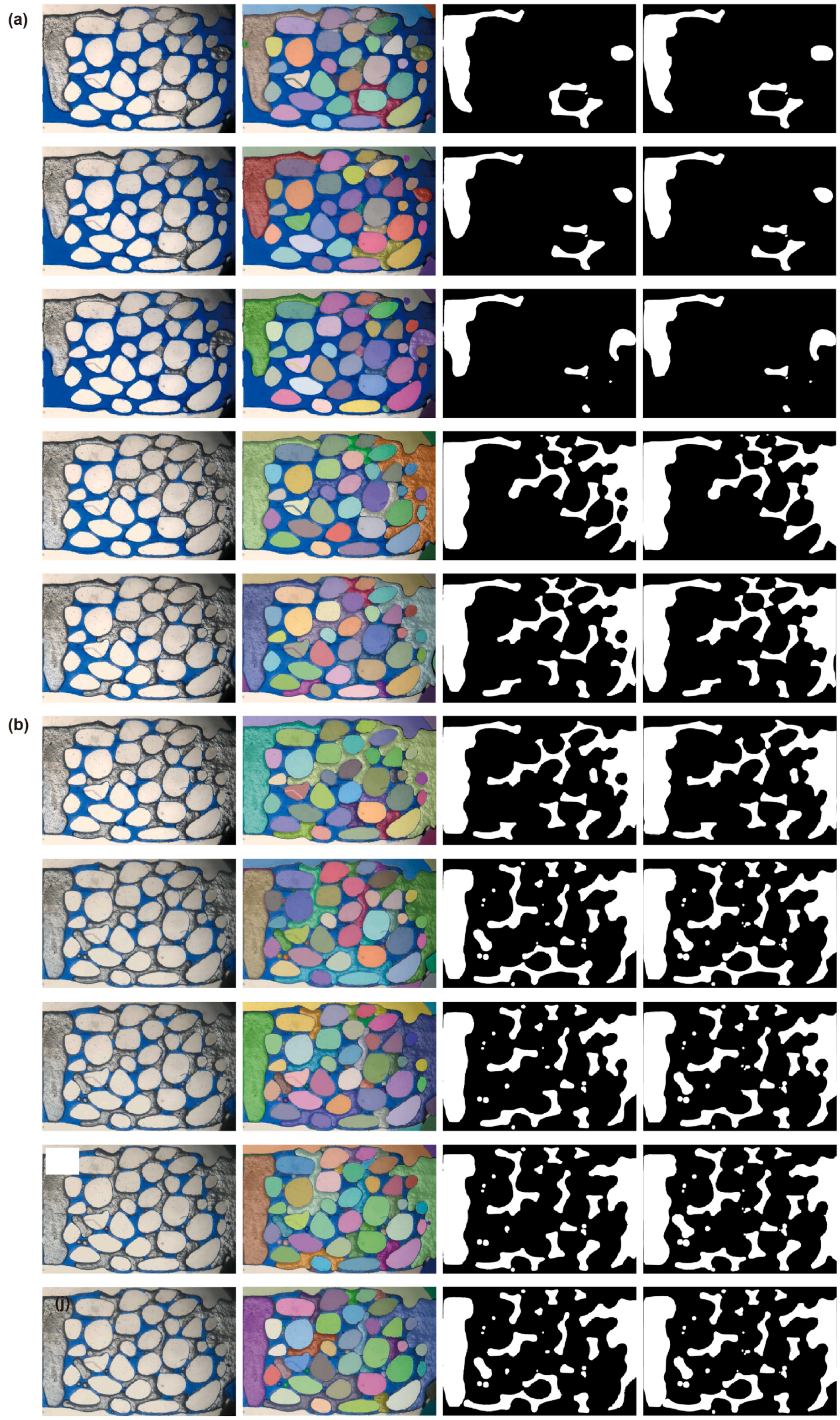

Fig. C1. Experimental images used for Figs. 10 and 11.

Appendix D

Manual segmentation process.

To provide a visual illustration of the traditional manual segmentation workflow, Fig. D1 presents representative examples of the annotation process used in this study. The displayed regions show local areas of the original images. In this method, annotators manually clicked along the edge of each bubble to form polygonal contours, demonstrating the manual segmentation process. Specifically, Fig. D1(a) and (e) shows the original bubble images; Fig. D1(b) and (f) presents the manual annotation process, where red points indicate the manually placed contour points; Fig. D1(c) and (g) displays the final segmentation results obtained through manual annotation; and Fig. D1(d) and (h) illustrates the segmentation results produced by the SAM2 model.

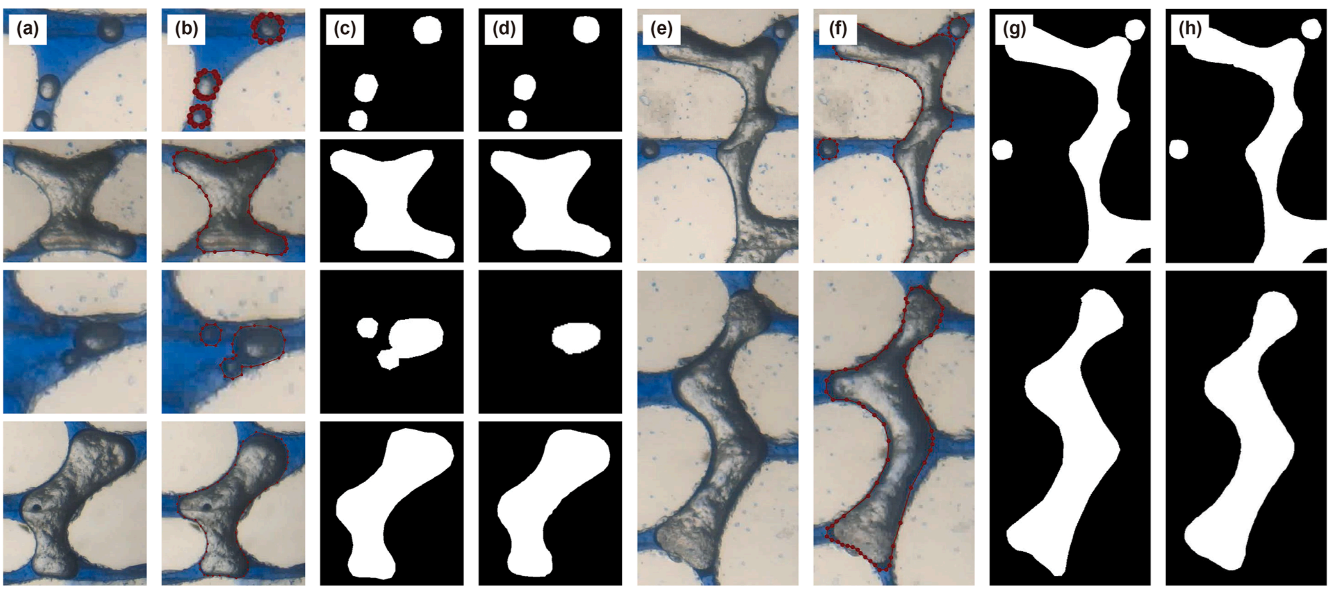

Fig. D1. Comparison between manual and SAM2 segmentation processes.

References

Alif, M.A.R., 2024. YOLOv11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems. https://doi.org/10.48550/arXiv.2410.22898.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8 (1), 53. https://doi.org/10.1186/s40537-021-00444-8.

Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., Liu, T., Heng, P.-A., Li, Q., 2024. MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation. Med. Image Anal. 98, 103310. https://doi.org/10.1016/j.media.2024.103310.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision–ECCV 2018. Springer International Publishing, Cham, pp. 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018b. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, Z., Chen, K., Lin, W., See, J., Yu, H., Ke, Y., Yang, C., 2020. Piou loss: Towards accurate oriented object detection in complex environments. In: Computer Vision–ECCV 2020: 16th European Conference. Springer, pp. 195–211. https://doi.org/10.1007/978-3-030-58558-7_12.

Durve, M., Tiribocchi, A., Bonaccorso, F., Montessori, A., Lauricella, M., Bogdan, M., Guzowski, J., Succi, S., 2022. DropTrack—Automatic droplet tracking with YOLOv5 and DeepSORT for microfluidic applications. Phys. Fluids 34 (8). https://doi.org/10.1063/5.0097597.

Fang, L., Lei, Y., Ning, J., Zhang, J., Feng, Y., 2024. A deep learning-based algorithm for rapid tracking and monitoring of gas–liquid two-phase bubbly flow bubbles. Phys. Fluids 36 (8). https://doi.org/10.1063/5.0222856.

Gavoille, T., Pannacci, N., Bergeot, G., Marliere, C., Marre, S., 2019. Microfluidic approaches for accessing thermophysical properties of fluid systems. React. Chem. Eng. 4 (10), 1721–1739. https://doi.org/10.1039/C9RE00130A.

Hamza, A., Hussein, I.A., Al-Marri, M.J., Mahmoud, M., Shawabkeh, R., Aparicio, S., 2021. CO_2 enhanced gas recovery and sequestration in depleted gas reservoirs: A review. J. Petrol. Sci. Eng. 196, 107685. https://doi.org/10.1016/j.petrol.2020.107685.

He, C., Sadeghpour, H., Shi, Y., Mishra, B., Roshankhah, S., 2024. Mapping distribution of fractures and minerals in rock samples using Res-VGG-UNet and threshold segmentation methods. Comput. Geotech. 175, 106675. https://doi.org/10.1016/j.compgeo.2024.106675.

He, D., Jia, C., Zhao, W., Xu, F., Luo, X., Liu, W., Tang, Y., Gao, S., Zheng, X., Li, D., Zheng, N., 2023. Research progress and key issues of ultra-deep oil and gas exploration in China. Petrol. Explor. Dev. 50 (6), 1333–1344. https://doi.org/10.1016/S1876-3804(24)60470-2.

Jahanbakhsh, A., Wlodarczyk, K.L., Hand, D.P., Maier, R.R.J., Maroto-Valer, M.M., 2020. Review of microfluidic devices and imaging techniques for fluid flow study in porous geomaterials. Sensors 20 (14), 4030. https://doi.org/10.3390/s20144030.

Jegham, N., Koh, C.Y., Abdelatti, M., Hendawi, A., 2024. Evaluating the evolution of YOLO (you only look once) models: A comprehensive benchmark study of YOLO11 and its predecessors. https://doi.org/10.48550/arXiv.2411.00201.

Jia, P., Li, Y., Guo, H., Feng, H., Cheng, L., 2023. Experimental study of pore-scale water flooding with phase change based on a microfluidic model in volatile

carbonate reservoirs. Appl. Sci. 13 (11), 6642. https://doi.org/10.3390/app13116642.
Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8, Version 8.0.0. https://github.com/ultralytics/ultralytics.
Jocher, G., Qiu, J., 2024. Ultralytics YOLO11. https://github.com/ultralytics/ultralytics.
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. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371.
Koeshidayatullah, A., 2023. Riding the wave: One-touch automatic salt segmentation by coupling SAM and SegGPT. In: Abu Dhabi International Petroleum Exhibition and Conference. https://doi.org/10.2118/216419-MS.
Lai-Dang, Q.-V., Lee, J., Park, B., Har, D., 2023. Sensor fusion by spatial encoding for autonomous driving. 2023 IEEE SENSORS 1–4. https://doi.org/10.1109/SENSORS56945.2023.10324977.
Li, Y., Liu, H.-Q., Peng, C., Jiao, P., Loh, W.L., Wang, Q., 2024. Insight into evolution of invasive patterns on fingering phenomenon during immiscible two-phase flow through pore structure. Pet. Sci. 21 (5), 3307–3325. https://doi.org/10.1016/j.petsci.2024.05.009.
Liu, N., Ju, B., Chen, X., Brantson, E.T., Mu, S., Yang, Y., Wang, J., Mahlalela, B., 2019a. Experimental study of the dynamic mechanism on gas bubbles migration, fragment, coalescence and trapping in a porous media. J. Petrol. Sci. Eng. 181, 106192. https://doi.org/10.1016/j.petrol.2019.106192.
Liu, N., Ju, B., Yang, Y., Brantson, E.T., Wang, J., Tian, Y., 2019b. Experimental study of different factors on dynamic characteristics of dispersed bubbles rising motion behavior in a liquid-saturated porous media. J. Petrol. Sci. Eng. 180, 396–405. https://doi.org/10.1016/j.petrol.2019.05.057.
Liu, X., Zhang, D., 2019. A review of phase behavior simulation of hydrocarbons in confined space: Implications for shale oil and shale gas. J. Nat. Gas Sci. Eng. 68, 102901. https://doi.org/10.1016/j.jngse.2019.102901.
Luo, M., Zhang, T., Wei, S., Ji, S., 2024. SAM-RSIS: Progressively adapting SAM with box prompting to remote sensing image instance segmentation. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2024.3460085.
Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. ISPRS J. Photogrammetry Remote Sens. 135, 158–172. https://doi.org/10.1016/j.isprsjprs.2017.11.009.
Nguyen, P., Mohaddes, D., Riordon, J., Fadaei, H., Lele, P., Sinton, D., 2015. Fast fluorescence-based microfluidic method for measuring minimum miscibility pressure of CO_2 in crude oils. Anal. Chem. 87 (6), 3160–3164. https://doi.org/10.1021/ac5047856.
Osco, L.P., Wu, Q., de Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Junior, J.M., 2023. The segment anything model (SAM) for remote sensing applications: From zero to one shot. Int. J. Appl. Earth Obs. Geoinf. 124, 103540. https://doi.org/10.1016/j.jag.2023.103540.
Ouyang, T., Tan, X., Yan, R., Xie, X., Liang, L., 2024. Two-phase flow characteristic and gas removal strategy of the paper-based microfluidic fuel cell. Energy Convers. Manag. 314, 118711. https://doi.org/10.1016/j.enconman.2024.118711.
Pan, S., Yang, B., Wang, S., Guo, Z., Wang, L., Liu, J., Wu, S., 2023. Oil well production prediction based on CNN-LSTM model with self-attention mechanism. Energy 284, 128701. https://doi.org/10.1016/j.energy.2023.128701.
Pavlidis, T., Liow, Y.-T., 1990. Integrating region growing and edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 12 (3), 225–233. https://doi.org/10.1109/34.49050.
Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Vasudev Alwala, K., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., Feichtenhofer, C., 2024. SAM 2: Segment anything in images and videos. https://doi.org/10.48550/arXiv.2408.00714.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.
Sakshi, Kukreja, V., 2023. Image segmentation techniques: Statistical, comprehensive, semi-automated analysis and an application perspective analysis of mathematical expressions. Arch. Comput. Methods Eng. 30 (1), 457–495. https://doi.org/10.1007/s11831-022-09805-9.
Shi, J., You, L., Wang, Y., Peng, B., 2024. Deep learning based method for dynamic tracking of bubbles in microfluidic gas-driven water experiments. Geoenergy Sci. Eng. 243, 213391. https://doi.org/10.1016/j.geoen.2024.213391.
Shi, Z., Chen, H., Zhang, D., 2025. Robustifying vision transformer for image forgery localization with multi-exit architectures. Pattern Recogn. 164, 111565. https://doi.org/10.1016/j.patcog.2025.111565.
Song, R., Wang, Y., Tang, Y., Jiajun, p., Liu, J., Yang, C., 2022. 3D printing of natural sandstone at pore scale and comparative analysis on micro-structure and single/two-phase flow properties. Energy 261, 125226. https://doi.org/10.1016/j.energy.2022.125226.
Sunkara, R., Luo, T., 2022. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. https://doi.org/10.1007/978-3-031-26409-2_27.
Tan, Q., Peng, H., Tian, J., Cao, Z., 2025. A review on mechanisms of CO_2-fluid-rock interaction during CO_2 injection into carbonate reservoirs. Geoenergy Sci. Eng. 249, 213773. https://doi.org/10.1016/j.geoen.2025.213773.
Tian, J., Kang, Y., You, L., Jia, N., Xi, Z., Luo, P., 2021. Investigation on water phase trapping mechanisms in tight gas reservoirs: Pore-scale visualization observation and core-scale flooding analysis. J. Petrol. Sci. Eng. 198, 108185. https://doi.org/10.1016/j.petrol.2020.108185.
Wang, C.-Y., Yeh, I.-H., Mark Liao, H.-Y., 2025. Yolov9: Learning what you want to learn using programmable gradient information. In: European Conference on Computer Vision. Springer, pp. 1–21. https://doi.org/10.1007/978-3-031-72751-1_1.
Wang, S., Li, C., Wang, R., Liu, Z., Wang, M., Tan, H., Wu, Y., Liu, X., Sun, H., Yang, R., 2021. Annotation-efficient deep learning for automatic medical image segmentation. Nat. Commun. 12 (1), 5915. https://doi.org/10.1038/s41467-021-26216-9.
Wang, Y.-J., You, L.-J., Yang, J., Kang, Y.-L., Chen, M.-J., Bai, J.-J., Tian, J., 2025. Adsorption and retention of fracturing fluid and its impact on gas transport in tight sandstone with different clay minerals. Pet. Sci. 22 (1), 370–383. https://doi.org/10.1016/j.petsci.2024.09.012.
Wang, Y., Zhao, Y., Petzold, L., 2024. An empirical study on the robustness of the segment anything model (SAM). Pattern Recogn., 110685 https://doi.org/10.1016/j.patcog.2024.110685.
Wang, Z., Li, C., Xu, H., Zhu, X., Li, H., 2025. Mamba YOLO: A simple baseline for object detection with state space model. Proc. AAAI Conf. Artif. Intell. 39 (8), 8205–8213. https://doi.org/10.1609/aaai.v39i8.32885.
Yang, Z., Tian, W., Deng, X., He, X., Wang, Z., Wang, J., Wang, Y., 2025. A deep learning-based segmentation method for multi-scale and overlapped bubbles in gas–liquid bubbly flow. Phys. Fluids 37 (3). https://doi.org/10.1063/5.0254679.
You, C., Dai, W., Liu, F., Min, Y., Dvornek, N.C., Li, X., Clifton, D.A., Staib, L., Duncan, J. S., 2024. Mine your own anatomy: Revisiting medical image segmentation with extremely limited labels. IEEE Trans. Pattern Anal. Mach. Intell. 46 (12), 11136–11151. https://doi.org/10.1109/TPAMI.2024.3461321.
Zhang, C., Jiang, W., Zhang, Y., Wang, W., Zhao, Q., Wang, C., 2022. Transformer and CNN hybrid deep neural network for semantic segmentation of very-high-resolution remote sensing imagery. IEEE Trans. Geosci. Rem. Sens. 60, 1–20. https://doi.org/10.1109/TGRS.2022.3144894.
Zhang, W., Wu, T., Li, Z., Li, Y., Shi, Y., 2021. Research on fracture recognition in well logging images: Adversarial learning with attention. 2021 IEEE Int. Geosci. Remote Sens. Symp. IGARSS 4079–4082. https://doi.org/10.1109/IGARSS47720.2021.9554264.
Zhou, T., Xia, W., Zhang, F., Chang, B., Wang, W., Yuan, Y., Konukoglu, E., Cremers, D., 2024. Image segmentation in foundation model era: A survey. https://doi.org/10.48550/arXiv.2408.12957.
Zhu, Y., Gao, X., Ke, B., Qiao, R., Sun, X., 2023. Coarse-to-Fine: Learning compact discriminative representation for single-stage image retrieval. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 11226–11235. https://doi.org/10.1109/ICCV51070.2023.01034.