

# YOLOv8-Based Visual Localization and Fixed-Point Immobilization of Cells Driven by Surface Acoustic Waves

Jiguang Han, Jun Wei

School of Automation and Electrical Engineering, University of Jinan, Jinan, Shandong, China

**Abstract:** Automated cell manipulation is a key technology in biomedical research. Realizing real-time detection and precise positioning of targets is an essential prerequisite for applications such as single-cell analysis and drug screening. Most existing methods rely on complex vision systems or multi-degree-of-freedom manipulators, which suffer from issues including low integration and limited automation. This paper proposes an automated system that combines deep learning-based object detection with Surface Acoustic Wave (SAW) technology driven by Interdigital Transducers (IDTs) for the real-time recognition and fixed-point immobilization control of zebrafish embryos. The system obtains the positional coordinates of embryos in the microchannel in real time using the YOLOv8n model. When an embryo enters the target area, SAW is automatically triggered to drive the embryo to the designated immobilization point and maintain its stability. Experimental results demonstrate that the mean average precision (mAP@0.5) of the YOLOv8n model for embryo detection reaches 0.994, with a single-frame inference time of 78 ms. In 15 repeated experiments, the system achieves a success rate of 93.3% in driving embryos to the target position. This method provides a compact, non-contact, and high-precision integrated solution for detection and control in automated cell manipulation.

**Keywords:** Object Detection; Automated Control; Surface Acoustic Wave; Interdigital Transducer.

## 1. Introduction

Automated cell manipulation is one of the core research directions in biomedical engineering and micromanipulation [1, 2], and has important application value in single-cell analysis, drug screening, embryonic development research and other fields. The realization of real-time detection and precise localization of target cells or embryos is a prerequisite for automated manipulation. However, due to the tiny cell scale, susceptibility to environmental interference, and high requirements for non-invasiveness during manipulation, traditional mechanical manipulation methods face the dual challenges of precision and invasiveness.

In recent years, significant progress has been made in technologies based on optical tweezers[3], electrorotation[4], magnetic manipulation[5], and acoustic fluidics. Among them, surface acoustic wave (SAW) technology has gradually become an important means for cell manipulation in the field of microfluidics owing to its advantages of non-contact actuation, good biocompatibility, fast response, and easy integration[6, 7]. SAW is generated via interdigital transducers on a piezoelectric substrate in the form of traveling waves, and realizes the actuation and concentration of microparticles through the acoustic streaming effect and acoustic radiation pressure. Nevertheless, most existing SAW manipulation systems adopt an open-loop control mode with preset fixed driving parameters, lacking the capability of real-time target position perception and closed-loop regulation. This results in limited manipulation precision and makes it difficult to achieve automated fixed-point immobilization.

Meanwhile, with the development of computer vision and deep learning technologies, cell recognition and localization methods based on object detection have shown broad application prospects in microfluidic systems. Traditional methods usually rely on image processing techniques such as image segmentation and edge detection to extract cell

contours and internal structural features, and then calculate their posture parameters. However, in microscopic imaging environments, factors such as uneven illumination, background noise, and the complexity of the cell structure itself pose significant challenges to traditional image processing methods in the recognition of samples such as zebrafish embryos, making stable recognition difficult to achieve. In contrast, deep learning methods based on convolutional neural networks can automatically learn multi-level structural features of targets and realize robust object recognition and posture estimation under complex backgrounds[8], providing strong support for high-precision automated manipulation of cells in microfluidic systems.

In response to the above issues, this paper proposes an automated control system combining deep learning-based object detection with SAW actuation technology for the real-time recognition and fixed-point immobilization of zebrafish embryos. Taking object detection as the core of visual perception, the system obtains the positional coordinates of embryos in the microchannel in real time. Through a closed-loop control strategy, it automatically triggers the IDT to generate SAW, which actuates the embryos to move toward the target position and stably immobilize there. This system is designed to provide a non-contact and integrated detection-control solution for automated cell manipulation.

## 2. Materials and Methods

### 2.1. System Hardware Composition

The system hardware adopts a modular integrated design, which mainly consists of four parts: the SAW driving module, the microscopic vision acquisition module, the host computer control module, and the microfluidic manipulation chamber. These modules work collaboratively to fulfill the full-process automated manipulation of zebrafish embryos, including image acquisition, recognition and positioning, acoustic wave

driving, and fixed-point immobilization.

The SAW driving module serves as the core execution unit of the entire system, which is composed of a standard straight-type IDT chip, a radio frequency (RF) signal generator (DG4202, RIGOL, China) and a power amplifier (LZY22+, Mini-Circuits, USA). The IDT chip uses 128° YX-cut lithium niobate (LiNbO<sub>3</sub>) as the piezoelectric substrate, adopts a straight interdigital electrode structure with 50 electrode pairs, and has a central resonant frequency of 15 MHz. According to the principle of SAW excitation, the electrode period corresponding to this frequency is determined by the acoustic velocity of the substrate (approximately 3980 m/s). Stable unidirectional traveling-wave SAW can be excited via the inverse piezoelectric effect under the excitation of RF signals. The RF signal generator is used to output sinusoidal excitation signals with adjustable frequency and voltage, which drive the IDT after being amplified by the power amplifier. The magnitude of the acoustic radiation force can be linearly controlled by adjusting the output power, so as to regulate the driving force applied to the embryos.

The microscopic vision acquisition module is responsible for real-time image acquisition of zebrafish embryos. The operating platform is equipped with a camera (NIB610, Nexcope, China). An inverted microscope is used to magnify and observe zebrafish embryos in the microfluidic chamber, with a magnification adjustable from 10× to 40× to ensure clear imaging of the morphology and position of embryos. The camera is coaxially connected to the microscope through a camera interface, collecting motion images of embryos in real time and transmitting them to the host computer to

provide raw visual data.

The host computer control module acts as the decision-making core of the system, consisting of a desktop computer and a Python-based control program. The computer configuration meets the requirements of lightweight inference for deep learning models, and mainly runs the YOLOv8n object detection model to complete embryo recognition, center coordinate extraction and position deviation calculation. Data interaction with the RF signal generator is realized through the serial port communication protocol. Start and stop commands are automatically sent according to the real-time positioning results of embryos, realizing closed-loop control of SAW driving.

The microfluidic manipulation chamber is the carrier for zebrafish embryo manipulation, which is fabricated with polydimethylsiloxane (PDMS). The chamber has dimensions of 9 mm in length, 1.5 mm in width and 1.5 mm in height, and can hold a small amount of liquid and zebrafish embryos. To achieve reliable sealed integration between the chamber and the IDT chip, oxygen plasma surface activation treatment is used to activate the bonding surface of the PDMS chamber and the surface of the IDT substrate (LiNbO<sub>3</sub>). The two form a stable sealed interface through irreversible chemical bonding at room temperature. This bonding method can effectively prevent liquid leakage during operation, thereby ensuring the smooth movement of embryos driven by acoustic radiation force and the stability of the manipulation environment.

## 2.2. Target Recognition and Localization

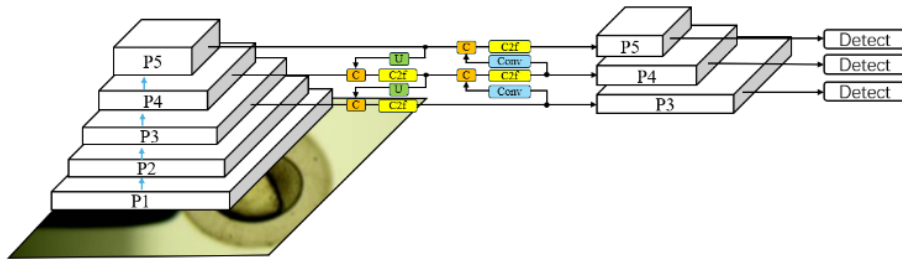


Fig 1. The network structure of the model

**Dataset Construction.** Zebrafish embryos were selected as the research object in the experiments. Since embryonic development involves multiple stages with significant morphological variations in embryonic structures across different stages, it is necessary to identify embryos at the target stage. The blastula stage was defined as the target stage, while other stages were regarded as non-targets. Embryos at the blastula stage exhibit a regular spherical shape with distinct structural features and a diameter of approximately 0.6 mm, facilitating visual recognition. A total of 260 original images were collected, covering embryo samples with different postures and spatial positions. To enhance the generalization ability of the model, data augmentation was performed on the original images, including random rotation, horizontal and vertical flipping, as well as brightness and contrast adjustment. After augmentation, the dataset was expanded to 520 images. Manual annotation was conducted using the LabelImg tool, and the dataset was divided into training set, validation set and test set at a ratio of 7:1.5:1.5.

**Model Selection and Training.** YOLOv8n was selected as the core detection model for the system. The network structure of the model is shown in Figure 1. The YOLOv8 series achieves a favorable balance between detection

accuracy and inference speed. As a lightweight version, YOLOv8n has approximately 3.2 million (3.2M) parameters and a computational complexity of about 8.7 billion (8.7B) FLOPs, making it suitable for real-time operation in a CPU environment and compatible with the hardware configuration of the system control computer (Intel i5-13400, 32GB RAM).

Table 1. Training Hyperparameters

Hyperparameter	Value
Batch size	8
Initial learning rate	0.01
Optimizer	SGD
Momentum	0.937
Weight decay	0.0005
Learning rate scheduler	Cosine annealing

The model training was based on the PyTorch deep learning framework, adopting a transfer learning strategy where the pre-trained weights of YOLOv8n on the COCO dataset were loaded for fine-tuning. The input image size was uniformly adjusted to 640×640 pixels. The training hyperparameters are shown in Table 1. The model was trained for 100 epochs with

an early stopping mechanism implemented: training was terminated if the validation loss did not decrease for 50 consecutive epochs.

### 2.3. System Working Flow

The system adopts an automated working mode integrating visual perception, deep learning-based localization, closed-loop decision-making and acoustic wave actuation, which independently accomplishes the recognition, localization, motion driving and fixed-point immobilization of zebrafish embryos. After power-on initialization, all modules start synchronously and collaborate in a fixed time sequence. The detailed workflow is as follows.

The image acquisition module continuously captures microscopic images of embryos in the microfluidic chamber at a frame rate of 10 FPS and transmits the frames to the host computer in real time. The lightweight YOLOv8n object detection model deployed on the host computer performs inference on each frame, outputs the bounding box of the embryo and calculates its central pixel coordinate. Combined with the pre-calibrated pixel-to-physical coordinate conversion relationship, the system maps the pixel coordinate to the actual physical coordinate in the chamber, achieving precise localization of the embryo. The control program compares the real-time acquired embryo position with the preset target immobilization point and calculates the distance deviation between them.

If the deviation exceeds the set immobilization accuracy threshold, the host computer sends a start command to the RF signal generator via serial communication. The signal generator outputs an RF signal with fixed frequency and voltage, which excites the standard straight IDT through the power amplifier to generate a unidirectional traveling-wave surface acoustic wave. Driven by the combined effects of acoustic streaming and acoustic radiation pressure, the embryo moves smoothly along the SAW propagation direction and gradually approaches the target immobilization point. The system continuously updates the position information via visual feedback, forming a closed-loop regulation.

When the embryo position deviation is reduced to within the threshold, the host computer immediately sends a stop command. The signal generator cuts off the output, the IDT ceases excitation, and the SAW disappears accordingly. The embryo stably remains at the target position without external force. The entire process constitutes a closed-loop control with real-time perception capability.

## 3. Results

### 3.1. IDT Performance Test

The reflection coefficient  $S_{11}$  of the standard straight-type IDT was characterized using a vector network analyzer (NanoVNA-FV2). The measurement results show that the IDT has an obvious resonance dip at 14.9 MHz, with the  $S_{11}$  value reaching  $-15.3$  dB. The measured resonant frequency is slightly different from the designed value, which is within the acceptable range of typical fabrication tolerances. At this resonant frequency, the reflection of the input radio-frequency signal is minimized, indicating that the electro-acoustic energy conversion efficiency reaches the maximum. In addition, the bandwidth of the IDT is about 1.8 MHz (from 14.1 MHz to 15.9 MHz), which confirms its good narrowband response characteristics.

### 3.2. Object Detection Performance

The trained YOLOv8n model was evaluated on the test set, with mean average precision (mAP) adopted as the core evaluation metric. The precision-recall (P-R) curve is shown in Fig. 2. The results demonstrate that the precision of the model remains at a level close to 1.0 throughout, and the area under the curve (AUC) is nearly 1, indicating that the model can maintain extremely high classification accuracy at different recall levels with almost no false detections. The mean average precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5) for all classes reaches 0.994, verifying the model's stable recognition capability for zebrafish embryo targets.

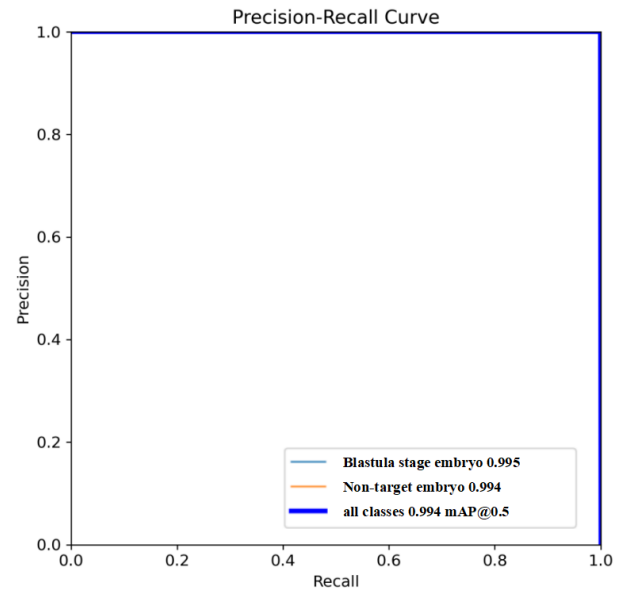


Fig 2. The precision-recall curve

Fig. 3 presents the training and validation loss curves, as well as the evaluation metric curves, during the model training process. The training loss decreases continuously and gradually stabilizes with the increase of training epochs, while the validation loss reduces synchronously with narrowed fluctuations. The loss trends of the training set and validation set are consistent, with no obvious overfitting or underfitting observed, demonstrating the stable training and good convergence of the model. Meanwhile, the precision, recall, and mAP@0.5 increase rapidly in the early stage of training and stabilize at a level close to 1.0 after approximately 20 epochs. The mAP@0.5:0.95 gradually rises to around 0.85 as training proceeds, indicating that the model still exhibits excellent detection performance under stricter IoU thresholds and possesses strong generalization ability.

### 3.3. Automated Manipulation Performance of Zebrafish Embryos

Experiments were carried out on the established microfluidic manipulation platform, with the target immobilization point preset on the right side of the chamber channel. The IDT driving parameters were set as follows: excitation frequency of 14.9 MHz and output power of  $-10$  dBm. To realize real-time detection and control of the system, the visual feedback frequency was set to 10 FPS. Before each experiment, embryos were manually placed in the initial area on the left side of the chamber, and then the automated control program was activated to drive the embryos toward the target

position.

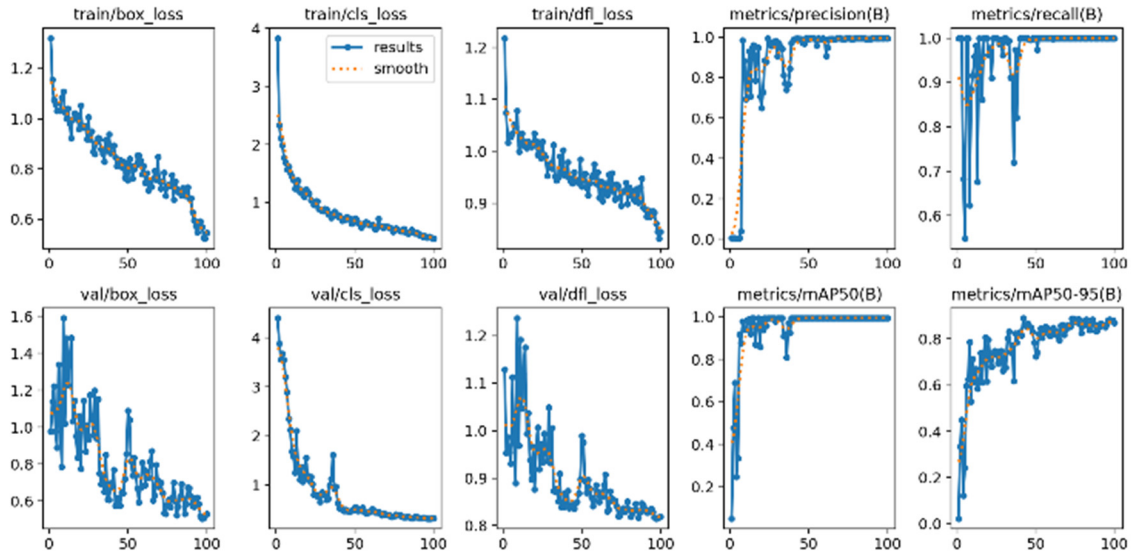


Fig 3. Training and validation loss curves for the model

Starting from the initial position, the embryos moved smoothly along the positive X-axis direction under SAW actuation. The motion process consisted of three phases: acceleration, uniform velocity, and deceleration. The velocity increased gradually within approximately 0.5 s after startup, followed by uniform motion at about 0.5 mm/s. The velocity decreased when approaching the target point, and the embryos finally stably immobilized in the target area at approximately 3.6 s. A total of 15 repeated experiments were conducted, and the system successfully drove and stably immobilized embryos at the target area in 14 trials, with a success rate of 93.3%. The only failure was caused by target loss due to frame rate delay when the embryo approached the target area during movement.

## 4. Conclusion

To address the demands for real-time target detection and precise position control in automated cell manipulation, this paper proposes an automated system that combines deep learning-based object detection with SAW actuation technology, and conducts experimental verification using zebrafish embryos. With target recognition as the core of visual perception, the system realizes real-time identification and coordinate localization of embryos in microchannels. Using IDT devices, it generates stable acoustic field actuation at a frequency of 14.9 MHz, achieving full-process closed-loop control from target perception and precise localization to fixed-point immobilization. Experimental results demonstrate that the mean average precision (mAP@0.5) of the YOLOv8n model on the test set reaches 0.994, and the system achieves a success rate of 93.3% in driving embryos to the target position in 15 repeated experiments. In the future, through IDT array integration and channel expansion, the system can

realize diversified cell manipulation and be applied to fixed-point immobilization, microscopic observation and precision drug delivery of more types of microscale biological samples.

## References

- [1] Gong H, Li L, Qiu J, et al. Automatic cell rotation based on real-time detection and tracking[J]. IEEE Robotics and Automation Letters, 2021, 6(4): 7909-7916.
- [2] Wang H, Bai K, Chen J, et al. Digital holography based three-dimensional multi-target locating for automated cell micromanipulation[J]. IEEE Transactions on Automation Science and Engineering, 2022, 21(1): 332-342.
- [3] Xie M. Autonomous robot-aided optical tweezer system for biological cell manipulation[J]. The International Journal of Advanced Manufacturing Technology, 2019, 105(12): 4953-4966.
- [4] Julius L A N, Akgül D, Krishnan G, et al. Portable dielectrophoresis for biology: ADEPT facilitates cell trapping, separation, and interactions[J]. Microsystems & Nanoengineering, 2024, 10(1): 29.
- [5] Hu H, Krishaa L, Fong E L S. Magnetic force-based cell manipulation for in vitro tissue engineering[J]. APL bioengineering, 2023, 7(3).
- [6] Huang J, Ren X, Zhou Q, et al. Flexible acoustic lens-based surface acoustic wave device for manipulation and directional transport of micro-particles[J]. Ultrasonics, 2023, 128: 106865.
- [7] Liu B, Ren X, Xue T, et al. A Slanted-Finger Interdigitated Transducer Microfluidic Device for Particles Sorting[J]. Micromachines, 2025, 16(4): 483.
- [8] Wang Y, Liu Y, Sun M, et al. Deep-learning-based polar-body detection for automatic cell manipulation[J]. Micromachines, 2019, 10(2): 120.