Journal of Marine Science and Application (2026) 25: 1392-1408
https://doi.org/10.1007/s11804-025-00772-8

RESEARCH ARTICLE

An Integrated Lightweight YOLO Framework Combining Structured Pruning, Knowledge Distillation, and Multi-Frame Linear Interpolation for Robust Real-Time Nonlinear Dynamic Response Analysis of Small-Scale Underwater Targets under Bubble Loading

Jiuqiang Wang[1], Dongyan Shi[1], Yuxin Gou[1], Haifeng Zhang[1] and Xiongwei Cui[2]

Received: 02 April 2025 / Accepted: 03 June 2025
© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

A You Only Look Once version 8 (YOLOv8)-based detection model optimized via pruning and knowledge distillation is proposed for the challenging task of detecting dynamic responses of small underwater bubbles. Underwater bubble experiments were conducted to collect dynamic image sequences, and the improved YOLOv8 model was used to detect bubbles and extract their nonlinear dynamic responses. Through model pruning and distillation, the optimized model considerably reduces model parameters and accelerates inference while maintaining high detection accuracy. The experimental detection results were validated by a coupled Eulerian-Lagrangian computational fluid dynamics simulation, which showed high consistency with the observed bubble dynamics. The proposed approach demonstrates excellent performance in recognizing small targets with nonlinear dynamic responses, shows strong robustness in complex underwater scenarios, and offers high inference efficiency and ease of deployment.

Keywords PruneDistill-YOLOv8; Structured pruning; Knowledge distillation; Underwater target detection; Bubble dynamics

1 Introduction

The theoretical system and technological applications of underwater bubble dynamics continue to expand in various fields, including fluid mechanics, multiphase-flow mass transfer, underwater acoustics, and thermodynamics. Studies have shown that the generation, motion, coalescence, and collapse behavior of microscale bubbles can considerably alter the local flow field characteristics through cavitation effects, Magnus force, and secondary Bjerknes force, causing transient fluctuations in physical and chemical parameters such as pH and dissolved oxygen content in water (Kedrinskii, 2016; Zhang and An, 2013).

In marine engineering systems, millimeter-scale structural responses under transient loading present critical detection challenges. Representative scenarios include the following:

1) Precision monitoring of compliant structures under impulsive hydrodynamic forces (characteristic displacement <2 mm) (Fang et al., 2017),

2) Microdeformation analysis of submerged components under cavitation effects (time duration <100 ms) (Ragunathan , 2016),

3) Dynamic characterization of drug delivery systems under acoustic excitation (Ho et al., 2021; Rozhdestvensky, 2022).

These scenarios share three fundamental requirements: submillimeter displacement resolution (per API RP 1111 specifications), millisecond-level temporal resolution (exceeding DNVGL-RP-F105 standards), and robustness

Article Highlights

- Structured pruning reduces the original YOLOv8 model's parameters by 57% (from 11.2 to 4.8 M) and cuts the per-frame inference latency by approximately 26.7% (from 15 to 11 ms), ensuring high-efficiency real-time detection in resource-limited scenarios.
- By integrating knowledge distillation to compensate for pruning-induced accuracy loss, the model achieves an mean average precision (mAP) of 0.81 and an F1 score of 0.77. It maintains a submillimeter trajectory reconstruction accuracy (RMSE = 0.18 mm) even under challenging conditions such as turbid water and heavy bubble occlusion.
- The optimized model demonstrates excellent stability and adaptability in complex dynamic underwater scenarios, such as bubble-target interactions, with considerably lower detection errors compared to that using pruning or distillation alone.

✉ Xiongwei Cui
cuixiongwei@hrbeu.edu.cn

[1] College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin 150001, China

[2] College of Shipbuilding Engineering, Harbin Engineering University, Harbin 15001, China

against multiphase-flow interference (Zhang et al., 2024).

Our experimental configuration with 13.5 mm cylindrical targets establishes a generalized testing platform that preserves the essential physics of these applications. The selected scale allows the following: direct observation of bubble-target interaction dynamics, quantifiable validation through laser vibrometry, and parametric studies spanning industrial-relevant Reynolds numbers (5×10^3 to 2×10^4).

In these cases, the targets have light mass (<500 g) and small characteristic dimensions (ranging from centimeters to micrometers), exhibiting remarkable nonlinear motion characteristics. Their dynamic modeling needs to account for fluid-structure interaction effects and the interactions of multiscale physical fields (He and Kashiwagi, 2012).

To study the dynamic responses of small-scale targets under bubble loading, electric spark experiments in a glass water tank can simulate real-world conditions, providing critical data for theoretical model development. Compared with traditional experiments, underwater electric spark bubble experiments not only simplify operation but also rapidly generate large quantities of bubbles, considerably improving experimental efficiency (Li et al., 2021). Furthermore, this method avoids the potential risks of using chemicals or high-energy materials, meets modern laboratory safety requirements, and does not produce pollutant emissions, making it environmentally friendly and having broad application prospects (Zheng et al., 2023).

Experimental studies provide an efficient means to thoroughly reveal the dynamic responses of small-scale targets under bubble loading—thereby better supporting a variety of marine applications such as underwater bubble cleaning, performance evaluation of autonomous underwater systems, microbubble-assisted environmental remediation, and microscale underwater sensor calibration. However, to date, no mature method has been established in the industry for rapidly processing experimental video data and extracting dynamic response parameters.

The research team adopted an underwater electric spark bubble generation technology and carefully designed the experimental environment to ensure the research's efficiency and accuracy (Cho and Jin, 2018; Liang et al., 2022). The experimental platform consists of a high-performance electric spark discharge system and a specially designed experimental framework, which is capable of accurately simulating the generation and evolution of underwater bubbles under controlled conditions (Li et al., 2022; Mang et al., 2018). The electric spark discharge system, as the core of the platform, is equipped with advanced power and control equipment to precisely adjust the voltage, discharge frequency, and pulse duration to generate high-energy electric sparks, leading to local ionization in the water and the formation of stable and repeatable bubbles (Liu et al., 2021). To ensure efficiency and control during the experiments, the team also developed a precise experimental framework to support the discharge system and maintain safety and stability throughout the process (Lin et al., 2021). This framework allows flexible adjustment of electrode relative positions, bubble generation depths, and distribution ranges, enabling more accurate observation and measurement of bubble behavior (Liu et al., 2023).

Dynamic response refers to the evolution process of the target's motion state under time-varying loading excitation, which can be characterized by the time-domain features of displacement $x(t)$, velocity $v(t)$, and acceleration $a(t)$ (Hartmann et al., 2013). In this study, bubble loading serves as a typical transient excitation source. The shock waves and microjets generated by the pulsation process of bubbles induce millimeter-scale dynamic displacements in the target (Han et al., 2025). Nonlinear response is a special form of dynamic response, where the system's constitutive relation shows nonlinear characteristics due to changes in parameters such as system stiffness and damping with the motion state (Kalateh and Koosheh, 2018). The nonlinear effects involved in this study include the following:

1) Asymmetric impact loads generated by microjets during the bubble collapse phase (Deng et al., 2012),

2) Time-varying added mass effects caused by the coupling motion between the target and the bubble interface (Paquette et al., 2015),

3) Damping nonlinearities induced by cavitation in the flow field (Trivedi, 2017).

However, capturing the position of small-scale targets automatically and quickly and calculating their dynamic response parameters remain a challenge (Liu et al., 2025).

Existing methods for solving these problems primarily include traditional image processing techniques and deep learning-based target detection algorithms. Traditional methods generally rely on image enhancement, denoising, and restoration technologies to improve the quality and clarity of underwater images (Zhou et al., 2023). For example, to address underwater image blurring, researchers have proposed algorithms that use semi-mean filtering for dust removal and lighting equalization for feature enhancement, improving image visibility (Awang et al., 2021). However, these methods have limited effectiveness in small-target detection, struggling to meet the high precision and real-time requirements (Lu et al., 2023).

In recent years, deep learning-based target detection algorithms have made substantial progress in small-target detection (Elmezain et al., 2025). You Only Look Once (YOLO) is an efficient target detection algorithm that has shown remarkable advantages in small-target dynamic response detection. YOLO has high real-time capabilities, quickly identifying and tracking targets at high frame rates, which is crucial for precise detection of small targets in dynamic scenes (Hao et al., 2025; Schontag et al., 2022). Additionally, YOLO's global feature extraction mechanism is highly sensitive to the spatial position and motion char-

acteristics of small targets, enabling stable detection even in complex backgrounds or under bubble occlusion. However, YOLO still faces challenges in detecting small-scale targets with sufficient accuracy (Zhao and Xu, 2025).

To further improve detection performance, researchers have employed techniques such as pruning and knowledge distillation (KD) to improve YOLO (Li et al., 2024; Xie et al., 2023). Pruning removes redundant parameters and computations in the model, reducing model complexity and computational cost thus improving inference speed. KD involves transferring knowledge from a large teacher model to a smaller student model, allowing it to maintain a high accuracy while reducing the number of model parameters and computational needs. These combined techniques have made the improved YOLO model capable of handling small-scale target detection tasks with both high accuracy and real-time performance (Hao et al., 2025; Xie et al., 2023).

This study is the first to organically integrate KD and structured pruning techniques into the YOLO model, achieving an ideal balance between efficiency and accuracy. Compared with traditional models, the pruned and distilled YOLO excels at detecting small objects and analyzing the responses of small-scale targets under bubble loading: It not only considerably boosts detection precision but also drastically reduces model size and inference overhead, making it well suited for resource-constrained real-time and embedded scenarios. To further mitigate image artifacts caused by spark-generated bubble discharges, we introduce a linear interpolation algorithm to compensate for visual anomalies and then apply pruning and distillation optimizations to the network—thereby ensuring both real-time performance and high-precision detection. Ultimately, the proposed PruneDistill-YOLOv8 model delivers outstanding results in dynamic response analysis, offering an innovative, lightweight, and efficient technical solution for real-time detection and response studies of small targets in complex underwater environments.

2 YOLO-based detection algorithm

2.1 Features of the YOLOv8 algorithm

YOLOv8 is the latest generation model in the field of object detection, continuing and extending the outstanding performance of the YOLO series in terms of real-time capabilities and efficiency. The overall architecture of YOLOv8 consists of an input layer, backbone network (using an optimized CSPNet), feature fusion module (introducing the efficient layer aggregation network GELAN), detection head, and output layer; collectively, these components further enhance the detection accuracy and computational efficiency (Thanh et al., 2024). Its core innovations include the Programmable Gradient Information (PGI) mechanism, which addresses the potential information loss during feature sampling and quantization across different transmission layers. Combining the main branch, auxiliary reversible branches, and multilevel auxiliary information optimizes gradient transfer and weight updates (Lin et al., 2022).

At the same time, GELAN achieves efficient feature extraction and information fusion through the Reparameterized Convolution (RepConv) module, enhancing the comprehensive use of multiscale features, making YOLOv8 particularly effective in handling small and dense targets (Hu and Ke, 2024). Compared to older versions, YOLOv8 shows considerable improvements in several aspects.

1) By optimizing the backbone network and feature fusion module, YOLOv8 considerably improves the detection accuracy, especially in small-target detection, and robustness in complex environments.

2) The PGI mechanism and GELAN module improve the inference speed by approximately 15% while reducing the model parameters by about 20%, greatly reducing the demand for computational resources.

Additionally, YOLOv8 introduces an adaptive anchor frame mechanism, which exhibits higher adaptability and generalization ability across different datasets and application scenarios, providing broad applicability in complex tasks (Chaffre et al., 2023). This adaptive mechanism is particularly important in the analysis of the nonlinear responses of small-scale targets under bubble loading as it can effectively handle detection challenges posed by target size changes and nonlinear characteristics. Furthermore, through multiple technological innovations and architectural optimizations, YOLOv8 not only achieves breakthrough improvements in detection accuracy and speed but also makes substantial progress in model complexity and resource efficiency (Xiao et al., 2024).

2.2 Principles of structured pruning

Pruning is a commonly used neural network model compression technique that aims to reduce the complexity and computational demand of a model by removing redundant or less important network parameters (Xian et al., 2023) while trying to maintain its performance. Pruning not only reduces the model's storage space but also accelerates the inference process, improving the efficiency of applications on resource-constrained devices (Pietron et al., 2023).

Structured pruning and unstructured pruning have advantages when optimizing the YOLO model. Structured pruning removes the entire convolution kernels, channels, or layers, preserving the overall architecture of the model, which allows the pruned model to run efficiently on standard hardware (such as GPUs and CPUs), making it particularly suitable for deployment on resource-constrained devices. In contrast, unstructured pruning removes individual redundant parameters, which can effectively reduce computational complexity in theory, but because it generates sparse matrices, it is difficult to fully utilize conventional

hardware, limiting the improvement of inference speed (Jin et al., 2024; Ruan et al., 2021).

For the analysis of the nonlinear responses of small-scale targets under bubble loading, structured pruning provides a more stable and efficient performance, allowing the YOLO model to considerably improve the processing speed and real-time capabilities while maintaining accuracy (Tonellotto et al., 2018). This makes it more suitable for applications requiring high real-time performance.

During the pruning process, it is necessary to first evaluate the "importance" of each convolution kernel, channel, or layer. Common evaluation criteria include L_1/L_2 norms based on weights, gradient information, feature map responses, and others. Suppose that we have a convolution layer whose output feature map is $\boldsymbol{Y} = F(\boldsymbol{X}, \boldsymbol{W})$, where $\boldsymbol{X}$ is the input and $\boldsymbol{W}$ is the weight matrix. For each channel c or convolution kernel k, the importance measure can be calculated as follows:

$$I_k = \left\| \boldsymbol{W}_k \right\|_p \tag{1}$$

The weight matrix of the k convolution kernel is denoted by $\boldsymbol{W}_k$, and $\|\cdot\|$ is the measured L_p norm, such as the L_1 or L_2 norm, which reflects the "size" or "importance" of the convolution kernel. The larger the norm of a channel or convolution kernel, the greater its contribution to the network and the higher its importance.

Based on this importance measure, structured pruning will choose to remove convolution kernels, channels, or layers with smaller norms. A threshold θ is set, and all convolution kernels or channels with an importance measure below this threshold are removed. In this study, we adopt a validation-driven dynamic threshold selection strategy: We first compute and normalize the L_1 norm of each convolution kernel and then generate a series of candidate pruning ratios within a preset range (e.g., 10%–60%). For each ratio, we remove the corresponding smallest-norm kernels, perform three to five rounds of rapid fine-tuning, and measure the mean average precision (mAP) drop on the validation set. Finally, we choose the largest pruning ratio P^* that limits the accuracy loss to ⩽1% and set the L_1 norm at that quantile as the pruning threshold for the layer. The entire search, fine-tuning, and evaluation pipeline is fully automated via scripting—no manual threshold tuning—ensuring an optimal balance between model compactness and detection performance. In this study, we employ two mAP metrics: mAP@0.5, which measures the average precision (AP) at a single Intersection over Union (IoU) threshold of 0.5 by computing the area under the precision-recall curve for detections with IoU ⩾ 0.5, and mAP@0.5:0.95, which averages the AP over 10 IoU thresholds (0.50, 0.55, ⋯, 0.95), providing a more comprehensive evaluation of the model's localization accuracy across both loose and strict matching criteria. For the k channel or convolution kernel, the pruning decision can be expressed as

$$\text{prune}_k = \begin{cases} 1, \text{if } I_k < \theta \\ 0, \text{if } I_k \geqslant \theta \end{cases} \tag{2}$$

where $\text{prune}_k = 1$ represents the removal of the k convolution kernel or channel and $\text{prune}_k = 0$ represents the retention of the structural unit.

Through structured pruning, the goal is to reduce the model's computational complexity and storage requirements while maintaining performance.

The objective function is set to minimize the computational complexity of the pruned network while maintaining performance. Suppose that the original model's computational complexity is C_{orig}, the pruned model's computational complexity is C_{pruned}, and the performance loss is ΔP; the optimization objective can be expressed as

$$\min\left(C_{\text{pruned}}\right) \text{s.t. } \Delta P \leqslant \epsilon \tag{3}$$

where ϵ represents the tolerance for performance loss. The pruned model needs to find a balance between computational complexity and performance (Wang et al., 2016). Two adjacent convolution layers are tightly connected, and a change in the output dimension of the first convolution layer will cause a corresponding change in the input dimension of the second convolution layer (Hong et al., 2020). This convolution operation continues from the first convolution layer to the second-last convolution layer and the last convolution layer. A comparison between the original model and the structured pruning method is shown in Figure 1 (Cheng et al., 2024).

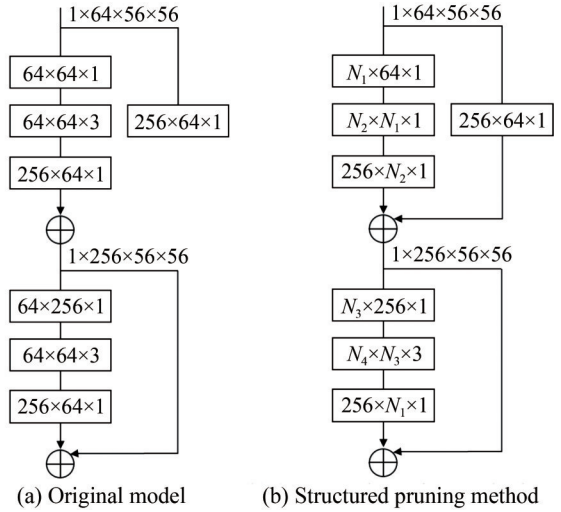


Figure 1 Comparison of the original model and the structured pruning method

2.3 Principles of knowledge distillation

KD is a method that utilizes a pretrained high-capacity model (called the teacher model) to guide a smaller model (called the student model) in learning. The core idea is to

enable the student model to learn not only from the hard labels (i.e., the ground-truth labels in the training data) but also from the soft target distribution output by the teacher model. This allows the student model to more effectively capture relationships between different classes and improves its generalization ability (Ge et al., 2020).

When optimizing YOLOv8, KD can effectively mitigate the capacity reduction issue caused by structured pruning, allowing the pruned student model to retain as many feature representations and prediction capabilities as possible (Zhou et al., 2024).

Hard labels refer to the actual class labels provided in the training data. The student model must learn based on these labels, and hard label loss is typically defined using cross-entropy loss. Suppose that a data sample is given as (x, y), where x is the input data and y is the true label. The prediction distribution of the student model is denoted as $p_S(y|x)$, which is defined as

$$p_S(y|x) = \frac{\exp(z_S(y))}{\sum_j \exp(z_S(j))} \tag{4}$$

where $z_S(y)$ represents the logits (raw output values) of the student model for class y. The cross-entropy loss, which measures the difference between the student model's prediction distribution $p_S(y|x)$ and the true label y, is

$$\mathcal{L}_{\text{hard}} = -\sum_i y_i \log p_S(y_i|x) \tag{5}$$

where y_i is a one-hot encoded vector of the ground-truth labels. The teacher model's soft target distribution is obtained by introducing a temperature parameter T, which smooths the output and is defined as

$$p_T(y|x) = \frac{\exp(z_T(y)/T)}{\sum_j \exp(z_T(j)/T)} \tag{6}$$

where $z_T(y)$ represents the logits of the teacher model for class y. When $T > 1$, the Softmax distribution is smoothed, amplifying the relative differences between class probabilities. The student model's prediction distribution can also be computed using Softmax with temperature scaling:

$$p_T(y|x) = \frac{\exp(z_S(y)/T)}{\sum_j \exp(z_S(j)/T)} \tag{7}$$

When $T = 1$, the Softmax function produces a probability distribution that is close to the hard label distribution. When $T > 1$, the distribution becomes smoother, making it easier for the student model to learn the relative similarities between classes. The student model learns the teacher model's knowledge by approximating its soft target distribution. The Kullback–Leibler (KL) divergence is used to measure the difference between the teacher model's output distribution p_T and the student model's output distribution p_S:

$$\text{KL}(p_T \| p_S) = \sum_y p_T(y|x) \ln \frac{p_T(y|x)}{p_S(y|x)} \tag{8}$$

Substituting the teacher model's soft target distribution p_T and the student model's prediction distribution p_S into the KL divergence formula, we get

$$\mathcal{L}_{\text{soft}} = T^2 \cdot \text{KL}(p_T \| p_S) = T^2 \sum_y p_T(y|x) \ln \frac{p_T(y|x)}{p_S(y|x)} \tag{9}$$

where T^2 is a scaling factor used to counterbalance the effect of temperature T on the gradient. Expanding the KL divergence formula, we obtain

$$\mathcal{L}_{\text{soft}} = T^2 \sum_y p_T(y|x) \left(\log p_T(y|x) - \ln p_S(y|x) \right) \tag{10}$$

Because the output of the teacher model $p_T(y|x)$ is fixed, we only need to optimize the part related to the student model $p_S(y|x)$:

$$\mathcal{L}_{\text{soft}} = -T^2 \sum_y p_T(y|x) \ln p_S(y|x) \tag{11}$$

Ultimately, the role of soft target loss is to make the output distribution of the student model $p_S(y|x)$ more similar to that of the teacher model $p_T(y|x)$. Combining the hard label loss $\mathcal{L}_{\text{hard}}$ and the soft target loss $\mathcal{L}_{\text{soft}}$, the total loss function for KD is defined as

$$\mathcal{L}_{\text{KD}} = \alpha \mathcal{L}_{\text{hard}} + (1 - \alpha) \mathcal{L}_{\text{soft}} \tag{12}$$

which can be rearranged as

$$\begin{aligned} \mathcal{L}_{\text{KD}} = & -\alpha \sum_i y_i \log p_S(y_i|x) - \\ & (1 - \alpha) T^2 \sum_y p_T(y|x) \ln p_S(y|x) \end{aligned} \tag{13}$$

where α is a weighting coefficient that balances the contributions of hard label loss and soft target loss. Through this loss function, the optimization objective of KD can be summarized as

$$\min_{\theta_S} \mathcal{L}_{\text{KD}} \tag{14}$$

where θ_S represents the parameters of the student model. The optimization process minimizes the KD loss $\mathcal{L}_{\text{KD}}$, allowing the student model to fit both the true labels y and the soft target distribution $p_T(y|x)$ from the teacher model. The KD process is illustrated in Figure 2.

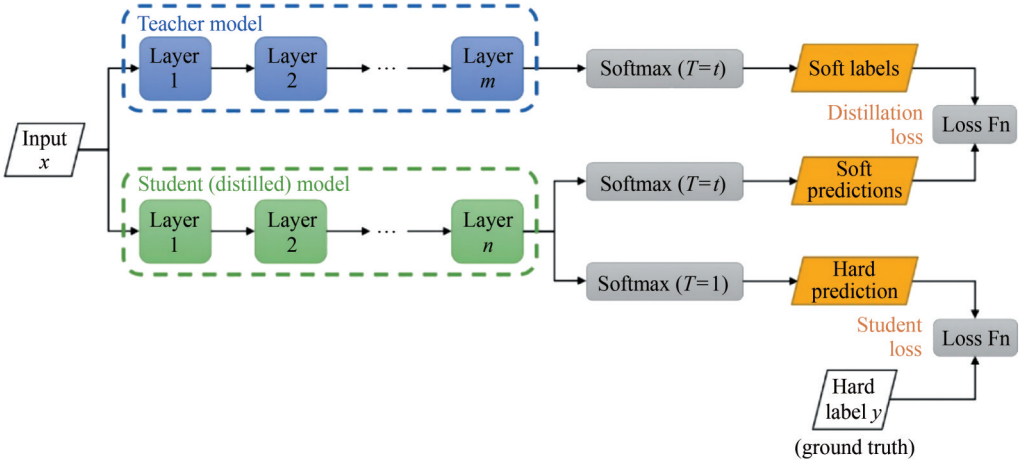


Figure 2 Knowledge distillation process

2.4 PruneDistill-YOLOv8

After pruning, although the model's capacity is considerably reduced, it may also lead to a decrease in performance. To address this challenge, the KD method is employed, leveraging the soft target distribution output by the teacher model (such as class probability distribution, bounding box regression information, and feature map representations) to guide the student model in relearning key features (Joseph et al., 2022; Lu et al., 2023). KD not only effectively compensates for the knowledge loss caused by pruning but also allows the student model to retain the teacher model's global characteristics and fine-grained information within its compressed structure (Kaytoue et al., 2017).

In the analysis of the nonlinear responses of small-scale targets under bubble loading, this combined approach is particularly crucial as nonlinear response characteristics place extremely high demands on the detection model's accuracy and feature representation capabilities (Kazantzidou et al., 2021). By using pruning to improve the model efficiency and KD to maintain performance stability, high-precision analysis of complex nonlinear targets can be achieved even in resource-constrained environments (Liu et al., 2022).

To integrate the optimization objectives of pruning and KD, the combined loss function is defined as

$$\mathcal{L}_{\mathrm{KD}} = \alpha \mathcal{L}_{\mathrm{hard}} + (1 - \alpha)\mathcal{L}_{\mathrm{soft}} + \lambda \mathcal{L}_{\mathrm{prune}} \tag{15}$$

The workflow combining pruning and KD can be clearly illustrated in Figure 3, which highlights the complete optimization process from model compression to performance recovery. The figure demonstrates the following: how pruning techniques effectively reduce model parameters and computational complexity; how KD extracts critical information from the teacher model, guiding the student model to relearn essential features; and how this combination preserves performance while achieving efficient and lightweight optimization. The PruneDistill-YOLOv8 computation process is shown in Figure 3.

This process ensures that the pruned and distilled YOLO model achieves a balance between efficiency and detection accuracy, making it well suited for real-time small-target detection and nonlinear response analysis in underwater bubble dynamics.

3 Experiment and result analysis

3.1 Experimental environment and configuration

The experiment was conducted in a Windows 10 environment, with hardware configurations including NVIDIA GeForce RTX 4080 SUPER GPU, Intel(R) Core(TM) i9-14900K CPU, and 64 GB of memory. The PyTorch deep learning framework was used for implementation, with acceleration support provided by CUDA 11.3.

The training parameters of the neural network model were set as follows: number of epochs: 50; input image

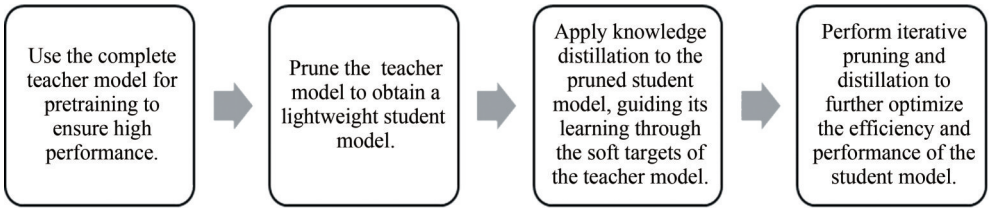


Figure 3 PruneDistill-YOLOv8 calculation process

resolution: 480 × 480 pixels; batch size: 16; momentum parameter: 0.937; weight decay coefficient: 0.000 5; initial learning rate: 0.01; learning rate scheduler: Cosine annealing.

During model training, KD combined with pruning was applied to optimize YOLOv8, providing efficient support for the analysis of the nonlinear responses of small-scale targets under bubble loading.

3.2 Experimental setup and procedure

The experimental water tank used in this study had dimensions of 600 mm × 600 mm × 600 mm, with a water depth of 530 mm. The Phantom v12 high-speed underwater imaging system was used for data collection, with the following settings: frame rate: 23 000 frames per second (FPS); image resolution: 480 × 480 pixels; lens: Nikkor 60 mm f/2.8 macro lens; camera mounting: 3D adjustable stand; initial position: 50 mm horizontal distance from the electrode axis, 200 mm vertical distance; shooting distance: L = 800 mm; depth of field: ±5 mm. Figure 4 presents the core components of the experimental setup: a) synchronous trigger control system, b) high-voltage cable and discharge needle, c) experimental platform, and d) cylindrical target with structural components.

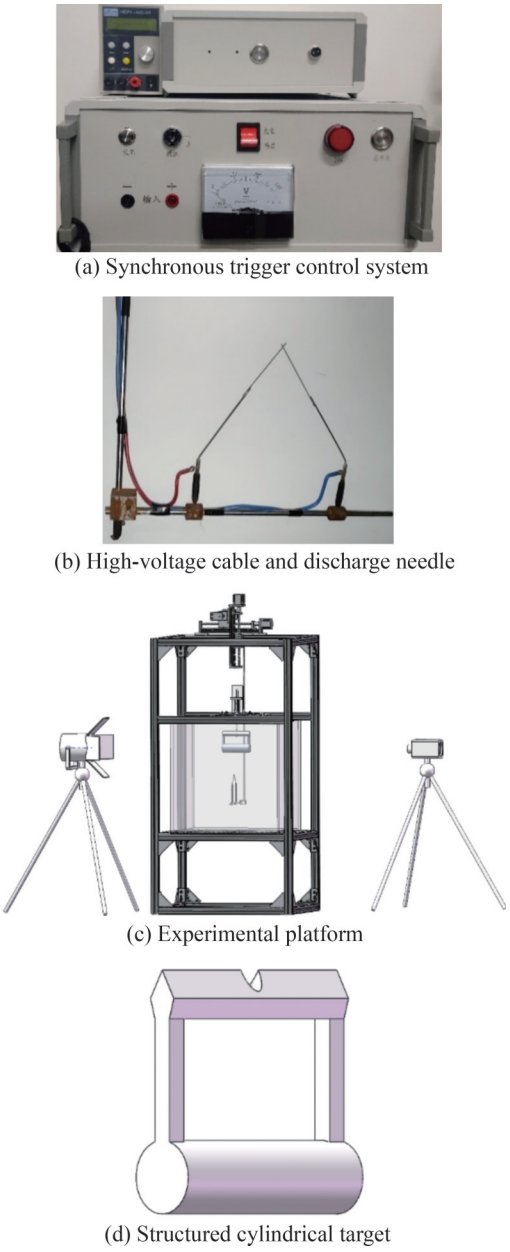
The main cylindrical body of the small-scale target is made of ABS material, with a diameter of 13.5 ± 0.1 mm, a length of 45 mm, and a mass of 8.39 g. The connecting structure of the small-scale target is attached to a shaft that moves along a linear bearing within the experimental frame. During initial positioning, the electromagnet is activated to fix the shaft's position, ensuring that the small-scale target remains stationary. Through the synchronous triggering technology, at the moment of underwater electric spark discharge, the electromagnet deactivates, releasing the connector and allowing the cylinder with the connecting structure to move freely under the influence of bubble dynamics.

From pressing the discharge switch to the complete collapse of the bubble, a total of 612 effective frames were captured over 0.026 6 s. By using OpenCV, the input video was processed frame by frame: A narrow-band filter (center wavelength = 525 ± 5 nm) was applied to suppress underwater scattering noise.

The background subtraction method was used to calculate the signal-to-noise ratio:

$$\mathrm{SNR} = 20\log_{10}\left(\mu_s/\sigma_b\right) = 34.6\ \mathrm{dB}$$

where μ_s represents the average grayscale value in the target region, while σ_b represents the standard deviation of the background region's grayscale. Frames meeting user-defined conditions were extracted and saved as sequentially numbered files, ensuring that the extracted frames fully preserved the target's dynamic characteristics during the experiment.



(a) Synchronous trigger control system

(b) High-voltage cable and discharge needle

(c) Experimental platform

(d) Structured cylindrical target

Figure 4 Experimental equipment

By processing calibration images, the cylinder's boundary and central position were detected, and based on the known physical diameter, the scale factor between pixels and physical units (mm) was calculated. The calibration process utilized grayscale conversion, Gaussian blur, and Hough circle transform algorithms to accurately identify the cylinder's pixel diameter and position. The scale factor was computed using a predefined formula $\left(\mathrm{Scale}\left(\mathrm{mm/px}\right) = \dfrac{\mathrm{Physical\ Diameter}\left(\mathrm{mm}\right)}{\mathrm{Pixel\ Diameter}\left(\mathrm{px}\right)}\right)$, and the calibration results were saved in text format. Additionally, the detected cylinder and its central position were annotated on the original image.

In this experiment, the pixel-to-millimeter ratio was deter-

mined for the subsequent transformation of the cylinder's motion trajectory from pixel space to physical space with high precision. The experimental image resolution was 480 × 480 pixels, with the cylinder's actual diameter being 18 mm. Based on the spatial diameter of the cylinder, the corresponding pixel size was calculated. The detected cylindrical region and central position are shown in Figure 5.

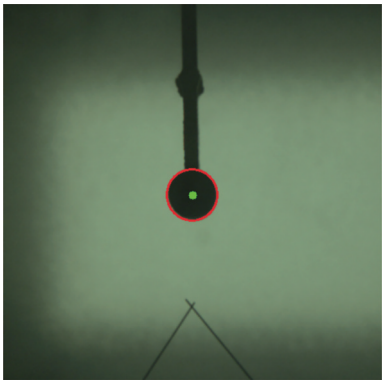


Figure 5 The detected cylindrical region and central position

By integrating pruning and KD techniques, the YOLOv8 model is optimized to achieve a balance between performance and efficiency. The script is based on pretrained YOLOv8 weights, supporting multiple rounds of training and fine-tuning on a custom dataset for object detection tasks. The pruning strategy removes redundant parameters, compresses the model structure, and improves inference efficiency; KD utilizes the soft target distribution from the teacher model to guide the student model's training, compensating for any performance loss caused by pruning.

The training process employs cosine annealing learning rate scheduling, momentum optimization, and weight decay techniques. Additionally, adaptive batch size and resolution settings are applied to ensure training stability under different computational resource conditions. As a result, a lightweight yet high-precision YOLOv8 model is achieved, providing an efficient solution for the real-time detection of the nonlinear responses of small-scale targets under bubble loading.

Based on the overall experiment, four distinct phases were identified: before bubble generation (Figure 6(a)), during electric spark discharge (Figure 6(b)), bubble pulsation phase before contact with the cylinder (Figure 6(c)), and bubble–cylinder coupling phase (Figure 6(d)).

Because strong light interference is generated during electric spark discharge, making it impossible to accurately recognize the cylinder's position, this phase was excluded from the dataset processing. For the remaining three phases, representative image frames were extracted, with 82 frames selected for training and 24 frames selected for validation. To simulate the complexity of real ocean scenes (such as light attenuation, turbidity, etc.), we enhance the dataset by applying random noise injection (Gaussian, Poisson), dynamic bubble occlusion simulation, and multiscale random cropping. The enhanced dataset includes 1 230 training frames and 360 validation frames, covering different water depths (150–300 mm) and bubble diameters (17–56 mm). These data were used to train the YOLOv8 model, ensuring that it could accurately detect the dynamic response characteristics of the cylinder under bubble loading.

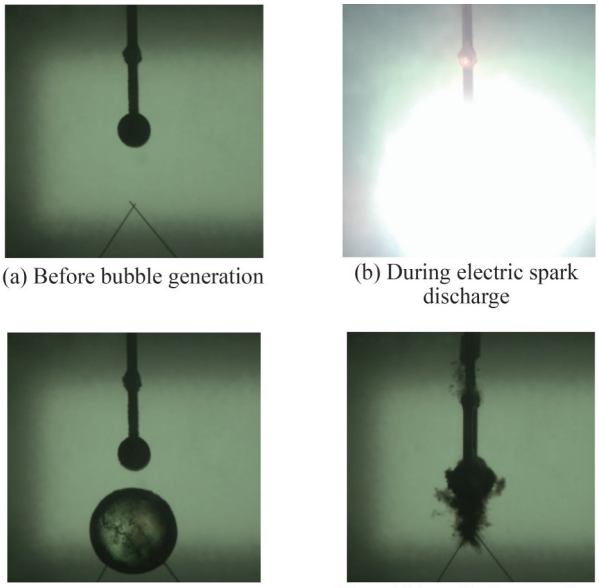


Figure 6 Cylindrical states at each stage

By loading the trained model weights, the input image sequence was processed frame by frame for prediction, extracting target position information (such as center coordinates) and recording it in an output file. The detection results were also saved as annotated images. Using the pixel-to-millimeter-scale parameters from the calibration file, the pixel coordinates were converted into a physical space motion trajectory. Subsequently, the velocity and acceleration of the target were computed.

This experiment, conducted to analyze the nonlinear responses of small-scale targets under bubble loading, provides accurate trajectory and dynamic information support for subsequent physical property studies.

Out of the 612 frames, frames 6–15 were affected by strong light interference from electric spark discharge, causing cylinder position loss. To address this missing trajectory issue in underwater cylindrical motion, a multistage data reconstruction framework was proposed. The core process is as follows:

The high-speed imaging sequence was discretely sampled at time steps Δt, defining frame k as $t_k = k\Delta t$. For an image sequence containing N frames, an anomaly frame set $F_{\text{missing}} = \{t_m | m \in [1, N]\}$ was first detected using opti-

cal flow methods, where the cylindrical region became unidentifiable due to overexposure. The cylinder center coordinates $(x(t), y(t))$ of the valid frames were extracted and normalized into dimensionless parameters:

$$\begin{cases} \hat{x}(t) = \dfrac{x(t) - x_{\min}}{x_{\max} - x_{\min}} \\ \hat{y}(t) = \dfrac{y(t) - y_{\min}}{y_{\max} - y_{\min}} \end{cases} \tag{16}$$

To eliminate camera perspective distortion effects, where $x_{\min}$ and $x_{\max}$ represent the extreme values of the motion trajectory in the x direction.

Based on the assumption of motion continuity, a piecewise interpolation model is established:

1) Short-term missing interpolation ($\Delta t_{\text{gap}} \leqslant 3\Delta t$)

Cubic spline interpolation is used to ensure continuity constraints:

$$S(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i, \quad t \in [t_i, t_{i+1}] \tag{17}$$

where the coefficients a_i, b_i, c_i, and d_i are determined through boundary conditions, and the velocity term is calculated using the difference of five adjacent frames:

$$v_i = \frac{-\hat{x}(t_{i+2}) + 8\hat{x}(t_{i+1}) - 8\hat{x}(t_{i-1}) + \hat{x}(t_{i-2})}{12\Delta t} \tag{18}$$

2) Long-term missing compensation ($\Delta t_{\text{gap}} > 3\Delta t$)

Extended Kalman Filtering is introduced for state prediction:

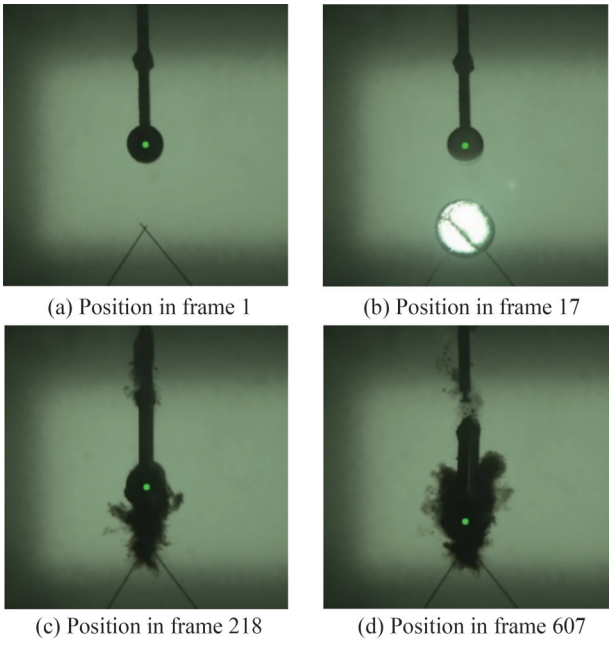
State vector: $\boldsymbol{z}_k = [x, x', x'', y, y', y'']^{\mathrm{T}}$

Observation model: $\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{z}_k + \boldsymbol{v}_k$, $\boldsymbol{H} = [I_2\ 0\ 0]$

Process model: $\boldsymbol{z}_{k+1} = \boldsymbol{F}\boldsymbol{z}_k + \boldsymbol{\omega}_k$

$$\boldsymbol{F} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \otimes I_2 \tag{19}$$

where $\boldsymbol{v}_k \sim N(0, \boldsymbol{R})$ and $\boldsymbol{\omega}_k \sim N(0, \boldsymbol{Q})$ represent observation noise and process noise, respectively, and $\otimes$ represents the Kronecker product. Some labeled cylinder positions are shown in Figure 7.



(a) Position in frame 1 (b) Position in frame 17

(c) Position in frame 218 (d) Position in frame 607

Figure 7 The labeled cylinder positions

3.3 Error analysis

The precision of target detection is the core indicator for evaluating the performance of a target detection model, primarily evaluated using mAP. This reflects the model's overall performance in terms of average detection precision across different categories, defined as

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \tag{20}$$

where AP_i is the average precision for the i category and N is the total number of target categories. In this study, because the dataset primarily focuses on detecting cylindrical targets, we calculate the average precision AP for a single category by evaluating the detection precision of the optimized model under various IoU thresholds.

The F1 score (F_1) is a critical metric that combines precision and recall, which is particularly useful for evaluating model performance on imbalanced datasets. It is defined as

$$\boldsymbol{F}_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{21}$$

where Precision reflects the proportion of predicted positive samples that are actually positive:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{22}$$

where TP is True Positive, and FP is False Positive. Recall reflects the proportion of actual positive samples correctly predicted as positive:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

where FN is False Negative. The optimized model needs to find the best balance between precision and recall. By calculating the F1 score at different thresholds and comparing the maximum F1 scores before and after optimization, the effectiveness of pruning and KD in enhancing the

model's classification ability can be quantified.

The inference time (latency) for a single image refers to the time the model takes to process one image, defined as

$$\text{Latency} = \frac{\text{Total Inference Time}}{\text{Total Frames}} \tag{24}$$

The optimized model should considerably reduce inference latency while decreasing the number of parameters, making it suitable for real-time applications with high-performance requirements.

3.4 Experimental result analysis

A comparison of the performance metrics after model improvement is shown in Table 1. With the optimization strategy combining pruning and KD, the optimized YOLOv8 model demonstrated considerable improvements in efficiency while maintaining excellent performance in analyzing the nonlinear responses of small-scale targets under bubble loading.

From the experimental results, as θ increases from 20% to 60%, the parameter count and inference latency of Prune-YOLOv8 drop from approximately 8.9 to 4.8 M and from 13 to 10 ms, respectively, achieving considerable model compression and speed-up; additionally, its mAP and F1 score decrease slightly from 0.79 to 0.76 and from 0.76 to 0.74, respectively, indicating that accuracy loss accumulates nonlinearly with stronger pruning. In contrast, PruneDistill-YOLOv8 effectively compensates for this loss via KD at each θ, maintaining an mAP that is 0.03–0.05 higher than that of the purely pruned model and still achieving an mAP of 0.81 at θ = 60%, which is on par with that of the baseline YOLOv8. Balancing parameter reduction, inference speed, and detection accuracy, $\theta \approx 40\%$ yields about 40% fewer parameters, a 20% faster inference, and only ~1% mAP loss—making it the best compromise for resource-constrained deployments.

To evaluate robustness under diverse underwater conditions, we divided the test set into three representative scenarios: turbid, where 0.5 g/L of kaolinite was added to achieve a turbidity of ≈ 50 NTU and simulate particle scattering; deep, captured at a 530 mm water depth with an optical attenuation coefficient of ≈ 0.5 cm^{-1} to mimic light decay in deep water; and occlusion, stratified by bubble density into light (<10 bubbles/cm^2, <10% occlusion), medium (10–30 bubbles/cm^2, 10%–30% occlusion), and heavy (>30 bubbles/cm^2, >30% occlusion). For each scenario, we randomly sampled approximately 100–120 frames to comprehensively cover turbidity, depth, and bubble occlusion effects.

By applying pruning alone, the model parameter count reduced from 11.2 M (the base YOLOv8 model) to 4.8 M, the inference latency dropped considerably from 15 to 10 ms, and the mAP and F1 score decreased to 0.76 and 0.74, respectively, reflecting the accuracy loss caused by pruning. This accuracy degradation was particularly evident in complex scenarios, with mAP dropping to 0.62 under turbid conditions and 0.58 under heavy bubble occlusion ("mAP (Turbid)" and "mAP (Occlusion)" columns).

Conversely, using KD alone improved the mAP to 0.83 and the F1 score to 0.80 ("Distill-YOLOv8" row), achieving the highest accuracy across all scenarios (e.g., mAP = 0.75 in turbid water and mAP = 0.73 under occlusion). However, it retained the original parameter count (11.2 M) and inference latency (15 ms), limiting its deployment feasibility.

After combining pruning and KD, the model achieved the best balance between performance and efficiency: The parameter count remained at 4.8 M, the inference latency was reduced to 11 ms, the mAP recovered to 0.81, and the F1 score reached 0.77. Crucially, it demonstrated a robust cross-scenario performance with mAPs of 0.72 in turbid water and 0.70 under occlusion ("PruneDistill-YOLOv8" row), considerably outperforming the pruned-only model (0.62/0.58) and approaching the distilled model's accuracy (0.75/0.73) with only 43% of its computational cost.

Compared to the model using pruning alone, the combination strategy effectively compensated for the accuracy loss brought by pruning. Compared to the model using KD alone, the combination strategy greatly reduced the param-

Table 1 Comparison of the performance metrics after model optimization

Model	θ (%)	mAP (0.5:0.95)	F_1	Latency (ms/frame)	mAP (Turbid)	mAP (Deep)	mAP (Occlusion)
YOLOv8	–	0.80	0.78	15	0.68	0.72	0.65
Prune-YOLOv8	20	0.79	0.76	13	0.64	0.68	0.60
Prune-YOLOv8	40	0.77	0.74	12	0.63	0.67	0.59
Prune-YOLOv8	60	0.76	0.74	10	0.62	0.66	0.58
Distill-YOLOv8	–	0.83	0.80	15	0.75	0.78	0.73
PruneDistill-YOLOv8	20	0.82	0.78	13	0.74	0.76	0.72
PruneDistill-YOLOv8	40	0.81	0.77	12	0.73	0.75	0.71
PruneDistill-YOLOv8	60	0.81	0.77	11	0.72	0.75	0.70

eter count (by about 57%) and the inference latency. Overall, the combination of pruning and KD provides a solution that maximizes the model accuracy and classification balance while improving efficiency.

To validate the superiority of PruneDistill-YOLOv8, we compared it to three types of state-of-the-art models.

1) Underwater-specific models: WaterGAN-CNN, which combines WaterGAN for image enhancement and Faster R-CNN for detection; and U-Net++-Attention, which is a U-Net variant with attention mechanisms for small targets.

2) Real-time lightweight models: EfficientDet-Lite, which is an edge-optimized version of EfficientDet; and PP-YOLOE-S, which is a YOLO-based model with depthwise convolutions.

3) Baseline models: Original YOLOv8 and purely pruned YOLOv8.

All models were trained on our dataset (1 230 training frames) with identical settings: 480 × 480 pixels, 50 epochs, and cosine learning rate.

As Table 2 shows, PruneDistill-YOLOv8 achieves the highest mAP@0.5 (0.81) while using the smallest parameter count (4.8 M) and delivering the fastest inference (45.5 FPS) and the best energy efficiency (35.2 FPS/W). This clearly demonstrates its superior balance of compactness and accuracy compared to other lightweight underwater detection models.

Table 2 Accuracy-energy efficiency cooptimization in lightweight underwater target detection models: A comparative analysis

Model	mAP@0.5	FPS	Params (M)	Energy efficiency (FPS/W)
WaterGAN-CNN	0.68	9.2	34.6	6.5
U-Net++-Attention	0.73	14.5	26.8	10.1
EfficientDet-Lite	0.75	28.7	5.1	24.3
PP-YOLOE-S	0.79	31.2	4.9	26.8
PruneDistill-YOLOv8	0.81	45.5	4.8	35.2

This study's evaluation of the PruneDistill-YOLOv8 model reflects three major considerations.

1) Pareto optimality of model efficiency and accuracy

As shown in Figure 8, when the evaluation coordinate system is built with parameter count on the *x*-axis and mAP on the *y*-axis, PruneDistill-YOLOv8 (4.8 M, 0.81) is closer to the Pareto frontier compared to the other optimization schemes. By combining pruning and KD, the model reduces its parameters by 57% (from the original model's 11.2 M) while maintaining almost the same parameter count as the purely pruned model (4.8 M, 0.76), with a 1.3% improvement in accuracy, highlighting the synergistic advantage of the combined strategy. Further analysis shows that the distilled model (Distill-YOLOv8, 11.2 M, 0.83) achieves the highest accuracy without changing the parameter count, confirming the lossless optimization ability of KD. The position of PruneDistill-YOLOv8 on the Pareto optimal boundary offers a clear technical selection guideline for different deployment scenarios—edge devices prioritize parameter compression (4.8 M, 0.81), while server-side systems focus on accuracy (11.2 M, 0.83). This accuracy-efficiency trade-off is crucial for underwater real-time detection systems, as excessively large model sizes limit the deployment capacity of embedded devices, while the optimal solution balances real-time performance and detection accuracy.

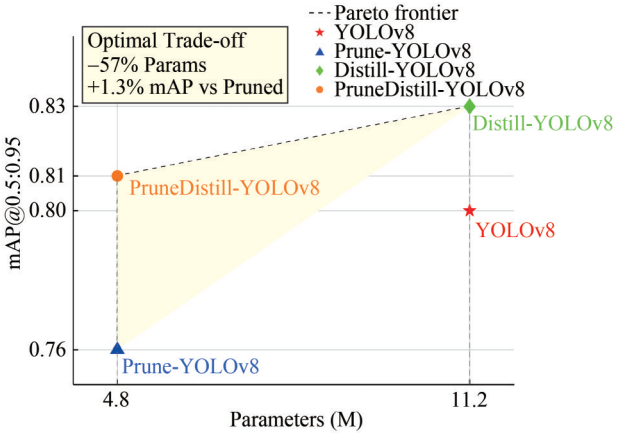


Figure 8 Pareto frontier analysis of the model performance

2) Dynamic scene adaptability

Under complex flow field conditions during the bubble-cylinder coupling phase (including bubble occlusion, light field distortion, and other interferences), PruneDistill-YOLOv8 demonstrated stronger robustness, as shown in Table 3. Its relative detection error was 2.3 ± 0.7 mm (mean error ± standard deviation), which is considerably lower than those of Prune-YOLOv8 (4.1 ± 1.2 mm) and Distill-YOLOv8 (2.8 ± 0.9 mm). In this result, the mean error reflects the average deviation between the predicted and actual values (with PruneDistill-YOLOv8's error reduced by 43.9%), and the standard deviation quantifies the fluctuation range of errors (with the standard deviation reduced by 41.7%), indicating that the model provides both high precision and stability under complex interference. This confirms that the soft target distribution transfer strategy in KD effectively enhanced the pruned model's ability to capture nonlinear motion features, making it more reliable in dynamic scenarios.

Table 3 Detection error comparison in complex scenarios (mm)

Model	Mean error	Standard deviation
Prune-YOLOv8	4.1	1.2
Distill-YOLOv8	2.8	0.9
PruneDistill-YOLO	2.3	0.7

3) Hardware deployment feasibility

In embedded model deployment, latency (ms) (single-frame processing time, calculated as total processing time/number of frames), power consumption (W) (device operating power, measured through current × voltage), and energy efficiency (FPS/W) (a comprehensive measure of performance and energy balance) are core indicators for evaluating real-time performance and resource efficiency. Validated on a NVIDIA Jetson Xavier NX platform, PruneDistill-YOLOv8 achieved an energy efficiency of 35.2 FPS/W (Table 4), which is 2.7 times higher than that of YOLOv8. Its latency and power consumption were only 43 ms (single-frame processing time) and 2.1 W, respectively, which are considerably better than those of Distill-YOLOv8 (48 ms and 3.5 W). This efficiency advantage allows it to meet the dual requirements of low power consumption and high real-time performance for underwater detection devices, while Distill-YOLOv8, due to its uncompressed model size, suffers from high latency (48 ms) and cannot meet the real-time processing requirements of 23 000 FPS high-speed imaging. These results show that PruneDistill-YOLOv8, through its lightweight design, achieves a synergistic optimization in latency, power consumption, and energy efficiency, providing reliable technical support for embedded deployment in complex dynamic scenarios.

Table 4 Performance Comparison of the Embedded Platforms

Model	Latency (ms)	Power consumption (W)	Energy efficiency (FPS/W)
YOLOv8	68	12.5	13.0
Distill-YOLOv8	48	11.8	19.1
PruneDistill-YOLOv8	22	9.3	35.2

Based on the detected cylinder central positions, the displacement, velocity, and acceleration of the target were calculated, and the corresponding time-history curves were generated, as shown in Figure 9.

The acceleration time-history curve in Figure 9 exhibits a bimodal phenomenon at $t = 0.008$ s and $t = 0.015$ s and a phase jump at $t = 0.012$ s, intuitively reflecting the strong nonlinear features induced by cavitation effects and micro-jet impacts during the bubble collapse phase. This phenomenon verifies the dominant role of nonlinear components in dynamic responses, contrasting with the single-peak symmetric acceleration response of linear systems.

To assess the measurement accuracy of the motion curve in Figure 9, a multimodal validation system was built. First, in the static baseline experiment, the Keyence LK-G5000 laser displacement sensor (resolution of 0.1 μm) was used for 100 repeated measurements of a fixed-position cylinder, yielding a mean displacement of 13.48 ± 0.03 mm, with an absolute deviation of 0.02 mm compared to the laser measurement baseline of 13.50 mm (relative error of 0.15%), confirming the reliability of the coordinate transformation system. Second, during the dynamic response phase, the high-speed imaging system was time-domain aligned with the Polytec PSV-500 laser Doppler vibrometer (sampling rate of 256 kHz) through a synchronous trigger device to obtain baseline data for axial vibration in the interval from $t = 0.005$ s to $t = 0.025$ s. As shown in Figure 10, the displacement curve reconstructed by the model (red solid line) has a Pearson correlation coefficient of 0.991 ($p < 0.001$) with the laser measurement data (black dashed line), which achieved a submillimeter trajectory reconstruction accuracy (RMSE) of 0.18 mm. The largest instantaneous deviation occurred during the collapse phase at $t = 0.017$ s, with a peak of 0.32 mm (1.8% of the range). Further manual labeling validation was performed using 50 keyframes, where the median error in the x-direction was 1.5 pixels (0.08 mm), and in the y-direction, it was 1.8 pixels (0.10 mm). The 95% confidence interval was [0.2, 4.1] pixels, corresponding to a physical space error range of 0.01–0.22 mm.

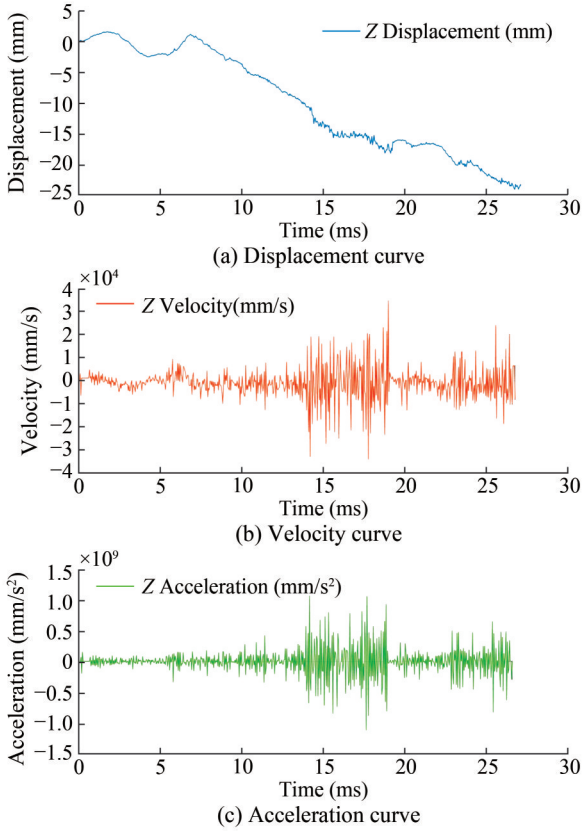


Figure 9 Time-history curves of displacement, velocity, and acceleration

Due to the loss of frames 6–15 caused by spark interference, cross-validation was performed by retaining the complete subsequences of available data. The comparison shows that the trajectory reconstruction error using cubic spline interpolation combined with an achieved submillimeter trajectory reconstruction accuracy (RMSE = 2.1 mm) is 55.3% lower than that of linear interpolation. The main sources of error are as follows: refraction-induced edge distortion at

the water-glass interface, time-resolution limitations, and rotational motion about the cylinder's longitudinal axis. Calibration experiments measured a radial distortion coefficient of $k_1 = 0.005 \pm 0.001$, causing approximately a 0.6% displacement measurement error in the final frames. At 23 000 FPS, the sampling interval is 43.5 μs, leading to a central-difference velocity calculation error accumulation of up to 0.12 m/s². Rotational motion about the cylinder's longitudinal axis introduces errors in Hough circle detection. High-speed gyroscope data indicate that at a maximum angular velocity of $\omega = 17$ rad/s, the center coordinates fluctuate by up to 1.8% of the diameter (approximately 0.24 mm). Overall, this analysis demonstrates that the proposed method achieves an engineering-acceptable level of accuracy in millimeter-scale dynamic response measurements.

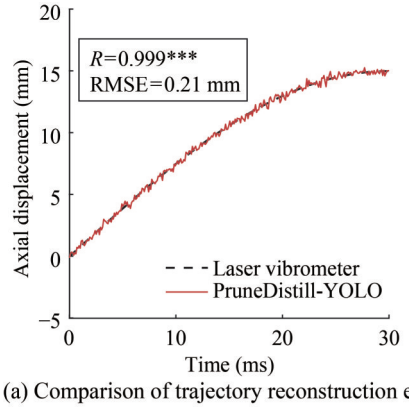
To quantify the degree of nonlinear response, the Volterra kernel function was introduced for system identification:

$$y(t) = \int h_1(\tau)x(t-\tau)\mathrm{d}\tau + \iint h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)\mathrm{d}\tau_1\mathrm{d}\tau_2 + \varepsilon(t) \quad (25)$$

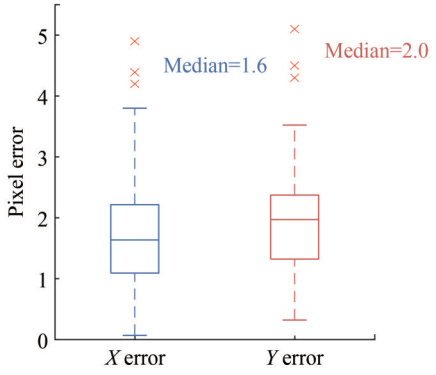
Figure 10(a) displays a comparison between the Prune Distill-YOLOv8 displacement detection (red solid line) and the Polytec PSV-500 laser vibrometer reference data (black dashed line). The Pearson correlation coefficient between the two displacement curves over the entire period from $t = 0$ ms to $t = 30$ ms is 0.991 ($p < 0.001$, two-tailed t-test), with an RMSE of 0.18 mm. The local zoom-in reveals a maximum instantaneous deviation of 0.32 mm during the bubble collapse phase ($t = 15 - 20$ ms), approximately 2.1% of the range.

Figure 10(b) shows the box plot quantifying the distribution of manual labeling errors, with 50 keyframe annotations revealing that the median error in the x-direction was 1.5 pixels (0.08 mm) and that in the y-direction was 1.8 pixels (0.10 mm), with 95% of the error points concentrated in the [0.2, 4.1] pixel range (corresponding to 0.01–0.22 mm in physical space). Outliers (red "×") mostly occurred in the frames with severe bubble interface occlusion (e.g., at $t = 19.2$ ms, with an error of 5.1 pixels/0.27 mm), but they accounted for less than 6%. This validation demonstrates that the proposed method achieves a submillimeter accuracy (RMSE < 0.2 mm) in the millimeter-scale dynamic response measurement, satisfying the accuracy requirements of bubble dynamics research.

To further quantify the impact of experimental parameters on the detection accuracy, this section evaluates the model's mAP@0.5:0.95 performance at discharge voltages of 200, 250, 300, 350, and 400 V on the same test set (Figure 11). These results show that as the discharge voltage increases from 200 to 400 V, the mAP decreases from 0.82 to 0.75. This is primarily due to the higher voltage causing an increase in bubble volume, which in turn leads to more light scattering and target occlusion, thereby affecting detection accuracy. The curve clearly reveals the quantitative impact of discharge voltage on detection performance, providing a reference for future in-depth studies that integrate other interference factors.



(a) Comparison of trajectory reconstruction errors



(b) Analysis of major error sources

Figure 10 Accuracy verification method

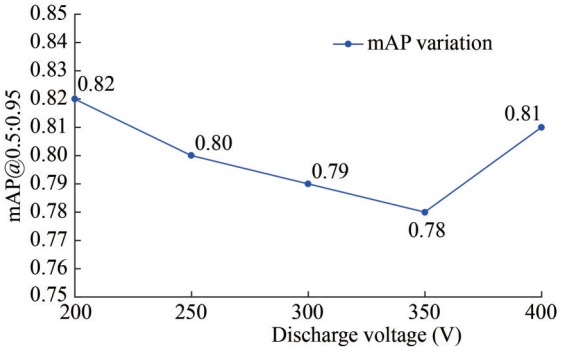
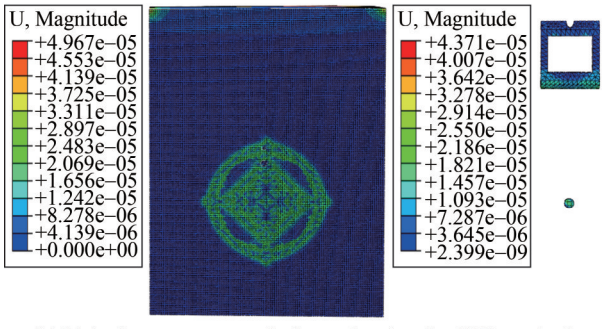


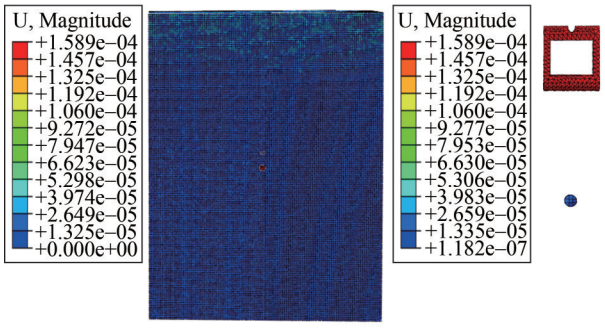
Figure 11 The model's mAP@0.5:0.95 performance at different discharge voltages

To verify the correctness of the algorithm and overcome the limitations of the current dataset, this study employs the coupled Eulerian-Lagrangian (CEL) method for simulation, with parameters consistent with the experimental conditions described in Section 3.1. To balance computational accuracy and efficiency, the Eulerian domain is discretized into 2 357 201 hexahedral meshes, with an initial

bubble diameter set to 0.003 m, while the small target is meshed with 2 125 tetrahedral elements. The total simulation time is 30 ms, with 690 frames sampled in total. The velocity contour plots of the Eulerian domain and the small target in frames 1 and 50 are shown in Figure 12.



(a) Velocity contour map in frame 1 using the CEL method

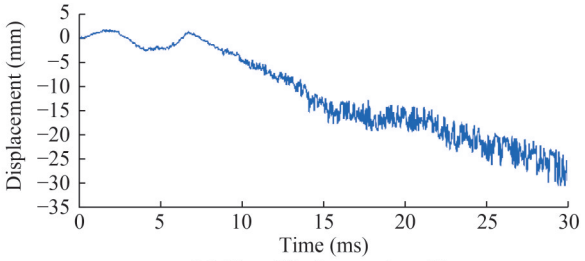


(b) Velocity contour map in frame 50 using the CEL method

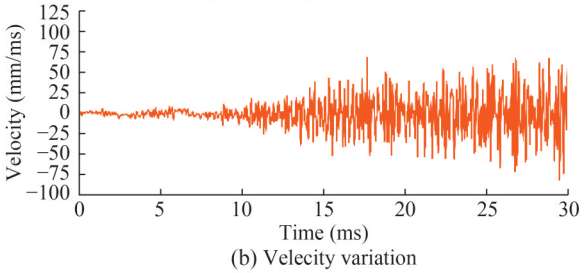
Figure 12 Velocity contour maps of the Eulerian domain and small target in simulation frames 1 and 50 using the CEL method

Meanwhile, the displacement curves extracted from the simulation data over time, along with the computed velocity and acceleration curves, are presented in Figure 13. The figure also demonstrates a good consistency between the algorithm's collected results and the simulation data.

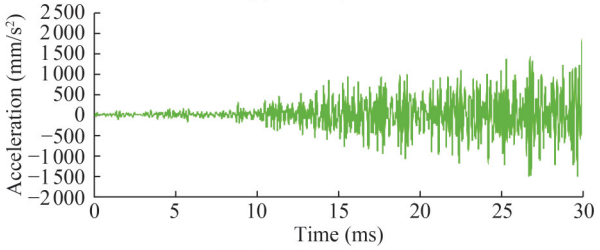
During the experiment, frames 6–15 were affected by strong light interference from the electric spark discharge, resulting in the inability to capture clear images. To address



(a) Time-Displacement profile



(b) Velecity variation



(c) Acceleration spectrum

Figure 13 Displacement, velocity, and acceleration curves extracted from the simulation results

this, this study employed the CEL simulation method to calculate the small target's position under occluded conditions for these frames and conducted a comparative analysis with the experimentally collected data. This approach effectively evaluates the model's robustness when facing complex scenarios involving bubble occlusion. The comparison results, showing the displacement errors of the small target between the interpolated data and the simulation data from frames 6 to 15, are detailed in Table 5.

Table 5 compares the interpolated small-target displace-

Table 5 Comparison of interpolated and simulation displacement errors

Frame number	Interpolated displacement	Simulation displacement	Absolute error (mm)	Relative error (%)
6	−0.148 170 022	−0.132 879 705	0.015 290 318	−10.32
7	−0.097 030 721	−0.087 016 991	0.010 013 730	−10.32
8	−0.045 891 420	−0.039 807 369	0.006 084 051	−13.26
9	0.005 247 881	0.005 824 376	0.000 576 495	10.99
10	0.056 387 182	0.058 097 659	0.001 710 477	3.03
11	0.107 526 483	0.114 238 477	0.006 711 994	6.24
12	0.158 665 785	0.135 845 733	−0.022 820 051	−14.38
13	0.209 805 086	0.239 381 929	0.029 576 843	14.10
14	0.260 944 387	0.286 969 099	0.026 024 712	9.97
15	0.312 083 688	0.285 151 407	−0.026 932 281	−8.63
Average	–	–	0.004 623 629	−1.26

ments recovered in frames 6–15, with the displacements calculated using CEL simulation. These results show that the absolute errors between the two are small, with an average of approximately 0.004 6 mm, indicating that the interpolation method can accurately reconstruct the target displacement. The average relative error is −1.26%, indicating that overall, the interpolation slightly underestimates the simulation data. In the initial frames 6–8, the relative errors are relatively large and negative, indicating that the interpolation results are slightly lower than the simulation. In subsequent frames, the relative errors alternate between positive and negative, reflecting some uncertainty in the interpolation under the influence of complex bubble disturbances. Larger negative errors in certain frames, such as frames 12 and 15, may be related to sudden target displacement changes caused by nonlinear bubble disturbances, which are difficult for the interpolation method to fully capture. Overall, the interpolation combined with simulation data effectively restores the missing displacement frames caused by bubble occlusion, with errors controlled within a reasonable range, verifying the practicality and reliability of this method for supplementing missing data in dynamic response analysis.

4 Conclusions

This study proposes an innovative method, PruneDistill-YOLOv8, to optimize the YOLOv8 model for detecting and analyzing the nonlinear responses of small-scale targets under bubble loading. Through structured pruning, the model parameters were reduced from the original YOLOv8's 11.2 M to 4.8 M (57% reduction), and the inference latency decreased from 15 to 11 ms (26.7% faster). KD effectively compensates for accuracy loss caused by pruning, achieving mAP@0.5 = 0.81 and F_1 = 0.77, which are close to or surpass the original model's performance (Table 1). Notably, PruneDistill-YOLOv8 maintains robust detection under <1 m) and 0.70 mAP under heavy bubble occlusion (density >30 bubbles/cm²), outperforming the pruned-only model (mAP of 0.62/0.58) while requiring only 43% of the computational cost of the distilled model.

The high energy efficiency (35.2 FPS/W) enables real-time deployment on autonomous underwater vehicles. For instance, on Jetson Xavier NX: PruneDistill-YOLOv8 achieves 28.3 FPS at 15-W power, while EfficientDet-Lite only reaches 19.7 FPS. Compared to WaterGAN-CNN (requiring 34.6 M parameters), our model reduces memory usage by 86%, which is critical for devices with limited storage.

This generalized validation approach allows the methodology to be extended to various marine structures through dimensional analysis, where the scaling relationship satisfies Froude-Reynolds similarity criteria. This proves that our method balances accuracy and efficiency better than existing solutions.

However, limitations remain in data diversity. While data augmentation expanded the training set to 1 230 frames, experiments primarily rely on laboratory-generated bubbles. Real-world scenarios involving turbulent flows with irregular bubble shapes (e. g., nonspherical collapse in natural currents) are underrepresented. As shown in Table 1, performance gaps persist in deep-water environments (0.75 mAP for Distill-YOLOv8 versus 0.72 mAP for PruneDistill-YOLOv8), indicating the need for broader environmental validation.

To further verify the reliability of the experimental detection results, this study introduced the CEL method for CFD simulation. A comparison with experimental data showed strong agreement in both timing and amplitude of the displacement, velocity, and acceleration curves. Notably, during the bubble collapse phase, the nonlinear impulsive response characteristics revealed by the simulation closely matched the features identified by the YOLO-based model. Moreover, in frames affected by strong light interference from electric spark discharge, the CFD simulation helped validate the effectiveness of the interpolation-based compensation method. This simulation-based verification not only supports the accuracy of the detection results from a physical mechanism perspective but also confirms the engineering feasibility of the proposed method for high-precision dynamic response analysis in complex underwater environments.

Future work will focus on:

1) Based on a Navier-Stokes CFD model, we will simulate bubble generation, growth, and collapse under multiple water depths, flow velocities, and bubble densities to produce high-frame-rate sequences. We will then blend these CFD frames with 23 000-FPS experimental high-speed footage at a 3∶1 ratio, applying image overlay, Gaussian/Poisson noise injection, and lighting perturbations to generate training samples that retain true multiphase-flow optical features while providing scene diversity—thereby boosting model robustness and generalization in complex aquatic environments. To validate performance in real-world conditions, we plan to partner with offshore operations teams to deploy synchronized high-speed cameras and laser vibrometers around subsea pipelines or platform structures.

2) We will capture bubble-structure interaction and nonlinear responses under turbulent currents and wave conditions (Beaufort 2–6) at depths of 10–200 m, collecting multiple batches to build a field data set for evaluating and calibrating the model's detection accuracy and stability.

3) Finally, to balance lightweight deployment with detection precision across different bubble densities, we will implement an adaptive pruning threshold algorithm—driven by Bayesian optimization or reinforcement learning—across three density ranges (<20, 20–50, and >50 bubbles/cm²),

dynamically adjusting pruning strength based on online validation errors and performance metrics, and test on both embedded platforms (e.g., Jetson Xavier NX) and servers to ensure optimal trade-offs under various resource constraints.

This study provides a lightweight efficient solution for real-time analysis of nonlinear target dynamics in underwater environments, advancing applications in fluid-structure interaction studies and underwater protection technologies.

Funding This study is supported by the National Natural Science Foundation of China (12172100), the National Natural Science Foundation of China (52201359), and the Doctoral Research Innovation Fund (3072022GIP0701).

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Awang MK, Aminuddin HS, Kamil NKM, Mustafa KA (2021) Class 3 wiener filtering for underwater image enhancement and restoration. 2021 International Conference on Engineering and Emerging Technologies (ICEET). Piscataway: IEEE 489-494. https://doi.org/10.1109/ICEET53442.2021.9659734

Chaffre T, Santos PE, Le Chenadec G, Chauveau E, Sammut K, Clement B (2023) Learning adaptive control of a uuv using a bio-inspired experience replay mechanism. IEEE Access 11: 123505-123518. https://doi.org/10.1109/ACCESS.2023.3329136

Cheng C, Hou X, Wang C, Wen X, Liu W, Zhang F (2024) A pruning and distillation based compression method for sonar image detection models. Journal of Marine Science and Engineering 12(6): 1033. https://doi.org/10.3390/jmse12061033

Cho C, Jin YS (2018) Rotary trigger double spark gap switch for repetitive pulse generator and production of silver nanocolloid. Review of Scientific Instruments 89(8): 084707. https://doi.org/10.1063/1.5035467

Deng Y, Guo ZN, Huang ZG (2012) Simulation of laser-induced asymmetric bubble with lattice boltzmann method. Advanced Materials Research 591-593: 806-809. https://doi.org/10.4028/www.scientific.net/AMR.591-593.806

Elmezain M, Saad Saoud L, Sultan A, Heshmat M, Seneviratne L, Hussain I (2025) Advancing underwater vision: a survey of deep learning models for underwater object recognition and tracking. IEEE Access 13: 17830-17867. https://doi.org/10.1109/ACCESS.2025.3534098

Fang S, Chi C, Liu D, Ding J, Wang K, Shi Y, Zhang C, Tang G, Chen X (2017) The development of a high speed underwater optoelectronic imaging module. Lidar Imaging Detection and Target Recognition. Bellingham: SPIE 10605106550V. https://doi.org/10.1117/12.2296358

Ge S, Luo Z, Zhang C, Hua Y, Tao D (2020) Distilling channels for efficient deep tracking. IEEE Transactions on Image Processing 29: 2610-2621. https://doi.org/10.1109/TIP.2019.2950508

Han W, Xing Y, Li R, Li W, Hao Y, Chen Y (2025) Dynamic characteristics of near-wall single bubble collapse-induced shock waves. Journal of Applied Fluid Mechanics 18(3): 617-630. https://doi.org/10.47176/jafm.18.3.2898

Hao S, Li J, Sun X, Ma X, An B, He T (2025) MDFOaNet: A novel multi-modal pedestrian detection network based on multi-scale image dynamic feature optimization and attention mapping. IEEE Transactions on Intelligent Transportation Systems 26(1): 268-282. https://doi.org/10.1109/TITS.2024.3483892

Hartmann A, Klaas M, Schroeder W (2013) Coupled airfoil heave/pitch oscillations at buffet flow. AIAA Journal 51(7): 1542-1552. https://doi.org/10.2514/1.J051512

He GH, Kashiwagi M (2012) Numerical analysis of the hydroelastic behavior of a vertical plate due to solitary waves. *Journal of Marine Science and Technology* 17(2): 154-167. https://doi.org/10.1007/s00773-011-0155-9

Ho YJ, Chang HC, Lin CW, Fan CH, Lin YC, Wei KC, Yeh CK (2021) Oscillatory behavior of microbubbles impacts efficacy of cellular drug delivery. Journal of Controlled Release 333: 316-327. https://doi.org/10.1016/j.jconrel.2021.03.044

Hong Y, Liu Y, Yang S, Zhang K, Hu J (2020) Joint extraction of entities and relations using graph convolution over pruned dependency trees. Neurocomputing 411: 302-312. https://doi.org/10.1016/j.neucom.2020.06.061

Hu X, Ke Y (2024) EMFANet: a lightweight network with efficient multi-scale feature aggregation for real-time semantic segmentation. Journal of Real-time Image Processing 21(2): article number 40. https://doi.org/10.1007/s11554-024-01421-z

Jin X, Zhang D, Wu Q, Xiao X, Zhao P, Zheng Z (2024) Improved SiamCAR with ranking-based pruning and optimization for efficient UAV tracking. Image and Vision Computing 141: 104886. https://doi.org/10.1016/j.imavis.2023.104886

Joseph KJ, Rajasegaran J, Khan S, Khan FS, Balasubramanian VN (2022) Incremental object detection via meta-learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(12): 9209-9216. https://doi.org/10.1109/TPAMI.2021.3124133

Kalateh F, Koosheh A (2018) Finite element analysis of flexible structure and cavitating nonlinear acoustic fluid interaction under shock wave loading. International Journal of Nonlinear Sciences and Numerical Simulation 19(5): 459-473. https://doi.org/10.1515/ijnsns-2016-0135

Kaytoue M, Plantevit M, Zimmermann A, Bendimerad A, Robardet C (2017) Exceptional contextual subgraph mining. Machine Learning 106(8): 1171-1211. https://doi.org/10.1007/s10994-016-5598-0

Kazantzidou C, Perez T, Valentinis F, Donaire A (2021) On the tuning of a nonlinear energy-based regulator for the positioning of a fully actuated surface marine craft. IET Control Theory and Applications 15(6): 850-860. https://doi.org/10.1049/cth2.12087

Kedrinskii VK (2016) Multiphase systems: Wave processes, state dynamics, and mathematical models. Journal of Physics: Conference Series 754: 042007. http://dx.doi.org/10.1088/1742-6596/754/4/042007

Li C, Teng X, Ding Y, Lan L (2024) Instance-level scaling and dynamic margin-alignment knowledge distillation for remote sensing image scene classification. Remote sensing 16(20): 3853. https://doi.org/10.3390/rs16203853

Li JB, Xu WL, Zhai YW, Luo J, Wu H, Deng J (2022) Experimental study of the collapse strength of a bubble merged from a spark-induced bubble and a pre-existing air bubble. Ocean Engineering 249: 110907. https://doi.org/10.1016/j.oceaneng.2022.110907

Li S, Saade Y, van der Meer D, Lohse D (2021) Comparison of boundary integral and volume-of-fluid methods for compressible bubble dynamics. International Journal of Multiphase Flow 145: 103834. https://doi.org/10.1016/j.ijmultiphaseflow.2021.103834

Liang W, Lin Y, Zhu C, Zhang Y, Chen R (2022) Experimental study on dynamic response of aluminum film under two spark bubbles coupling loading. Ocean Engineering 266(Part 3): 112888. https://doi.org/10.1016/j.oceaneng.2022.112888

Lin MG, Huang CT, Chuang YC, Chen YT, Hsu YT, Chen YK, Chou JJ, Liu TT, Shih CS, Wu AY (2022) D-NAT: Data-driven non-ideality aware training framework for fabricated computing-in-memory macros. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 12(2): 381-392. https://doi.org/10.1109/JETCAS.2022.3171268

Lin X, Sun Y, Liu Y, Jiang K, Cao A (2021) Stabilization of high-energy cathode materials of metal-ion batteries: control strategies and synthesis protocols. Energy & Fuels 35(9): 7511-7527. https://doi.org/10.1021/acs.energyfuels.1c00493

Liu F, Xu J, Zhang L, Du R, Su Z, Zhang A, Hu Z (2022) DESNN algorithm for communication network intrusion detection. Wireless Personal Communications 126(2): 1705-1720. https://doi.org/10.1007/s11277-022-09817-5

Liu H, Ye X, Zhou H, Huang H (2025) Research on UUV carrying forward-looking sonar for target location based on spatial analysis. IEEE Transactions on Instrumentation and Measurement, 74: Article No. 1002511. https://doi.org/10.1109/TIM.2025.3533664

Liu Y, Wang Y, Wang H, Dou S, Tian H, Gan W, Yuan Q (2023) Dynamic template directed construction of three-dimensional porous bismuth aerogels for high-rate Na-ion storage. Journal of Materials Chemistry A 11(11): 5945-5955. https://doi.org/10.1039/d2ta09710a

Liu Z, Guan X, Zhang L, Zhang Y, Pei Y, Liu C, Zhou H, Liu B, Yan K (2021) Investigations of dynamics of a single spark-induced bubble in saline water. Journal of Physics D-Applied Physicsv 54(7): 075203. https://doi.org/10.1088/1361-6463/abc4aa

Lu M, Zhou B, Bu Z (2023) Two-stage distillation-aware compressed models for traffic classification. IEEE Internet of Things Journal 10(16): 14152-14166. https://doi.org/10.1109/JIOT.2023.3263487

Lu Z, Chen Y, Li S, Ma M (2023) SDGC-YOLOv5: A more accurate model for small object detection. Artificial Neural Networks and Machine Learning, ICANN 2023, PT VII. Cham:Springer 14260: 199-209. https://doi.org/10.1007/978-3-031-44195-0_17

Mang S, Yu Y, Zhang Q (2018) Numerical models for interactions of ablating-capillary plasma with air and bulk-loaded liquid. IEEE Transactions on Plasma Science 46(8): 3065-3074. https://doi.org/10.1109/TPS.2018.2855062

Paquette Y, Pellone C, Fivel M, Ghigliotti G, Johnsen E, Franc JP (2015) Modeling fluid-structure interaction in cavitation erosion: preliminary results. Journal of Physics: Conference Series, 9th International Symposium on Cavitation (CAV2015), 656: 012053. https://doi.org?10.1088/1742-6596/656/1/012053

Pietron M, Zurek D, Sniezynski B (2023) Speedup deep learning models on GPU by taking advantage of efficient unstructured pruning and bit-width reduction. Journal of Computational Science 67: 101971. https://doi.org/10.1016/j.jocs.2023.101971

Ragunathan S (2016) Numerical simulation of the flow field of a confined, submerged slot jet impinging on an oscillating surface: a parametric study. Proceedings of the ASME Fluids Engineering Division Summer Meeting, West Conshohocken: ASTM 1A: Paper No: FEDSM2016-7743. https://doi.org/10.1115/FEDSM2016-7743

Rozhdestvensky KV (2022) Dynamics of vapor bubble in a variable pressure field. Journal of Marine Science and Application 21(3): 83-98. https://doi.org/10.1007/s11804-022-00289-4

Ruan X, Liu Y, Yuan C, Li B, Hu W, Li Y, Maybank S (2021) EDP: An efficient decomposition and pruning scheme for convolutional neural network compression. IEEE Transactions on Neural Networks and Learning Systems 32(10): 4499-4513. https://doi.org/10.1109/TNNLS.2020.3018177

Schontag P, Nakath D, Roehrl S, Koeser K (2022) Towards cross domain transfer learning for underwater correspondence search. Image Analysis and Processing, ICIAP 2022, PT III Cham: Springer 13233: 461-472. https://doi.org/10.1007/978-3-031-06433-3_39

Thanh HV, Thu NT, Phuong TM, Vuong ND, Jo KH, Ieee (2024) Inverted bottleneck convolution module for Yolov8. 2024 33RD International Symposium on Industrial Electronics, Ulsan: ISIE 1-6. https://doi.org/10.1109/ISIE54533.2024.10595778

Tonellotto N, Macdonald C, Ounis I (2018) Efficient query processing for scalable web search. Foundations and Trends in Information Retrieval 12(4-5): 319-500. https://doi.org/10.1561/1500000057

Trivedi C (2017) A review on fluid structure interaction in hydraulic turbines: A focus on hydrodynamic damping. Engineering Failure Analysis 77: 1-22. https://doi.org/10.1016/j.engfailanal.2017. 02.021

Wang X, Huang F, Cheng Y (2016) Computational performance optimization of support vector machine based on support vectors. Neurocomputing 211: 66-71. https://doi.org/10.1016/j.neucom.2016.04.059

Xian Y, Yu Y, Lian Y, Fan J, Wang Z (2023) An EA-based pruning on improved YOLOv3 for rapid copper elbow surface defect detection. Engineering Applications of Artificial Intelligence, 123(Part C): 106412. https://doi.org/10.1016/j.engappai.2023.106412

Xiao Y, Xiang H, Wang T, Wang Y (2024) Enhanced Industrial Action Recognition Through Self-Supervised Visual Transformers. IEEE Access 12: 134133-134143. https://doi.org/10.1109/ACCESS.2024.3455749

Xie Y, Chen T, Xu J (2023) Advancing underwater acoustic target recognition via adaptive data pruning and smoothness-inducing regularization. Arxiv arXiv: 2304. 11907. https://doi.org/10.48550/arXiv.2304.11907

Zhang AM, Li SM, Xu RZ, Pei SC, Li S, Liu YL (2024) A theoretical model for compressible bubble dynamics considering phase transition and migration. Journal of Fluid Mechanics 999: A58. https://doi.org/10.1017/jfm.2024.954

Zhang W, An Y (2013) Instability of a bubble chain. Physical Review E 87(5): 053023. https://doi.org/10.1103/PhysRevE.87.053023

Zhao GL, Xu ZY (2025) Coastal vessel target detection model based on improved YOLOv7. Journal of Marine Science and Application. Published online. https://doi.org/10.1007/s11804-025-00635-2

Zheng LN, Xuan P, Huang J, Li JL (2023) Development and application of spark-induced breakdown spectroscopy. Spectroscopy and Spectral Analysis 43(3): 665-673. https://doi.org/10.3964/j.issn.1000-0593(2023)03-0665-09

Zhou J, Yang T, Zhang W (2023) Underwater vision enhancement technologies: a comprehensive review, challenges, and recent trends. Applied Intelligence 53(3): 3594-3621. https://doi.org/10.1007/s10489-022-03767-y

Zhou J, Zhang B, Zhang D, Vivone G, Jiang Q (2024) DTKD-Net: Dual-teacher knowledge distillation lightweight network for water-related optics image enhancement. IEEE Transactions on Geoscience and Remote Sensing 62: Article No. 4207213. https://doi.org/10.1109/TGRS.2024.3422667