Journal of Marine Science and Application (2026) 25: 197-215
https://doi.org/10.1007/s11804-025-00617-4

RESEARCH ARTICLE

Automated Pipe Defect Identification in Underwater Robot Imagery with Deep Learning

Mansour Taheri Andani[1] and Farhad Ameri[2]

Received: 15 July 2024 / Accepted: 19 March 2025

Abstract

Underwater pipeline inspection plays a vital role in the proactive maintenance and management of critical marine infrastructure and subaquatic systems. However, the inspection of underwater pipelines presents a challenge due to factors such as light scattering, absorption, restricted visibility, and ambient noise. The advancement of deep learning has introduced powerful techniques for processing large amounts of unstructured and imperfect data collected from underwater environments. This study evaluated the efficacy of the You Only Look Once (YOLO) algorithm, a real-time object detection and localization model based on convolutional neural networks, in identifying and classifying various types of pipeline defects in underwater settings. YOLOv8, the latest evolution in the YOLO family, integrates advanced capabilities, such as anchor-free detection, a cross-stage partial network backbone for efficient feature extraction, and a feature pyramid network+ path aggregation network neck for robust multi-scale object detection, which make it particularly well-suited for complex underwater environments. Due to the lack of suitable open-access datasets for underwater pipeline defects, a custom dataset was captured using a remotely operated vehicle in a controlled environment. This application has the following assets available for use. Extensive experimentation demonstrated that YOLOv8 X-Large consistently outperformed other models in terms of pipe defect detection and classification and achieved a strong balance between precision and recall in identifying pipeline cracks, rust, corners, defective welds, flanges, tapes, and holes. This research establishes the baseline performance of YOLOv8 for underwater defect detection and showcases its potential to enhance the reliability and efficiency of pipeline inspection tasks in challenging underwater environments.

Keywords YOLO8; Underwater robot; Object detection; Underwater pipelines; Remotely operated vehicle; Deep learning

1 Introduction

Underwater object detection refers to the process of identifying and locating objects submerged in water through various technologies and methods. This process has broad applications across various domains, including marine life research, environmental conservation, infrastructure inspection and health monitoring, and offshore renewable energy. Underwater environments present unique challenges, such as light scattering, poor visibility, and object variability, and thus require innovative detection methods. Techniques and advancements from other fields, such as agricultural object detection, can serve as inspiration for underwater applications. Lightweight deep learning models, such as improved You Only Look Once (YOLO) v5s, have demonstrated robust performance in detecting objects, such as pitaya fruits, under challenging conditions and effectively addressed uneven illumination and light scattering (Li et al., 2024). Similarly, spatiotemporal convolutional neural networks have been employed for tasks, including unmanned pineapple picking, overcoming occlusions, and variability in object appearance (Meng et al., 2023). These methodologies highlight the adaptability and efficiency of object detection models in addressing complex environments and offer insights into tackling the unique challenges in underwater object detection.

Detection of objects underwater is inherently challenging due to factors such as poor light conditions, varying visibility, turbidity, and the optical properties of water that can considerably distort images (Han et al., 2020). The

Article Highlights

- YOLOv8 X-Large was successfully applied to underwater pipeline defect detection, and it demonstrated superior accuracy in identifying cracks, rust, defective welds, and other structural issues.
- A custom underwater pipeline image dataset was created using a remotely operated vehicle, and it addressed the lack of publicly available datasets for underwater defect detection research.

✉ Farhad Ameri
farhad.ameri@asu.edu

[1] Department of Engineering Technology, Texas State University, San Marcos, TX 78666, USA

[2] School of Manufacturing Systems and Networks, Arizona State University, Tempe, AZ 85283, USA

detection methodologies and technologies for underwater objects have been substantially enhanced due to advances in computer hardware, artificial intelligence algorithms, and computing and sensing technologies. Furthermore, advances in autonomous underwater vehicles and robotic systems have further facilitated underwater exploration and data extraction processes (Zhong et al., 2023). Remotely operated vehicles (ROVs) with advanced sensors and digital cameras are used extensively for safe and efficient inspection and monitoring of subsea objects and infrastructures, such as pipelines (Zhang et al., 2021). These challenges can hinder existing object detection technologies, as light scattering reduces contrast, and turbidity introduces noise, leading to low accuracy and missed detections in traditional methods. As a result, the availability of subaquatic imagery has increased exponentially (Fayaz et al., 2022).

The evolution of deep learning methods has provided researchers and practitioners with reliable frameworks for the analysis of image datasets collected in underwater environments (Jin and Zheng, 2020). Since the invention of region-based convolutional neural networks (R-CNNs) (Girshick et al., 2014), they have become crucial elements in object detection solutions. Such networks fall under two categories, namely, two- and single-stage detection methods. The two-stage methods, exemplified by R-CNN and its descendants, namely, Fast R-CNN (Zhang et al., 2016) and Faster R-CNN (Ren et al., 2015), first generate potential bounding boxes that may contain objects; then, these proposed regions are classified into specific object categories. Despite their high precision, given the two-step process, these models are usually slower and less efficient for applications such as real-time underwater pipeline inspection and monitoring. This shortcoming has been addressed using single-stage methods, such as single-shot multibox detector (SSD) (Liu et al., 2016) and You Only Look Once (YOLO) (Rahman et al., 2023). YOLO is an object detection algorithm that simultaneously predicts bounding boxes and class probabilities for multiple objects in an image and thus provides real-time and efficient detection.

YOLOv8, the latest version of YOLO, incorporates inherent architectural advancements, including anchor-free detection, cross stage partial network (CSPNet) backbone for efficient feature extraction, and feature pyramid network (FPN) + path aggregation network (PAN) neck for superior multiscale detection. These features make YOLOv8 well-suited for challenging environments, such as underwater pipelines, which this study leveraged to detect and classify defects. YOLOv8 X-Large, the largest variant in this family, was optimized for accuracy and intended for use in high-end systems where computational resources are abundant.

YOLO8 X-Large was applied to identify and classify defects in underwater pipes. Underwater pipelines operate in rough environments characterized by unpredictable geological changes, continuous seawater corrosion, high pressure, and sediment accumulation. These issues pose a crucial threat to the integrity of pipelines over time, which necessitate regular inspections to ensure their continued health and functionality. Traditional pipe inspection methods involve section-by-section inspections using specialized ships, which is time-consuming and escalates in cost with the expansion of the pipeline network. ROVs are preferred alternatives in this context. With their enhanced stability and autonomous features, ROVs offer an efficient and cost-effective approach to navigating the complex underwater environment and ensuring the systematic inspection and maintenance of pipelines. The performance of ROV cameras is often compromised by shadow noise resulting from inconsistent illumination, camera shake, complex background interference, and the diversity of target types. In addition, light scattering in seawater frequently diminishes or eliminates color features, which further complicates reliable object detection in underwater environments. In this work, an image dataset was generated using a FIFISH V6 ROV equipped with a UHD camera in a controlled environment. Several pipe samples with various defects and features, such as holes, cracks, bad welds, tape, and rust, were used in the simulation of underwater pipelines.

The remainder of this paper is organized as follows: Section 2 provides an overview of related works in underwater object detection. Section 3 discusses the details of the image dataset used for training and validation. Section 4 briefly focuses on the specifications and capabilities of the underwater robot used in this research. Sections 5 and 6 elaborate on the methodology and results, respectively. The paper ends with discussions and conclusions.

2 Related work

This literature review explored the evolution and current state of underwater object methodologies and technologies. Its scope ranges from early sonar-based methods to the most recent advances in artificial intelligence-driven image analysis. This review is guided by the objective, that is, determining how advances in deep learning and image processing meet the unique challenges in the underwater environment**,** such as low observability, light scattering, and intricate backgrounds. This review highlights the major contributions in the field and identifies existing gaps and potential areas for future research.

Karimanzira et al. (2020) worked on challenging underwater environments to achieve remarkable recognition rate improvements obtained by integrating CNNs for enhanced feature extraction using extreme learning machines for efficient classification. Their method addresses inherent challenges of underwater imagery, such as poor visibility and remarkable light reflections. However, the scalability and real-time application of their approach, particularly in

pipeline inspection with its high variability in defect types and environmental conditions, remains a challenge. The YOLO8 algorithm emerges as a viable solution to bridging this gap as it offers a flexible and comprehensive approach to underwater pipeline defect detection.

Chen et al. (2023) have made notable contributions to enhancing underwater image quality through their development of OECBNet. This work focused on image enhancement, which is distinct from object detection, and complemented the capabilities of the YOLO algorithm in defect detection in underwater pipelines. The integration of OECBNet's enhanced image clarity with the detection technology of YOLO8 can enrich the field of underwater object detection. This combined approach offers a robust method for the precise and efficient identification of defects in underwater environments, which demonstrates the potential synergy between advancements in image enhancement and object detection.

Deep learning for underwater image processing was examined extensively in the work of Vidhya and Deelthi (2023), and the results revealed its adaptability to challenges, such as light distortion and color variation in underwater environments. A limited emphasis has been observed on object detection models, such as YOLO, which is essential for precise tasks such as underwater pipeline defect detection. The lack of attention on this issue highlights a critical gap in the literature, which underscores the need for specialized research focusing on the nuanced requirements of underwater object detection.

Wang et al. (2023) addressed the complex challenge of underwater object detection in forward-looking sonar images through the integration of an improved YOLOv8 model with attention mechanisms and advanced preprocessing techniques. Their study tackled the limitations of traditional methods, which suffer from high noise levels and low-resolution sonar imagery, by employing a novel preprocessing module that includes artifact removal, contrast enhancement, and noise filtering. The incorporation of the convolutional block attention module (CBAM) further enhanced feature extraction and noise suppression while maintaining computational efficiency. The present research highlights the potential of YOLOv8 in adapting to challenging underwater environments by effectively balancing accuracy and resource constraints. The innovative use of these techniques aligns with the objectives of this study and showcases how tailored advancements can tremendously improve object detection capabilities in underwater settings.

Burguera and Bonin-Font (2022) have made remarkable contributions to machine learning for autonomous underwater navigation and sensor data processing. However, their work did not specifically address defect detection in underwater pipelines, which is a key focus of our research. Integrating their findings with the YOLO algorithm's defect detection capabilities can substantially enhance autonomous underwater robotics, particularly in specialized tasks such as pipeline inspection and maintenance. This approach represents a promising direction for future research, where the combination of autonomous navigation and precise defect detection can lead to the development of efficient and accurate maintenance solutions.

Orinaitė et al. (2023) examined the difficulty of using machine learning to detect cracks in underwater concrete structures. The problem they addressed is important because of the notable difficulties in underwater environments, such as limited visibility and complex concrete surfaces underwater. To recognize cracks in images, they used image processing techniques, especially machine learning algorithms and improved algorithm accuracy and robustness using a dataset augmentation strategy. An important finding of their study was the effectiveness and accuracy of this approach in identifying cracks in underwater concrete structures. They demonstrated that machine learning approaches, specifically those that use AlexNet and SqueezeNet through transfer learning on an augmented dataset, offer a reliable means of detecting and monitoring such cracks. This approach can enhance the safety and reliability of underwater structures and prevent catastrophic failures. However, a limitation of their study can be the challenges in adapting these techniques to various underwater conditions and types of concrete structures.

Xu et al. (2023) presented a systematic analysis and review of deep learning techniques for underwater object detection and addressed the unique challenges imposed by underwater environments. They identified key issues affecting the accuracy and effectiveness of traditional object detection methods in these settings, including high noise levels, low visibility levels, and color deviation. Xu et al. (2023) reported the extremely complex processing of underwater images, and advanced deep learning approaches that can adapt to these challenging conditions, such as YOLO8, played crucial roles. Their research provided valuable insights into the current state of underwater object detection and its limitations.

Zhang et al. (2023) developed an improved YOLOv4-based pipeline defect detection method, specifically for sewer environments. Their enhancements consisted of the integration of a spatial pyramid pooling (SPP) module and modifications to the loss function for improved precision and recall. These advancements enabled their model to achieve remarkable improvements in the detection of small defects (such as cracks) and achieved a 4.6% increase in the mean average precision (mAP) over the standard YOLOv4. Despite the strong performance of their method in pipeline inspection, the study focused on sewer pipelines, which differ greatly from underwater environments in terms of challenges, such as light scattering, turbidity, and visibility. This highlights a gap in the adaption of their

improvements to underwater defect detection, where the environmental conditions demand additional preprocessing and architectural considerations.

The proposal by Wang et al. (2023) addresses the limitations of traditional CCTV-based pipeline defect detection. Their work integrated involution operators for lightweight modeling, CBAM for enhanced feature extraction, and knowledge distillation for improved generalization capabilities of their YOLOv5s-based model. This approach considerably reduced computational overhead and improved detection accuracy, with a mAP@0.5 of 80.5%. However, their study was primarily focused on sewer pipelines and did not explicitly address the unique challenges of underwater environments, such as light scattering, color distortion, and varying visibility conditions. This limitation opens avenues for the application of similar enhancements in underwater pipeline defect detection scenarios.

Chi and Zhang (2024) proposed a biological vision-inspired underwater image enhancement method to address challenges, including color distortion, low visibility, and poor image quality, which are commonly encountered in underwater imaging tasks. Their approach integrates two key modules: a LAB color space-based correction module and a visibility enhancement module utilizing Type-II fuzzy sets. This combination allows for improved clarity and color fidelity in underwater images. Their method demonstrated notable improvements in enhancing visibility in challenging underwater environments, with its performance validated on benchmark underwater datasets. However, although their work focused on image enhancement, it excluded object detection or defect classification tasks, which left a gap in the direct application of their methods to real-time underwater pipeline inspection. The integration of such enhanced imaging techniques with advanced detection algorithms such as YOLOv8 can further improve the accuracy and reliability of underwater defect detection.

Gašparović et al. (2022) used deep learning methods to address the challenges of underwater pipeline detection tailored for harsh subsea environments. They focused on mitigating issues, such as light scattering, attenuation, and poor visibility, which often degrade image quality in underwater settings. By training and testing six different CNN detectors, including five YOLO-based architectures and one Faster R-CNN model, they evaluated performance on a custom dataset of underwater pipeline images. YOLOv4 achieved the highest mAP of 94.21%, which showcases its robustness in handling varying environmental conditions. This study highlights the adaptability and efficiency of YOLO-based methods in underwater object detection and provides a strong foundation for extending pipeline inspection capabilities. Gašparović et al.'s (2022) findings align with the objectives of this research, which demonstrates the viability of advanced CNN architectures for underwater inspection tasks.

Jin and Zheng (2020) proposed a method that uses the YOLOv3 algorithm to detect oil spill points in underwater pipelines. Their approach addresses common underwater imaging challenges, including distortion, blur, and low contrast, through image enhancement techniques, such as Gaussian filtering and histogram equalization. Jin and Zheng (2020) concentrated on oil leakage detection, whereas this research extends these capabilities to a broader range of defect types in underwater pipelines, demonstrating the evolving and adaptive nature of YOLO algorithms in underwater object detection.

Table 1 provides a comparative analysis of the performance metrics of various YOLO-based methods in underwater object detection. The table also includes key performance indicators, such as precision, recall, and mAP50, for different YOLO models applied across various datasets.

This literature review illustrates the evolution of underwater object detection and highlights important advances and present challenges, particularly in real-time and accurate defect detection. Underwater environmental conditions, such as variable lighting, suspended particles, and water optical properties, persistently challenge the efficacy of existing technologies and often compromise image quality. To address these challenges, this study comprehensively evaluated the YOLO8 algorithm for its efficacy in detecting defects within underwater pipeline structures.

Table 1 Comparative performance metrics of YOLO models in underwater object detection across various datasets

Model	Precision (%)	Recall (%)	mAP50 (%)	Dataset
YOLOv3	87.0	67.0	75.27	Fish (Al Muksit et al., 2022)
YOLOv4	80.0	76.0	81.02	Fish (Al Muksit et al., 2022)
YOLOv5s	81.5	75.9	80.50	Pipe (Wang et al., 2023)
YOLOv3-Tiny	80.0	78.0	87.18	Underwater Life (Asyraf et al., 2021)
YOLOv5	91.0	58.5	69.30	Trash (Pavani et al., 2023)
YOLOv8	60.7	47.2	43.60	Trash (Pavani et al., 2023)
YOLOv7-AC	90.0	84.2	89.60	Seafood (Liu et al., 2023)
YOLOv3	94.0	90.0	78.80	Underwater Life (Shankar and Muthulakshmi, 2023)
YOLOv5	89.0	89.0	80.00	Underwater Life (Shankar and Muthulakshmi, 2023)

3 Dataset

In deep learning methods, the size and quality of the input dataset contribute greatly to the accuracy of final outcomes. The available public datasets for underwater pipelines failed to meet the requirements of this project on image quality, diversity of pipe defects, and ambient conditions. Therefore, the required images were generated internally in a controlled underwater environment for training, validation, and test purposes. Dataset generation involved submerging various defective and nondefective pipes within a round pool, with dimensions measuring 365 cm in diameter and a depth of 121 cm, and utilizing an underwater robot camera to capture high-resolution images of these pipes under various arrangements and lighting conditions.

FIFISH V6 ROV was selected as the underwater robot, given its advanced imaging capabilities, including a UHD camera and integrated lighting system. The ROV's six degrees of freedom enable precise control over orientation and positioning (roll, pitch, and yaw) and ensure consistent and diverse image capture. These features are crucial for the simulation of real-world conditions, such as varying light scattering and visibility, which enhance the dataset's representativeness.

The metadata associated with each image, including depth, temperature, imaging time, and robot orientation (roll, pitch, and yaw), were recorded. The chemical properties of water were also measured and recorded during each imaging session. The pipes used in this experiment were collected from a scrap metal yard. Some of the features of interest, such as holes, cracks, bad welds, and bends, were already available on the pipes, but additional defects were added manually in the laboratory. The final high-resolution dataset contained 2 467 images, which were representative of various underwater inspection scenarios[①]. This dataset serves as a critical resource for testing and validating object detection models such as YOLOv8.

Figure 1(a) and Figure 1(b) show some representative images of the dataset. Nine features, including bad weld, corners, cracks, flanges, good weld, holes, pipe rust, and tape, were designated the classes of interest in this classification problem. Each image included at least one of these nine classes.

Data labeling

All generated images were labeled on the Roboflow website. Each image was assigned manually, and the relevant labels related to the objects of interest. These labels included the object type (class) and rectangular coordinates of objects in the image (top, bottom, left, and right positions of the rectangle). Figure 2 summarizes the class frequency and geometric distribution of the data set generated after labeling. Figure 2(a) displays the number of occurrences of class labels in the dataset. Accordingly, 2 079 bad welds, 1 083 corners, 2 310 cracks, 1 228 flanges, 1 778 good welds, 4 613 holes, 4 725 pipes, 4 146 rusts, and 3 087 tapes were labeled in the entire dataset, with a total of 25 046 target objects. Figure 2(b) reveals the size distribution of the label box. As displayed in Figure 2, a majority of labeled targets were of smaller dimensions during the detection process, which demonstrated the prevalence of smaller targets within the dataset. Figure 2(c) reveals a normalized label-box center-coordinate point distribution. The coordinates 0 to 1 reveal the middle coordinate points in the image. The center coordinates, which covered the entire image from 0 to 1. According to the random distribution of underwater object features in the image, the middle coordinates (x, y) of the dataset showed a relatively vast distribution. Figure 2(d) illustrates the height and width distribution of the normalized label boxes. It shows that the distribution is usually centered around smaller values, with the highest concentration occurring in the 0.0 to 0.1 range, which suggests the relatively small proportion of the target within the image. A total of 70% of the dataset was selected for training (1 763 images), 10% for testing (293 images), and 20% for validation (411 images).

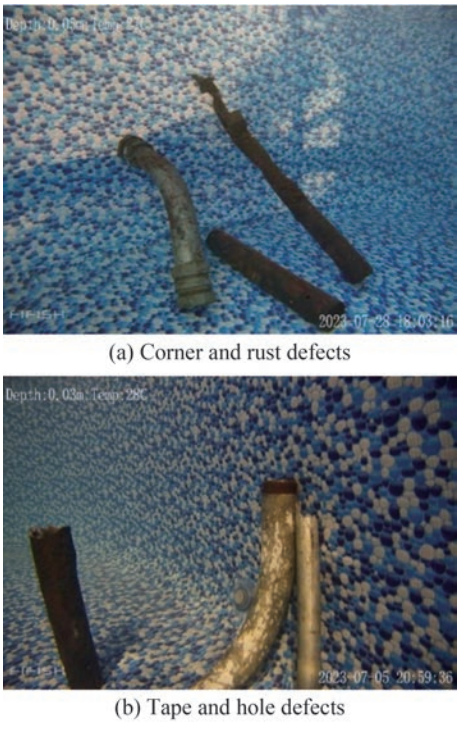
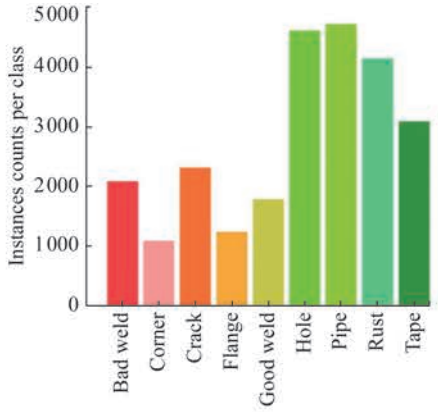


Figure 1 Dataset containing nine classifications

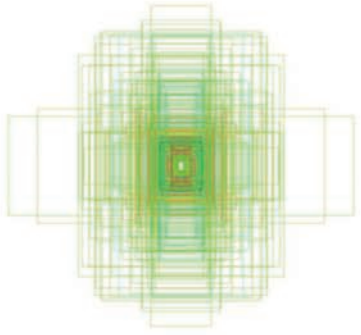
4 Underwater robot

FIFISH V6 (Figure 3) was used in this study. FIFISH V6 is an OMNI-directional ROV equipped with a 4K UHD

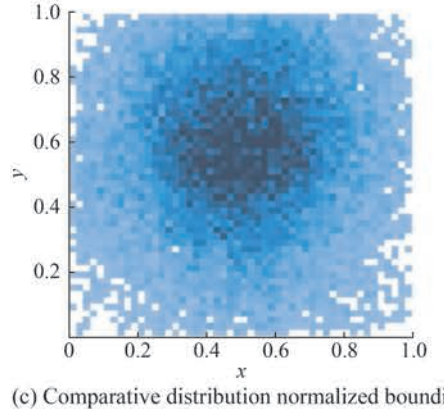
[①] The dataset is available at: github.com/infoneerTXST/Pipe-datasets



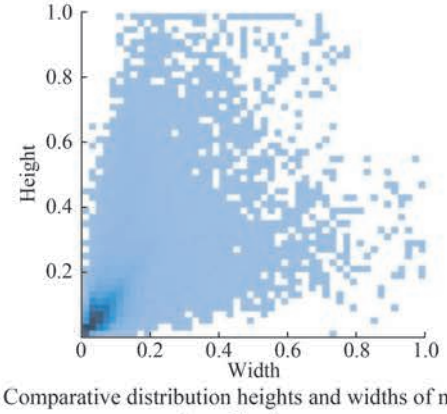
(a) Classification-based object count distribution



(b) Distribution of bounding box dimensions



(c) Comparative distribution normalized bounding



(d) Comparative distribution heights and widths of normalized bounding boxes

Figure 2 Dataset distribution

camera that boasts a 166° viewing angle, and it captures images in JPEG and DNG formats at a resolution of 4 000 × 3 000. A wireless remote control facilitates its comprehensive maneuverability, including 360° rotations on all axes and a 90° tilting capability. The ROV features onboard storage for image data, which can be transferred to a computer for analysis. It also provides real-time information, including temperature and orientation, through a dedicated application. V6 is enhanced with LED lighting and various sensors for improved visibility and data accuracy in underwater environments. Its modular design allows the integration of additional sensors, including those measuring dissolved oxygen levels, pH, water quality, and salinity, which enriches the data acquisition process. These features play a crucial role in identifying areas prone to rust and evaluating the quality of welding joints. FIFISH V6, which can operate at depths up to 328 ft, was used to inspect underwater infrastructure, particularly pipelines. Table 2 provides a detailed summary of the FIFISH V6 specifications.

Figure 3 FiFISH V6 underwater robot (QYSEA, 2022)

Table 2 Detailed specifications of the FIFISH V6 ROV highlighting its advanced features for underwater imaging and data collection

Parameter	Specification
Camera	4K UHD, 4 000 × 3 000 resolution
Field of view (FOV)	166°
Lighting	Dual 4000-lumen LED lights
Depth rating	328 ft (100 m)
Degrees of freedom	6 (360° rotation, 90° tilt)
Battery life	Up to 4 h
Control method	Wireless remote control via mobile application
Image formats	JPEG and DNG
Sensors	Dissolved oxygen, pH, water quality, salinity
Onboard data storage	Supported, transferable to a computer
Weight	4 kg
Dimensions	13.7″ × 12.4″ × 5.9″

5 Methodology

This section discusses in detail the methodology in terms of detection algorithm, data preprocessing, model preparation, evaluation metrics, and model training.

5.1 Detection algorithm

The YOLO series algorithms are the premier choice in industrial applications, with their combination of speed and accuracy establishing them as industry leaders (Li et al., 2022). Through a single evaluation, the unified neural network computes class probabilities and bounding boxes directly from an entire set of images (Redmon et al., 2016). YOLO series unified detection approach stands in considerable contrast to traditional object detection strategies that process images through sliding windows or region proposals. YOLO processes an image in a single pass, hence the name "You Only Look Once". This approach was used as a basis to divide the image into a grid system. It also directly identifies bounding boxes and class probabilities from the pixels within the image. The implementation of the YOLO algorithm in object detection for underwater images is primarily based on two major advantages: the combined high speed with efficiency and enhanced capability to recognize the global context (Raza and Hong, 2020). The first advantage is speed and efficiency: YOLO can deliver real-time object detection, which is necessary for applications such as self-driving vehicles and real-time ocean monitoring. It also processes images rapidly without substantially compromising accuracy. Therefore, YOLO is often the top choice when timing is critical. The second advantage is global context recognition: The YOLO algorithm processes an entire image one at a time, a distinctive aspect furnishing a comprehensive understanding of the scene. This global perspective reduces false positives, which results in improved accuracy and reliable predictions. This is particularly beneficial in certain situations, especially in complex and varied underwater environments where accurate object detection can be a crucial challenge (Li et al., 2022). Through its integrated speed, efficiency, and global context, YOLO serves as a suitable choice for object detection in underwater images (Zhao et al., 2022). To further enhance multiscale object detection in underwater imagery, YOLOv8 employs the FPN and path aggregation network (PAN) neck architecture. This component combines semantic information across multiple feature scales, which improves the model's capability to detect objects of varying sizes and shapes, such as small cracks or large rust patches on underwater pipelines. Through the aggregation of high- and low-level features, the FPN+PAN neck architecture also addresses challenges, such as light scattering and complex object appearances, in underwater environments, which results in accurate defect localization (Yaseen, 2024).

5.2 Data preprocessing

To enhance the computational efficiency and model performance, we applied two distinct preprocessing steps to all images in the dataset acquired using ROV cameras.

5.2.1 Resizing images

In our project, we adopted an essential preprocessing step where we altered the initial dimensions of captured images to a uniform size of 640 × 640 pixels. This transformation was achieved using a specific programming function. Resizing ensured compatibility with the YOLOv8 model architecture, which requires fixed input dimensions for optimal performance and effective utilization of the CSPNet backbone during feature extraction. However, this step alone does not address underwater-specific challenges, such as light scattering or turbidity, which remain prominent issues affecting image clarity and defect visibility. Future research on preprocessing enhancements should explore more adaptive techniques to mitigate these challenges. Resizing was executed across all images, which allowed standardization of the dataset and simplification of the training and detection process.

This crucial adjustment minimized the computational complexity and potential for distortions that may occur with varying image sizes. The uniform resolution also aids the YOLOv8 architecture, particularly the FPN+PAN neck, in integrating multiscale features effectively during detection. This step ensured that the model can detect defects of varying sizes, such as small cracks or larger patches of rust, with consistent accuracy. Moreover, a uniform resolution must be maintained to strike a harmonious balance between computational efficiency and image quality, which is particularly important in the context of real-time applications, where speed and accuracy are imperative.

5.2.2 Image normalization

The pixel values of images were normalized to a range of 0 to 1 to enhance the training process. This adjustment improved the object detection model's performance by ensuring that all pixel values have the same scale, which allowed the model to learn more easily. Normalization also helped stabilize gradient updates during training, which reduced the risk of exploding or vanishing gradients in the neural network. This preprocessing step is particularly beneficial in underwater environments, as it reduces the effects of varying lighting conditions often encountered in these settings. Through standardization of pixel values, the model became more robust to alterations in brightness and contrast, which improved its capability to detect subtle defects, such as cracks and faint rust patches. Despite its benefits, normalization failed to fully address underwater image issues, such as reduced contrast and color distortion caused by light absorption and scattering. Advanced methods, including contrast enhancement and color correction, will be considered in future work to further improve the quality of training data (Park and Eom, 2024; Soorma et al., 2023).

5.2.3 Neural network architecture

To develop an adept underwater object detection system, we selected the YOLOv8 model architecture, a state-of-

the-art framework renowned for its efficiency and accuracy (Aboah et al., 2023). The architecture integrates advanced components, such as the CSPNet backbone for efficient feature extraction and the FPN + PAN neck for multiscale detection. These features are particularly useful for underwater defect detection and enable the model to detect objects of varying sizes with precision, even in complex environments (Zhang et al., 2024). This architecture is fundamentally based on a deep CNN, which contains several layers with different dimensions and numbers of filters. The anchor-free detection mechanism of YOLOv8 further simplifies the detection process and improves bounding box predictions of defects with varying shapes and aspect ratios (Du and Song, 2024).

The CSPDarknet backbone, an advanced variant of CSPNet, enhances feature extraction by splitting feature maps into two parts and fusing them across stages. This design reduces computation without sacrificing accuracy, which makes it well-suited for the intricate visual patterns of underwater defects, such as cracks, rust, and poor welds. By effectively capturing low- and high-level visual cues, CSPDarknet improves the model's capacity to detect subtle features in noisy underwater environments (Zhou et al., 2024).

The FPN + PAN neck, an enhanced version of the PAN, integrates features from various network layers. This multiscale feature fusion considerably improves the detection of small objects, such as holes and cracks, in underwater pipelines. Its capability to aggregate high- and low-resolution features ensures robust performance in scenarios with complex backgrounds and varying object scales (Fu et al., 2023).

The output layer of YOLOv8 generates predictions for bounding boxes, class probabilities, and confidence scores. For this study, the model was configured to detect nine defect classes, including cracks, rust, and flanges. The anchor-free detection mechanism simplifies the prediction of bounding boxes via direct learning of object center points and scales, which improves the accuracy of detecting irregularly shaped defects (Gao et al., 2024).

5.2.4 Initial model weights

To speed up the training process and improve performance, we initialized our model using pretrained weights from YOLOv8, which is a proven image detection model trained on an extensive dataset. In addition to the acceleration of training convergence, transfer learning enables the model to inherit prior knowledge of generic features (e.g., edges and textures), which is useful in fine-tuning the model on the specialized underwater pipeline defect dataset. This approach considerably reduces the computational resources and training time while increasing the model's adaptability to new underwater images.

5.3 Evaluation metrics

Several standard evaluation metrics, such as precision, recall, and mAP, and various loss functions, such as regression, confidence, and classification losses, were used to assess the performance of the object detection model. This section discusses in more detail the aforementioned metrics and presents their associated formulas. The precision (P) formula is shown in Equation (1):

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

where TP (true positive) refers to the number of correctly predicted positive instances by the model. FP (false positives) denotes the number of instances that the model predicted as positive but were actually negative. The recall(R) formula is computed using Equation (2):

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

where FN (false negative) indicates the number of samples in which the target was not detected. The mAP formula is provided in Equations (3) and (4).

$$\mathrm{mAP} = \frac{1}{C}\sum_{J=1}^{C} \mathrm{AP}_J \tag{3}$$

$$\mathrm{AP} = \frac{1}{N}\sum_{i=1}^{N} \cdot P_i \tag{4}$$

The AP (average precision) metric signifies the mean value of the accurate prediction probability for every individual class. The symbol N stands for the cumulative count of images that encompass target attributes, and P denotes the likelihood of accurate prediction of these target attributes within each respective image. C shows the number of classes. In our case, we identified nine distinct classes that were crucial in our analysis: pipes, good welds, tape, rust, bad welds, corners, cracks, flanges, and holes. The mAP illustrates the average precision computed across the AP values of all available classes. The intersection over union (IoU) measures the overlap between two bounding boxes, and it is calculated by dividing the intersection area by the union area. This variable is one of the most important metrics for the evaluation of the accuracy of object detection models. The formula is represented as Equation (5).

$$\mathrm{IoU} = A_{\mathrm{overlap}}/A_{\mathrm{union}} \tag{5}$$

where A_{overlap} represents the area of overlap, A_{union} is the area of union.

The loss regression formula is shown in Equation (6):

$$L_{\text{regression}} = \sum_{i=1}^{N}\sum_{j=1}^{M} 1_{ij}^{\text{obj}}\left[\left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 + \left(\sqrt{w}_i + \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}}_i\right)^2\right] \quad (6)$$

where $L_{\text{regression}}$ (regression loss) is a vital factor in comprehending the performance and accuracy of the object detection process. The $L_{\text{regression}}$ metric quantifies the disparity between ground truth boxes and predicted boxes generated by the model. It involves parameters, such as N and M, which represent the total number of ground truth and predicted boxes, respectively. The indicator function 1_{ij}^{obj} discerns correct detections; it interacts with the coordinates $\left(x_i, y_i, w_i, h_i\right)$ and $\left(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i\right)$, which define the dimensions and central points of the actual and predicted boxes, respectively. This evaluation assists in honing the model through the identification and minimization of errors in object localization.

Confidence loss was calculated to quantify the difference between the actual labels of the bounding boxes and the confidence scores predicted by the model. The formula for confidence loss is provided in Equation (7). The confidence loss metric, denoted as $L_{\text{confidence}}$, was used to quantify the discrepancy between the predicted confidence scores and actual labels associated with bounding boxes in object detection tasks. This metric involves several parameters and indicator functions. Specifically, 1_{ij}^{obj} is an indicator function that takes the value of 1 if an object is detected within the i-th cell, and 0 otherwise. Conversely, 1_{ij}^{noobj} implies the absence of an object in the i-th cell, with a value of 1 assumed when no object is detected, and 0 otherwise. p_i, which denotes the predicted confidence score for the same box, represents the ground truth confidence score of the i-th bounding box. Here, N stands for the total count of bounding boxes analyzed in the procedure. In addition, λ serves as a regularization parameter, and it was employed to balance the loss contributions from cells containing and lacking objects. This balancing act ensured the harmonious integration of loss values from varying scenarios and fostered an optimized and accurate object detection model. Equation (8) calculates the classification loss. The classification loss, denoted by $L_{\text{Classification}}$, measures the discrepancy between the predicted and actual class labels within bounding boxes. Here, N represents the total count of bounding boxes analyzed. The indicator function 1_{ij}^{obj} takes the value 1 if an object is detected in the i-th bounding box and 0 otherwise. The variable c signifies each class in the possible set of categories. Meanwhile, $p_i(c)$ refers to the genuine probability that the i-th bounding box encompasses an object belonging to class c. Conversely, $\hat{p}_i(c)$ stands as the forecasted probability that the i-th bounding box contains an object of class c, as predicted by the model. This loss metric aids in refining the classification aspect of the object detection model through the minimization of the errors between predicted and actual class probabilities.

$$L_{\text{confidence}} = \sum_{i=1}^{N} 1_i^{\text{obj}}\left(p_i - \hat{p}_i\right)^2 + \lambda \sum_{i=1}^{N} 1_i^{\text{noobj}}\left(\hat{p}_i\right)^2 \quad (7)$$

$$L_{\text{classification}} = \sum_{i=1}^{N} 1_i^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2 \quad (8)$$

5.4 Training the model

5.4.1 Determining the epoch number

A total of 100 epochs were used in our training process, with each representing a full cycle where the model was trained using the entire training dataset. The number of epochs was selected to ensure sufficient opportunities for the model to learn and adjust its weights optimally. Such a condition would allow it to extract meaningful patterns from the data. Setting the number of epochs at 100 helped strike a balance between training time and performance, which ensured satisfactory convergence of the model while avoiding overfitting. This approach provided the model with sufficient exposure to the data to generalize effectively to unseen samples, which is crucial for robust underwater defect detection.

5.4.2 Model training

As a consequence of the model design and data learning, we selected 70% for training, 10% for testing, and 20% for validation. This partitioning ratio is commonly employed in machine learning applications to balance the need for sufficient training data while reserving a portion for unbiased evaluation. By allocating 70% of data for training, the model gains access to a diverse and comprehensive subset of the dataset to learn underlying patterns effectively. The other 10%, which was reserved for testing, ensured that model performance was evaluated on observed data, which provided an accurate measure of its generalization capabilities. The remaining 20%, which was allotted for validation, was used to tune hyperparameters and monitor the model's performance during training, which prevented it from overfitting the training set. This particular ratio was selected for this study to accommodate the relatively small size of the dataset while ensuring sufficient data points of each partition for reliable training, validation, and testing. The choice reflects a practical trade-off between computational efficiency and statistical reliability, especially in underwater object detection, where the generation of large datasets can be challenging (Nguyen et al., 2021). Therefore, the training dataset contained 1 763 images, and the test and validation sets contained 293 and 411 images, respectively. Training data were used to update the model's parameters through backpropagation during each epoch. The data were fed into the network, the error

(or loss) was calculated, and network parameters were adjusted accordingly. The loss function played a critical role in quantifying the difference between the model's predictions and actual labels, which enables the adjustment of weights to minimize this discrepancy over time. This iterative process gradually fine-tuned the model to reduce errors and improved its predictive precision and recall for detecting underwater pipeline defects.

To further refine the training process, we analyzed three critical loss elements—box, classification, and deep feature learning (DFL) losses—to evaluate the model's capability to localize and classify defects accurately. Box loss represents the error in bounding box predictions, and classification loss quantifies the error in assigning correct defect categories. The DFL loss focuses on feature representation, and it aids the model in learning detailed and distinguishing patterns essential for accurate defect detection. This multifaceted approach ensures that the model captures the spatial and semantic characteristics of underwater defects.

Figure 4 shows the analytical progression of the performance of our model, which highlights the training and validation stages for these critical loss elements. To enhance interpretability, we applied a smoothing technique to the graphs, which eliminates minor fluctuations and results in cleaner, more readable curves. In this specific instance, the orange curve demonstrates a smoother and more stable trend, as opposed to the blue curve, which exhibits more fluctuations. The stabilization of the orange curve over epochs reflects the model's capability to converge effectively, which reduces overfitting and ensures reliable predictions during validation.

We implemented the trained model on the validation data set and undertook a detailed evaluation of the outcomes. The results are articulated through the metrics of precision, recall, and mAP, offering comprehensive insights into the model's performance.

Figure 5 illustrates the model's validation performance, that is, its precision, recall, and mAP50 metrics and its capability to detect objects accurately.

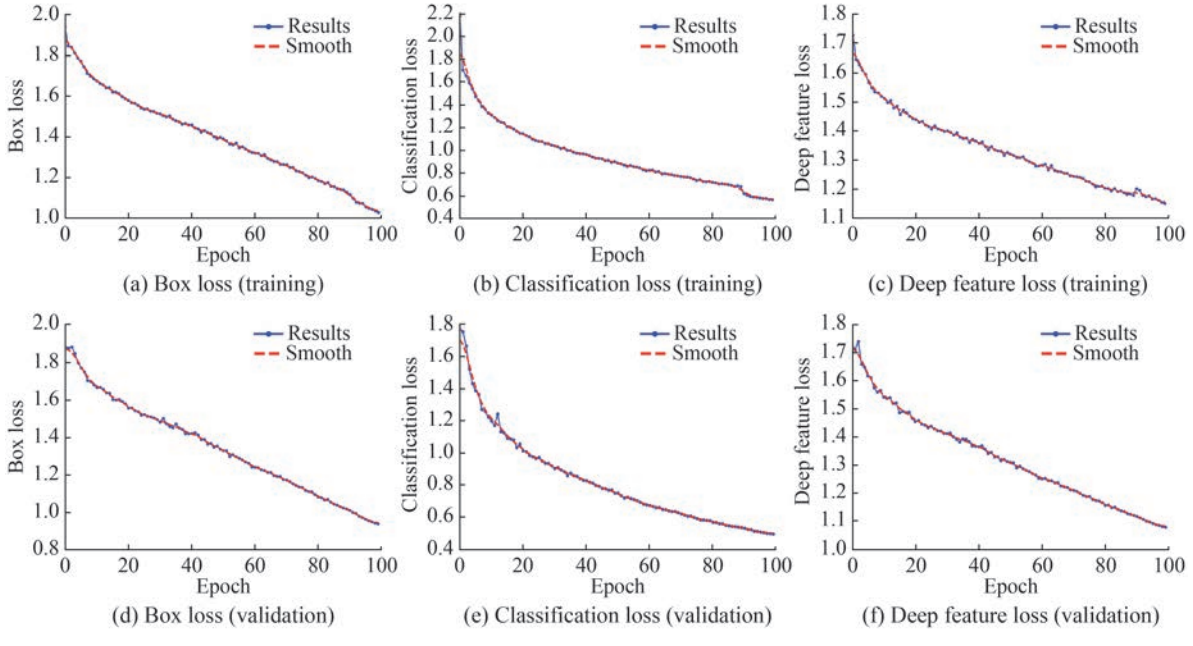


Figure 4 Loss function optimization trends

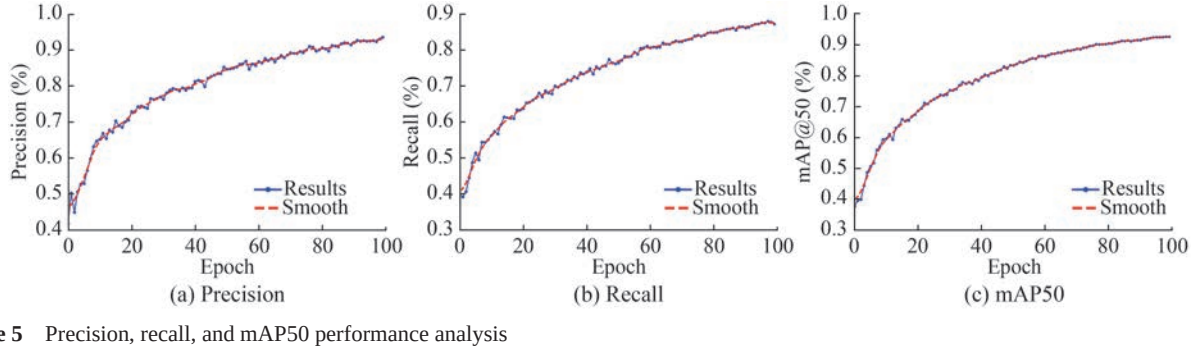


Figure 5 Precision, recall, and mAP50 performance analysis

5.5 Utilizing the model for detection and prediction

Input images were used to train the model in detecting and categorizing objects present within them. In addition to pinpointing objects with comprehensive rectangular coordinates, the model associates them with their associated class labels. The rectangular bounding boxes represent the spatial location of each detected object within the image, and the class labels identify specific defect types, such as cracks, rust, or weld issues. This dual output provides actionable insights into underwater pipeline inspection. YOLOv8's anchor-free detection mechanism enables the simplification of the bounding box prediction process via the direct estimation of the center and dimensions of objects, which improves the model's capability to localize defects accurately, even in challenging underwater environments. This approach ensures a structured and effective method for object recognition and detection, which allows for precise identification of pipeline defects despite variations in size, shape, and environmental conditions.

Figure 6 illustrates the model's functionality. First, it accurately labels various defect classes within a single image, which demonstrates its multiclass detection capability. This capability includes the detection of multiple defects, such as a crack and rust patch within the same frame, which highlights the model's capability to handle complex underwater scenes. The same framework is then deployed to predict potential object detections, which showcases a cohesive blend of object detection and prediction processes. This capability ensures that the model can not only identify existing defects but also predict their likelihood in similar contexts, making it a valuable tool for proactive pipeline monitoring.

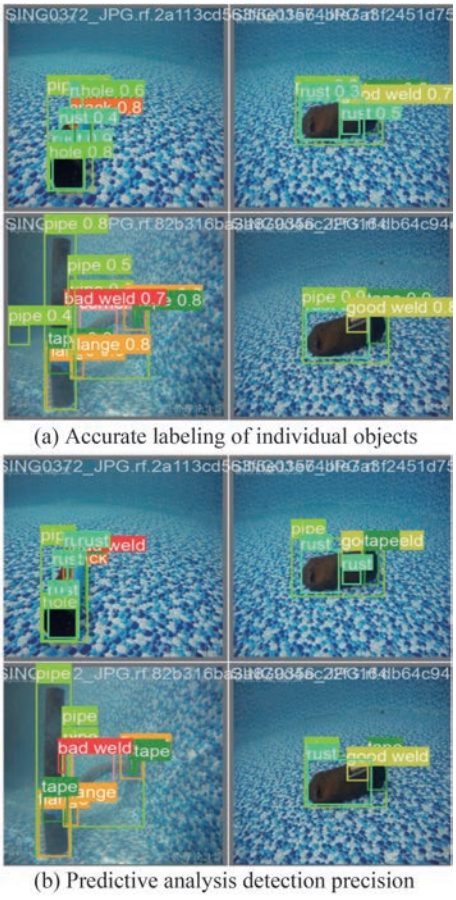


Figure 6 Image labeling and predictive analysis

6 Experimental results

To optimize the YOLOv8 model for underwater object detection, Kim et al. (2023) proposed a high-speed detection approach. The YOLOv8 model comprises six different sizes: Nano, Small, Medium, Large, and X-Large. Each configuration represents a trade-off among computational efficiency, speed, and detection performance, which allows users to select the most suitable model based on available hardware and application requirements.

The YOLOv8 Nano model, which is designed for environments with limited hardware resources, prioritizes speed and lightweight performance while maintaining the capability to identify and localize defects effectively (Zhang and Ni, 2023). The YOLOv8 Medium model possesses upgraded speed, accuracy, and efficiency over its predecessors. This model employs advanced features, such as multiscale image fusion, to enhance its detection capabilities, particularly for medium-complexity datasets, while remaining computationally efficient (Kim et al., 2022). The YOLOv8 Large model provides enhanced precision and detection capabilities and is optimized for systems with substantial hardware resources. This model demonstrates superior performance in detecting small and complex defects, such as cracks and weld issues, by leveraging its deeper architecture to capture fine-grained details in underwater imagery.

Finally, the YOLOv8 X-Large model, which has the most accurate configuration, is tailored for high-end hardware systems and detailed datasets. This model achieves the highest precision and recall metrics, which make it ideal for detailed underwater pipeline inspections where accuracy is paramount.

Table 3 illustrates the performance variations across different YOLOv8 configurations. As the model's complexity increases from nano to X-Large, improvements in precision, recall, and mAP scores are observed. These metrics reflect the models' effectiveness in the accurate localization and classification of defects, such as small cracks, rust patches, and flanges. However, this increase in accuracy comes at the cost of longer processing times, which must be balanced against real-time detection requirements in practical applications.

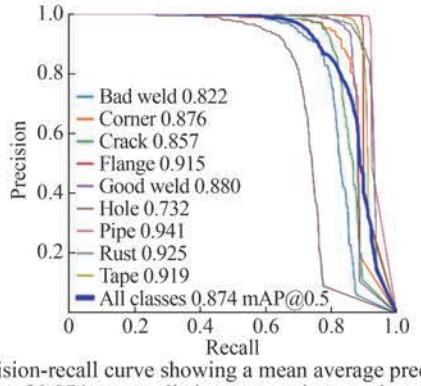
Table 3 Performance of different YOLOv8 configurations

Configuration	Precision	Recall	mAP50	mAP50-95	Time (ms)
Nano	0.844	0.720	0.800	0.491	12.5
Small	0.822	0.831	0.893	0.612	15.1
Medium	0.952	0.862	0.908	0.670	18.4
Large	0.960	0.892	0.933	0.711	21.2
X-Large	0.961	0.929	0.966	0.746	22.5

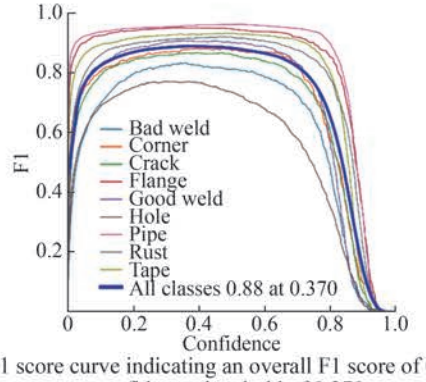
This scenario showcases the trade-off between detection speed and accuracy (Avsar et al., 2023), which emphasizes the importance of selecting the right configuration based on project requirements. The YOLOv8 X-Large configuration showed excellent performance among the tested configurations. It demonstrated robust detection capabilities in identifying small and complex defects, such as holes and bad welds, while maintaining a balance between precision and recall. To enrich this analysis, we introduced a set of analytical curves to the X-Large model (Figure 7), which provided a comprehensive illustration of the model's nuanced behaviors and efficiencies. These visual insights bridge quantitative metrics with interpretability and offer a detailed understanding of how the model performs across various defect classes under different conditions. The precision–recall curve in Figure 7(a) illustrates the balancing act between model accuracy and inclusivity. Specifically, the "bad weld" class emphasizes the importance of precision in detection. A failure to accurately identify a "bad weld" can lead to substantial environmental and structural issues, such as rust propagation, underwater leaks, or catastrophic failures. On the other hand, the recall curve in Figure 7(d) underscores the criticality of detecting the "hole" class early. A high recall ensures swift identification of structural vulnerabilities, enabling proactive maintenance and reducing operational costs by minimizing the need for frequent manual inspections.

The curves show that each class, which is represented by different colors, experiences a trade-off where elevating precision occasionally dampens recall, and vice versa. This inverse relationship between precision and recall is intrinsic to classification tasks. A stricter confidence threshold improves precision by reducing false positives but may also exclude true positives, which reduces recall. Figure 7(b), which depicts the F-1 curve, shows how the F-1 score varied at different confidence thresholds. The F-1 score, which is the harmonic mean of precision and recall, reflects the overall balance between these metrics and provides a single measure for evaluating the model's performance for each class.

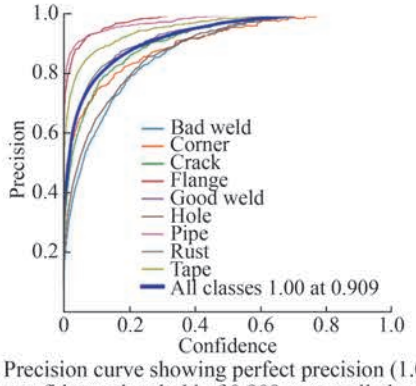
The precision-confidence curve in Figure 7(c) demonstrates how the model's precision changes with varying confidence levels. Segmented by class, this curve high-



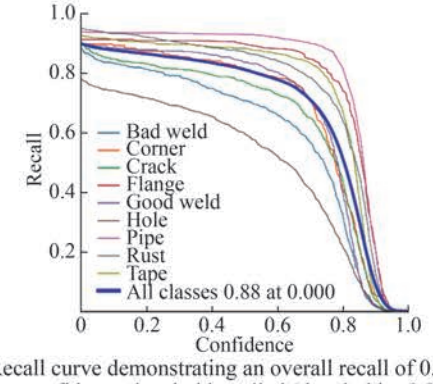
(a) Precision-recall curve showing a mean average precision (mAP) of 0.874 across all classes at an intersection over union threshold of 0.5



(b) F1 score curve indicating an overall F1 score of 0.88 at a confidence threshold of 0.370



(c) Precision curve showing perfect precision (1.00) at a confidence threshold of 0.909 across all classes



(d) Recall curve demonstrating an overall recall of 0.88 with no confidence threshold applied (threshold = 0.000)

Figure 7 Performance curves of X-Large model

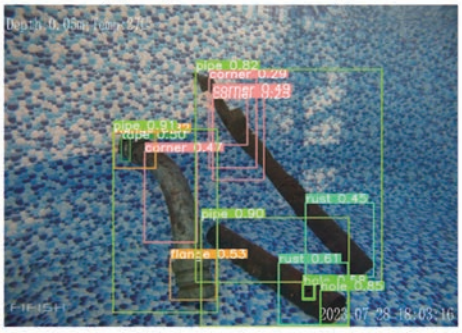
lights the model's adaptability to shifting thresholds, which is useful in determining optimal confidence levels for practical deployment. Finally, Figure 7(d) shows the recall curve, which offers a comprehensive view of the model's capability to detect objects across various confidence thresholds.

In conclusion, the interplay between precision and recall, particularly for the "bad weld" and "hole" classes, highlights the model's strength in balancing detection accuracy with comprehensive coverage. This balance is critical for effective defect identification and reliable underwater pipeline inspections, which render the YOLOv8 X-Large configuration a valuable tool in this domain.

6.1 Visual interpretation of detection results

Visual interpretation was an essential tool during the evaluation of our detection model. This section compares our detection performance against raw, unlabeled images, which showcases the model's precision. As displayed in Figure 8, the original image presented a wide range of objects and defects commonly found in underwater environments. Specifically, elements such as a "pipe", "corner", and "flange" can be discerned. In addition, noticeable signs of "rust" and "tape" were observed.

(a) Original unlabeled image



(b) Detected objects with corresponding classifications

Figure 8 Original and visual interpretation images

When analyzed by our model, the processed image successfully detected and highlighted these elements and marked them with distinct bounding boxes and class labels. The bounding boxes provided precise spatial localization of each defect, and the associated class labels clearly categorized them into their respective defect types. A visual representation such as this illustrates the model's capability to pinpoint and categorize underwater elements and anomalies, even in challenging underwater environments with poor visibility and complex object backgrounds. This capability demonstrates the effectiveness of YOLOv8 X-Large in supporting defect detection tasks critical for underwater pipeline maintenance and marine research.

6.2 Understanding the confusion matrix

Confusion matrices are commonly used to assess the effectiveness of classification models. With test images used as a basis, a number of performance metrics, including F1 score, accuracy, precision, and recall, can be computed using confusion matrices. These metrics offer a quantitative overview of the model's capability to correctly classify each defect type, which helps in the identification of areas of strength and improvement.

Confusion matrices were calculated by comparing the actual class label with the predicted class label (Selcuk and Serif, 2023). Figure 9 shows the confusion matrix for the YOLOv8 X-Large model. In this matrix, rows represent the true classes (ground truth labels), and columns indicate the predicted ones (model outputs). The confusion matrix uses color intensity to represent quantity: darker shades of blue indicate higher numbers, which mark an improved model performance when appearing on the diagonal from the top left to the bottom right. Lighter shades, especially those off the diagonal, highlight areas where the model struggles to make correct classifications.

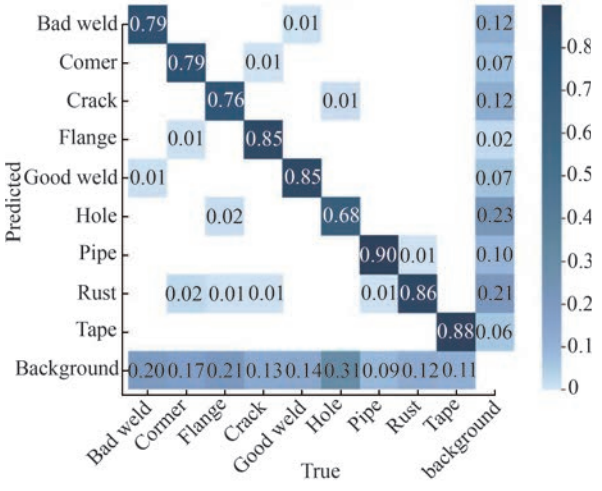


Figure 9 Normalized confusion matrix for X-Large model

The matrix shows that the model performed well in identifying "flange", "goodweld", "pipe", "rust", and "tape" classes. This finding suggests that the model effectively learned to distinguish these defects, possibly due to their

distinct visual characteristics and sufficient representation in the dataset. However, the model faced challenges with regard to the "background" and "hole" classes. The confusion between these classes may arise due to their similar shapes and appearances and the small bounding box sizes associated with "hole" defects. In addition, the limited representation of these classes in the dataset contributed to the difficulty of classification. Addressing these issues can involve the expansion of the dataset with more examples of "hole" defects and improvement of the labeling accuracy for small bounding boxes.

7 Discussion

This section compares the performance of various models on the designated dataset. The selected models, including YOLOv8 (Nano, Small, Medium, and Large, X-Large), YOLOv3 variants (YOLOv3-sppu, YOLOv3-tinyu, and YOLOv3u), YOLOv5 versions (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5xl), and SSD models (SSD-MobileNETV2, and SSD-EfficientNET), were evaluated to present a comprehensive analysis. Each model was analyzed to comprehend its distinct strengths, weaknesses, and optimal application scenarios. Key performance indicators, such as precision, recall, and mAP50, were pivotal in ascertaining the model's efficiency, accuracy, and overall efficacy in detecting and identifying objects within the underwater dataset.

YOLOv3 was initially introduced in ArXiv in 2018 (Redmon and Farhadi, 2018). With its expanded architecture and enhanced performance, YOLOv3 is often used as a benchmark in object detection. Recognized for integrating predictions across different scales, the model can provide predictions across multiple grid sizes, which ensures a high level of detail and enhances small-object and defect recognition and detection. The fundamental idea behind YOLOv3 is Darknet-53, which replaces traditional max-pooling layers with stridden convolutions and is enriched with residual connections, resulting in a composition of 53 convolutional layers (He et al., 2015). Given this architectural skill, the results obtained include enhanced bounding box prediction and class recognition, which improve detection tasks' accuracy and detail. The SPP bit enhances YOLOv3's capability to take in deep features at multiple scales, which improves the model's adaptability and effectiveness huang2020dc (He et al., 2015). The Tiny-YOLOv3 emerges with a reduced number of convolutional layers. This structure ensures memory efficiency and detection acceleration. However, it necessitates a compromise on detection accuracy (Adarsh et al., 2020). YOLOv3-Ultralytics improves the original model's detection, especially for small objects. With support for multiple pretrained models and additional customization, it increases bendiness, which works well across different detection situations and scenarios (Shen et al., 2023). YOLOv5 powerfully evolves in the YOLO series. It offers tailored versions, such as Nano, Small, Medium, Large, and X-Large, which meet varied operational and computational requirements. With its core YOLO algorithm at its heart, YOLOv5 detects small objects in remote sensing images while balancing accuracy with speed. Object detection is efficient and precise across diverse applications due to the optimization of each version for specific scenarios and cases (Jocher et al., 2022).

The SSD-MobileNetV2 model, which is designed for maximum speed and efficiency, combines the SSD with the MobileNetV2 model. SSD ensures accurate object location and classification, and MobileNetV2 facilitates feature extraction and parameter reduction, which makes the model extremely efficient. Consequently, SSD-MobileNetV2 can be effectively deployed in embedded devices, which renders it suitable for applications with limited hardware resources. The adoption of transfer learning enhances the model's performance by incorporating information from pretrained models, which ensures rapid and efficient object detection while maintaining a compact and efficient architecture (Cheng, 2022). With SSD-EfficientNet, SSD's accuracy is combined with EfficientNet's multiscale capabilities. However, given the latter's layer complexity, memory consumption increases, and the training period is slower. By balancing improved detection performance with efficiency, improvements in the loss and activation functions have mitigated these challenges (Cao et al., 2021; Liu et al., 2016). Table 4 comprehensively presents various object detection models, including variants of YOLOv8, YOLOv3, SSD-EfficientNet, and SSD-MobileNetV2 and provides details on their performance metrics. A number of metrics were used to evaluate each model, including precision, recall, mAP50, mAP50- 95, and time. These metrics provide insights into the models' capability to accurately detect objects, their speed, and overall efficiency and provide a multifaceted perspective for a thorough comparison. As shown in Table 4, the different models exhibited superior performance in certain areas. The YOLO8 X-Large scored 0.961, which indicates a strong capability to minimize false positives. With an impressive 0.929 recall score, the YOLO8 X-Large was the most proficient in capturing a high number of true positives. Using mAP50, YOLO8 X-Large again stood out with a score of 0.966, which illustrates its exceptional object localization capabilities at an IoU threshold of 50%. As a result, the mAP50-95 metric was also high and exhibited a balanced performance across various IoU thresholds, with a score of 0.746. A quick processing speed of 12.4 ms distinguishes YOLOv3-Tiny as the top choice for applications requiring real-time analytics and instant results.

Table 4 Performance comparison of different YOLOv models and SSD varian

Model	Precision	Recall	mAP50	mAP50-95	Time (ms)
YOLOv8 Nano	0.844	0.720	0.800	0.491	12.5
YOLOv8 Small	0.922	0.831	0.893	0.612	15.1
YOLOv8 Medium	0.952	0.862	0.908	0.670	18.4
YOLOv8 Large	0.960	0.892	0.933	0.711	21.2
YOLOv8 X-Large	0.961	0.929	0.966	0.746	22.5
YOLOv3-SPP	0.960	0.906	0.948	0.738	21.7
YOLOv3-Tiny	0.827	0.969	0.760	0.485	12.4
YOLOv3-U	0.960	0.915	0.956	0.742	14.1
YOLOv5 Nano	0.816	0.717	0.791	0.472	12.6
YOLOv5 Small	0.906	0.839	0.909	0.598	13.2
YOLOv5 Medium	0.947	0.887	0.942	0.659	16.1
YOLOv5 Large	0.956	0.848	0.910	0.654	19.2
YOLOv5 X-Large	0.950	0.839	0.889	0.644	28.5
SSD-MobileNETV2	0.884	0.810	0.890	0.560	13.0
SSD-EfficientNET	0.930	0.925	0.965	0.688	31.2

Underwater dataset challenges

Obtaining high image quality is a typical challenge in most underwater environments. In this experiment, the additional challenge was due to water contamination caused by pipe rust, which renders it murky and obscuring the image. To combat this issue, we instituted regular pool cleanings to ensure that images were captured in relatively clear water. However, lighting inconsistencies also posed a challenge in clean water. The level of exposure of an image depends on various factors, such as the time of the day, source of light (natural or artificial), and light angle. This problem was addressed by collecting data under various lighting conditions to provide a diverse dataset. Despite implementing these steps to improve image quality, challenges remained. The intricacies of underwater imaging mean that even slight visibility issues can affect the level of detail captured, which complicates the annotation process and, by extension, model training.

Data augmentation played a crucial role in refining our dataset. Figure 10 illustrates the enhancement in dataset diversity and showcases the noticeable improvement in the YOLOv8 X-Large model's performance metrics. Extend-

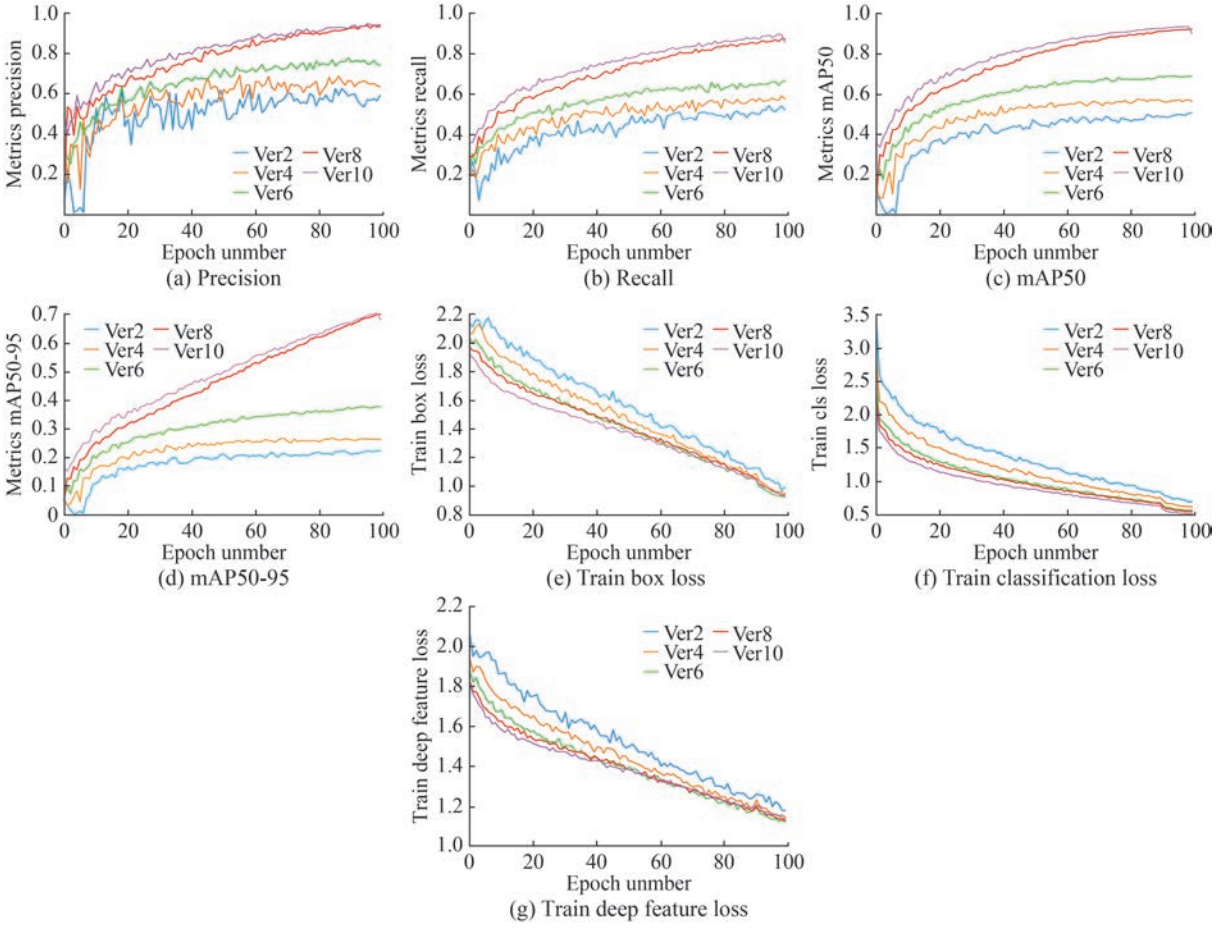


Figure 10 Performance and loss metrics of YOLOv8x-large with increased dataset size

ing the dataset proved instrumental in elevating precision, recall, mAP50, and mAP50-95 and contributed to reduced box, class, and DFL losses. These findings highlight the substantial impact of a rich, varied dataset on the optimization of the efficiency and accuracy of underwater object detection models.

To overcome the limitations of the dataset further, future research should focus on the diversification of the dataset using synthetic image generation techniques. Generative adversarial networks can be employed to create realistic underwater images simulating varying levels of turbidity, light scattering, and object orientations. These synthetic images can complement real-world datasets to address gaps in variability and representation.

Figure 10(a) illustrates a positive correlation between dataset size and model precision. Each colored line represents a different version of the dataset, with the green line (version 10) displaying optimal precision, which highlights the advantage of a larger dataset. As shown in Figure 10(b), recall increased with the increase in dataset size, which demonstrates the model's capability to identify relevant instances with dataset expansion. The green line, which represents version 10, outperformed the others. This finding indicates that the expansion of the dataset effectively increased the recall rate of the model.

In addition, employing advanced data acquisition techniques, such as multi-angle imaging and spectral filtering, can further enhance dataset quality. These approaches can mitigate limitations caused by object occlusions and color distortion, which enable the capture of more detailed visual information.

Figure 10(c) illustrates the comparison of the mAP50 of different model versions. Version 10, depicted by a green line, clearly displayed a superior performance. At an intersection over the union of 50%, enhanced mean and average precision correlated with an enhanced dataset. Based on the mAP50-95 metric shown in Figure 10(d), Ver10 (green line) outperformed the other model versions, demonstrating more accurate detection and localization of objects and comprehensive dataset enrichment.

Figure 10(e) shows a distinct contrast in the training box loss among various versions, with Ver10 exhibiting the lowest loss, which is indicative of superior object localization. Ver2, despite showing a reduction in loss, lagged behind, which underscores Ver10's efficiency. Each version's trajectory provides insights into their respective learning efficiencies and object localization accuracy. In the training class loss graph, Ver10 outperformed the others, which denotes its enhanced capability to classify objects accurately. Figure 10(f) illustrates the comparative efficiency of each version in learning and classifying objects. Figure 10(g) highlights Ver10's initial low error and showcases its efficient object count parameter tuning in early training. Ver6 closely followed, showing a refined accuracy over epochs, whereas Ver2 consistently lagged. By the 50th epoch, Ver6 almost matched Ver10, which proves that increased training data effectively reduced DFL loss and improved object count precision in images.

8 Conclusions

This research evaluated the performance of the YOLOv8 X-Large model in detecting and classifying defects in underwater pipelines. The results demonstrate that YOLOv8 X-Large can effectively identify critical defects, including holes and bad welds, and balance precision and recall. This model also showed a strong performance across a wide range of pipeline defects, which was particularly reflected by its high recall rate. YOLOv8 performed well in identifying defect types, such as cracks, rust, corners, welds, flanges, and tapes. In comparison with other models, including YOLOv3 variations (SPP, Tiny, and U), YOLOv5 versions (Nano, Small, Large, XLarge, and Medium), and SSD models, the YOLOv8 X-Large model achieved higher precision and recall rates, which demonstrate its superiority in defect detection under challenging underwater conditions (Table 4). This outcome was evident in the detection of small objects, such as holes, and subtle features, such as cracks. Key architectural features of YOLOv8, such as its CSPNet backbone for efficient feature extraction and anchor-free detection, enabled its processing of high-resolution images with minimal computational overhead, which ensured the precise detection of small defects. The FPN+PAN neck facilitated multiscale detection and effectively addressed light scattering and distorted object details in underwater imagery. Altogether, these features allowed YOLOv8 to adapt well to the challenges of underwater environments, including complex backgrounds and low visibility.

This study also introduced a custom dataset on underwater pipeline defects, and it was tailored to address challenges, such as varying defect types, light scattering, and visibility issues. Although the dataset was designed for this study, it has the potential to serve as a benchmark for future research on underwater object detection. The findings of this paper not only establish a baseline performance for YOLOv8 in underwater environments but also provide insights into leveraging advanced object detection frameworks for sub-aquatic systems. CSPNet further enhanced YOLOv8's capability to represent objects in low-contrast environments, and its focal loss function prioritized the detection of subtle or rare features, such as small cracks or faint defects, which ensure robust performance under adverse conditions.

In conclusion, the YOLOv8 X-Large model is an effective algorithm for underwater exploration and structural health monitoring. It paves the way for further research on

leveraging deep learning to improve the reliability and cost-efficiency of pipeline inspections. To overcome the challenges of light scattering and varying visibility, this study employed advanced preprocessing techniques, such as contrast normalization and adaptive resizing, to standardize input images and reduce distortions. The FPN + PAN neck architecture enabled the aggregation of multi-scale features and ensured accurate detection of small objects, such as cracks and holes, in noisy underwater conditions. In addition, the anchor-free detection mechanism simplified bounding box predictions, which enabled robust performance under varying object scales and environmental complexities.

Future work

Future work will refine the YOLOv8 X-Large model to improve its accuracy in detecting underwater defects, guided by insights obtained from the confusion matrix in our current study. This research identified specific challenges associated with the detection of certain types of defects accurately, such as holes (Figure 9). As a result, the model's performance in real-world scenarios must be improved through the acquisition of a more realistic and diverse dataset. Advanced data enhancement techniques should be implemented to enhance the quality of training data. Given the unique challenges posed by underwater environments, including variable lighting and visibility conditions, increasingly sophisticated image preprocessing methods must be employed. Advanced preprocessing techniques, such as adaptive histogram equalization or gray-world color correction, can be explored to mitigate the effects of light scattering and color distortions. Moreover, noise reduction filters can be employed to address artifacts caused by turbidity and suspended particles, which will enhance the overall quality of the input images.

The model's architecture also necessitates further refinements. Spatial and temporal attention mechanisms can be integrated to enable the model to focus on key features and improve object tracking across sequences of images. Incorporating adaptive weighted feature pyramids can enhance the model's capability to handle variations in object scale, which will ensure the consistent detection of small defects such as holes.

For dealing with specific underwater challenges, preprocessing techniques, such as color channel compensation and denoising algorithms, can be adopted to improve image clarity further. Incorporating domain-specific filters to handle turbidity artifacts will also ensure better data quality for training robust models.

To further improve its capability to distinguish between different types of defects, we aimed to refine the model's parameters based on the current findings, including the adjustment of thresholds and tuning of hyperparameters. Dataset augmentation is also critical and involves the creation of synthetic images that simulate underwater conditions, such as low light, high turbidity, and uneven illumination. Including more examples of rare defects, such as faint cracks, will ensure better class representation and reduce misclassification rates.

Exploration of domain adaptation techniques will allow the more effective generalization of the model to unobserved underwater conditions. In addition, comparative evaluations with other state-of-the-art models, such as YOLO variants and Faster R-CNN, can help benchmark the YOLOv8 X-Large model's performance across metrics, including precision, recall, and mAP.

By addressing these areas, critical structural flaws can be detected more effectively, which will contribute to the safety and long-term functionality of underwater infrastructures. These refinements will ensure that the YOLOv8 X-Large model achieves greater accuracy and reliability and thus pave the way for more effective underwater inspection and monitoring solutions.

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

References

Aboah A, Wang B, Bagci U, Adu-Gyamfi Y (2023) Real-time multi-class helmet violation detection using few-shot data sampling technique and YOLOv8. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE 5349-5357. https://doi.org/10.1109/CVPRW59228.2023.00564

Adarsh P, Rathi P, Kumar M (2020) YOLO v3-Tiny: Object detection and recognition using one stage improved model. Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Piscataway: IEEE 687-694. https://doi.org/10.1109/ICACCS48705.2020.9074315

Al Muksit A, Hasan F, Emon MFHB, Haque MR, Anwary AR, Shatabda S (2022) YOLO-fish: A robust fish detection model to detect fish in realistic underwater environment. Ecological Informatics 72: 101847. https://doi.org/10.1016/j.ecoinf.2022.101847

Asyraf MS. Isa IS, Marzuki MIF, Sulaiman SN, Hung CC (2021) CNN-based YOLOv3 comparison for underwater object detection. Journal of Electrical and Electronic Systems Research 18: 30-37. https://doi.org/10.24191/jeesr.v18i1.005

Avsar E, Feekings JP, Krag LA (2023) Estimating catch rates in real time: Development of a deep learning based Nephrops (Nephrops norvegicus) counter for demersal trawl fisheries. Frontiers in Marine Science 10: 1129852. https://doi.org/10.3389/fmars.2023.1129852

Burguera A, Bonin-Font F (2022) Advances in autonomous underwater robotics based on machine learning. Journal of Marine Science and Engineering 10(10): 1481. https://doi.org/10.3390/jmse10101481

Cao C, Yu Y, Xie Y, Sun C (2021) An efficient approach for gastric polyps detection based on improved SSD. Proceedings of the 2021 China Automation Congress (CAC). Piscataway: IEEE 852-857. https://doi.org/10.1109/CAC52703.2021.9727623

Chen Y, Li Q, Lu D, Kou L, Ke W, Bai Y, Wang Z (2023) A novel

underwater image enhancement using optimal composite backbone network. Biomimetics 8(3): 275. https://doi.org/10.3390/biomimetics8030275

Cheng C (2022) Real-time mask detection based on SSD-MobileNetV2. In Proceedings of the 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). Piscataway: IEEE 761-767. https://doi.org/10.1109/AUTEEE56487.2022.9994442

Chi Y, Zhang C (2024) Underwater image enhancement methods using biovision and type-II fuzzy set. Journal of Marine Science and Engineering 12(11): 2080. https:/ /doi.org/10.3390/ jmse12112080

Du P, Song X (2024) Lightweight target detection: An improved YOLOv8 for small target defect detection on printed circuit boards. Proceedings of the 2024 International Conference on Generative Artificial Intelligence and Information Security (GAIIS). New York: ACM 329-334. https://doi.org/10.1145/3665348.3665404

Fayaz S, Parah SA, Qureshi GJ (2022) Underwater object detection: Architectures and algorithms–A comprehensive review. Multimedia Tools and Applications 81(15): 20871-20916. https://doi.org/10.1007/s11042-022-12502-1

Fu C, Liu R, Fan X, Chen P, Fu H, Yuan W, Zhu M, Luo Z (2023) Rethinking general underwater object detection: Datasets, challenges, and solutions. Neurocomputing, 517: 243-256. https://doi.org/10.1016/j.neucom.2022.10.039

Gao Y, Liu W, Chui H-C, Chen X (2024) Large span sizes and irregular shapes target detection methods using variable convolution-improved YOLOv8. Sensors 24(8): 2560. https://doi. org/10.3390/s24082560

Gašparović B, Lerga J, Mauša G, Ivašić-Kos M (2022) Deep learning approach for objects detection in underwater pipeline images. Applied Artificial Intelligence 36(1): 2146853. https://doi.org/10.1080/08839514.2022.2146853

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE 580-587. https://doi.org/10.1109/CVPR.2014.81

Han M, Lyu Z, Qiu T, Xu M (2020) A review on intelligence dehazing and color restoration for underwater images. IEEE Transactions on Systems, Man, and Cybernetics: Systems 50(5): 1820-1832. https://doi.org/10.1109/TSMC.2017.2788902

He K, Zhang X, Ren S, Sun J (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9): 1904-1916. https://doi.org/10.1007/978-3-319-10578-9_23

Jin ZZ, Zheng YF (2020) Research on application of improved YOLO V3 algorithm in road target detection. Journal of Physics: Conference Series, 1654: 012060. https://doi.org/10.1088/1742-6596/1654/1/012060

Jocher G, Chaurasia A, Stoken A, Borovec J, NanoCode012, Kwon Y, Michael K, Taoxie, Fang J, Imyhxy, Lorna, Zeng Y, Wong C, Abhiram V, Montes D, Wang Z, Fati C, Nadar J, Laughing, UnglvKitDe, Sonck V, tkianai, YxNong, Skalski P, Hogan A, Nair D, Strobel M, Jain M (2022) ultralytics/yolov5: v7.0-YOLOv5 SOTA realtime instance segmentation. Zenodo. https://doi.org/10.5281/zenodo.7347926

Karimanzira D, Renkewitz H, Shea D, Albiez J (2020) Object detection in sonar images. Electronics, 9(7): 1180. https://doi.org/10.3390/electronics9071180

Kim JH, Kim N, Won CS (2023) High-speed drone detection based on YOLO-V8. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Piscataway: IEEE 1-2. https://doi.org/10.1109/ICASSP49357.2023.10095516

Kim N, Kim JH, Won CS (2022) FAFD: Fast and accurate face detector. Electronics 11(6): 875. https://doi.org/10.3390/electronics11060875

Li C, Li L, Jiang H, Weng K, Geng Y, Li L, Ke Z, Li Q, Cheng M, Nie W, Li Y, Zhang B, Liang Y, Zhou L, Xu X, Chu X, Wei X, Wei X (2022) YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint, arXiv: 2209.02976. https://arxiv.org/abs/2209.02976

Li H, Gu Z, He D, Wang X, Huang J, Mo Y, Li P, Huang Z, Wu F (2024) A lightweight improved YOLOv5s model and its deployment for detecting pitaya fruits in daytime and nighttime light-supplement environments. Computers and Electronics in Agriculture 220: 108914. https://doi.org/10.1016/j.compag.2024. 108914

Liu K, Sun Q, Sun D, Peng L, Yang M, Wang N (2023) Underwater target detection based on improved YOLOv7. Journal of Marine Science and Engineering 11(3): 677. https://doi.org/10.3390/jmse11030677

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single shot multiBox detector. Computer Vision – ECCV 2016: 14th European Conference. Berlin: Springer Part I: 21-37. https://doi.org/10.48550/arXiv.1512.02325

Meng F, Li J, Zhang Y, Qi S, Tang Y (2023) Transforming unmanned pineapple picking with spatio-temporal convolutional neural networks. Computers and Electronics in Agriculture 214: 108298. https://doi.org/10.1016/j.compag.2023.108298

Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Le HV, Tran VQ, Prakash I, Pham BT (2021) Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Mathematical Problems in Engineering 2021: 4832864. https://doi.org/10.1155/2021/4832864

Orinaitė U, Karaliūtė V, Pal M, Ragulskis M (2023) Detecting underwater concrete cracks with machine learning: A clear vision of a murky problem. Applied Sciences 13(12): 7335. https://doi.org/10.3390/app13127335

Park CW, Eom IK (2024) Underwater image enhancement using adaptive standardization and normalization networks. Engineering Applications of Artificial Intelligence 127(Part A): 107445. https://doi.org/10.1016/j.engappai.2023.107445

Pavani D, Reddy ANN, Saw N, Prasad S, Naik, SM (2023) Octacleaner: Underwater trash detection through YOLO. Proceedings of the 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC). Piscataway: IEEE 1-6. https://doi.org/10.1109/ICMNWC60182.2023.10435715

QYSEA (2022) FIFISH V6 plus official website. https://www.qysea.com/v6-expert.html (Accessed: 23 August 2022)

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE 779-788. https://doi.org/10.1109/CVPR.2016.91

Rahman S, Rony JH, Uddin J, Samad MA (2023) Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography. Journal of Imaging 9(10): 216. https://doi.org/10.3390/jimaging9100216

Raza K, Hong S (2020) Fast and accurate fish detection design with improved YOLO-v3 model and transfer learning. International Journal of Advanced Computer Science and Applications 11(2): 7-16. https://doi.org/10.14569/IJACSA.2020.0110202

Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement.

arXiv preprint, arXiv: 1804.02767. https://doi.org/10.48550/arXiv.1804.02767

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 28 (NeurIPS) arXiv: 1506.01497. https://doi.org/10.48550/arXiv.1506.01497

Selcuk B, Serif T (2023) A comparison of YOLOv5 and YOLOv8 in the context of mobile UI detection. In Proceedings of the International Conference on Mobile Web and Intelligent Information Systems 161-174. https://doi.org/10.1007/978-3-031-39764-6_11

Shankar R, Muthulakshmi M (2023) Comparing YOLOv3, YOLOv5 & YOLOv7 architectures for underwater marine creatures detection. Proceedings of the 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). Piscataway: IEEE 25-30. https://doi.org/10.1109/ICCIKE58312.2023.10131703

Shen L, Tao H, Ni Y, Wang Y, Stojanovic V (2023) Improved YOLOv3 model with feature map cropping for multi-scale road object detection. Measurement Science and Technology 34(4): 045406. https://doi.org/10.1088/1361-6501/acb075.s

Soorma MS, Chaudhary A, Sonali S, Pal S, Upadhyay DK (2023) Underwater image processing with normalized AttUNet. 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3). Piscataway: IEEE 1-5. https://doi.org/10.1109/IC2E357697.2023.10262737

Vidhya SK, Deepthi PS (2023) A comprehensive analysis of underwater image processing based on deep learning techniques. Proceedings of the 2023 International Conference on Control, Communication and Computing (ICCC). Piscataway: IEEE 1-6. https://doi.org/10.1109/ICCC57789.2023.10165168

Wang T, Li Y, Zhai Y, Wang W, Huang R (2023) A sewer pipeline defect detection method based on improved YOLOv5. Processes 11(8): 2508. https://doi.org/10.3390/pr11082508

Xu S, Zhang M, Song W, Mei H, He Q, Liotta A (2023) A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing 527: 204-232. https://doi.org/10.1016/j.neucom.2023.01.056

Yaseen M (2024) What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector. arXiv preprint arXiv: 2409.07813. https://arxiv.org/abs/2409.07813

Zhang H, Zhang S, Wang Y, Liu Y, Yang Y, Zhou T, Bian H (2021) Subsea pipeline leak inspection by autonomous underwater vehicle. Applied Ocean Research 107: 102321. https://doi.org/10.1016/j.apor.2020.102321

Zhang H, Dai C, Chen C, Zhao Z, Lin M (2024) One stage multi-scale efficient network for underwater target detection. Review of Scientific Instruments 95(6): 065108. https://doi.org/10.1063/5.0206734

Zhang J, Liu X, Zhang X, Xi Z, Wang S (2023) Automatic detection method of sewer pipe defects using deep learning techniques. Applied Sciences 13(7): 4589. https://doi.org/10.3390/app13074589

Zhang L, Lin L, Liang X, He K (2016) Is faster R-CNN doing well for pedestrian detection? Computer Vision–ECCV 2016: 14th European Conference. ECCV 2016. Lecture Notes in Computer Science Part II: 443-457. https://doi.org/10.1007/978-3-319-46475-6_28

Zhang Y, Ni Q (2023) A novel weld-seam defect detection algorithm based on the S-YOLO model. Axioms 12(7): 697. https://doi.org/10.3390/axioms12070697

Zhou H, Kong M, Yuan H, Pan Y, Wang X, Chen R, Lu W, Wang R, Yang Q (2024) Real-time underwater object detection technology for complex underwater environments based on deep learning. Ecological Informatics 82: 102680. https://doi.org/10.1016/j.ecoinf.2024.102680

Zhong J, Gao C, Tian Y, Zhang M (2023) Research on the influence of hydrodynamic analysis to dynamic modeling of underwater manipulator. Proceedings of the 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Piscataway: IEEE 6: 982-986. https://doi.org/10.1109/ITNEC56291.2023.10082093

Zhao S, Zheng J, Sun S, Zhang L (2022) An improved YOLO algorithm for fast and accurate underwater object detection. Symmetry 14(8): 1669. https://doi.org/10.3390/sym14081669