Theory and Practice of Science and Technology
2025, VOL. 6, NO. 10, 229-233
DOI: 10.47297/taposatWSP2633-456945.20250610

Enhancing Drowning Detection Efficiency: A YOLOv11 Approach with Lightweight Improvements

Benbo Cao

Nanjing University of Information Science & Technology, Nanjing, Jiangsu, 210044, China

ABSTRACT
Swimming, while a popular activity worldwide, poses significant safety risks, with drowning being a leading cause of accidental death, particularly among youth. Traditional surveillance methods, such as lifeguard monitoring, are prone to fatigue and high costs, highlighting the urgent need for automated drowning detection systems. This study explores the application of the YOLOv11 model, a state-of-the-art one stage object detector, for real-time drowning detection. To enhance its suitability for edge devices, this paper introduces a lightweight C3K2 module through structural improvements and incorporates a BiFPN architecture into the neck network. These modifications reduce the number of parameters and enhance the inference speed of the model. Experimental results demonstrate that the pruned YOLOv11 model achieves a high mean average precision (mAP) on a custom dataset containing three critical categories (drowning, swimming, and floating). The study validates the feasibility of the proposed approach, which effectively balances performance and efficiency, and provides a practical, scalable solution for enhancing water safety through intelligent surveillance.

KEYWORDS
Drowning detection; YOLOv11; StarNe; DW conv; BiFPN; Real-time detection; Swimming safety

1 Introduction

Swimming remains one of the most popular physical activities worldwide, yet it continues to pose significant safety risks [1]. With rising summer temperatures, the number of people swimming in both open and enclosed venues has increased—and consequently, so have drowning incidents. According to World Health Organization (WHO) statistics from 2024, drowning accounts for approximately 300,000 fatalities annually, making it a critical global health issue [2]. In China, it is the leading cause of accidental death among children and adolescents [3]. These alarming figures underscore the urgent need for more effective safety mechanisms, especially in aquatic environments.

Currently, many swimming facilities rely primarily on human lifeguards for surveillance. This approach is not only labor-intensive and costly but also susceptible to limitations such as visual fatigue, distractions, or difficulties in monitoring large crowds. As a result, a considerable number of drowning incidents go undetected until it is too late [4]. This gap has motivated research into automated drowning detection systems that can operate reliably in real time, thereby complementing or even replacing traditional human oversight. Early approaches to drowning detection often depended on physical sensors, such as pressure-based wearable devices or underwater acoustic sensors [5]. While useful in specific scenarios, these methods require users to wear equipment, which may be intrusive, prone to consumption or damage, and limited in coverage. More recently, computer vision-based techniques have emerged as a non-invasive and highly scalable alternative [6]. These methods leverage surveillance cameras and deep learning algorithms to analyze swimmers ' behavior without any physical contact. Within computer vision, object detection models are broadly categorized into two-stage and one-stage frameworks [7]. Two-stage detectors (e. g., Faster R-CNN) first generate region proposals and then classify them, yielding high accuracy but at a slower inference speed [8]. In contrast, one-stage models (e. g., the YOLO family) perform localization and classification in a single forward pass, making them significantly faster and better suited for real-time applications [9]. Some one-stage detectors can be more than three times faster than their two-stage counterparts, which is critical in time-sensitive drowning prevention first.

Owing to this need for rapid response, recent research in vision-based drowning detection has largely focused on one-stage architectures [10]. Based on the improvements of deformable convolution, deformable attention mechanism, auxiliary detection heads, and InnerIOU loss function, a Swimming-YOLO model was applied to enhance the accuracy of drowning person detection in complex swimming scenarios [11]. A GEF-YOLO model which replaced the C2f modules in the backbone with C2f-EMBC modules containing EMBConv was used on detecting drowning risks at sea [12]. While these methods marked important advances, they still faced challenges in terms of detection accuracy, robustness under complex water conditions, and computational efficiency. YOLOv11 introduces architectural innovations like the C3k2 module and C2PSA component, significantly enhancing feature extraction capabilities and improving detection accuracy, particularly for complex scenes and small objects [13]. It also optimizes model design and training strategies, reducing the number of parameters while maintaining high precision and achieving faster inference speeds, thus balancing efficiency and performance [13].

Given these advantages, we adopt YOLOv11 as our primary detection framework for recognizing drowning incidents. To further optimize the model for real-time use in edge device environments with limited computational power, we also incorporate StarNet and BiFPN to decrease the computation load and the amount of the parameters. The star operation possesses the unique advantage of performing computations in a low-dimensional space while simultaneously generating high-dimensional features, a characteristic that highlights its significant potential in the design of efficient network architecture. StarNet, built upon the star operation, significantly reduces the reliance on manual intervention with its minimalist structural design [14]. Despite its simple architecture, it delivers outstanding performance, fully validating the effectiveness of the star operation [15]. Furthermore, this work replaces standard convolution layers with depthwise separable convolution (DWConv), a common lightweight networking technique, thereby effectively reducing the computational complexity of the prediction process [16]. In the Neck component, the introduction of Bidirectional Feature Pyramid Network (BiFPN) effectively reduces the parameter quantity by simplifying the network structure and optimizing connection paths [17]. Specifically, BiFPN removes nodes with only one input edge and adds extra connections between input and output nodes at the same level, thereby eliminating redundant elements while enhancing feature fusion efficiency [18]. This streamlined design not only decreases computational complexity but also strengthens bidirectional information flow across different scales, contributing to improved model performance.

Overall, by integrating YOLOv11's state-of-the-art detection capabilities with strategic model improvement, our approach realizes a high-performance, real-time drowning detection system that significantly enhances safety in aquatic environments. Firstly, applying the YOLOv11 model to drowning detection leads to major breakthroughs in both detection speed and accuracy. Secondly, lightweight improvements considerably reduces the model's parameter count while further improving detection speed, making it more suitable for deployment on edge devices. This integrated solution demonstrates strong potential for practical implementation in real-world aquatic safety monitoring systems.

The rest of the paper is organized as follows. Sec. 2 presents the design of the drowning detection model. Sec. 3 analyzes the performance of the model. Sec. 4 concludes this paper.

2 Drowning Detection Model Design

2.1 YOLOv11 model introduction

The YOLOv11 model, introduced by the Ultralytics team in 2024 as the latest iteration in the YOLO series, retains the three-stage architecture of Backbone-Neck Head while achieving significant performance improvements over its predecessors. This paper utilizes YOLOv11 for drowning detection and further applies pruning techniques to enhance its rapid response capability.

2.2 Lightweight Modifications to the Backbone component

The star operation typically denotes element-wise multiplication, used to fuse two features, and its mathematical relationship can be described by specific Eq. 1.

$$\begin{aligned} & w_1^{\mathrm{T}} x * w_2^{\mathrm{T}} x \\ & = \left(\sum_{i=1}^{d+1} w_1^i x^i \right) * \left(\sum_{j=1}^{d+1} w_2^j x^j \right) \\ & = \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j \end{aligned}$$

In this formula, $x \in \mathrm{R}^{d+1}$ denote the input feature vector, where the additional dimension accounts for a bias term. Let $w_1, w_2 \in \mathrm{R}^{d+1}$ be two independent weight vectors associated with distinct linear projections. The indices I and j are summation indices iterating from 1 to $d+1$, and w_1^i, w_2^i, x_i, and x_j represent the i-th or j-th components of the corresponding vectors. The operator $*$ denotes the star operation, defined as the scalar multiplication of the two dot products $w_1^{\mathrm{T}} x$ and $w_2^{\mathrm{T}} x$. The expansion reveals that the operation computes a double summation over all pairwise products of the input components and their corresponding weights, effectively modeling second-order feature interactions.

This operation significantly enhances feature representation capability without introducing excessive computational overhead in a single-layer structure. StarNet, constructed by stacking multiple star operation layers, effectively maps features into a high-dimensional space while maintaining low computational complexity, thereby alleviating the inherent trade-off between computational efficiency and model performance in traditional efficient network designs. By simplifying the network architecture, StarNet achieves efficient feature representation without relying on complex multi-branch structures or elaborate fusion mechanisms. In this work, we replace the BottleNet layer in the original C3K2 module with StarNet to construct a lightweight version named C3K2_light, whose structure is illustrated in Figure 1.

Furthermore, Depthwise Separable Convolution (DWConv), as a common lightweight convolutional technique, decomposes standard convolution into depthwise convolution and pointwise convolution, significantly reducing both

parameter count and computational burden. To further advance network lightweighting, we replace the standard convolutional layers in the Backbone with DWConv structures. The optimized Backbone component is depicted in Figure 2.

2.3 BiFPN introduction to the Neck component

In the Neck component, this study replaces the traditional FPN architecture with BiFPN (Bidirectional Feature Pyramid Network), an efficient multi-scale feature fusion network optimized based on the conventional Feature Pyramid Network (FPN). BiFPN introduces efficient bidirectional cross-scale connections, allowing information to flow and integrate more effectively between features of different scales through both top-down and bottom-up paths. This design enables a more comprehensive combination of multi-scale features. The network structure is simplified by removing nodes with only one input edge and adding extra connections between original input and output nodes at the same level. Each bidirectional path is treated as a feature network layer and can be repeated multiple times to achieve higher-level feature fusion while reducing parameters. Furthermore, BiFPN incorporates weighted feature fusion, introducing learnable weights to assess the importance of different input features, thereby enhancing the effectiveness of feature integration. By adopting BiFPN, the model achieves a further increase in inference speed and a reduction in parameter count, effectively meeting the requirements for network lightweight design.

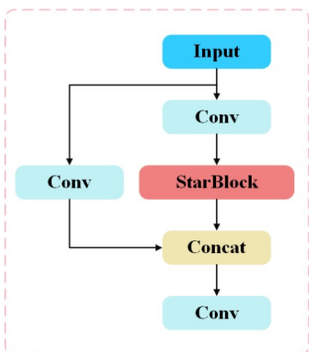

Figure 1 C3K2_light

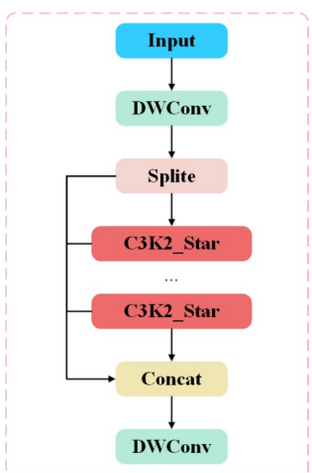

Figure 2 Improved Backbone Component

3 Result and analysis

3.1 Experimental Environment and Dataset

The experimental environment of this study: a 64-bit Ubuntu 20.04 operating system, with an Intel(R) Xeon(R) Gold 6330 CPU, and an NVIDIA GeForce RTX 3090 GPU with 24GB VRAM, along with ample system RAM. The programming language used was Python 3.8. The experiment was conducted using a deep learning framework based on PyTorch 2.0.0 and run in an NVIDIA GPU environment, specifically utilizing CUDA 11.8 and the cuDNN acceleration library to ensure compatibility. The primary dependency was the Ultralytics YOLO library for model training and validation. During training, the batch size was set to 16, the initial learning rate was 0.001, the weight decay coefficient was 0.01, and the momentum was 0.95. These parameters were adjusted using the SGD optimizer, and Automatic Mixed Precision (AMP) was enabled to accelerate training and reduce GPU memory usage. The number of iterations (epochs) for the model was set to 60, while the input image size was specified as 640x640 pixels, and the number of data loading workers was set to 0 to avoid

potential conflicts in parallel processing. The dataset used in this experiment is a swimming and drowning detection dataset, which includes three annotated categories: drowning person, swimming person, and person floating on the water surface. The scenes are set in an indoor swimming pool. The dataset comprises a total of 8572 images, with 7000 images in the training set, and 1572 images each in the validation set and test set.

3.2 Performance Evaluation Indicators

This paper selects mAP (mean Average Precision), Parameters, and FLOPs (computational load) as the key metrics for evaluating the performance of YOLOv11 combined with pruning technology in drowning detection. Among these, mAP is used to measure the detection accuracy of the model, with a higher value indicating better precision. Parameters and FLOPs are employed to assess the model's complexity—fewer parameters and lower computational load typically mean the model is more suitable for deployment in edge device environments with limited computing resources.

3.3 YOLOv11 Feasibility Analysis

By applying the YOLOv11 model to drowning detection, this study preliminarily validates its feasibility in this scenario; it demonstrates high accuracy and recall rates in recognizing three key states (drowning, swimming, and floating). The performance result can be seen in Table I. However, the YOLOv11 model still has considerable room for optimization in terms of detection speed and parameter quantity, especially when facing challenges in computational efficiency and real-time performance for deployment on edge devices. To address this, this study employs lightweight improvement to the original model, reducing redundant parameters and computational load through improving the backbone and neck component, significantly improving inference speed while maintaining accuracy, thereby enhancing the model's adaptability in resource-constrained edge device environments.

Table 1 Drowning detection model performance

Model	P/%	R/%	mAP50/%	mAP50-95/%	Params/10^6	GFLOPs
YOLOv11n	95.3	95.7	97.9	69.8	2.58	6.3
+C3k2_Light	96.9	97.3	98.9	72.9	1.91	4.9
+C3k2_Light+BiFPN	96.3	97.1	98.6	74.0	1.49	4.4

3.4 Comparison of Model Improvements Before and After

Through the lightweight improvements, this study effectively reduced the parameter count and computational complexity of the YOLOv11 model while largely maintaining or even slightly improving its detection accuracy (Table. I). Specifically, by introducing the C3K2 module to reconstruct the backbone network for lightweight purposes, the model achieved a significant reduction in both parameters and computations, alongside a certain improvement in detection accuracy, with the decrease in computational complexity being particularly notable. Furthermore, after integrating the BiFPN structure into the feature fusion network, although the accuracy metrics showed a slight decrease compared to the C3K2 version, they remained superior to the original YOLOv11 model, and the model's parameter size was further compressed. This series of structural optimizations enabled the model to achieve a better balance between efficiency and accuracy.

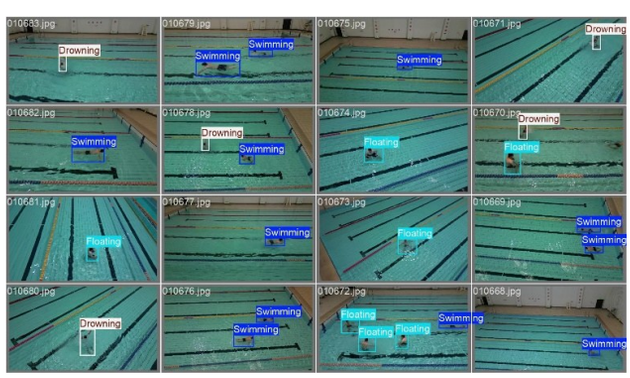

Figure 3 Sample Visualization Labels

Figure 3 and 4 respectively present some test sample visualization labels and detection results of the drowning detection model used in this study under real swimming pool scenarios. The figures intuitively demonstrate the model's recognition effectiveness for various water behaviors such as swimming, floating, and drowning. Each sample image is annotated with the behavior category predicted by the model and its corresponding confidence score. As shown in the figures, the model can effectively distinguish between different poses and maintain high detection accuracy in complex water environments, verifying the practicality of the algorithm proposed in this paper for real-world applications.

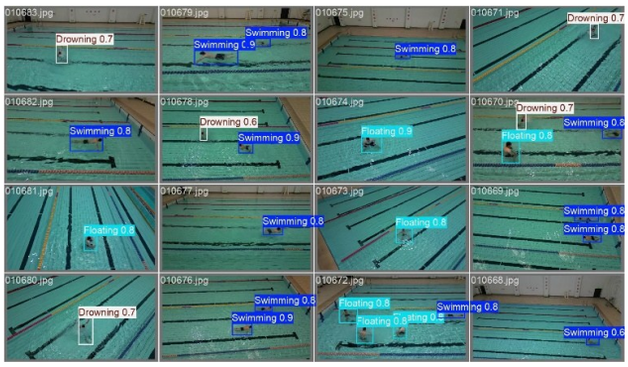

Figure 4 Sample Visualization Tests

4 Conclusion

This study focuses on drowning detection based on YOLOv11 and model lightweighting improvements, aiming to enhance the automation level and real-time performance of water area safety monitoring. The results demonstrate that the optimized YOLOv11 model achieves high accuracy and recall rates in recognizing three critical states: drowning, swimming, and floating. Meanwhile, the model lightweighting improvement significantly reduces both the model parameter count and computational complexity. The compressed model improves inference speed while maintaining accuracy, proving its feasibility for deployment on edge devices. Future work may explore multi-modal sensor fusion, robustness optimization in complex environments.

References

[1] Massey H, Gorczynski P, Harper C M, et al. Perceived impact of outdoor swimming on health: web-based survey[J]. Interactive Journal of Medical Research, 2022, 11(1): e25589.

[2] World Health Organization (WHO). Global Drowning statistics reports[R/OL]. (2024) [2025-09-11]. https://www. who. int/ne ws-room/fact-sheets/detail/drowning.

[3] People's Online Public Opinion Data Center. 2022 China Youth Drowning Prevention Big Data Report[R/OL]. (2022) [2025-09- 11]. https://fcy.618cloud.com.cn/news/info?data=48844efbadb 140acb17aca7c33c3cb85&active=1.

[4] PREVENTION P. Prevention of drowning[J]. Pediatrics, 2019, 143(5).

[5] Jalalifar S, Belford A, Erfani E, et al. Enhancing water safety: exploring recent technological approaches for drowning detec- tion[J]. Sensors, 2024, 24(2): 331.

[6] Eng H L, Toh K A, Yau W Y, et al. DEWS: A live visual surveillance system for early drowning detection at pool[J]. IEEE transactions on circuits and systems for video technology, 2008, 18(2): 196-210.

[7] Zhang Y, Li X, Wang F, et al. A comprehensive review of one-stage networks for object detection[C]//2021 IEEE Inter- national Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2021: 1-6.

[8] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real- time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[9] Vijayakumar A, Vairavasundaram S. Yolo-based object detec- tion models: A review and its applications[J]. Multimedia Tools and Applications, 2024, 83(35): 83535-83574.

[10] Song Q, Yao B, Xue Y, et al. MS-YOLO: A lightweight and high-precision YOLO model for drowning detection[J]. Sensors, 2024, 24(21): 6955.

[11] Jiang X, Tang D, Xu W, et al. Swimming-YOLO: a drowning de- tection method in multi-swimming scenarios based on improved YOLO algorithm[J]. Signal, Image and Video Processing, 2025, 19(1): 161.

[12] Yuan J K, Liu F, Jiang L, et al. GEF-YOLO: an enhanced model for detecting drowning risks at sea[J]. Journal of Real- Time Image Processing, 2025, 22(4): 1-19.

[13] Khanam R, Hussain M. Yolov11: An overview of the key ar- chitectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.

[14] Wang X, Deng H, Gao L, et al. Scale-Aware Pre-Training for Human-Centric Visual Perception: Enabling Lightweight and Generalizable Models [J]. arXiv preprint arXiv:2503.08201, 2025.

[15] Ma X, Dai X, Bai Y, et al. Rewrite the stars[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 5694-5703.

[16] Wang H, Ju X, Zhu H, et al. SEFormer: A Lightweight CNN- Transformer Based on Separable Multiscale Depthwise Convo- lution and Efficient Self-Attention for Rotating Machinery Fault Diagnosis[J]. Computers, Materials & Continua, 2025, 82(1).

[17] Doherty J, Gardiner B, Kerr E, et al. Bifpn-yolo: One-stage object detection integrating bi-directional feature pyramid net- works[J]. Pattern Recognition, 2025, 160: 111209.

[18] Chen J, Mai H S, Luo L, et al. Effective feature fusion network in BIFPN for small object detection[C]//2021 IEEE international conference on image processing (ICIP). IEEE, 2021: 699-703.