Journal of Computers Vol. 36 No. 2, April 2025, pp. 91-104
https://doi.org/10.63367/19911599202504360200 7

Adaptive Spatial Cross-Stage Pooling Network based Yolov7 for Water Level Detection

Chenyang Zhang[1], Yuangao Ai[1], Sui Guo[1], Wentao Zhang[1], Changming Xu[1], and Houming Shen[2*]

[1] China Yangtze Power Co., Ltd, Yichang City, Hubei, China, 44300
{zhang_chenyang, ai_yuangao, guo_sui, zhang_wentao, xu_changming1}@ctg.com.cn

[2] Wuhan NARI Limited Liability Company, State Grid Electric Power Research Institute Co., Ltd., Wuhan, Hubei, China, 430070
shenhouming@sgepri.sgcc.com.cn

Received 14 November 2024; Revised 20 November 2024; Accepted 20 November 2024

Abstract. Real-time water level detection inside the pumps of hydropower stations is significant for timely water resource control and electrical energy production work, which has gained a lot of attention in industry and academia. Traditional water level detection algorithms are unable to detect the water level in real-time due to large errors and high costs, and therefore, artificial intelligence technology is needed to improve the efficiency of water level detection. In this work, based on the latest target detection model YOLO (You Only Look Once), we propose an improved Yolov7 framework based on an adaptive spatial cross-stage pooling network to achieve real-time accurate and robust water level detection. First, to dynamically extract finer targets, an adaptive network detection head is proposed; then, to simplify the feature extraction process and improve the feature extraction speed, an improved SPPFCSPC (Spatial Pyramid Pooling Faster Cross-Stage Partial Channel) module is proposed; finally, to improve the object detection speed, an improved ELAN (Efficient Layer Aggregation Networks) module is proposed. Sufficient experiments have proved that the method proposed in this paper can quickly realize the accurate detection of water levels and provide an effective solution for the efficient management of hydropower stations.

Keywords: water level detection, YOLOv7 model, cross-stage partial channel, layer aggregation network

1 Introduction

The intelligent and secure management of hydropower stations has become a key issue of concern for water conservancy and hydropower projects at all levels [1]. However, due to the scattered distribution and remote location of hydropower stations, there are problems such as information blockage and delayed response in management. Therefore, strengthening the efficient monitoring and management of subordinate hydropower stations at all levels, timely obtaining on-site information, quickly and effectively responding to emergencies, making reasonable use of hydropower generation, arranging safety production tasks for each power station in combination with flood control and irrigation needs, and ensuring national livelihood safety has always been a top priority for government departments at all levels [2-3]. Specifically, intelligent monitoring, analysis, calculation, and control of real-time water levels in the internal collection wells of hydropower stations are important components for ensuring the safe operation of hydropower stations and are the cornerstone for ensuring dam safety and internal safety production of hydropower plants [4-5]. Real-time water level detection for hydropower stations can effectively monitor water level changes, such as the rise and fall of the water level, etc. In addition, it helps the relevant staff of the hydropower station grasp real-time water level information and guide the rational deployment and utilization of water resources, which provides a strong guarantee for real-time control of water resources and the production of electric energy. In recent years, with the rapid development of artificial intelligence technology centered on deep learning, the [6-7] intelligent recognition technology of water level has also been developed significantly [8]. Although these methods combine deep neural networks for water ruler and water level detection, these deep learning models are not specifically designed for target detection and therefore have poor accuracy [9].

* Corresponding Author

The YOLO family of models is a series of frameworks designed specifically for object detection, and since its proposal in 2016 [10], it has now evolved to the YOLOv8 version [11]. Among the various object detection algorithms, the YOLO framework stands out for its balance of speed and accuracy in recognizing objects in images quickly and reliably [12]. Since the YOLOv7 model [13] has the highest accuracy (up to 56.8% on average) among this family of models, the YOLOv7 model was chosen and improved for this project for the task of real-time water level detection inside the pumps of a hydropower plant. First, because the size of the feature map output by the network detection head of YOLOv7 is fixed and cannot be flexibly adjusted according to the size of the input image, the adaptive network detection head is designed in this project, which is capable of adaptively adjusting the size of the output feature map according to the size of the input image to capture a more detailed target; then, because the SPPCFPC module in the YOLOv7 model is more complicated, even though it is used frequently, but the structure is more complicated, this project improves this module, replacing it with the more efficient SPPFCSPC module and changing the activation function so that the whole module is more efficient; finally, because the ELAN module in YOLOv7 has a large computational overhead, this project proposes the BE-ELAN module, improves the convolutional layer and uses a more efficient attention mechanism so that the model can capture more fine-grained targets. More efficient attention mechanisms are used so that the model speed is greatly improved to meet the requirements of real-time water level detection. The experimental results show that the method proposed in this paper is not only able to monitor the water level inside the pump of a hydropower plant in real-time, but also has a higher accuracy than the direct use of YOLOv7, while also remaining robust under weak illumination.

In short, the main contributions of this study are summarized as follows:

(1) This study has proposed the latest object detection framework based on the improved and optimized YOLOv7 framework, which includes an improved SPPFCSPC (Spatial Pyramid Pooling Faster Cross Stage Partial Channel) module and an improved ELAN (Efficient Layer Aggregation Networks) module for Water Level Detection.
(2) The proposed improved SPPFCSPC module is used to dynamically extract finer targets and simplify the feature extraction process. The ELAN module is devoted to improving the object detection speed.
(3) Extensive experiments have been conducted on the mixed public and self-built datasets, which fully prove the promising results of the method in this study.

The organization of this manuscript is as follows. First, Section 2 describes the related work for water level detection. Section 3 gives the specific method for our algorithm. The experimental results and analysis are introduced in Section 4. Finally, concluding remarks are presented in Section 5.

2 Related Work

Existing works based on water level detection have achieved progress performance and have gained widespread attention in the academic community. Ding et al. [14] proposed a water level detection algorithm with a joint contextual attention mechanism, which utilizes the CAM-UNet mode based on the contextual attention mechanism and the least-squares polynomial fitting function to realize the intelligent recognition of water level information under the complex background. Sun et al. [15] proposed a water level detection algorithm based on computer vision, which detects and corrects the water ruler with edge detection and affine transformation and obtains the water surface height based on the keyword processing and edge feature detection results. Li et al. [16] proposed a water level detection algorithm incorporating a transformer and residual channel attention mechanism, mainly dealing with intelligent water level detection in harsh scenes. Zhang et al. [17] trained a deep full convolutional neural network to predict the water ruler image for pixel-by-pixel classification and detected the position of the water level line in the semantically image segmentation based on the idea of deep learning image semantic segmentation. Wu [18] proposed a research method of foam line position detection algorithm based on depth learning in the A/O pool, which aims at the interference of image detection background and other factors and realizes water level detection through BYOLOv5x, a foam line detection algorithm with enhanced feature information. Yin et. al [19] proposed a water level prediction method based on the coupled model of long short-time memory (LSTM) and Isolation forest (IForest), which was used to model the correlated time data with the combination of variables equipped with the IForest algorithm. Jannatul et.al [20] proposed a Particle swarm optimization-based LSTM (long short-term memory) with Sutcliffe efficiency (NSE) coefficient and root mean square error (RMSE) for water level forecasting. To encourage researchers to carry out intelligent water level detection tasks, Li et al. [21] proposed a water level detection dataset called GLH-water, which consists of 250 satellite images and 40.96

billion pixels of labeled surface water annotations, including water level states in various scenarios (such as rivers, lakes, ponds in forests, irrigated fields, bare land, and urban areas), providing important research basis for the development of this task. Yao et al. [22] proposed a hybrid CEEMDAN BiGRU SVR MWOA (CBSM) framework for predicting lake water levels based on multiple environmental, hydrological, and meteorological factors.

Although current researchers have conducted extensive research on water level detection, they mainly focus on completing recognition tasks under simple scenes and ideal lighting conditions. However, in real conditions, water level detection faces more complex scenes (such as rainy and snowy weather, water vapor diffusion, etc.) [5]. Different from the existing methods, our work adopts a framework based on improved Yolov7 and completes water level detection by improving SPPFCSPC and ELAN models for more complex application scenes.

3 Methodology

3.1 YOLOv7 Model

The YOLO series of algorithms, as a typical representative of one-step target detection, has been widely researched and developed in recent years. The YOLOv7 model, published by Wang et al. [13] in 2023, outperforms previous versions of the YOLO model in terms of speed and accuracy and simultaneously achieves a balance between the model's accuracy and inference performance. The innovations of the YOLOv7 model compared to the previous models in the same series include Model structure re-referencing and dynamic label assignment. For model structure re-referencing, YOLOv7 analyzes the propagation path of the gradient to optimize the structure re-referencing for different layers in the network with different planning; in terms of dynamic label assignment, the authors propose a label assignment strategy guided from coarse to fine to solve the problem that it is difficult to dynamically assign targets to different branches during the training because the model has multiple output layers. In addition, the authors propose an extended and composite scaling model, which can control the amount of parameters and computation more efficiently and not only reduce the number of parameters that need to be trained substantially, but also improve the inference speed as well as the detection accuracy effectively.

3.2 Adaptive Network Detection Header

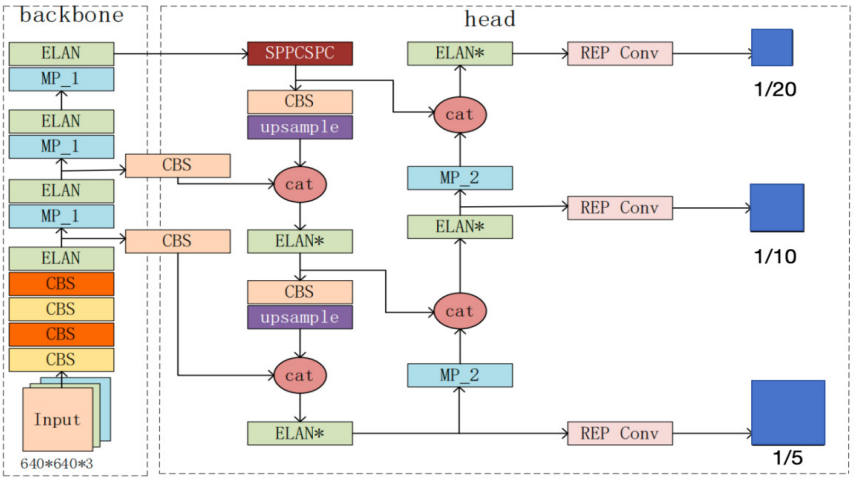

Fig. 1. Adaptive network detection header (blue block)

The YOLOv7 network outputs three feature maps with sizes of 20×20, 40×40, and 80×80, which are used to detect large, medium, and small objects in the input image, respectively. The size of these three feature maps has limited extraction capability for fine targets. According to the characteristics of the hydroelectric power station

water level pictures, this project uses an adaptive network detection head to adaptively obtain three feature maps of different sizes according to the size of the input picture. The size of the three feature maps are 1/20, 1/10, and 1/5 of the size of the input image, respectively. e.g., a 400×400 input image has feature map sizes of 20×20, 40×40 and 80×80, and a 200×200 input image has feature map sizes of 10×10, 20×20 and 40×40, respectively. The method is to use the original backbone network in the ELAN module. The number of CBS modules required is calculated according to the size of the input image, thus realizing the adaptively sized feature map output. The YOLOv7 network applying the improved adaptive network detection head is shown in Fig. 1.

3.3 Improved SPPCSPC Module

The SPPCSPC is a module to extract image features with the YOLOv7 model. It adds serial and parallel multiple maximal pooling operations to a string of convolutions and avoids image distortion due to pre-processing of the image, which solves the problem of duplicated features extracted from the picture by the convolutional neural network. The SPPCSPC module is the same as the one used in the YOLOv7 Spatial Pyramid Pooling (SPP) structure, but the number of parameters and computations are improved compared to the original SPP structure, which is shown in Fig. 2.

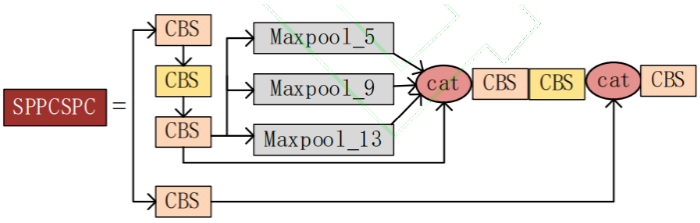

Fig. 2. Original SPPCSPC module

This work replaces the SPP module in the original YOLOv7 with an improved SPPFCSPC module, which improves the feature extraction speed while keeping the same sensory field as the original SPPCSPC. The structure of the module is shown in Fig. 3.

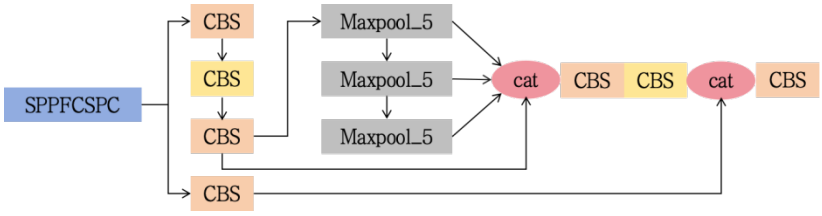

Fig. 3. Improved SPPCSPC module

As can be seen in Fig. 3, we unified the different sizes of pooling windows (5, 9, 13, respectively) in the original SPPCSPC module to a pooling window of size 5 and then performed serial pooling operations on the input features passing through the three CBS modules, respectively. Finally, we concatenated the output features of the three pooling layers with the features passing through the three CBS modules.

In addition, we change the activation function of the CBS layer in the SPPFCSPC module from ReLU to FReLU, whose expression of FReLU is shown in Eq (1).

$$f(x)=\begin{Bmatrix}\alpha_1 x+\beta_1,\ if\ x\geq 0\ and\ x<t_1\\ \alpha_2 x+\beta_2\ if\ x\geq t_1\ and\ x<t_2\\ \cdots\\ \alpha_n x+\beta_n,\ if\ x\geq t_{n-1}\end{Bmatrix}, \tag{1}$$

where f(x) is the spatial context feature extractor, and α, β, t are the adjustable parameters. The original CBS layer, i.e. Conv+BN+SiLU, after changing the ReLU function [7] to FReLU, which changes the input into a one-dimensional vector, then applies the RelU activation function to each element of the vector. Finally, it recovers the output features into the shape as they were at the time of the original input. FReLU not only does not require additional parameters and can reduce the number of parameters of the model, but also improves the representation ability of the model without adding parameters, reducing the occurrence of overfitting.

3.4 Improved ELAN Module

The improved ELAN module used to be in this project with the original module in the backbone network, which is used for image feature extraction and controlling the number of channels. Due to the low efficiency of the ELAN module in YOLOv7, which is improved to increase the detection speed. The main method is to introduce Blueprint Separation Convolution (BSConv) and Efficient Channel Attention (ECA) mechanisms. The original ELAN module can be called BE-ELAN, and the structure of the ELAN module is shown in Fig. 4.

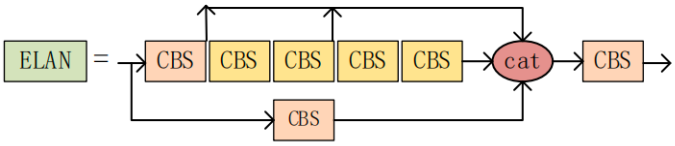

Fig.4. Original ELAN module

In this work, we have improved and optimized the ELAN model, whose specific structure of improved ELAN is shown in Fig 5. First, this work replaces the original 3×3 convolution in the ELAN module with the Blueprint Separable Convolution (BSConv). The original 3×3 convolution makes the whole model slower due to a higher computational overhead. BSConv makes an upgraded version of depth separable convolution (DSConv), which better utilizes kernel internal correlation for efficient separation. BSConv is a decomposition of a standard convolution into a 1×1 convolution and a channel-by-channel convolution. Each filter of the CNN is split from $M\times K\times K$ into $MK\times K$ filters; principal component analysis is performed on each of the split M filters, and the variance of the filter is determined from the first principal component. Such a structure utilizes the strong correlation between the CNN depth-direction convolution kernels and is therefore computationally efficient. This lightweight convolutional model is the current trend in developing depth models. In the original standard convolution, the input features are $U\in R^{M\times Y\times X}$, the output features after convolution are $V\in R^{M\times Y\times X}$, and the corresponding formula is:

$$V_{n=}U * F^{(n)}, n\in\{1,\ldots,N\}, \tag{2}$$

Among them, $F^{(n)}$ is the value of the $M\times K\times K$ the convolution kernel. Based on the principle of BSConv, each convolution kernel $F^{(n)}$ can be represented as s blueprint $B^{(n)}$ and weights ω:

$$F_m^{(n)}=\omega_{n,m}\cdot B^{(n)}, m\in\{1,\ldots,m\}, n\in\{1,\ldots n\}, \tag{3}$$

The above BSConv convolution compares the standard convolution's $M \times N \times K^2$ parameters, which only have $N \cdot K^2 + M \cdot N$ parameters to be trained.

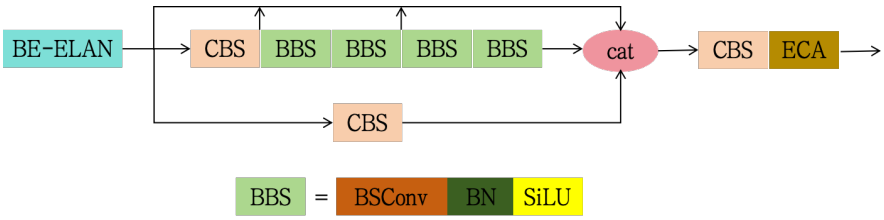

Fig. 5. Improved BE-ELAN module

The attention mechanism is usually used to deal with information screening and integration in image data so that the model can focus on the regions that need to be modified, which plays an important role in the improvement of detection accuracy. In this work, we add an efficient channel attention mechanism ECA [23] to the ELAN module. The ECA attention mechanism efficiently realizes local cross-channel information interaction with one-dimensional convolution to extract the dependencies between image feature channels, avoiding the problem of not being able to learn the dependencies between channels caused by channel compression in the SE attention mechanism. This attention mechanism first pools the input features globally on average then performs a one-dimensional convolution operation with a convolution kernel size of k. Like [24], we compute the weights of each channel using the Sigmoid activation function with the following formula (4):

$$\omega = \sigma\left(C1D_k\left(y\right)\right), \tag{4}$$

where ω is the computed weight of each channel, and σ denotes the Sigmoid activation function [17]. C1D denotes the one-dimensional convolution operation, and y is the input feature. Finally, the weights are multiplied by the corresponding elements in the original feature map to obtain the output feature map.

3.5 Optimization and Training

For the specific water level detection process in this paper, the first step is to detect the water level boundary line and then carry out the corresponding numerical processing. This paper mainly uses the classification loss function to complete the corresponding model training, and its specific process is expressed as follows:

$$L_{cls} = -g \log p - (1-g)\log(1-p), \tag{5}$$

where g is the label corresponding to the input sample (such as 1 for this sample and 0 for negative samples). In this method, g then corresponds to the water level line coming from the actual detection marker, and p is the probability that the model predicts that this input sample is positive (correctly detected water level). In this process, the corresponding water level value is detected by the Yolov7 model, and the actual water level is predicted after focusing on the coordinates.

In addition, in practical scenarios, the water level line has the characteristic of being infinitely large in its extension direction and infinitely small in other directions. This article will also improve the center coordinate sampling of the water level anchor box predicted by Yolov7 and perform a least square fitting to obtain the expression of the water level line as follows:

$$k = av + b, \tag{6}$$

$$b = \bar{k} - \bar{a}v, \tag{7}$$

$$a = \frac{n\sum_{i=1}^{n} v_i k_i - \sum_{i=1}^{n} v_i \sum_{i=1}^{n} k_i}{n\sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} v_i)^2}, \tag{8}$$

where v_i and k_i are the coordinates of the midpoint in the prediction box; a is the slope of the fitted line; b is the intercept of the fitted line; N is the number of predicted boxes; $\bar{v}$ and $\bar{k}$ are the mean coordinates of the points in the predicted box. The specific water level prediction loss function is expressed as follows:

$$L_{wat} = \sum_{i}^{M} \sqrt{(K-k)}, \tag{9}$$

where K represents the actual water level value, k represents the water level value predicted and fitted by the least squares method, M represents the number of samples, and i is a single sample point.

In all, the whole loss function L of our method is represented as follows:

$$L = \lambda L_{cls} + (1-\lambda) L_{wat}. \tag{10}$$

where $\lambda \in (0,1)$ is the control parameter to prevent overfitting of the model due to biased classification or direct water level recognition during the training process, to achieve the training balance of the entire framework while ensuring the lightweight and efficient role of the model.

4 Experimental Results and Analysis

This section mainly describes the experimental results and relevant analysis. Firstly, the dataset, implementation details, and evaluation metrics are introduced. Then, we describe the comparison with the latest methods. At last, the ablation study is described.

4.1 Datasets

To effectively validate our proposed method, this paper adopts the algorithm validation based on the publicly available dataset Water Level Computer Vision (WLCV) [25] and the dataset collected in the field. Given that the publicly available dataset WLCV has only 256 water level image data, which is insufficient to complete the effective training of the algorithm model, we collected 15,000 pieces of real-time digital data at different times and angles in the field, which are mainly categorized into four scenarios, namely, Daytime, Nighttime, Rainy day, and Vapor. Specifically, there are 7000 samples for daytime scenes, 5000 samples for nighttime scenes, 1500 samples for rainy days, and 1500 samples for indoor scenes with water vapor. We combine the data collected in the field and the publicly available online data collection to form a diversified whole. To ensure the diversity and completeness of the training data, and to cover as many data samples as possible in all kinds of scenes, the field collection data and the online public data collection are merged to form an overall diversified data set. Pre-processing methods such as flipping and segmentation are used to expand the overall data to 20,000 images, of which 5,000 are images of the above four scenes. 15,000 are taken as training and 5,000 for testing. The specific dataset samples are shown in Fig. 6, where Fig. 6(a) represents the daytime scenes, Fig. 6(b) is the Nighttime scenes, Fig. 6(c) denotes the Rainy day scenes, and Fig. 6(d) represents the Vapor ones.

Fig. 6. Instance samples in datasets (public data and self-collected dataset collections)

4.2 Implementation Environment and Evaluation Metrics

Implementation Environments: The experimental environment of this paper is mainly based on the Pytorch platform and the operating system is Ubuntu20.04. The hardware environment is an Intel I7 9700 processor and the GPU is Nvidia GTX2080Ti. In addition, the software uses Python 3.8.12.

Parameters Setting: In this work, the parameters are set as follows: the Batch Size is set to 10, the epoch of each time is 100, the momentum is set to 0.85, the initial learning rate is 0.005, the learning rate period is 0.2, the weight decay factor is 0.0005, and λ is set as 0.55.

Evaluation Metrics: Generally speaking, the target detection algorithm is mainly based on the mean Average Precision (mAP) and Accuracy (Acc) [26], as well as other metrics to complete the performance evaluation of the realized algorithm. The water level detection also belongs to one kind of target detection, but in the actual situation, because the extension of the water level line in the horizontal plane presents an infinite expansion of the status quo, in addition to the above usual evaluation indexes, this paper also adopts another kind of water level height computation method to realize real-time water level detection. By taking the original real water level scale as a criterion, the algorithm predicts the water level height and the relative accuracy of the real water level height S_a, as the evaluation index.

$$S_a = \left(1 - \frac{|h_r - h|}{h_r}\right) \times 100\% \tag{11}$$

where h is the algorithm's predicted water level height and h_r is the true water level height. The true water level elevation value is the manually extracted labeled value.

4.3 Experimental Results and Analysis

This section focuses on comparing the performance of the proposed algorithm of this method with the performance of the existing techniques and the ablation experiments in two parts.

Comparison with State-of-the-Art Methods. Firstly, we compare the proposed method in this paper with the existing algorithms based on water level detection and complete the corresponding performance validation from four scenarios, namely, daytime, nighttime, rainy day, and foggy, whose specific results are shown in Table 1. We mainly compare them with the latest models such as ResNet-50 [27], PVT-S [28], TwinsP-S [29], GoogleNet

[30], Transformer [31], Yolov5 [32], Yolov7 [13], Trans-LSTM [8], and JCAM [14]. From the results, we can see that the method in this paper achieves the optimal results both in terms of average accuracy and final accuracy.

Table 1. Comparison of algorithmic results with the latest benchmark model

Methods	Year	Daytime		Nighttime		Rainy day		Vapor	
		mAP	Acc	mAP	Acc	mAP	Acc	mAP	Acc
ResNet-50 [27]	2021	0.85	0.90	0.80	0.87	0.83	0.85	0.78	0.83
PVT-S [28]	2021	0.75	0.82	0.70	0.77	0.73	0.75	0.70	0.73
TwinsP-S [29]	2021	0.77	0.81	0.72	0.77	0.70	0.75	0.72	0.76
GoogleNet [30]	2016	0.80	0.87	0.74	0.76	0.72	0.78	0.71	0.71
Transformer [31]	2018	0.90	0.95	0.85	0.90	0.90	0.88	0.89	0.93
Yolov5 [32]	2021	0.88	0.91	0.83	0.84	0.85	0.87	0.88	0.90
Yolov7 [13]	2023	0.90	0.92	0.89	0.89	0.85	0.86	0.88	**0.93**
Trans-LSTM [8]	2024	0.91	0.93	**0.91**	0.90	0.87	**0.90**	0.88	0.91
JCAM [14]	2024	0.89	0.92	0.88	0.89	0.88	0.88	0.89	0.90
Ours	2024	**0.92**	**0.96**	0.90	**0.92**	**0.91**	0.89	**0.91**	**0.93**

As shown in the above table, the text has been recognized by the corresponding algorithms with different backbone networks as the base network of the segmentation model, and the performance comparison has been carried out from the three dimensions of mean Accuracy (Mean AP, mAP) and Accuracy (Acc), respectively, from which it can be seen that the optimal performance of the improved Yolov7 based proposed in this paper has been obtained.

Ablation Study. In addition, to verify the validity of our approach as a whole, in this section, relevant ablation experiments are conducted, specifically from the four sub-modules that make up the framework proposed in this paper. Based on the water level in the above four scenarios of day, night, rain and fog, the algorithm predicts the water level height from h, the true water level height h, and the relative accuracy of the predicted water level height to the real water level height S. The three dimensions are used to evaluate the proposed method in this paper separately, and the specific results are shown in Table 2 ("SPPSSPC+" denotes the improved SPPSSPC, and "ELAN+" denotes the improved ELAN, "**Ses.**" denotes different scenes). As can be seen from the table, each sub-module composed of the framework proposed in this paper achieves good performance in the actual algorithmic performance identification. This is especially the case of certain real height. The best performance is finally achieved by the joint learning model in this paper.

Table 2. Performance analysis of the sub-modules of the proposed framework (cm, S_a(%))

Ses. / Sub-model	Daytime			Nighttime			Rainy day			Vapor		
	h	h_r	S_a	h	h_r	S_a	h	h_r	S_a	h	h_r	S_a
Adaptive-head	32.8	33.2	98.8	29.7	33.2	89.5	30.7	33.2	92.3	31.6	33.2	95.2
SPPSSPC+	32.8	33.2	98.8	30.2	33.2	91.0	31.3	33.2	96.7	30.8	33.2	92.8
ELAN+	31.9	33.2	96.1	31.2	33.2	93.4	32.1	33.2	96.7	30.9	33.2	93.1
Ours	**33.0**	33.2	**99.4**	**32.7**	33.2	**98.6**	**32.9**	33.2	**99.1**	**31.8**	33.2	**95.8**

To further validate the effectiveness of the method proposed in this paper, in addition to the above quantitative evaluation indexes, this paper also carries out the intermediate process of validation of the YOLOv7 model, in which the results of the water level recognition detected in some daytime and nighttime scenarios are visualized and displayed, which are shown in Fig. 7.

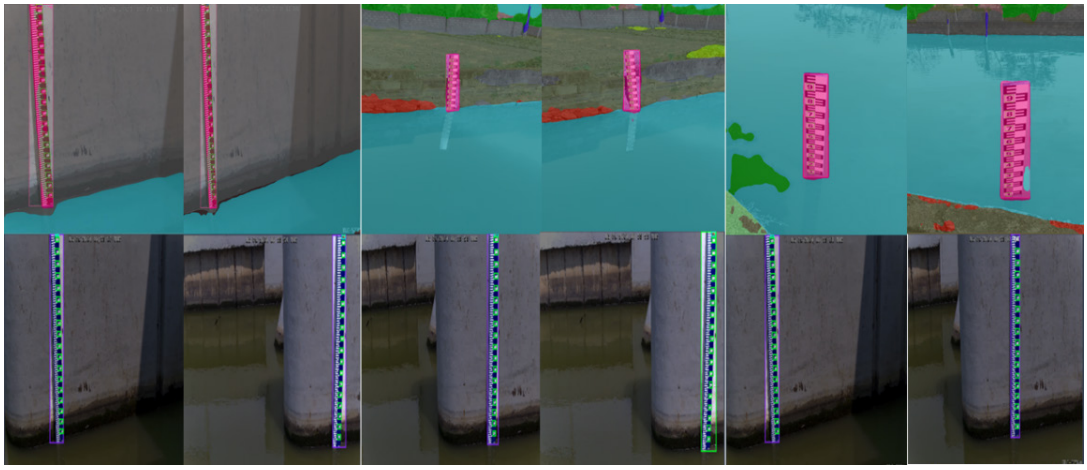

Fig. 7. Visualization of water level recognition results

As can be seen from the figure, some of the original water level monitoring images, after marking recognition, can get some relatively "clean" recognition results, especially for the scale of the water level demarcation line, which also verifies the effectiveness of the algorithm proposed in this paper.

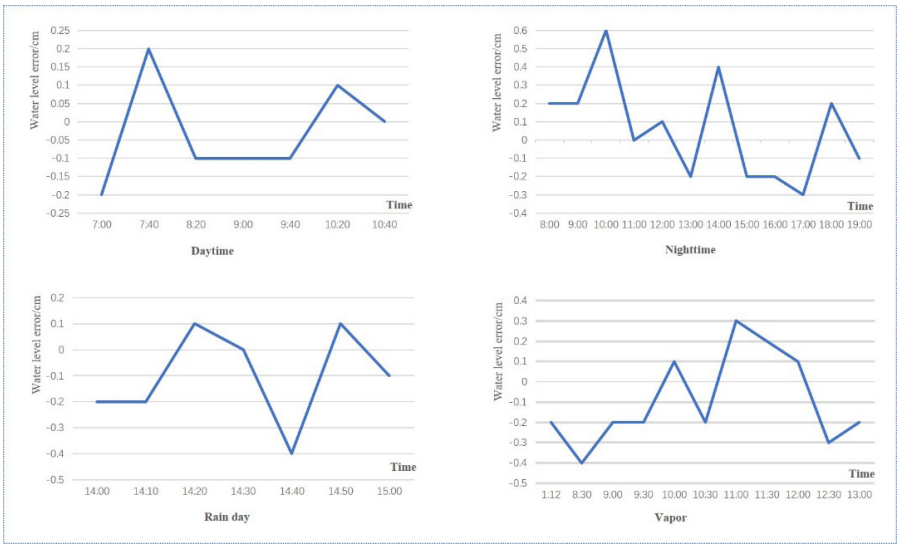

Fig. 8. Comparison between the algorithm of this paper and the actual manual reading error in our scenarios

At the same time to verify the applicability of this paper's method, in four different scenarios, with time, this paper's algorithm real-time recognition of the water level value and the error of manual reading towards the trend graph. As can be seen from the figure, in the daytime scene, the error of manual reading is the smallest error relative to the manual, especially at noon with ideal light conditions. The algorithm proposed in this paper and the manual reading data does not have much difference. In addition, in other scenarios, although there is some error, overall and the actual water level is still the same and the overall difference is not large, thus proving the robustness and practicality of the method in this paper. Finally, to validate the training process of the method in this paper, we give the graphs of the loss function change curves of the model on the training and validation sets, as shown in Fig. 9, where the left plot is the loss change curve on the training set, and the right one is the change result on the validation set. As can be seen from the figure, the model proposed in this paper still conforms to the process of loss function change in general.

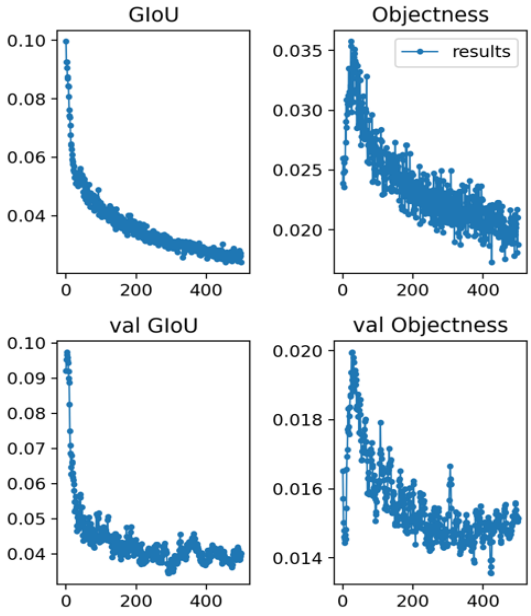

Fig. 9. Plot of the variation of the loss function

Meanwhile, we compared the average accuracy of the method proposed in this article with the real manual readings, as shown in Fig. 10. The red line segment represents the water level value predicted by the algorithm in this article, and the blue line segment represents the actual water level value. As shown in the figure, over time, the overall performance predicted by the method proposed in this paper is equivalent to the real water level values, which once again verifies the practicality of the algorithm proposed in this work. In Fig. 10, the horizontal axis represents time and the vertical axis represents the specific indication of water level. It can be seen that over time, the predicted values and actual water level values alternate in magnitude.

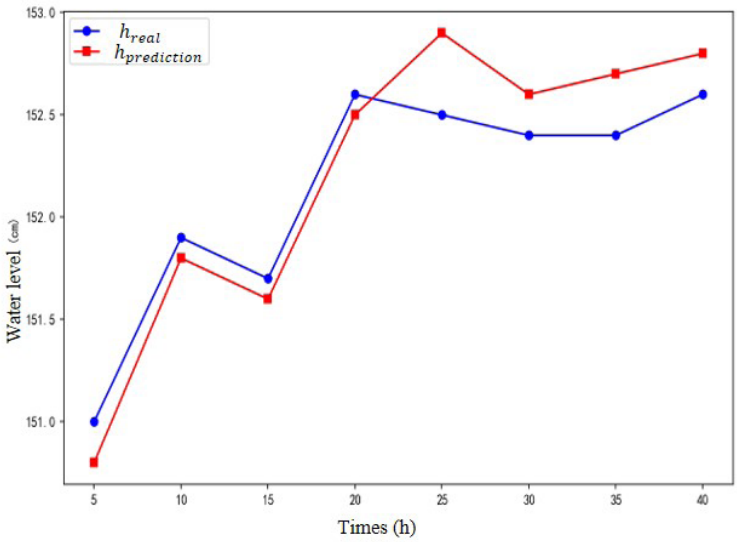

Fig. 10. Overall comparison of water level detection

In addition, to verify the timeliness of our algorithm model, we also conducted corresponding experiments to demonstrate its complexity ("Comp." denotes "Complexity") and real-time performance. The paper mainly compares the model's parameters, the time required to identify a single sample, the training time of the model, and the number of floating-point operations per second (Flops) with the latest existing methods. The specific results are shown in Table 3. According to Table 3, although our model has a slightly larger number of parameters compared to the GoogleNet model, we achieved optimal performance in all other dimensions. It is of great significance for the overall real-time algorithm deployment.

Table 3. Comparison of the latest models for complexity

Comp. / Models	Parameters (M)	Recognition time for a single sample (s)	Flops (G)	Training time (h)
ResNet-50	23.5	5.0×10^{-4}	4.56	6.3
PVT-S	32.1	4.7×10^{-4}	5.23	11.2
TwinsP-S	38.5	5.7×10^{-4}	4.75	17.2
GoogleNet	**5.00**	3.7×10^{-4}	5.67	8.34
Transformer	77.00	2.9×10^{-4}	8.92	20.34
Yolov5	27.00	4.5×10^{-4}	7.84	15.34
Yolov7	40.00	3.5×10^{-4}	8.47	17.39
Ours	25.32	$\mathbf{2.7\times10^{-4}}$	**9.54**	**5.9**

Table 4. Impact of λ for the performance

λ	Daytime		Nighttime		Rainy day		Vapor	
	mAP	Acc	mAP	Acc	mAP	Acc	mAP	Acc
0.15	0.65	0.70	0.60	0.67	0.53	0.65	0.58	0.53
0.25	0.68	0.72	0.62	0.67	0.63	0.68	0.60	0.60
0.35	0.75	0.77	0.72	0.70	0.70	0.72	0.66	0.70
0.45	0.85	0.90	0.84	0.86	0.82	0.88	0.84	0.87
0.55	**0.92**	**0.96**	**0.90**	**0.92**	**0.91**	**0.89**	**0.91**	**0.93**
0.65	0.86	0.92	0.87	0.84	0.87	0.85	0.87	0.90
0.75	0.80	0.82	0.79	0.79	0.80	0.82	0.78	0.89
0.85	0.66	0.70	0.60	0.65	0.61	0.65	0.62	0.63
0.95	0.60	0.67	0.60	0.62	0.59	0.65	0.55	0.57

To verify the impact of control parameters on the performance of our algorithm, we conducted a validation analysis on the value of λ in formula (10), specifically analyzing the four scenarios (Daytime, Night, Rainy day, and Vapor) mentioned in this article. The specific results obtained are shown in Table 4. From the table, it can be seen that when λ takes the middle value between 0 and 1, the performance is relatively good. When it takes the values at both ends, the overall performance decreases. The function of λ is to balance the entire algorithm model, ensuring that the loss of recognition and classification is balanced, so as not to lean towards one side and cause model imbalance and overfitting, which will affect the final algorithm performance.

5 Conclusion and Future Work

In this work, we propose an improved Yolov7 framework based on an adaptive spatial cross-stage pooling network to achieve real-time accurate and robust water level detection. A finer target detection is realized by proposing an adaptive network detection head, and then the feature extraction speed is improved based on the improved SPPFCSPC and ELAN modules. The experimental results show that the method proposed in this paper can effectively complete the water level detection in four different scenarios, and all of them have achieved good

recognition accuracy. The algorithm implementation has a certain degree of practicability and operability, which can meet the practical application requirements.

In the future, we will use the latest models in the Yolov series (such as Yolov8 and Yolov9) for water level detection research in multiple scenarios and construct corresponding water level datasets to provide technical support for the development and application of this field.

6 Acknowledgement

This work is supported by China Yangtze Power Co., Ltd. (No. Z212302030).

References

[1] S.C. Olisa, C.N. Asiegbu, J.E. Olisa, B.O. Ekengwu, A.A. Shittu, M.C. Eze, Smart two-tank water quality and level detection system via IoT, Heliyon 7(8)(2021) e07651. https://doi.org/10.1016/j.heliyon.2021.e07651

[2] J. Liang, Y. Yi, X. Li, Y. Yuan, S. Yang, X. Li, Detecting changes in water level caused by climate, land cover and dam construction in interconnected river-lake systems, Science of the Total Environment 788(2021) 147692. https://doi.org/10.1016/j.scitotenv.2021.147692

[3] A.T. Assaf, K.N. Sayl, A. Adham, Surface water detection method for water resources management, Journal of Physics: Conference Series 1973(1)(2021) 012149. https://doi.org/10.1088/1742-6596/1973/1/012149

[4] L. Yang, J. Driscol, S. Sarigai, Q. Wu, C.D. Lippitt, M. Morgan, Towards synoptic water monitoring systems: a review of AI methods for automating water body detection and water quality monitoring using remote sensing, Sensors 22(6)(2022) 2416. https://doi.org/10.3390/s22062416

[5] F. Chen, X. Chen, T. Van de Voorde, D. Roberts, H. Jiang, W. Xu, Open water detection in urban environments using high spatial resolution remote sensing imagery, Remote Sensing of Environment 242(2020) 111706. https://doi.org/10.1016/j.rse.2020.111706

[6] J. Khan, E. Lee, A.S. Balobaid, K. Kim, A comprehensive review of conventional, machine learning, and deep learning models for groundwater level (GWL) forecasting, Applied Sciences 13(4)(2023) 2743. https://doi.org/10.3390/app13042743

[7] G. Li, Z. Liu, J. Zhang, H. Han, Z. Shu, Bayesian model averaging by combining deep learning models to improve lake water level prediction, Science of The Total Environment 906(2024) 167718. https://doi.org/10.1016/j.scitotenv.2023.167718

[8] N.A.A.B.S. Bahari, A.N. Ahmed, K.L. Chong, V. Lai, Y.F. Huang, C.H. Koo, J.L. Ng, A. El-Shafie, Predicting sea level rise using artificial intelligence: a review, Archives of Computational Methods in Engineering 30(7)(2023) 4045-4062. https://doi.org/10.1007/s11831-023-09934-9

[9] P.Y. Kow, J.Y. Liou, M.T. Yang, M.H. Lee, L.C. Chang, F.J. Chang, Advancing climate-resilient flood mitigation: Utilizing transformer-LSTM for water level forecasting at pumping stations, Science of The Total Environment 927(2024) 172246. https://doi.org/10.1016/j.scitotenv.2024.172246

[10] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: A unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. https://doi.org/10.1109/CVPR.2016.91

[11] A. Kumar, G.S. Lehal, Layout Detection of Punjabi Newspapers using the YOLOv8 Model. International Journal of Performability Engineering 20(3)(2024) 186-193. https://doi.org/10.23940/ijpe.24.03.p7.186193

[12] P. Singh, R. Krishnamurthi, AgriGuard: IoT-Powered Real-Time Object Detection and Alert System for Intelligent Surveillance, International Journal of Performability Engineering 20(4)(2024) 232-241. https://doi.org/10.23940/ijpe.24.04.p5.232241

[13] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023. https://doi.org/10.1109/CVPR52729.2023.00721

[14] X.Ding, Y. Geng, Water level detection algorithm featured by a context attention mechanism, Beijing Water (2)(2024) 66-72. https://doi.org/10.19671/j.1673-4637.2024.02.012

[15] W. Sun, D. Wang, S. Xu, J. Wang, Z. Ma, Water level detection algorithm based on computer vision, Journal of Applied Sciences 40(3)(2022) 434-447. https://doi.org/10.3969/j.issn.0255-8297.2022.03.007

[16] X. Li, C. Sun, Y. Wei, Y. Yuan, Z. Wu, Y. Li, Water level intelligent detection method based on fuse Transformer residual channel attention mechanism in harsh environments, Journal of Electronic Measurement and Instrumentation 37(1)(2023) 59-69. https://doi.org/10.13382/j.jemi.B2205896.

[17] W.J. Zhang, Z. Zhang, J. Huang, Y. Zhou, Y. Jiang, Intelligent water-level monitoring method based on image semantic segmentation, Journal of Hehai University (Natural Scientific Edition) 51(5)(2023) 24-30.

https://doi.org/10.3876/j.issn.1000-1980.2023.05.004
[18] Y. Wu, A Foam Line Position Detection Algorithm Research for A/O Pool Based on Deep Learning, [dissertation] Taiyuan: Tai-yuan University of Science and Technology, 2024. https://doi.org/10.27721/d.cnki.gyzjc.2024.000392
[19] H. Yin, Q. Wu, S. Yin, S. Dong, Z. Dai, M.R. Soltanian, Predicting mine water inrush accidents based on water level anomalies of bore-hole groups using long short-term memory and isolation forest, Journal of Hydrology 616(2023) 128813. https://doi.org/10.1016/j.jhydrol.2022.128813
[20] J.F. Ruma, M.S.G. Adnan, A. Dewan, R.M. Rahman, Particle swarm optimization based LSTM networks for water level forecasting: A case study on Bangladesh river network, Results in Engineering 17(2023) 100951. https://doi.org/10.1016/j.rineng.2023.100951
[21] Y. Li, B. Dang, W. Li, Y. Zhang, Glh-water: A large-scale dataset for global surface water detection in large-size very-high-resolution satellite imagery, in: Proceedings of the AAAI Conference on Artificial Intelligence, 38(20)(2024) 22213-22221. https://doi.org/10.1609/aaai.v38i20.30226
[22] Z. Yao, Z. Wang, T. Wu, W. Lu, A hybrid data-driven deep learning prediction framework for lake water level based on fusion of meteorological and hydrological multi-source data, Natural Resources Research 33(1)(2024) 163-190. https://doi.org/10.1007/s11053-023-10284-3
[23] D. Yarotsky, Error bounds for approximations with deep ReLU networks, Neural networks 94(2017) 103-114. https://doi.org/10.1016/j.neunet.2017.07.002
[24] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. https://doi.org/10.1109/CVPR42600.2020.01155
[25] J.M.G.P. Isidoro, R. Martins, R.F. Carvalho, J.L.M.P. de Lima, A high-frequency low-cost technique for measuring small-scale water level fluctuations using computer vision, Measurement 180(2021) 109477. https://doi.org/10.1016/j.measurement.2021.109477
[26] I. Varsadan, A. Birk, M. Pfingsthorn, Determining map quality through an image similarity metric, in: Proc. RobotCup 2008: Robot Soccer World Cup XI 12, 2009. https://doi.org/10.1007/978-3-642-02921-9_31
[27] B. Koonce, ResNet 50, in: Convolutional neural networks with swift for TensorFlow: image recognition and dataset categorization, Apress, Berkeley, CA, 2021 (63-72). https://doi.org/10.1007/978-1-4842-6168-2_6
[28] W. Wang, E. Xie, X. Li, D.P. Fan, K. Song, D. Liang, Pyramid vision transformer: a versatile backbone for dense prediction without solutions, in: Proceedings of the IEEE/ CVF international conference on computer vision, 2021. https://doi.org/10.1109/ICCV48922.2021.00061
[29] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, C. Shen, Twins: Revisiting the design of spatial attention in vision transformers, in: Proc. Advances in neural information processing systems, 2021.
[30] P. Ballester, R. Araujo, On the performance of GoogLeNet and AlexNet applied to sketches, in: Proceedings of the AAAI conference on artificial intelligence 30(1)(2016) 1124-1128. https://doi.org/10.1609/aaai.v30i1.10171
[31] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, D. Tran, Image transformer, International conference on machine learning. PMLR 80(2018) 4055-4064.
[32] Y. Liu, B.H. Lu, J. Peng, Z. Zhang, Research on the use of YOLOv5 object detection algorithm in mask wearing recognition, World Science Research Journal 6(11)(2020) 276-284. https://doi.org/10.6911/WSRJ.202011_6(11).0038