Received: 23 October 2022 | Revised: 10 January 2023 | Accepted: 5 February 2023

IET Image Processing

DOI: 10.1049/ipr2.12765

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

A video drowning detection device based on underwater computer vision

Tingzhuang Liu | **Xinyu He** | **Linglu He** | **Fei Yuan**

Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Xiamen, China

Correspondence
Fei Yuan, Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Xiamen 361005, China.
Email: yuanfei@xmu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 62071401; Xiamen Ocean and fishery Development Special Fund project, Grant/Award Number: 21CZB015HJ10

Abstract

Drowning is a significant public health concern. A video drowning detection algorithm is a helpful tool for finding drowning victims. However, there are three challenges that drowning detection research typically encounters: a lack of actual drowning video data, subtle early drowning traits, and a lack of real time. In this paper, the authors propose an underwater computer vision based drowning detection device composed of embedded AI devices, camera, and waterproof case to solve the above problems. The detection device utilizes the high-performance computing of Jetson Nano to realize real-time detection of drowning events through the proposed drowning detection algorithm on the acquired underwater video stream. The proposed drowning detection algorithm primarily consists of two stages: in the first step, to successfully solve the interference of the surroundings and to give a trustworthy basis for video drowning detection, the YOLOv5n network is used to detect the near-vertical human body based on the characteristics of the drowning person. In the second stage, the authors propose a lightweight drowning detection network (DDN) based on a deep Gaussian model for fast feature vector detection. The lightweight DDN is combined with the Gaussian model to detect anomaly in the high-level semantic features, which has higher robustness and solves the lack of drowning videos. The experimental results show that the proposed drowning detection algorithm has good comprehensive performance and practical application value.

1 | INTRODUCTION

Drowning is a significant global public health problem [1]. According to the report of World Health Organization, about 372,000 people die from unintentional drowning worldwide each year [2, 3]. Thus, to reduce drowning accidents, most swimming pools are equipped with professional lifeguards, but lifeguards are unable to stay focused for a long time [4] and cannot detect drowning persons in time, leading to death due to a lack of timely assistance. Therefore, an effective and fast drowning detection device will play an important role in pool management and drowning rescue.

In the past, many researchers have provided various methods to detect drowning events in swimming pools, but there are still many challenges in drowning detection. The existing drowning detection methods fall into two main categories. The first is a wearable sensor-based method and the second is a vision-based method. The wearable sensor-based methods use different sensors attached to the swimmer's body to detect drowning by the swimmers' physiological indicators or time in the water [5–11]. People sometimes feel uncomfortable wearing such sensors. The vision-based method avoids the limitations of wearable devices and mainly extracts features from the swimmer's video to accomplish drowning detection. Many vision-based studies also detect drowning based on the swimmer's position, time, and speed in the water [12–21], but the drowning person is quiet in the early stage of drowning. So, these methods lack robustness and are not accurate.

Recently, it has become a trend to use deep neural networks to solve industry problems [22–26]. Some researchers have applied neural networks to drowning detection tasks. However, due to the lack of real drowning sample data, most of these researches first obtain drowning sample data by simulating drowning. Then, the features of drowning data are extracted by learning

methods for performing supervised classification. But, drowning behaviour is difficult to simulate truly, and the features of stimulative drowning behaviour lack realism. Drowning is an anomalous event. The vision-based drowning detection is similar to the video anomaly detection (VAD) task. Both complete the task of detecting abnormal events in the lack of anomalous event videos. Most researchers have used unsupervised deep learning techniques in VAD tasks [27–32], which also motivates us to use unsupervised deep learning techniques in drowning detection tasks. Most of the existing unsupervised deep learning-based VAD first reconstructs or predicts frames and then determines anomalies based on the pixel-level reconstruction error or prediction error. This method is more susceptible to environmental noise interference and is unsuitable for pool videos. So, in this paper, we proposed a deep convolutional neural network combined with the Gaussian model. This neural network can distinguish normal frames from drowning frames in high-level semantic features and has higher robustness.

There are four main contributions of this paper:

1. A drowning detection device based on underwater computer vision is designed, which consists of embedded AI device, camera, waterproof case, and other components. The proposed algorithm is deployed on a high-performance embedded AI device to achieve real-time detection of drowning events.
2. A drowning detection algorithm based on a deep Gaussian model is proposed, which detects drowning in high-level semantic features and has higher robustness.
3. The strategies for underwater near-vertical human detection are proposed to screen out most non-drowning and incomplete humans, providing a reliable basis for drowning detection.
4. A lightweight drowning detection network (DDN) is proposed, which guarantees that the feature of human spatial-temporal cube (STC) can be extracted quickly.

2 | RELATED WORK

Most researchers have explored contact sensor-based and vision-based drowning detection methods. In this paper, we focus on vision-based drowning detection methods.

2.1 | Traditional vision-based method

The 'Poseidon' Drowning Alarm System [12] is the most representative early drowning detection system. The swimmers' activity location and time are monitored by cameras installed on the wall of pools. Kam et al. [33] provide new insights into robust human tracking and semantic event detection for drowning detection. End et al. [34] propose an efficient segmentation algorithm based on threshold hysteresis which can realize fast and reliable detection of swimmers, and then define a set of swimmers and drowning descriptors according to professional knowledge. Fei et al. [15] present a drowning detection method based on background subtraction. Zhang et al. [16] detect drowning according to the rate of change of the human body area in the alert zone. Salehi et al. [35] use the HSV threshold mechanism and contour detection function to track swimmers and count time with a camera installed above swimming pools. Prakash et al. [18] describe a near-drowning early prediction technique using novel equations. Hou et al. [20] propose swimming target detection and tracking technology based on a discrete cosine transform algorithm to analyze motion parameters for drowning detection. Lei et al. [21] analyze the spatial relationship between the target's location information and the swimming/drowning area of swimming pool to further determine the swimmer's drowning or swimming behaviour.

The traditional vision-based drowning detection methods mainly detect human bodies through the videos. Then the bodies' features are extracted to detect drowning persons, including motion, position and time, which can't also detect the early drowning phenomenon.

2.2 | Supervised learning vision-based method

Lu et al. [13] tracked swimmers with a camera and analyzed the characteristics of swimmers' movements and body features by a finite state machine. In 2004, Lu et al. [36] propose a new drowning detection method, which uses a visual module to detect and track swimmers, and uses an event reasoning module based on a finite state machine to analyze swimmers' video sequences and detection drowning behaviours. The DEWS team [37] develops a module containing data fusion and hidden Markov models to learn the characteristics of different swimming behaviours. Pavithra et al. [38] built a drowning detection system using Raspberry Pi with USB camera, and used Faster RCNN to classify normal swimmers and drowning people. Li et al. [19] obtain the final swimming pool intelligent-assisted drowning detection results through the YOLO principle. Hasan et al. [39] propose a water behaviour dataset curated to support the design of image-based methods for drowning detection.

These supervised vision-based drowning detection technologies extract the features of drowning behaviour by simulative drowning events. The state of swimmers is classified by the supervised classification. This lacks robustness and authenticity. Moreover, the current drowning detection methods are relatively complicated.

2.3 | Unsupervised learning vision-based method

Few researchers have used methods based on unsupervised deep learning to detect drowning events. Unsupervised learning is a popular method in VAD. In the VAD task, real abnormal samples are very few. And in the drowning detection task, real drowning video samples are also very few. In recent years, many VAD methods based on unsupervised deep learning are proposed. They hypothesize that the anomalies are hard to be

reconstructed and predicted well by the autoencoders trained only on normal data. Hasan et al. [27] propose a convolutional autoencoder-based anomaly detection method that works with non-supervision. Park et al. [28] add a memory module to the convolutional autoencoder with a new updated scheme. The items in memory record the prototype patterns of normal data, improving the discrimination of features from normal data. LV et al. [29] propose a dynamic prototype unit (DPU) to encode normal dynamics without additional memory cost, introducing meta-learning into DPU to form a novel few-sample normal learner. Liu et al. [30] first propose predicting future frames and using prediction error as an anomaly indicator. Lu et al. [31] introduce a Conv-LSTM for prediction. Hu et al. [40] propose a self-attention prototype unit to encode the normal latent space as prototypes. They also introduce a circulative attention mechanism to backbone to form a novel feature extracting learner. Recently, Liu et al. [32] develop a hybrid VAD framework that first extracts foreground and integrates optical flow reconstruction and frame prediction. However, this method is only applicable to anomalous events with large changes in motion velocity and is not suitable for detecting early quiet drowning phenomena. Moreover, judging anomalies based on pixel-level reconstructed error or prediction error is more susceptible to environmental noise interference and unsuitable for pool videos.

In this paper, we proposed a drowning detection algorithm via the deep Gaussian model to solve the problems of noisy underwater video environments and inconspicuous early drowning features in drowning detection work. The proposed algorithm can distinguish normal and drowning frames in high-level semantic features, with higher robustness and fast computing speed. Meanwhile, we design a drowning detection device based on underwater computer vision, which can realize real-time and edge drowning detection by using embedded AI devices.

3 | THE STRUCTURE OF THE VIDEO DROWNING DETECTION DEVICE

The structure of our proposed video drowning detection device is shown in Figure 1, which is mainly composed of a camera, an embedded AI device, a buck module, an LED light strip, and a waterproof casing. The embedded AI device uses NVIDIA's Jetson Nano artificial intelligence development board, which is responsible for carrying and running the drowning detection algorithm. The camera uses Sony MIX219 with a resolution of 800 W, which is responsible for acquiring underwater video streams. The buck module reduces the 12 V power supply to 5 V to power the embedded AI device, and the continuous output current is 6 A. The LED light strip adopts the 12 V 2835 type SMD light strip, which can improve the imaging effects of the camera in the low light state, and provide the necessary lighting for the swimming pool at night. The waterproof case has a diameter of 180 mm and a height of 61 mm. There are many through holes on one side for placing cameras and LED strips, and they are sealed and covered with a panel made of glass. The other side is equipped with a buckle for fixing the detection device to the pool wall. The detection device is powered by 12 V power supply. In order to ensure its waterproofness and integrity, the interior of the detection device is completely filled with waterproof silicone rubber. We communicate with the detection device through a reserved unshielded twisted pair network cable. This drowning detection device can carry and run the proposed drowning detection algorithm on an embedded AI device and complete the drowning detection work without uploading to the server.

The technical roadmap of the proposed drowning detection device is shown in Figure 2. The camera will transmit the captured underwater video stream to the Jetson Nano through the CSI or USB interface. Jetson Nano will sequentially perform underwater video enhancement, underwater human detection, unsupervised anomaly detection, and the final decision according to the deployed algorithm flow. When the detection device detects a drowning person, an early warning signal is sent through the twisted pair network cable, and the lifeguard is notified in time to pay attention to the state of swimming pool to ensure the drowning person's safety. In addition, in order to protect the privacy of swimmers, the detection device will notify and upload the video of the drowning person only when a drowning person is detected, and will not save and upload any underwater video under normal conditions.

4 | DESIGN OF DROWNING DETECTION ALGORITHM

As illustrated in Figure 3, the proposed drowning detection algorithm is composed of two stages: underwater near-vertical human detection and unsupervised anomaly detection based on deep Gaussian model. The whole algorithm is trained on normal data. At test time, the Mahalanobis distance from the feature vector extracted by DDN to the standard multivariate Gaussian model is used for drowning detection.

In the following sections, we introduce underwater near-vertical human detection, deep Gaussian model, and DDN in detail. Finally, we show how to use the proposed algorithm for drowning detection.

4.1 | Underwater near-vertical human detection

Underwater videos in swimming pools may appear the phenomenon of colour recession, low contrast, and blurred details, affecting subsequent drowning detection. To solve this problem, contrast limited adaptive histogram equalization (CLAHE) [41] is applied for underwater image preprocessing. It can quickly improve the contrast of underwater images, making human bodies more prominent and enhancing information about the area related to the drowning detection task. The CLAHE algorithm is applied for each of three RGB channels.

From our investigation [42], we find that drowning people often present near-vertical posture, so reliable near-vertical

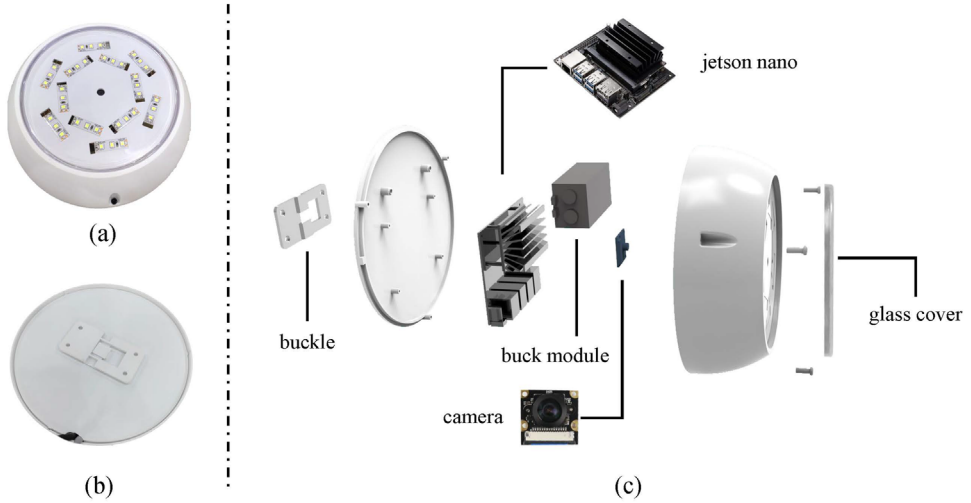

FIGURE 1 The structure of the drowning detection device. (a) The front of the detection device. (b) The back of the detection device. (c) The structure diagram of the internal components of the detection device.

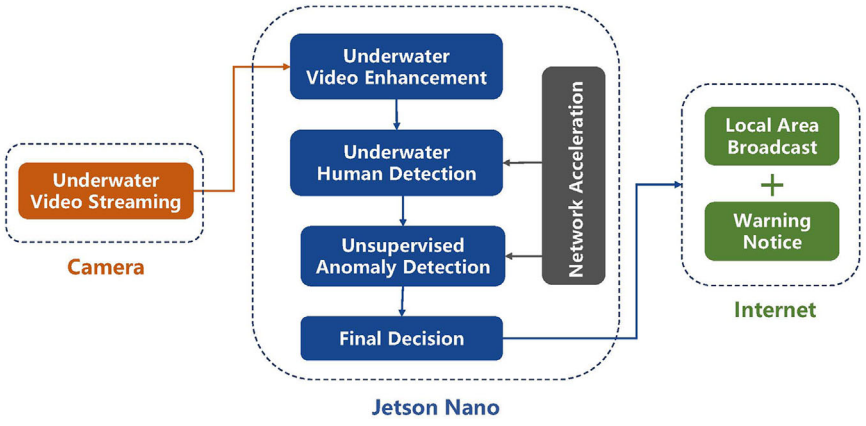

FIGURE 2 The technical road-map of the proposed drowning detection device.

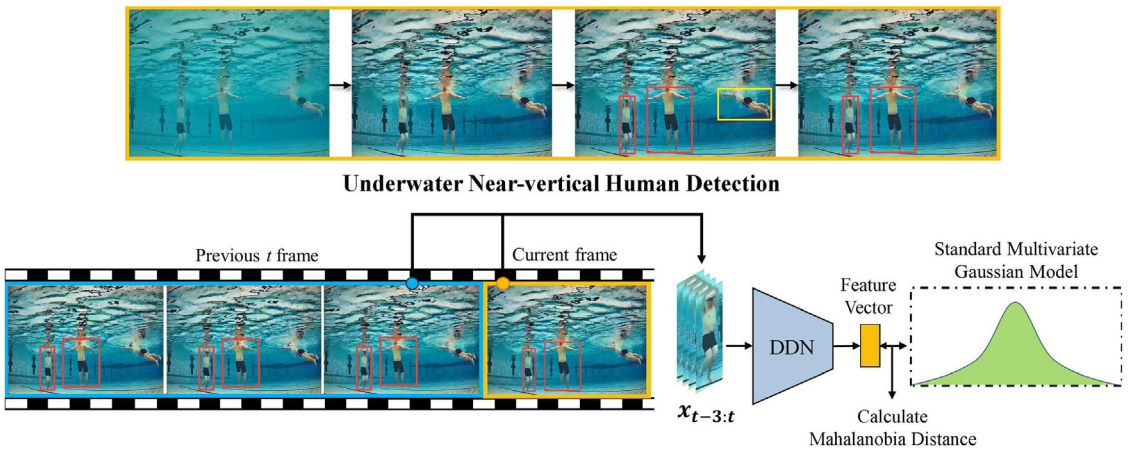

FIGURE 3 The proposed algorithm flow for drowning detection, which contains underwater near-vertical human detection and unsupervised anomaly detection based on deep Gaussian model.

human detection is an important foundation for drowning detection. YOLOv5n [43] is a lightweight network version of YOLOv5, with simple design and good performance. Through our experiments, compared with other lightweight detection networks, YOLOv5n can achieve a good balance between accuracy and reasoning speed. The YOLOv5n is trained on the provided underwater object detection dataset. For YOLOv5n, the loss function is composed of three losses as shown below:

$$L_{YOLOv5n} = l_{box} + l_{obj} + l_{cls} \tag{1}$$

In this case, l_{box} is the bounding box loss, l_{obj} is the object confidence loss, and l_{cls} is the class loss. The bounding box loss is defined as follows:

$$l_{\text{box}} = \lambda_{\text{box}} L_{\text{giou}} \tag{2}$$

λ_{box} is the factor of bounding box loss. L_{giou} is the GIoU loss [44]. For any two rectangular boxes A and B, first find a smallest rectangular box C that can enclose them. Then calculate the ratio of the area of $C \backslash (A \cup B)$ to the area of C (the area of $C \backslash (A \cup B)$ is the area of C minus the area of $A \cup B$). Finally, use the IoU value of A and B to minus this ratio to get GIoU. Thus, GIoU Loss is defined as follows:

$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|} \tag{3}$$

$$L_{\text{giou}} = 1 - GIoU \tag{4}$$

The object confidence loss is defined as follows:

$$\tilde{C}_i = \hat{C}_i \log \sigma(C_i) + \left(1 - \hat{C}_i\right) \log\left(1 - \sigma(C_i)\right) \tag{5}$$

$$l_{obj} = -\sum_{i=0}^{S^2} \sum_{j=0}^{M} I_{ij}^{obj} \tilde{C}_i - \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{M} I_{ij}^{noobj} \tilde{C}_i \tag{6}$$

where $\hat{C}$ is the confidence of the predicted box and C is the confidence of the ground truth box. σ is the Sigmoid function. S is the size of the grid, M is the number of anchor boxes, λ_{noobj} is the factor used to decrease the probability of a bounding box without an object. If the box at (i, j) has an object, I_{ij}^{obj} is 1, otherwise it is 0. I_{ij}^{noobj} and I_{ij}^{obj} are opposites.

The class loss is defined as follows:

$$P_{i,c} = \hat{p}_i(c) \log \sigma\left(p_i(c)\right) + \left(1 - \hat{p}_i(c)\right) \log\left(1 - \sigma\left(p_i(c)\right)\right) \tag{7}$$

$$l_{cls} = -\lambda_{co} \sum_{i=0}^{S^2} I_i^{obj} \sum_{c \in cla} P_{i,c} \tag{8}$$

where $\hat{p}(c)$ is the predicted box classification function, and $p(c)$ is the ground truth box classification function. $\lambda_{co}$is the factor that increases the probability of a bounding box with an object. $cla$is the number of classes. S, σ, and I_i^{obj} have the same meaning in Formulas (5) and (6).

Our research shows people's posture is nearly horizontal when swimming, and nearly vertical when drowning and treading water. Therefore, our drowning detection algorithm will focus on near-vertical people, which will greatly solve the previous problem of susceptibility to environmental changes and make the results of drowning detection more accurate. To quickly filter out the complete near-vertical human body, we propose the following near-vertical human extraction strategies:

1. The width-to-height ratio of the human bounding box must be larger than 110%.
2. The left/right boundaries of human bounding box must be larger than 20 pixels from the left/right boundaries of the input image.
3. The ratio of the intersection area of the human bounding box with other human bounding boxes to its area is not larger than 20%.
4. The area of human bounding box is not less than 5000 pixels.

Figure 4 shows that the proposed human detection method can better extract the complete near-vertical human, which can screen out many irrelevant non-drowning humans and improve the performance of drowning detection methods.

Then each near-vertical human is identified by a region of interest (ROI) bounding box. For each near-vertical human ROI, we build a STC that contains the object in the current frame and the contents in the same bounding box of previous t frames by stacking ROI in the RGB channel dimension, where $t = 3$, as shown in Figure 5. The width and height of STC are both resized to 32.

4.2 | Unsupervised anomaly detection based on deep Gaussian model

Near-vertical human STCs are obtained from the bounding box for unsupervised anomaly detection. In drowning detection, these anomaly humans are treated as drowning people, which solves the lack of drowning videos and the inauthenticity of simulative videos.

Real drowning samples are very few and drowning is an anomaly event. So, the vision-based drowning detection is similar to VAD. Most of the existing unsupervised deep learning-based VAD first reconstruct or predict frames, and then determine anomalies based on the pixel-level reconstruction error or prediction error. However, the difference between drowning human STCs and normal human STCs on the pixel level is not clear enough sometimes. These methods are still more susceptible to the interference of environmental noise and are not suitable for pool videos, causing undesirable results in drowning detection. To solve these problems, the deep Gaussian model is proposed. In machine learning, Gaussian model-based anomaly detection is a classical approach. It can

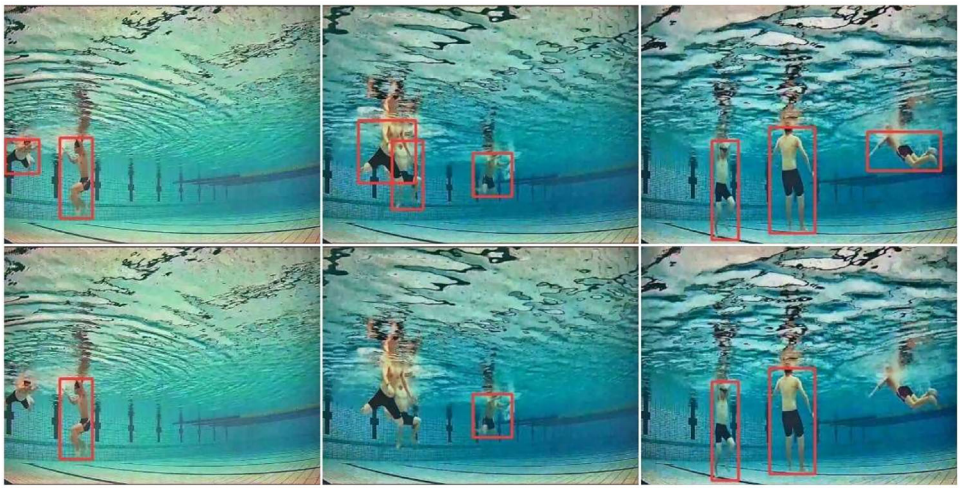

FIGURE 4 Extraction results of a near-vertical human body. Top: Yolov5n detection results. Bottom: Results after filtering by the proposed near-vertical human strategies.

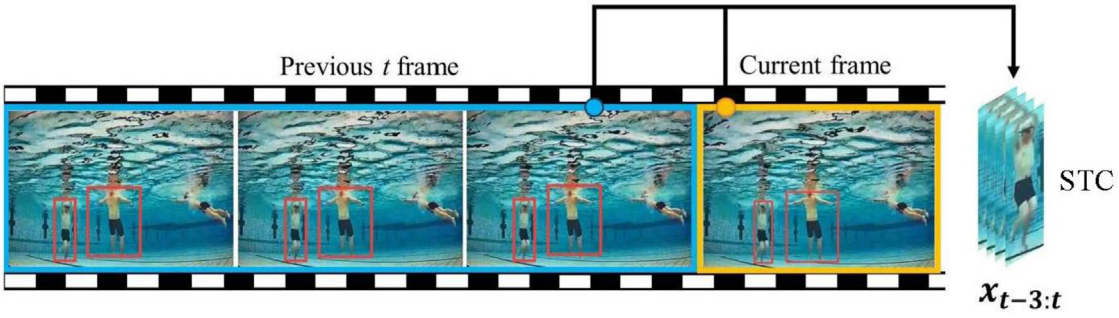

FIGURE 5 STC is formed by the current frame and previous t frames. STC, spatial-temporal cube.

automatically figure out possible relations among features by constructing a multivariate Gaussian model of the features of normal data. The Mahalanobis distance from the feature vector of the test sample to the Gaussian model is calculated. If the Mahalanobis distance is larger than the threshold, it indicates that the test sample is abnormal and that unsupervised anomaly classification is achieved.

However, traditional Gaussian model-based anomaly detection manually sets the features of the data, which is troublesome and poorly generalized. Considering human STCs in swimming pools have complex features, we put forward deep Gaussian model that uses the deep convolutional neural network to extract features of human STCs without setting features manually. Next, the details of the proposed depth Gaussian model will be introduced in detail. The following is a detailed description of deep Gaussian model.

First, a lightweight DDN is proposed for extracting the feature of human STCs. We improve ShuffleNetv2 [45] to build DDN. ShuffleNetv2 is originally a lightweight convolutional neural network for image classification and provides a good balance between running speed and classification accuracy. DDN, with fewer parameters and faster running speed, meet the real-time requirements of drowning detection tasks. Figure 6 shows the architecture of the proposed DDN. DDN mainly consists of two convolution layers, three max pooling layers, two Shuffle Blocks, and a fully connected layer. The normalization strategy is instance normalization (IN) [46]. The activation function is leaky ReLU (LReLU). The Max pooling layer is used for downsampling. The size and stride of the kernels of all max pooling layers are set to 2×2.

Figure 7 shows the detailed network architecture of Shuffle Block. In Shuffle Block, the input feature map is first divided into two branches in the channel dimension. The ratio of the number of bottom branch channels to the number of input channels is c where $c = 0.5$. Then the bottom branch is mapped equally, and the top branch goes through 1×1 convolution, 3×3 depthwise separable convolution (DWConv), and 1×1 convolution successively, which can reduce the parameters very well. Finally, the outputs of the two branches are concatenated together for channel shuffle to ensure information exchange between the two branches.

Second, a standard multivariate Gaussian model is built to fit all the feature vectors of normal human STCs extracted by DDN. It is as a modelling target of the feature vectors of normal

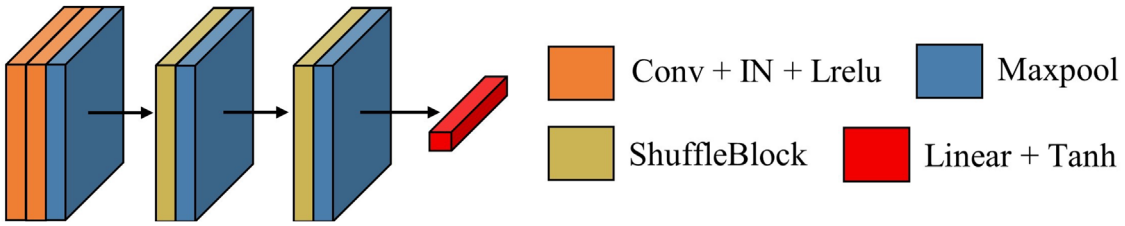

FIGURE 6 The architecture of the proposed drowning detection network.

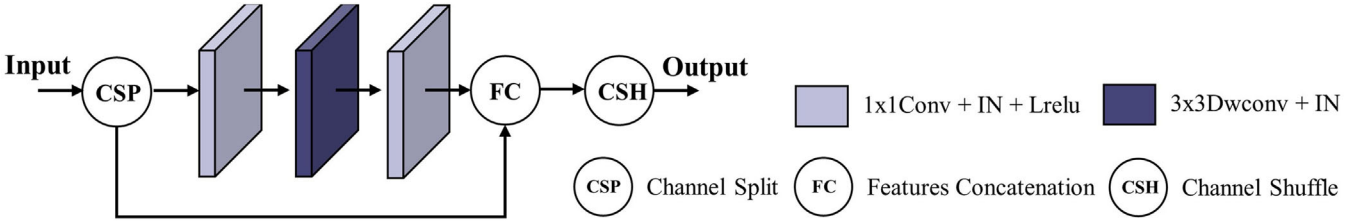

FIGURE 7 Detailed architecture of shuffle block.

human STCs. The probability density function of the standard multivariate Gaussian model is shown below:

$$X = -\frac{1}{2}(\phi(x_{t-3:t}) - \mu)^T \Sigma^{-1} (\phi(x_{t-3:t}) - \mu) \tag{9}$$

$$p(x_{t-3:t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp(X) \tag{10}$$

where $x_{t-3:t}$ is the input STC, $\phi(x_{t-3:t})$ is the output of DDN. X is an intermediate variable, T is the matrix transpose, t is the number of previous frames, and n is the dimensionality of the feature vector. Here $t = 3$, $n = 128$. Because the mean vector and the covariance matrix of the standard multivariate Gaussian model are zero vector and identity matrix, μ is a zero vector and Σ is an identity matrix.

Then, as shown in (a) of Figure 8, we want that the extracted features of normal human STCs can be closer to the standard multivariate Gaussian model, so DDN is trained on a dataset with only normal human STCs in pools by using the Mahalanobis distance from the output feature vector of DDN to the standard multivariate Gaussian model. The loss function of DDN is as follows:

$$L = (\phi(x_{t-3:t}) - \mu)^T \Sigma^{-1} (\phi(x_{t-3:t}) - \mu) + \lambda \parallel W \parallel_2^2 \tag{11}$$

where λ is the hyperparameter of regularization. W is the weight parameter of the model. Other variables have the same meanings as above.

Finally, as shown in (b) of Figure 8, in the testing phase, the feature vectors of normal human STC extracted by DDN after training are very close to the standard multivariate Gaussian model. The standard multivariate Gaussian model is regarded as the model of normal human features. The square of Mahalanobis distance between the feature vector of human STC and the standard multivariate Gaussian model can reflect the degree of deviation of input human body from normal human body. The anomaly score is the square of Mahalanobis distance from the feature vector of the test sample to the standard multivariate Gaussian model. The proposed algorithm can better extract high-level semantic features of normal human STCs and distinguish them at a high level, which is beneficial to perform anomaly detection of drowning human STCs. It has greater robustness than the pixel level-based method. The anomaly score is described as follows:

$$Score = \left\| (\phi(x_{t-3:t}) - \mu)^T \Sigma^{-1} (\phi(x_{t-3:t}) - \mu) \right\| \tag{12}$$

where the meaning of each variable is the same as above.

Higher anomaly scores indicate a greater possibility of human abnormality or drowning. The maximum anomaly score in STCs is considered as the anomaly score of a frame. An optimal threshold on the receiver operating characterisitc (ROC) curve will be calculated based on the Youden Index, and frames with an anomaly score higher than the threshold indicate drowning.

5 | EXPERIMENTS

In this section, the experimental results of the proposed drowning detection device and algorithm via the deep Gaussian model are shown. We evaluate the proposed drowning detection algorithm and other methods with the dataset collected in pools. Finally, to realize deployment and edge computing on embedded platforms, we prune and accelerate the proposed and used network.

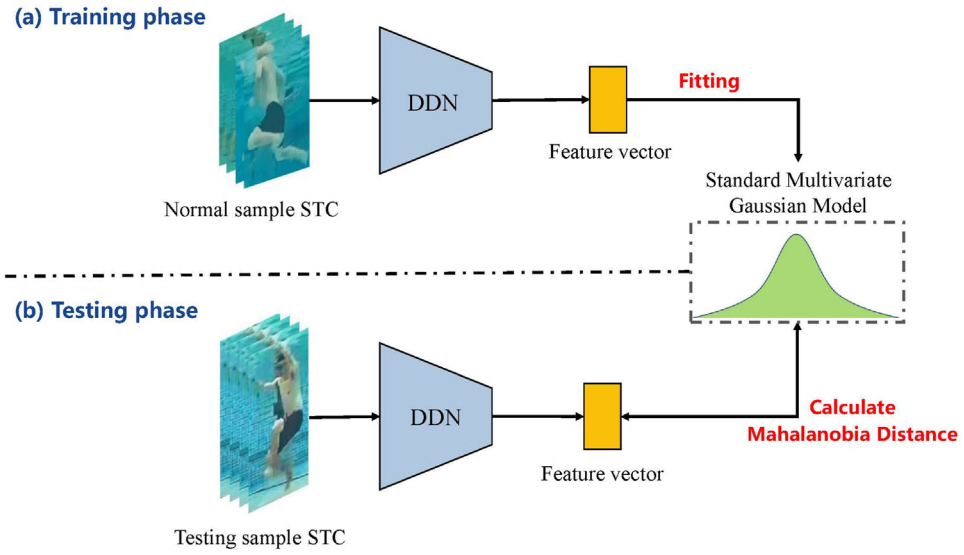

FIGURE 8 Unsupervised drowning detection training and testing phase. (a) The training phase of DDN. (b) The testing phase of DDN. DDN, drowning detection network.

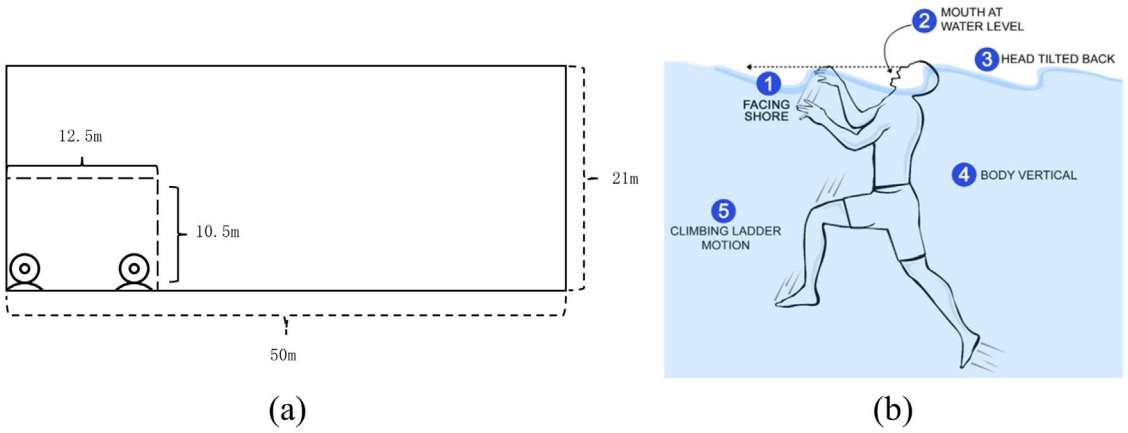

FIGURE 9 Experimental details. (a) Schematic diagram of the location of the underwater cameras. (b) The five characteristics that often appear when drowning.

5.1 | Datasets

Nowadays there are no public datasets about crowd swimming. In this work, a dataset named Pool is collected with underwater cameras installed on the pool wall. The standard swimming pool is 50m × 21m. Therefore, two cameras in a small area of 12.5m × 10.5m are installed according to the pool's size and the underwater cameras' parameters. The position diagram of underwater cameras is shown in Figure 9a. The recording is performed at 25 frames per second (FPS) with a resolution of 640 × 480.

The volunteers perform normal actions such as swimming and treading water in the pool. Meanwhile, the volunteers perform some simulative drowning actions according to the five drowning characteristics in Figure 9b for testing. Videos with the obvious and clear crowd are selected to add into the dataset. Crowd density changes from sparse to crowded and the size of persons changes greatly. The dataset has 27 training clips and 13 test clips. There are 5374 frames in the training set and 2301 in the testing set. The behaviour in training clips consists only of normal swimming and treading. That in testing clips consist of normal swimming, treading, and abnormal drowning. Some samples of the Pool dataset are shown in Figure 10. A part of the training set in the Pool dataset is used to train the YOLOv5n model. The YOLOv5n model uses the pre-trained model weights on the COCO dataset as the initial weights of training. This underwater human detection dataset only contains 1400 images of normal human bodies.

5.2 | Evaluation metric

To evaluate the proposed algorithm, ground truth is manually created for each frame of each video sequence to indicate whether there a drowning event is occurring in that frame. The

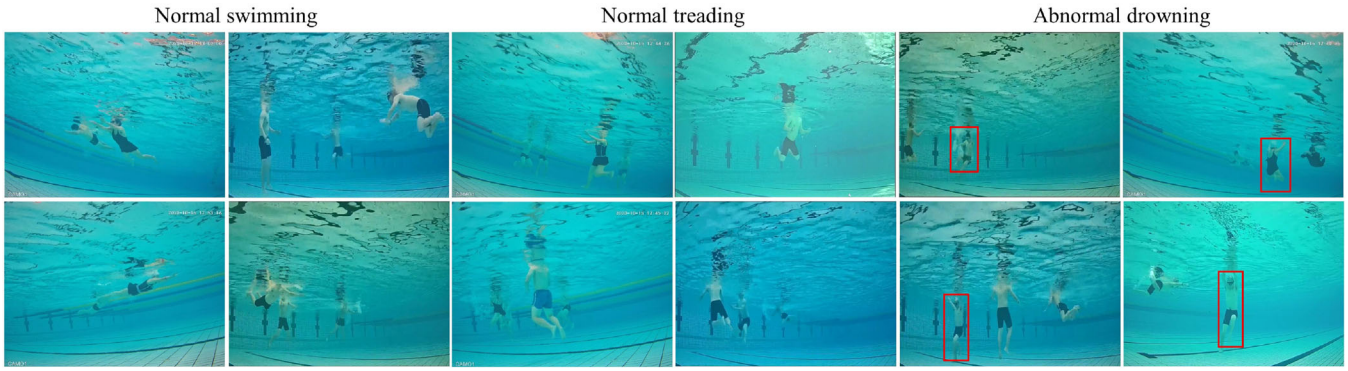

FIGURE 10 Some samples including normal and abnormal drowning frames in the Pool dataset are illustrated. Red boxes denote drowning in abnormal frames.

ROC curve is generated by constantly changing the threshold to calculate the true positive rate (TPR) and false positive rate (FPR). The average area under the ROC curve (AUC) is calculated, which represents the algorithm's ability to rank positive and negative samples. A drowning frame is positive. A high AUC indicates that scores of drownings predicted by the algorithm are usually higher than those of normal occasions and that the algorithm has better classification performance. Besides, the average precision (AP) is calculated. AP is the ability to predict drowning correctly. Higher AP indicates higher accuracy in the classification of the algorithm. In addition, we also use the equal error rate (EER) to evaluate the performance. EER indicates the probability of misclassification. Speed is measured in FPS.

5.3 | Performance evaluation

5.3.1 | Implementation details

We adopt Adam optimizer to optimize the parameter of YOLOv5n and DDN. The input image size of YOLOv5n is 640 × 640. The learning rate, batch size, and epoch number of YOLOv5n are set to (1×10^{-3}, 64, 100). Then, the learning rate, batch size, and epoch number of DDN are set to (1×10^{-3}, 256, 60). The regularization hyperparameter λ is 1×10^{-2}. The Adam optimizer is used to optimize the network parameters of the DDN. The two neural network models are trained on the Pool training set with only the normal scene. The experiments are performed on an NVIDIA GTX 1080Ti GPU and Intel Core (TM) i7-7800X CPU.

5.3.2 | Detection performance

First DDN is trained on the pool training set with only normal humans. Then we visualize the output feature vectors of DDN by t-SNE [47] dimensionality reduction technique on the Pool test set including normal human STCs and drowning human STCs. Figure 11 shows that the output feature vectors of normal human STCs are closer to the centre than those of drowning human STCs, indicating that the output feature vectors of drowning human STCs are further away from the standard multivariate Gaussian model and deviate more from the normal pattern of the training set. Therefore, the output feature vectors of drowning human STCs will produce a larger Mahalanobis distance and a larger anomaly score. Figure 12 shows the intuitive anomaly curves of two testing videos. An anomaly curve shows the anomaly scores of all video frames sequentially. The proposed method can produce higher and stable anomaly scores when a drowning event occurs and better detect anomalous drowning events.

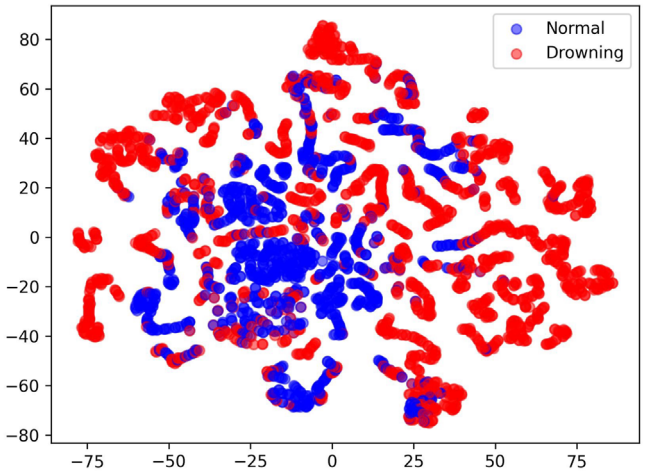

FIGURE 11 Output feature vectors of DDN by 2D t-SNE dimensionality reduction in the Pool test set. DDN, drowning detection network.

Then we compared the performance of the proposed method with other state-of-the-art VAD methods on the Pool dataset. Figure 13 shows that the ROC curves of the proposed method are above the ROC curves of other methods. The proposed method outperforms other methods and has better drowning detection performance.

Ano-Pred [48], MNAD [28], and MPN [29] detect anomalies by reconstructing the whole video frame and calculating reconstructed error. APN [40] and sRNN-AE[49] predict the whole frame and detect anomalies by the compactness error of feature reconstruction term and frame prediction error.

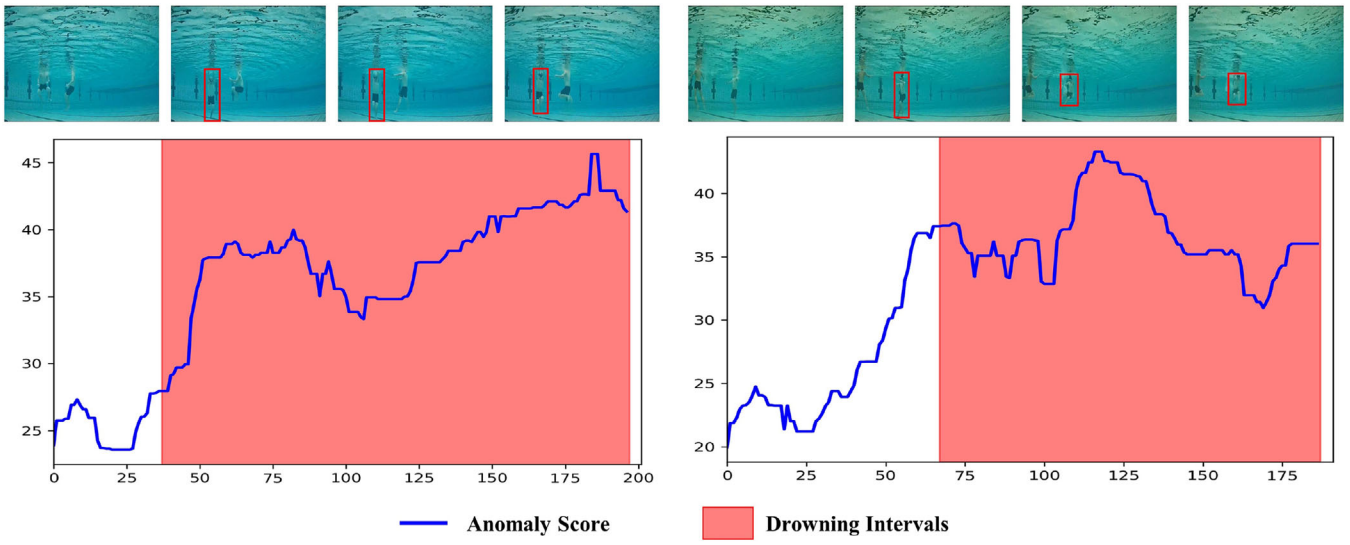

FIGURE 12 Drowning detecting examples on Pool dataset. The horizontal axis denotes time, while the vertical axis denotes anomaly score (a higher value indicates more possibility to be abnormal).

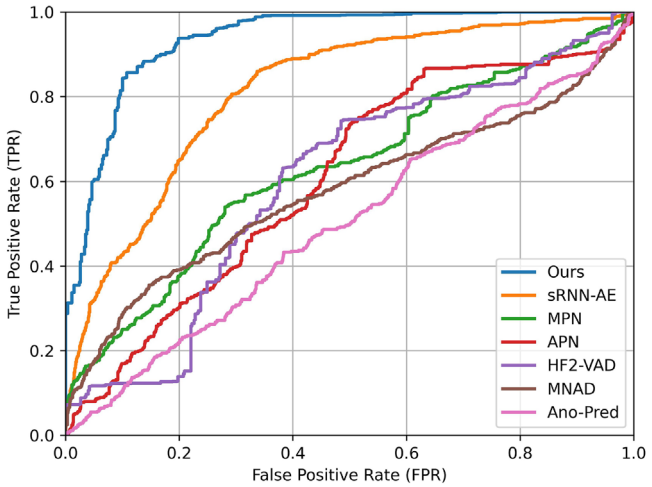

FIGURE 13 ROC of various methods on the Pool dataset.

TABLE 1 Performance metrics of different methods on the Pool dataset

	AUC (%)	EER (%)	AP (%)	FPS	FPS (edge)
APN	60.7	43.2	30.4	19	1.0
MNAD	52.6	50.3	27.8	33	1.7
MPN	62.2	36.7	34.9	31	1.1
HF2-VAD	60.1	38.0	34.9	3	0[a]
Ano-Pred	54.3	55.9	24.9	25	1.1
sRNN-AE	81.2	25.1	59.0	13	0.5
Ours	**93.6**	**12.8**	**97.5**	**127**	**13**

[a]HF2-VAD cannot run smoothly on the Jetson Nano due to its huge network model.
AP, average precision; AUC, area under the ROC curve; EER, equal error rate; FPS, frames per second.

They are only suitable for scenes with relatively stable video frame backgrounds. However, in the underwater environment of swimming pools, light and ripples in the water surface often change irregularly. These have a large adverse effect on the performance of the whole-frame reconstruction-based VAD. So, they are easily disturbed by environmental noise and not very robust. HF2-VAD [32] first extracts foreground objects in frames and then predicts human STCs with reconstructed optical flow. Still, it does not screen out incomplete humans and non-drowning humans in horizontal swimming, which will largely interfere with its modelling of normal humans and reduce its anomaly detection performance. Moreover, HF2-VAD uses the prediction error and the reconstructed optical flow error of STCs to detect an anomaly. These are low-level semantic features and still easily disturbed by environmental noise. Also, the calculation of inter-frame optical flow is very time-consuming.

The proposed method first performs underwater near-vertical human detection, which largely reduces the interference caused by environmental noise. Then four consecutive frames of ROI are composed into an STC, which cleverly incorporates spatial-temporal information. The proposed lightweight DDN can also extract the features of STCs quickly. Then the Mahalanobis distance from the feature vector to the standard multivariate Gaussian model is calculated, which can detect anomaly in high-level semantic features. The proposed method has high robustness. We tested the performance and speed of these algorithms on server devices and edge devices (Jetson Nano), respectively. Table 1 shows that the proposed drowning detection method has a good comprehensive performance of drowning detection in the test set, and can achieve the highest AUC, AP, and lowest EER. At the same time, the speed of our algorithm performs well on both server and edge devices, basically meeting the real-time requirements.

Figure 14 shows the qualitative results of the proposed drowning detection algorithm, where the bounding boxes of all

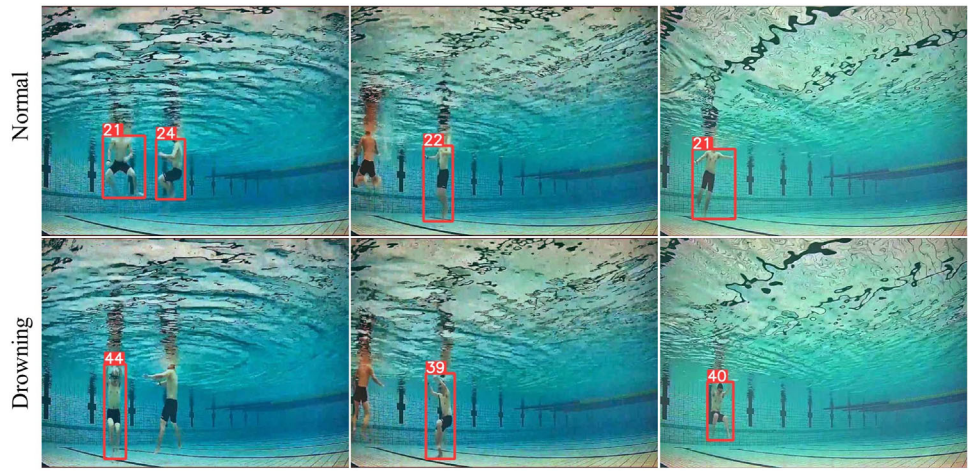

FIGURE 14 Qualitative performances of the proposed drowning detection algorithm in normal and drowning events.

detected near-vertical people are kept. The number on the rectangular boxes represents anomaly scores. The anomaly score of drowning people is higher than that of normal treaders, indicating that the proposed drowning detection framework can distinguish the two well.

5.3.3 | Deployment and acceleration on embedded AI devices

According to the drowning detection algorithm proposed above, we combined embedded AI device, camera, and other equipment to design a drowning detection device based on underwater computer vision. The device can greatly protect the privacy of swimmers. Only when drowning is detected, the underwater video of drowning person be uploaded to alert lifeguards. Embedded AI device is the primary carrier of edge computing. Therefore, we test the speed of the proposed drowning detection algorithm on Jetson Nano (an NVIDIA embedded AI device). Meanwhile, TensorRT, a high-performance deep learning inference framework, is adopted to accelerate YOLOv5n and DDN, the two neural network models. To further improve the speed, we also use TensorRT C++ Application Programming Interface (API). Figure 15 shows that the inference speed is greatly improved. Finally, the proposed algorithm with acceleration runs at a speed of 31 FPS in Jetson Nano, which meets the real-time requirement.

Then, the performance of the proposed algorithm with and without acceleration is compared. A drowning threshold is set and F1 score is calculated. Table 2 shows that AUC, AP, EER, and F1 scores of the proposed algorithm have little difference after acceleration.

TABLE 2 Performance metrics on Pool dataset after acceleration

	AUC	EER	AP	F1	FPS
w/o acceleration	93.6%	12.8%	97.5%	90.8%	15
Acceleration	93.5%	13.0%	97.4%	90.6%	31

AP, average precision; AUC, area under the ROC curve; EER, equal error rate; FPS, frames per second.

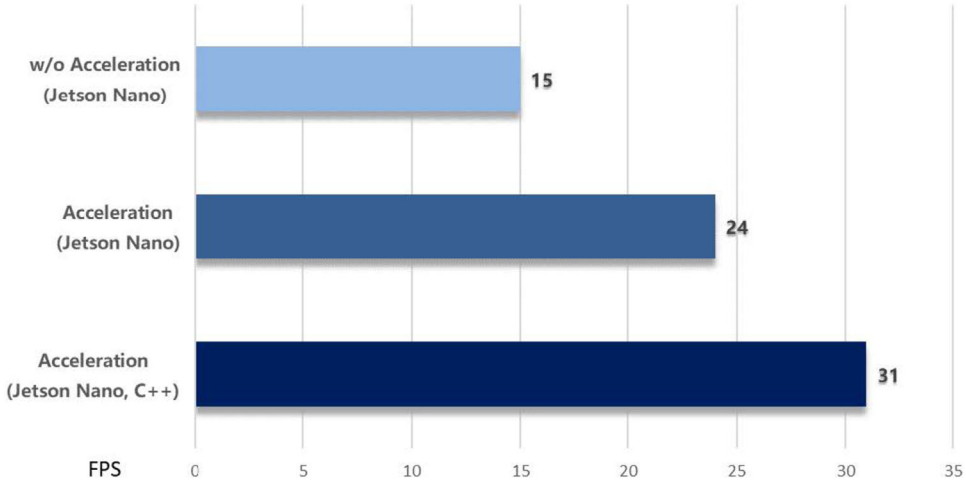

FIGURE 15 Speed comparison before and after model acceleration.

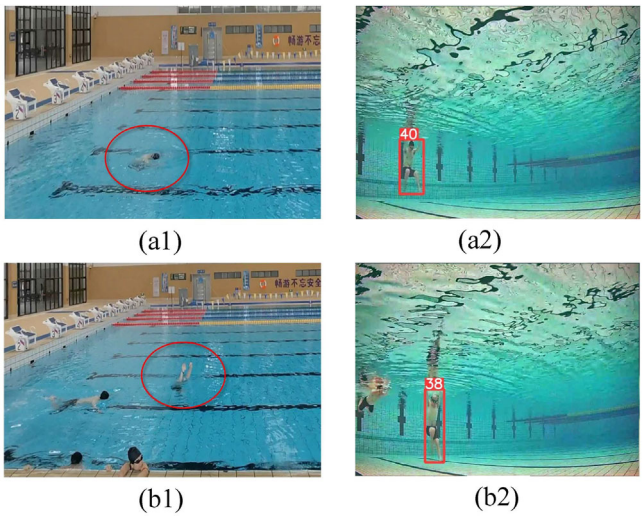

FIGURE 16 Part of the field test results of the detection device. (a1), (b1) The positions of the drowning persons. (a2), (b2) The underwater video received by the terminal.

Next, we field test the designed drowning detection device in a swimming pool. First, a detection device is fixed on the pool wall to capture the underwater video stream and enable drowning detection. Then use the laptop as a terminal on the shore to receive the early warning information and underwater video transmitted by the detection device. We invited several volunteers to perform normal swimming and treading water in the swimming pool. We also let them simulate drowning action according to the posture characteristics of the drowning person. In this test, the threshold of abnormal score is set as 29. When the abnormal score of somebody is greater than or equal to this threshold, the terminal will receive warning information and an underwater drowning video. Part of the test results are shown in Figure 16 The sub-figures (a1) and (b1) mark the position of the drowning person, and the sub-figures (a2) and (b2) show the underwater video received by the terminal. Meanwhile, we calculated the F1 score, precision and recall of multiple videos based on the test results. Figure 17 shows the evaluation results of each video, and the F1 score, precision, and recall of each video area are stable at around 90%. The experimental results show that the proposed algorithm also has good robustness and drowning detection performance in practical applications.

6 | CONCLUSION

In this paper, we propose a drowning detection device based on underwater computer vision, composed of Jetson Nano, camera, waterproof case, and other components. The real-time detection of drowning events on the captured underwater video stream is realized through the proposed deep Gaussian model-based drowning detection algorithm. The proposed drowning detection algorithm includes two main stages. The first stage is underwater near-vertical human detection. The proposed near-vertical human detection strategies can effectively solve the interference of the environment and provide a reliable basis for drowning detection. The second stage is unsupervised anomaly detection based on deep Gaussian model. A lightweight DDN is proposed for extracting the feature vectors of human STCs quickly. Then the Mahalanobis distance from the feature vector to the standard multivariate Gaussian model is calculated. It can achieve unsupervised drowning detection in the high-level semantic features to solve the lack of drowning videos and the inauthenticity of simulative videos. The proposed algorithm has higher robustness than the pixel level-based method. Also, a dataset named Pool including many pools underwater videos is collected and the proposed method can process pool videos in real time at the edge.

The experimental results show that the proposed algorithm has an excellent good comprehensive performance in drowning detection. At the same time, by accelerating the proposed algorithm, the drowning detection device we designed can complete the real-time accurate drowning detection work on the

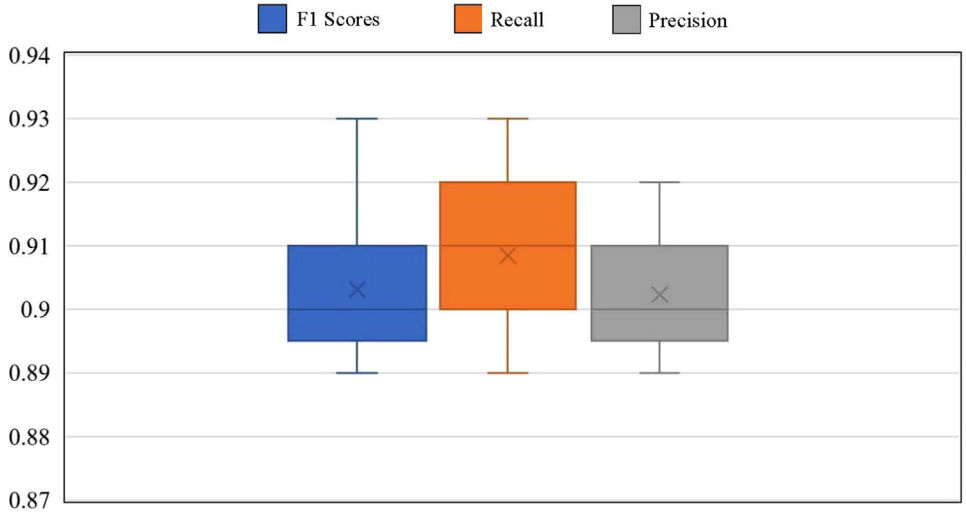

FIGURE 17 Field test multi-segment test results.

underwater video stream, which has a more excellent practical application value. However, there are still some things that could be improved in our research. For example, the proposed method cannot detect too large or too small humans. Also, it cannot detect drowning effectively when people cover each other for a long time. These limitations of the proposed methods are the direction of future improvement.

CREDIT CONTRIBUTION STATEMENT

Tingzhuang Liu: Software, Validation, Writing—review & editing. Xinyu He: Project administration, Investigation, Resources. Linglu He: Writing—original draft, Visualization. Fei Yuan: Methodology, Supervision, Conceptualization

ACKNOWLEDGEMENTS

The authors would like to thank the National Natural Science Foundation of China (62071401) and Xiamen Ocean and fishery Development Special Fund project (21CZB015HJ10).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest in this research work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Fei Yuan https://orcid.org/0000-0002-8614-8756

REFERENCES

1. Organization, W.H.: Preventing drowning: an implementation guide (2017).
2. Organization, W.H., Bloomberg, L.P.: Global Report on Drowning: preventing a Leading Killer (2014).
3. UNICEF, Bank, W., Division, U.P.: Levels and trends in child mortality 2013. Lancet 243(6288), 317 (2014).
4. Laxton, V., Crundall, D.: The effect of lifeguard experience upon the detection of drowning victims in a realistic dynamic visual search task. Appl. Cogn. Psychol. 32, 14–23 (2018)
5. Kharrat, M., Wakuda, Y., Koshizuka, N., Sakamura, K.: Near drowning pattern recognition using neural network and wearable pressure and inertial sensors attached at swimmer's chest level. In: 2012 19th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), IEEE, pp. 281–284 (2012).
6. Roy, A., Srinivasan, K.: A novel drowning detection method for safety of swimmers. In: 2018 20th National Power Systems Conference (NPSC), IEEE, pp. 1–6 (2018).
7. Lei, Y., Chen, M., Sun, T., et al. Application of BeiDou navigation satellite system in anti-drowning system[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing. 012009 (2018).
8. John, S.N., Ukpabio, I.G., Omoruyi, O., Onyiagha, G., Noma-Osaghae, E., Okokpujie, K.O.: Design of a drowning rescue alert system. Int. J. Mech. Eng. Technol. 10(1), 1987–1995 (2019).
9. Dehbashi, F., Ahmed, N., Mehra, M., Wang, J., Abari, O.: Swimtrack: Drowning detection using rfid. In: Proceedings of the ACM SIGCOMM 2019 Conference Posters and Demos, pp. 161–162 (2019)
10. Monish, P., Darshan, R., Ponvalavan, K., Bharathi, M.: Drowning alert system using rf communication and gprs/gsm. J. Phys. Conf. Ser. 1997, 012044 (2021)
11. Sneha, M.: An automatic drowning detection and rescue system. Int. J. Res. Appl. Sci. Eng. Technol. 9, 1021–1028 (2021)
12. Meniere, J.: System for monitoring a swimming pool to prevent drowning accidents. Google Patents. US Patent 6,133,838, 2000
13. Lu, W., Tan, Y.-P.: A camera-based system for early detection of drowning incidents. In: Proceedings International Conference on Image Processing, IEEE, vol. 3 (2002).
14. Eng, H.-L., Toh, K.-A., Yau, W.-Y., Wang, J.: Dews: A live visual surveil lance system for early drowning detection at pool. IEEE Trans. Circuits Syst. Video Technol. 18, 196–210 (2008)
15. Fei, L., Xueli, W., Chen, D.: Drowning detection based on background subtraction. In: 2009 International Conference on Embedded Software and Systems, pp. 341–343 (2009).
16. Zhang, C., Li, X., Lei, F.: A novel camera-based drowning detection algorithm. In: IGTA (2015).
17. Salehi, N., Keyvanara, M., Monadjemmi, S.A.: An automatic video-based drowning detection system for swimming pools using active contours. Int. J. Image Graph. Signal Process. 8, 1–8 (2016)
18. Prakash, B.D.: Near-drowning early prediction technique using novel equations (neptune) for swimming pools. ArXiv: abs/1805.02530 (2018)
19. Li, K.: Construction method of swimming pool intelligent assisted drown ing detection model based on computer feature pyramid networks. J. Phys. Conf. Ser. 2137, 012065 (2021)
20. Hou, J., Li, B.: Swimming target detection and tracking technology in video image processing. Microprocess. Microsyst. 80, 103535 (2021)
21. Lei, F., Zhu, H., Tang, F., Wang, X.: Drowning behavior detection in swimming pool based on deep learning. Signal Image Video Process. 16, 1683–1690 (2022)
22. Blanco-Filgueira, B., Garcia-Lesta, D., Fernández-Sanjurjo, M., Brea, V. M., López, P.: Deep learning-based multiple object visual tracking on embedded system for IoT and mobile edge computing applications. IEEE IoT J. 6(3), 5423–5431 (2019).
23. Wang, Y., Tang, C., Cai, M., Yin, J., Wang, S., Cheng, L., … Tan, M.: Real-time underwater onboard vision sensing system for robotic gripping. IEEE Trans. Instrum. Meas. 70, 1–11 (2020).
24. Lu, J., Yuan, F., Yang, W., Cheng, E.: An imaging information estimation network for underwater image color restoration. IEEE J. Oceanic Eng. 46(4), 1228–1239 (2021).
25. Ming, Q., Miao, L., Zhou, Z., Dong, Y.: CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–14 (2021).
26. Hu, Q., Hu, S., Liu, S.: BANet: A balance attention network for anchor-free ship detection in SAR images. IEEE Trans. Geosci. Remote Sens. 60, 1–12 (2022).
27. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–742 (2016).
28. Park, H., Noh, J., Ham, B.: Learning memory-guided normality for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14372–14381 (2020)
29. LV, H., Chen, C., Cui, Z., Xu, C., Li, Y., Yang, J.: Learning nor mal dynamics in videos with meta prototype network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021).
30. Liu, W., Luo, W., Lian, D., Gao, S.: Future frame prediction for anomaly detection–a new baseline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6536–6545 (2018).
31. Lu, Y., Kumar, K.M., shahabeddin Nabavi, S., Wang, Y.: Future frame prediction using convolutional vrnn for anomaly detection. In: 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, pp. 1–8 (2019).
32. Liu, Z., Nie, Y., Long, C., Zhang, Q., Li, G.: A hybrid video anomaly detection framework via memory-augmented flow reconstruction and flow-guided frame prediction. In: Proceedings of the IEEE International Conference on Computer Vision (2021)
33. Kam, A.H., Lu, W., Yau, W.-Y.: A Video-Based Drowning Detection System. ECCV (2002).
34. Eng, H.-L., Toh, K.-A., Kam, A.H., et al. An automatic drowning detection surveillance system for challenging outdoor pool environments. In:

Proceedings of the IEEE International Conference on Computer Vision, pp. 532–532 (2003).
35. Salehi, N., Keyvanara, M., Monadjemmi, S.A.: An automatic video-based drowning detection system for swimming pools using active contours. Int. J. Image, Graph. Signal Process. 8(8), 1–8 (2016).
36. Lu, W., Tan, Y.-P.: A vision-based approach to early detection of drowning incidents in swimming pools. IEEE Trans. Circuits Syst. Video Technol. 14(2), 159–178 (2004).
37. Eng, H.-L., Toh, K.-A., Yau, W.-Y., Wang, J.: Dews: A live visual surveil lance system for early drowning detection at pool. IEEE Trans. Circuits Syst. Video Technol. 18(2), 196–210 (2008).
38. Pavithra, P., Nandini, S., Nanthana, A., et al.: Video based drowning detection system. In: 2021 International Conference on Design Innovations for 3Cs Compute Communicate Control. IEEE, pp. 203–206 (2021).
39. Hasan, S., Joy, J., Ahsan, F., Khambaty, H., Agarwal, M., Mounsef, J.: A water behavior dataset for an image-based drowning solution. In: 2021 IEEE Green Energy and Smart Systems Conference (IGESSC), pp. 1–5 (2021).
40. Hu, C., Wu, F., Wu, W., Qiu, W., Lai, S.: Normal learning in videos with attention prototype network. In: Computer Vision and Pattern Recognition (2021).
41. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. J. VLSI Sig. Proc. Syst. Signal, Image Video Technol. 38(1), 35–44 (2004).
42. Vittone, M. (2010). Drowning doesn't look like drowning. Repéré à http://mariovittone.com/2010/05/154
43. Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., NanoCode012, TaoXie, Kwon, Y., Michael, K., Changyu, L., Fang, J., Abhiram, V., Laughing, tkianai, yxNONG, Skalski, P., Hogan, A., Nadar, J., imyhxy, Mammana, L., AlexWang1900, Fati, C., Montes, D., Hajek, J., Diaconu, L., Minh, M.T., Marc, albinxavi, fatih, oleg, wanghaoyang0106: ultralytics/yolov5: V6.0 - YOLOv5n 'Nano' Models, Roboflow Integration, TensorFlow Export, OpenCV DNN Support. https://doi.org/10.5281/zenodo.5563715
44. Rezatofighi, S.H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 658–666 (2019).
45. Ma, N., Zhang, X., Zheng, H.-T., Sun, J.: Shufflenet v2: Practical guide lines for efficient cnn architecture design. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 116–131 (2018).
46. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016).
47. Van der Maaten, L., Hinton, G.: Visualizing data using T-SNE. J. Mach. Learn. Res. 9(11), 2579–2605 (2008).
48. Liu, W., Luo, W., Lian, D., Gao, S.: Future frame prediction for anomaly detection–a new baseline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6536–6545 (2018).
49. Luo, W., Liu, W., Lian, D., Tang, J., Duan, L., Peng, X., Gao, S.: Video anomaly detection with sparse coding inspired deep neural networks. IEEE Trans. Pattern Anal. Mach. Intell. 43(3), 1070–1084 (2019).

How to cite this article: Liu, T., He, X., He, L., Yuan, F.: A video drowning detection device based on underwater computer vision. IET Image Process. 17, 1905–1918 (2023). https://doi.org/10.1049/ipr2.12765