Recognition of Unsafe Onboard Mooring and Unmooring Operation Behavior Based on Improved YOLO-v4 Algorithm

Zhao, Changjiu; Zhang, Wenjun; Chen, Changyuan; Yang, Xue; Yue, Jingwen; Han, Bing

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Article

Recognition of Unsafe Onboard Mooring and Unmooring Operation Behavior Based on Improved YOLO-v4 Algorithm

Changjiu Zhao [1,2], Wenjun Zhang [1,2,*], Changyuan Chen [3], Xue Yang [1,2], Jingwen Yue [1,2] and Bing Han [4,5]

1   Navigation College, Dalian Maritime University, Dalian 116026, China
2   Key Laboratory of Safety & Security Technology for Autonomous Shipping, Dalian University, Dalian 116024, China
3   Department of Civil Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands
4   Shanghai Ship and Shipping Research Institute Co., Ltd., Shanghai 200135, China
5   College of Physics and Electronic Information Engineering, Minjiang University, Minhou County, Fuzhou 350108, China
*   Correspondence: wenjunzhang@dlmu.edu.cn

Abstract: In the maritime industry, unsafe behaviors exhibited by crew members are a significant factor contributing to shipping and occupational accidents. Among these behaviors, unsafe operation of mooring lines is particularly prone to causing severe accidents. Video-based monitoring has been demonstrated as an effective means of detecting these unsafe behaviors in real time and providing early warning to crew members. To this end, this paper presents a dataset comprising videos of unsafe mooring line operations by crew members on the M.V. YuKun. Additionally, we propose an unsafe behavior recognition model based on the improved You Only Look Once (YOLO)-v4 network. Experimental results indicate that the proposed model, when compared to other models such as the original YOLO-v4 and YOLO-v3, demonstrates a significant improvement in recognition speed by approximately 35% while maintaining accuracy. Additionally, it also results in a reduction in computation burden. Furthermore, the proposed model was successfully applied to an actual ship test, which further verifies its effectiveness in recognizing unsafe mooring operation behaviors. Results of the actual ship test highlight that the proposed model's recognition accuracy is on par with that of the original YOLO-v4 network but shows an improvement in processing speed by 50% and a reduction in processing complexity by about 96%. Hence, this work demonstrates that the proposed dataset and improved YOLO-v4 network can effectively detect unsafe mooring operation behaviors and potentially enhance the safety of marine operations.

Keywords: maritime safety; Mobilenet-v3; unsafe behavior; YOLO-v4

1. Introduction

1.1. Background

In world trade, about 90% of freight is transported by ships, and unsafe crew operations have already led to various accidents, including ship collisions, ship grounding, and occupational accidents. The statistics reveal that 80–85% of accidents are directly or indirectly caused by human error or unsafe behavior [1]. Mooring operations are among the most dangerous tasks carried out aboard ships [2]. To date, many fatalities and fatal injuries originating from unsafe mooring practices aboard have been reported, with statistics from the European Harbor Masters' Committee showing that ropes and wires cause 95% of personal injury incidents, and 60% of these injuries happen during mooring operations [3]. According to the Pilbara ports authority, mooring line failures/parted mooring line incidents pose a significant risk to personnel, infrastructure, and operations in the port. Over the past few years, many mooring line incidents have been reported in the Port of Hedland, forming approximately 20% of all marine incidents [4]. If the mooring operation is conducted improperly, there is also a risk of the vessel colliding with other ships or

damaging the shore structure. Therefore, ensuring safe mooring practices is essential for accident prevention.

Currently, the primary way to ensure safety during mooring and unmooring is to enforce the safety procedures, including proper use of personal protective equipment (PPE) during works on deck, clearance of deck prior to operations, and monitoring of the dangerous zones (e.g., snap-back zones and potential pinch points) and deck activities. The monitoring mainly relies on manual supervision by experienced seafarers, who purposely search for hazards that may cause injury, such as trip and trap hazards and mooring lines under tension, and immediately issue warnings to the crew. However, mooring is a dynamic and complicated process. Thus, in practice, the supervisor is often involved with other operations and cannot monitor effectively. Furthermore, the supervisor has human sensory and attention limitations, potentially delaying the warning.

In the maritime industry, despite the considerable amount of research that has been conducted on unsafe behaviors, the focus is primarily on error-producing conditions for unsafe acts and prediction. One limitation of focusing primarily on error-producing conditions and prediction is that it may not be effective in addressing more complex or dynamic safety issues. In some cases, the factors that contribute to accidents and incidents may be difficult to predict or may involve a combination of individual, organizational, and systemic factors that are difficult to identify and address. Another limitation is that a focus on prediction and prevention may not be sufficient to ensure safety in all circumstances. Even with the best efforts to predict and prevent accidents and incidents, there is always a risk that something unexpected could happen. Therefore, it is important to have monitor systems in place to manage and respond to emergencies when they do occur. For example, Celik and Cebi used the Human Factors Analysis and Classification System (HFACS) to model the factors that affect human errors in ship accidents and found that the technical environment plays the most crucial role [5]. Moreover, Xue Yang used the Bayesian network and tree-augmented naive Bayes (TAN) search algorithms to build a prediction model to decide the operational limits of a given operation [6]. Additionally, Ung employed the fuzzy cognitive reliability and error analysis method (CREAM) to analyze accidents caused by personnel factors [7]. Furthermore, Salihoglu and Bal Beşikçi developed the functional resonance analysis (FRAM) model to analyze offshore oil spill accidents [8]. All the methods mentioned above analyzed the results to determine the factors causing the accidents and provided suggestions for improving maritime safety.

To address the above problems, this paper develops a computer vision approach to monitor the behavior of rope handling on the deck. There are a number of reasons why a computer vision approach to monitoring rope handling on the deck may be considered a better approach:

1. Automation: It allows for automated tracking and analysis of human behavior, reducing the time and resources required [9].
2. Objectivity: Analysis is not influenced by human biases, providing accurate and consistent results [10].
3. Non-intrusive: Monitoring can be performed through cameras or other image-capturing devices, minimizing disturbance to workers [11].

Computer vision, also known as machine vision, is a field of artificial intelligence (AI) enabling computers and systems to derive meaningful information from digital images, video, and other visual inputs and then take actions or make suggestions based on that information. In recent years, computer vision-based approaches and deep learning techniques have been widely applied in high-risk industries to improve onsite safety by recognizing and warning against unsafe behaviors. For instance, Fang developed a computer vision scheme to monitor the unsafe behavior of workers passing through the scaffolding during the construction of the engineering structure [12]. Additionally, Lee used a camera to detect pedestrian behaviors for early accident prevention in an advanced driver assistance system (ADAS) [13], and Luo used real-time intelligent video surveillance to prevent people from entering hazardous work areas during the construction of projects

in urban areas [14]. Additionally, Cyganek and Gruszczyński built a system for driver eye recognition and fatigue monitoring, ensuring traffic safety [15].

With the development of video surveillance technology and convolutional neural networks, video monitoring has been widely applied for safety monitoring [16]. For example, Han and Lee utilized computer vision to monitor people's unsafe behavior by developing a framework for unsafe behavior detection and monitoring [17]. This computer vision-based method included two CNN models to monitor whether safety belts were appropriately used while workers were working at heights [18]. An image-skeleton-based method to monitor construction workers' unsafe behaviors has also been proposed [19], and the detection algorithm comprises two modules. The first module predicts the bounding boxes of the driver's right hand and right ear from RGB images. The second module takes the bounding boxes as input and predicts the type of distraction. This method efficiently detects normal driving, using a touchscreen, and talking to a phone. Additionally, Pramanik proposed a video surveillance-based system for improving road safety [20]. The system integrates five algorithms within one surveillance system to detect traffic pre-events. These algorithms combine spatial and temporal features and thus are more informative than solely relying on spatial or temporal features. This method generates an alarm in the control room to allow precautionary measures to be taken and improve road safety.

The video surveillance technology and machine learning approaches applied for automatic unsafe behavior detection and recognition in various industries have the following merits:

1. Noncontact and safer measurement, avoiding damaging the observer and the observed objects/personnel, reducing the risk of injury due to participating in the monitoring process.
2. Working stably for a long time. It is difficult for humans to observe the same object for a long time due to fatigue and distractions, while machine vision can measure, analyze, and recognize an object for a long time.
3. Reduced cost. The machine vision system improves the recognition speed and reduces the labor in monitoring, significantly reducing the overall cost.

Despite the above merits, research on mooring operation monitoring in the maritime industry has not been investigated yet. However, intelligent monitoring will significantly improve crew safety, assist the shipping companies' safety management systems, and reduce the supervisors' stress and help them focus more on their work, reducing the risk caused by fatigue and distraction during mutual supervision. At the same time, monitoring mooring can also provide early warnings in situations with fewer crew members than required on the deck.

You Only Look Once version 4 (YOLO-v4) is a real-time object detection algorithm that uses a single CNN to identify objects within an image; it is known for its high accuracy and fast processing speed. It uses a multi-scale architecture and anchor boxes to improve detection performance and adjust the size and shape of the predicted bounding boxes. Arunabha M. Roy presents an improved version of the YOLO-v4 algorithm for high-performance real-time fine-grain object detection in plant disease detection that addresses obstacles such as dense distribution, irregular morphology, multi-scale object classes, and textural similarity, resulting in better detection accuracy and speed than existing state-of-the-art models [21]. Arunabha M. Roy presents a high-performance real-time fine-grain object detection framework, based on an improved YOLO-v4 algorithm, that addresses several obstacles in plant disease detection, providing an effective and efficient method for detecting different plant diseases in complex scenarios that can be extended to different fruit and crop detection, generic disease detection, and various automated agricultural detection processes [22]. Arunabha M. Roy presents WilDect-YOLO, resulting in superior detection of endangered wildlife under challenging environments with a mean average precision (mAP) value of 96.89%, F1 score of 97.87%, and precision value of 97.18% at a detection rate of 59.20 FPS, outperforming current state-of-the-art models [23].

The current work proposes a real-time object detection framework Dense-YOLO-v4 that includes DenseNet in the backbone and a modified path aggregation network (PANet) to preserve fine-grain localized information, which has been applied to detect different growth stages of mango with a high degree of occultation in a complex orchard scenario and different automated agricultural applications.

1.2. Objective

This paper proposes a method based on an improved YOLO-v4 network for the crew's automatic monitoring of mooring operations to reduce occupational accidents and reduce shipping companies' supervision costs. This method aims to provide early warning when the crew behaves unsafely during mooring to remind the crew to correct their behavior in time, avoiding accidents effectively. The technical contributions of this paper are as follows:

1. Building a dataset containing unsafe mooring line operation behaviors.
2. Proposing a new monitoring method with a low computational cost and fast detection speed that monitors unsafe behavior in real time.

1.3. Organization of the Paper

The remainder of this paper is organized as follows: Section 2 introduces the status of computer vision applications in the field of safety management. Section 3 describes the proposed method, including the collection of datasets and algorithm improvement, while Section 4 presents the experimental results. Finally, Section 5 discusses the results and concludes this work.

2. Computer Vision for Behavior-Based Safety

Computer vision aims to study the ability of machine vision and enable a machine to visually analyze its surrounding environment and features. Machine vision typically evaluates images or videos and automatically extracts, analyzes, and understands the valuable information of a single image or a series of images. Computer vision involves several classification schemes, such as image classification, image localization, object detection, and semantic segmentation.

Object detection is the core and fundamental challenge in computer vision [24]. It has many applications in video surveillance, biomedicine, automatic driving, and other fields. In the past few years, convolutional neural networks (CNNs) have been widely used to solve the problem of image classification. Due to the success of CNNs, researchers used CNNs to a greater extent for video classification. So far, the research in this field can be divided into two types (Shinde et al., 2018):

1. Human behavior recognition using dual-stream CNNs, namely space stream and time stream [25]. The dual-stream CNN is mainly trained on a multi-frame dense optical flow, where the temporal and spatial streams deal with human behavior in the form of dense optical flow and static video frames, respectively.
2. Human motion recognition based on skeleton tracking [26]. This method calculates the video stream using a skeleton tracking algorithm. Then, motion recognition is performed according to the motion between the selected joints and their respective angular velocities [27].

Although the above two methods have achieved remarkable results, considering the recognition speed and computational cost in practical applications, the two algorithms use more information than required because the huge amount of image data requires much space and time for saving and processing [28].

The ship is a confined area with limited equipment allowed for installation aboard and with many applications competing for computational resources in the system. In addition, there is a lot of equipment on the deck. Thus, obtaining a full view of the seafarers' bodies while working is challenging. Therefore, skeleton tracking, which needs more computational resources and has strict requirements for the full body view, is unsuitable for this research. Since installing a camera on the deck and adopting an algorithm that

consumes less computation is feasible, this study adopts the first approach, i.e., dual-stream CNN, for the above reasons and judges whether the action is unsafe through the video frames without requiring optical flow data.

3. Method

This research began by defining the unsafe mooring operation and summarizing seven typical unsafe mooring operations for automatic detection and recognition (Section 3.1). A dataset covering these unsafe behavior images was established by onsite shooting using the installed Huawei D2150-10-siu infrared pinhole camera aboard the ship (Section 3.2.1). Based on the processed dataset, a Mobile-v3-YOLO-v4 network was developed (Section 3.2.2) and trained to establish its weights (Section 3.2.3) and realize the unsafe behavior detection on the video stream (Section 3.2.4). The research steps are illustrated in Figure 1 and described in detail in the following subsections.
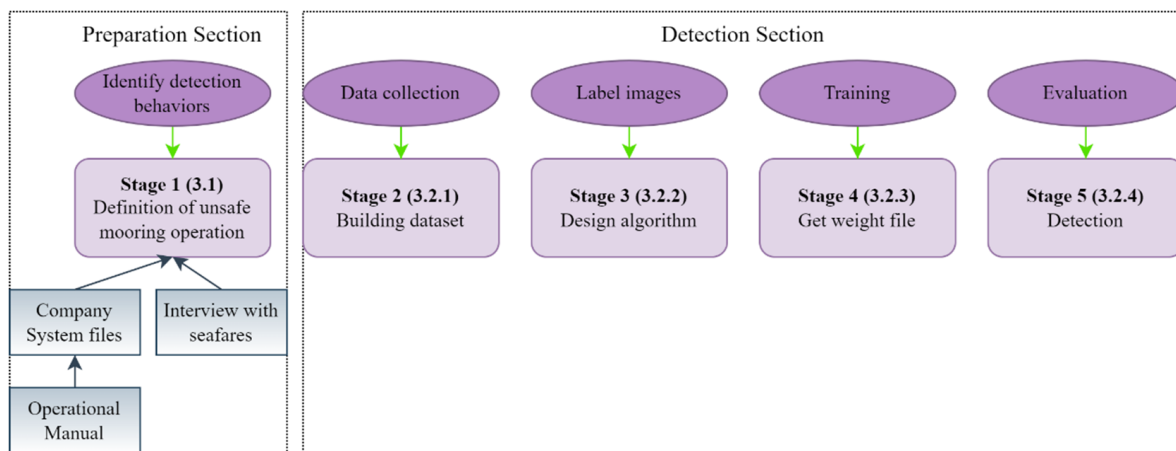

Figure 1. Research progress.

3.1. Subsection

Preparation Section

The International Maritime Organization (IMO) defines an unsafe act as "an error or violation that is committed in the presence of a hazard or potential unsafe condition" [29]. The error is further described as a planned sequence of activities failing to achieve the desired goals and deviating from necessary practices to maintain safety [30]. Two critical factors in defining unsafe behavior can be interpreted from the above definitions: "undesired goals" and "practices deemed necessary". Therefore, in defining the unsafe behavior of mooring operations, the primary sources are safety management system (SMS) documentation from shipping companies, official operational procedures developed based on International Regulations for Preventing Collisions at Sea (COLREGs), and interviews with seafarers.

After collecting and analyzing that information, seven representative unsafe mooring operations on board were identified (Table 1). The summarized unsafe behaviors are also emphasized in Mooring Operations Manual developed by Pacific International Lines company (PIL). For example, walking across the mooring rope and standing in a loop or "bight" of any rope may lead to severe damage to the seafarers (Figure 2). In addition, many seafarers are prone to sit on the bollard or stamp on the taut mooring rope to rest when working on the deck. Although they do not carry out mooring operations while resting, it is considered a hazardous state that has led to severe injuries. Such behavior is prohibited in shipping companies' SMS documentation. Still, unsafe behaviors, as shown in Figure 3, are uncommon in practice but were revealed by the interviews with crews on the M.V. YuKun. An experienced sailor also emphasizes that while holding a rope, it is

essential to remember never to put a hand over the rope because its vibration may damage the hand, and it is hard to pull the hand out when the rope is broken (Figure 4) [31].

Table 1. The source of unsafe behaviors.

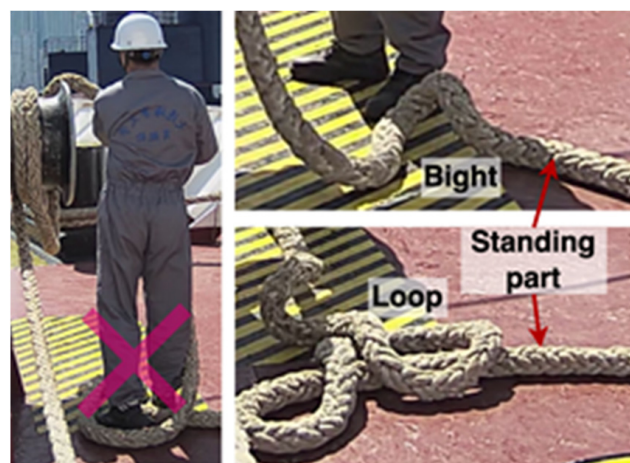| Unsafe Behavior	Source
Sitting on the bollard	Interview and SMS
Stamping on the cable	Interview and SMS
Stamping on the bollard	Interview and SMS
Across the cable	Operation manual and SMS
Across the taut cable	Operation manual and SMS
Hand over the cable	Interview
Standing in the cable	Operation manual and SMS

Figure 2. Unsafe behavior of standing in a loop or "bight" of any rope.

Figure 3. Unsafe behavior of having a rest on board.

Figure 4. A crew member utilizing a mooring rope.

It is worth noting that current research focuses only on the seven representative unsafe behaviors presented in Table 1, based on safety management documentation, relevant reports, and interviews. Moreover, utilizing these unsafe behaviors in the automatic detection model is limited by the camera's detection ability. In practice, other less frequent unsafe behaviors of the mooring operation also exist; they require further research but are not covered here to keep this paper's length reasonable.

3.2. Detection Section

3.2.1. Building Dataset

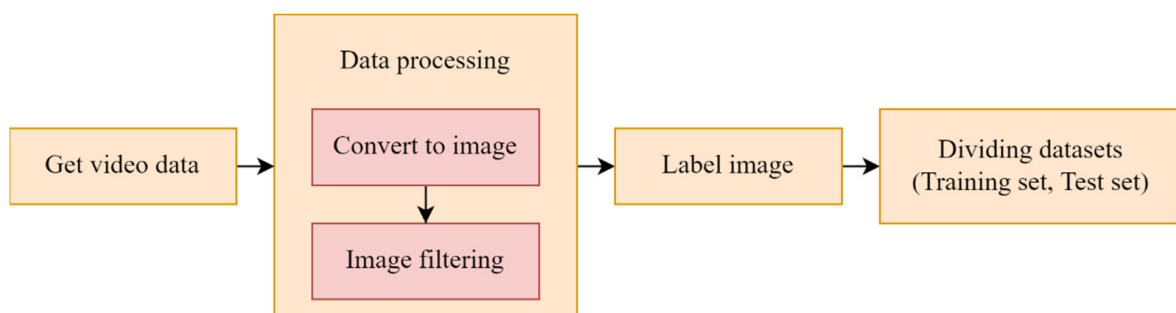Figure 5 illustrates the dataset establishment process.

Figure 5. The dataset establishment process.

(a)   Obtaining video data

The video stream data of seafarers conducting mooring operations on the M.V. YuKun (Figure 6) of the Dalian Maritime University were collected using a Huawei D2150-10-siu infrared pinhole camera. The vessel's main features are reported in Table 2.

Figure 6. Research and training ship "YuKun".

Table 2. Main particulars of M.V. YuKun.

Parameters	Description	Value
LOA	Length overall (m)	116
LPP	Length between perpendiculars (m)	105
B	Breath molded (m)	18
D	Depth to main deck (m)	8.35
g	Gross tonnage (t)	6106
Vref	Design speed (kn)	16.9
T	Design draft (m)	5.4

The video recordings involve the boatswain and sailors conducting mooring operations, with both safe and unsafe behaviors. The camera was set up on the deck next to the bridge to enhance coverage of the whole deck. The camera's tripod height was 1.5 m, and its downward-looking angle was 15 degrees to the deck to avoid the reflected light from the sea and the vertical light from the sun influencing the recording. The camera arrangement is depicted in Figure 7.
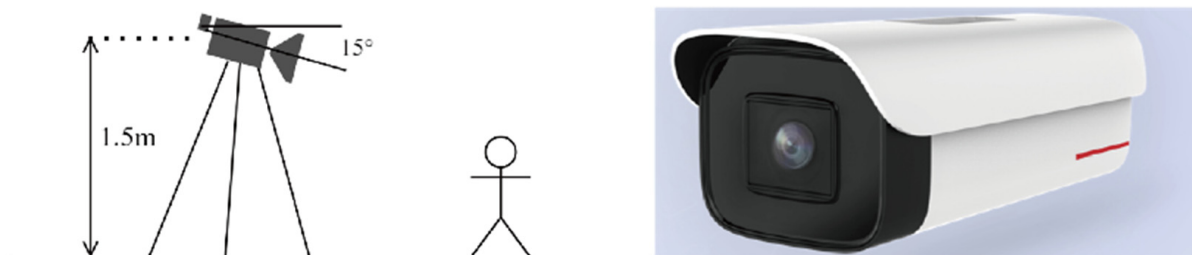

Figure 7. Schematic diagram of the video recording setup.

The shooting area covers the entire deck, including the mooring line windlass, each fairway, and the bollard. The captured video size was 1920 × 1080 pixels, and the frame rate was 60 frames per second (FPS). Images with unsafe behaviors were extracted from video data every ten seconds. In total, 3000 images with unsafe behaviors were selected to comprise the dataset.

(b)   Data processing

After the video data were collected, the active video was stored for approximately one hour, including the seven representative types of unsafe mooring line operations. Then

the video was decomposed into a sequence of pictures frame by frame, and the pictures were filtered; then, 3300 RGB images containing obvious unsafe behaviors were selected for further processing. The seven types of unsafe behaviors are presented in Table 3.

Table 3. Label names and numbers of the dataset.

Label Name	Label Number
Sitting on the bollard	885
Stamping on the cable	824
Stamping on the bollard	432
Across the cable	426
Across the taut cable	425
Hand over the cable	439
Standing in the cable	453

(c)   Label image

The dataset was annotated using the open-source script *labeling* from GitHub (https://github.com/tzutalin/labelImg (accessed on 25 July 2022), and images with unsafe behaviors were manually labeled. After labeling, an XML file was generated containing the image name corresponding to the labeled file, the image size, the name of unsafe behavior, and the coordinate information of the labeled actions. An example of an XML file is presented in Figure 8.
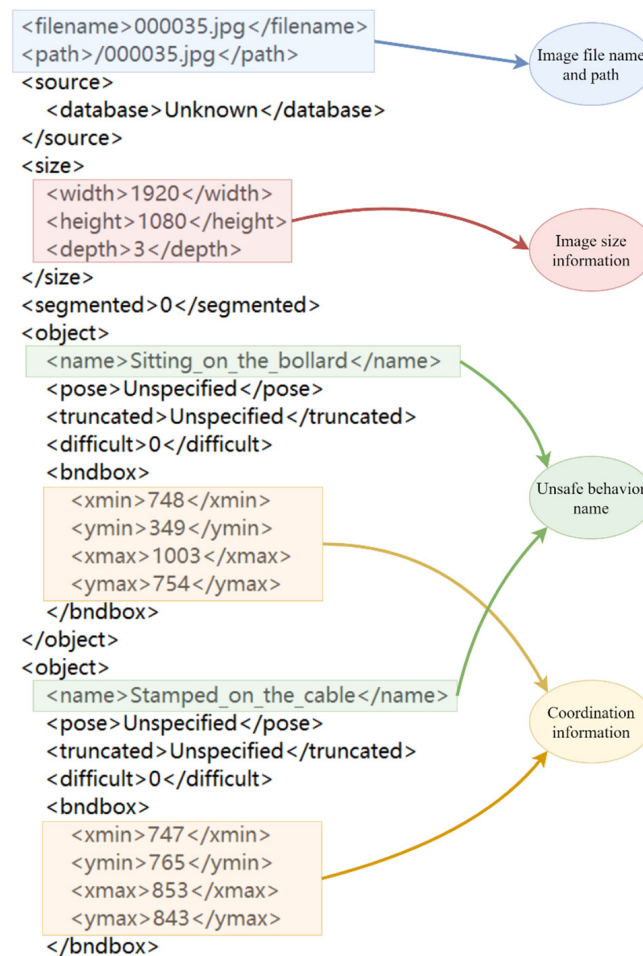

Figure 8. Example of an XML file.

(d)　Dividing dataset

The experiment involved 4213 labeled samples generated from the 3300 images; the 3000 images covering the seven types of unsafe behaviors were used as a training dataset, and the remaining 300 pictures were used for testing. The testing images contained all seven unsafe behaviors. Based on the principle that the training dataset should be enlarged as much as possible to ensure detection and training quality, we selected the 3000 images as the training dataset rather than the 300 images.

3.2.2. Design Algorithm (Improved YOLO-v4 Models)

We utilized the YOLO-v4 convolutional neural network (CNN) for the detection network, but we replaced CSPDarknet53 in the original YOLO-v4 with Mobilenet-v3 as the backbone network for faster detection and less resource consumption. This subsection briefly introduces YOLO-v4 and Mobilenet-v3 and explains the improved YOLO-v4.

(a)　YOLO-v4

The YOLO-v4 network is an improvement of YOLO-v1 [32], YOLO-v2 (Yolo-9000) [33], and YOLO-v3 [34]. All YOLO algorithms adopt a one-stage strategy, i.e., the CNN directly predicts the category and location of the various targets. Compared with the two-stage algorithms, such as faster R-CNN, the detection speed is greatly improved. The principle of the one-stage algorithms is the transformation of the detection problem into a regression problem [35]. Precisely, if the image is divided into an n × n mesh, the target is predicted if it is located in the grid, while the target position and confidence degree are predicted through a regression method. Figure 9 illustrates the YOLO-v4 architecture.
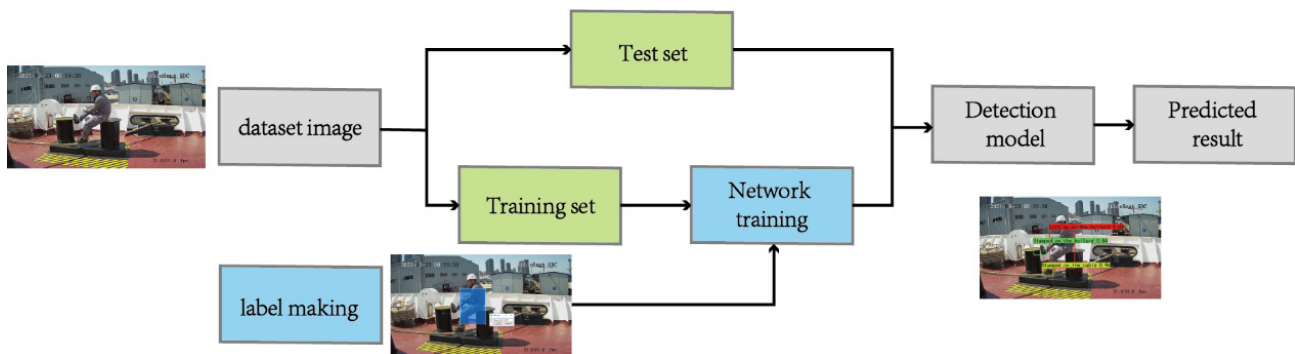

Figure 9. Object detection flowchart using YOLO-v4.

(b)　Mobilenet-V3

Mobilenet-v3 [36] is an extension of Mobilenet-v1 [37] and Mobilenet-v2 [38], enhancing the recognition accuracy compared to the previous versions. Mobilenet-v3 has two versions: Mobilenet-v3 large and small. This paper utilizes Mobilenet-v3 large, which, compared with Mobilenet-v2, achieves the same accuracy on the COCO dataset but is 25% faster. Mobilenet-v3 has the following characteristics:

1.　Squeeze and Excitation (SE) structure

The SE structure is added to the bottleneck structure and placed after the depth-wise filtering. Since the SE structure has a certain amount of processing complexity, the channel of the expansion layer is reduced to $\frac{1}{4}$ of the original SE structure [36]. This way, the accuracy is improved without further increasing the execution time.

2.　Simple tail structure

In Mobilenet-v2, before the AVG pooling, there is a 1 × 1 convolution layer. The structure of Mobilenet-v3 is different after the AVG pooling. The AVG pooling layer reduces the size of the feature graph from 7 × 7 to 1 × 1 to reduce the computation burden by

$7 \times 7 = 49$ times. Moreover, to further reduce computational complexity, the Mobilenet-v3 discards the 3×3 and 1×1 convolutions of the previous spindle convolution.

3. Modified number of channels

The number of convolution channels in the head is reduced, as Mobilenet-v2 uses $32 \times 3 \times 3$, while Mobilenet-v3 reduced it to $16 \times 3 \times 3$. The speed is reduced by 3 ms, but the accuracy is the same.

4. The change of nonlinear transformation

The h-swish replaced swish with the related formulas presented below [39]. The swish function has a lower bound, smoothness, non-monotonicity, and no upper bound and is superior to ReLU in the deep model. However, since the sigmoid function is computationally complex, Mobilenet-v3 approximates the swish function with h-swish.

$$\text{swish} x = x \cdot sigmoid(x); \; sigmoid = (1 + exp(-x))^(-1)) \tag{1}$$

$$\text{h} - \text{swish}[x] = x \frac{\text{ReLU6}(x + 3)}{6} \tag{2}$$

(c) Improved YOLO-v4

The improved YOLO-v4 is illustrated in Figure 10, and the entire network structure comprises three parts:

1. Backbone feature extraction network, utilizing Mobilenet-v3.
2. Strengthening the feature extraction network using SPP and PANet.
3. YOLO head, a prediction network that employs the obtained features for prediction.

The architecture of YOLO-v4, as illustrated in Figure 10, is composed of three distinct parts: (a) the backbone feature extraction network, (b) the enhanced feature extraction network, and (c) the prediction network. The backbone feature extraction network conducts initial feature extraction and generates three preliminary feature layers. The enhanced feature extraction network subsequently refines these layers by fusing them together in order to extract more effective features, resulting in three additional feature layers. The prediction network then utilizes these more effective layers to produce the final predictions.

While both parts (a) and (b) can be modified with ease, the third part (c) is more rigid and cannot be modified easily. However, this is not a significant concern as it primarily comprises a combination of 3×3 and 1×1 convolutions. To further enhance the feature extraction process, the Mobilenet series network is employed for classification, with its central component utilized specifically for feature extraction. As a result, in YOLO-v4, CSPDarknet53 is replaced with Mobilenet, allowing for the modification of the exact shape of the three preliminary feature layers as illustrated in Figure 11.

3.2.3. Creating the Parameter Weight File

In this step, the designed algorithm trains the detection model using the training set presented above to create a parameter weight file. Table 4 introduces the experimental setup.
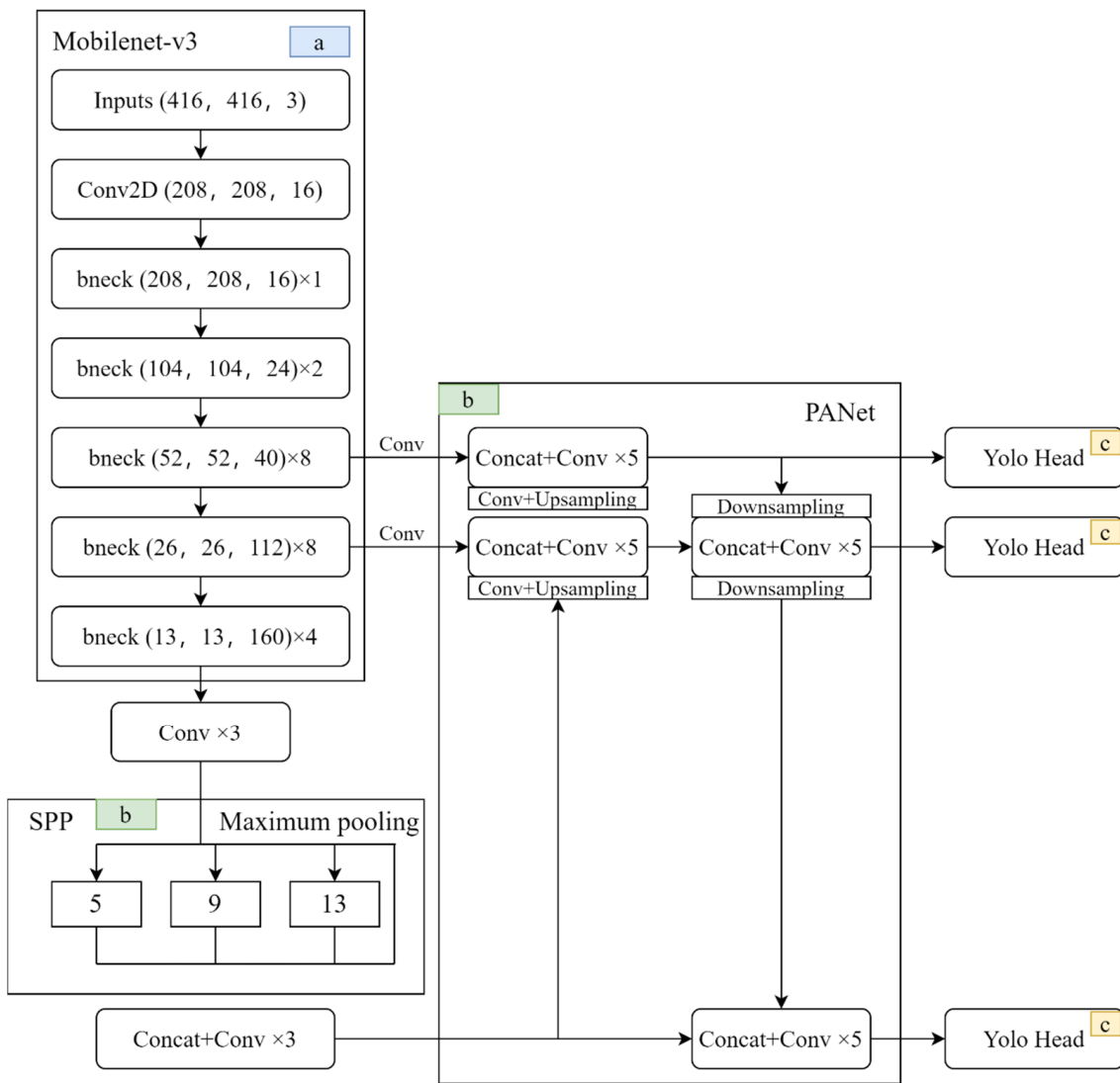
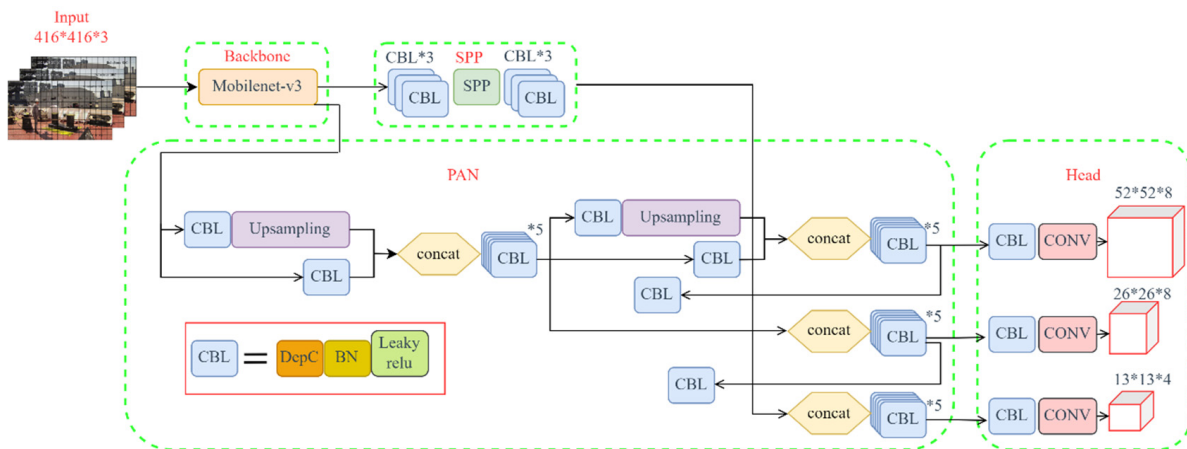**Figure 10.** Structure of the improved YOLO-v4 (**a**) Mobilenet-v3, (**b**) PANet, SPP Maximum pooling, (**c**) Yolo Head.

Figure 11. The improved YOLO-v4 process.

Table 4. The experimental setup.

Configuration	Parameter
Collection tools	Huawei D2150-10-SIU camera
CPU	Intel(R) Xeon(R) CPU E5-2687W v4
GPU	Quadro M4000
Operating system	Window 10
Accelerated environment	CUDA 10.1 CUDNN 8.0.4
Development environment	Pycharm 2020.3.2
Library	Opencv 4.5.1.48

In this research, the training and testing performance of the improved model depends on the Mobilenet backbone. The testing images and videos have the same resolution as the training input. Given that the results obtained are compared against YOLO-v3 and YOLO-v4, and YOLO-v4 comprises five down-sampling stages, the input image must be a multiple of 32. Therefore, the input image size was 416×416 during all the experiments. Additionally, parameters such as momentum, initial learning rate, and weight decay are the original parameters of the YOLO-v4 network, while the remaining parameters during training are reported in Table 5. An example of the labeled image is presented in Figure 12, and the training structure is depicted in Figure 13.

Table 5. Training parameters.

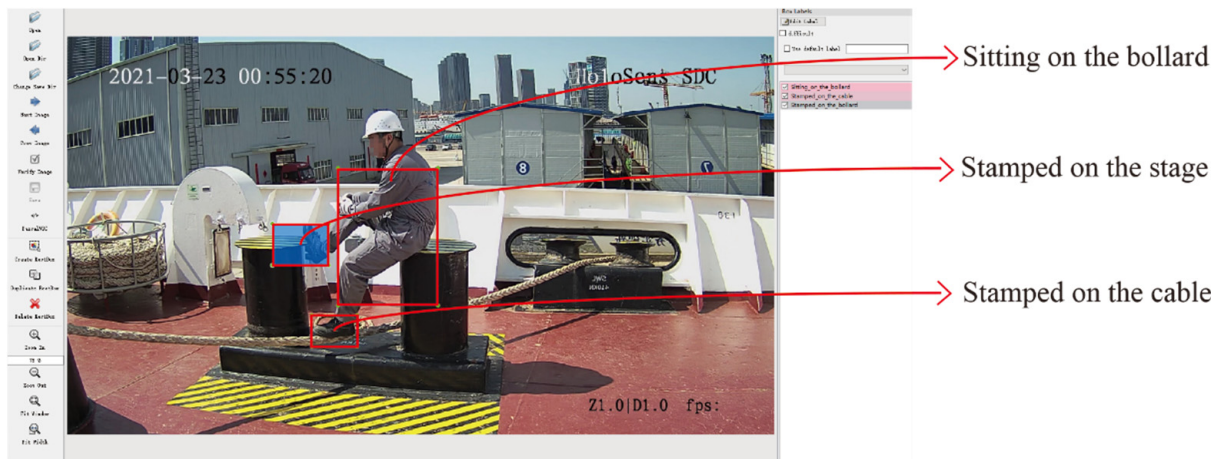| Parameters	Value
Initial generation	0
Frozen generation	50
Initial learning rate	0.0001
Batch size	16
Iteration	100

Figure 12. Example of a labeled image.

In this experiment, the training involved 100 epochs where the weight file was constantly updated, providing one file per epoch. These weights record the eigenvalues in the training set, which are used to predict the test set. During training, the weight file's loss value decreases but will fluctuate after reducing to a certain extent. We chose the weight file providing the lowest loss value, as it achieves the best training effect. The corresponding results are reported the Table 6.
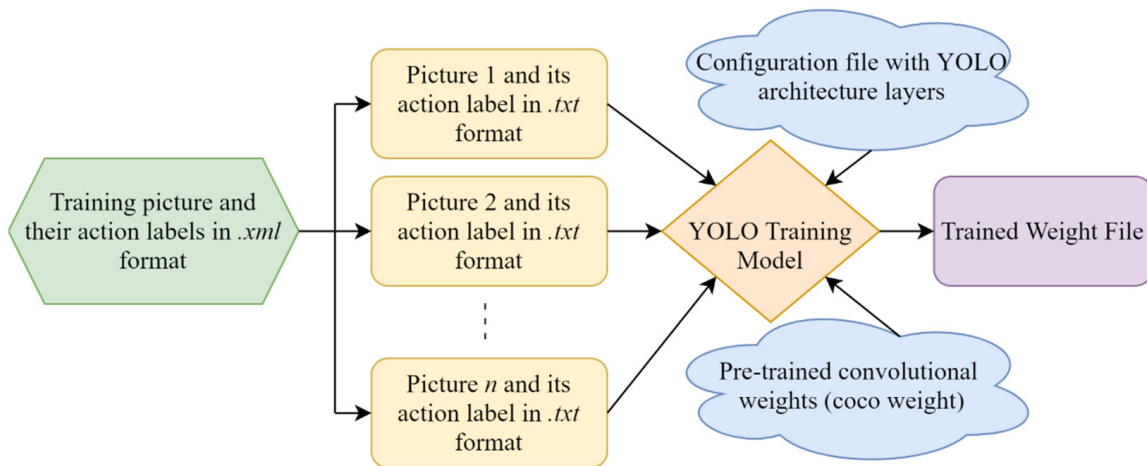
Figure 13. Training structure of the improved YOLO-v4.

Table 6. The employed weight file during detection.

Trained Method	Chosen Epoch	Val Loss	Total Loss
YOLO-v3	98	3.2516	3.3527
Mobilenet-v1-YOLO-v4	94	2.5634	1.4056
Mobilenet-v2-YOLO-v4	96	2.2842	1.5314
Mobilenet-v3-YOLO-v4	91	2.2842	1.5314

3.2.4. Detection

After the weight file was obtained, the test set and the collected video data were used to test the training's effectiveness and verify the algorithm's performance. The corresponding monitoring results are illustrated in Figure 14, highlighting that the algorithm can identify unsafe behaviors of various mooring line operations, such as crossing the cable, stamping on the cable, and sitting on the bollard. This is important as camera monitoring can provide an early warning to effectively avoid occupational and ship accidents.
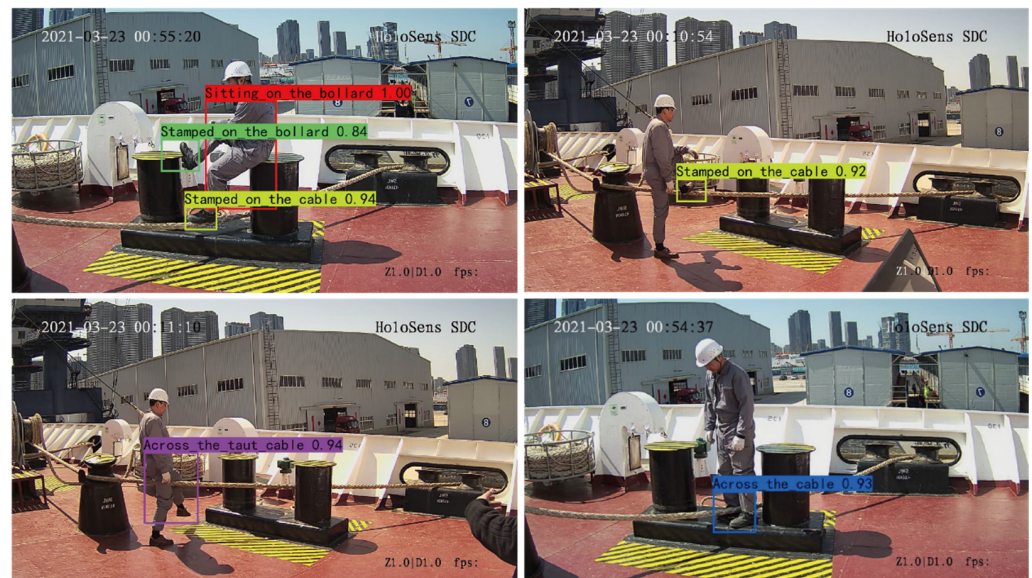

Figure 14. The detection result example.

4. Analysis of Results

This section evaluates the training effect to ensure the accuracy of the detection algorithm. The algorithm's performance is also analyzed by evaluating its processing burden, and the detection accuracy and speed are compared with other mainstream algorithms to verify the algorithm's feasibility in practical applications.

4.1. Training Result Analysis

The loss value of YOLO-v4 *Loss*(*object*) comprises the boundary box regression loss *Loss*(*coord*), confidence loss *Loss*(*conf*), and classification loss *Loss*(*cls*), formulated as follows [29]:

$$Loss(object) = Loss(coord) + Loss(conf) + Loss(cls) \tag{3}$$

$$Loss(coord) = \lambda_{coord} \sum_{i=0}^{k*k} \sum_{j=0}^{M} I_{ij}^{obj} (2 - w_i * h_i)[L_{CIOU}] \tag{4}$$

$$
\begin{aligned}
Loss(conf) = &- \sum_{ij}^{K*K} \sum_{j=0}^{M} I_{ij}^{obj} \big[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i)\big] \\
&- \lambda_{noobj} \sum_{i=0}^{K*K} \times \sum_{j=0}^{M} I_{ij}^{noobj} \big[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i)\big]
\end{aligned}
\tag{5}
$$

$$Loss(cls) = - \sum_{i=0}^{K*K} I_{ij}^{obj} \sum_{c \in classes} [\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c))] \tag{6}$$

where I represents the i-th unit of the feature graph, K is the size of the network, j is the prediction box responsible for the j-th box, and W and H are the width and height of the ground truth bounding box, respectively. C_i represents the grid's confidence, I_{ij}^{obj} denotes whether there is an object in the i-th cell, and L_{CIOU} is the bounding box regression loss function. In each image, several mesh cells do not contain any object, pushing the "confidence" score of these cells to 0 and typically exceeding the gradient of the cells containing the object. This scenario leads to model instability and divergence during an early training stage. To solve this problem, Redmon increased *Loss(coord)* and reduced *Loss(conf)* for boxes that do not contain objects by setting weights $\lambda_{coord} = 5$ and $\lambda_{noobj} = 0.5$ (Redmon et al., 2015).

The loss curve of the five network training processes is illustrated in Figure 15, revealing that the loss value drops rapidly for up to 20 iterations, and the computational burden increases as the iterations increase. Then, the loss slowly oscillates downward and tends to stabilize, and the final loss value is stable at about 1. However, due to the complexity of the unsafe behavior action, the loss value cannot continue to decrease. The loss values of these five networks gradually decrease and tend to stabilize, the model converges gradually, and the training results achieve the expected results. However, the results of training the three versions of Mobilenet as YOLO-v4's backbone are slightly inferior to those of the original YOLO-v4 but better than those of YOLO-v3, and the loss value stabilizes at about 1.

4.2. Algorithm Performance Evaluation

4.2.1. Computational Cost

The unsafe mooring line monitoring system utilizes the ship's computer, with its CPU and GPU not being powerful and having to run other navigation-related software. Therefore, to ensure the stability of the monitoring operation, the network must be lightweight. Table 7 reports the calculation consumption of five mainstream network structures.
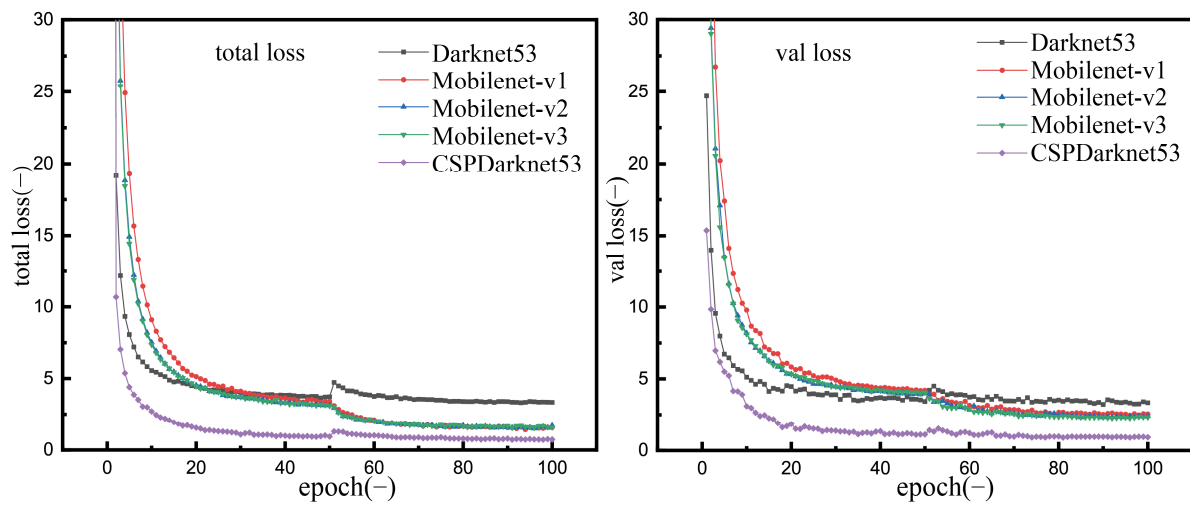
Figure 15. The loss curve of the five networks.

Table 7. Comparison of the calculation consumption of the five network structures.

Net Index	Darknet53	Mobilenet-v1	Mobilenet-v2	Mobilenet-v3	CSPDarknet53
Total params	61,545,274	4,231,976	3,504,872	5,483,032	64,363,101
Total memory	127.74 MB	57.72 MB	74.25 MB	51.23 MB	176.57 MB
Total Madd	18.97 G	1.16 G	627.69 M	450.2 M	17.48 G
Total FLOPS	9.5 G	583.87 M	320.24 M	229.22 M	8.75 G

The *total params* index represents the number of parameters in the whole network; Mobilenet-v2 has the fewest parameters (3,504,872), and Mobilenet-v3 is the second smallest. However, this does not represent the network's computational burden, as it compares the network's size and has a minor impact on the computational burden. *Total memory* refers to the network's total memory, and Mobilenet-v3 requires the least (51.23 MB). *Total Madd* denotes the multiply–accumulate operation, which in computing and especially in digital signal processing is a common step that computes the product of two numbers and adds it to an accumulator. The hardware unit that performs the operation is a multiplier–accumulator (MAC, or MAC unit). The operation itself is also called a MAC or a MAC operation. The *total FLOPS* (floating point operations per second) index refers to the number of floating point operations, realized as the number of computations, and is typically used to measure the complexity of an algorithm or a model.

Figure 16 compares the above parameters, revealing that using the Mobilenet series as the backbone involves far fewer calculations than Darknet. Of the three Mobilenet versions used as a backbone, version 3 is the lightest. Regarding the calculations of the five competitor algorithms, Mobilenet-v3 requires 96% less than Darknet53 and CSPDarknet53, 44% less than Mobilenet-v1, and 12% less than Mobilenet-v2. The calculations of Mobilenet-v3 are far less than those of the other algorithms, playing a significant role in practical applications on board the vessel and thus being more suitable for unsafe behavior monitoring on the ship. However, the applicability of Mobilenet-v3 depends on the accuracy and recognition speed.
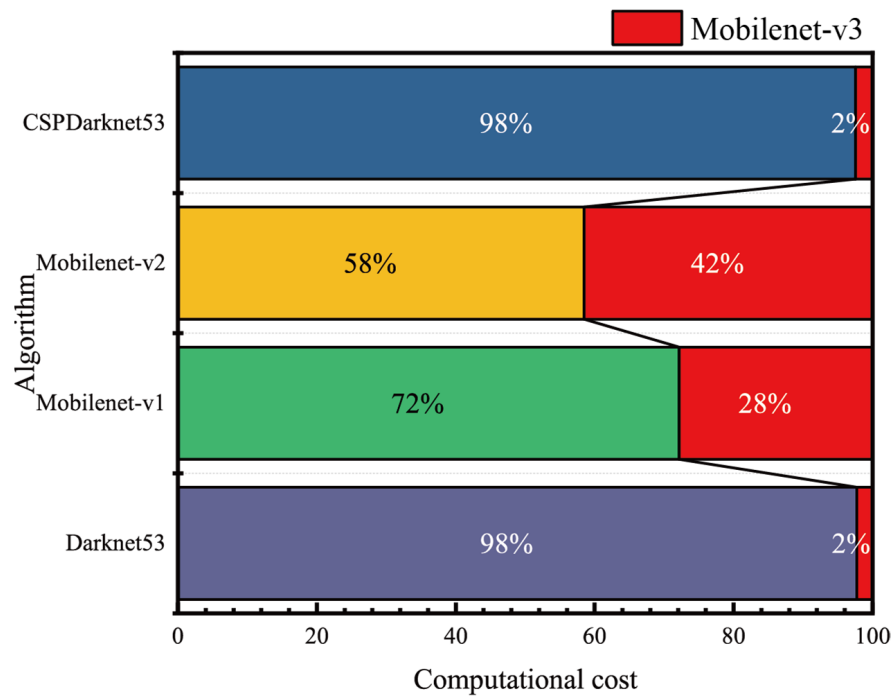
Figure 16. Comparison of the computational cost.

4.2.2. Precision

Evaluating a detection algorithm primarily depends on two criteria: (1) whether the object category in the frame is correctly predicted or not and (2) the coincidence degree of the predicted and the manual label boxes. The quantitative indicators are mean average precision (mAP) and intersection over union (IoU).

The intersection-over-union ratio is the ratio between the intersection and union of the predicted bounding box and the reference bounding box. This statistic, also known as the Jaccard index, was first proposed in [40] in the early 20th century. IoU measures the coincidence degree of the predicted object frames and real frames. The schematic diagram of IoU is illustrated in Figure 17, where the green box represents the actual box and the red box is the prediction box. When the relationship between the two boxes must be judged, IoU judges the coincidence degree of the two boxes. So, the IoU value is obtained by calculating the ratio of intersection to union formulated in Equation (7):

$$IoU = \frac{\text{Intersection of BBox}_{\text{pred}} \text{ and BBox}_{\text{gt}}}{\text{Union of BBox}_{\text{pred}} \text{ and BBox}_{\text{gt}}} \tag{7}$$

$$mAP = \frac{\text{Number of IoUs greater than 0.5}}{\text{Total number of IoUs}} \tag{8}$$

where $\text{BBox}_{\text{pred}}$ and BBox_{gt} represent the boundary boxes of the classification and ground truth, respectively. The intersection over union (IoU) calculates the intersection of the two bounding boxes divided by their union.
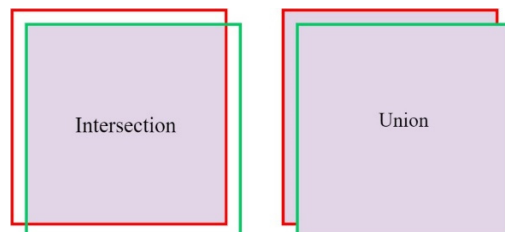

Figure 17. The schematic diagram of IoU.

Mean average precision (mAP) measures the average detection precision (AP) of each category and then calculates the average AP value of all categories (Equation (8)). The mAP represents a comprehensive evaluation metric of the average precision of the detected target. AP refers to the area under the curve using the combination of different precision and recall points.

The target detection algorithm requires a performance evaluation index to evaluate the related model, with neural networks typically employing the accuracy, recall rate, *F1* score, and average accuracy (AP) as evaluation indexes. In order to calculate the calculations of neural networks, the model's FLOPS are also calculated. The precision, recall, and *F1* score metrics are defined in Equations (9)–(11), where *TP* is the actual positive number of detections, i.e., correctly identified unsafe behavior actions; *FN* denotes the number of occurrences falsely unidentified; *TN* is the number of true negative instances correctly identified as background; and *FP* is the number of false positives, where the samples are incorrectly identified as unsafe behaviors. However, these metrics are designed for binary classification and tend to overestimate the error rate classification. Thus, this paper relies on the AP curve for a more comprehensive assessment.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

$$P = \int_0^1 Precision * Recall \, dr \tag{12}$$

Equation (12) presents the calculation formula of AP, where *r* is the integral variable between 0 and 1 and AP denotes the area under the precision–recall (PR) curve. Figure 18 compares the AP diagrams resulting after the detection of several unsafe behaviors. The formula calculating the confidence metric is as follows:

$$Pr(Class_i|Object) * Pr(Object) * IOU_{pred}^{truth} = Pr(Class_i) * IOU_{pred}^{truth}, \ Pr(Object) \in \{0,1\} \tag{13}$$

The PR curves of the YOLO-v4 algorithm using the Mobilenet-v3 as backbone are compared against the other competitor algorithms. Four different unsafe actions are selected for comparison, revealing that the closed area of the improved YOLO-v4's PR curve is the same as that for the other algorithms. This means that using the Mobilenet-v3 as the backbone does not significantly reduce the recognition accuracy.

Figure 19 compares the overall accuracy rate of mAP obtained by the five backbones. It can be seen that, overall, the accuracy rates of YOLO-v3 using Darknet53 as the backbone and YOLO-v4 using CSPDarknet53 as the backbone are slightly higher than those of employing YOLO-v4 and Mobilenet as the backbone. However, the difference is slight, only 0.16%, because the Mobilenet backbone affords lightweight processing, slightly decreasing the accuracy rate.
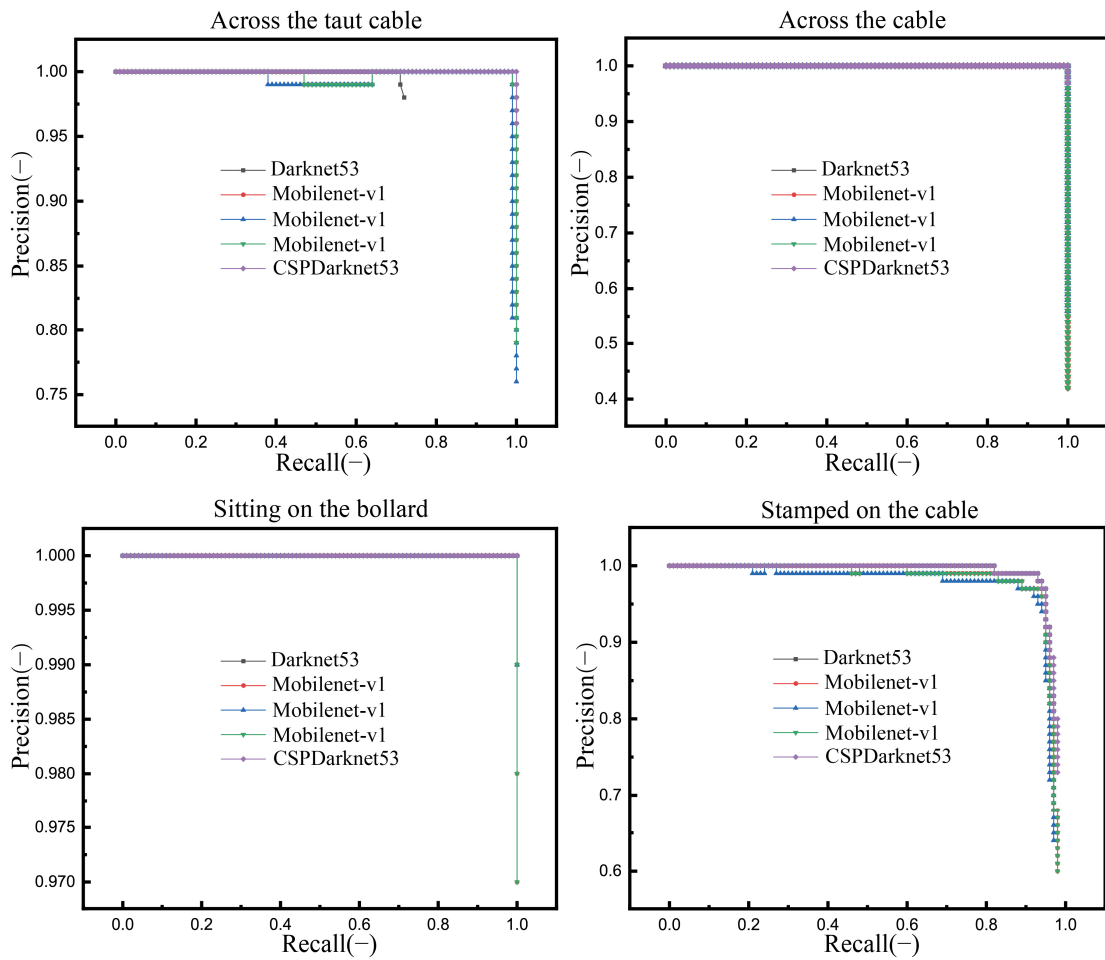
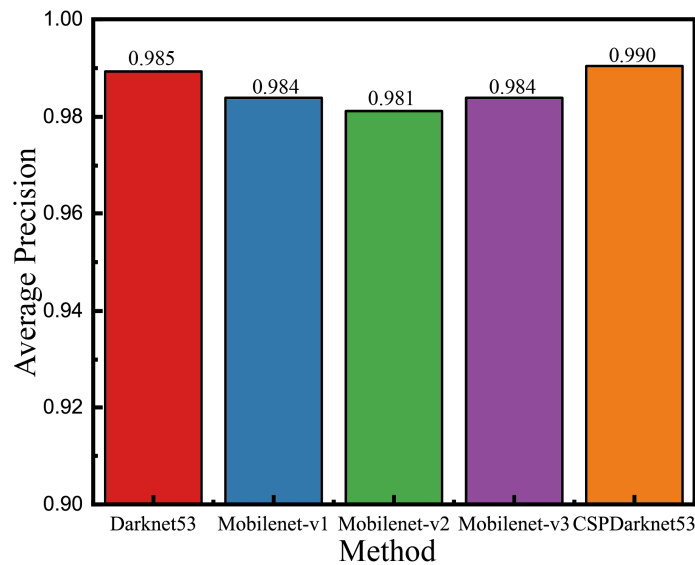**Figure 18.** Comparison of AP diagrams after several unsafe behaviors.

Figure 19. Comparison of the overall accuracy rate.

Figure 20 compares the accuracy rates of several selected unsafe actions with the algorithm's advantages and disadvantages determined by comparing the accuracy rates of individual actions. The corresponding results indicate that, for example, the accuracy of the Mobilenet series algorithm for actions such as "across the taut mooring line" and

"sitting on the bollard" is not much different from that of the CSPDarknet53 and Darknet53 algorithms. At the same time, the effect is slightly worse for some actions of "stamping on the bollard" recognition. However, this comparison reveals that the accuracy of Mobilenet-v3 is improved to a certain extent compared with Mobilenet-v1 and v2.
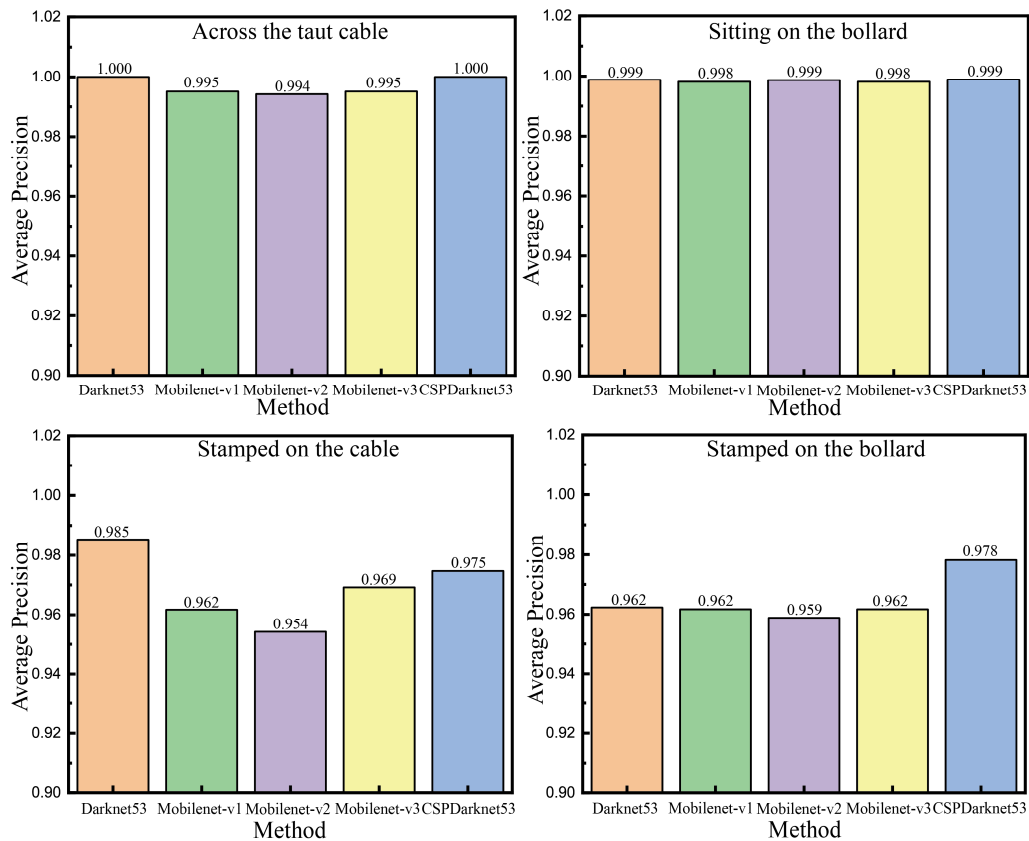

Figure 20. Comparison of the accuracy rates of several selected unsafe actions.

4.2.3. Speed Comparison

In the actual scene, it is necessary to provide an early warning of the unsafe behavior of the crew, so the monitoring system must achieve real-time monitoring and send an early warning. Thus, comparing the recognition speed among all competitor methods is necessary. There are two indexes for speed comparison: tact time and FPS for video processing. Tact time is the time the neural network requires to process a single picture, while FPS is the number of frames per second transmitted, which is the number of animated pictures or videos. FPS measures the amount of information used to save and display dynamic video. The higher the FPS value, the smoother the action displayed.

Figure 21 compares five algorithms' single image processing speed, highlighting that YOLO-v4 based on Mobilenet is faster than the original YOLO-v3 and YOLO-v4. However, the processing speed of v3 is slower than that of v1 and v2, which is inconsistent with the previous algorithm analysis. Therefore, we still use the same video to detect the processing speed of the Mobilenet backbone in video processing. The FPS comparison chart of CSPDarknet53, Darknet53, and Mobilenet-v3 is depicted in Figures 22 and 23, revealing that the recognition speed of Mobilenet-v3 is significantly higher than that of CSPDarknet53 and Darknet53. Additionally, the FPS value of Mobilenet-v3 is about 12, while CSPDarknet53 and Darknet53 afford 8.5 FPS. Figure 22 highlights that Mobilenet-v3 is better than v1 and v2 considering average processing speed and video frame rate stability.
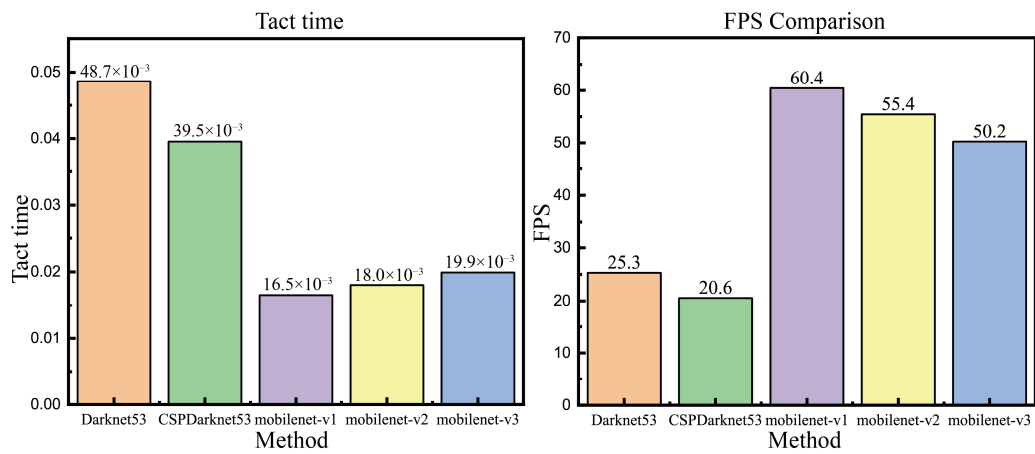
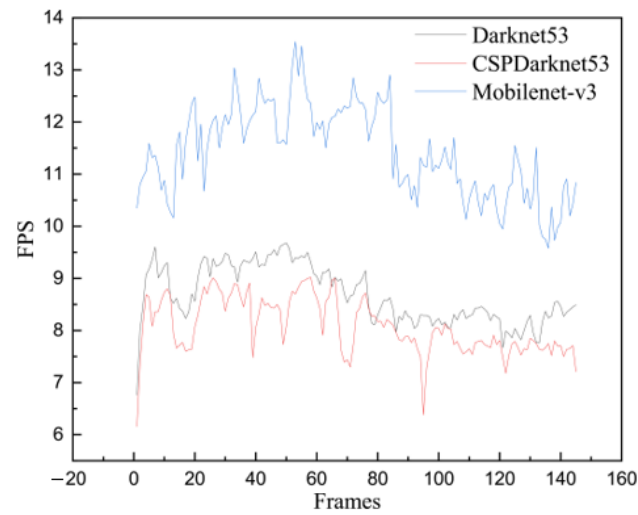**Figure 21.** Comparison of the single image processing speed of five algorithms.

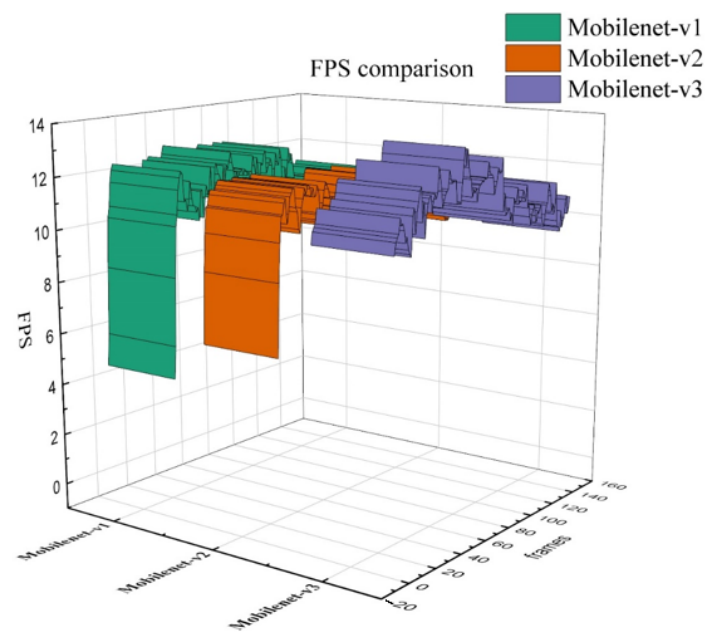**Figure 22.** The FPS comparison of SCPDarknet53, Darknet53, and Mobilenet-v3.

Figure 23. FPS comparison of Mobilenet-v1, v2, and v3.

5. Discussion and Conclusions

In order to monitor the unsafe behavior of seamen's mooring operation and give early warning, this paper proposes an improved YOLO-v4 network that quickly identifies unsafe actions with less computational cost. At the same time, a dataset containing seven representative unsafe behaviors of mooring operation is established to train the model. This study primarily focuses on improving the monitoring speed and decreasing the algorithm's computational burden to afford real-time monitoring of the mooring operation and provide early warnings to reduce occupational and ship accidents. Specifically, we improve YOLO-v4 by replacing the standard backbone network to improve recognition speed and reduce the calculations so that the algorithm can be applied to practical scenarios. We demonstrate that the recognition effect is feasible through the actual tests aboard a vessel. Moreover, the unsafe behavior dataset of mooring line operation provides a basis for other researchers to improve the monitoring of the seamen's mooring operation.

The study conducted a comprehensive evaluation of the improved YOLO-v4 algorithm in terms of recognition speed, precision, and computational cost in comparison to other competitor methods. The results indicated that the implementation of the Mobilenet backbone in the YOLO-v4 algorithm improved the recognition speed, whilst also reducing the calculations required for the algorithm to run. Furthermore, the use of the Mobilenet-v3 backbone was found to provide a significant increase in frames per second in comparison to the CSPDarknet53 and Darknet53 backbones, making it a suitable option for real-time monitoring and early warning in the context of seamen's mooring operations. Additionally, it was noted that the precision of the improved YOLO-v4 algorithm was maintained at a comparable level to that of other algorithms.

5.1. Contribution of This Work

5.1.1. Theoretical Implication

At the technical level, the paper's innovations are as follows:

1. An unsafe behavior dataset of mooring operations is produced by employing multi-party information fusion. The dataset contains seven representative types of unsafe behavior to detect unsafe behavior of mooring operations that potentially lead to navigation and occupational accidents. The proposed dataset applies video surveillance in ship safety and lays the foundation for research on unsafe behavior of mooring operations based on video monitoring.

2. This work improves the YOLO-v4 algorithm. YOLO-v4 is a competent target detection algorithm widely used in various fields. However, real-time monitoring and early warning are mandatory in the unsafe behavior monitoring of crew cable operations. Since CSPDarknet53 in the original YOLO-v4 algorithm is quite complex and consumes many resources, the original YOLO-v4 is unsuitable for monitoring mooring operations in practical scenarios. To solve this challenge, this paper replaces the original backbone CSPDarknet53 with Mobilenet-v3, increasing the processing speed by 50% and reducing the calculations by about 96%. Hence, the proposed improvement makes the algorithm more suitable for practical use on ships.

5.1.2. Practical Implication

This paper applies computer vision technology to the monitoring of unsafe crew behavior. Currently, the research on unsafe crew behavior is mainly limited to theoretical research, and the critical points of unsafe behavior were discussed through theoretical analysis. However, implementing the research conclusions in the actual work did not provide any reasonable warning. To date, the most common way is relying on people to remind each other.

Spurred by the current research gap, this paper develops an unsafe crew behavior early warning method based on video monitoring, effectively reducing the work-related accidents caused by mooring operations on the ship. Considering personal security, intelligent identification can improve supervision efficiency, reduce labor costs, and provide early

warnings. Regarding company safety management, reducing the occupational accident rate can aid in better recording the existing hidden dangers and carrying out training and education. At the same time, this work can inspire future research on the unsafe behavior of the crew in other scenes on the ship. In the future, technology based on computer vision can also be used in other scenes, e.g., the bridge, to ensure the ship's and crew's safety.

5.2. Limitation of the Study

Due to time constraints, the data collection and testing have only been carried out aboard the M.V. YuKun. However, enlarging the data samples will afford a more accurate and comprehensive recognition performance. Furthermore, the algorithm lacks systematic debugging and only monitors the actual operation on the M.V. YuKun. According to the project, the detection system will be installed on the M.V. Yangtze 3, where several sea trials will further demonstrate the algorithm's feasibility.

5.3. Further Work

The application of computer vision in crew and ship safety is still limited. Hence, to promote the development and application of computer vision technology in the shipping safety field, the following research will be carried out:

1. Dataset expansion and adjustment. This research will be carried out under the "Platform for safety risk identification and prevention and control of ships at sea", with several cameras installed on the M.V. Yangtze 3. Utilizing the cameras near the deck mooring equipment, we will further expand the size of the datasets involving different weather conditions to achieve higher recognition accuracy. Meanwhile, new unsafe behaviors are constantly found during ship operations and will be added to the dataset to expand the types of unsafe behaviors to achieve better early warning effects in practical applications.
2. Commissioning of the algorithm. The algorithm will be further debugged, and the alarm sensitivity will be adjusted according to the actual early warning requirements.
3. Algorithm migration for other scenes aboard the ship, such as the bridge, the gangway, and other places. We will produce an unsafe behavior dataset for different scenes, where the developed system and algorithm will be evaluated.
4. Through the bracelet and other means, we will study further the crew's unsafe behavior. This project provides opportunities to obtain physiological information from each crew member by collecting data from their wearable sensors. Through multi-source information fusion, the video and sensor information will be combined to monitor the crew's working state comprehensively.

Author Contributions: Methodology, C.Z.; Validation, X.Y.; Resources, W.Z.; Data curation, C.C. and J.Y.; Funding acquisition, B.H. All authors have read and agreed to the published version of the manuscript.

References

1. Walls, L.; Revie, M.; Bedford, T. Risk, Reliability and Safety. In Proceedings of the ESREL 2016, Glasgow, UK, 25–29 September 2016; CRC Press: Boca Raton, FL, USA, 2016. ISBN 9781498788984.
2. AMSA. *Thinking Mooring Safety*; AMSA: Canberra, Australia, 2015.
3. DNV. Taking a new look at mooring safety—Industry insights. In *Maritime Impact*; DNV: Bærum, Norway, 2020.
4. Tyson, J. Mooring Line and Mooring Systems Management: Mooring Line and Mooring System Management. *Port Hedland Mar. Saf. Bull.* 2021. Available online: https://www.pilbaraports.com.au/PilbaraPortsAuthority/media/Documents/Port%20of%20Port%20Hedland/Safety%20and%20Security/Marine%20Safety%20Bulletins/2021/PH-01-2021-Mooring-Line-and-Mooring-Systems-Management.pdf (accessed on 25 July 2022).
5. Celik, M.; Cebi, S. Analytical HFACS for investigating human errors in shipping accidents. *Accid. Anal. Prev.* **2009**, *41*, 66–75. [CrossRef] [PubMed]
6. Yang, X.; Utne, I.B.; Holmen, I.M. Methodology for hazard identification in aquaculture operations (MHIAO). *Saf. Sci.* **2020**, *121*, 430–450. [CrossRef]
7. Ung, S.-T. A weighted CREAM model for maritime human reliability analysis. *Saf. Sci.* **2015**, *72*, 144–152. [CrossRef]
8. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [CrossRef]
9. Cheng, J.; Wong, P.K.-Y.; Luo, H.; Wang, M.; Leung, P.H. Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Autom. Constr.* **2022**, *139*, 104312. [CrossRef]
10. Zhang, M.; Shi, R.; Yang, Z. A critical review of vision-based occupational health and safety monitoring of construction site workers. *Saf. Sci.* **2020**, *126*, 104658. [CrossRef]
11. Ahn, J.; Park, J.; Lee, S.S.; Lee, K.-H.; Do, H.; Ko, J. SafeFac: Video-based smart safety monitoring for preventing industrial work accidents. *Expert Syst. Appl.* **2023**, *215*, 119397. [CrossRef]
12. Salihoglu, E.; Bal Beşikçi, E. The use of Functional Resonance Analysis Method (FRAM) in a maritime accident: A case study of Prestige. *Ocean Eng.* **2021**, *219*, 108223. [CrossRef]
13. Lee, E.J.; Ko, B.C.; Nam, J.-Y. Recognizing pedestrian's unsafe behaviors in far-infrared imagery at night. *Infrared Phys. Technol.* **2016**, *76*, 261–270. [CrossRef]
14. Luo, H.; Liu, J.; Fang, W.; Love, P.E.; Yu, Q.; Lu, Z. Real-time smart video surveillance to manage safety: A case study of a transport mega-project. *Adv. Eng. Inform.* **2020**, *45*, 101100. [CrossRef]
15. Cyganek, B.; Gruszczyński, S. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing* **2014**, *126*, 78–94. [CrossRef]
16. Guo, B.H.; Zou, Y.; Fang, Y.; Goh, Y.M.; Zou, P.X. Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Saf. Sci.* **2021**, *135*, 105130. [CrossRef]
17. Han, S.; Lee, S. A vision-based motion capture and recognition framework for behavior-based safety management. *Autom. Constr.* **2013**, *35*, 131–141. [CrossRef]
18. Fang, W.; Ding, L.; Luo, H.; Love, P.E. Falls from heights: A computer vision-based approach for safety harness detection. *Autom. Constr.* **2018**, *91*, 53–61. [CrossRef]
19. Yu, Y.; Guo, H.; Ding, Q.; Li, H.; Skitmore, M. An experimental study of real-time identification of construction workers' unsafe behaviors. *Autom. Constr.* **2017**, *82*, 193–206. [CrossRef]
20. Pramanik, A.; Sarkar, S.; Maiti, J. A real-time video surveillance system for traffic pre-events detection. *Accid. Anal. Prev.* **2021**, *154*, 106019. [CrossRef]
21. Roy, A.M.; Bose, R.; Bhaduri, J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. *Neural Comput. Appl.* **2022**, *34*, 3895–3921. [CrossRef]
22. Roy, A.M.; Bhaduri, J.; Kumar, T.; Raj, K. WilDect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. *Ecol. Inform.* **2022**, *in press*. [CrossRef]
23. Roy, A.M.; Bhaduri, J. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. *Comput. Electron. Agric.* **2022**, *193*, 106694. [CrossRef]
24. Xudong, Z.; Xi, K.; Ningning, F.; Gang, L. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Comput. Electron. Agric.* **2020**, *178*, 105754. [CrossRef]
25. Bhattacharjee, P.; Das, S. Two-Stream Convolutional Network with Multi-level Feature Fusion for Categorization of Human Action from Videos. In *Pattern Recognition and Machine Intelligence*; Shankar, B.U., Ghosh, K., Mandal, D.P., Ray, S.S., Zhang, D., Pal, S.K., Eds.; Springer International Publishing: Cham, Switzedrland, 2017; pp. 549–556. ISBN 978-3-319-69899-1.
26. Yu, B.X.; Liu, Y.; Chan, K.C.; Yang, Q.; Wang, X. Skeleton-based human action evaluation using graph convolutional network for monitoring Alzheimer's progression. *Pattern Recognit.* **2021**, *119*, 108095. [CrossRef]
27. Yang, Z.; Li, Y.; Yang, J.; Luo, J. Action Recognition with Spatio–Temporal Visual Attention on Skeleton Image Sequences. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 2405–2415. [CrossRef]
28. Liu, C.; Li, X.; Li, Q.; Xue, Y.; Liu, H.; Gao, Y. Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model. *Neurocomputing* **2021**, *430*, 174–184. [CrossRef]
29. IMO. *Amendments to the Code for the Investigation of Marine Casualties and Incidents (Resolution A.849(20))*; International Maritime Organization: London, UK, 1999.

30. Reason, J. *Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*; CRC Press: Boca Raton, FL, USA, 2008.
31. Sluiskes, B. *Safety in Mooring*. 2016. Available online: https://www.iadc-dredging.com/wp-content/uploads/2017/02/article-safety-in-mooring-143-2.pdf (accessed on 25 July 2022).
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640v5.
33. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242v1.
34. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767v1.
35. Hu, X.; Liu, Y.; Zhao, Z.; Liu, J.; Yang, X.; Sun, C.; Chen, S.; Li, B.; Zhou, C. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network. *Comput. Electron. Agric.* **2021**, *185*, 106135. [CrossRef]
36. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. *arXiv* **2019**, arXiv:1905.02244v5.
37. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861v1.
38. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
39. Avenash, R.; Viswanath, P. Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), Prague, Czech Republic, 25–27 February 2019.
40. Jaccard, J. The Distribution of the Flora in the Alpine Zone. *New Phytol.* **1912**, *11*, 37–50. [CrossRef]