IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 January 2024, accepted 31 January 2024, date of publication 5 February 2024, date of current version 20 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3362636

RESEARCH ARTICLE

Research on Intelligent Oil Drilling Pipe Column Detection Method Based on Improved Lightweight Target Detection Algorithm

BIN PENG AND KE NIU
School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China

Corresponding author: Bin Peng (pengb2000@lut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51675254 and Grant 51966009, in part by the National Key Research and Development Program of China under Grant SQ2020YFF0420989, in part by the Talent Innovation and Entrepreneurship Program of Lanzhou under Grant 2020-RC-23, in part by the Science and Technology Program of Gansu Province under Grant 20YF8GA057, and in part by the Central Guided Local Science and Technology Development Funding Program under Grant 23ZYQA312.

ABSTRACT The inadequate automation level in the transit of oil drilling tubular columns has led to significant inefficiencies and safety issues. To address these challenges, a real-time detection algorithm, ECS-YOLOv5s, has been proposed. This algorithm aims to improve the accuracy of drill pipe identification during operational processes, facilitating the automation of tubular column handling It has the potential to reduce drilling cycles and overall drilling costs. ECS-YOLOv5s enhance the detection accuracy of drill pipes by incorporating a Bidirectional Feature Pyramid Network (BiFPN) architecture with an improved multi-scale feature fusion network. The use of EfficientNet as the backbone network reduces the number of parameters and computations while effectively merging features from different layers. Additionally, the Spatial Pyramid Pooling (SPP) structure in the Neck is replaced with SPPF, and a Convolutional Block Attention Module (CBAM) is introduced to improve model robustness, reduce parameters and computations, and enhance the model's ability to detect dense targets. The ECS-YOLOv5s algorithm exhibits superior performance in drill pipe inspection, achieving a mean Average Precision (mAP) of 90.2%, a frame rate of 125 FPS, and a parameter count of only 37%. It achieves an accuracy of 98.6%, outperforming the original model by 9.2%. The comparative analysis demonstrates that the improved algorithm surpasses traditional models such as YOLOv5s, SSD, Faster-RCNN, and YOLOv7-tiny in both performance and accuracy. These findings provide valuable insights for the research on automated processing of tubular columns in intelligent oil drilling platforms.

INDEX TERMS Yolov5s, oil pipe column, deep learning, target detection, ESC-YOLOv5s.

I. INTRODUCTION

Drilling is a key link in the discovery, exploration, and exploitation of oil and gas resources, but the existing drilling technology is unable to meet the development needs of complex oil and gas resources in terms of economy, safety, high efficiency, and environmental protection. It is necessary to develop a new generation of transformative drilling technology. Intelligent drilling technology is a transformative drilling technology that integrates the theories and technologies of large data, artificial intelligence, information engineering, and downhole control engineering. Intelligent drilling technology is in its infancy, and for this reason, based on the systematic analysis of the current status of the development of key technologies and equipment for intelligent drilling, the key direction of intelligent drilling is discussed. To promote the basic theory research of intelligent drilling and improve the intelligent drilling technology system [1]. With the increase in available data and the rapid development of artificial intelligence (AI) technology. A large number of

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

machine learning studies have been conducted in different drilling applications. Data-driven models based on machine learning methods can provide greater advantages than traditional analytical or numerical models. Such as flexible model inputs, better prediction accuracy, and the ability to discover hidden patterns [2].

At present, the level of automation in oil drilling is relatively low, with most of the operation process still requiring manual labor. This creates a high demand for personnel and contributes to efficiency and safety problems. Therefore, the development of oil drilling tubular column automation processing technology is crucial in reducing manpower consumption, work intensity, and operation cycles, while improving operation efficiency and quality. In intelligent drilling pipe column inspection, the detection algorithm must possess strong robustness and adaptability to accurately identify drilling targets amid complex environments such as temperature, pressure, and vibration. High real-time performance is also required during the detection process, which puts significant demands on the algorithm's efficiency and speed. The continuous detection of tubular columns requires large data processing and analysis, which can be complex and require model training and optimization, thus presenting a challenge for model data storage and high-performance computing power. Moreover, different drilling geology and environmental conditions can significantly impact target recognition, necessitating strong environmental adaptability for different drilling scenarios. Completing target detection before complex processing for specific targets can effectively reduce the difficulty of subsequent processing and improve efficiency and accuracy. The image target detection technology, which mainly utilizes distinguishable target image features to complete target detection, is one of the key technologies in vision technology and is widely used in the field of automatic detection.

Currently, the level of automation in oil drilling remains low. On drilling platforms, most operations still require manual labor, resulting in high personnel demand, efficiency, and safety issues. Therefore, the development of oil drilling tubular column automation processing technology is crucial in reducing manpower consumption, work intensity, and operation cycles while improving efficiency and operation quality. During intelligent drilling pipe column inspection, the algorithm requires robustness and adaptability to handle the complex environment, including temperature, pressure, and vibration. The algorithm must also possess high real-time performance to ensure accurate identification of drilling targets. Processing and analyzing large amounts of data during continuous tubular column detection often require complex model training and optimization, posing challenges for model data storage and high-performance computing power. Different drilling geologies and environmental conditions can impact target recognition, necessitating strong environmental adaptability for different drilling scenarios. Target detection involves identifying all interested targets in the image and completing classification and localization accordingly. Utilizing distinguishable target image features, image target detection technology is a key aspect of vision technology and is widely used in the field of automatic detection.

The use of traditional target detection models may have problems such as low accuracy, slow speed, and large model size, while the improved models can achieve better performance in different fields. Therefore, the use of improved target detection models adapted to the needs of various domains is necessary to enhance the detection accuracy, response speed, and deployment efficiency in application scenarios to better meet operational requirements. Fan et al. [5] replaced the up-sampling module in the original model with the CARAFE up-sampling module in the Neck layer to improve the accuracy and average precision of the model to recognize honeysuckle. Gu et al. [6] proposed an improved YOLOv5 (AYOLOv5) based on the attention mechanism to improve the recognition rate of cell detection. Huang et al. [7] added a small target detection layer in the Neck of the network structure. The CBAM attention mechanism was added to the convolutional module to optimize the model, and the accuracy of the model was assessed by sensory evaluation, texture profile analysis, and chromaticity analysis. Jia et al. [8] improved the C3 module in the YOLOv5s feature fusion network and designed the C3CA module by combining the coordinated attention mechanism, which improved the accuracy of the metal corrosion recognition. Li et al. [9] introduced the EIoU and the Quality Focal Loss to optimize the loss function of the network, which solved the problem of accuracy reduction caused by sample inhomogeneity, and at the same time, accelerated the training convergence speed and improved the regression accuracy. Li et al. [10] improved the CSP, FPN, and NMS modules in YOLOv5s, which eliminated the influence of the external environment, enhanced the ability of multiscale feature extraction, and improved the detection distance and detection performance. Li et al. [11] proposed an underwater scallop recognition algorithm based on improved Yolov5s, designed a new lightweight backbone network model, and utilized group convolution and inverse residual block to replace the original Yolov5s backbone network, which improved the detection accuracy and accelerated the detection speed. Yu and Shin [12] proposed an improved scheme based on YOLOv5s, which combines coordinate concern blocks and uses a bidirectional feature pyramid network for better feature fusion, showing the effectiveness and applicability of the model in SAR image ship detection. Yuan et al. [13] proposed a novel YOLOv5s-CBAM-DMLHead method based on YOLOv5s to improve the performance of the model, which reduces the computation of the original model and the detection time. Chuang [30] summarized the current problems of the YOLO algorithm in the field of target detection and the future research trends.

The aforementioned research has proposed several algorithms that have been applied to various fields for detection purposes. These algorithms have made significant breakthroughs in engineering, agriculture, medicine, and other areas of identification. However, there have been very few

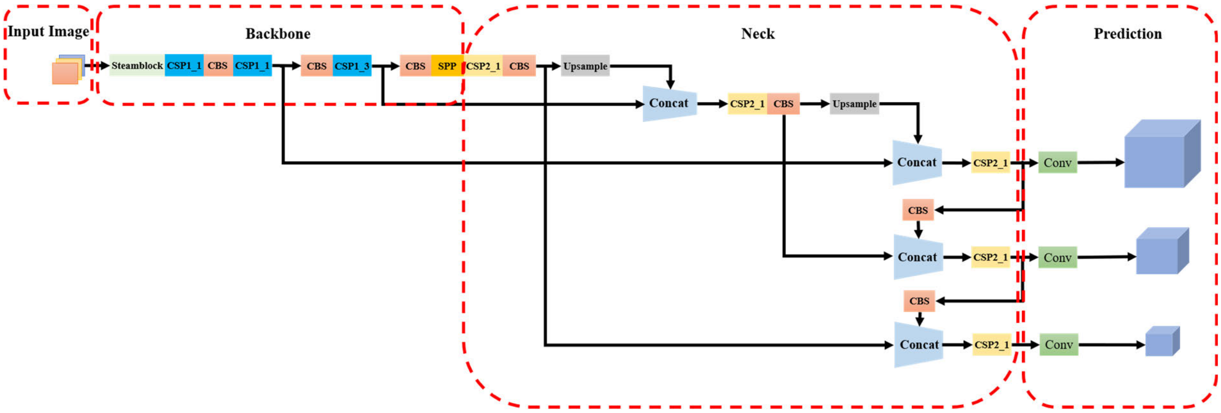


FIGURE 1. Traditional YOLOv5s model structure.

studies conducted on the recognition and detection of drilling pipe columns. The algorithm of ECS-YOLOv5s, presented in this paper, provides accurate and efficient detection of drilling pipe columns. Its robustness and performance have been experimentally verified and compared in different environments during the writing process.

II. CONVOLUTIONAL NEURAL NETWORK TARGET DETECTION ALGORITHM

A. YOLO TARGET DETECTION ALGORITHM

Convolutional Neural Network target detection algorithms include two main categories: region-based methods and single-stage methods.1) Region-based Methods: R-CNN (Region-based Convolutional Neural Networks) is one of the milestones in the field of target detection. It first extracts candidate regions using a selective search algorithm, then performs convolutional feature extraction for each candidate region, and finally performs target classification by a support vector machine classifier. Fast R-CNN [14]improves on R-CNN by feeding the entire image into the network and extracting features from the candidate regions through a RoI pooling layer, which reduces repetitive computation and memory consumption. Faster R-CNN introduces a candidate region generation network Region Proposal Network, RPN, based on R-CNN for fast generation of candidate regions and joint training with shared convolutional features. This allows for a more accurate extraction of candidate regions at different scales and aspect ratios. Mask R-CNN [15] adds segmentation of target instances to Faster R-CNN [16], which not only detects and classifies the targets but also generates an accurate segmentation mask for each target instance. 2) Single-shot Methods): YOLO [17], [18], [19], [20] is a very fast target detection algorithm that performs dense prediction directly on the image and accomplishes target classification and localization in a single stage. It divides the image into grids and directly predicts the class probability and bounding box information for each grid by using a convolutional neural network. SSD [21] (Single Shot MultiBox Detector) is another single-stage target detection algorithm, that employs multi-scale feature maps for detecting targets of different scales and aspect ratios. By applying multiple predefined anchor boxes on different levels of the feature map, SSD enables accurate detection of targets with different shapes and sizes. Region-based methods usually have better performance in terms of accuracy but are slower, while single-stage methods have faster speed but may be slightly less accurate. Depending on the specific needs, a suitable algorithm can be chosen to realize the target detection task.

B. YOLOv5 ALGORITHM

YOLOv5 is a deep learning-based target detection algorithm, which is the fifth version of the YOLO series of algorithms.YOLOv5 is characterized by fast speed and high accuracy and is suitable to be deployed on devices with limited resources. Arranged from smallest to largest model size, they are YOLOv5n, YOLOv5s, YOLOv5 m, YOLOv5l, and YOLOv5x. The different widths and depths of these models make YOLOv5 applicable to different datasets, which makes it easy for users to make choices. Because only one category of drill pipe is recognized in this study, and considering the need for real-time detection and easy deployment, the YOLOv5s model, which has fewer parameters and less computation, is used as the base model.YOLOv5s adopts the idea of a single-stage detector, which means that target detection is divided into two parts: firstly, we get the bounding box of the object, and then we classify and localize the bounding box. Moreover, YOLOv5s adopts the backbone network as CSPDarknet, which is a network architecture that improves accuracy with less computation and parameters through a new grouped convolution scheme. The structure of the YOLOv5s model is shown in Figure. 1.

YOLOv5s has the following key features: 1) Network structure design: YOLOv5s uses a lightweight network structure design that contains a backbone network, a feature pyramid network, and a prediction head. The backbone network uses CSPDarknet53 as a feature extractor, which builds more complex feature representations by using convolutional blocks and cross-layer connections. The feature pyramid network is used to process feature maps at different scales and generate multi-scale predictions. The prediction head is then responsible for predicting target classification and localization for features at different scales.2) Multi-scale training and inference: to improve detection accuracy and robustness, YOLOv5s employs a multi-scale training and inference strategy. In the training phase, YOLOv5s randomly scales the input images to different scales and performs data augmentations to increase data diversity. In the inference phase, YOLOv5s can receive an image of any size as input resize the image to a fixed size by interpolation and padding operations, and then perform target detection. 3) Adaptive Data Enhancement: YOLOv5s introduces an adaptive data enhancement strategy by adjusting the image enhancement according to the size and location of the target. Smaller targets are more strongly augmented to increase their importance in training, while larger targets avoid over-enhancement to prevent information loss.4) Detection and tracking strategy: YOLOv5s provides target tracking by applying a Kalman filter in the prediction process. The filter predicts the position of the target and provides continuous target tracking until the detection results are stabilized. YOLOv5s strikes a good balance between detection speed and accuracy compared to previous versions. It can achieve relatively accurate target detection and localization while ensuring high detection speed. Due to its lightweight network structure, YOLOv5s is more advantageous for embedded devices and mobile deployments. The algorithm flowchart of YOLOv5s is shown in Figure 2. First, the image is resized to a specified size and normalized to a pixel value within the range of 0 to 1. Simultaneously, data augmentation techniques like random image flipping, rotating, and cropping are applied to augment the training data and enhance the model's generalization ability. Next, the image is passed through a feature extraction network that's typically based on convolutional neural networks (CNNs) to extract useful features from the image. In YOLOv5s, CSPDarknet53 is used as the feature extraction network, where CSP refers to Cross cross-stage partial connections. This architecture balances computational efficiency with accuracy. The feature extraction network produces different levels of feature maps, and YOLOv5s uses a feature pyramid to detect targets of varying sizes. The feature pyramid merges all levels of feature maps using pooling and upsampling operations. By joining shallow feature maps with deeper feature maps, a feature pyramid with high-dimensional semantic information is produced. A detection head is then applied to the feature pyramid to detect objects in the image. YOLOv5s uses a CenterNet-based detection head that can predict the centroid and aspect ratio of the object simultaneously, thereby improving the accuracy of target detection. The detection head also employs optimization techniques like Feature Pyramid Networks (FPN) and Path Aggregation Network (PANet) to enhance detection performance. Finally, the prediction results undergo several post-processing steps to determine the final detection frames and category labels. These post-processing steps include the de-duplication of frames and confidence screening to improve the accuracy and reliability of the prediction results.

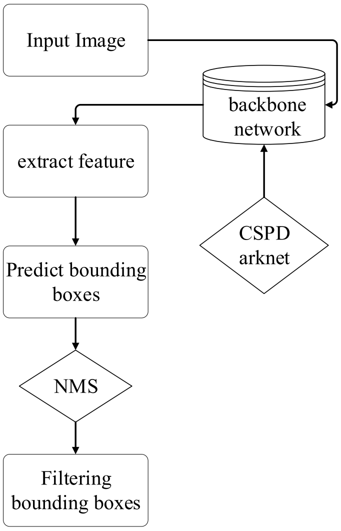


FIGURE 2. Algorithm flowchart of YOLOv5s.

Compared with the YOLOv4 [22] algorithm, the improvement of YOLOv5s is mainly in two aspects: 1) the backbone network adopts CSPDarknet, which not only improves the accuracy, but also has higher computational efficiency; 2) during the training process, YOLOv5s adopts Mosaic data augmentation, Self-Adaptive Training [23], and Label Smoothing [24] to enhance the model's generalization ability and robustness.

Input Layer: The input layer mainly consists of Mosaic image enhancement, adaptive anchor frame computation, and adaptive image scaling. The structure of the input layer is shown in Figure 3.

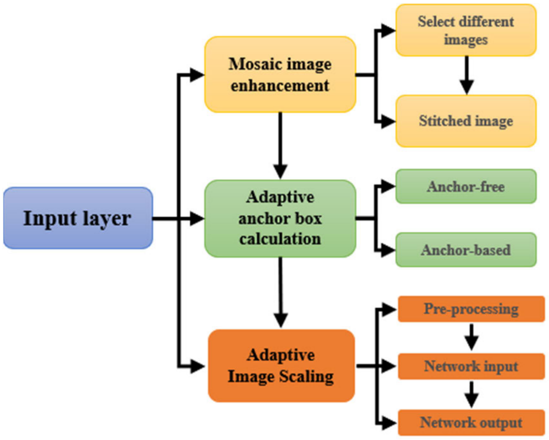


FIGURE 3. Input layer structure diagram.

- *The Backbone:* YOLOv5s backbone network is a deep convolutional neural network consisting of several convolutional layers, pooling layers, and activation functions. It is used in the YOLOv5 target detection algorithm and is one of the core parts of the algorithm.

(1) The Focus structure in YOLOv5 is a convolutional neural network layer used for feature extraction, which is used to compress and combine the information in the input feature maps to extract higher-level feature representations. The diagram of the Focus structure is shown in Figure. The Focus structure is a special convolutional operation in YOLOv5, which is used as the first convolutional layer in the network to downsample the input feature maps to reduce the amount of computation and the number of parameters. The Focus structure diagram is shown in Figure 4.

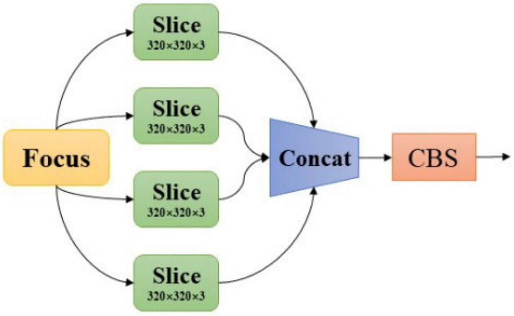


FIGURE 4. Focus structure diagram.

(2) CSP (Cross Stage Partial) structure is an important component in YOLOv5, which can effectively reduce the network parameters and computation while improving the efficiency of feature extraction. The core idea of the CSP structure is to split the input feature map into two parts, one of which is processed by a small convolutional network (called a sub-network), and the other part is directly processed in the next layer of processing. The two parts of the feature maps are then stitched together and used as input for the next layer. The structure is shown in Figure 5.

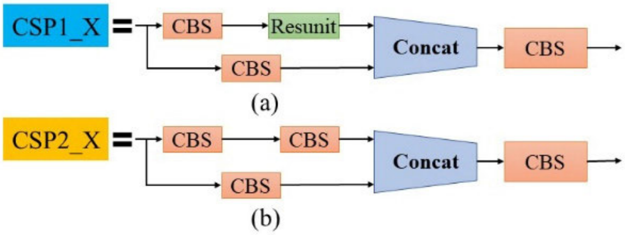


FIGURE 5. CSD structure diagram.

The neck network in YOLOv5 refers to the intermediate feature extraction network added on top of the backbone network, which is mainly used to enhance the feature expression ability of the model and further improve the detection performance of the model. Two different Neck network structures are used in YOLOv5: SPP and PAN.

(2) SPP structure

Spatial Pyramid Pooling (SPP) structure is a pyramid pooling structure that can pool feature maps of different sizes to enhance the model's ability to perceive targets at different scales. The main idea of the SPP structure is to fuse information at different scales by pooling input feature maps of different sizes. Its structure is shown in Figure 6. As can be seen from the figure, the SPP module consists of maximum pooling with three different pooling kernel sizes, a jump join, and a stacking operation, where kerbel size $=\{1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13\}$.

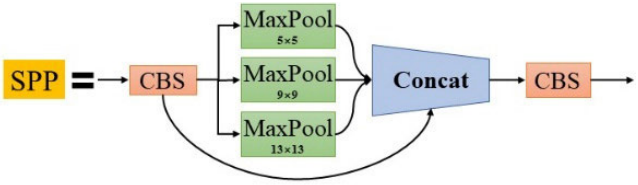


FIGURE 6. SPP sturcture diagram.

(2) PAN structure

Path Aggregation Network (PAN) is a feature pyramid network structure for target detection. The PAN structure is mainly composed of two modules, feature pyramid, and feature fusion, which aims to improve the model's ability to perceive targets at different scales through multi-level feature fusion.

• Outputs: The output of YOLOv5 is mainly prediction boxes, each of which consists of the following information:

1) confidence score: the probability of whether the target exists in the box, ranging from 0 to 1; 2) class probabilities: the probability that the target in the box belongs to each class, generally the number of predefined classes; 3) bounding box coordinates: indicates the position and size of the target, usually represented by a rectangular box.

The output layer in YOLOv5 generally includes three feature maps at different scales, each feature map corresponding to a prediction box at a different scale, and each prediction box containing information as described above. Specifically, YOLOv5 predicts the location and size of the bounding box of the target in the output layer by using an anchor box, and at the same time calculates the category probability by using a softmax function for the prediction result corresponding to each anchor box.

(1)Bounding box

The Bounding box loss function in YOLOv5 uses the IoU loss function, which is mainly used to measure the difference between the predicted bounding box and the true bounding box.

IoU loss is a variant of Intersection over Union (IoU), which is a metric used to measure the degree of overlap between the predicted bounding box and the true bounding box. In target detection, IoU is often used to evaluate the overlap between predicted and true boxes to determine whether the predicted boxes are correct.

Specifically, for each predicted bounding box, we calculate its IoU value with all the real bounding boxes, and then select the real bounding box with the largest IoU as its corresponding matching target, to calculate its IOU loss. its specific use of the DIOU loss function with the following formula:

$$L_{DIOU} = 1 - IOU + \frac{P^2\left(b, b^{gt}\right)}{c^2} \tag{1}$$

$$L_{CIOU} = 1 - IOU + \frac{P^2\left(b, b^{gt}\right)}{c^2} + av \tag{2}$$

$$a = \frac{v}{(1 - IOU) + v} \tag{3}$$

$$v = \frac{4}{\pi^2}\left(\tan^{-1}\frac{\mathrm{w}^{gt}}{\mathrm{h}^{gt}} - \tan^{-1}\frac{\mathrm{w}}{\mathrm{h}}\right) \tag{4}$$

(2) NMS non-maximum suppression

In a target detection task, an object may be detected by multiple prediction frames, to avoid multiple detections of the same object, the duplicate prediction frames need to be filtered, this process is Non-maximum suppression (NMS). The core of the NMS algorithm is to remove the redundant prediction frames by comparing the IOUs between the duplicate prediction frames and retaining the optimal result of prediction frames and retaining the optimal result. In YOLOv5, NMS can avoid the problem of repeated detection of the same object and improve the accuracy and efficiency of detection.

III. BASED ON IMPROVED YOLOv5s

A. AN IMPROVED NECK STRUCTURE

The SPPF (Spatial Pyramid Pooling Fusion) structure further introduces a feature fusion module based on the SPP structure to improve the model's perceptual ability and detection performance. Specifically, the SPPF structure first performs different sizes of pooling operations on the input feature maps, then fuses the pooling results of different scales by convolution operations, and finally outputs the fused feature maps. The structure is shown in Figure 7.

The SPPF structure is characterized by the ability to adaptively fuse feature information of different scales, thus enhancing the feature expression and sensing ability of the model.

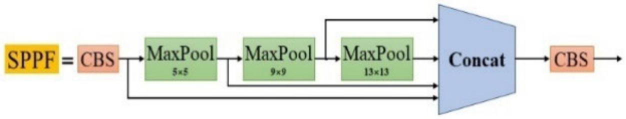


FIGURE 7. SPPF sturctures diagram

B. IMPROVED MULTI-SCALE FEATURE FUSION NETWORK

As the number of network layers deepens, the semantic information of the feature graph becomes richer and richer, but the detail information gradually decreases, which is extremely unfavorable for small target detection. However, if only shallow features with rich detailed information are used for detection, it will reduce the detector's performance. To overcome the shortcomings of deep features and shallow features, the fusion of deep and shallow features is generally used to obtain more comprehensive and rich feature information. As shown in the following figure, three typical design forms of multi-scale feature fusion networks are demonstrated, and these three structures also represent the development process of feature fusion networks to some extent. The characteristics of each network structure are introduced one by one below. The multi-scale feature fusion network architecture is shown in Figure 8.

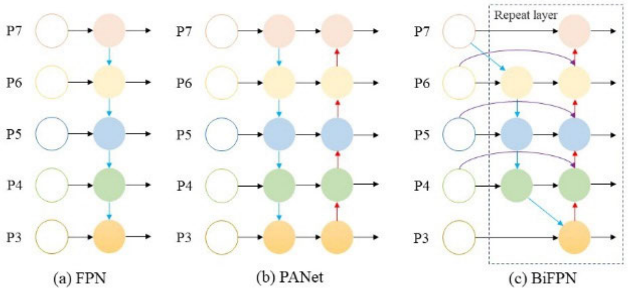


FIGURE 8. Multi-scale feature fusion network architecture.

A Feature Pyramid Network (FPN) is a widely used network structure for target detection tasks that address the inadequacy of single-scale features. It utilizes bottom-up and top-down connections to flow information from the feature pyramid to detect targets at different scales. The FPN structure is shown in (a). From the figure, it can be seen that FPN can only convey strong semantic information from the top down, and fuses the deeper feature maps with the shallower ones after up-sampling them.

Features are extracted from the input image by a backbone network, usually a convolutional neural network, to obtain a series of feature maps at different levels, with progressively lower resolutions and different semantic information. Starting from the highest resolution feature maps, the resolution of the feature maps is increased by up-sampling, and a summing operation is performed with lower resolution feature maps to obtain the fused feature maps. Specifically, each up-sampled feature map is summed with the neighboring lower-resolution feature maps to form a pyramid structure, and a 1×1 convolution operation adjusts the number of channels of the neighboring feature maps to ensure the channel consistency of the feature maps in the fusion process. However, in deep neural networks, the path to pass shallow features to deep features is generally very long, and most of the information of the target may have been lost in the downsampling process, and this approach fails to perform feature fusion well.

Path Aggregation Network for Instance Segmentation (PANet) improves based on FPN to further improve the performance of the feature pyramid. Unlike FPN, PANet introduces horizontal and vertical feature fusion mechanisms, and the structure of PANet is shown in (b), which contains two feature fusion paths, top-down and bottom-up. This shortens the distance from the shallow features to the deeper features optimizes the feature fusion of the FPN network to a certain extent, and improves the effect of target detection. Still, at the same time, it also increases the number of parameters and computation amount, which is different from FPN. Similar to the lateral connection of FPN, PANet also adjusts the channel number of neighboring feature maps by 1 × 1 convolution operation. However, unlike FPN, PANet also introduces a feature supplementation module, i.e., fusing the laterally connected feature maps with the upper-level feature maps to enhance the lower-resolution features. In FPN, the bottom-up feature extraction network simply generates a feature pyramid, whereas PANet introduces a cascade fusion module, which fuses the features of each level of the feature pyramid with the upper-level features.

Weighted Bidirectional Feature Pyramid Network (BiFPN) is a further improvement of the feature pyramid structure for target detection tasks. It adds a multi-level feature fusion mechanism based on PANet and introduces cross-level feature connections to better balance the feature information at different levels. The structure of BiFPN is shown in (c), which verifies the effectiveness of bidirectional feature fusion in the PANet network, but the structure is simpler.

Bottom-up construction: similar to the traditional feature pyramid network, different levels of feature maps are extracted through a backbone network. BiFPN introduces the connection of top and bottom branches, and in the process of bottom-up connection, each resolution level can connect the feature maps of its upper and lower levels as needed. Such a bidirectional connection can realize the transmission and fusion of information. Based on bi-directional connection, BiFPN further introduces feature fusion operation. Specifically, each node calculates the weighted sum between the feature maps of different resolution levels, where the weights are adaptively calculated according to the resolution levels and can be adjusted based on learning to balance the feature information contribution of different levels. To further enhance the information transfer, BiFPN also introduces additional cross-layer connections. Specifically, for each resolution level, it can not only connect the feature maps of the previous and next levels but also connect with the levels that are far apart. This extends the range of information flow and increases the perceptual capability of target detection.

In 2020, Tan and his colleagues proposed a novel network called BiFPN [39], which is an extension of the PANet network that aims to further optimize its performance. BiFPN mainly makes four key improvements to the PANet architecture. Firstly, it removes nodes with only one input, which do not contribute to feature fusion and increase computational complexity. Secondly, it increases jump connections between features extracted from the backbone network and those involved in downsampling fusion. This can fuse more features without increasing the cost and effectively alleviate the feature loss phenomenon caused by too many network layers. Thirdly, BiFPN considers bi-directional paths as a single unit, allowing more advanced features to be fused. Lastly, it proposes a weighted feature map fusion strategy that assigns learnable weights to the features of different scales involved in fusion, thus regulating the importance of features. These improvements enable BiFPN to achieve better performance than the original PANet network.

C. HYBRID CBAM ATTENTION MECHANISM

The attention mechanism distinguishes the importance of feature information by adjusting the size of the weights. Given that the fusion of secondary features does not contribute much to the detection results and also increases the computational effort, the weights of the secondary features have to be lowered while the important features are given higher weights. Since YOLOv5 first fuses the features to different degrees at the Neck side, and then the detection head directly outputs the prediction results on the fused feature map, it is necessary to add an attention mechanism to boost the critical features and suppress the irrelevant features before the next feature fusion at the Neck side.

The network structure after adding the CBAM [42] module is shown below. The YOLOv5s-P2-CBAM model adds a CBAM module after each C3 module at the Neck side and before the convolution operation. That is, between two feature fusions, the attention mechanism is added to enhance the network's attention to small targets. The network after adding the CBAM module performs a feature enhancement operation on the feature map before the next feature fusion. So that the network ignores the interference of irrelevant information, focuses on the key features, and fuses the relatively important features. This not only makes the fused feature map contain more effective information and improves the accuracy of small target localization but also achieves the purpose of reducing the amount of computation and improving the speed of the model.

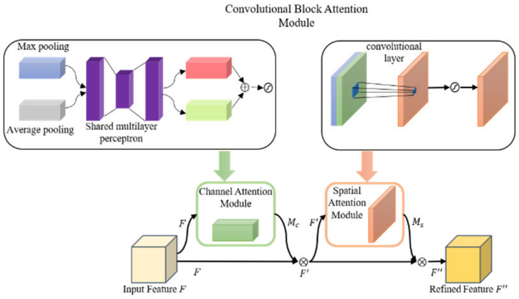


FIGURE 9. Convolutional black attention module.

As can be seen from Figure 9, after adding the CBAM module, the feature processing flow at the Neck end is as follows: 1) In the FPN structure, up-sampling operations are continuously performed on the deep feature maps to fuse the shallow feature maps of larger scales. For each fused feature map, the C3 operation is performed first to reduce the computation of the network. Then the CBAM attention mechanism is added to enhance the network's ability to pay attention to key channels and key regions in the feature map. Then operations such as convolution and upsampling are performed for the next feature fusion. 2) The CBAM attention mechanism is also introduced into the PAN structure to update the feature map after the last fusion, to enhance the network's attention to important features, and to incorporate more useful shallow geometric features into the deeper features; and then the feature map is downsampled for fusion with the small resolution feature map.

D. EFFICIENTNET LIGHTWEIGHT NETWORK

EfficientNet is a convolutional neural network architecture and scaling method that uses composite coefficients to uniformly scale all dimensions of depth/width/resolution. It aims to reduce the complexity and computational cost of the model while maintaining high performance. The network was proposed by a team of Google researchers and achieves efficient and effective modeling by using a composite scaling parameter (composite scaling) approach that simultaneously scales the depth, width, and resolution of the network at different network layers. Unlike traditional approaches that arbitrarily scale these factors, the EfficientNet scaling method uses a fixed set of scaling coefficients to uniformly scale network width, depth, and resolution. EfficientNet employs composite coefficients to uniformly scale network width, depth, and resolution in a principled manner. Experiments have justified the composite scaling method in that if the input image is larger, then the network needs more layers to increase the receptive domain and more channels to capture finer-grained patterns on the larger image. The underlying EfficientNet-B0 network is based on the reverse bottleneck residual block of MobileNetV2, as well as the squeeze and excitation blocks. The EfficientNets also have good migration accuracies on CIFAR-100(91.7%), Flowers (98.8%), and three other migration learning datasets, and with an order of magnitude fewer parameters.

The EfficientNetB0 structure is a powerful and efficient neural network architecture, where "B" denotes the basic version of the network. EfficientNetB was proposed by the Google Brain team in 2019 to design a neural network with excellent performance along with efficient computation and parameter counts. It is the smallest EfficientNet model so far with less number of parameters and computational complexity. The composite scaling of the EfficientNetB0 model is computed as in Eq.

$$
\begin{aligned}
&depth : d = \alpha^{\varphi} \\
&width : \omega = \beta^{\varphi} \\
&resolution : r = \gamma^{\varphi} \\
&s.t. \begin{cases} \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ \alpha \geq 1, \quad \beta \geq 1, \gamma \geq 1 \end{cases}
\end{aligned} \tag{5}
$$

where α, β, γ are constants that can be determined from a mini-grid search. φ is a user-specified factor that controls how many resources are available for model scaling, while α, β, and γ specify how to allocate these additional resources to network width, depth, and resolution, respectively.

The core component of EfficientNetB0 is a convolutional neural network whose architecture is based on a variant of the Inception model. It employs a range of convolutional operations, which include standard convolution, depth-separable convolution, and channel-by-channel linear combinations to improve the model's representational power and computational efficiency. An important feature of the EfficientNetB0 architecture is the use of a design methodology known as a depthwise separable convolutional network (DSCN). Network) design methodology. This approach combines two techniques, Depthwise Separable Convolutional, and Channel-by-Channel Linear Combination, to reduce the amount of computation and the number of parameters while maintaining the model's representational power. Specifically, the EfficientNetB0 model consists of multiple stacked convolutional layers containing depth-separable convolutional layers, channel-by-channel linear combination layers, and standard convolutional layers. The entire network structure is divided into multiple stages, each of which consists of multiple convolutional modules. Each convolutional module contains a depth-separable convolutional layer and a channel-by-channel linear combination layer for extracting features at different levels. The basic network structure is shown in Table 1. In addition, EfficientNetB includes some additional techniques such as the Swish activation function and the SE [36] (Squeeze-and-Excitation) module. The Swish activation function provides better nonlinear modeling capability while maintaining high computational efficiency. The SE module can adaptively adjust the weights of different feature channels to further improve the model's characterization ability, achieving a smaller number of parameters and computational complexity while maintaining a strong feature extraction capability.

TABLE 1. EfficientNet-B0 baseline network.

Stage/ i	Operator/ $\hat{F}_i$	Resolution/ $\hat{H}_i \times \hat{W}_i$	Channels/ $\hat{C}_i$	Layers/ $\hat{L}_i$
1	Conv3×3	224 × 224	32	1
2	MBConv6, k 3×3	112 × 112	16	1
3	MBConv6, k 3×3	112 × 112	24	2
4	MBConv6, k 5×5	56 × 56	40	2
5	MBConv6, k 3×3	28 × 28	80	3
6	MBConv6, k 5×5	28 × 28	112	3
7	MBConv6, k 5×5	14 × 14	192	4
8	MBConv6, k 3×3	7 × 7	320	1
9	Conv1×1&Pooling&FC	7 × 7	1280	1

MBConv is a modularized convolutional layer structure used in EfficientNet, and its structure is shown schematically in Figure 10. The core idea of the MBConv structure is to build a lightweight and efficient neural network by using a combination of deeply separable convolution and dilation convolution. It combines the concepts of Inverted Residual and Bottleneck and aims to improve the efficiency and performance of the model. In Table 1, MBConv1 does not expand the number of channels of the features during feature extraction, while MBConv6 indicates that the number of channels of the features is expanded to six times the number of channels of the input features, aiming at expanding the features in

TABLE 2. Improved process of the YOLOv5s.

```
Algorithm 1: Improved automated processing algorithm for tube and column based on YOLOv5s
Require: D: Data of column connection images
Ensure: R: Detection results
Initialize the YOLOv5s network with random weights and$\beta$.
Set the training stage: $num\_epochs$ = 200, $batch\_size$ = 16.
Prepare the normal dataset norm VOC_norm_trainval.
for $i$ in $num\_epochs$ do
  repeat
    Take a batch images M from VOC_norm trainval;
      for $i$ in $batch\_size$ do
        if $random.randint$(0,2)>0 then
         Generate the foggy image, where $A$ = 0.5, $k$ = random, $\beta$=0.01×k+0.05//for foggy conditions
          Step 1: Data Preprocessing
          $D_{preprocessed} \leftarrow$ Preprocess (D)
          Step 2: Feature Extraction
          F ← ExtractFeatures (Dpreprocessed)
          Step 3: Spatial Pyramid Pooling (SPPF)
          $F_{SPPF} \leftarrow$ ApplySPPF (F)
          Step 4: Channel Attention Mechanism (CBAM)
          A ← ApplyCBAM ($F_{SPPF}$ )
          Step 5: Multi-Scale Feature Fusion
          Fusion ← Fusion ($F_{SPPF}$, A)
          Step 6: YOLOv5s Object Detection
          Candidates ← DetectObjects ($F_{fusion}$)
          Step 7: Pruning and Filtering
          R ← PruneAndFilter R candidates)
          Step 8: Post-processing and Visualization
          R ← PostProcess (R)
        end if
      end for
  Compute DIP params by $P_N = P^{\theta}(image\_batch)$;
  Perform DIP filter processing by $image\_batch$ = DIP ($image\_batch$, $P_N$);
  Send $image\_batch$ to improved YOLOv5s network D.
  until all images have been fed into training models
end
```

the channel dimension; k3 × 3 indicates that convolutional kernel of size 3 × 3 is used in the MBConv structure, and k5 × 5 indicates that convolutional kernel of size 5 × 5 is used in the MBConv structure. k5 × 5 denotes a convolution with kernel size 5 × 5 in the MBConv structure [5].

The MBConv structure consists of three parts: dilation convolution, depth separable convolution, and projection convolution. The first is the dilation convolution, which applies a convolution operation with a dilation rate greater than 1 on the input feature map. Dilation convolution has a larger sensory field than traditional convolution and helps the model to better capture the global information in the image. The next step is depth-separable convolution, which is divided into two steps: depthwise convolution and pointwise convolution. Depthwise convolution first applies spatial convolution on each input channel and then performs point-by-point convolution on the channels to reduce the amount of computation and the number of parameters. This operation allows for better feature extraction with relatively minimal computing expense. Finally, there is the projection convolution, which is used to adjust and match the number of channels of the input and output feature maps. The projection convolution is a point-by-point convolution used to map the number of channels of the input feature map to the number of channels of the output feature map. With the combination of these three components, the MBConv architecture provides higher parametric and computational efficiency, thus excelling in model lightweight and efficiency. It has already achieved remarkable results in EfficientNet and some other excellent lightweight networks.

E. LIGHTWEIGHT DRILL PIPE TARGET INSPECTION MODEL

Replacing the traditional YOLOv5s backbone network with the EfficientNet lightweight model, to reduce computational and storage resources while maintaining high performance and avoiding accuracy degradation. The proposed method also incorporates an improved multi-scale feature fusion network BiFPN architecture to enable the fusion of features from different feature layers, enhancing the model's ability to extract features and improving training accuracy. Additionally, the SPP module is replaced with the SPPF module,

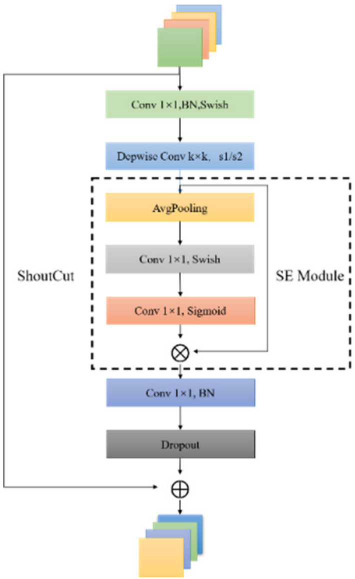


FIGURE 10. **MBConv structure diagram.**

and the CBAM attention mechanism module is introduced to enhance the model's robustness, reduce parameters and computation, and further improve feature extraction ability and detection accuracy. The proposed lightweight YOLOv5s model is named ECS-YOLOv5s. The improved process of the ECS-YOLOv5s model is shown in Table 1.

Therefore, the improved network model aims to improve the model detection accuracy and robustness while reducing the number of model parameters and computation, which is suitable for lightweight scenarios and better able to identify drill pipes in complex operating environments. The structure of the improved lightweight YOLOv5s model is shown in Figure 11.

IV. TEST METHODS

A. IMAGE DATA PROCESSING AND ENVIRONMENT SETTINGS

The experimental validation of this study was conducted under Windows 10 operating system with Intel(R) Core(TM) i7-10700F CPU @ 2.90GHz, 64-bit operating system, NVIDIA GeForce GTX 1060 6GB for GPU, 32GB of host computer RAM, Pytorch deep learning framework, Pycharm IDE, software environment is CUDA10.2, torch version 1.7.1, Python3.7 programming language to run.

B. NETWORK MODEL SELECTION

There are many convolutional neural network models for target detection, such as the early R-CNN family of models, SSD, CenterNet [48], RetinaNet [49], Mask R-CNN, EfficientDet, etc., as well as new algorithmic models published in the last few years such as MobileNet [50], EfficientNet, etc., the network models are getting smaller and more compact, and the recognition speed as well as the recognition accuracy of the models are getting higher and higher. Therefore, in this paper, we use AlexNet, a network that has been used more times in previous studies and a newly published network in recent years, as a deep convolutional neural network model for recognizing pipe columns, and by adjusting the hyperparameters of the network model such as the learning rate, the number of training cycles, and the size of the sample batch, the images of pipe columns are recognized.

C. NETWORK INFRASTRUCTURE

The convolutional neural network mainly consists of the input layer, convolutional layer, pooling layer, fully connected layer, and output layer. In this paper, the recognition of drill pipe is based on five kinds of convolutional network models, YOLOv5s, Faster-RCNN, SSD, improved light-weight YOLOv5s, and YOLOv7-tiny [55], and through the comparison of the experiments, we find out the network model suitable for the recognition of pipe columns. The flowchart of the training and testing process. is shown in Figure 12.

D. NETWORK MODEL EVALUATION METRICS

In the process of target detection, the model outputs several prediction frames after calculating the input image, and these prediction frames can be categorized into four categories True Positive (TP), false positive (FP), false negative (FN), and True Negative(TN) according to their classifica-tion results [28], and the corresponding evaluation indexes are calculated according to the prediction frame judgment is correct or not to calculate the corresponding evaluation index. Where TP indicates that the true target A is correctly predicted as target A, TN indicates that the true target B is correctly predicted as target B, FN indicates that the true target A is incorrectly predicted as target B, and FP indicates that the true target B is incorrectly predicted as target A. The four categories are shown in Figure 13.

Based on the above four categories of samples, various evaluation metrics of the model can be obtained, such as Recall, Precision, Average Precision (AP), mean Average Precision (mAP), and Frame Rate (FPS).

1) Recall rate, also known as check all rate, refers to the ratio of the number of correctly detected objects to the total number of objects in the test set, and the expression is shown in Eq:

$$Recall = \frac{TP}{TP + TN} \tag{6}$$

2) The accuracy rate, also known as the detection rate, refers to the ratio of the number of correctly detected objects to the total number of detected objects, and the expression is shown in the formula:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

3) The average precision represents the average detection precision of the single-category model, which is the area enclosed under the P-R curve formed by the coordinate system established with the recall rate as the horizontal coordinate and the precision rate as the vertical coordinate,

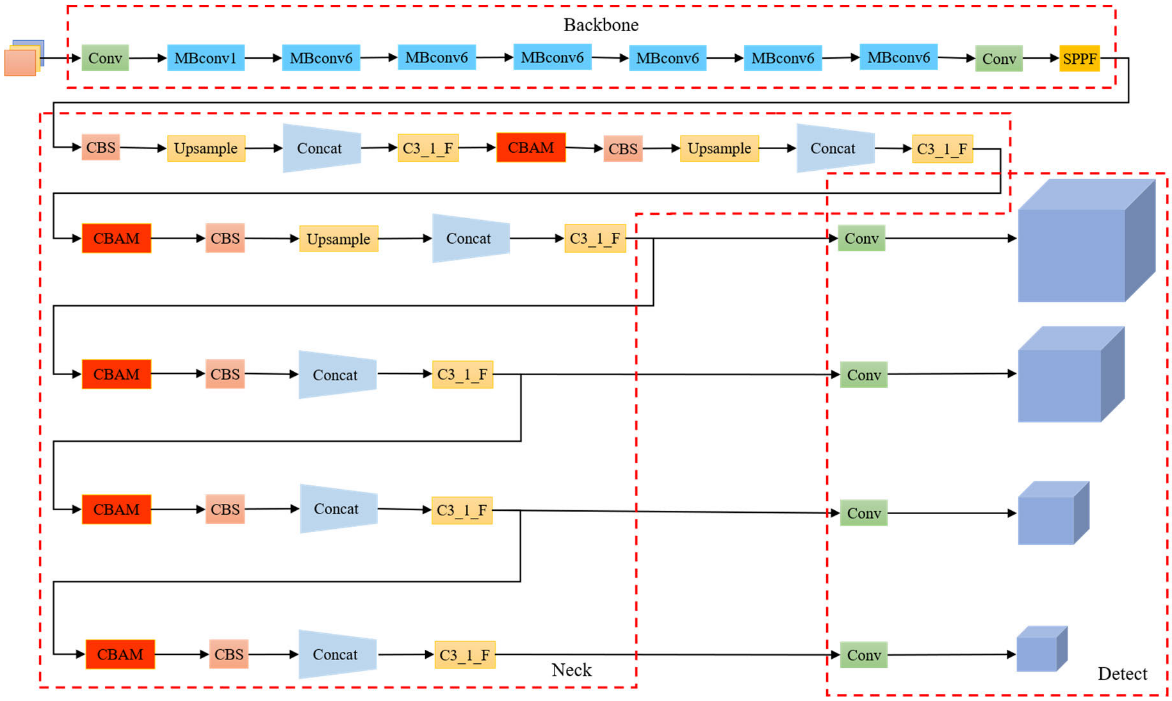


FIGURE 11. Lightweight YOLOv5s network structure.

based on a certain threshold, with the expression shown in Eq:

$$AP = \int_0^1 prdRe \tag{8}$$

4) The average precision mean value indicates the average precision of all categories N. The larger value indicates the higher precision of the target detection model, and the specific expression is shown in Eq:

$$mAP = \frac{\sum AP}{\mathrm{N}} \tag{9}$$

5) The frame rate, i.e., the number of frames detected by the model per second, is used as an important indicator of the real-time performance of the model.

V. MODEL TRAINING AND RESULT ANALYSIS

A. TRAINING PARAMETER SELECTION

Hyperparameter selection is a very critical step in Convolutional Neural Networks (CNN), which directly affects the performance of the model and the training process. The following are some common methods for hyperparameter selection: 1) Grid Search: For each hyperparameter, define a range and a set of candidate values, and then use the grid search method to try all possible hyperparameter combinations. The best-performing hyperparameter combination is selected through cross-validation or other evaluation metrics. 2) Random Search: In contrast to Grid Search, Random Search uses random sampling to assess the performance of each set of hyperparameters inside the stated space of hyperparameters. Finding the hyperparameters with improved performance requires iterating through several combinations. 3) Experience-based selection: based on previous experiments and experiences, the commonly used hyperparameter values are selected as the initial tuning values. On this basis, fine-tuning and optimization are carried out. 4) Automatic tuning algorithms (such as Bayesian optimization, genetic algorithms, etc.): automatic tuning algorithms are used to search the hyper-parameter space, and optimize the hyper-parameters according to the evaluation indexes. These algorithms can intelligently adjust the values of hyperparameters based on previous attempts and results, thus speeding up the search process. 5) Use of heuristic rules: Heuristic rules are used to select the range and initial values of hyperparameters based on the characteristics of the CNN architecture and the task. For example, the number of filters in the convolutional layer is usually a power of 2, the initial value of the learning rate can be set to 0.1, etc. 6) Visualization and analysis: Observe the effects of different hyperparameters on the model by visualizing and analyzing the model performance and loss curves during the training process, and then make adjustments. When making a hyperparameter selection, it is necessary to weigh the computational resources and time. Usually, a small portion of data can be used for initial hyperparameter search and experimentation, and then a few most promising hyperparameter combinations can be selected for more in-depth training and evaluation. In this way, better hyperparameter settings can be found with limited resources. After practical validation, the training parameters are shown in Table 3.

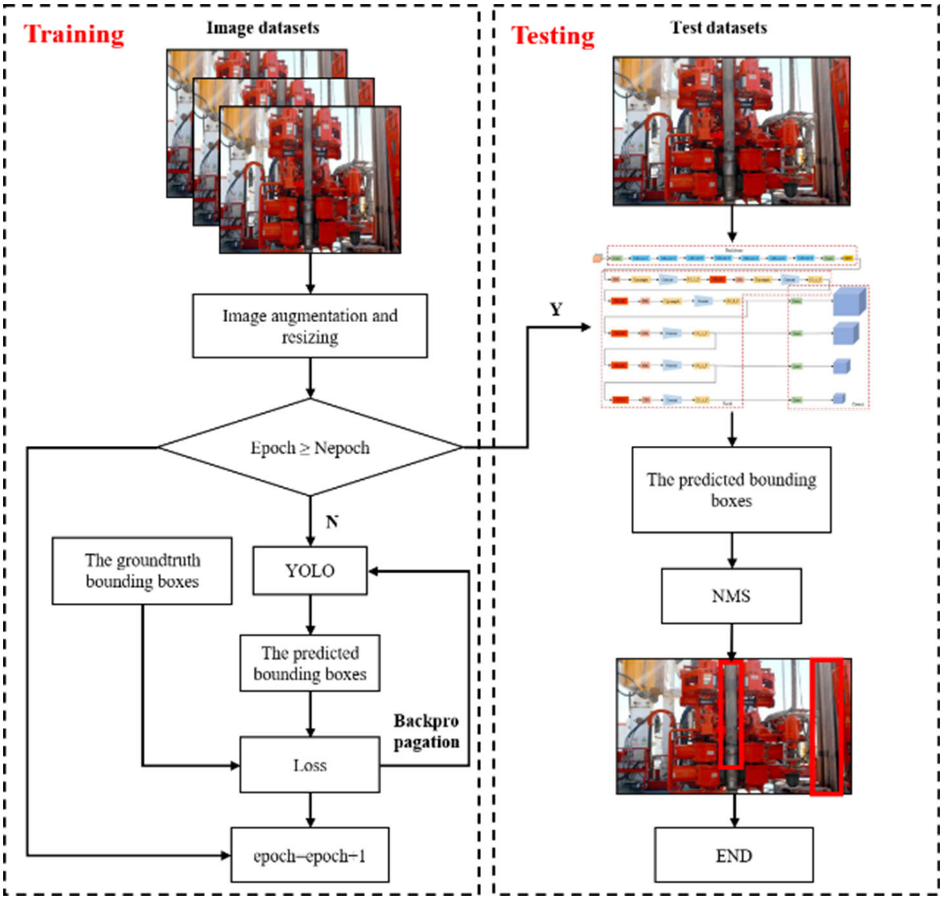
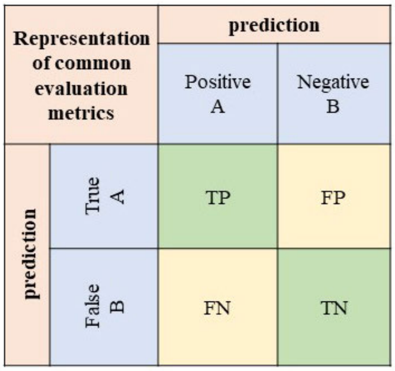

Training
Image datasets
Image augmentation and resizing
Epoch ≥ Nepoch
N
Y
The groundtruth bounding boxes
YOLO
The predicted bounding boxes
Backpropagation
Loss
epoch=epoch+1
Testing
Test datasets
The predicted bounding boxes
NMS
END

FIGURE 12. The flowchart of the training and testing process.



Representation of common evaluation metrics		prediction	
		Positive A	Negative B
prediction	True A	TP	FP
	False B	FN	TN

FIGURE 13. Evaluation indices.

B. DATA SET CONSTRUCTION

1) IMAGE PROCESSING

This study mainly verifies the detection effect of the improved algorithm on oil drill pipe and lays a good foundation for the subsequent automated processing of drilling pipe columns. The detection-grabbing flowchart is shown in Figure 14.

This study presents a self-constructed dataset of drill pipes due to the limited availability of photographs and datasets for this specific type of equipment. The imaging data of the drill pipes was collected by Lan Shi Petroleum Equipment Engineering Co. Ltd. in Lanzhou City, Gansu Province. Images of the drill pipes in different environments were captured using a cell phone under natural light, which included various lighting and positioning scenarios. The acquired images were saved in .jpg format and underwent data enhancement preprocessing, such as image mirroring, noise, and rotation, to increase dataset diversity and improve the overall generalization ability of the model. This approach effectively avoided overfitting problems and

TABLE 3. Model training parameter.

Parameters	Values
Epochs	500
Batch size	128
Learning rate	0.001
Dimension of picture	640

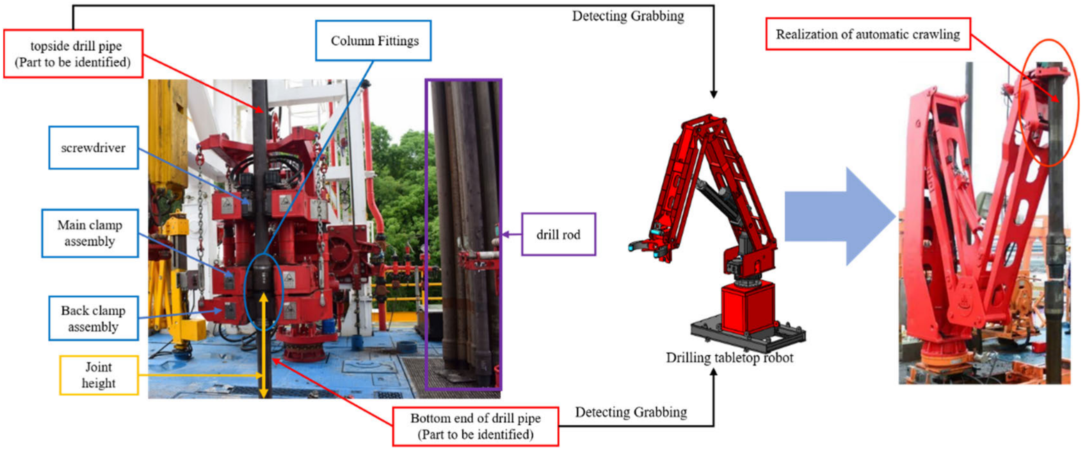
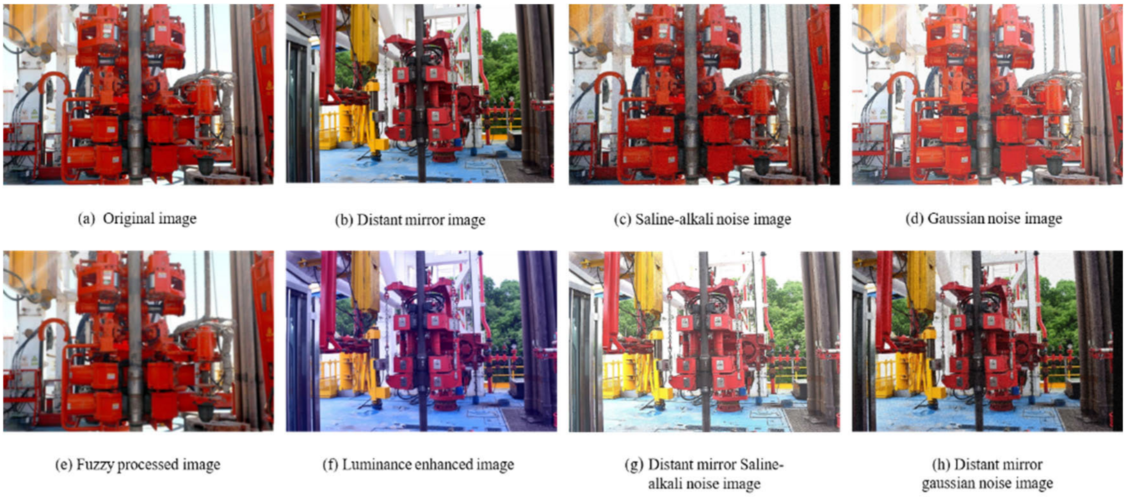


FIGURE 14. Drill pipe inspection-gripping flowchart.


(a) Original image

(b) Distant mirror image

(c) Saline-alkali noise image

(d) Gaussian noise image

(e) Fuzzy processed image

(f) Luminance enhanced image

(g) Distant mirror Saline-alkali noise image

(h) Distant mirror gaussian noise image

FIGURE 15. Image processing diagram.

insufficient training due to a lack of sample images. In total, 2009 images with a resolution of 4032 × 3024 were obtained. The processed images of some of the datasets are shown in Figure 15.

2) DATA SET LABELING

Label the 2009 drill pipe images in the dataset using Labelimg labeling software, using horizontal rectangular boxes to label the drill pipes in the images individually. Save the labeling information in the text format. Finally, the dataset images are randomly divided into 1406 images for the training set and 603 images for the validation set according to the ratio of 7:3, the specific dataset directory structure is shown in Figure 16.

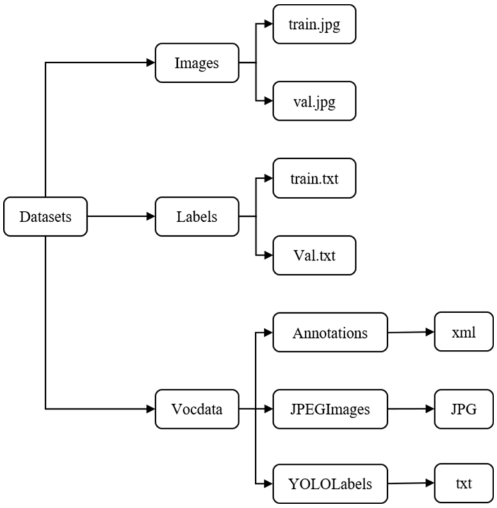


FIGURE 16. Data set directory structure.

C. ANALYSIS OF MODEL TRAINING AND VALIDATION

Figure 17 and Figure 18 present the bounding box regression loss and confidence loss curves for the training and validation sets of YOLOv5s and ECS-YOLOv5s. As shown in Figure 1, the loss function curves for each part of the loss function exhibit a clear declining trend with each iteration, and when the iteration reaches 500, the value of each loss is reduced and

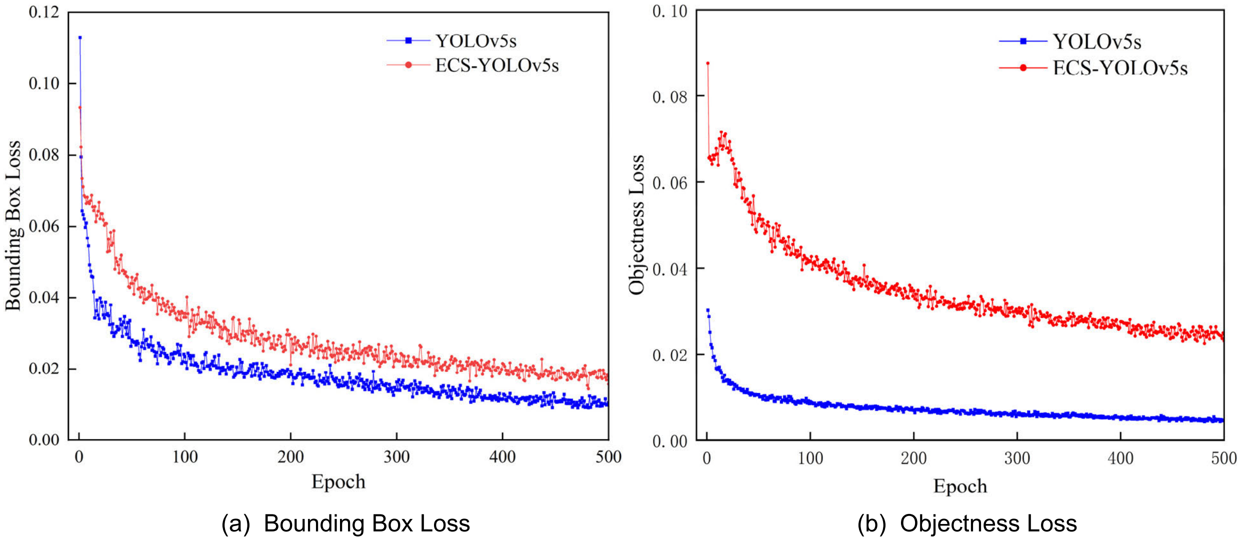


FIGURE 17. Comparison of train loss curves for the YOLOv5s with ECS-YOLOv5s.

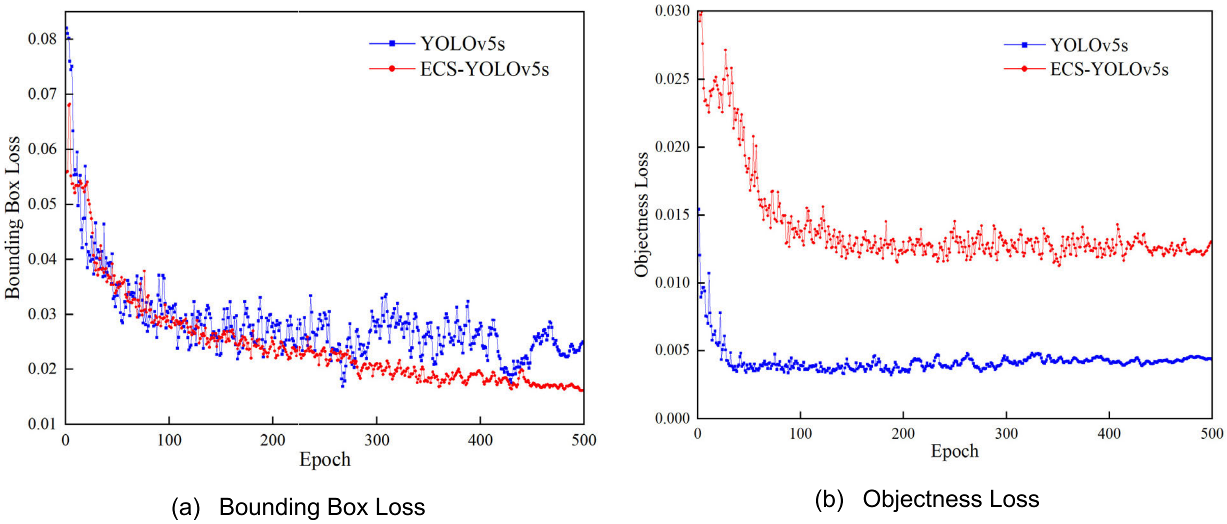


FIGURE 18. Comparison of validation loss curves for the YOLOv5s with ECS-YOLOv5s.

tends to stabilize. Specifically, the bounding box regression loss values for the training and validation sets of YOLOv5s and ECS-YOLOv5s are 0.0105, 0.0247, and 0.0157, 0.0161, respectively, when the iteration is 500 times. It is worth noting that the bounding box loss of ECS-YOLOv5s is smaller, indicating that the model's prediction of the target bounding box position is more accurate than that of YOLOv5s. Regarding the confidence loss, when iterating 500 times, the confidence loss values for the training and validation sets of YOLOv5s and ECS-YOLOv5s are 0.0047, 0.0044, and 0.0231, 0.0130, respectively. Again, the confidence loss of ECS-YOLOv5s is smaller, indicating the model's high target detection accuracy. The above analysis confirms that the detection accuracy of ECS-YOLOv5s surpasses that of YOLOv5s, making it a superior performer.

VI. EXPERIMENTAL VALIDATION ANALYSIS

A. EXPERIMENTAL ANALYSIS OF LIGHTWEIGHT NETWORK ABLATION

The research paper proposes a drill pipe identification model, which is divided into two groups - a lightweight network group and a traditional network group. The model is subjected to ablation experiments, where certain features are removed, and model hierarchy and parameter adjustments are made for each group. The study employs an improved multi-scale feature fusion network BiFPN architecture, which replaces the backbone network in the Backbone layer of YOLOv5s with the backbone network of EfficientNet. Furthermore, the SPPF module and the CBAM attention module are introduced. The performance of the lightweight network for drill pipe identification is evaluated using metrics

TABLE 4. Ablation study of different components in ECS-YOLOv5s model.

Model	SPPF	CBAM	EfficientNet	Parameters/×10^6	Precision(%)	Recall(%)	mAP (%)
YOLOv5s	×	×	×	7.06	89.4	86.2	88.4
YOLOv5s-SPPF	√	×	×	7.22	92.3	89.1	88.2
SC- YOLOv5s	√	√	×	7.23	90.2	87.2	86.3
E- YOLOv5s	×	×	√	7.12	89.42	78.56	80.64
SE- YOLOv5s	√	×	√	4.44	93.7	78.7	84.6
ECS- YOLOv5s	√	√	√	**4.48**	**98.6**	**84.6**	**90.2**

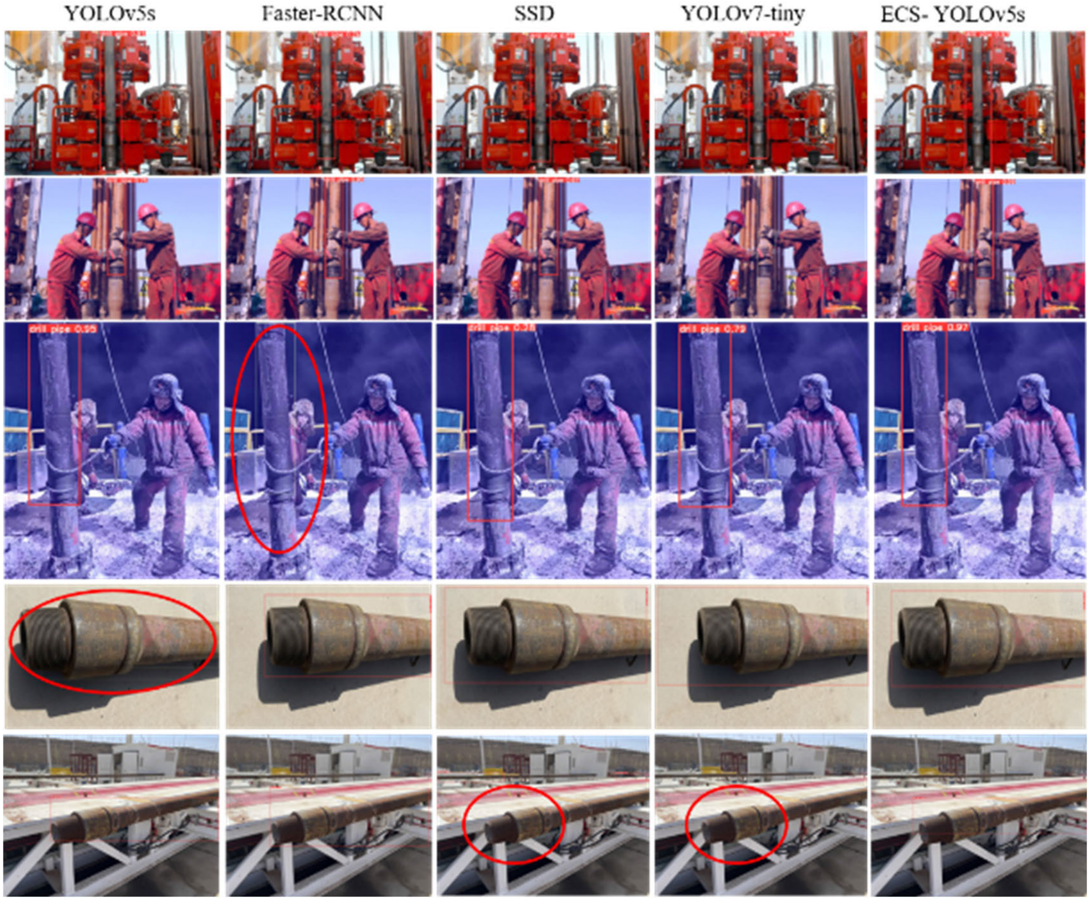


FIGURE 19. Recognition effects of different models on drill pipe.

such as accuracy, recall, and number of parameters. Additionally, the models are compared based on their training time, memory occupation, and computational resource consumption. The experimental results are presented in Table 4.

From the experimental results, we can observe that the revised model has significantly improved the recognition accuracy. The recognition accuracy has increased from 89.4% to 98.6% compared to the original model. Table 3 presents the results of the ablation experiments, which concluded

TABLE 5. AYOLOv5 and other network detection performance comparison.

Model	Parameters/×10^6	Computation/GFLOPs	Model size/MB	mAP (%)	FPS
YOLOv5s	7.06	15.9	14.4	84.5	**156**
Faster-RCNN	28.12	946.12	108.2	78.8	12
SSD	23.64	274.53	88.6	82.5	38
YOLOv7-tiny	6.01	**13.2**	**12.3**	86.2	105
ECS- YOLOv5s	**4.48**	44.3	90.4	**90.2**	125

that replacing the EfficientNet model's backbone network with the original model of YOLOv5s' backbone network can reduce the model's parameters and computation to a larger extent. This can also reduce the size of the weight file generated by the model while maintaining the mAP enhancement and slightly reducing the recall. The SPPF module introduces spatial pyramid pooling to extract features of different scales, thereby improving the model's ability to adapt to different sizes of drill pipes and enhancing the detection accuracy. The CBAM module introduces the channel attention mechanism to enhance the useful features and inhibit the useless ones. This improves the feature expression ability and robustness of the model, leading to improved recognition of drill pipes in different situations. The backbone network of the EfficientNet model uses depth-separable convolution, which effectively reduces the number of parameters and computation volume generated by the model while ensuring the lightweight of the network and high recognition accuracy. The introduction of modules such as SPPF, CBAM, and EfficientNet improves the model's performance and accuracy. However, this increases the amount of computation, which needs to be weighed against the performance to select the appropriate network model and technique according to the specific task and resource constraints. The ablation test results show that the recall decreases slightly after the lightweight improvement of the model. Lightweight treatment reduces the complexity of the network to a certain extent, which reduces the model's ability for network feature extraction. However, in this study, only one category, drill pipe, is recognized, and recall is relatively less important in drill pipe recognition. Precision and average precision are more important. Therefore, the effect of a slight decrease in the recall can be ignored on the premise that both precision and average precision have been improved. The primary objective is to achieve a lightweight treatment of the model. The above analysis shows that the performance of the lightweight YOLOv5 network in drill pipe identification detection by adding the attention mechanism can prove that the improved YOLOv5s method in this research is reasonable and correct.

B. IDENTIFICATION RESULTS

Several sets of data were set up to be detected with different models, and the results are shown in Figure. 19. Among the algorithms, YOLOv5s, Faster-RCNN, SSD, and YOLOv7-tiny dun algorithms have missed detections. The missed drill pipe has been marked with red ellipses in the figure. The improved algorithm has targeted miss-detection cases. It shows that the improved algorithm has successfully improved the performance of detection.

To conduct a comparative analysis of the improved ECS-YOLOv5s network model with different algorithm models, and to investigate the performance of various algorithms, this study selected the current mainstream target detection algorithms, such as YOLOv5s, Faster RCNN, SSD, and YOLOv7-tiny, for conducting comparative experiments. The results of the experiment are presented in Table 4, which highlights the superiority of the improved algorithms over the existing ones.

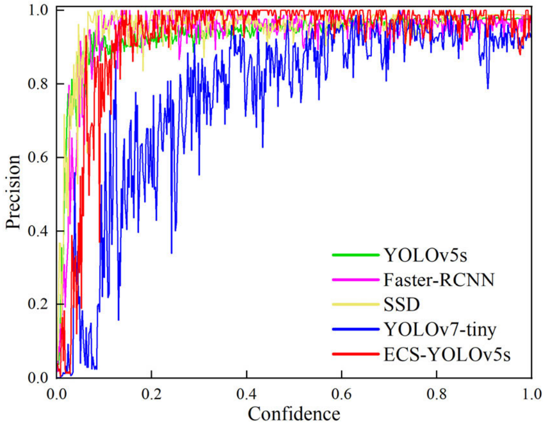


FIGURE 20. ESC-YOLOv5s precision.

Table 5 combined with Figure. 20 shows that based on the evaluation criteria, the traditional YOLOv5s model appears to be a practical option for lightweight application settings due to its high accuracy and smaller number of parameters and computation. However, the Faster-RCNN network model has a relatively poor detection accuracy despite having a higher number of parameters and computations, generating a larger weight file size. On the other hand, the SSD network model exhibits improved detection accuracy with decreased parameters and processing and a smaller weight size compared to Faster-RCNN. Although the YOLOv7-tiny network model has the smallest computational effort, it has poor accuracy and is therefore not suitable for the drill pipe recognition task. Meanwhile, the ECS-YOLOv5s model has the highest recognition accuracy and optimal mAP score, but its number of parameters, computational volume, and weight file size is higher compared to the traditional YOLOv5s model. Nonetheless, the improved network model ECS-YOLOv5s has a detection speed that only slightly differs from the original YOLOv5s model and performs better than all other models in terms of real-time detection requirements. Considering the requirements of the drill pipe application scenario, the ESC-YOLOv5s network model may be adopted to balance accuracy and computational efficiency.

VII. CONCLUSION

The proposed ECS-YOLOv5s algorithm is a new lightweight target detection algorithm that is specifically designed for the automatic detection of tubular columns in intelligent oil drilling platforms. The algorithm combines several advanced techniques, including SPPF, CBAM, EfficientNet, and an improved multiscale feature fusion network, BiFPN, based on the traditional YOLOv5s model. This approach enhances the model's ability to detect dense targets and improves its accuracy in complex environments with varying angles, distances, different light conditions, different orientations of the drill pipe, and different operating conditions.

The experimental comparative analysis shows that the ECS-YOLOv5s model outperforms the traditional YOLOv5s model, the SSD model, the Faster-RCNN model, and the YOLOv7-tiny model in both performance and accuracy. The model achieved a mAP of 90.2%, an accuracy of 98.6%, and a frame rate of 125 FPS. Furthermore, the number of parameters is only 37% of the traditional model, which demonstrates the algorithm's lightweight, high efficiency, and high accuracy.

Despite the promising results, the study acknowledges that the current research is limited by the size and quality of the dataset. Therefore, future research will focus on enhancing the model's adaptability to complex backgrounds, light changes, and occlusion, among other factors. It is also necessary to consider the processing of large-scale data and ensure real-time performance while maintaining high detection accuracy. The study provides valuable insights into the automatic tubular column detection in smart oil rigs and provides a reference for further research in this field.

REFERENCES

[1] L. Gensheng, S. Xianzhi, and T. Shouwei, "Research status and development trend of intelligent drilling technology," *Oil Drilling Technol.*, vol. 48, no. 1, pp. 1–8, 2020.

[2] R. Zhong, C. Salehi, and R. Johnson, "Machine learning for drilling applications: A review," *J. Natural Gas Sci. Eng.*, vol. 108, Dec. 2022, Art. no. 104807.

[3] W. Haige, H. Hongchun, and B. Wenxin, "Progress and prospect of oil and gas drilling technology in deep wells and ultra-deep wells," *Natural Gas Ind.*, vol. 41, no. 8, pp. 163–177, 2021.

[4] Z. Xiaoming, *Research on Visual Localization Technology of Slender Pipe Columns Based on Three-Dimensional Point Clouds*. Wuhan, China: Huazhong Univ. Science and Technology, 2021.

[5] T. H. Fan, Y. N. Gu, and W. B. Wang, "Lightweight honeysuckle recognition method based on improved YOLOv5s," *J. Agricult. Eng.*, vol. 39, no. 11, pp. 192–200, 2023.

[6] W. Gu and K. Sun, "AYOLOv5: Improved YOLOv5 based on attention mechanism for blood cell detection," *Biomed. Signal Process. Control*, vol. 88, Feb. 2024, Art. no. 105034.

[7] Y. Z. Huang, L. Han, and X. Yang, "Enhanced batch sorting and rapid sensory analysis of mackerel products using YOLO5vs algorithm and CBAM: Validation through TPA, colorimeter, and PLSR analysis," *Food Chem., X*, vol. 19, Jun. 2023, Art. no. 100733.

[8] Z. Jia, M. Fu, X. Zhao, and Z. Cui, "Intelligent identification of metal corrosion based on corrosion-YOLOv5s," *Displays*, vol. 76, Jan. 2023, Art. no. 102367.

[9] T. Li, M. Sun, Q. He, G. Zhang, G. Shi, X. Ding, and S. Lin, "Tomato recognition and location algorithm based on improved YOLOv5," *Comput. Electron. Agricult.*, vol. 208, May 2023, Art. no. 107759.

[10] J. Li, Y. Qiao, S. Liu, J. Zhang, Z. Yang, and M. Wang, "An improved YOLOv5-based vegetable disease detection method," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107345.

[11] S. Li, C. Li, Y. Yang, Q. Zhang, Y. Wang, and Z. Guo, "Underwater scallop recognition algorithm using improved YOLOv5," *Aquacultural Eng.*, vol. 98, Aug. 2022, Art. no. 102273.

[12] C. Yu and Y. Shin, "SAR ship detection based on improved YOLOv5 and BiFPN," *ICT Exp.*, Mar. 2023.

[13] C. Yuan, T. Liu, F. Gao, R. Zhang, and X. Seng, "YOLOv5s-CBAM-DMLHead: A lightweight identification algorithm for weedy Rice (*Oryza sativa* f. Spontanea) based on improved YOLOv5," *Crop Protection*, vol. 172, Oct. 2023, Art. no. 106342.

[14] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[16] S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2016, doi: 10.1109/tpami.2016.2577031.

[17] B. Alexey, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detectio," 2020, *arXiv:2004.10934*.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[19] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[20] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[21] W. Liu, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.

[22] B. Alexey, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[23] L. Huang, C. Zhang, and H. Zhang, "Self-adaptive training: Beyond empirical risk minimization," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2020, pp. 19365–19376.

[24] C.-B. Zhang, "Delving deep into label smoothing," *IEEE Trans. Image Process.*, vol. 30, pp. 5984–5996, 2021.

[25] Q. Li, Y. Shi, R. Lin, W. Qiao, and W. Ba, "A novel oil pipeline leakage detection method based on the sparrow search algorithm and CNN," *Measurement*, vol. 204, Nov. 2022, Art. no. 112122.

[26] P. Lin, H. Yang, S. Cheng, F. Guo, L. Wang, and Y. Lin, "An improved YOLOv5s method based bruises detection on apples using cold excitation thermal images," *Postharvest Biol. Technol.*, vol. 199, May 2023, Art. no. 112280.

[27] Y. Lin, T. Chen, S. Liu, Y. Cai, H. Shi, D. Zheng, Y. Lan, X. Yue, and L. Zhang, "Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning," *Comput. Electron. Agricult.*, vol. 197, Jun. 2022, Art. no. 106938.

[28] A. Lu, R. Guo, Q. Ma, L. Ma, Y. Cao, and J. Liu, "Online sorting of drilled lotus seeds using deep learning," *Biosystems Eng.*, vol. 221, pp. 118–137, Sep. 2022.

[29] C. Qiumeng, *Target Detection and Ranging System Based on Binocular Vision*. Daqing, China: Northeast Petroleum Univ., 2022.

[30] G. Chuang, S. Pinde, and C. Lijia, "Research progress of YOLO algorithm in target detection," *J. Weapon Equip. Eng.*, vol. 43, no. 9, pp. 162–173, 2022.

[31] J. Xin, Z. Jianjun, and X. Ziheng, "Lightweight YOLOv5s network under-vehicle hazard recognition algorithm," *J. Zhejiang Univ. Eng. Ed.)*, vol. 57, no. 8, pp. 1516–1526+1561, 2023.

[32] L. Yanzhou, H. Yanzhou, and Q. Feng, "Convolutional neural network-based recognition of M. Intermedia," *Chin. J. Agricult. Mech. Chem.*, vol. 44, no. 4, pp. 159–166, 2023.

[33] H.-T. Zhang, C.-J. Zhao, and C.-Y. Wang, "Research on the detection method of wheat ears in the field based on convolutional neural network," *J. Wheat Crops*, vol. 43, no. 6, pp. 798–807, 2023.

[34] H. Jie et al., "Sprout eye detection method for seed potato based on improved YOLOv5s," *Trans. Chin. Soc. Agricult. Eng.*, vol. 39, no. 9, pp. 172–182, 2023.

[35] Z. Haimin et al., "Rapid grading detection of hybrid Rice shoot seeds based on YOLOv5 improved model," *J. South China Agricult. Univ.*, vol. 44, no. 6, pp. 960–967, 2023.

[36] X. Zhu, F. Chen, and Y. Zheng, "Integration of bilinear network and attention mechanism for olive variety identification," *J. Agricult. Eng.*, vol. 39, no. 10, pp. 183–192, 2023.

[37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[38] P. Luo, X. Zhang, and Y. Wan, "Lightweight YOLOv5 model based small target detection in power engineering," *Cognit. Robot.*, vol. 3, pp. 45–53, Jan. 2023.

[39] B. Mahaur and K. K. Mishra, "Small-object detection based on YOLOv5 in autonomous driving systems," *Pattern Recognit. Lett.*, vol. 168, pp. 115–122, Apr. 2023.

[40] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[41] M. Ren, X. Zhang, X. Chen, B. Zhou, and Z. Feng, "YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 120, Jun. 2023, Art. no. 103335.

[42] S. Woo, J. Park, and J. Y. Lee, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[43] Y. Shang, X. Xu, Y. Jiao, Z. Wang, Z. Hua, and H. Song, "Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments," *Comput. Electron. Agricult.*, vol. 207, Apr. 2023, Art. no. 107765.

[44] L. Shen, J. Su, R. He, L. Song, R. Huang, Y. Fang, Y. Song, and B. Su, "Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s," *Comput. Electron. Agricult.*, vol. 206, Mar. 2023, Art. no. 107662.

[45] C. Wang, G. Liu, Z. Yang, J. Li, T. Zhang, H. Jiang, and C. Cao, "Downhole working conditions analysis and drilling complications detection method based on deep learning," *J. Natural Gas Sci. Eng.*, vol. 81, Sep. 2020, Art. no. 103485.

[46] P. Wu, H. Weng, W. Luo, Y. Zhan, L. Xiong, H. Zhang, and H. Yan, "An improved YOLOv5s based on transformer backbone network for detection and classification of bronchoalveolar lavage cells," *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 2985–3001, Jan. 2023.

[47] J. Xu, J. Ye, S. Zhou, and A. Xu, "Automatic quantification and assessment of grouped pig movement using the XGBoost and YOLOv5s models," *Biosyst. Eng.*, vol. 230, pp. 145–158, Jun. 2023.

[48] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6568–6577.

[49] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, p. 531, Mar. 2019.

[50] Y. Chen, X. Dai, D. Chen, M. Liu, X. Dong, L. Yuan, and Z. Liu, "Mobile-former: Bridging MobileNet and transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5260–5269.

[51] Y. Yang and X. Wang, "Recognition of bird nests on transmission lines based on YOLOv5 and DETR using small samples," *Energy Rep.*, vol. 9, pp. 6219–6226, Dec. 2023.

[52] Z. Ying, Z. Lin, Z. Wu, K. Liang, and X. Hu, "A modified-YOLOv5s model for detection of wire braided hose defects," *Measurement*, vol. 190, Feb. 2022, Art. no. 110683.

[53] P. Zhang and D. Li, "Automatic counting of lettuce using an improved YOLOv5s with multiple lightweight strategies," *Expert Syst. Appl.*, vol. 226, Sep. 2023, Art. no. 120220.

[54] Y. Zhang, Y. Yang, J. Sun, R. Ji, P. Zhang, and H. Shan, "Surface defect detection of wind turbine based on lightweight YOLOv5s model," *Measurement*, vol. 220, Oct. 2023, Art. no. 113222.

[55] A. Hosseiny and H. Jahanirad, "Hardware acceleration of YOLOv7-tiny using high-level synthesis tools," *J. Real-Time Image Process.*, vol. 20, no. 4, p. 75, Aug. 2023.

[56] G. Zhao, R. Yang, X. Jing, H. Zhang, Z. Wu, X. Sun, H. Jiang, R. Li, X. Wei, S. Fountas, H. Zhang, and L. Fu, "Phenotyping of individual apple tree in modern orchard with novel smartphone-based heterogeneous binocular vision and YOLOv5s," *Comput. Electron. Agricult.*, vol. 209, Jun. 2023, Art. no. 107814.

[57] S. Zhao, F. Kang, and J. Li, "Concrete dam damage detection and localisation based on YOLOv5s-HSC and photogrammetric 3D reconstruction," *Autom. Construct.*, vol. 143, Nov. 2022, Art. no. 104555.

BIN PENG received the Ph.D. degree in mechanical engineering from the Lanzhou University of Technology, Lanzhou, China, in 2007. From 2008 to 2009, he was sponsored by the China Scholarship Council to go to the Herrick Laboratory, Purdue University, USA, for one year of exchange study. From 2014 to 2015, he was sponsored by the China Scholarship Council to go to the Thermal Power Laboratory, University of Liège, Belgium, for one year of exchange study. He is currently a Professor with the Lanzhou University of Science and Technology, and a Flying Scholar Distinguished Professor in Gansu Province. His current research interests include modern design methods, theory, and applications; low-temperature waste heat power generation; theory and application of vortex machinery; rotor dynamics, and fault diagnosis.

KE NIU is currently pursuing the M.S. degree in mechanical engineering with the Lanzhou University of Technology. From 2021 to 2022, he is engaged in Technical Research and Development at Lanzhou Power Vehicle Research Institute, China, Machinery Industry Corporation (CMIC). In September 2023, he studied at the laboratory of the Shenzhen Institute of Artificial Intelligence and Robotics (AIRS). He is also participated in research projects with the group of The Chinese University of Hong Kong, Shenzhen. His current research interests include computer vision and deep learning.