ISSN: 2688-8645

Research on Safe Wearable Target Detection Technology for Power Production Scenarios based on YOLOv5-C3CA

Chiyi Ma, Xingzhong Xiong, Jun Liu

Sichuan University of Science and Engineering, Sichuan 643000, China

Abstract

Aiming at the establishment of a new generation of substation auxiliary equipment detection platform intelligence, this study proposes a power production industry personnel safety wear detection technology based on improved Yolov5-C3CA to meet the demand. First, the network performance is improved by adding the CA attention mechanism module to Yolov5s network; moreover, the CA module is modified to C3CA module and added to the model for more effective addition of the attention mechanism; finally, the AFPN network structure is used to replace the original feature pyramid network structure, which further effectively improves the utilization efficiency of shallow and deep features. The experimental results show that the mean average accuracy of the designed network is improved by 3.9% to 93.1% compared with the original network, and other evaluation indexes are also improved. It can be seen that the model modification in this paper has improved the performance of the detection network, and the improved network meets the requirements of the new generation of substation auxiliary equipment detection platform, which has a positive effect on the safety and reliability of the power production industry.

Keywords

Target Detection; YOLOv5; Attention Mechanism; AFPN.

1. Introduction

With the continuous development of the power system, the construction of the new generation of substations needs to integrate more artificial intelligence technology. And the safety of power production scene has always been a hot issue of concern in the power industry, and the detection of safety wear has an important research significance for guaranteeing the safety of the personnel in the power industry and the safe operation of the power production scene. Therefore, this paper designs a target detection technology based on YOLOv5-C3CA for safety wear in power production scene.

The current rapid development of deep learning in natural language processing[1] image recognition[2] image segmentation[3] and image segmentation, etc. It has shown its excellent performance in many fields. It also improves a new path for safety wear detection based on deep learning. This study is oriented to the problem of safe wear in power production scenarios, and proposes the research of safe wear based on YOLOv5 improved network target detection technology, aiming at solving the problem of safety hazards in the power production industry, and improving the safety of the power production industry . Image-based target detection technology has been widely researched in the power industry, and the main methods are traditional image processing and machine learning[4] . Wang [5] inspection of high voltage transmission towers by yolov3 in a home-made high voltage transmission tower dataset, Kang[6] detects the wearing of helmets of personnel in the electric power industry through yolov4 in the homemade dataset. In order to enhance the detection model capability, many scholars have also proposed various improvements to the detection framework, Souza[7]

ISSN: 2688-8645

detects transmission lines through YOLOv5 and then classifies the detected objects through ResNet-18 classifier, which gives better F1 scores than the results of the original model, and deploys the designed model to UAVs. Qi[8] For the problem of dense and small targets of safety helmets, the SSP (Spatial pyramid pooling) module is optimized based on YOLOv5, a shallow fusion channel is added to the fusion layer, and finally the CA attention mechanism is added to the image input module so that the detection network focuses on the positional information of the target object earlier. Although the above improved algorithm involves many directions for safety detection in the power industry, the current research content for personnel safety wear is mainly focused on the detection of safety helmets, and the research on the detection of other safety wear is relatively small.

Therefore, this paper proposes a more complete detection method for the safety wear of the personnel in the power industry. This study realizes the safety and security of the personnel in the power industry through safety wear detection. The main work and contribution of the study includes:

1. A new secure wearable dataset is collected and organized for training the network. The dataset consists of a total of 6,641 sheets with the labels "helmet", "boot", and "reflective vest", of which the number of helmet labels is 12,226; the number of boot labels is 2,081; and the number of reflective vest labels is 4,171.

2. In order to enhance the network detection capability, this design adds CA attention module to the backbone network and allocates attention resources to critical areas, thus reducing the background interference in complex environments.

3. In order to better combine the CA attention mechanism, this design combines the CA attention mechanism to design the C3CA module to replace the C3 module in the YOLOv5 network, which improves the detection capability of the network.

4. In order to enhance the network ability to detect small targets, this design replaces the feature fusion layer with AFPN fusion network to realize the enhancement of small target detection ability.

2. Yolov5 Target Detection Algorithm

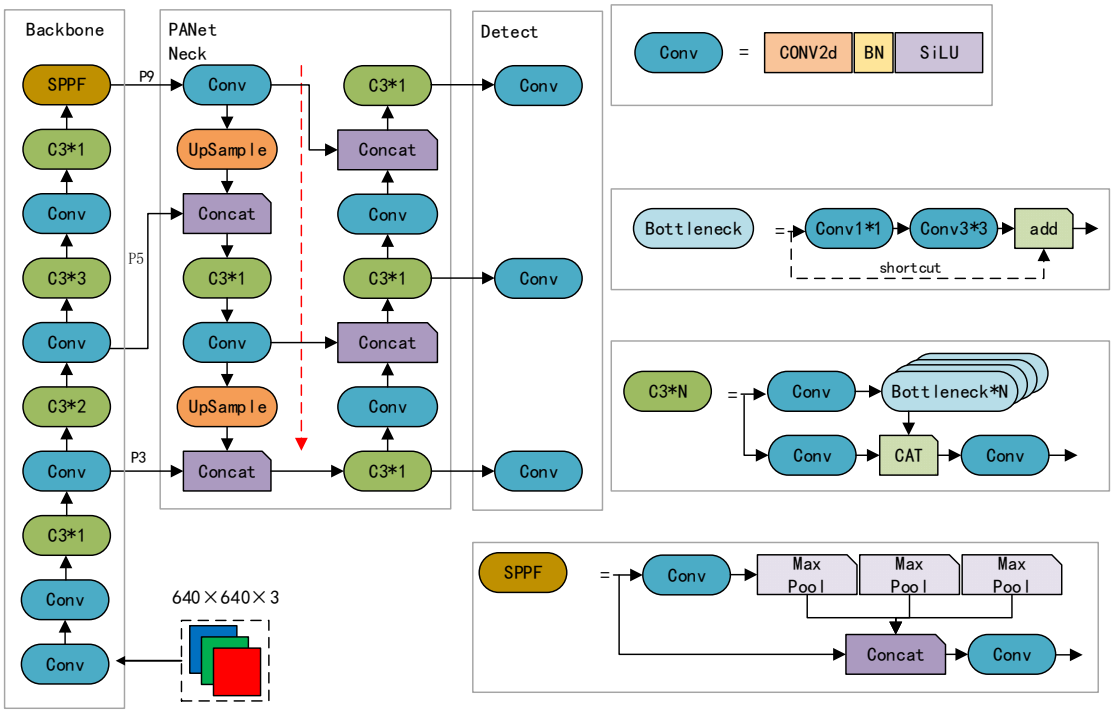


Figure 1. YOLOv5 network architecture diagram

ISSN: 2688-8645

Yolov5 is a deep learning based target detection algorithm based on the YOLO (You Only Look Once) family.[12,11,10,9] (You Only Look Once) family of target detection algorithms, which is a deep learning based detection algorithm.Yolov5 adopts a lightweight architecture with higher speed and accuracy. It mainly consists of Input, Backbone and Head. The Input segment employs Mosaic operation, Adaptive Anchor Frame, and Adaptive Picture Size in three ways to enhance the data. Backbone network is simplified to be spliced by three modules including Conv, C3, SPPF (Spatial Pyramid Pooling).

The SPPF module is a modification of the SPP[13] (Spatial Pyramid Pooling) improvement. The detection head part includes a neck feature fusion part and a detection part. The feature fusion layer adopts the PANet[14] (Path Aggregation Network) network structure, which is based on the FPN[15] (Feature Pyramid Network) network, which is based on FPN and adds a bottom-up path. The structure of Yolov5 network is shown in Figure 1.

3. Module Improvements

3.1. Introduction of Attention Mechanisms

Due to the complexity of the power production environment and the background interference, this paper chooses to opt for the inclusion of the attention mechanism to improve the target detection effect. Attention modules commonly used in the feature backbone network layer and feature fusion network layer are: SE[16] (Squeeze and excitation), CBAM[17] (Convolutional block attention module), BAM[18] GAM (Global Attention Mechanism).[19] (Global Attention Mechanism), CA[20] (Coordinate attention), ECA-Net[21] (CA).

Where CA positional attention decomposes channel attention into two 1-dimensional features by aggregating features along 2 spatial directions respectively. It captures remote dependencies along one spatial direction, while it can retain precise location information along the other spatial direction. The generated feature maps are then encoded as a pair of direction-aware and location-sensitive ATTENTION MAPS, respectively, which can be applied complementarily to the input feature maps to enhance the representation of the object of attention. In this paper, we choose to use the CA attention module, as Figure 2 shown.

The X-scale of the input feature map is C(the number of channels of the input feature map)*H(the height of the input feature map)*W(the width of the input feature map), firstly, the global average pooling from the direction of the height and width of the two dimensions to get the $Z_c^h \in \mathbb{R}^{C*H*1}$ and $Z_c^w \in \mathbb{R}^{C*1*W}$ feature maps; and then, the Z_c^w feature map is permuted, and the Z_c^h spliced together, and then the 1*1 convolution kernel is used for the downscaling to get the $f \in \mathbb{R}^{C/r*(H+W)*1}$ feature map, which also includes the nonlinear operation to improve the nonlinear expression capability. Enhance the nonlinear expression ability; finally, along the spatial dimension, the f feature map is split into $f^h \in \mathbb{R}^{C/r*H*1}$ and $f^w \in \mathbb{R}^{C/r*1*W}$, and then the 1*1 convolutional transform F_h and F_w are used to carry out the dimensionality upgrading operation, and the final attention vectors $g^h \in \mathbb{R}^{C*H*1}$ and $g^w \in \mathbb{R}^{C*1*W}$ are obtained through the activation function sigmoid, as(1) The final attention vectors and are obtained through the activation function sigmoid, as shown.

$$g^h = \sigma(F_h(f^h))$$
$$g^w = \sigma(F_w(f^w)) \quad (1)$$

where σ is the sigmoid function. The final CA output equation is

$$y_c(i,j) = x_c(i,j) * g_c^h(i) * g_c^w(j) \quad (2)$$

ISSN: 2688-8645

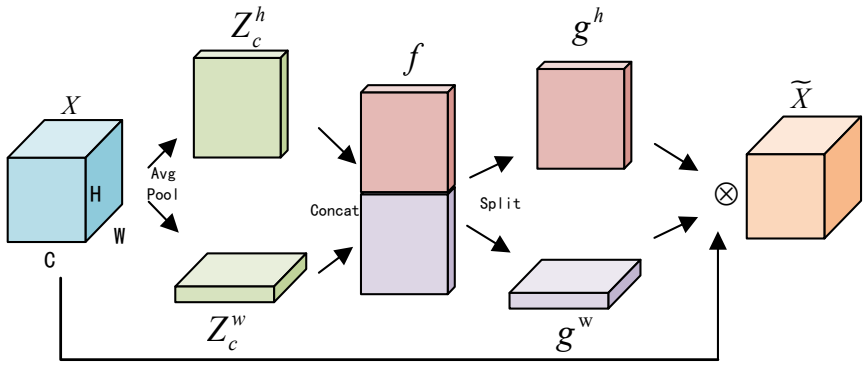


Figure 2. Flowchart of the CA network

3.2. Modification of the C3CA Backbone Network Module

In order to better incorporate the attention mechanism, this paper replaces the C3 module in the backbone network with the C3CA module with attention mechanism, and the improvement effect diagram of this module is as follows Figure 3 shown. Replacing the Bottleneck module in C3 with CA Bottleneck module, the idea of CA Bottleneck module comes from CA attention module. The most advanced part of its module idea is to convert the traditional global pooling operation into two one-dimensional feature tensors by dividing it more carefully into horizontal pooling and vertical pooling.

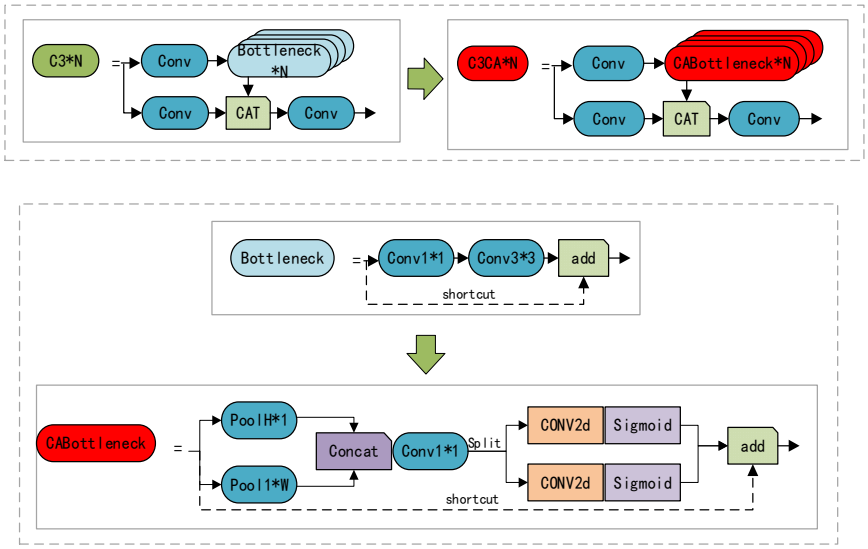


Figure 3. C3CA module

3.3. Asymptotic Characteristic Pyramid Network

In YOLOv5 algorithm, the feature pyramid structure is located in the Neck fusion layer part.YOLOv5 previously used FPN feature pyramid as in Figure 4 (1) shows, its a technique used to deal with multi-scale target detection because the size and position of the object in the image is uncertain so a mechanism is needed to deal with targets of different scales and sizes.

Although this fusion mechanism improves the use and fusion of features, it does not learn the features with larger contributions for fusion. In order to better fuse the feature information at each scale Liu[22] proposed the PANet structure, whose main idea is to add a bottom-up fusion path on the basis of the FPN network, and its structure is as follows Figure 4 (2) shows

Although PANet obtained better accuracy than FPN fusion network, but at the cost of more parameters and computation, in order to improve the model , Tan[23] proposed the design of

ISSN: 2688-8645

a weighted bidirectional feature pyramid network (Bi FPN), which allows for simple and fast multiscale feature fusion with a unified composite scaling method for scaling the resolution, depth, and width of all the backbone, feature networks, and box and class prediction networks. For multiscale fusion, the authors of propose a simple and efficient weighted (similarly to ATTENTION) bi-directional feature pyramid network (BiFPN), which introduces learnable weights to learn the importance of different input features, while iteratively applying top-down and bottom-up multiscale feature fusion. Its network structure is shown in Figure 4 (3) shows.

In the process of feature fusion, the semantic information between non-adjacent scale features is not effectively fused, so Yang[24] proposed AFPN (Asymptotic Feature Pyramid Network) fusion network, which improves the semantic fusion ability of the network by iteratively fusing low-level and high-level features to generate richer feature maps. The network structure is shown in Figure 4 (4) shows . The comparative experiments show that Table 6, in this paper, we choose to use AFPN fusion network for model modification.

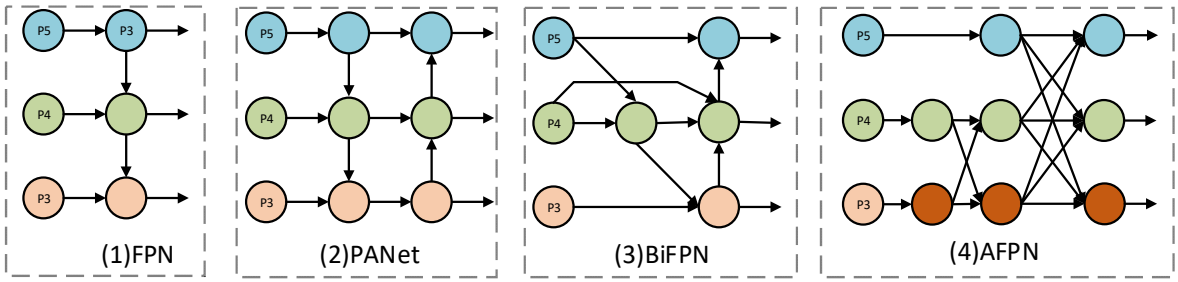


Figure 4. Schematic diagram of the structure of the feature fusion network

4. Experimentation and Analysis

4.1. Data Set

Due to the limited number of open source datasets for this type of scenario and the limited number of labeling categories, this paper uses both web-collected and autonomously-collected datasets to build the Safety wear Dataset. It contains 5427 web-collected images and 1214 self-collected images, totaling 6641 images. And LabelImg is used to label the self-made dataset in YOLO format, in which there are three labels, namely: "helmet", "boot", "reflective vest", in which the number of helmet labels is 12,226; the number of boot labels is 2,081; the number of reflective vest labels is 4171.

4.2. Evaluation Criteria for Experimental Equipment and Network Model Parameters

4.2.1. Experimental Equipment

The hardware and frame parameters of this experimental server are as follows Table 1 shows.

Table 1. Server parameters

manipulate systems	frame	CPU	random access memory (RAM)	GPUs	display memory
Ubuntu 18.04.5	Pytorch 1.13.0	Intel(R) Xeon(R) Platinum 8350C	42G	RTX 3080 Ti	12GB

ISSN: 2688-8645

Table 2. Network training parameters

imgsz	epochs	lr0	lrf	batch-size
640×640	100	0.01	0.01	16

4.2.2. Network Parameter Setting

The YOLOv5-based network training hyperparameters are set as in Table 2 is shown. where imgsz denotes the size of the training images; epochs denotes the number of network training rounds; lr0 denotes the initial learning rate of network training; lrf denotes the periodical learning rate of network training; and batch-size denotes the number of images that are passed into the network training each time.

4.2.3. Evaluation Criteria

In order to evaluate the performance of the model and the accuracy of image detection for safety wearable devices, this paper adopts mean average precision (mAP), precision (P), and recall (R) as the evaluation metrics. The formulas are shown below.

$$\begin{cases} Precision(P) = \dfrac{TP}{TP + FP} \\ AP = \int_0^1 P(R)dR \\ Recall(R) = \dfrac{TP}{TP + FN} \\ mAP = \dfrac{\sum_{i=1}^{m} AP_i}{m} \end{cases} \tag{3}$$

where TP (true positives) denotes the amount of positive categories judged as positive, i.e., correctly predicted in the model; FP (false positives) denotes the amount of negative categories judged as positive, i.e., incorrectly predicted in the model; FN (false negative) denotes the amount of positive categories judged as negative, i.e., the amount of otherwise positive samples that were incorrectly detected in the model; AP is the area enclosed by the P-R curve (P is the vertical coordinate and R is the horizontal coordinate), and mAP is the mean value of the average precision AP across all categories, where m is the number of detected categories.

4.2.4. Model Selection

There are four choices of YOLOv5 algorithmic models: the YOLOv5s, YOLOv5n, YOLOv5m, and YOLOv5l, of which all four are based on the extension of the YOLOv5s model. According to the experiments obtained Table 3, this design chooses to use YOLOv5s model as the base network for network improvement , and all the following use YOLOv5s model as the base model.

4.3. Experiments to Introduce Attention Mechanisms

4.3.1. Backbone Network

By adding different attention mechanisms into the YOLOv5s backbone network, we get Table 4 Experimental results, through the experimental results, it can be seen that the introduction of the attention mechanism has a great improvement on the performance of the detection network, and there is a significant improvement in the results of the correct rate, the recall rate and the mean evaluation precision. And it is found that the referencing effect of CA positional attention mechanism is better than the referencing effect of attention such as SE, CBAM, GAM, etc. Therefore, in this paper, the CA attention mechanism is referenced to effectively improve the network's anti-jamming ability as well as to improve the network's feature extraction ability.

ISSN: 2688-8645

Table 3. Results of yolov5 model comparison experiments

Model	PR mAP@0.5/%	P Precision/%	R Recall/%
yolov5l	84.4	91.9	90
yolov5m	86.7	97.1	88
yolov5n	87.3	96.1	**91**
yolov5s	**88.8**	**97.6**	90

Table 4. Comparative experimental results of different attention mechanisms added to the backbone network

Model	PR mAP@0.5/%	P Precision/%	R Recall/%
YOLOv5s +CBAM	90.2	94.3	86.7
YOLOv5s +GAM	89.5	92.3	87
YOLOv5s +SE	89.6	95.3	88.4
YOLOv5s +CA	**89.7**	**92.1**	**89.4**

(a) (b) (c)

Figure 5. Comparison of the effect of introducing CA mechanism network detection

Figure 5 (a) shows the original image. Figure 5 (b) shows the network detection effect of the original YOLOv5s, which shows a leakage in the recognition of secure wearable devices with fuzzy problems. Figure 5 (c) shows the introduction of CA attention mechanism network, which can also achieve effective detection for the security wearable device target with fuzzy problem, so the reference of CA attention mechanism effectively improves the model's utilization of location information, greatly enhances the model's anti-interference ability, reduces the leakage rate, and improves the detection effect.

4.3.2. Modify the Backbone Network Module

From the above experiments, it can be seen that the introduction of CA attention mechanism has significantly improved the detection effect of the network, in order to better introduce the CA attention mechanism, this experiment modifies the C3 module of the YOLOv5s detection network into the C3CA module which adds the CA attention mechanism. Through the experiment, the results are obtained as Table 5.

As a result, it can be seen that the network after adding the C3CA module improves the mean average precision by 1.1% over the original network, improves the correct rate by 2.6% over the original network, and improves the recall rate by 3.2%. It can be seen that by modifying the C3 module in the backbone network and adding in the CA attention idea has a significant effect on the improvement of the network.

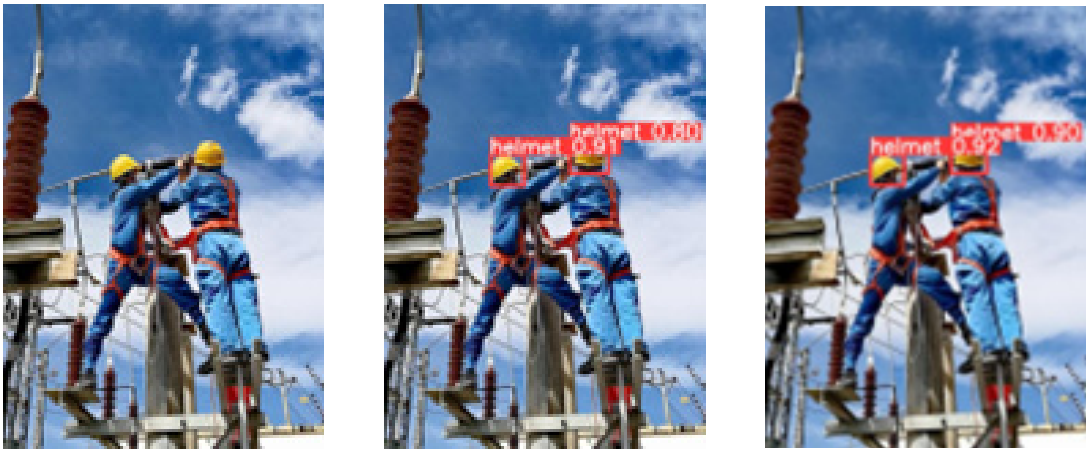
ISSN: 2688-8645

Table 5. Experimental results of adding C3CA module

Model	PR mAP@0.5/%	P Precision/%	R Recall/%
YOLOv5s	89.2	89.7	86.2
YOLOv5s +C3CA	**90.9**	**92.3**	**89.4**

Table 6. Experimental comparison of improved fusion feature layers

Model	PR mAP@0.5/%	P Precision/%	R Recall/%
YOLOv5s +ASFF	89.5	90.1	**87**
YOLOv5s +BiFPN	92.2	90.2	86.7
YOLOv5s +AFPN	**92.7**	**90.4**	86.9



(a) (b) (c)

Figure 6. Experiments with contrasting feature fusion layers

4.4. Experiments with Asymptotic Characteristic Pyramid Networks

As Table 5 As shown in Table 35, the fusion networks in the Neck layer are replaced by ASSP, Bi FPN and ASPN in the original network, which are adaptive spatial feature fusion network, weighted bidirectional feature fusion network, respectively.[25] AFPN, weighted bidirectional feature fusion network Bi FPN and asymptotic feature fusion network ASPN. It is found that the model using AFPN has the most obvious improvement in all indexes, and is better than the model adding ASSP and BiFPN, Therefore, AFPN is chosen as the feature fusion network in this paper. The asymptotic fusion of deep and shallow features by this fusion network makes the features fully utilized and provides effective support for the model to improve the detection of secure wearable devices.

Figure 6 (a) shows the original YOLOv5s detection effect. Figure 6 (b) shows the effect of introducing AFPN detection. The model replacing the AFPN fusion network recognizes the target more accurately and with higher confidence, which also indicates that the asymptotic fusion of high-level features in the AFPN feature fusion layer has improved the performance of the detection model .

ISSN: 2688-8645

4.5. Ablation Experiment

As shown inTable 7, through the introduction of C3CA coordinate attention mechanism module and AFPN feature fusion network module, the YOLOv5s algorithm model is improved in all the indexes, and it can be seen from the experimental data that the introduction of the C3CA coordinate attention module has improved the accuracy of the detection network; and secondly, through the change to the AFPN feature fusion neck to effectively improve the utilization rate of the network feature fusion, and effectively improve the utilization efficiency of the whole network on features to achieve better detection effect.

Table 7. Results of ablation experiments

Model	CA	C3CA	AFPN	GFLOPs	Parameters/10^6	AP			PR mAP@0.5/%	P Precision/%	R Recall/%
						helmet	boot	reflective vest			
Origin				16.0	7.0	91.4	88.6	87.6	89.2	89.7	86.2
1	√			16.0	7.1	91.6	89.1	88.4	89.7	92.1	**89.4**
2		√		**15.5**	**6.7**	91.7	90.1	90.9	90.9	92.3	89.4
3			√	16.6	7.2	92.9	90.4	94.7	92.7	90.4	86.9
4	√	√		15.5	6.8	91.6	90	92	91.2	93.9	88.4
5	√		√	16.6	7.2	93.3	91.1	94.6	92.5	**94.5**	86.9
6		√	√	16.1	6.9	93.7	90.9	**95**	93.2	92.5	87.2
7	√	√	√	16.1	6.9	**94.6**	**91.2**	95	**93.6**	91.3	89.2

Table 8. Mainstream algorithm comparison experiments

Model	PR mAP@0.5/%	P Precision/%	R Recall/%
SSD	78.2	67.3	67.5
Efficient Det	84.3	86.2	76.3
YOLOv3	83.5	82.3	79.7
YOLOv4	82.3	74.4	77.3
YOLOv5	89.2	89.7	86.2
Ours	**93.6**	**91.3**	**89.2**

4.6. Comparison Experiment

In order to evaluate the network model performance more objectively, the SSD[26] , EfficientDet [23] , YOLOv3[11] YOLOv4[12] YOLOv5, YOLOv5 and the improved model of this paper are compared, and the results are shown in Table 8, which shows that the model of this paper is better than other mainstream detection network models in all the indexes, which also indicates that the modification of the attention mechanism and the asymptotic feature fusion mechanism has a great effect on the performance enhancement of the YOLOv5s network model.

5. Summarize

In this paper, we propose a YOLOv5-C3CA network for detecting safe wear of personnel in power production industry. The network detection accuracy is effectively improved by creating a homemade safety wear dataset, adding the CA attention mechanism to YOLOv5s network and replacing the backbone network C3 module with the C3CA module; and replacing the feature fusion network structure with an AFPN. It is concluded through experiments that the addition of C3CA attention module and the replacement of AFPN fusion network have a great effect on the optimization of the original network, and significantly improve the recognition rate of

security wearable target detection. The method of this paper can quickly and accurately detect the personnel safety wearing situation, thus reducing the power production industry personnel due to unsafe wearing and causing accidents.

Acknowledgments

Supported by The Innovation Fund of Postgraduate, Sichuan University of Science & Engineering. (D10501637).

References

[1] RogersAnna, GardnerMatt, AugensteinIsabelle. QA Dataset Explosion: a Taxonomy of NLP Resources for Question Answering and Reading Comprehension [J]. ACM Computing Surveys, ACM, 2023.

[2] Narayan V, Awasthi S, Fatima N, et al. Deep Learning Approaches for Human Gait Recognition: a Review[A]. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC)[C]. IEEE, 2023: 763-768.

[3] Malhotra P, Gupta S, Koundal D, et al. Deep Neural Networks for Medical Image Segmentation[J]. Journal of Healthcare Engineering, Hindawi, 2022, 2022: e9580991.

[4] Ang L, Rahim S K N A, Hamzah R, et al. YOLO algorithm with hybrid attention feature pyramid network for solder joint defect detection[J]. arXiv, 2024.

[5] Wang H, Yang G, Li E, et al. High-Voltage Power Transmission Tower Detection Based on Faster R-CNN and YOLO-V3[A]. 2019 Chinese Control Conference (CCC)[C]. 2019: 8750-8755.

[6] Kang F, Li J. Research on the Detection Method of Electric Power Workers not Wearing Helmets based on YOLO Algorithm[A]. Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence[C]. New York, NY, USA: Association for Computing Machinery, 2023: 66-71.

[7] Souza B J, Stefenon S F, Singh G, et al. Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV[J]. International Journal of Electrical Power & Energy Systems, 2023, 148: 108982.

[8] QI Zezheng, XU Yinxia. Research on helmet wearing detection with improved YOLOv5s algorithm[J]. Computer Engineering and Application, 2023: 1-10.

[9] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Las Vegas, NV, USA: IEEE, 2016: 779-788.

[10] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. arXiv, 2016.

[11] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. : 6.

[12] Bochkovskiy A, Wang C-Y, Liao H-Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv, 2020.

[13] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[14] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation[A]. 2018: 8759-8768.

[15] Lin T-Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[A]. 2017: 2117-2125.

[16] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[17] Woo S, Park J, Lee J-Y, et al. Cbam: Convolutional block attention module[A]. Proceedings of the European conference on computer vision (ECCV)[C]. 2018: 3-19.

[18] Park J, Woo S, Lee J-Y, et al. BAM: Bottleneck Attention Module[J]. arXiv, 2018.

ISSN: 2688-8645

[19] Liu Y, Shao Z, Hoffmann N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions[J]. arXiv, 2021.

[20] Hou Q, Zhou D, Feng J. Coordinate Attention for Efficient Mobile Network Design[A]. 2021: 13713-13722.

[21] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[J]. arXiv, 2020.

[22] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation[A]. 2018: 8759-8768.

[23] Tan M, Pang R, Le Q V. Efficientdet: scalable and efficient object detection[A]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition[C]. 2020: 10781-10790.

[24] Yang G, Lei J, Zhu Z, et al. AFPN: Asymptotic Feature Pyramid Network for Object Detection[J]. 2023.

[25] Liu S, Huang D, Wang Y. Learning Spatial Fusion for Single-Shot Object Detection[J]. arXiv, 2019.

[26] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[A]. B. Leibe, J. Matas, N. Sebe, et al. Computer Vision - ECCV 2016[C]. Cham: Springer International Publishing, 2016: 21-37.