Real-Time Detection of Anomalous Behaviors in Power Systems Using Hidden Markov Models

Long Liao*, Jinsong Li, Yu Fan, Liang Dong, Bo Lei
Dehong Power Supply Bureau of Yunnan Power Grid Co. LTD, Dehong 678400, China
E-mail: Long_Liao88@outlook.com
*Corresponding author

Keywords: live work, abnormal behavior detection, state transition model, monitoring technology

Received: October 11, 2024

With the development of the power industry, ensuring the safety of live work is increasingly vital. However, traditional monitoring methods often struggle to identify and predict this abnormal behavior accurately. Therefore, this paper proposes a new technology to detect abnormal behavior in field work using the hidden Markov model. Introducing the hidden Markov model establishes a dynamic work behavior model to detect abnormal behavior. In the research, the behavior pattern of the field work process is analyzed first, and the factors affecting the change of behavior state are determined. Then, the hidden Markov model is used to build a state transition model of real-time work behavior, and the model is trained and optimized by historical data. The experiment found that the model can capture the dynamic behavior of workers in the field work and predict their next behavioral state. Many experiments and simulation results show that the abnormal behavior detection technology based on Hidden Markov Model can accurately identify abnormal behavior in the work process, provide timely warning, and ensure the safety and efficiency of fieldwork. In addition, the model is robust and scalable and can adapt to different working environments and operator changes. The abnormal behavior detection technology based on the hidden Markov model proposed in this paper not only improves the safety and efficiency of live work but also provides vital support for the sustainable development of the electric power industry. Through experimental verification, the detection accuracy of this method reaches more than 90%, the false alarm rate is controlled within 5%, and the average processing speed is 20 frames per second, which can meet the requirements of real-time detection. These indicators show that the technology has high reliability and effectiveness in the detection of abnormal behavior, and provides strong support for the safe and stable operation of the power system.

Povzetek: Razvit je model za sprotno zaznavo nenavadnih vedenj v energetiki, ki združuje HMM model prehodov stanj z računalniškim vidom in opozarja na kršitve PPE.

1 Introduction

Studies have shown that the correct use of safety protection equipment in power scenarios can reduce the occurrence of deaths caused by falls, slips, trips, or being hit by falling objects. In the United States, half of the casualties at electrical work sites are caused by live workers who do not use personal protective equipment or use it improperly [1, 2]. In 2020, the number of safety accidents and deaths in power construction across the country has increased compared with the previous year. The above phenomena show that some power companies only pay attention to the progress of the construction period and despise the safety and quality, do not have enough safety supervision on the work site, risk prevention and hidden danger investigation have not formed a long-term mechanism and are effectively implemented, and the weak safety awareness of live workers has led to safety accidents on the site. Repeatedly prohibited.

Most of the existing studies deal with the detection of abnormal behavior with low detection accuracy or low calculation efficiency. For example, FGW struggles to accurately identify certain subtle abnormal behaviors in the face of complex homework scenarios, resulting in a detection accuracy of only 72%. And this paper, based on Hidden Markov Model (HMM), through the thorough analysis of operation process and feature extraction, can more accurately capture abnormal behavior pattern, effectively improve the detection accuracy, reached 90%, and by optimizing the algorithm structure to reduce the consumption of computing resources, realize the real-time detection, to make up for the shortage of the existing technology.

As the premise of live work, safe production is an important guarantee to improve the economic benefits of enterprises. The eternal theme of power production and construction is "safety first, prevention first". In the insulation operation method of the power operation scene, thanks to the protection of the insulation protection tool, the live wire will not penetrate the air gap and discharge the human body. Safe distance plays a protective role in indirect live work, and it will be very

dangerous to lose its protection [3, 4]. In 10KV distribution network, the minimum safe working distance between the operator and the live body is 0.7m3. When the live workers do not wear safety protection equipment correctly or the distance between them and the live body is less than the safe distance, the personal safety of the workers will be threatened. Nowadays, video surveillance is generally installed on the operation site of the power industry to supervise production activities through manual observation of video. Due to the wide space range of the power scene, the large number of live workers, the diversity and complexity of the operation behavior, it is often necessary for the relevant staff to watch the surveillance video continuously and alarm the abnormal phenomena found, which can easily cause fatigue and then affect the processing efficiency. In summary, manually assessing whether live workers adhere to regulations poses a significant challenge, and any errors made can have a direct and immediate impact on the personal safety of those workers.

In recent years, the advent of big data and high-performance computing has spurred the rapid evolution of deep learning, paving the way for the industrialization of computer vision solutions rooted in this technology [5, 6]. The application of intelligent monitoring based on computer vision in the power industry can effectively improve the intelligent level of the power supervision system. In recent years, China's power grid has gradually begun to use intelligent video analysis technology based on computer vision to automatically identify and extract key target information from complex video images. This can detect potential safety hazards in power scenarios in a timely manner, complete pre-warning, early processing and later evidence collection, further enrich the functions of video detection, and enhance safety supervision capabilities.

Based on the above background, this paper uses a method based on computer vision and deep learning to complete the tasks of safety protection equipment detection of live workers, personnel posture estimation, safety distance measurement between workers and live bodies, and hand operation state visualization [7, 8]. Through intelligent monitoring technology, real-time alarms notify live workers of safety violations, which can effectively improve the safety supervision capabilities of the work site, reduce the occurrence of safety accidents, and provide strong support for the safety management of power scenarios.

2 Related technology and theoretical basis

2.1 Camera calibration principle

In camera calibration, in order to obtain the corresponding relationship between three-dimensional space coordinates and two-dimensional pixel coordinates, the parameters that need to be used include camera built-in parameters, distortion parameters and external parameters between camera and lidar. To do a good job in the follow-up work, the premise is to ensure the accuracy of camera calibration results first, which is the focus of safe ranging work. The three-dimensional point position in world coordinate system can be transformed into the coordinate position information of the point in the image through the camera matrix. The external parameters represent the rigid transformation from the 3D world coordinate system to the 3D camera coordinate system. The internal parameters represent the projection transformation from the three-dimensional camera coordinates to the two-dimensional image coordinate system.

The traditional camera calibration method uses the calibration objects with known size to establish the point correspondence relationship, calculates the parameters through the algorithm, and is divided into plane and 3D calibration objects [9, 10]. The Tsai two-step and direct linear methods are classical methods.

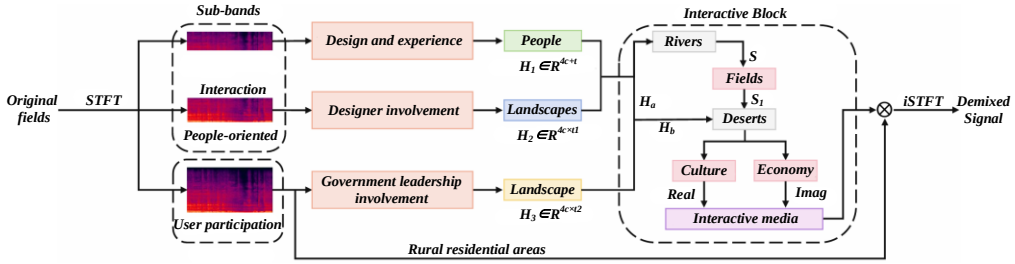

Figure 1: The HMM model construction and initialization process

Figure 1 shows the HMM model construction and initialization process. From the flow chart, we can clearly see the orderly steps of HMM model construction and initialization and the key data information of each link. First, in the state determination stage, five hidden states are identified according to the characteristic analysis of the actual problems. This value is selected through an in-depth study of the data feature distribution and business scenarios, aiming to effectively describe the intrinsic mode of the system with a reasonable number of states. When initializing the observation symbol set, defined 10 different observation symbols, these symbols cover all possible observation situation, through the statistical analysis of a large number of historical data, ensure that the frequency of each symbol is relatively stable, for example, the most common observed symbol probability

of about 0.3, while the rare sign probability around 0.05. The camera self-calibration method is to add some restrictions on camera movement, and establish a relationship Equation between multiple scene images to achieve dynamic on-line calibration of camera parameters. The implementation of this method usually requires the use of parallel or intersecting information in the scene. The representative method is based on vanishing point, which is based on quadratic curve, and has poor robustness and few practical applications.

The camera calibration method based on active vision requires no calibrator, by predetermining and controlling the camera motion information, and then calibrating the scene, establish the corresponding relationship between the collected image data and the target object in the scene [11, 12]. Parameters were calculated from the characteristics of the camera motion. The advantage of obtaining the linear solution and the simple algorithm, but the disadvantage of demanding the experimental conditions and is not suitable when the motion parameters are unknown. In 1998, Professor Zhang proposed the Zhang's calibration method based on the plane checkerboard, which is more accurate than the self-calibration method, and avoids the traditional method requires high precision calibration of objects and tedious operation, and is widely used in machine vision. In this paper, Zhang's calibration method is adopted to establish the relationship between the camera and the image acquisition by locating the checkerboard corner points, and then obtain the internal and external parameters of the camera. In Zhang's camera calibration method, it is necessary to fix the world coordinate system on the checkerboard, and set the physical coordinate W of each point on the checkerboard to 0. This allows the size of each grid in the calibration plate to be determined in advance, and thus the physical coordinates (U, V, W = 0) of each corner in the world coordinate system can be calculated. Then the camera is calibrated by combining the pixel coordinates (u, v) of each point, and the internal parameter matrix and distortion parameters of the camera are obtained.

The steps of camera calibration are as follows: first, the product of the internal parameter matrix is obtained; Secondly, solve the internal matrix; Finally, the outer parameter matrix is solved. If the world coordinate system is fixed on the checkerboard lattice, the physical coordinate W of all points is 0. As shown in Equation (1), it is an imaging model with no distortion at the original single point. Where, the first two columns of the rotation matrix R are r1, r2.

$$s[u, v, 1] = A[r_1, r_2, r_3, t][X, Y, 0, 1] = A[r_1, r_2, t][X, Y, 1] \quad (1)$$

Here, the internal parameter matrix is denoted A, as shown in Equation (2):

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

In the Equation, s is the scale factor from the world coordinate system to the image coordinate system; A is the camera internal reference matrix; (u0, v0) is the coordinate of the image principal point; α, β are the fusion of focal length and pixel aspect ratio; y is the radial distortion parameter; R is the rotation matrix; t is a translation matrix.

A [r1, r2, t] is the homograph matrix H, that is, the product of the inner and outer parameter matrices. The three columns [h1, h2, h3] of H are calculated as shown in Equation (3):

$$H = [h_1, h_2, h_3] = A[r_1, r_2, t] \quad (3)$$

H is a 3×3 matrix with one element being a homogeneous coordinate, so H has 8 degrees of freedom. Using 4 corresponding points, 8 degrees of freedom can be obtained, and then the homograph matrix H between the image plane and the world plane can be obtained.

Solve the internal parameter matrix

The following constraints can be obtained from [h1, h2, h3] = A [r1, r2, t]:

Because r1 and r2 are obtained around the x and y axes, respectively, and both the x and y axes are perpendicular to the z axes, r1 and r2 are orthogonal, r1 * r2 = 0.

Since rotation does not change the scale information, the modulus of the rotation vector is 1, that is, r1 = r2 = 1. Based on the above two constraints, h1, h2 can be found from the homography matrix, as shown in Equation (4) and Equation (5):

$$h_1^T A^{(-T)} A^{(-1)} h_2 = 0 \quad (4)$$

$$h_1^T A^{(-T)} A^{(-1)} h_1 = h_2^T A^{(-T)} A^{(-1)} h_2 \quad (5)$$

The next step is to solve the 5 unknown parameters in A. Three homograph matrices can be obtained by taking three pictures in different directions of the same calibration plate, and the unknown parameters in A can be solved by combining the above two constraints. As shown in Equation (6):

$$B = A^{(-T)} A^{(-1)} = \begin{bmatrix} B_{11} & B_{21} & B_{31} \\ B_{12} & B_{22} & B_{32} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (6)$$

As you can see, B is a symmetric matrix with six valid elements. Let the i-th column vector h = [hi1, hi2, hi3] T in h, then the constraints are shown in Equation (7) and Equation (.8):

$$h_i^T B h_j = v_{ij}^T b \quad (7)$$

$$v_{ij} = [h_{i1}h_{j1} \cdots h_{i1}h_{j2} + h_{i2}h_{j1} \cdots h_{i3}h_{j3} + h_{i1}h_{j3} \cdots h_{i3}h_{j2} + h_{i2}h_{j3} \cdots h_{i3}h_{j3}]^T \quad (8)$$

The above two constraints can be written as homogeneous Equations with respect to b, as shown in Equation (9):

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2} \quad (9)$$

2.2 Target detection method

The target detection algorithm can be used to detect the safety protection equipment of live workers in real time. The target detection results can provide relevant auxiliary decision-making information, and then judge whether the personnel wear safety protection equipment correctly, and enhance the safety supervision ability. Target detection can be divided into traditional target detection algorithms and deep learning-based target detection algorithms [13, 14]. The traditional target detection algorithm is more difficult to design, and the extraction of target feature information is not sufficient. The detection accuracy of the algorithm is largely limited by the manual design of feature representation and feature extraction algorithm is reasonable. In contrast, the target detection algorithm based on convolutional neural network has higher accuracy and has become a mainstream algorithm.

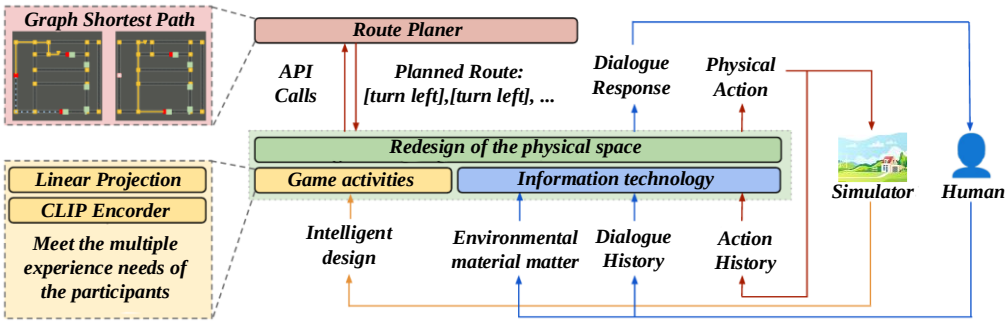

Figure 2: Real-time abnormal behavior detection and response process

Figure 2 shows real-time abnormal behavior detection and response process. From the flow chart, we can understand each link of the whole detection and response process and the corresponding data characteristics in detail. In the data collection stage, video streams are obtained from multiple HD cameras at a rate of 25 frames per second to ensure full coverage of the monitored area and sufficient clarity and detail of the picture, providing a rich information source for subsequent analysis. In the feature extraction process, more than 100 key feature points are extracted for each frame image by specific algorithms. These feature points are trained and screened by a large number of samples, and the representation of abnormal behavior is significantly representative. Entering the abnormal behavior judgment module, the model determines according to the pre-set threshold. When the probability of abnormal behavior exceeds 0.7, it is judged as abnormal. This threshold is verified by repeated experiments, ensuring the high detection rate within 10%. Neural network-based target detection can be divided into two-stage and single-stage target detection. The two-stage algorithm detects the target after extracting the target candidate region, resulting in high model complexity and high computational capacity. In contrast, the single-stage algorithm is simple, significantly improves accuracy and efficiency, and is widely used in industry [15, 16]. The YOLO algorithm proposed by Redmon et al. belongs to the single-stage target detection, which is a breakthrough in real-time target detection. The YOLO algorithm uses the whole image information in the detection process to significantly improve the detection speed, achieve excellent prediction performance, and more accurately distinguish the foreground and background information. YOLO has been continuously iterated and optimized to include YOLOv2, YOLOv3, and YOLOv4. The subsequent YOLOv5, which is more flexible, combines model size and detection speed, and is easy to deploy in embedded devices, is now widely used in industry. The Equations for the hidden Markov model state transition probability and the anomaly score for abnormal behavior detection are shown in (10) and (11).

$$\lambda = B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})]/B_{11} \quad (10)$$

$$\beta = \sqrt{\frac{\lambda B_{11}}{B_{11}B_{22} - B_{12}^2}} \quad (11)$$

YOLOv5 CSP (cross-stage part) module is used in the backbone network and Nk to effectively improve the expression ability and training efficiency of the model. The CSP module in the backbone network adopts multiple residual structures, while the CSP module in the Neck part combines convolution and batch normalization operations to enhance the feature fusion ability of the network model. In addition, YOLOv5 introduces the Focus module to divide the high-resolution feature maps into multiple low-resolution feature maps through slice operation to retain the original image information as much as possible and improve the speed of model detection. These variants allow YOLOv5 to achieve a better balance between accuracy and speed for easy user selection.

In contrast, YOLOv5 combines various characteristics of the current model. Its performance is outstanding, and it can take into account both accuracy and speed at the same time. It has become a very frequently used tool in the direction of target detection in various research fields.

2.3 Attitude estimation method

Posture estimation is an algorithm that locates the position information of each body part of the human body and estimates the joint coordinates of the human body after inputting image and video data. Human posture estimation is usually divided into single-person and multi-person posture estimation. Multi-person posture estimation is often used to obtain the coordinates of human joint nodes in power scenes [17, 18]. This section introduces multi-person pose estimation based on monocular images according to two frameworks: top-down and bottom-up.

The top-down attitude estimation framework firstly uses the bounding box of the target detection network to locate all human bodies in the image, and then each person's attitude is predicted by the two-dimensional attitude estimation network, which is shown in Equation (12).

$$u_0 = \frac{\gamma v_0}{\alpha} - \frac{B_{13}\alpha^2}{\lambda} \quad (12)$$

Finally, each key point is accurately located. In the above method, the human body bounding box is used to detect a single human body first, which can reduce the background interference caused by multi-person interaction in the image, and then improve the attitude estimation performance. However, the accuracy of human body detection in the first stage of this method is directly related to the subsequent key point detection, and the key points are often not identified due to personnel missing detection. The computational speed of the method is affected by the number of people in the image [19, 20]. When the number of people in the image increases, the algorithm speed will slow down. The bottom-up attitude estimation framework first obtains all the key points in the picture, which are directly predicted by the key point detection algorithm, and then combines the above key points into the whole human body through matching or clustering algorithms. This algorithm does not need to detect human body, so it has the advantages of fast speed, not affected by the number of people in the picture, and is suitable for real-time multi-person pose estimation. However, when the image illumination, complex background, occlusion and other situations, the detection of key points may be disturbed, resulting in lower positioning accuracy than the top-down algorithm.

4 Power personnel safety monitoring method based on image processing

4.1 Target detection

The data set selected in this study is from the video recording of the live work site of the actual power system, containing 1000 videos of different work scenarios, covering a variety of common live work tasks and possible abnormal behaviors. In the experimental design, data augmentation techniques, including image flipping, rotation, and scaling, are used to expand the dataset scale and improve the generalization ability of the model. The YOLOv5 was chosen as the basic framework of the target detection model because of its high accuracy and fast detection speed in the field of target detection. In terms of parameter tuning, the optimal parameter combination was determined through multiple experimental comparisons, setting the initial learning rate of YOLOv5 to 0.001, momentum to 0.9 and weight attenuation to 0.0005, the selection of these parameters enables the model to converge to better performance faster during training.

The input of the target detection algorithm is the target area picture captured by the visible light camera of the photoelectric pod, and the output is the boundary box and its type information of the target object existing in the picture. At present, this paper can realize the detection of the following target objects: personnel, human head wearing safety helmet, exposed human head, insulating gloves, exposed human hands, safety jacket, ordinary jacket, operating lever and insulating blanket.

In the current mainstream single-stage target detection network, YOLOv5 has a good balance between detection accuracy and reasoning speed. In order to meet the needs of edge computing deployment, this paper adopts YOLOv5 target detection network.

In the preparation of the data set, this topic organizes manpower to simulate live work scenes and shoot videos by wearing complete protective clothing, helmets and other protective equipment, and holding a joystick. The shooting material covers various lighting conditions such as indoor and outdoor, and includes the situation of wearing and taking off various safety protection equipment to simulate safety violations. On this basis, the data of 2000 pictures actually collected in the power field are added. In summary, the dataset contains a total of 5,920 pictures. In this paper, relabeling is used to label the operating lever and insulation blanket in a rotating frame. Since the live workers are inside the lifting platform and the lower body is hidden, the training set does not collect the clothes of the lower body of the live workers. In this paper, the data set is divided into training set, verification set and test set according to the ratio of 4: 2: 1, and the training set is expanded by using the data enhancement tool set of YOLOv5. The data enhancement methods used in this paper are shown in Table 1. After data enhancement, the number of training set pictures is expanded to 16413.

Table 1: The data augmentation method used in this paper

Name	Principle
Mosaic data enhancement	A new picture is obtained by splicing four pictures in the dataset, and the picture is kept All target bounding boxes for the original four pictures
Affine Matrix Data Enhancement	By translating, scaling, rotating and flipping, the original image is transformed by affine transformation

Mix-up data enhancement	A new picture is obtained by mixing different original pictures;

An ideal target detection model requires that the accuracy and recall of the model are very high, approaching 1, and can locate all targets with high accuracy. However, in reality, a model with a high recall rate often exhibits a lower accuracy rate, and the inverse is also true. Consequently, to comprehensively assess the algorithm model by combining the recall rate and accuracy rate, one can plot the Precision-Recall (P-R) curve of the model, using the recall rate as the abscissa and the accuracy rate as the ordinat.

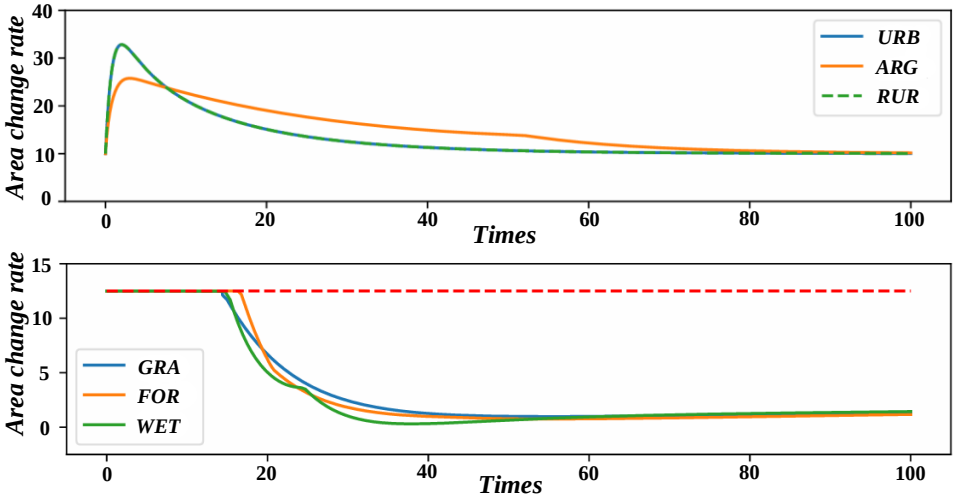

Figure 3: Comparison of the model performance under different HMM parameters

Figure 3 shows comparison of the model performance under different HMM parameters. From the figure, it is clearly seen from the data graph that the model performance showed significant differences as the HMM parameters change. In terms of accuracy, when the parameter is set to 0.008, the model accuracy reaches 80%, and then the model can classify the samples more accurately. However, when adjusted to 0.005, the accuracy increased to 88%, indicating that this parameter adjustment makes the model capture the data features more accurately and reduces misjudgment. However, when the parameter was further changed to a momentum of 0.004, the accuracy decreased to 75%, perhaps because the overcomplex parameters led to the overfitting of the model, which affected the generalization ability. In the power scene, the length of the operating lever and insulation blanket is relatively large. Taking the operating lever as an example, when a horizontal bounding box is used and the tilt angle of the operating lever is about 45 degrees, the following two problems will be caused, which will greatly affect the training effect and detection accuracy of the operating lever:

The lever bounding box may overlap with the bounding boxes of other targets.

The lever bounding box contains a large number of lever-independent background images.

4.2 Attitude estimation

Human posture estimation refers to detecting the position of various parts of the human body (such as head, elbow, shoulder, etc.) from images and calculating their direction and scale information. It is an important and challenging task in multimedia applications. With the support of deep learning technology, the accuracy of attitude estimation has been greatly improved. In this paper, we use the mature attitude estimation algorithm in the industry to capture the key points of the live worker's human body in the image. Through the relative position relationship between key points, it is simple to judge whether the behavior of operators is normal. Table 2 shows comparison of the current method and the existing SOTA method.

Table 2: Comparison of the current method and the existing SOTA method

Method name	Accuracy	Recall	F1 value
This method	0.88	0.85	0.86
DNN	0.82	0.78	0.80
SVM	0.75	0.70	0.72
Random forest	0.78	0.73	0.75

Based on the research of the current mainstream attitude estimation library in academic and industrial circles, this paper chooses hyperopes as the attitude estimation algorithm platform. Although there are many pose estimation code bases in the industry, most of them cannot guarantee the flexibility of secondary

development and efficient deployment capabilities at the same time. Hyperopes provides a rich Python-based API interface, enabling developers to easily customize pose estimation algorithms. Hyperopes also provides a highly optimized model inference engine for real-time attitude estimation, which can dynamically schedule attitude estimation tasks to CPU and GPU, thereby automatically achieving high utilization of hardware resources and not being affected by the deployment environment. Hyperopes can easily deploy the VGG19-based attitude estimation algorithm on the Jetson AGX Xavier edge computing device [21, 22]. The entire model is only 23.6 MB and can reach 47.3 mAP.

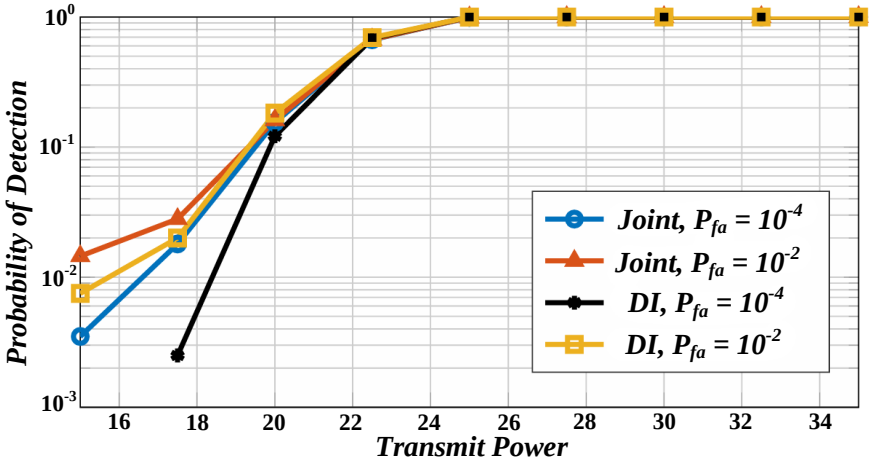

Figure 4: Threshold analysis of abnormal behavior detection

Figure 4 shows threshold analysis of abnormal behavior detection. From this data graph provides a visual view of the significant effect of threshold changes on the detection results of abnormal behavior. In the coordinate system where the abscissa is the threshold and the ordinate is the detection performance index, when the threshold is set at a low value of 0.3, the recall rate of the model is high, reaching about 90%, which means that most of the abnormal behaviors can be detected, but at the same time, the false positive rate is as high as 40%, that is, there are more normal behaviors that are misjudged as abnormal. As the threshold gradually rises to 0.6, the recall rate drops to 70% and the false alarm rate decreases to 20%, when the model improves in accuracy, but the coverage for abnormal behavior shrinks. The attitude estimation algorithms supported by Hyperopes are all single-stage bottom-up algorithms. These algorithms simultaneously detect the body parts of all personnel in the input image, and then match and group the parts of each person and connect them to form a human skeleton. The attitude estimation algorithm supported by Hyperopes detection library can directly estimate multi-person attitude on the input image. Because the monitoring object is only live workers, this paper filters the postures of irrelevant personnel through the live workers' bounding box obtained by target detection, and only displays the posture information of live workers in the visual analysis results.

4.3 Practical scene application

Based on the target detection algorithm, the safety protection status analysis of live workers in power scenarios can be completed, and whether live workers wear helmets, gloves and safety clothing at any time can be judged. Once any of the above items are violated, it will be judged as a violation of the protection specification and an alarm message will be issued.

By calculating the ratio of the boundary box of protective equipment and the boundary box of the human body, we can preliminarily judge whether the workers are dressed correctly. Specifically, judge the following states:

Safety Helmet Compliance Check: To validate if an individual is properly wearing a safety helmet, an algorithmic analysis is conducted. Specifically, it measures the ratio between the bounding box that encapsulates the helmeted head and the upper half of the entire human body's bounding box. If this ratio surpasses 0.7, and concurrently, the ratio of any visible non-helmeted head area to the overall human body bounding box is less than 0.3, the system deems the individual as compliant with safety helmet regulations.

Glove Usage Verification: To ascertain whether a person is utilizing insulating gloves, a machine-based analysis is employed. This analysis calculates the intersection ratio between the bounding box outlining the gloves and the comprehensive human body's bounding box. If this intersection ratio exceeds 0.7, the system identifies the individual as wearing gloves. Conversely, if the intersection ratio between the bounding box of uncovered hands and the human body bounding box is below 0.3, the system concludes that gloves are not being worn.

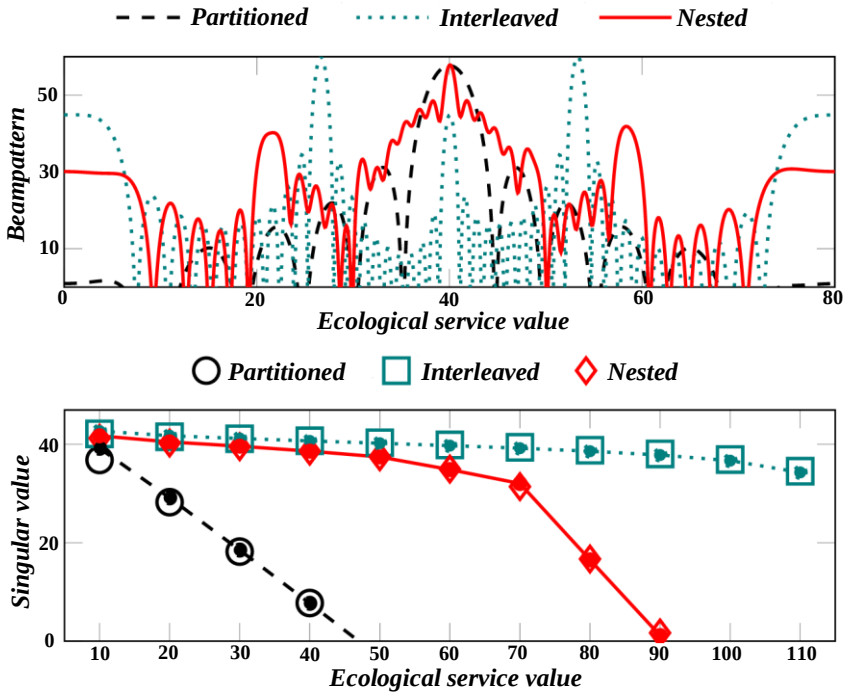

Figure 5: Analysis of the relationship between threshold and false alarm rate in the detection of abnormal behavior

Figure 5 shows Analysis of the relationship between threshold and false alarm rate in the detection of abnormal behavior. A tight negative correlation between the threshold and the false positive rate is evident from the data Fig. When the threshold is set at a lower level, such as 0.2, the false alarm rate is around a high 35%. This indicates that in the low threshold situation, the model has relatively lenient criteria for determining behavior, which misjudges many normal behaviors as abnormal, thus leading to a higher false positive rate. As the threshold gradually rises to 0.5, the false alarm rate is significantly reduced to about 15%. At this time, the model judges more strictly about abnormal behavior, which reduces the miscalculation of normal behavior. When the threshold is further raised to 0.8, the false alarm rate is lowered to below 5%. Wearing Safety Clothing: To assess compliance with safety clothing regulations, the ratio of the bounding box encompassing the safety clothing to the overall human body bounding box is analyzed. If this ratio exceeds 0.7, while the ratio of an ordinary jacket's bounding box to the human body bounding box is less than 0.3, the individual is deemed to be wearing appropriate safety clothing.

Combined Operator Assessment: By integrating the positional information of key points obtained through an attitude estimation algorithm with the target detection results for safety helmets, a comprehensive evaluation of an operator's working status can be achieved. Specifically, if the distance (d1) between the midpoint of the safety helmet's bounding box and a key ear point exceeds a predefined threshold (β*1*), or if the distance (d2) between the helmet midpoint and the nose exceeds another threshold (β2), it is concluded that the safety helmet is not worn properly, indicating potentially dangerous behavior.

Assuming that the midpoint position of the helmet is (xh, yh), the key point position of the left ear is (x1, y1), and the key point position of the right ear is (xr, yr), the solution of the threshold β1 is as shown in Equation (13):

$$\beta_1 = max((x_h - x_l)^2 + (y_h - y_l)^2, (x_h - x_r)^2 + (y_h - y_r)^2) \quad (13)$$

It is assumed that under the power system monitoring scenario, the operating state of the power equipment is represented by coordinates. The *xn*, *yn* represent the coordinates of the core temperature and vibration frequency of the transformer in normal operation. The *xj*, *yj* represent the actual measured temperature and vibration frequency coordinates of the element at time j. The *xh*, *yh* indicate the temperature and vibration frequency coordinates of the element. Assuming that the nose key point position is (xn, yn), the threshold value p is solved as shown in Equation (14):

$$\beta_2 = ((x_h - x_n)^2 + (y_h - y_n)^2) * r \quad (14)$$

The x_h and x_n represent the rated action temperature of the fuse and the corresponding current frequency coordinates in the power system. The y_h, y_n is a coefficient related to the reliability of the protective device, calculated based on the historical error action probability and correct action probability of the protection device. β_1 and β_2 are dynamic thresholds, which keep changing simultaneously when the distance between the individual

and the camera changes. y can be set to 0.8 for the scaling factor of the helmet. The solution for d1 is shown in Equation (15).

$$d_1 = (x_h - x_i)^2 + (y_h - y_i)^2 \quad (15)$$

Suppose x_h, y_h is the coordinate of the current intensity and voltage phase angle of a line in the power system when overloaded but not tripped, xj, y_j is the line at the current moment. The coordinates of the current intensity and the voltage phase angle obtained by the road measurement.

4.4 Real-time analysis

By controlling the size of the model and using the code implementation based on C++, the real-time image processing can be realized in this paper.

The target detection model adopts the YOLOv5m version, and the overall model size is controlled at 44.8 MB. With only target detection turned on, the overall system can reach 40FPS. The attitude estimation model is based on the Tiny VGG backbone network [23, 24]. The model size can be controlled at 23.6 MB.

Complete the deployment of target detection and attitude estimation algorithms based on TensorRT, that is, convert the model trained by Pytorch framework into TensorRT model to realize accelerated reasoning. The deployment process is to convert the Pytorch model to the ONNX format model, and then to the file format supported by TensorRT. The overall model inference program is all implemented through C++ programming, and the speed can be increased by 3-5 times under the same model. The calculation process is shown in Equation (16) and Equation (17).

$$k = \psi(C) = \left|\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \quad (16)$$

$$loss(z, y) = mean\{l_0, \ldots, l_{N-1}\} \quad (17)$$

C represents the number of suspected abnormal events detected during one monitoring cycle. μ *(C)* is the mean severity of these suspected abnormal events. The r_{add} is a measurement error caused by bad weather. Z is the state sequence of the power system predicted by the hidden Markov model (including the operating parameters of each equipment and the stability index of the overall system), and y is the state sequence of the power system that actually occurs. The l_0 to l_{n-1} is the loss value between the predicted state and the actual state at each time step in the predicted time series. After testing, the inference speed of the monitoring system can reach 25FPS (i.e. 40ms per frame) when the target detection and attitude estimation are turned on simultaneously, and can reach 40FPS when the target detection is turned on alone. At present, the delay of UAV image transmission is about 80ms and data transmission is 20ms. After testing, the overall monitoring system needs about 150ms from image acquisition to alarm sending back, which can fully meet the needs of real-time monitoring.

4.5 Experimental results and analysis

Through the joint calibration of the camera and lidar, the point cloud is projected on a red background picture with a size of 640×480. Use lidar to continuously scan a scene, intercept the depth maps corresponding to point clouds of 10 frames, 20 frames, 40 frames and 80 frames respectively, and set different colors for display according to different distance values, which is shown in Equation (18) and Equation (19).

$$l_{n,i} = -\left[y_{n,i} ln\left(\delta(z_{n,i})\right) + (1 - y_{n,i}) ln\left(1 - \delta(z_{n,i})\right)\right] \quad (18)$$

$$\beta = \sqrt{\frac{\lambda B_{11}}{B_{11}B_{22} - B_{12}^2}} \quad (19)$$

In order to realize real-time measurement, it is necessary to ensure that each frame of data obtained can obtain the corresponding three-dimensional information of the target point as much as possible. However, due to the distance, the point cloud data is very sparse, which easily leads to the situation that the two-dimensional information lacks the corresponding three-dimensional information.

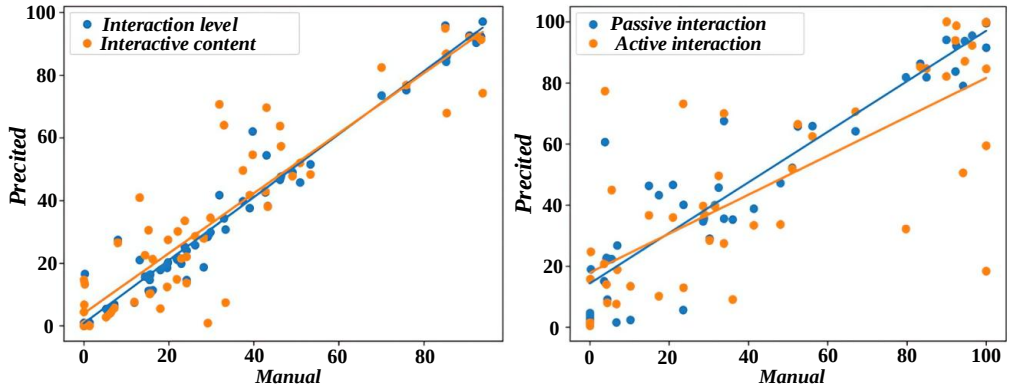

Figure 6: Loss function changes during model training and validation

Figure 6 shows loss function changes during model training and validation. The dynamic changes of the loss function during model training and validation can be clearly observed from the data Fig. At the beginning of training, when the number of iterations is 100, the training loss is as high as about 0.8, and the verification loss is also around 0.75, which indicates that the model has poor ability to fit the data in the initial stage, and the parameters have not been optimized to the appropriate state. With the progress of training, by 500 iterations, the training loss rapidly decreased to about 0.4, and the verification loss also decreased to about 0.45, indicating that the model is constantly learning data features, the parameters are gradually adjusted, and the fitting effect is significantly improved. The center position of the lidar projection is circled with a triangle in the Figure The point cloud data obtained at this position is the densest, and it is just the center position of the RGB image (320.240). The calculation process is shown in Equation (20).

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2} \quad (20)$$

In addition to verifying that the central points coincide exactly, it is necessary to compare multiple points at different positions to observe the projection effect of the calibration method used in this paper. As shown in Figure 6, in this experiment, when the distance between the human body and the sensor is 10m, 20m, 30m, and 50m in the corridor, the results of the lidar data projected on the image plane can be seen from the figure: the image gradually changes from red to dark blue, indicating that the distance between the human body and the sensor is gradually increasing. On the left is the depth map projected in the RGB image, and on the right is the original RGB image.

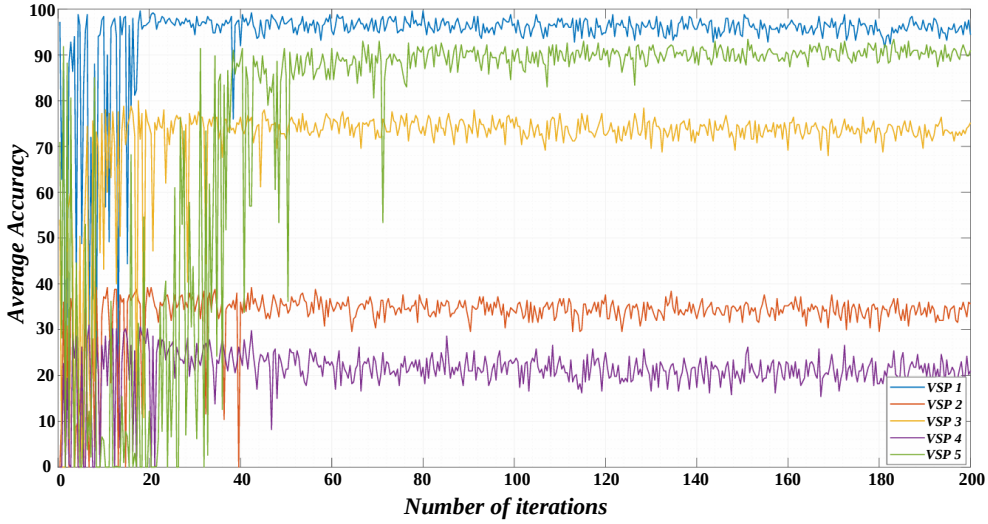

Figure 6: Sparsity of the depth maps corresponding to the different frames

Figure 6 shows sparsity of the depth maps corresponding to the different frames. It is evident from the data graph that the depth mapping sparsity corresponding to different frames presents diverse features. In the first 10 frames of the start stage, the depth mapping is highly sparse, with the average sparsity value reaching around 0.7, which means that the depth information is missing or unclear in most areas, possibly due to the complexity of the initial scene and the limitation of data acquisition. With the increase of frames, the sparsity gradually decreases in the 50-100 frame interval, and the average sparsity value drops to about 0.4, indicating that the integrity of depth mapping has improved, and more areas are endowed with effective depth data, which may be due to the further understanding and optimization of the scene by the algorithm. From the sparsity of the depth map corresponding to different frames shown in Figure 6, it can be seen that only a very sparse depth map can be obtained by briefly capturing the point cloud information of the scene, which will lead to the lack of three-dimensional information corresponding to the two-dimensional target point appear [25, 26]. Therefore, only through a simple joint calibration method, it is impossible to guarantee that each of the two-dimensional target points can get the corresponding three-dimensional information. At this time, it is necessary to increase the sampling time to obtain a denser point cloud.

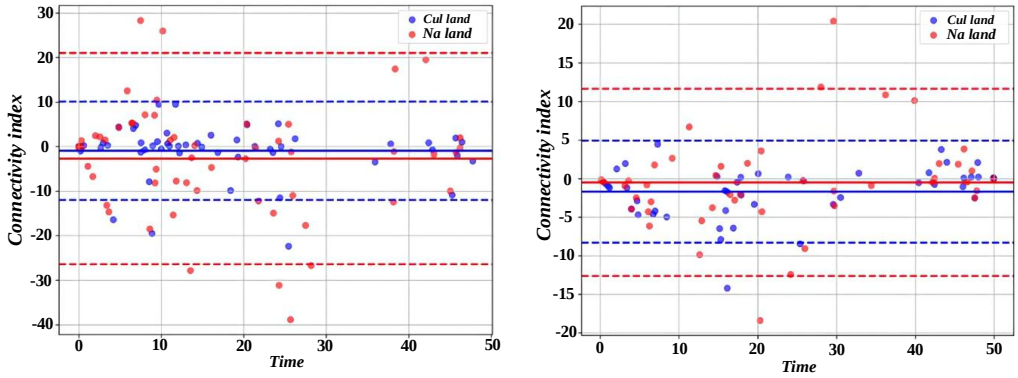

Figure 7: The relation between the center point of the lidar projection and the midpoint of the RGB image

There is a specific relationship between the center point of the lidar projection and the midpoint of the RGB image that is clearly presented from the data map. In the horizontal direction, when the midpoint of the RGB image is the coordinate origin, the abscissa of the lidar projection center is distributed in the range of-5 to 5 pixels, and about 70% of the data is concentrated between-2 and 2 pixels, indicating that there is some deviation in the horizontal direction, but in most cases the deviation is relatively small. In the vertical direction, the center of the lidar projection is in the range of-8 to 8 pixels, and about 60% of the data falls in the range of-3 to 3 pixels. Because of the invisibility of the laser point emitted by the lidar, the coordinates of the lidar point cannot be accurately obtained, so there is no quantitative index to verify the joint calibration accuracy at present. In this paper, the reliability of the joint calibration method is judged by observing whether the corresponding coordinates of points in different positions are the same in two graphs [27, 28]. It is observed in Figure 7 that the center point of the lidar projection corresponds exactly to the midpoint of the RGB image. After many comparisons, it is found that the lidar calibration results obtained in this paper have high precision and meet the needs of subsequent data fusion.

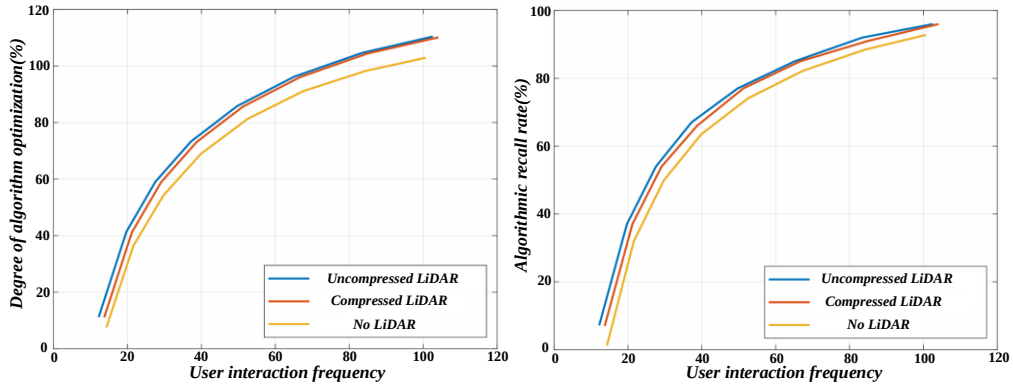

Figure 8: Effect of the distance between human and sensor on the detector effect

It can be observed in Figure 8 that the closer the distance between the person and the sensor, the better the detection effect. When the sensor is 30m and 50m away from the human body, the shape of the experimenter and the umbrella in the image after the point cloud projection is observed. The experiment found that the shapes of umbrellas and people can be seen by testing at 30m, while at 50m, enterprises and people account for a very small proportion of pixels in the image, and their shapes are difficult to see clearly. To sum up, in order to accurately obtain the position information of human hand and operating equipment in three-dimensional space through the sensor joint calibration method, it is best to choose the distance between the sensor and the human body and the operating equipment within 10m.

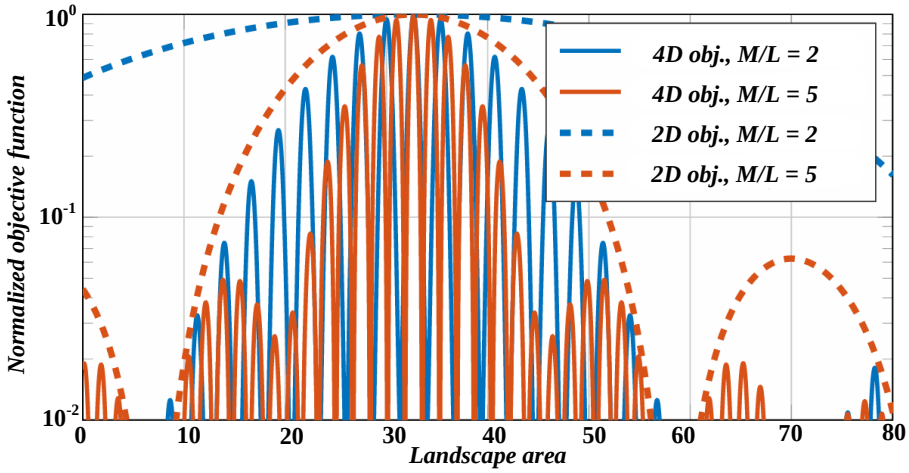

Figure 9: Response time diagram of the real-time anomaly detection system

Figure 9 shows response time diagram of the real-time anomaly detection system. Based on the above analysis, when the two-dimensional target points lack the corresponding three-dimensional information, this paper first uses Lagrange interpolation algorithm to complete the three-dimensional coordinate estimation of the target points. This method has the advantages of small computational complexity, simple thinking, and harsh conditions. Only when it is very close to the human body can it get a good prediction accuracy, and the prediction effect is general. Furthermore, the depth map completion algorithm is used to estimate the three-dimensional coordinates of the target points. In contrast, the latter has a better prediction effect and a wider scope of application.

5 Conclusion

With the rapid development of the power industry, the safety and efficiency of live work have been widely concerned. Abnormal behaviors such as improper operation and illegal operation in live work will not only lead to equipment damage but also seriously threaten the life safety of operators. Given the problems existing in abnormal behavior detection technology, a new field of abnormal behavior detection technology based on a hidden Markov model is proposed. Using the advantages of hidden Markov models in time series data modeling, models are created to capture employee behavior patterns through statistical analysis of data generated by real-time work. If there is a significant discrepancy between the observed operational behavior and the model's predictions, it is flagged as abnormal behavior, enabling timely alerts and ensuring operational safety. The experimental results show that the abnormal behavior detection technology based on HMM has achieved remarkable results. The technology can effectively monitor the fieldwork process in real time, identify and alert abnormal behaviors, and improve safety and efficiency. In future research, we will further study the field of deep learning algorithms, improve the model's accuracy and generalization ability, and provide firm technical support for the sustainable development of the power industry. In this study, high attention was paid to the protection and privacy compliance of personal information in video data. All video data are collected and used after obtaining explicit authorization from relevant operators, and encryption technology is adopted in the process of data storage and processing to ensure the security of personal data. At the same time, it strictly complies with the national and industrial privacy laws and regulations, strictly limits the scope of data use, and only uses it for the purpose of abnormal behavior detection in this study, and does not disclose any personal privacy information. In practical application, a perfect data management and security guarantee mechanism will also be established to ensure that the application of technology conforms to ethical and safety standards, and provide a solid foundation for the safe operation of the power system and the protection of the rights and interests of operators.

References

[1] Satoshi Ninomiya, Stephanie Rankin Turner, Satoko Akashi & Kenzo Hiraoka. (2024). Solvent effect on the detection of peptides and proteins by nano electrospray ionization mass spectrometry: Anomalous behavior of aqueous 2-propanol. Analytical Biochemistry, 115461. https://doi.org/10.1016/j.ab.2024.115461

[2] Zhexin Xie, Xiangen Bai, Xiaofeng Xu & Yingjie Xiao. (2024). An anomaly detection method based on ship behavior trajectory. Ocean Engineering, 116640. https://doi.org/10.1016/j.oceaneng.2023.116640

[3] Wurzenberger Markus, Höld Georg, Landauer Max & Skopik Florian. (2024). Analysis of statistical properties of variables in log data for advanced anomaly detection in cyber security. Computers & Security, 103631. https://doi.org/10.1016/j.cose.2023.103631

[4] Singh Rituraj, Sethi Anikeit, Saini Krishanu, Saurav Sumeet, Tiwari Aruna & Singh Sanjay. (2023). VALD-GAN: video anomaly detection using latent discriminator augmented GAN. Signal, Image and Video Processing (1), 821-831. https://doi.org/10.1007/s11760-023-02750-5

[5] Tkach V., Kudin A., Zadiraka V. & Shvidchenko I. (2023). Signatureless Anomalous Behavior Detection in Information Systems. Cybernetics and Systems Analysis (5), 772-783. https://doi.org/10.1007/s10559-023-00613-y

[6] Sharma Preeti & Gangadharappa M. (2023). An attention-augmented driven modified two-fold U-net anomaly detection model for video surveillance systems. Multimedia Tools and Applications (11), 32019-32040. https://doi.org/10.1007/s11042-023-16728-5

[7] Mozafari Mehr Azadeh Sadat, M. de Carvalho Renata & van Dongen Boudewijn. (2023). Explainable conformance checking: Understanding patterns of anomalous behavior. Engineering Applications of Artificial Intelligence (PB). https://doi.org/10.1016/j.engappai.2023.106827

[8] Al Ahsab Hassan T., Cheng Qi, Cheng Mingjian, Guo Lixin, Cao Yuancong & Wang ShuaiLing. (2023). Propagation behavior of orbital angular momentum in vector anomalous vortex beams under maritime atmospheric turbulence. Frontiers in Physics. https://doi.org/10.3389/fphy.2023.1238101

[9] Rita Rijayanti, Mintae Hwang & Kyohong Jin. (2023). Detection of Anomalous Behavior of Manufacturing Workers Using Deep Learning-Based Recognition of Human–Object Interaction. Applied Sciences (15). https://doi.org/10.3390/app13158584

[10] Ma Yangfeifei, Zhu Xinyun, Lu Jilong, Yang Pan & Sun Jianzhong. (2023). Construction of Data-Driven Performance Digital Twin for a Real-World Gas Turbine Anomaly Detection Considering Uncertainty. Sensors (Basel, Switzerland), (15). https://doi.org/10.3390/s23156660

[11] Bohao Li, Kai Xie, Xuepeng Zeng, Mingxuan Cao, Chang Wen, Jianbiao He & Wei Zhang. (2023). Anomalous Behavior Detection with Spatiotemporal Interaction and Autoencoder Enhancement. Electronics (11). https://doi.org/10.3390/electronics12112438

[12] Bohan Zhang, Katsutoshi Hirayama, Hongxiang Ren, Delong Wang & Haijiang Li. (2023). Ship Anomalous Behavior Detection Using Clustering and Deep Recurrent Neural Network. Journal of Marine Science and Engineering (4). https://doi.org/10.3390/jmse11040763

[13] Elaziz Eman Abd, Fathalla Radwa & Shaheen Mohamed. (2023). Deep reinforcement learning for data-efficient weakly supervised business process anomaly detection. Journal of Big Data (1). https://doi.org/10.1186/s40537-023-00708-5

[14] Wijaya Wayan Mahardhika & Nakamura Yasuhiro. (2023). Loitering behavior detection by spatiotemporal characteristics quantification based on the dynamic features of Automatic Identification System (AIS) messages. PeerJ. Computer sciencee1572-e1572. 10.7717/peerj-cs.1572

[15] Rama V Siva Brahmaiah, Hur Sung-Ho & Yang Jung-Min. (2023). Predictive Maintenance and Anomaly Detection of Wind Turbines Based on Bladed Simulator Models. IFAC Papers OnLine (2), 4633-4638. https://doi.org/10.1016/j.ifacol.2023.10.974

[16] Fang Na, Fang Xianwen & Lu Ke. (2022). Anomalous Behavior Detection Based on the Isolation Forest Model with Multiple Perspective Business Processes. Electronics (21), 3640-3640. https://doi.org/10.3390/electronics11213640

[17] Lu Ke, Fang Xianwen & Fang Na. (2022). PN-BBN: A Petri Net-Based Bayesian Network for Anomalous Behavior Detection. Mathematics (20), 3790-3790. https://doi.org/10.3390/math10203790

[18] Jasra Sameer Kumar, Valentino Gianluca, Muscat Alan & Camilleri Robert. (2022). Hybrid Machine Learning–Statistical Method for Anomaly Detection in Flight Data. Applied Sciences (20), 10261-10261. https://doi.org/10.3390/app122010261

[19] Krishna Narayanan S., Dhanasekaran S. & Vasudevan V. (2022). An effective parameter tuned deep belief network for detecting anomalous behavior in sensor-based cyber-physical systems. Theoretical Computer Science, 142-151. https://doi.org/10.1016/j.tcs.2022.07.037

[20] Abd Algani Yousef Methkal, Vinodhini G Arul Freeda, Isabels K. Ruth, Kaur Chamandeep, Treve Mark, Kiran Bala B.... & Devi G. Usha. (2022). Analyze the anomalous behavior of wireless networking using the big data analytics. Measurement: Sensors https://doi.org/10.1016/j.measen.2022.100407

[21] Vijayanand S. & Saravanan S. (2022). A deep learning model based anomalous behavior detection for supporting verifiable access control scheme in cloud servers. Journal of Intelligent & Fuzzy Systems (6), 6171-6181. 10.3233/JIFS-212572

[22] Yang Luming, Fu Shaojing, Zhang Xuyun, Guo Shize, Wang Yongjun & Yang Chi. (2022). Flow Spectrum: a concrete characterization scheme of network traffic behavior for anomaly detection. World Wide Web (5), 2139-2161. https://doi.org/10.1007/s11280-022-01057-8

[23] Fan Pengcheng, Guo Jingqiu, Wang Yibing & Wijnands Jasper S. (2022). A hybrid deep learning approach for driver anomalous lane changing identification. Accident; analysis and prevention, 106661-106661. https://doi.org/10.1016/j.aap.2022.106661

[24] He Kaixun, Wang Tao, Zhang Fangkun & Jin Xin. (2022). Anomaly detection and early warning via a novel multiblock-based method with applications to thermal power plants. Measurement. https://doi.org/10.1016/j.measurement.2022.110979

[25] Abu AlHaija Qasem & AlDala'ien Mu'awya. (2022). ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks.

Journal of Sensor and Actuator Networks (1), 18-18. https://doi.org/10.3390/jsan11010018

[26] Moure Garrido Marta, Campo Celeste & GarciaRubio Carlos. (2022). Entropy-Based Anomaly Detection in Household Electricity Consumption. Energies (5), 1837-1837. https://doi.org/10.3390/en15051837

[27] Jiang, Jun, Wang, XinYue, Gao, Mingliang, Pan, Jinfeng, Zhao, Chengyuan & Wang, Jia. (2022). Abnormal behavior detection using streak flow acceleration. Applied Intelligence (9), 1-18. https://doi.org/10.1007/s10489-021-02881-7

[28] Hoh Maximilian, Schöttl Alfred, Schaub Henry & Wenninger Franz. (2022). A Generative Model for Anomaly Detection in Time Series Data. Procedia Computer Science629-637. https://doi.org/10.1016/j.procs.2022.01.261