# Enterprise Employee Work Behavior Recognition Method Based on Faster Region-Convolutional Neural Network

Lu Zhang
School of Economics and Management, Nanchang Institute of Science and Technology, China
zhangl052@outlook.com

**Abstract:** *The original Faster Region-Convolutional Neural Network (R-CNN) model based on Convolutional Neural Network (CNN) is not effective enough to solve the challenge of identifying employee work behavior data sets. To overcome this limitation, an innovative optimization strategy is proposed in this paper. First, we replace the traditional Visual Geometry Group (VGG) network with a Residual Network (ResNet) to ensure that the features extracted from the image are more comprehensive and detailed. Then, multi-scale feature fusion technology combined with Convolutional Block Attention Module (CBAM) is used to predict the multi-layer feature layers and further strengthen the fused feature maps. This method makes the feature map contain both high-level semantic information and low-level detail information, and provides a richer feature description for small size targets. In order to further improve the accuracy of target detection, Regional of Interest (ROI) align technology is selected to replace the traditional ROI pooling method. Through these improvements, an enhanced version of Faster R-CNN algorithm is successfully constructed. In comparison experiments, the performance of the improved Faster R-CNN algorithm is evaluated against Support Vector Machine (SVM), Extreme Learning Machine (ELM), Single Shot MultiBox Detector (SSD) and the original Faster R-CNN algorithm. The results show that under the condition of similar recognition speed, the improved Faster R-CNN shows significant advantages in recall rate, accuracy rate and accuracy rate. Specifically, the processing time of the algorithm is kept within 0.40 seconds, and the accuracy and accuracy of the algorithm have reached a high level of more than 90%.*

**Keywords:** *Enterprise staff, work behavior identification, Faster R-CNN, multi-scale features, K-means++ algorithm.*

*Received June 24, 2024; accepted January 03, 2025*
*https://doi.org/10.34028/iajit/22/2/7*

## 1. Introduction

With the rapid evolution of Artificial Intelligence (AI) technology, especially the breakthrough in the field of deep learning, more and more industries begin to use AI to optimize and monitor work processes, especially the identification and analysis of employee behaviors [32]. The key behind this trend is the ability of deep neural networks to process large amounts of data and automatically learn features from it without relying on artificially designed rules or features, which is difficult to achieve in traditional machine learning methods such as Back Propagation (BP) neural networks or Support Vector Machine (SVM) [6]. Deep learning models, such as Convolutional Neural Network (CNN), have demonstrated superior capabilities in image classification and object recognition tasks, where they are able to recognize and classify complex visual patterns with great accuracy and efficiency [11]. Compared with BP algorithm, deep learning methods usually have less computational overhead, especially when dealing with large-scale data sets, because deep learning can automatically capture higher-order abstract features in the data through multi-layer nonlinear transformations, thus simplifying the entire recognition process [30]. In the field of employee behavior recognition, the application of deep learning is not limited to improving recognition accuracy, but also to providing more detailed behavior analysis, such as action recognition, emotion detection and even predicting employee behavior. These advances are revolutionizing areas such as human resource management, safety monitoring, and productivity optimization, enabling organizations to make more informed decisions based on data. With the deepening of research, deep learning models and algorithms customized for specific scenarios continue to emerge, further enhancing the robustness and adaptability of the identification system, paving the way for the future integration of AI and workplace [35].

In recent years, with the advent of the era of big data and the leap in computing hardware capabilities, deep learning technology has made revolutionary breakthroughs in the field of image recognition and analysis, especially in the face of large-scale and complex data sets, its performance far exceeds that of traditional image processing methods [14]. This is mainly due to deep learning models, such as CNN, which can automatically learn highly abstract feature representations from massive data to achieve accurate understanding and classification of image content [13]. As a landmark algorithm in the field of object detection, Faster Region-Convolutional Neural Network (R-CNN) has become the basis and inspiration source for many subsequent studies since it was proposed [8]. Still, even Faster R-CNN faces serious challenges when it comes to behavioral object detection in complex scenarios. Complex background, mutual occlusion among targets, and diversity and variability of targets themselves greatly increase the difficulty of detection tasks [31]. In addition, the acquisition of high-quality and large-scale behavioral target annotation data is particularly difficult, mainly because behavioral targets are often small and changeable in shape, and even human eyes are difficult to accurately identify in some cases, which undoubtedly increases the complexity and cost of data annotation, thus limiting the training and optimization of deep learning models. For special objects such as behavioral objects, due to the low content of pixel-level information, traditional convolution operation is easy to lead to the loss of key details in the process of feature extraction, which is particularly unfavorable to the recognition of behavioral objects. Therefore, the development of more advanced and targeted algorithms has become the focus of current research, aiming to overcome the above problems and improve the accuracy and reliability of behavioral target detection by enhancing the model's detail capture ability and scene understanding. Although deep learning technology has shown great potential in the field of image detection, efficient and accurate detection of behavioral targets in complex scenes still requires continuous exploration and innovation by researchers, with a view to achieving more extensive and in-depth applications in the future [17].

In today's rapidly changing business environment, companies are facing unprecedented challenges, one of the most central challenges is how to effectively manage and stimulate the potential of employees to cope with increasing work stress and uncertainty. In the modern workplace, the explosion of workload, the ambiguity of role definition and the decline of job security together constitute a complex environment, which seriously erodes employees' mental health and work enthusiasm, and directly affects the overall efficiency and sustainable development of enterprises [10].

With the flattening of enterprise architecture and the reduction of management levels, employees are given greater responsibility and freedom. Although this helps to improve work efficiency and innovation, it also increases the complexity and uncertainty of work, making the role positioning and responsibility boundaries of employees more ambiguous [22]. In this context, talents become the core resources of enterprise competition. How to attract, train and retain high-quality employees, stimulate their enthusiasm and creativity, so that they can not only complete basic tasks, but also actively contribute added value, becomes the key to building competitive advantage.

In view of this, the importance of employee work behavior has become increasingly prominent. It is no longer just a spontaneous behavior at the individual level, but has become an indispensable part of enterprise management strategy. Managers are actively exploring various means to create a positive working atmosphere and guide employees to show behavioral patterns conducive to the long-term development of the organization through effective incentive mechanisms and cultural construction [9]. Research shows that positive work behavior can not only significantly improve the performance of individuals and teams, but also enhance the internal cohesion of the organization, attract more outstanding talents to join, thus forming a virtuous circle, and promote the steady growth of enterprises and the continuous improvement of market competitiveness [15]. In short, in the face of complex and changeable external environment, enterprises must attach importance to and optimize employees' work behavior, regard it as the core driving force to promote organizational change and development, and build a healthy, efficient and dynamic workplace ecology through scientific management practices and humanized work design [21].

In this paper, by adding Convolutional Block Attention Module (CBAM), the network pays more attention to the size and position of the target, and obtains richer feature information for the small-size target. Based on the analysis of employee work behavior data, K-Means ++ algorithm is used to cluster the real employee work behavior data set collected in this paper. The experimental results strongly prove that the optimized Faster R-CNN algorithm has achieved a significant jump in performance, showing obvious advantages compared with the four algorithms. This improvement not only ensures that the identification speed of the algorithm can meet the needs of actual engineering application scenarios, but more importantly, it marks a key step in the field of intelligent identification and analysis of employee work behavior, and steadily advances towards a more refined and intelligent direction.

The rest of this article is organized as follows. Related work is discussed in section 2. In section 3, an improved Faster R-CNN algorithm is designed. In section 4, experiments are carried out and the results are analyzed. Section 5 summarizes the full text.

## 2. Related Work

In the field of deep learning, in view of the limitations of CNN in processing variable size inputs, researchers creatively designed the Spatial Pyramid Pooling Net (SPP-Net) and incorporated the innovative mechanism of spatial pyramid layer into it. This breakthrough enables CNN to directly process input images of any size and output feature vectors of uniform length without pre-scaling the candidate region, greatly expanding the application range and flexibility of CNN [7]. Thanks to the efficient processing power of SPP-Net, the speed of target detection has been significantly improved. At the same time, the recognition accuracy is not reduced, and the double guarantee of speed and accuracy is realized.

Then, on the basis of SPP-Net, Fast R-CNN further simplifies and optimizes the entire detection process, especially in the speed to achieve a qualitative leap, compared with R-CNN, its efficiency advantage is particularly prominent. However, despite the excellent performance of Fast R-CNN in many aspects, the generation of candidate regions is still a key link restricting its overall performance, which has become the bottleneck to further improve the recognition speed.

In order to solve this problem, Faster R-CNN came into being, which cleverly introduced Region Proposal Network (RPN), effectively overcame the speed barrier in the process of candidate region extraction, and realized the leap from “fast” to “faster”. Faster R-CNN not only inherits the advantages of the previous generation model, but also finds a new balance between speed and accuracy, marking that object detection technology has entered a new stage of development, and opens up broader possibilities for subsequent computer vision applications.

After Fast R-CNN, researchers quickly introduced Faster R-CNN, which marked the first time in the field of deep learning to approach the real-time processing speed of object detection [19]. A key innovation of Faster R-CNN is the introduction of RPN, which integrates feature extraction, candidate region generation and bounding box adjustment into a unified framework and greatly optimizes the detection process [25]. The dual-function feature of RPN, that is, the classification confidence and location information of the target are calculated simultaneously, further improves the detection efficiency and accuracy [33].

To overcome the limitation of single-scale feature extraction, researchers designed Feature Pyramid Network (FPN), a multi-scale network designed to capture semantic information from different scales, which is often used as a supplement to other detection networks to achieve more comprehensive object detection [28]. FPN has been widely used to prove its effectiveness in multi-scale feature extraction and is widely integrated in many target detection models.

On this basis, Mask R-CNN comes into being. It adopts Residual Network (ResNet) and FPN as feature extractors, which not only improves the detection accuracy, but also can perform instance segmentation, that is, distinguish different object instances in the image [18]. In addition, Graphic Feature Pyramid Network (GraphFPN), as a novel GraphFPN, aims to dynamically adjust according to the inherent structure of images, while promoting the collaborative fusion of cross-scale features, significantly enhancing the adaptability and robustness of the model to complex scenes [1]. In order to further improve the performance, the researchers also introduced the local channel attention mechanism, which effectively strengthened the expression of multi-scale features, thus improving the overall performance of the detection model [24].

Recently, the in-depth exploration of loss function has become one of the key factors to improve the efficiency of target detection system. Aiming at the common problem of positive and negative sample imbalance in multi-scale detection, researchers proposed an innovative adaptive variance weighting strategy, which effectively balanced the loss contributions at different scales, and dynamically adjusted the training weights combined with reinforcement learning algorithms to significantly enhance the generalization ability of the model [4]. Another study focused on the challenge of class imbalance and developed a ranking loss function known as Rank and Sort (RS) loss, which not only simplified the training process, but also specifically improved the stability and detection accuracy of the model when processing long-tail distributed data [2]. In addition, some scholars advocate the combination of natural language understanding and computer vision, and propose an end-to-end modulation detector, which aims to enrich the training materials of target detection by integrating text descriptions, so as to optimize the recognition effect of the model [5].

In the field of organizational behavior, scholars have explored the relationship between Organizational Citizenship Behavior (OCB) and Counter-productive Work Behavior (CWB). According to the traditional view, high OCB level often corresponds to low CWB level, and the two are negatively correlated [16]. However, subsequent research has revealed the complexity of this relationship, pointing out that in some situations both may be high at the same time, challenging the previous simple linear assumption [29]. Further analysis shows that OCB contributes to the construction and maintenance of employees’ social psychological state and indirectly promotes the improvement of task performance [26]. This means that by actively cultivating civic behavior within the organization, companies can virtually optimize the team atmosphere, which in turn increases overall productivity. In order to accurately assess OCB, researchers have carefully designed a comprehensive scale with two sub-dimensions, aiming to more comprehensively capture the diversity of this behavior pattern and its potential impact on organizational performance [20].

# 3. Methods

## 3.1. Faster Improvement of R-CNN Feature Extraction Module

The target detection and classification process of Faster R-CNN can be summarized as follows: First, RPN is responsible for generating a series of Bounding Boxes on images that may contain target objects. These candidate boxes are then mapped onto the feature maps previously generated by the CNN to obtain their respective corresponding feature representations. Then, the dimensions of these feature matrices are standardized by the Regional Of Interest (ROI) pooling layer to ensure that they can be uniformly input into the subsequent full connection layer. After being processed by the two fully connected layers, the features are fed into the classification branch and the bounding box regression branch. The former is used to predict the category to which each candidate region belongs, while the latter is responsible for adjusting the position of the bounding box to more precisely locate the target.

In the Visual Geometry Group (VGG) network architecture, a smaller 3x3 convolution kernel is used to replace the traditional 5x5 or 7x7 large size convolution kernel. This design not only maintains a receptive field range comparable to that of large convolution kernel, but also reduces the computational complexity of the network effectively due to the small number of parameters. More importantly, the use of multiple 3x3 convolution cores in series introduces additional nonlinear operations, which enhances the hierarchical representation of the network and the ability to learn complex features. VGG16 network, with its depth and high efficiency, has become the preferred backbone network in many target detection tasks, providing a powerful feature extraction basis for subsequent detection and classification.

Among the CNN architectures, ResNet stands out because of its excellent capability of deep feature extraction, showing better model performance. In order to achieve more comprehensive and accurate feature capture, ResNet50 and ResNet101 are preferred as upgrade solutions for feature extraction components. These two variants enhance the model's learning of complex patterns by increasing the depth of the network, while avoiding the gradient disappearance or explosion problems caused by increasing depth.

When constructing a very deep neural network, if the error gradient of each layer is more than 1, the deep layer of the network is prone to gradient explosion, resulting in unstable weight updating. In addition, with the application of nonlinear activation functions such as ReLU, the information flow will undergo irreversible transformation in the forward propagation process, which may cause information loss and affect the network performance. To overcome these problems, ResNet introduces a residual module.

The core idea of the residual module is that the network no longer learns the mapping relationship directly from input to output, but instead learns the residual mapping between the input and the ideal output. This strategy simplifies the learning process, allowing the network to effectively learn shallow features even at deep structures, without having to learn the complex mapping of each layer from scratch. Therefore, even when the network depth is significantly increased, ResNet can still maintain or even improve the model performance, avoid the performance degradation problem commonly faced by deep networks, and ensure the effectiveness and stability of deep networks. The mathematical formula of the ResNet structure is as Equation (1):

$$\begin{cases} x_{l+1} = f(y_l) \\ y_l = F(x_l, W_l) + H(x_l) \end{cases} \tag{1}$$

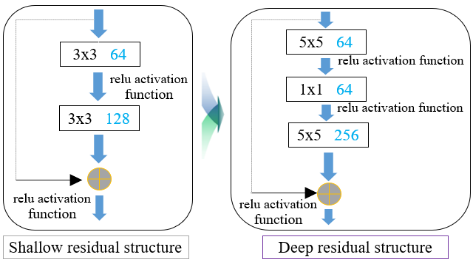
$H$ stands for identity mapping. The learning characteristics from shallow to deep level can be expressed as Equation (2):

$$x_L = x_l + \sum_{i=l}^{L} F(W_i, x_i) \tag{2}$$

By using the chain normal form for derivation, the gradient of the backpropagation process is obtained as Equation (3):

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \bullet \left[ 1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L} F(W_i, x_i) \right] \tag{3}$$

The result of this derivation must always be greater than 1, and the continuous multiplication operation in the previous derivation becomes a continuous addition operation. Two residual structures are used in the ResNet network, as shown in Figure 1.



Shallow residual structure

Deep residual structure

Figure 1. ResNet different network structures.

When the dimensions of the input and output are different, a 5×5 convolution kernel is added during the jump to increase the dimension. This particular network design, often referred to as a "bottleneck" structure, cleverly uses 1x1 convolution nuclei to regulate the number of channels in the feature graph. Specifically, it first reduces the number of channels in the input feature map by 1x1 convolution, then extracts the core feature by 3x3 convolution, and finally restores the number of channels to the original size or makes appropriate adjustments by 1x1 convolution. In this way, the operation of 3x3 convolutional kernel is no longer limited by the width of input features, so that the number of parameters and computational complexity of the model are effectively controlled while ensuring the efficiency of information transmission, and the high-performance and lightweight design of deep networks is realized. The 1×1 convolution layer in the middle passes through the 1×1 convolution kernel, thus reducing the computational effort.

Using the ResNet network architecture, its unique shortcut connections can significantly reduce the difficulty and complexity of model training when faced with low complexity data sets.

## 3.2. Improvement Based on Multi-Scale Features

In the Faster R-CNN architecture, the deepest feature map of the network is usually used for target prediction. As the depth of the network increases, the extracted features are gradually rich in semantic information. However, this also means that the size of the feature map is relatively reduced, and the spatial resolution decreases accordingly, which undoubtedly increases the difficulty of detecting those targets that occupy a small area in the image.

To address this challenge, an effective strategy is to perform two up-sampling operations on the shallow feature map and then fuse it with the current deep feature map. This approach cleverly combines rich semantic detail at the top level with high-resolution information at the bottom, significantly improving detection of small or dense targets, especially when dealing with data sets containing tiny objects.

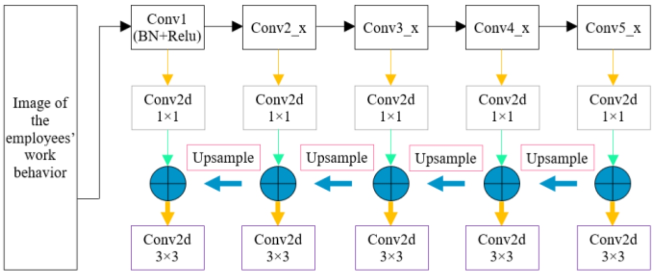


Figure 2. Feature fusion network.

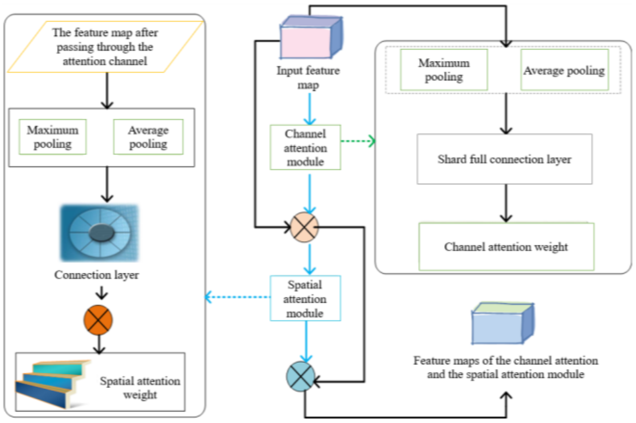


Figure 3. CBAM module structure.

Figure 2 shows an example of feature fusion based on ResNet50 as the backbone network, and clearly shows how to achieve feature representation that is both fine and rich in context information through cross-level feature fusion, so as to optimize the detection performance of behavioral targets by Faster R-CNN. By predicting different feature layers and performing feature fusion, rich details and semantic information are aggregated. The CBAM module is shown in Figure 3.

The two dimensions of the feature channel and space are processed in a layer-by-layer approach. CBAM refines the input feature map through its built-in channel attention mechanism, enabling the deep learning model to learn the relative information between different feature channels. This process enables the network to dynamically adjust during the training process and learn to assign weights to each channel in the input feature layer, thereby highlighting key information, suppressing irrelevant or redundant features, and improving the focus perception ability and decision-making efficiency of the model. The channel attention module is as Equation (4):

$$M_C(F) = \sigma\left[MLP(\max\ pool(F)) + 2MLP(avgpool(F))\right] \quad (4)$$

The Sigmoid activation function is applied directly to the output of 8x8 convolution kernel to generate a spatial attention diagram. This attention map is then multiplied element-by-element (dot product) with the original input feature map. Spatial attention modules can be represented as Equation (5):

$$M_C(F) = \sigma\left[Conv^{8*8}(\max\ pool_C(F), avgpool_C(F))\right] \quad (5)$$

## 3.3. Improvement of ROI Pooling Module

In object detection frameworks for deep learning, such as Faster R-CNN, the RPN network is responsible for generating a set of candidate Regions that may contain target objects, called ROI. Because RPNs generate ROIs of varying sizes and shapes, they cannot be fed directly to fully connected layers that require fixed-size inputs. To this end, the ROI Pooling module is introduced to standardize the size of these ROIs.

Specifically, ROI Pooling works as follows:

The candidate box of RPN output is mapped back to the feature map extracted by the backbone network to obtain the corresponding feature region.

These feature regions are evenly divided, usually into n×n subregions, where n is predefined to determine the size of the final output feature vector. A maximum pooling operation is performed within each subregion, which ensures that the information for each subregion is preserved while being compressed into a fixed-size feature vector.

The whole process involves two operations: The first is when mapping the candidate boxes in the original image coordinate system to the feature map coordinate system, because the feature map is usually obtained by downsampling, so the coordinates need to be converted, and this conversion may lead to decimal coordinates, which need to be converted to integer coordinates. The second time occurs when the candidate box on the feature map is divided into subregions, and the boundaries of the subregions also need to be quantized with integers in order to facilitate calculation.

When dealing with large-size targets, because the target occupies a relatively large area on the feature map, the influence of two processing is small, so the accuracy of target detection is limited. However, when the target is very small, especially when it is close to or smaller than a feature map pixel, this quantization error can significantly affect the detection performance. Because on small targets, each pixel can carry important information. Therefore, when designing detectors for dealing with small targets, special attention needs to be paid to reducing the impact of such quantization errors.

However, for some relatively small size targets, when the operation is performed, the candidate box position on the real image will have a large error. In Mask-RCNN network, ROI Align is proposed to improve the accuracy of ROI Pooling detection on behavioral targets. Compared with ROI Pooling, the difference between ROI Aligin and ROI Pooling is that there is no direct quantitative processing like ROI Pooling, resulting in low precision. It converts the whole process of feature aggregation into continuous operations. The detailed steps are shown in Figure 4.

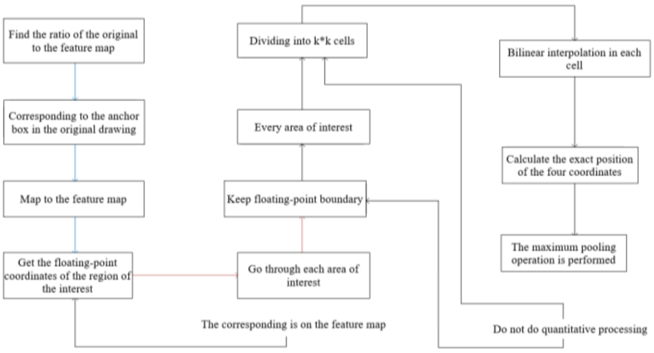

Find the ratio of the original to the feature map
Corresponding to the anchor box in the original drawing
Map to the feature map
Get the floating-point coordinates of the region of the interest
Dividing into k*k cells
Every area of interest
Keep floating-point boundary
Go through each area of interest
The corresponding is on the feature map
Bilinear interpolation in each cell
Calculate the exact position of the four coordinates
The maximum pooling operation is performed
Do not do quantitative processing

Figure 4. Improvement steps of the ROI pooling module.

## 3.4. Faster R-CNN Loss Function Analysis

### 3.4.1. RPN Loss Function

In anchor box-based object detection algorithms such as Faster RPN or YOLO, anchor box adjustment is a key step that includes two main tasks: classification of anchor boxes and regression of bounding boxes. The loss function of the RPN network is shown as Equation (6):

$$L(t_i ; p_i) = \lambda \bullet N_{cls} \bullet \sum_i L_{cls}(p_i^{'}, p_i) + N_{reg} \bullet \sum_i p_i^{'} \bullet L_{reg}(t_i^{'}, t_i) \quad (6)$$

The classification loss function $L_{cls}$ is a Binary Cross Entropy (BCE) with only two categories of foreground and background, and its function expression is as Equation (7):

$$L_{cls}(p_i^{'} ; p_i) = p_i^{'} \bullet \log(p_i) - (1-p_i^{'}) \bullet \log(1-p_i) \quad (7)$$

The bounding box regression loss function adopts SmoothL1 loss function, whose function expression is as Equations (8) and (9):

$$\begin{cases} t_i = \begin{pmatrix} t_x & t_y & t_w & t_h \end{pmatrix} \\ t_i^{'} = \begin{pmatrix} t_x^{'} & t_y^{'} & t_w^{'} & t_h^{'} \end{pmatrix} \end{cases} \quad (8)$$

$$L_{reg}(t_i^{'}, t_i) = \sum_i Smooth_{L_1} \left| t_i^{'} - t_i \right| \quad (9)$$

$Smooth_{L1}$ loss function is expressed as Equation (10):

$$Smooth_{L_1}(x) = \begin{cases} x^2 / 2 & -1 < x < 1 \\ |x| - 1 & x \le -1, x \ge 1 \end{cases} \quad (10)$$

### 3.4.2. Classification Regression Loss Function

The functional expression of the classification regression loss function is as Equation (11):

$$L(p, u, v, t^u) = L_{cls}(p_i^{'}, p_i, v) + \lambda \bullet u \bullet p_i^{'} \bullet t^u \quad (11)$$

Here, $p$ represents the Softmax probability distribution predicted by the classifier $p$=($p_0$,... $p_k$), tu represents the regression parameters of the corresponding class u predicted by the bounding box regressor: $v_x$, $v_y$, $v_h$, and $v_w$, used to adjust the specific importance of the two loss functions so that the two loss functions can achieve an equilibrium effect. The only difference between the classification loss function and the RPN classification loss function is the Softmax Cross Entropy (SCE), expressed as Equation (12):

$$L_{cls}(p_k, u) = -\log p_u \quad (12)$$

In target detection, if the size of the defect is too large, the Euclidean distance will reduce the accuracy of positioning. Therefore, 1-Intersection Over Union (IOU) is used as the definition of distance in this paper.

## 3.5. Improved Anchor Frame Based on K-Means++ Algorithm

In Faster R-CNN, the RPN aims to generate a series of candidate regions (or anchors) that may contain target objects. The original RPN designed a series of prior candidate boxes with fixed dimensions and proportions to cover targets of different sizes and shapes.

The original RPN network's default size and scale Settings (for example, 128x128, 256x256, 512x512 sizes and 1:1, 1:2, 2:1 ratios) may not be fully applicable to specific data sets, such as the work behavior data set of enterprise employees. When these prior candidate boxes do not match the size and shape of the target in the dataset, the candidate regions generated by the RPN may not be accurate enough, which will affect the subsequent classification and bounding box regression performance.

In other words, if the anchor size and scale used by the RPN network are not consistent with the statistical characteristics of the target dataset, the RPN may produce many candidate boxes that are significantly different from the actual location and size of the target. These inaccurate candidate boxes not only increase the difficulty of subsequent classification and regression tasks, but also may cause the network to learn wrong feature representations, thus affecting the performance of the entire detection system.

In the K-means++algorithm, clustering is performed using the Euclidean distance between samples and cluster centers. However, for object detection, if the size of the defect is too large, the Euclidean distance will reduce the accuracy of localization and lower the performance of the object detection algorithm. Therefore, this article uses 1-IOU as the definition of distance. The smaller the value d (box, centroid) from the target label box to the cluster center label box, the smaller the distance between the two. The function expression of AvgIOU is as Equation (13):

$$p = \arg\max \sum_{i=1}^{k-1} \sum_{j-1}^{m_i-1} I_{IOU}(box, centroid) / m \qquad (13)$$

To assess the match between anchor boxes and ground truth bounding boxes, it is common to calculate the IOU value between a set of anchor boxes and each labeled box. The IOU is a commonly used metric to measure the proportion of the total area where two rectangles overlap, with a value ranging from 0 to 1, with the closer the value to 1, the better the match.

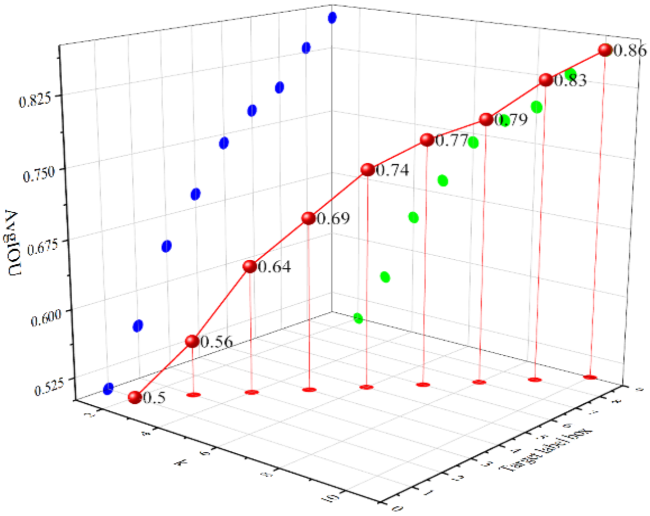


Figure 5. AvgIOU under different k-values.

Based on the recognition of employee work behavior in enterprises, the K-means++algorithm is used to read the annotated XML file from the dataset and obtain the width and height of the target annotation box. We randomly selected k values from the obtained data as the initial values for k anchor boxes, and calculated the IOU values for each annotation box and the initial k anchor boxes. By comparing the IOU values of the annotation box and the k anchor boxes set, each annotation box is assigned to the nearest anchor box until the anchor box no longer changes. In this paper, k from 2-10 was selected for cluster analysis [12], and AvgIOU values under different k-values were obtained, as shown in Figure 5.

The trend observed in Figure 5 shows that as the number of K-means clustering centers (k-value) increases, the AvgIOU value shows an upward trend, but its growth rate gradually slows down. This phenomenon reflects the delicate balance between anchor frame design and object detection performance. At the initial stage, increasing the type of anchor frame can significantly improve the adaptability of the model to different object sizes and proportions, thus improving the detection accuracy. However, as the k-value continues to increase, although the AvgIOU value is still improving, the marginal benefit is diminishing, meaning that the contribution of the additional anchor frame type to performance is gradually diminishing. Considering the trade-off between computational efficiency and detection accuracy of the model, you customized a set of nine anchor frames with different sizes and proportions for Faster R-CNN. The logic behind this decision is that by properly setting the diversity of anchor frames, the performance of the detection system can be maximized without significantly increasing the computational burden. In practice, the model makes an initial prediction of the candidate region (i.e., the anchor frame) on multiple feature layers. Each feature layer is responsible for object detection at different scales, which benefits from the design concept of FPN, that is, feature maps at different levels are used to capture multi-scale information, and the size of anchor box is obtained when k value is 9.

## 4. Experimental Results and Analysis

### 4.1. Experimental Environment and Parameter Settings

The experimental platform is built on an HP server equipped with high-end hardware, running 64-bit Windows operating system, equipped with a six-core Intel i7-7800X processor, the main frequency of up to 3.5GHz, supplemented by NVIDIA GeForce RTX 2080 GPU, providing powerful graphics processing capabilities. In terms of storage, the system integrates 2TB traditional hard disk and 512GB solid state drive to ensure high-speed and stable data reading and writing, and is also equipped with 64GB large-capacity memory, providing sufficient resources for the operation of deep learning algorithms. This study uses Python programming language and relies on Pytorch 1.3.0 deep learning framework to build the model, which has become the preferred tool for deep learning researchers with its concise and easy to understand API and excellent debugging support. In order to accelerate parallel computing on the GPU, CUDA 10.0 and CUda Deep Neural Network (CUDNN) 11.0 are installed to fully leverage the computing potential of the GPU. It is worth mentioning that Pytorch 1.2.0 has been optimized in the TorchScript environment to enhance the flexibility and efficiency of model building, allowing developers to focus more on algorithmic innovation rather than tedious code debugging.

The experimental data came from the actual enterprise scenario, collected a series of pictures reflecting the employee's work behavior, and constructed a proprietary dataset, covering five categories: maintenance behavior, in-role work behavior, out-of-role work behavior, counterproductive production behavior and turnover behavior. This article divides the dataset into a training set and a testing set in an 8:2 ratio.

In order to ensure the training quality of the model, the data set is preprocessed, including image normalization and data enhancement, to improve the generalization ability of the model. During model training, standard deep learning practices are followed. After each epoch, model performance is evaluated using validation sets, changes in loss values are closely monitored, and training strategies are adjusted in time.

In order to filter the high-quality prediction results, the confidence threshold of the detection algorithm is set to 0.55, and only those boundary boxes whose confidence is higher than the threshold is kept, which effectively reduces the redundant calculation. At the same time, the Non-Maximum Suppression (NMS) policy is set to 0.3 to eliminate overlapping detection boxes and further refine the output. In the stage of experimental evaluation, you pay attention to the identification accuracy, recall rate and processing time of the model, which comprehensively reflect the performance of the model in identifying the work behavior of enterprise employees. By comparing the performance of different models, the effectiveness of the algorithm can be evaluated objectively, and provide valuable reference for subsequent research. This series of experimental design fully embodies the powerful ability of deep learning in solving complex visual recognition tasks, and opens up a new path for improving the intelligent level of enterprise human resource management.

## 4.2. Learning Rate Automatic Adjustment Results

In the training process of deep learning model, the choice of learning rate is particularly critical, which directly affects the convergence speed and final performance of the model. If the learning rate is too high, although it can accelerate the training process, it may make the optimization algorithm "jump" on the surface of the loss function, and it is difficult to reach the global minimum, resulting in unstable convergence of the model or falling into the local optimal solution. On the contrary, although too low learning rate can ensure that the algorithm smoothly approaches the optimal solution, it will significantly prolong the training cycle, and even make the model stall due to too small step size, and cannot continue to learn.

For this reason, the ideal strategy is to have the learning rate dynamically adjust with the training progress, rather than being fixed. In this study, the ReduceLROnPlateau learning rate scheduler provided by PyTorch framework was adopted to take the val_loss on the verification set as the monitoring index. Specifically, when val_loss does not decrease significantly for 10 consecutive epoches, the learning rate is automatically lowered, each time by half of the current value, until the minimum limit of 1.0e-6 is reached. This strategy takes into account the need of fast convergence and preventing overfitting, and reduces the learning rate in time to promote the model to explore more carefully in the later training period.

The learning rate of the employee work behavior recognition model established in this paper starts from the initial value of 6.0e-6, doubles each time, and finally drops to 1.0e-6. In addition, the validation set loss fluctuates significantly due to the fluctuation of the learning rate, which also indicates the effectiveness of the training strategy selected in this paper for automatically adjusting the learning rate of the employee work behavior recognition model. Figure 6 shows how the learning rate and loss value change with the number of iterations.

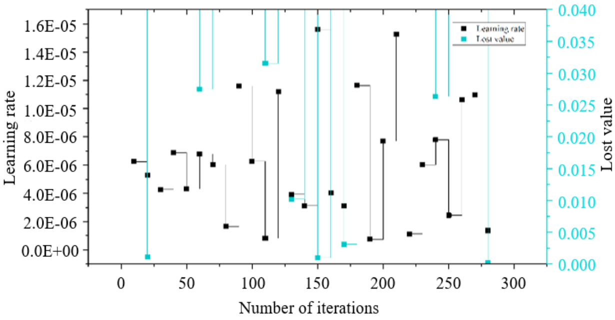


Figure 6. Changes of learning rate and loss value with the number of iterations.

In addition to dynamically adjusting the learning rate according to val_loss, the study also introduced the "learning rate warm-up" strategy, which is an effective means to optimize the initial performance of the model training. In the initial stage of model training, because most of the weight parameters are in the random initialization state, the performance of the model is often unstable and may experience large fluctuations in loss value. At this time, if the preset learning rate is directly used for training, the instability may be aggravated, resulting in the difficulty for the model to quickly enter the stable learning state.

To address this, the warm-up learning strategy advocates setting the learning rate at a relatively low level for the first few epoches of training. The implementation of this strategy is equivalent to giving the model an “adaptation period,” allowing it to gradually adjust the parameters under a small learning step, so as to smoothly transition to the normal training stage. With the gradual stability of the model, after the end of the warm-up phase, the model returns to the original planned learning rate level. At this time, the model has a good initial state, and can carry out deep learning more efficiently and stably, and pursue better performance.

## 4.3. Training Effect Analysis

Figure 7 shows the loss function value curve of the improved Faster R-CNN employee work behavior recognition algorithm. The loss function of the improved algorithm starts at 0.0026 and reaches about 0.00001 after 280 epochs.

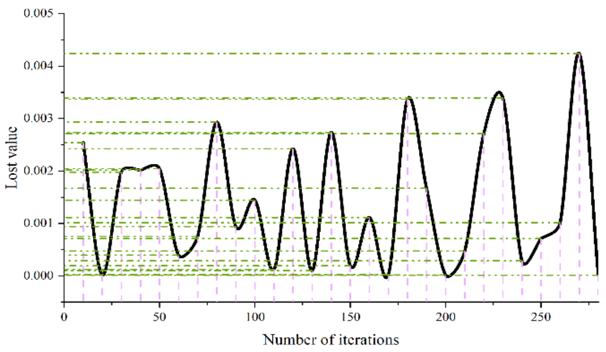


Figure 7. Loss value of the improved Faster R-CNN algorithm.

## 4.4. Analysis of Identification Accuracy and Accuracy

Figure 8 shows the recall accuracy (P-R) curves for testing five types of employee work behaviors (maintenance behavior, on-the-job behavior, off-the-job behavior, behavior harmful to production, and turnover behavior) using SVM [3], Extreme Learning Machine (ELM) [23], Single Shot MultiBox Detector (SSD) [27, 34], Faster R-CNN, and improved Faster R-CNN algorithms.

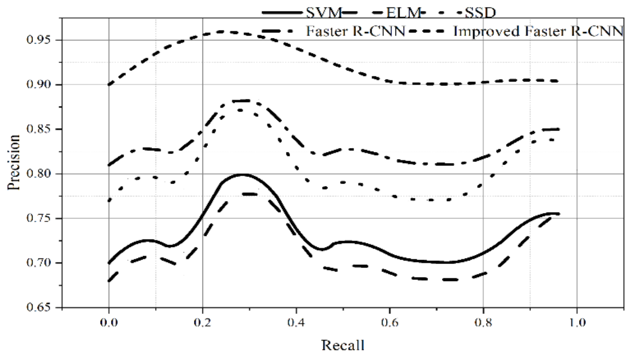


Figure 8. Comparison of P-R curves of different algorithms.

Through careful analysis of the precision-recall curve of employees’ work behavior, it can be clearly observed that the optimized Faster R-CNN algorithm not only exceeds the original Faster R-CNN in terms of the area enclosed by the curve, it is also superior to other competing algorithms such as SVM, ELM and SSD. This result intuitively indicates that the improved Faster R-CNN algorithm shows better detection accuracy in terms of identifying the work behaviors of enterprise employees, and its ability to capture subtle behavior characteristics and reduce false positives and missing positives is significantly improved.

In order to further verify the performance of the algorithm in practical applications, a set of 1000 employee behavior images from real enterprise scenarios are used as a test set. The improved Faster R-CNN algorithm also performs well in this series of tests, not only accurately identifies various work behaviors, but also shows a stable recognition rate, which proves its robustness and practicability in complex environments. The comparison of recognition accuracy rates shown in Figure 9 further consolidated the leading position of the improved algorithm in the field of employee behavior recognition, and provided a strong technical support for improving the efficiency of enterprise management and monitoring.

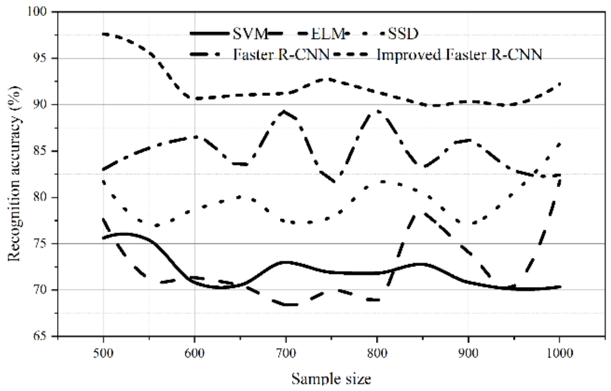


Figure 9. Comparison of the recognition accuracy of employees' work behaviors by different algorithms.

## 4.5. Time Consuming Analysis

Under the same hardware environment, SVM, ELM, SSD, Faster R-CNN and improved Faster R-CNN algorithm are used to detect the work behavior images of enterprise employees in the test set. The comparison of the time required for recognition of different models is shown in Figure 10.

The improved Faster R-CNN algorithm generally takes a longer time, up to 0.40 seconds, because the improved Faster R-CNN algorithm is more complex than other algorithms. It introduces attention mechanisms and FPN structures that increase recognition time. However, in terms of time, the improved Faster R-CNN algorithm still meets the practical application requirements of employee work behavior identification.

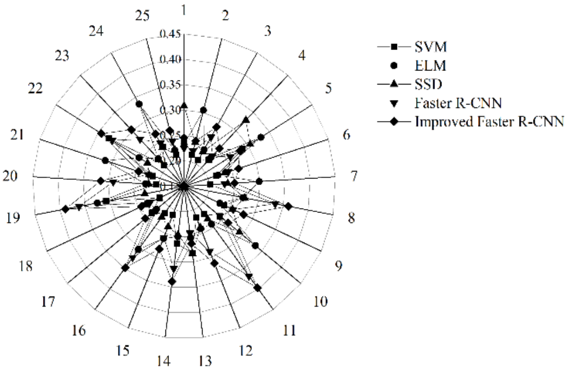


Figure 10. Comparison of time required by different algorithms.

## 5. Conclusions

The original Faster R-CNN model is unable to identify employees' work behavior effectively. To solve this problem, the feature extraction network was improved and a deeper ResNet50 residual network was used to replace the original VGG16. The newly generated feature map is obtained, and richer feature information is obtained. In this study, a detailed set of employee work behavior data is collected, and then the optimized K-Means++ algorithm is used to conduct in-depth analysis and clustering of these data. Based on the results of cluster analysis, the size of the anchor frame is dynamically adjusted and optimized, which significantly improves the training efficiency of the model and the speed and accuracy of target recognition. By comparing the experimental results of SVM, ELM, SSD, Faster R-CNN and the improved Faster R-CNN algorithm, it is concluded that the improved Faster R-CNN algorithm performs better than the other four algorithms in terms of accuracy and precision when there is no significant difference in recognition speed.

## Acknowledgements

This work was supported and funded by the project: Education Science 14th "Five Year" Plan Project of Jiangxi province in 2023, Exploration and Research on Data Literacy Education Based on SNA Social Network Analysis-Taking Jiangxi Universities as an example (23QN074).

## References

[1] Aboramadan M., "The Effect of Green HRM on Employee Green Behaviors in Higher Education: The Mediating Mechanism of Green Work Engagement," *International Journal of Organizational Analysis*, vol. 30, no. 1, pp. 7-23, 2022. https://doi.org/10.1108/IJOA-05-2020-2190

[2] Afzal C., Khan S., Baig F., and Ashraf M., "Impact of Green Human Resource Management on Environmental Performance: The Mediating Role of Green Innovation and Environmental Strategy in Pakistan," *Review of Applied Management and Social Sciences*, vol. 6, no. 2, pp. 227-242, 2023. https://doi.org/10.47067/ramss.v6i2.311

[3] Akpinar M., Adak M., and Guvenc G., "SVM-based Anomaly Detection in Remote Working: Intelligent Software SmartRadar," *Applied Soft Computing*, vol. 109, pp. 107457, 2021. https://doi.org/10.1016/j.asoc.2021.107457

[4] Asad M., Samad A., Khan A., and Khan A., "Green Human Resource Management Perception in the Corporate Sectors of Khyber Pakhtunkhwa, Pakistan," *Journal of Environmental Science and Economics*, vol. 1, no. 4, pp. 51-60, 2022. https://doi.org/10.56556/jescae.v1i4.397

[5] Bahuguna P., Srivastava R., and Tiwari S., "Two-Decade Journey of Green Human Resource Management Research: A Bibliometric Analysis," *Benchmarking: An International Journal*, vol. 30, no. 2, pp. 585-602, 2023. https://doi.org/10.1108/BIJ-10-2021-0619

[6] Bibi S., Khan A., Hayat H., Panniello U., Alam M., and Farid T., "Do Hotel Employees Really Care for Corporate Social Responsibility (CSR): A Happiness Approach to Employee Innovativeness," *Current Issues in Tourism*, vol. 25, no. 4, pp. 541-558, 2022. https://doi.org/10.1080/13683500.2021.1889482

[7] Chouchane R., Fernet C., Austin S., and Zouaoui S., "Organizational Support and Intrapreneurial Behavior: On the Role of Employees' Intrapreneurial Intention and Self-Efficacy," *Journal of Management and Organization*, vol. 29, no. 2, pp. 366-382, 2023. DOI:10.1017/jmo.2021.14

[8] Elsayed A., Zhao B., Goda A., and Elsetouhi A., "The Role of Error Risk Taking and Perceived Organizational Innovation Climate in the Relationship between Perceived Psychological Safety and Innovative Work Behavior: A Moderated Mediation Model," *Frontiers in Psychology*, vol. 14, pp. 1042911, 2023. DOI:10.3389/fpsyg.2023.1042911

[9] Gong Y. and Wang L., "Teacher Professional Identity, Work Engagement, and Emotion Influence: How Do they Affect Teachers' Career Satisfaction," *International Journal of Education, Science, Technology, and Engineering*, vol. 6, no. 2, pp. 80-92, 2023. https://doi.org/10.36079/lamintang.ijeste-0602.611

[10] He P., Zheng W., Zhao H., Jiang C., and Wu T., "Citizenship Pressure and Knowledge Hiding: The Mediating Role of Citizenship Fatigue and the Moderating Role of Supervisor-Subordinate Guanxi," *Applied Psychology*, vol. 73, no. 2, pp. 565-598, 2024. https://doi.org/10.1111/apps.12490

[11] Khan A., Qureshi M., Hussain K., Abbas Z., and Munawar S., "Corporate Social Responsibility

Promotes Organisation Citizenship and Pro-Environmental Behaviours: The Employee's Perspective," *Organizacija*, vol. 56, no. 2, pp. 106-124, 2023. https://sciendo.com/de/article/10.2478/orga-2023-0008

[12] Li H. and Wang J., "Collaborative Annealing Power K-Means++ Clustering," *Knowledge-Based Systems*, vol. 255, pp. 109593, 2022. https://doi.org/10.1016/j.knosys.2022.109593

[13] Li M., Khan H., Chughtai M., and Le T., "Innovation Onset: A Moderated Mediation Model of High-Involvement Work Practices and Employees' Innovative Work Behavior," *Psychology Research and Behavior Management*, vol. 15, pp. 471-490, 2022. DOI:10.2147/PRBM.S340326

[14] Liu W., Zhu Y., Chen S., Zhang Y., and Qin F., "Moral Decline in the Workplace: Unethical Pro-Organizational Behavior, Psychological Entitlement, and Leader Gratitude Expression," *Ethics and Behavior*, vol. 32, no. 2, pp. 110-123, 2022. https://doi.org/10.1080/10508422.2021.1987909

[15] Liu X., Yu J., Guo Q., and Li J., "Employee Engagement, its Antecedents and Effects on Business Performance in Hospitality Industry: A Multilevel Analysis," *International Journal of Contemporary Hospitality Management*, vol. 34, no. 12, pp. 4631-4652, 2022. https://doi.org/10.1108/IJCHM-12-2021-1512

[16] Lysova E., Tosti-Kharas J., Michaelson C., Fletcher L., Bailey C., and McGhee P., "Ethics and the Future of Meaningful Work: Introduction to the Special Issue," *Journal of Business Ethics*, vol. 185, no. 4, pp. 713-723, 2023. https://link.springer.com/article/10.1007/s10551-023-05345-9

[17] Ma H., Tang S., and Zhao C., "CEOs' Leadership Behaviors and New Venture Team Stability: The Effects of Knowledge Hiding and Team Collectivism," *Frontiers in Psychology*, vol. 13, pp. 1001277, 2022. https://doi.org/10.3389/fpsyg.2022.1001277

[18] Malibari M. and Bajaba S., "Entrepreneurial Leadership and Employees' Innovative Behavior: A Sequential Mediation Analysis of Innovation Climate and Employees' Intellectual Agility," *Journal of Innovation and Knowledge*, vol. 7, no. 4, pp. 100255, 2022. https://doi.org/10.1016/j.jik.2022.100255

[19] Maqsoom A., Umer M., Alaloul W., Salman A., Fahim Ullah., Ashraf H., and Musarat M., "Adopting Green Behaviors in the Construction Sector: The Role of Behavioral Intention, Motivation, and Environmental Consciousness," *Buildings*, vol. 13, no. 4, pp. 1-20, 2023. https://www.mdpi.com/2075-5309/13/4/1036

[20] Nabi M., Liu Z., and Hasan N., "Examining the Nexus between Transformational Leadership and Follower's Radical Creativity: The Role of Creative Process Engagement and Leader Creativity Expectation," *International Journal of Emerging Markets*, vol. 18, no. 10, pp. 4383-4407, 2023. https://doi.org/10.1108/IJOEM-05-2021-0659

[21] Ogunfowora B., Nguyen V., Steel P., and Hwang C., "A Meta-Analytic Investigation of the Antecedents, Theoretical Correlates, and Consequences of Moral Disengagement at Work," *Journal of Applied Psychology*, vol. 107, no. 5, pp. 746-775, 2022. DOI: 10.1037/apl0000912

[22] Pu B., Ji S., and Sang W., "Effects of Customer Incivility on Turnover Intention in China's Hotel Employees: A Chain Mediating Model," *Journal of Hospitality and Tourism Management*, vol. 50, pp. 327-336, 2022. https://doi.org/10.1016/j.jhtm.2022.02.004

[23] Reyes-Menendez A., Saura J., and Martinez-Navalon J., "The Impact of e-WOM on Hotels Management Reputation: Exploring Tripadvisor Review Credibility with the ELM Model," *IEEE Access*, vol. 7, pp. 68868-68877, 2019. https://ieeexplore.ieee.org/document/8723076

[24] Shafiq M., Ramzan M., Faisal M., and Iqbal S., "Exploring the Relationship between Green Human Resource Management and Green Creativity: The Moderating Influence of Green Behavioral Intention," *Pakistan Journal of Humanities and Social Sciences*, vol. 11, no. 1, pp. 426-439, 2023. https://doi.org/10.52131/pjhss.2023.1101.0362

[25] Shah S., Fahlevi M., Rahman E., Akram M., Jamshed K., Aljuaid M., and Abbas J., "Impact of Green Servant Leadership in Pakistani Small and Medium Enterprises: Bridging Pro-Environmental Behaviour through Environmental Passion and Climate for Green Creativity," *Sustainability*, vol. 15, no. 20, pp. 1-16, 2023. https://doi.org/10.3390/su152014747

[26] Sinurat V. and Widhianto C., "The Influence of Job Satisfaction and Employee Retention on Employee Performance Mediated by Perceptions of Leadership Style," *International Journal of Social Service and Research*, vol. 3, no. 10, pp. 2672-2680, 2023. https://doi.org/10.46799/ijssr.v3i10.570

[27] Stepanyan I. and Hameed S., "A Neuro Phenotypic Evolution Algorithm for Recognizing Human Motion Type," *The International Arab Journal of Information Technology*, vol. 21, no. 6, pp. 1015-1028, 2024. DOI:10.34028/iajit/21/6/6

[28] Surucu L., Maslakci A., and Sesen H., "Inclusive Leadership and Innovative Work Behaviors: A Moderated Mediation Model," *Leadership and Organization Development Journal*, vol. 44, no. 1,

pp. 87-102, 2023. https://doi.org/10.1108/LODJ-05-2022-0227

[29] Susanto P., Sawitri N., and Susita D., "Job Satisfaction and Employee Turnover: Analysis Recruitment, Career Development, Organizational Culture," *Dinasti International Journal of Digital Business Management*, vol. 4, no. 3, pp. 619-629, 2023. https://doi.org/10.31933/dijdbm.v4i3.1825

[30] Sutardi D., Nuryanti Y., Kumoro D., Mariyanah S., and Agistiawati E., "Innovative Work Behavior: A Strong Combination of Leadership, Learning, and Climate," *International Journal of Social and Management Studies*, vol. 3, no. 1, pp. 290-301, 2022. https://ijosmas.org/index.php/ijosmas/article/view/114

[31] Usmanova K., Wang D., Sumarliah E., Khan S., Khan S., and Younas A., "Spiritual Leadership as a Pathway toward Innovative Work Behavior via Knowledge Sharing Self-Efficacy: Moderating Role of Innovation Climate," *VINE Journal of Information and Knowledge Management Systems*, vol. 53, no. 6, pp. 1250-1270, 2023. https://doi.org/10.1108/VJIKMS-04-2021-0054

[32] Wen J., Hussain H., Waheed J., Ali W., and Jamil I., "Pathway toward Environmental Sustainability: Mediating Role of Corporate Social Responsibility in Green Human Resource Management Practices in Small and Medium Enterprises," *International Journal of Manpower*, vol. 43, no. 3, pp. 701-718, 2022. https://doi.org/10.1108/IJM-01-2020-0013

[33] Younas A., Wang D., Javed B., and Ul Haque A., "Inclusive Leadership and Voice Behavior: The Role of Psychological Empowerment," *The Journal of Social Psychology*, vol. 163, no. 2, pp. 174-190, 2023. DOI:10.1080/00224545.2022.2026283

[34] Zhou R., Peng H., and Liu S., "Research on Employee Abnormal Behavior Detection Algorithm Based on Improved SSD," *Advances in Computer, Signals and Systems*, vol. 8, no. 3, pp. 129-136, 2024. DOI:10.23977/acss.2024.080318

[35] Zonghua L., Junyun L., Yulang G., Ming Z., and Xu W., "The Effect of Corporate Social Responsibility on Unethical Pro-Organizational Behavior: The Mediation of Moral Identity and Moderation of Supervisor-Employee Value Congruence," *Current Psychology*, vol. 42, no. 17, pp. 14283-14296, 2023. https://link.springer.com/article/10.1007/s12144-022-02722-x

**Lu Zhang** graduated from Nanchang Hangkong University with a Master's Degree in Management. A Lecturer at Nanchang Institute of Technology. Main research areas include Management Science and Engineering, Business Management, Technological Innovation, Big Data Management and Application. Once studied at the Shanghai Institute of Scientific and Technical Information for two months. Published several papers in domestic authoritative journals such as "Journal of Intelligence," presided fund projects like the Jiangxi Provincial Education Science Planning Project.