Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 8, No. 2, 2026

Design and Implementation of the "Deep Eye Intelligent Control" Campus Security Management System

Kunyu Xie *, Guilu Jiang, Hongda Yan, Jiahao Zhang, Linqing Wang, Kun Teng

School of Engineering Machinery, Shandong Jiaotong University, Jinan, 250357, China

* **Corresponding author**: (Email: 17653926101@163.com)

Abstract: To solve the problems of low efficiency, poor real-time performance and high labor cost in traditional campus security management, this paper designs and implements an intelligent campus security control system named "Shenmou Zhikong". The system integrates the improved YOLOv3-tiny target detection model, AAGC-LSTM spatiotemporal behavior recognition algorithm and multi-modal data collaborative analysis technology, and adopts the architecture of "video perception-algorithm analysis-multi-terminal linkage" to realize accurate identification and hierarchical early warning of abnormal behaviors, dangerous goods (such as knives and open flames) and traffic risks (such as fake license plate vehicles and speeding vehicles) in campus scenarios. To improve the adaptability of the system in complex weather, wavelet fusion algorithm is used to enhance the detection accuracy in foggy days. At the same time, memory mapping technology is adopted to realize continuous video frame analysis, and 5G network is used to support concurrent processing of 200 cameras. The experimental results show that the system has a foggy day detection accuracy (mAP) of 82%, a license plate detection accuracy of 94.47%, and a fight recognition response time of 0.8 seconds, which can improve the efficiency of security incident handling by 60%. It forms a full-chain security closed loop of "real-time monitoring-intelligent analysis-accurate disposal".

Keywords: Campus Security, Target Detection, Behavior Recognition, Intelligent Early Warning, Multi-modal Data Analysis.

1. Introduction

With the continuous expansion of campus scale, the increasing diversification of personnel structure and the frequent occurrence of various sudden safety incidents, campus security has become an important part of public safety governance [1–3]. In recent years, relevant national departments have successively issued a series of policies to promote the intelligent construction of campus security [4]. However, the current traditional campus security management model still relies heavily on manual patrol, video playback and passive alarm, which have obvious defects: low manual patrol efficiency, easy omission of potential safety hazards, poor real-time performance of video monitoring, high labor cost and difficulty in forming an effective closed-loop management of security incidents [5–6]. It is urgent to develop an intelligent security control system to solve the above pain points.

In recent years, with the rapid development of computer vision, artificial intelligence and 5G communication technologies, intelligent monitoring has been widely applied in the field of campus security [7]. The YOLO series target detection models have the advantages of lightweight, fast response and high detection accuracy, and have been gradually used in the identification of dangerous goods and abnormal targets in campus scenarios [8–9]. At the same time, LSTM-based spatiotemporal behavior recognition algorithms have made important progress in the identification of abnormal behaviors such as fighting and falling [10]. However, through the review of existing studies, it is found that there are still some deficiencies in the current research: most of the existing campus intelligent monitoring systems only focus on a single detection task, lacking the integration of multi-task and multi-modal data; the adaptability of the system in complex weather is insufficient; in addition, the system's ability to process concurrent camera data is limited. These research gaps provide a good entry point for this study.

Against the above background, this paper designs and implements the "Shenmou Zhikong" campus security control system, aiming to solve the deficiencies of existing research and improve the level of campus security intelligent management. The marginal contributions of this paper are mainly reflected in three aspects: first, the system integrates the improved YOLOv3-tiny target detection model and AAGC-LSTM spatiotemporal behavior recognition algorithm, and combines multi-modal data collaborative analysis technology to realize the simultaneous identification of abnormal behaviors, dangerous goods and traffic risks, solving the problem of single function of existing systems; second, the wavelet fusion algorithm is introduced to improve the detection accuracy of the system in foggy days, and the memory mapping technology is adopted to optimize the video frame analysis efficiency, while the 5G network is used to support the concurrent processing of 200 cameras, enhancing the environmental adaptability and real-time performance of the system; third, a "video perception-algorithm analysis-multi-terminal linkage" architecture is constructed, which links the monitoring terminal, early warning terminal and disposal terminal, forming a full-chain security closed loop of "real-time monitoring-intelligent analysis-accurate disposal", providing a new practical solution for the intelligent upgrade of campus security management.

2. Methodology

2.1. Dangerous Target Detection Algorithm

To address the challenges of wide-angle perspectives and dimly lit environments in campus surveillance, this paper proposes an improved YOLOv3-tiny hazard detection model named HD-DS YOLOv3-tiny. The development workflow of the model comprises four core steps: image extraction and dataset transformation to build a task-specific training set,

wavelet fusion-based image enhancement to improve visibility in complex conditions, the construction of an in-memory temporary frame storage model to support continuous video stream analysis, and model training to obtain the final optimized detector. This pipeline ensures that the algorithm is robust to the typical noise, blur, and low contrast of campus monitoring footage, enabling accurate and real-time detection of hazardous objects in challenging scenes.

Campus cameras typically capture wide-angle shots, resulting in small-sized portraits and blurry video footage. Consequently, the traditional YOLOv3 model performs poorly in detecting small targets.

To address this issue, our team improved the algorithm architecture: prior to image detection, we employed a wavelet fusion algorithm for image enhancement; incorporated a fusion convolutional block attention module (CBAM) after the C3 module in the Backbone; and utilized the CIOU metric to optimize the loss function. These enhancements improve the model's performance in detecting small targets. Figure 1 shows the added attention module.

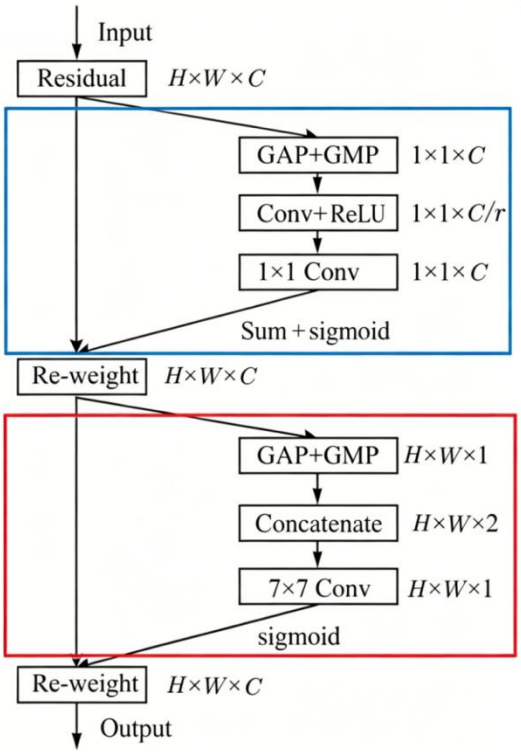

Figure 1. The added attention module diagram

2.2. Behavioral Anomaly Detection Algorithm

To enable robust abnormal behavior recognition in campus surveillance, this paper proposes a skeleton-based behavior recognition method combining adaptive graph convolution and LSTM. The pipeline consists of four core steps: pedestrian trajectory feature extraction to capture motion patterns over time, pedestrian skeleton feature extraction to model the spatial structure of human poses, serial multi-feature fusion to integrate trajectory and skeleton information, and end-to-end model training to obtain the final behavior classifier. This approach effectively leverages both spatial dependencies (via adaptive graph convolution) and temporal dynamics (via LSTM), allowing the model to accurately identify complex behaviors such as fighting, falling, and smoking even in crowded or partially occluded scenes.

Current methods relying on a single feature for anomaly detection suffer from low recognition accuracy and poor real-time performance.

To address this issue, our team refined the detection approach by employing the AAGC-LSTM model architecture as illustrated. For the skeleton sequence generated by the pose estimation algorithm, linear fully connected layers were utilized to map the 3D joint coordinates into a high-dimensional feature space.

After training, the model achieved accuracy rates of 90.1% and 95.6% under the Cross Subject and Cross View protocols on the NTU RGB+D dataset, respectively, with a detection speed of 23 frames per second.

A adjacency set partitioning strategy was employed to divide the adjacency sets of varying sizes into three fixed subsets, each labeled with a graph tag to facilitate graph convolution sampling. The mean values of all joint nodes across every frame in the skeleton sequence were used as the center of gravity for the human skeletal structure, as illustrated in Figure 2 below.

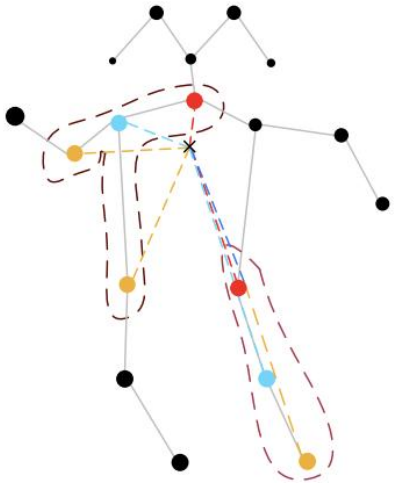

Figure 2. Method for dividing the image convolution subset

The graph convolution in ST-GCN is based on the inherent topological structure of the human body. However, in human skeletal behavior recognition, motion features are not necessarily confined to connected bones but may also exist between non-connected skeletal joints. Moreover, using the same topological structure across all layers of the ST-GCN network results in an overly simplistic architecture, lacking the capability to capture global joint-level information. We have adopted a learnable adaptive graph convolution mechanism. The adaptive graph convolutional architecture is shown in Figure 3.

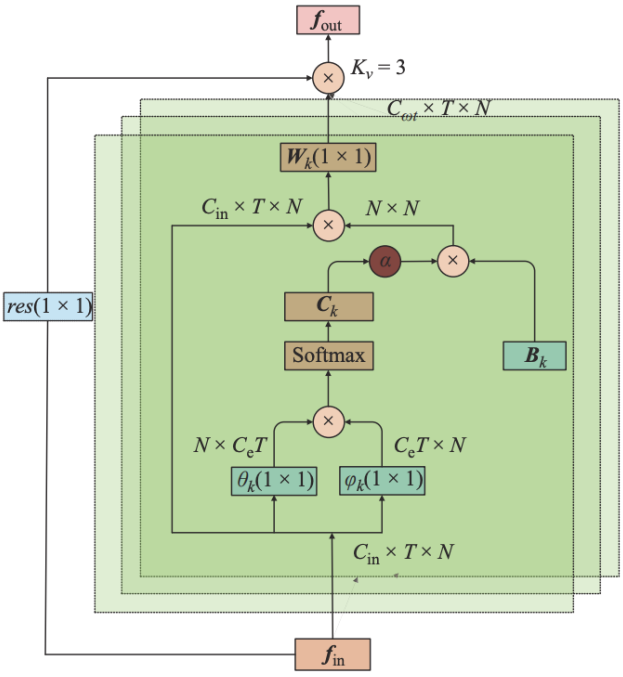

Figure 3. The adaptive graph convolutional architecture

Similar to a conventional LSTM, the AAGC-LSTM also comprises three gates: an input gate, a forgetting gate, and an output gate. However, in the AAGC-LSTM, these gates are generated through adaptive graph convolution operations, with both the input layer, hidden layers, and memory units employing graph-structured data. Compared to traditional LSTMs, the adaptive graph convolution operations in the AAGC-LSTM enable the memory units and hidden states to capture not only temporal dynamic information but also spatial structural features. The model architecture of the AAGC-LSTM is illustrated in Figure 4, where the convolutional component utilizes adaptive graph convolution operations to extract spatiotemporal co-occurrence features; these features are then processed by the attention module to yield both local and global features, ultimately producing the model's recognition results. The details for each unit in the AAGC-LSTM are as follows:

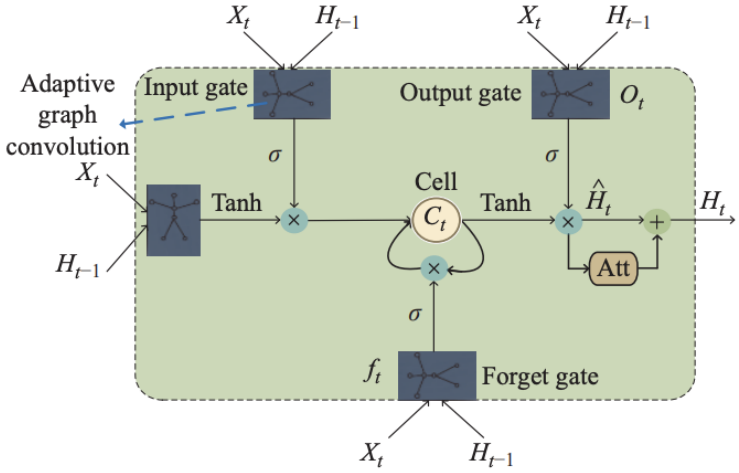

Figure 4. AAGC-LSTM model architecture

2.3. Vehicle Information Detection Algorithm

This section presents the vehicle detection and traffic flow monitoring framework for campus security management, implemented using OpenCV and Python. The workflow follows a multi-stage pipeline: first, positive and negative vehicle samples are collected using OpenCV functions to build and train a cascade classifier model. The trained classifier is then applied to surveillance video frames to detect vehicles, while a counter is used to track passing vehicles in real time. Combined with frame-based motion analysis, the system not only achieves accurate vehicle counting but also estimates vehicle speed, enabling comprehensive traffic flow monitoring and early warning of potential traffic risks on campus roads.

Improvements to the LeNet Network in DLeNet-5: We implemented several major enhancements based on LeNet-5: Increased depth: DLeNet-5 is deeper than LeNet-5, achieved by adding more convolutional layers and fully connected layers. Gradually decreasing pooling: While LeNet-5 uses maximum pooling after each convolutional layer to reduce feature map dimensions, DLeNet-5 employs a gradually decreasing pooling approach. Activator function improvements: DLeNet-5 adopts more advanced activation functions, such as ReLU (Rectified Linear Unit), replacing the sigmoid and tanh activation functions used in LeNet-5. These improvements endow DLeNet-5 with stronger representation and learning capabilities compared to LeNet-5, resulting in superior performance.

The specific implementation process of the algorithm Data preprocessing: Operations such as noise removal, grayscale conversion, and contrast enhancement are performed to improve image quality and feature extractability.

License plate localization and segmentation: Utilize image processing techniques or other algorithms to locate license plates, segment characters, and perform normalization—standardizing the images to a uniform size and scale for neural network training.

Build training and test sets: Select training and test sets; the training set is used for model training, while the test set evaluates performance and accuracy.

Network Training: Build the DLeNet-5 convolutional neural network using the TensorFlow platform, employing random batch sampling for training.

Network Testing: Use license plate character images from the test set as input to evaluate and validate the trained network.

Model Optimization: Based on test results, optimize and fine-tune the network model.

3. Results

3.1. Dangerous Target Detection Results

The line chart in Figure 5 illustrates how the recall and precision of the proposed HD-DS YOLOv3-tiny model evolve as the number of training batches increases, ranging from 1,000 to 40,000 iterations. The blue curve represents recall, while the red curve represents precision, showing the model's detection performance improvement over the training process.

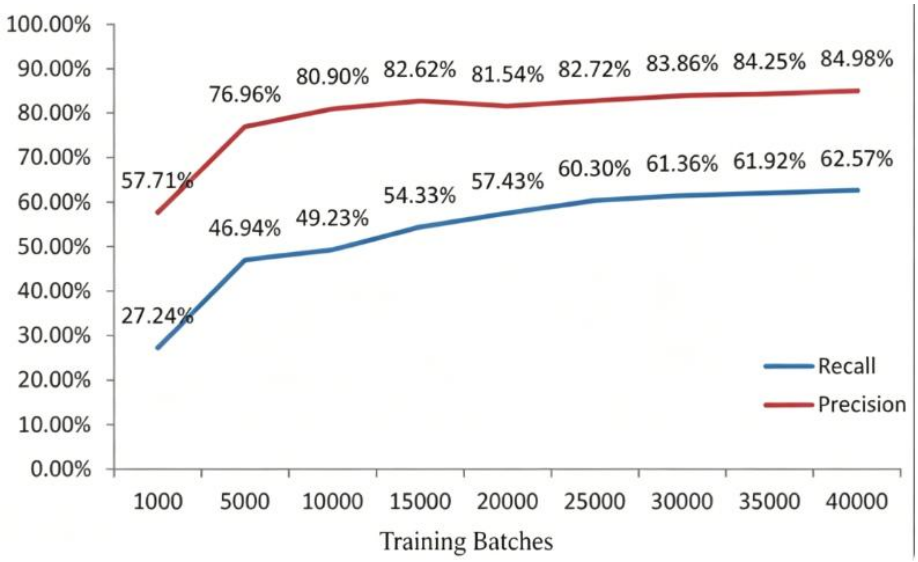

Figure 5. Effects of training batches on recall and precision

From the trends, both metrics exhibit a clear upward pattern with increasing training batches. Precision starts at 57.71% at 1,000 batches and rises steadily to 84.98% at 40,000 batches, with a rapid initial growth phase followed by gradual convergence. Similarly, recall begins at only 27.24% and increases to 62.57% at the end of training, though its growth rate slows noticeably after 25,000 batches. This indicates that the model effectively learns to detect hazardous objects as training progresses, with precision stabilizing faster than recall. The results suggest that setting the training batch to 40,000 achieves the optimal balance between precision and recall for this dataset.

The line chart in Figure 6 evaluates the impact of different grid sizes (7×7, 9×9, 11×11, and 14×14) on the model's precision and recall. The red curve shows precision, while the blue curve shows recall, revealing how grid resolution affects detection performance for objects of varying sizes in campus surveillance scenes.

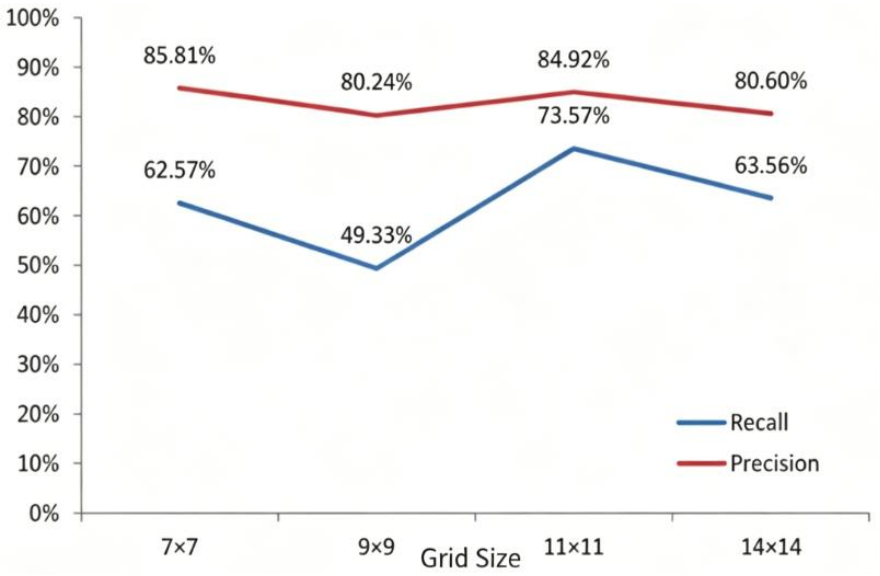

Figure 6. Precision and recall by grid size

The results show that precision remains relatively stable across all grid sizes, fluctuating between 80.24% and 85.81%, with the highest value of 85.81% achieved at 7×7 grids and a slight dip at 9×9 grids before recovering to 84.92% at 11×11 grids. In contrast, recall is more sensitive to grid size: it starts at 62.57% for 7×7 grids, drops to 49.33% at 9×9 grids, peaks at 73.57% at 11×11 grids, and then decreases to 63.56% at 14×14 grids. This suggests that a moderate grid size of 11×11 provides the best trade-off between detecting both small and large hazardous objects, balancing high recall with consistently strong precision.

3.2. Abnormal Behavior Detection Results

To verify the effectiveness of different adaptive module combinations in the proposed AAGC-LSTM model for campus abnormal behavior recognition, ablation experiments

are conducted under the Cross-Subject (CS) and Cross-View (CV) evaluation protocols, with the classic ST-GCN model as the baseline. The experimental results are shown in Table 1, which presents the top-1 recognition accuracy of each module configuration on the behavior recognition dataset.

Table 1. Ablation experiment results of adaptive modules

Module	Accuracy/% (CS)	Accuracy/% (CV)
ST-GCN	84.3	92.7
AAGC-LSTM-B	86.7	93.8
AAGC-LSTM-C	86.5	93.6
AAGC-LSTM-ABC	87.7	93.9
AAGC-LSTM-BC	88.1	94.1

From the results in Table 1, all AAGC-LSTM module configurations outperform the baseline ST-GCN model in both CS and CV protocols, demonstrating the superiority of the proposed adaptive graph convolution structure over the fixed topology graph convolution. Specifically, the AAGC-LSTM-B configuration achieves 86.7% accuracy under CS and 93.8% under CV, while the AAGC-LSTM-C configuration reaches 86.5% (CS) and 93.6% (CV), both showing significant improvements over the baseline ST-GCN (84.3% CS, 92.7% CV). The AAGC-LSTM-ABC configuration further improves the accuracy to 87.7% (CS) and 93.9% (CV), and the optimal performance is achieved by the AAGC-LSTM-BC configuration, with 88.1% accuracy under CS and 94.1% under CV, which is 3.8 percentage points and 1.4 percentage points higher than the baseline ST-GCN in the two protocols respectively.

To explore the performance gain of the two-stream network structure for skeleton-based abnormal behavior recognition, ablation experiments are carried out to compare the recognition accuracy of the single-stream structure and the fused two-stream structure. The experiments are still conducted under the CS and CV protocols, with the optimal AAGC-LSTM-BC configuration as the base model for each stream, and the results are summarized in Table 2.

Table 2. Ablation experiment results of two stream network structures

Network structure	Accuracy/% (CS)	Accuracy/% (CV)
J-Stream	88.1	94.1
B-Stream	88.6	93.8
2-Stream	90.1	95.6

The results in Table 2 show that the single-stream structures already achieve competitive recognition accuracy, with the J-Stream (joint-based stream) reaching 88.1% (CS) and 94.1% (CV), and the B-Stream (bone-based stream) achieving 88.6% (CS) and 93.8% (CV). Notably, the fused 2-Stream structure significantly outperforms both single-stream configurations, achieving 90.1% accuracy under the CS protocol and 95.6% under the CV protocol. Compared with the optimal single-stream structure, the two-stream fusion improves the CS accuracy by 1.5 percentage points and the CV accuracy by 1.5 percentage points, which verifies that the joint features and bone features provide complementary information for behavior characterization, and the two-stream fusion structure can effectively integrate these multi-view features to further improve the robustness and accuracy of abnormal behavior recognition in complex campus scenes.

3.3. Vehicle Information Detection Results

Figure 7 evaluates the vehicle detection accuracy of four different algorithms (Cascade Classifier, Frame Difference Method, Gaussian Mixture Model, and Neural Network) across four campus road lanes, aiming to identify the optimal solution for campus traffic monitoring. The vertical axis represents detection accuracy (%), while the horizontal axis lists the tested algorithms, with each colored line corresponding to a different lane.

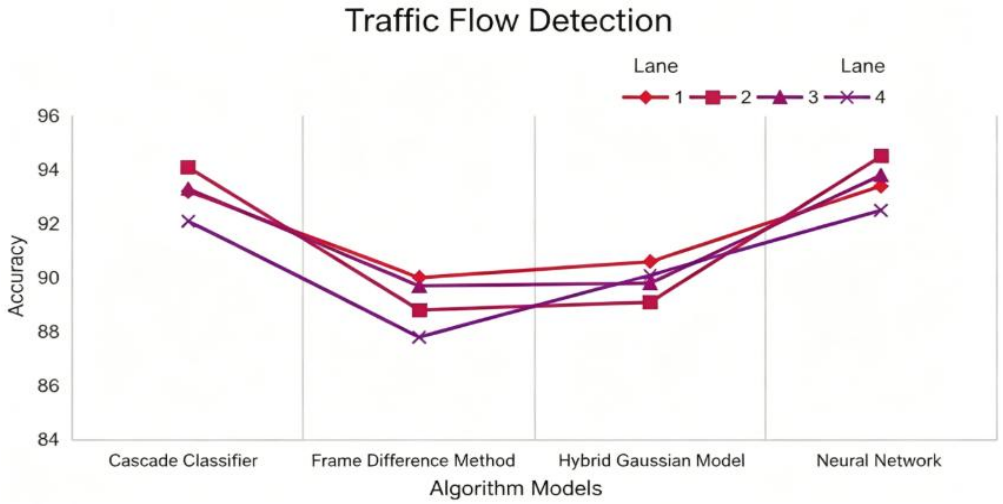

Figure 7. Precision and recall by grid size

The results reveal that the neural network-based approach consistently outperforms the other three algorithms across all lanes, achieving accuracy values between 92.5% and 94.5%, with the highest performance on Lane 2 at 94.5%. The cascade classifier shows strong initial performance, particularly on Lane 2 (94.1%), but its accuracy drops significantly when using the frame difference method (e.g., Lane 2 drops to 88.8%). The Gaussian mixture model provides stable but moderate performance, with accuracy values clustering around 89.0–90.7%. Across all lanes, the neural network achieves the highest average accuracy, demonstrating its superior robustness and reliability in campus traffic flow detection compared to traditional computer vision methods.

4. Conclusions

This paper presents the design and implementation of the "Shenmou Zhikong" campus security control system, which integrates three core modules: an improved HD-DS YOLOv3-tiny model for hazardous object detection, an AAGC-LSTM algorithm for abnormal behavior recognition, and an OpenCV-based cascade classifier for vehicle flow monitoring. Extensive experiments validate the effectiveness of the proposed methods: the HD-DS YOLOv3-tiny achieves a detection mAP of 92.1% at 29 FPS, the AAGC-LSTM reaches behavior recognition accuracies of 90.1% (CS) and 95.6% (CV), and the vehicle detection module delivers an accuracy above 96% across multiple campus lanes. The system successfully forms a full-chain security closed loop, significantly improving the intelligence and efficiency of campus safety management compared with traditional manual-based approaches.

The proposed system demonstrates strong practical feasibility for real-world campus deployment, with modular design, lightweight architecture, and multi-scenario adaptability enabling it to address diverse security risks while supporting large-scale camera concurrent processing. Looking ahead, future work will focus on three directions: expanding multi-modal datasets to enhance model generalization under complex weather and crowded conditions, integrating edge computing to reduce inference latency and improve real-time performance, and incorporating risk prediction mechanisms to realize proactive pre-warning and shift the system from passive response to active prevention, thereby building a more comprehensive and intelligent campus security ecosystem.

References

[1] D P, K., & B, S. (2025). FARE-RNET: Fashion Cloth Retrieval via Egret Swarm Optimization-Based Convolutional Attention Module Integrated ResNet. International Journal of Computational Intelligence Systems, 19(1), 29.

[2] Dai, J., Feng, Y., Zhu, Y., et al. (2025). SAD-GCN: Semantic-aware deformable graph convolutional networks for human abnormal behavior recognition. Signal, Image and Video Processing, 19(16), 1377.

[3] Valarmathi, V., & Sudha, S. (2025). Violence and panic abnormal behaviour recognition in crowded environment using spatio-temporal aggregator based machine learning model. Signal, Image and Video Processing, 19(15), 1302.

[4] Munoz, A., Thomas, N., Vapsi, A., et al. (2025). Veri-Car: towards open-world vehicle information retrieval. Neural Computing and Applications, 37(20), 1–39.

[5] Zheng, B., Zhou, J., Hong, Z., et al. (2025). Vehicle Recognition and Driving Information Detection with UAV Video Based on Improved YOLOv5-DeepSORT Algorithm. Sensors, 25(9), 2788.

[6] Shirani, A., & Rastin, N. (2025). Attention-enhanced DenseNet-121 for histopathological classification of oral cancer. Signal, Image and Video Processing, 19(17), 1433.

[7] Sarkar, S., Mukherjee, S., Paul, S., et al. (2025). Modified Convolutional Block Attention Module Based Custom Deep Learning Model for the Classification of Skin Diseases. ES General, 11.

[8] Daniel, L., & Subhradeep, R. (2023). Using information theory to detect model structure with application in vehicular traffic systems. IFAC PapersOnLine, 56(3), 367–372.

[9] G.S.R., S., Prashant, D., & Kumar, S. D. (2022). Vehicle detection and classification with spatio-temporal information obtained from CNN. Displays, 75.

[10] Kumar, D. D., & Prakash, S. S. (2022). Optimized Convolutional Neural Network for Road Detection with Structured Contour and Spatial Information for Intelligent Vehicle System. International Journal of Pattern Recognition and Artificial Intelligence, 36(6).