Real-time Detection Of Abnormal Pedestrian Behavior In Scenic Areas Using Hybrid Gaussian Models And Regional Feature Analysis

Ping Xu[1] and Aziguli Abudouwaili[2*]

[1] Shanxi Vocational University of Culture and Tourism
[2] Moray (Beijing) Data Technology Co., Ltd.
[*] Corresponding author. E-mail: aziguli_abd92@outlook.com; aziguli@haimanyun.com

Received: Mar. 14, 2026; Accepted: Apr. 29, 2026

The rapid development of the tourism industry has made the need for managing tourists' behaviors in scenic spots more urgent. The irregularities of the tourists, such as unauthorized climbing, writing graffiti, and entering the forbidden regions, can cause security hazards and negatively impact the reputation of the location. In this paper, a novel framework is introduced for recognizing abnormal behaviors of pedestrians in the region of interest using the characteristics extracted from surveillance videos. To achieve a robust background subtraction, a mixture of Gaussians is applied, and then the bounding boxes of motion objects are computed using minimum bounding rectangles. To recognize periodic changes in aspect ratios, curve fitting is employed to find abnormal behaviors.

Keywords: Abnormal behavior detection; video surveillance; hybrid Gaussian model; regional features; real-time monitoring.

http://dx.doi.org/10.6180/jase.202610_33.044

1. Introduction

Tourism has increased rapidly in recent years, making scenic and cultural sites popular among domestic and international visitors. However, this growth has led to challenges in managing visitor behavior at such destinations. Instances of unsafe and abnormal activities, such as climbing monuments, graffiti, and trespassing, have become more common [1–3]. These behaviors not only create safety risks but also negatively affect the long-term sustainability and reputation of these sites [4, 5].

In response to these concerns, efficient real-time monitoring systems are increasingly required for effective tourism management. Traditional approaches such as manual surveillance are labor-intensive and prone to human error [6, 7]. They depend on continuous human monitoring of video feeds, which limits scalability and delays response time. Moreover, detecting abnormal behavior manually in crowded and dynamic environments is inefficient, where quick intervention is essential for ensuring visitor safety [8, 9].

To address these challenges, there has been an increasing focus on automating the detection of abnormal pedestrian behaviors through advanced video surveillance and computer vision techniques. A graph- and ensemble learning-based method improves non-English text classification and emotion analysis by leveraging relational features, achieving higher accuracy and robustness than traditional models [10]. The goal is to enable real-time automatic detection and alerting of unsafe activities, improving accuracy, reducing false positives, and minimizing human intervention for efficient and scalable monitoring [11].

Despite advances, abnormal pedestrian detection remains challenging due to GMM sensitivity to lighting and crowd variations, causing false alarms and reduced accuracy [12].

Recent studies use CNN-based models to improve behaviour detection accuracy, but they require high computational resources, are noise-sensitive, and face challenges in real-time large-scale surveillance [13].

Region-based methods using bounding box aspect ratios are computationally efficient for detecting simple anomalies but are sensitive to noise, lighting changes, and occlusions in crowded scenes. This study presents a real-time method combining hybrid Gaussian modeling and region-based features, using aspect ratio and curve fitting to accurately detect abnormal pedestrian behavior under dynamic conditions [14].

An improved YOLOv5-based method is proposed for detecting anomalous behaviors in classroom environments, enhancing feature extraction and object localization for better accuracy [15]. Gollavilli and Thanjaivadivel [16] also note that detecting abnormal pedestrian behavior remains challenging, as GMM-based methods are affected by lighting changes, crowd dynamics, and environmental variations, leading to false alarms and reduced accuracy.

2. Methods and materials

The background in stationary landscape surveillance is modeled using Gaussian distributions, with hybrid Gaussian models offering the most comprehensive representation over time. The proposed hybrid Gaussian model is equivalent to a Gaussian Mixture Model (GMM) with online adaptive parameter updating.

2.1. Gaussian hybrid model

The pixel value probability of the current observation is determined as follows: Each pixel sequence over time $\{X_1, X_2, \ldots, X_n\}$ is modelled using a hybrid Gaussian model.

$$P\left(X_t\right) = \sum_{k=1}^{K} \omega_{k,t} \eta\left(X_t; \mu_{k,t}, \Sigma_{k,t}\right) \tag{1}$$

where $\sum$ denotes the summation over all Gaussian components ($k = 1$ to K), K is the total number of Gaussian components, $\omega_{k,t}$ represents the weight of the kth Gaussian component at time t, X_t is the observed pixel value at time t, $\mu_{k,t}$ is the mean vector, $\Sigma_{k,t}$ is the covariance matrix of the kth component, and where $\eta(\cdot)$ denotes the Gaussian probability density function defined in Equation (2). The proposed hybrid Gaussian model uses adaptive online updates of weights, means, and variances, providing more stable and robust background modelling than conventional GMM, ViBe, and SuBSENSE in dynamic environments

$$\begin{aligned}\eta(X_t; \mu, \Sigma) = {} & \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \\ & \times \exp\left(-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)\right)\end{aligned} \tag{2}$$

where μ is the mean vector, Σ denotes the covariance matrix, $|\Sigma|$ represents the determinant of the covariance matrix, and Σ^{-1} is the inverse of the covariance matrix, $(X_t - \mu)^T$ denotes the transpose operation, and d is the dimensionality of the feature space.

The distribution of each hybrid model must be updated because the background of longrunning video programs invariably changes [14]. A pixel that does not match any model is classified as foreground with a new Gaussian component added, while a matching pixel updates the model's mean, variance, and weight. The following changes were made to the mixing models by Stauffer and Grimson:

$$\begin{aligned}\omega_{k,t} &= (1-\alpha)\omega_{k,t-1} + \alpha M_{k,t} \\ \mu_{k,t} &= \left(1-\rho_{k,t}\right)\mu_{k,t-1} + \rho_{k,t} X_t \\ \sigma_{k,t}^2 &= \left(1-\rho_{k,t}\right)\sigma_{k,t-1}^2 + \rho_{k,t}\left(X_t - \mu_{k,t}\right)^2\end{aligned} \tag{3}$$

Where $\rho_{k,t} = \alpha \cdot \eta\left(X_t \mid \mu_{k,t-1}, \Sigma_{k,t-1}\right)$ is the learning rate that controls the adaptation speed of the Gaussian model parameters, and $M_{k,t}$ is the matching indicator, which takes the value 1 if the current pixel matches the kth Gaussian component and 0 otherwise. A higher value of $\rho_{k,t}$ allows the model to quickly adapt to recent changes in the background, while a lower value ensures more stable updates by giving higher importance to past observations.

The learning rate α is empirically tuned to balance adaptability and stability, enabling optimal performance under varying environmental conditions.

2.2. The outside rectangle's lowest aspect ratio

We use hybrid Gaussian background modelling in Hue-Saturation-Value (HSV) space because the aspect ratio of the moving target's minimal outer rectangle is extremely sensitive to the shadows cast by the target. The RGB image is converted into HSV color space using the standard RGB-to-HSV color space conversion method described in the literature [17]. HSV reduces shadow and illumination sensitivity by separating intensity (V) from chromatic components (H, S), improving segmentation stability and reducing false detections.

$$S = \frac{V - \min(R, G, B)}{V} \tag{4}$$

This shows that saturation depends on relative intensity differences, which makes HSV less sensitive to illumination changes (shadow effects mainly affect overall brightness, not this ratio).

Compared to other colour spaces such as YCbCr and CIELAB , HSV provides a more intuitive separation of chromatic and intensity components for surveillance applications. While YCbCr separates luminance and chrominance, and CIELAB offers perceptual uniformity, HSV is more robust in handling illumination variations in outdoor scenes due to its explicit decoupling of the HSV component from hue and saturation, thereby maintaining better consistency under diverse lighting conditions as shown in Table 1.

Following the hybrid Gaussian model's segmentation of the motion target, concatenated component analysis and morphological filtering can provide a binary image that only contains the motion target and its minimum outer rectangle box. To characterize the geometric properties of the detected pedestrian region, the aspect ratio of the bounding box is computed.

$$R = \frac{h}{w} \tag{5}$$

h = height of the bounding box and w = width of the bounding box

In cases of partial occlusion, the minimum bounding rectangle may not fully capture the complete pedestrian silhouette, leading to slight deviations in aspect ratio estimation. However, the proposed method mitigates this limitation by relying on temporal consistency across consecutive frames, which helps stabilize geometric feature variations under occluded conditions.

As shown in Fig. 1, this picture is then projected onto the original Fig. 1(a) and (b). In the adaptive hybrid Gaussian modelling approach is used because, in HSV mode, the chromaticity signal varies less when shadows are covered, while the pixel brightness changes dramatically. This property makes the HSV color space robust to illumination changes in outdoor environments, as hue and saturation are less sensitive to lighting variations. This improves detection accuracy and reliability in surveillance applications.

The aspect ratios for pedestrians standing, jogging, and sitting were obtained for this study; the outcomes are displayed in Fig. 2.

This research employs this property to identify pedestrian behavior based on tests. In particular, there is an analytical minimum outside the rectangular box variation of typical pedestrians in video sequences. For instance, the rectangular box's aspect ratio is periodic when walking regularly; when sprinting, it is comparable to walking, but the frequency and magnitude of changes are different, as shown in Fig. 2. Posture-based aspect ratio variations and their temporal changes effectively indicate abnormal behaviors such as loitering, sudden running, or unauthorized actions in surveillance environments.

The method extracts pedestrian bounding boxes using hybrid Gaussian modelling and uses temporal variations in aspect ratio across frames for behaviour classification.

Temporal and trajectory-based motion patterns across frames enable reliable detection of abnormal behaviours such as climbing, trespassing, wandering, and repeated entry.

2.3. Fitting curves

For periodic functions, the fitted function frequently takes the following shape:

$$f(x) = a + \sum_{k=1}^{K} \left[A_k \sin(\omega_k x) + C_k \cos(\omega_k x)\right] \tag{6}$$

Given the complexity and real-time nature of the calculation, this study uses the fitting form below.

$$F(x) = a + b_1 \sin(\omega_1 x + \phi_1) + b_2 \cos(\omega_2 x + \phi_2) \tag{7}$$

Given the actual situation, we limited the fitting to the constraint using the least squares criterion. Trigonometric curve fitting with regularized least squares models periodic pedestrian motion and improves robustness against noise and outliers.

The periodic characteristics of pedestrian motion are analyzed using Fourier-based frequency domain transformation.

$$X(f) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi f n/N} \tag{8}$$

The dominant motion frequency is identified by selecting the peak value in the frequency spectrum.

$$f_{\text{peak}} = \arg\max |X(f)| \tag{9}$$

Walking and running are distinguished based on dominant frequency f_{peak} , where running exhibits higher frequency values compared to walking.

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \|y_i - F(x_i; \theta)\|^2 \tag{10}$$

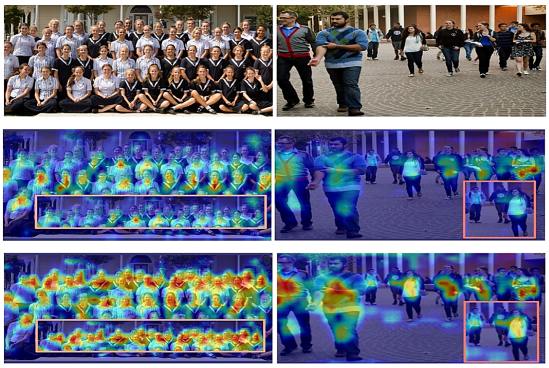
where θ represents the model parameters and N is the number of samples.

Table 1. Performance comparison of HSV, YCbCr, and LAB color spaces in handling illumination variations

Color Space	Illumination Robustness	Computational Cost	Suitability
HSV	High	Low	Excellent
YCbCr	Medium	Low	Good
LAB	High	High	Moderate

(a) (b)

Fig. 1. Motion target segmentation using the proposed hybrid Gaussian model: (a) original frame, (b) detected foreground with bounding box.



(a) (b) (c)

Fig. 2. Feature extraction results for pedestrian scenes: (a) original images, (b) intermediate feature responses, and (c) enhanced feature representations.

3. Results and discussion

To evaluate the system's recognition ability, the Region(s) of Interest (ROIs) of segmented continuous infrared images were input into a LeNet-7 convolutional neural network, which is designed for fast CPU operation. An Intel i7 9750H processor (4.5GHz, 32 GB RAM) powers the machine. The Terrific Motion IR database and the Ohio State University Object Tracking and Classification Beyond Visible Spectrum (OTCBVS) benchmark dataset were used for the tests. To guarantee that the test set had the same size as the training set, all ROIs were standardized to 32×32.

Three infrared datasets were used for pedestrian detection, including the OSU Thermal Pedestrian Database, Terrific Motion IR database, and OSU Color Thermal Database, with varying numbers of targets and occlusion conditions. The system was able to successfully detect pedestrians even under partial blockage, and the detected regions were highlighted using green bounding boxes, demonstrating its effectiveness.

The performance was evaluated using TP, FP, and FN metrics and compared with SVMHOG and Fast R-CNN models, including VGG-CNN-M1024, ZF, and VGG16. The improved LeNet-7 network showed competitive performance against these existing methods in infrared pedestrian detection. The evaluation metrics used in this study are defined as follows.

$$\text{Quantity} = TP + FN \tag{11}$$

$$\text{Detection Rate} = \frac{TP}{TP + FN} \times 100 \tag{12}$$

Table 2 displays the comparison results. All compared methods, including HOG+SVM, Fast R-CNN, and the proposed approach, were evaluated under identical experimental conditions.

The same datasets, preprocessing steps, and input image resolutions (32×32 ROIs) were used for all methods to ensure a fair comparison. In addition, all experiments were conducted on the same hardware and testing environment to maintain consistency in performance evaluation.

Compared to conventional methods, the proposed approach achieves a higher detection rate with fewer false alarms, as shown in Table 2. It records zero false alarms in Test Set 2 despite high occlusion. In contrast, Fast R-CNN performs less effectively in complex scenes, showing lower accuracy and occasional false alarms in Test Set 1.A more detailed analysis of false alarm rates reveals that most errors occur under challenging conditions such as occlusion, lighting variations, and pedestrian overlap. Occlusion can lead to partial visibility of targets, resulting in misclassification. Lighting variations may affect pixel intensity and introduce noise into the background model. In addition, overlapping pedestrians can produce merged bounding regions, reducing detection accuracy. Despite these challenges, the proposed method maintains a low false alarm rate across all test scenarios, demonstrating its robustness.

Classical HOG+SVM has the longest detection time because it uses fixed-size, singleshape feature vectors as linear classification inputs and has too many traversal windows in complex scenarios. Convolutional neural networks, as opposed to conventional feature matching, greatly shorten the detection time. In the single-line human test set 3 , the detection time of the method used in this study is roughly one-third of that of Fast R-CNN, which significantly improves the detection efficiency (see Fig. 3). The detection time of the methodology used in this paper is roughly half that of Fast R-CNN.

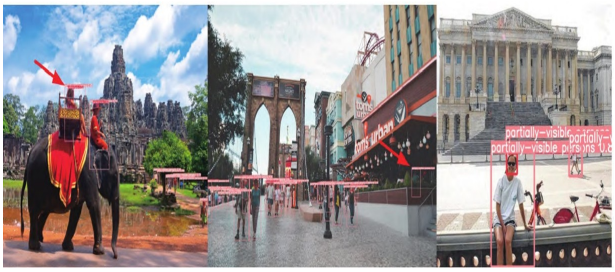


Fig. 3. Results of a partial experiment

Fig. 4(a) and (d) display the segmentation results of the identification of human behavior in the picturesque settings.

As illustrated in Fig. 4, the grayscale vertical projection curve is obtained by vertically projecting the segmented images in Fig. 4(b) and (e). The grayscale area in Fig. 4(b) shows a clear peak without the sticking phenomenon, and luminance bands are identified using the rising and falling curves around the peaks. The projection curve in Fig. 4(e) requires additional processing to handle the sticking issue. When these luminance bands are applied to the original image in Fig. 4(c), the image is segmented into multiple high-brightness rectangular regions.

Although deep learning models such as Fast R-CNN achieve high detection accuracy, they often suffer from limitations, including high computational cost, increased energy consumption, and limited interpretability. These factors make them less suitable for edge-based and timesensitive deployments. In contrast, the proposed lightweight approach achieves a balanced trade-off between accuracy and efficiency, enhancing its suitability for resource-constrained surveillance applications.

4. Conclusion

This study proposes a novel approach for detecting abnormal pedestrian behavior in scenic areas using hybrid Gaussian models and region-based feature analysis. With the growth of tourism, there is a need for efficient real-time monitoring systems, as traditional manual methods are insufficient. The proposed method combines adaptive background modeling and aspect ratio analysis to improve detection accuracy while reducing false alarms.

Results from experiments show that our approach performs better than more conventional approaches like HOG+SVM and Fast R-CNN, especially in difficult, crowded environments. The system's real-time processing capability ensures practical applicability for large-scale deployment. Results from experiments show that the proposed approach performs better than conventional methods such as HOG+SVM and Fast R-CNN, especially in crowded and complex environments. The system's fast processing capability ensures practical applicability for large-scale deployment. The proposed method achieves a detection rate of up to 99.9% with a false alarm rate of 0%, outperforming HOG+SVM and Fast R-CNN across all test sets. Additionally, it reduces detection time to as low as 10.26 ms , demonstrating superior efficiency for time-sensitive applications.

Future work will explore the integration of multimodal inputs, such as thermal and RGB data, to improve detection performance under low-light and challenging environmental conditions. This can enhance the robustness and reliability of the proposed framework in complex surveillance scenarios. Future work will focus on integrating alert mechanisms to improve detection response in surveillance applications and linking anomaly detection outputs with automated response systems to enable timely interventions.

Table 2. Results of the Experiment

Test Set	Method	Quantity	TP	FP	FN	Detection Rate (%)	False Alarm Rate (%)	Time (ms)
	HOG+SVM	99	97	3	2	97.98	1.8	402.52
1	Fast R-CNN	100	99	2	1	99.00	1.0	31.06
	Proposed	101	100	2	1	99.01	1.0	14.89
	HOG+SVM	450	450	8	0	100.00	0.2	319.05
2	Fast R-CNN	452	452	6	0	100.00	0.0	28.56
	Proposed	455	455	4	0	100.00	0.0	14.50
	HOG+SVM	200	200	0	0	100.00	0.0	204.60
3	Fast R-CNN	200	200	0	0	100.00	0.0	28.23
	Proposed	200	200	0	0	100.00	0.0	10.26

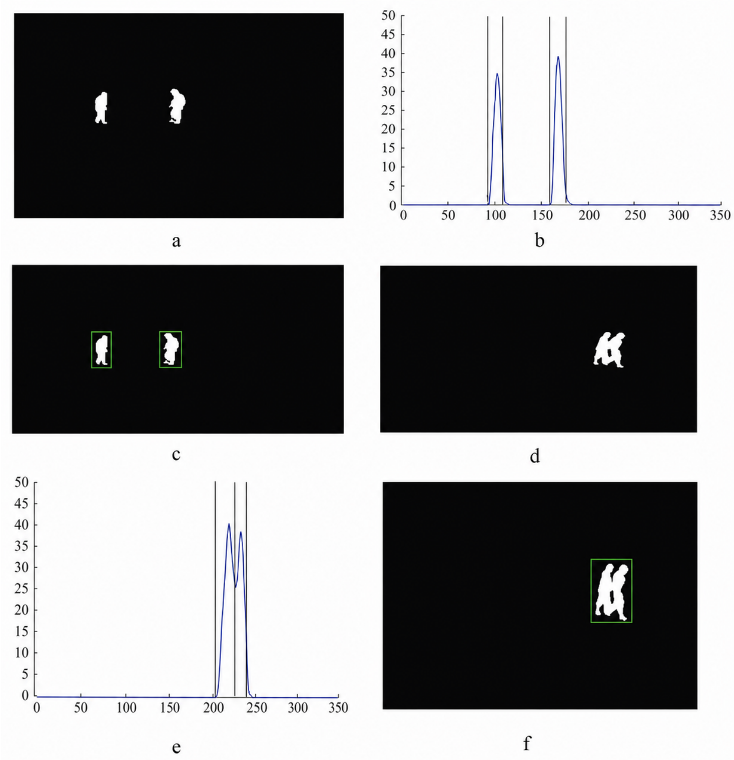


Fig. 4. Variations in the traits of pedestrians

Declarations

Funding: Project type: Shanxi Province art science planning project, Project name: Empirical research on the integration of culture and tourism to enhance the value of smart tourism. Project number: 23AA008

Conflict of interest: The authors declared that they have no conflicts of interest regarding this work.

Data Availability: All data generated or analyzed during this study are included in the manuscript.

Code Availability: Not applicable.

Author's contributions: Ping Xu, Aziguli Abudouwaili is contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

References

[1] J. M. Korn, F. G. Kuruvilla, S. A. McCarroll, A. Wysoker, J. Nemesh, S. Cawley, and D. Altshuler, (2008) *"Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs"* **Nature Genetics** ***40***(10): 1253–1260. DOI: 10.1038/ng.237.

[2] L. M. M. Visuwasam and D. P. Raj, (2023) *"Spatio Temporal Tourism Tracking System Based on Adaptive Convolutional Neural Network"* **Computer Systems Science & Engineering** ***45***(3): 2435–2446. DOI: 10.32604/csse.2023.024742.

[3] E. Zhao, P. Du, E. Y. Azaglo, S. Wang, and S. Sun, (2023) *"Forecasting daily tourism volume: A hybrid approach with CEMMDAN and multi-kernel adaptive ensemble"* **Current Issues in Tourism** ***26***(7): 1112–1131. DOI: 10.1080/13683500.2022.2048806.

[4] W. Chen, W. Wang, H. Yin, L. Zhao, and X. Zhou, (2023) *"HFUL: A hybrid framework for user account linkage across location-aware social networks"* **The VLDB Journal** ***32***(1): 1–22. DOI: 10.1007/s00778-022-00730-8.

[5] G. Bharathi and G. Anandharaj, (2023) *"Classification of review information based on tourist's places of interest in India using support vector machine"* **International Journal of Business Information Systems** ***44***(2): 202–218. DOI: 10.1504/IJBIS.2023.134951.

[6] S. Wang, Y. Yu, J. Chen, and J. Liu, (2023) *"Impact of climate change on cherry blossom viewing tourism: Analysis and simulation based on Weibo proxy data"* **Current Issues in Tourism** ***26***(5): 718–734. DOI: 10.1080/13683500.2022.2049711.

[7] I. Afyouni, Z. Al Aghbari, and R. A. Razack, (2022) *"Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey"* **Information Fusion** ***79***: 279–308. DOI: 10.1016/j.inffus.2021.10.013.

[8] W. Yu, (2021) *"Research on feature extraction of vehicle abnormal driving behaviour based on 5G internet of vehicles"* **International Journal of Vehicle Design** ***86***(1–4): 124–142. DOI: 10.1504/IJVD.2021.122256.

[9] J. Ali, R. H. Jhaveri, M. Alswailim, and B.-H. Roh, (2023) *"ESCALB: An effective slave controller allocation-based load balancing scheme for multi-domain SDN-enabled-IoT networks"* **Journal of King Saud University - Computer and Information Sciences** ***35***(6): 101566. DOI: 10.1016/j.jksuci.2023.101566.

[10] F. Gholami, Z. Rahmati, A. Mofidi, and M. Abbaszadeh, (2023) *"On enhancement of text classification and analysis of text emotions using graph machine learning and ensemble learning methods on non-English datasets"* **Algorithms** ***16***(10): 470. DOI: 10.3390/a16100470.

[11] M. Qaraqe, A. Elzein, E. Basaran, Y. Yang, E. B. Varghese, W. Costandi, and N. Alam, (2024) *"PublicVision: A secure smart surveillance system for crowd behavior recognition"* **IEEE Access** ***12***: 26474–26491. DOI: 10.1109/ACCESS.2024.3366693.

[12] I. Nosheen, A. Naseer, and A. Jalal. "Efficient vehicle detection and tracking using blob detection and kernelized filter". In: *Proceedings of the 5th International Conference on Advancements in Computational Sciences (ICACS)*. IEEE, 2024, 1–8. DOI: 10.1109/ICACS60934.2024.10473292.

[13] H. Yang, Q. He, L. Cui, and A. M. Mohamed Taha, (2023) *"Exploring the spatial relationship between urban vitality and urban carbon emissions"* **Remote Sensing** ***15***(8): 2173. DOI: 10.3390/rs15082173.

[14] A. Zahid, T. Qasim, N. Bhatti, and M. Zia, (2024) *"A data-driven approach for road accident detection in surveillance videos"* **Multimedia Tools and Applications** ***83***(6): 17217–17231. DOI: 10.1007/s11042-023-16193-0.

[15] F. Yao, (2025) *"Classroom anomalous behavior detection based on improved YOLOv5"* **Journal of Applied Science and Engineering** ***28***(7): 1473–1481. DOI: 10.1080/0000000.

[16] X. Liu, (2024) *"Detection of abnormal tourist behavior in scenic spots based on optimized Gaussian model for background modeling"* **Journal of Intelligent Systems** ***33***(1): 20240092. DOI: 10.1515/jisys-2024-0092.

[17] G. Saravanan, G. Yamuna, and S. Nandhini. "Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models". In: *Proceedings of the International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2016, 462–466. DOI: 10.1109/ICCSP.2016.7754179.