Enhancing Urban Traffic Management: Advanced Strategies in Image Recognition-Based Intelligent Traffic Monitoring

Jianli Pang

College of Electronic Information, Huanghuai University, Zhumadian 463000, China

Corresponding Author Email: pangjianli@huanghuai.edu.cn

ABSTRACT

As urban traffic pressures escalate, intelligent traffic monitoring systems have become instrumental in mitigating congestion and enhancing road safety. These systems, founded on image recognition technology, are central to the evolution of intelligent urban traffic management by providing real-time, precise vehicle identification and traffic violation detection. Despite their significance, current technologies exhibit limitations in license plate recognition accuracy and in the precision and timeliness of traffic violation identification. This study addresses these critical bottlenecks in existing intelligent traffic monitoring systems and proposes optimization strategies. Enhanced template matching techniques are employed to improve the recognition process of English letters and numerals in license plates. Concurrently, novel neural network algorithms are introduced to increase the accuracy of Chinese character recognition, a significant advancement considering the diversity and complexity of these characters. Additionally, the study innovates in traffic violation detection by integrating new relational connections and spatial attention modules into graph convolutional networks (GCNs). This integration substantially improves the system's understanding of complex traffic scenarios and enhances its processing efficiency, crucial for real-time monitoring applications. The research efforts detailed in this study not only elevate the capabilities of license plate recognition and traffic violation detection but also offer substantial technical support for the practical deployment and application of intelligent traffic systems. These advancements contribute significantly to intelligent traffic monitoring, presenting reliable and efficient solutions for contemporary urban traffic management challenges.

1. INTRODUCTION

With the acceleration of urbanization, traffic management has emerged as a complex challenge in modern city operations [1, 2]. Traditional traffic monitoring systems are often inadequate in processing the extensive volume of traffic image data efficiently and accurately [3-5]. In this context, intelligent traffic monitoring systems, leveraging advanced computer vision technology, have been developed. These systems enhance the efficiency and precision of real-time monitoring and management of traffic conditions [6-8]. The essence of these systems is rooted in the efficient recognition of vehicle images, encompassing the automatic identification of license plates and the accurate detection of traffic violations, both crucial for reducing traffic congestion and improving road safety.

The research and enhancement of intelligent traffic monitoring systems are critically important for the realization of sophisticated urban traffic management, reduction of traffic accident rates, and improvement of road utilization efficiency. The technology of license plate recognition, a cornerstone of intelligent traffic systems, is pivotal for effective traffic monitoring and automated enforcement [9-12]. Furthermore, the automated detection of traffic violations aids traffic managers in responding promptly, thus enabling timely interventions and bolstering technical support for urban traffic management [13-16].

Nonetheless, current image recognition-based algorithms for license plate recognition necessitate improvements in adaptability and accuracy across diverse license plate environments. This is particularly challenging in identifying a variety of fonts and formats of English letters, numbers, and Chinese characters [17-20]. In the domain of traffic violation detection, existing image processing techniques also reveal substantial potential for enhancement in both accuracy and real-time performance, especially in complex scenarios. Such limitations partially impede the practical application effectiveness of intelligent traffic monitoring systems [21-24].

This paper addresses a method for license plate recognition that utilizes refined template matching techniques to optimize the recognition efficiency of English letters and numbers. Innovations in neural network algorithms are introduced to augment the accuracy of Chinese character recognition. Furthermore, this study proposes a novel framework for traffic violation detection by incorporating new edge connections into GCNs, along with spatial attention modules. This framework is designed to enhance the system's comprehension and processing speed in complex traffic scenarios. These

research efforts not only address the technical gaps in the field of intelligent traffic monitoring but also provide robust technical support and a theoretical foundation for the development of more efficient and intelligent urban traffic management systems.

2. LICENSE PLATE RECOGNITION BASED ON INTELLIGENT TRAFFIC MONITORING IMAGE PROCESSING

In the methodology section, a method utilizing enhanced neural networks and refined template matching for license plate recognition is proposed. This approach significantly improves the recognition of English letters and numerals on license plates, thereby enhancing the system's adaptability and accuracy across diverse traffic environments. The development of a neural network algorithm, specifically tailored for the recognition of Chinese characters, addresses the limitations inherent in traditional license plate recognition technologies, particularly when processing complex Chinese scripts. Such an advancement is vital for intelligent traffic monitoring systems that operate in environments featuring license plates in multiple languages. The method's effectiveness extends beyond mere vehicle identification; it contributes to optimizing the management strategies of intelligent traffic monitoring systems. Consequently, this leads to an increase in both the precision and response speed of vehicle control, and provides substantial data support for more intelligent and automated traffic management. Figure 1 in the document illustrates the structural design of the license plate recognition model.

The methodology revolves around character normalization, a process that adjusts characters in the license plate image to uniform sizes, orientations, and grayscale ranges. This normalization is essential for counteracting image deformations attributable to varying camera angles, lighting changes, and inconsistencies in license plate sizes, thereby ensuring a standardized input for subsequent recognition stages. It is posited that for a pixel processed and represented as (l, b), with the original image dimensions being $Z \times T$ and the processed dimensions $L \times B$, its coordinates in the original image are delineated as follows:

$$z = l \times \frac{Z}{L}, t = b \times \frac{T}{B} \qquad (1)$$

Adjacent to this, the methodology calculates grayscale values for the midpoints between the closest neighboring pixels to pixel o, denoted as s, n, v, and f, with respective coordinates (z_0, t_0), (z_{0+1}, t_0), (z_0, t_{0+1}), and (z_{0+1}, t_{0+1}) and grayscale values $h(s)$, $h(n)$, $h(v)$, and $h(f)$. The formulas for determining the grayscale values of the midpoints between s and n, and between v and f, are articulated as:

$$h(r) = \overline{sr}\left[h(n) - h(s)\right] + h(s) \qquad (2)$$

$$\begin{aligned} h(d) &= \overline{vd}\left[h(f) - g(v)\right] + h(v) \\ z_0 &= [z], t_0 = [t] \\ z_1 &= z_0 + 1, t_1 = t_0 + 1 \\ \overline{sr} &= z - z_0, vd = z - z_0, \overline{ro} = t - t_0 \end{aligned} \qquad (3)$$

The grayscale value for pixel o is calculated using the formula:

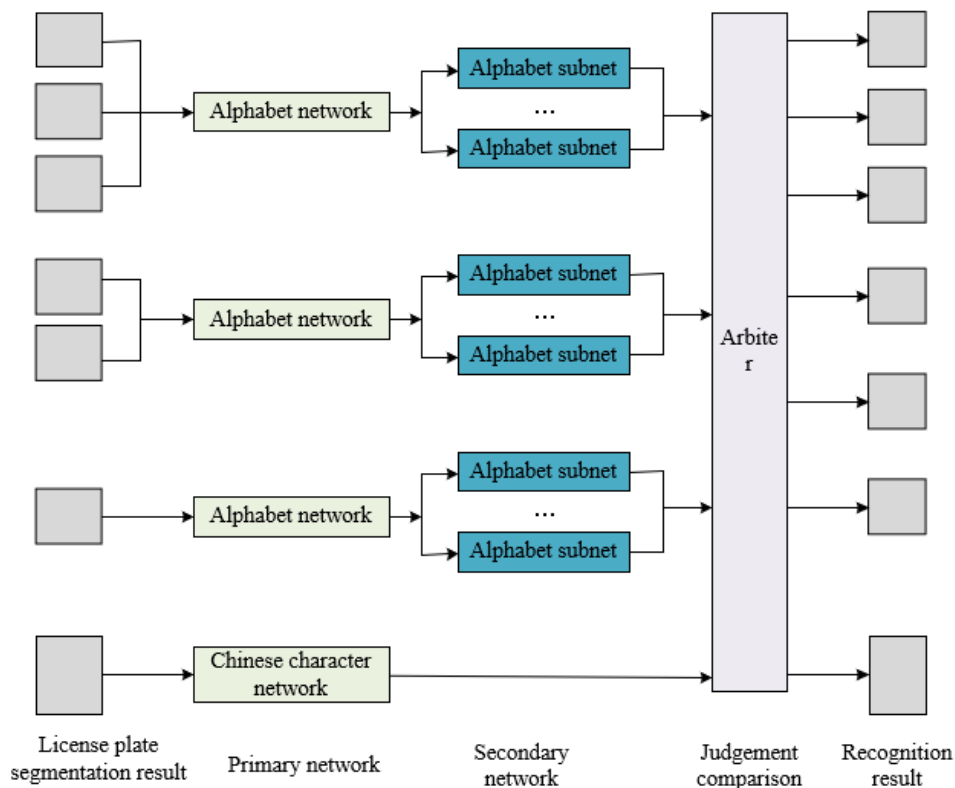$$h(o) = EP\left[h(d) \cdot h(r)\right] + h(r) \qquad (4)$$

Figure 1. Structure of the license plate recognition model

In the methodology of this study, a critical step involves the normalization of license plate images, as depicted in Figure 2. Subsequent to this normalization, a process of character refinement is undertaken, specifically targeting the extraction of stroke information from Chinese characters within the license plate imagery. This refinement step aims to enhance the precision of Chinese character recognition. It typically encompasses the transformation of character images into a skeletal form, effectively reducing each character to its medial axis. The primary objective of this process is the retention of essential recognition information, while concurrently eliminating extraneous noise and details. The initial phase of this refinement involves the computation of the average width of Chinese characters. This calculation is generally conducted by analyzing the distribution of pixel widths within characters and determining their mean value. This initial computation furnishes crucial parameters for the subsequent application of skeleton extraction algorithms. Among these algorithms are thinning algorithms and distance transform algorithms, utilized to ascertain the medial axis of each character stroke. A part of this process entails calculating the average count of black pixels along segments perpendicular to the medial axis, thereby aiding in the determination of the stroke's width and shape. A black pixel is identified as the central point of a horizontal stroke if it fulfills conditions such as $vU>n3$, $vU>n7$, $v5>n3$, $v5>n7$, and $(vU+v5)>2\times(n3+n7)$, with $|n3+n7|\leq1$. Similarly, the central point of a vertical stroke is determined if a black pixel satisfies conditions like $v3>nU$, $v3>n5$, $v7>nU$, and $v3+v7>2\times(nU+n2)$, with $|n1-n5|\leq1$. In instances where the directly extracted medial axis does not sufficiently represent the key structural elements of a character, adjustments are implemented through a process of weighted processing. This process may involve allocating varying weights to pixels in proximity to the medial axis, thus more accurately reflecting the actual width of the strokes. In the final stage of this refinement process, distinct types of strokes, such as horizontal, vertical, and oblique, are delineated using varying grayscale values. This differentiation clearly distinguishes the stroke types, providing distinctive features essential for the training of subsequent neural networks.
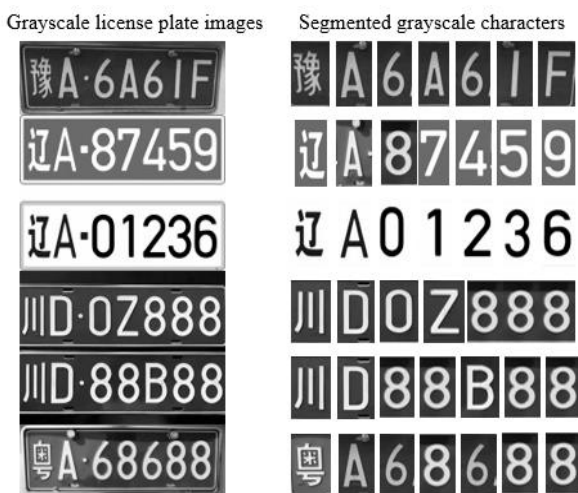

Figure 2. Normalized license plate images

Enhancing the recognition of English letters and numerals is achieved through an improved template matching process, which involves establishing a template library and setting a threshold. This process begins with preprocessing the

segmented license plate image, incorporating noise reduction and contrast enhancement to improve recognizability. The preprocessed image is then resized to a uniform dimension, crucial for the subsequent template matching as it necessitates the input image and template to be of the same scale. A comprehensive template library, containing various English letters and numerals, is created with these templates normalized to target sizes and styles. Assuming the input function is represented by $d(z,t)$, and the standard template by $D(z,t)$, the output of the comparison is denoted by $Y(z,t)$, with the correlator output expression given as follows:

$$Y(z1-z2,t1-t2) = \iint d(z,t)D(z+(z1-z2),t+(t1-t2))\,fzft \tag{5}$$

When $z1=z2$ and $t1=t2$, the formula is expressed as:

$$Y(0,0) = \iint d(z,t)D(z,t)\,fzft \tag{6}$$

When $d(z,t)$ equals $D(z,t)$, the equation is articulated as:

$$Y(0,0) = \iint d(z,t)d(z,t)\,fzft \tag{7}$$

It is established that if the peak value $Y(0, 0)$ is greater than or equal to $Y(z, t)$, then the main peak of $Y(z, t)$ is identified at $Y(0, 0)$. This phenomenon is pivotal for character recognition, as it becomes feasible when the main and secondary peaks are distinct. A threshold, determined by existing positional information and prior knowledge, is set to assess whether the degree of match meets the established standard. The process involves comparing the normalized character with templates in the library and calculating the degree of match. In instances where the highest match degree of a template falls below the designated threshold N, the character is temporarily deemed unrecognizable. Conversely, if a single template achieves the highest match degree, the corresponding character is identified. When multiple similar characters are discerned, additional feature description methods are employed for differentiation. Should no template match degree surpass the threshold, this indicates an unsatisfactory match, potentially attributable to character wear or noise interference.

The template is defined as a discrete function, denoted as $Y_{z,t}$. Under the assumption that Gaussian white noise, represented as $U_{z,t}$, with a standard deviation σ, affects the license plate image, the probability of matching corresponding pixels is calculated.

$$o_{u,k}(z,t) = \frac{1}{\sqrt{2\tau}\sigma}e^{-\frac{1}{2}\sum_{(z,t)\in\Pi}\left(\frac{U_{z+u,t+u}-U_{z,t}}{\sigma}\right)^2} \tag{8}$$

This calculation leads to:

$$M_{u,k} = \prod_{(z,t)\in Q} o_{u,k}(z,t) \tag{9}$$

i.e.,

$$M_{u,k}(z,t) = \left(\frac{1}{\sqrt{2\tau}\sigma}\right)^b e^{-\frac{1}{2}\sum_{(z,t)\in\Pi}\left(\frac{U_{z+u,t+u}-Y_{x,y}}{\sigma}\right)^2} \tag{10}$$

This culminates in the formulation of a minimization problem:

$$MIN\ r = \sum_{(z,t \in Q)} \left(U_{z+u,t+u} - Y_{z,t} \right)^2 \qquad (11)$$

Let Y be the character image of the standard template. The correlation operator, crucial for the matching process, is defined between the image to be detected, represented by $D(l, b)$, and the sub-image of the character to be detected, denoted as $D_{zt}(l, b)$. The correlation operator quantifies the similarity between $D(l, b)$ and $D_{zt}(l, b)$, facilitating precise character recognition.

$$E(u,k) = \frac{\sum_{l=1}^{L}\sum_{b=1}^{B}\left[d_{zt}(l,b) \times Y(l,b) \right]}{\sqrt{\sum_{l=1}^{L}\sum_{b=1}^{B}\left[D_{zt}(l,b) \right]^2}\sqrt{\sum_{l=1}^{L}\sum_{b=1}^{B}\left[Y(l,b) \right]^2}} \qquad (12)$$

Despite the application of normalization and standard template matching, the structural similarities of characters and image quality issues such as stroke adhesion, breaks, or blurring can still pose significant challenges in character recognition. To address this, feature description methods have been employed, offering an in-depth analysis of characters by extracting critical features like shape contours, stroke direction, and the distribution of corner points. These methods generate a feature vector that accurately describes and distinguishes the unique attributes of characters. For instance, numerals and letters with similar appearances, such as "8" and "B", may pose differentiation challenges in traditional template matching. However, feature description methods, by extracting detailed attributes like the size and shape of internal spaces and the connections between strokes, effectively differentiate between such characters. In cases of stroke adhesion, these methods discern individual characters by analyzing spatial relationships between strokes. For characters that are broken, the continuity of feature points is utilized to reconstruct the original appearance of the character. In instances of blurred images, character recognizability is enhanced through the emphasis on edge features provided by feature description methods.

The process begins with the extraction of key visual features from license plate character images. These features, not limited to but including edge features, corners, contours, stroke direction, and spacing, describe the geometric and structural properties of characters. They maintain a level of independence and distinctiveness, even in situations where characters may be adhered or damaged. The extracted features are then quantified into feature vectors, effectively transforming the character images into numerical descriptors within a high-dimensional feature space. Subsequently, these feature vectors undergo a process of training and classification. During training, the algorithm distinguishes the differences between feature vectors of various characters, constructing a classification model. For each character's feature vector to be recognized, the model outputs a classification result, signifying the most likely character category corresponding to that feature vector. When characters are similar, the model utilizes the differences learned in features to make final determinations. Post-processing steps, such as the application of contextual information of characters or statistical data, might be necessary after initial recognition results are obtained

to further enhance recognition rates.

To capture the contours of license plate characters more comprehensively, the Canny edge detection operator is incorporated into the matching process. This operator employs two distinct thresholds to identify strong and weak edges. The lower threshold is used for detecting weak edges, which may not be prominent due to factors like image blurring, noise, or suboptimal lighting conditions, while the higher threshold is dedicated to recognizing strong, clearly defined edges. This method enables a more thorough capture of license plate character contours, including edges that might be less visible under insufficient lighting or poor image quality. The spatial relationship between the characters and the edges of the license plate, a unique physical structural feature, is effectively identified and utilized by the Canny operator for edge extraction, thereby enhancing the prominence of characters. The presence of clear edge contours allows the system to utilize edge information for character positioning and recognition, even when matching templates are not available.

In addressing the complex task of Chinese character recognition, traditional gradient descent algorithms have been further refined to enhance the efficiency and stability of neural network training. The inherent structural complexity and stroke diversity of Chinese characters contribute to a more rugged and uneven loss surface in the optimization problem. Conventional gradient descent algorithms, characterized by fixed learning rates and gradient magnitudes, may result in slow convergence or oscillations at local minima, thereby hindering the attainment of global optima. The innovation in the proposed algorithm lies in its use of only the signs of the partial derivatives to indicate the direction of weight changes. This simplification of gradient information effectively reduces computational complexity. An adaptive mechanism for adjusting the update value is employed, thereby regulating the learning step size. The constancy of the sign of a partial derivative signifies that the current direction of weight updating is effective. In such scenarios, an increase in the update value can expedite the optimization process. Conversely, a change in the sign of the partial derivative indicates the necessity to reassess the optimization direction. Here, a reduction in the update value aids in preventing excessive adjustments and circumvents oscillations. This strategic approach enables the neural network to adjust weights more responsively, particularly in intricate optimization landscapes. It achieves a balance between exploration and exploitation, thereby enhancing the speed of convergence and ensuring a more stable training process. The iterative process, assuming the previous update value is denoted by $\Delta Z^{(j)}$, is elucidated as follows:

$$Z^{(j+1)} = Z^{(j)} - \Delta Z^{(j)} * SIGN\left(\nabla dZ^{(j)} \right) \qquad (13)$$

3. TRAFFIC VIOLATION DETECTION BASED ON INTELLIGENT TRAFFIC MONITORING IMAGE PROCESSING

In the realm of intelligent traffic monitoring, traditional GCNs, predominantly based on the static physical structure of human skeletons, are often inadequate for fully capturing the intricate interactions and dependencies among bones during dynamic movements. This limitation can adversely affect the efficiency of data transmission and the accuracy of motion

recognition. To address these challenges, this study proposes an innovative approach by integrating novel relational connections, specifically edge connections, into the conventional GCN model. Such connections are designed to capture the synergistic effects between bones during specific actions, thereby enabling a more accurate simulation and analysis of human motion patterns. Figure 3 exemplifies the architecture of the conventional adaptive GCN model.

In the developed model for detecting traffic violations within the framework of intelligent traffic monitoring image processing, three types of relational connections are employed: self-connection, physical structure connection, and edge connection. Self-connections are utilized to preserve the intrinsic features of each node within the model, maintaining the original feature information of the nodes and preventing excessive smoothing during graph convolution. This ensures the unique characteristics of each node are accurately captured. Physical structure connections are established based on the physical proximity of vehicles within images, reflecting the direct spatial relationships between them. These connections

are instrumental in enabling the model to learn about the interactions and mutual constraints among vehicles, which is essential for comprehending the relative movements and positional changes of vehicles in complex traffic scenarios. The introduction of edge connections, which are predicated on relational reasoning, forms new edges that abstract relationships such as the continuity and predicted paths of vehicle movements. This feature effectively captures dynamic information across consecutive frames, thus augmenting the model's capability to track and interpret trends in vehicle behavior. Collectively, these relational connections facilitate the model's ability not only to identify traffic violations within single-frame images but also to analyze dynamic behavioral changes in vehicles over time. This multifaceted approach significantly enhances the accuracy and efficiency of traffic violation detection, aligning closely with the objectives of real-time, dynamic monitoring in intelligent traffic monitoring systems. Figure 4 presents the comprehensive framework of the traffic violation detection model.
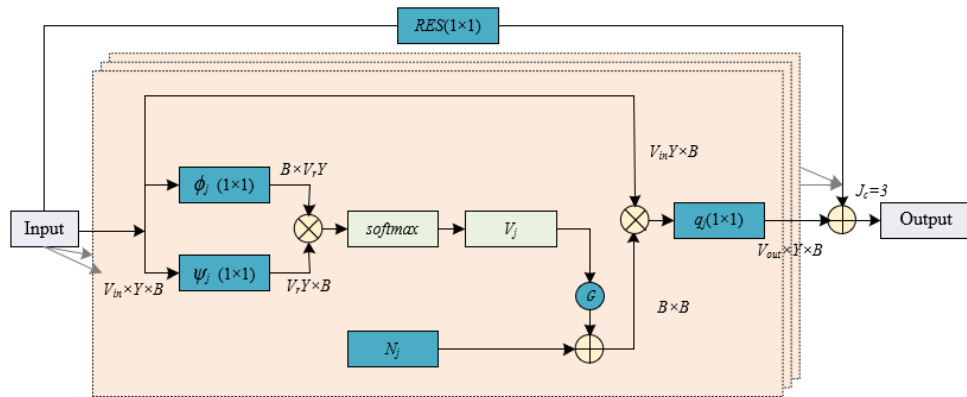

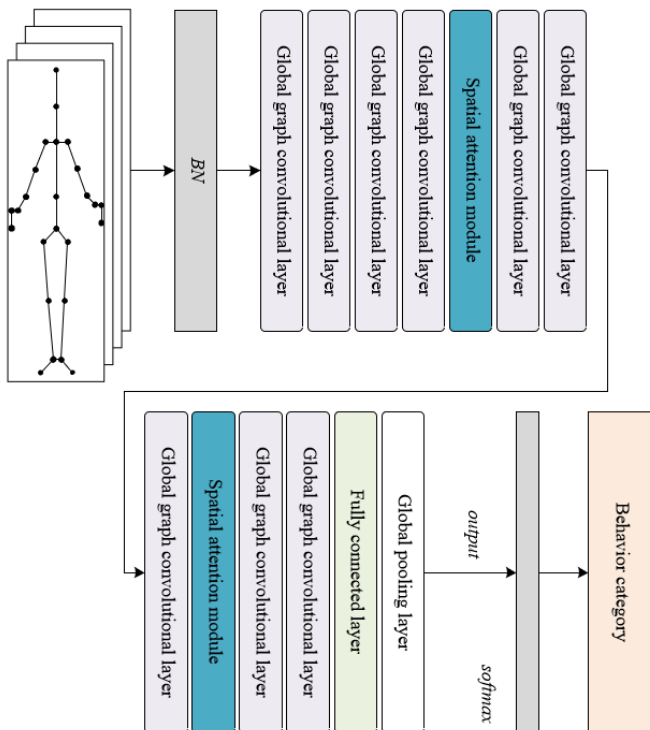**Figure 3.** Architecture of the conventional adaptive GCN model

Figure 4. Overall framework of the traffic violation detection model

The three distinct types of relational connections, namely, self-connection, physical structure connection, and edge connection, are represented by respective vertex connection matrices, denoted in the study as S^-_{LO}, S^-_{PH}, and S^-_{ED}. S^-_{LO} is conceptualized as a $B \times B$ identity matrix, and its operational mechanism within the model is delineated by the following expression:

$$\overline{S}_{LO} = U_{B \times B} \tag{14}$$

S^-_{PH} is formulated to represent the $B \times B$ adjacency matrix that corresponds to the physical structure of the human skeleton. This matrix is integral to understanding the spatial relationships within the skeletal framework, with its expression provided as:

$$\overline{S}_{PH(uk)} = \begin{cases} 1, & \textit{Connection between vertices } c_u \textit{ and } c_k \\ 0, & \textit{else} \end{cases} \tag{15}$$

The innovative integration of edge connection S^-_{ED} into the intelligent traffic monitoring system is a key focus of this study. Its primary objective is to enhance the model's comprehension of the dynamic nature of vehicle behaviors in traffic scenarios, thereby elevating the performance of traffic violation detection. Edge connections are designed to model potential relationships between vehicles, such as interactions

in traffic flow and following behaviors, thus more accurately capturing the sequentiality and coherence of vehicle behaviors. These dynamic relationships are crucial for predicting future vehicle positions and behavioral patterns, as well as for understanding potential traffic violations under specific traffic conditions. For instance, edge connections are adept at capturing reactions from trailing vehicles when a leading vehicle executes a sudden braking maneuver or the reciprocal adjustments among vehicles during lane-changing scenarios, insights typically elusive with conventional static image-based methods.

To achieve optimal functionality, the relational matrices S^-_{LO}, S^-_{PH}, and S^-_{ED} necessitate normalization through respective degree matrices. In the constructed traffic violation detection model, each type of relational connection, self-connection, physical structure connection, and edge connection, is associated with a corresponding normalized relational matrix. These matrices are instrumental in the spatial domain update rules of the model. The normalization matrix for self-connection is vital in maintaining the significance of each node's intrinsic information, preventing the neglect of self-features during feature propagation, and thus preserving the stability of node states. The normalization matrix for physical structure connection reflects direct interactions between nodes (vehicles) based on their physical proximity, facilitating the model's ability to update a node's state using features of adjacent nodes, thereby capturing spatial interrelations among vehicles. The formulas for these normalized relational matrices, S_{LO}, S_{PH}, and S_{ED}, are articulated as follows:

$$S_{LO} = \overline{S}_{LO} \tag{16}$$

$$S_{PH} = \Lambda_o^{-\frac{1}{2}} \overline{S}_{PH} \Lambda_o^{\frac{1}{2}} \tag{17}$$

$$S_{ED} = \Lambda_o^{-\frac{1}{2}} \overline{S}_{ED} \Lambda_o^{\frac{1}{2}} \tag{18}$$

In the developed traffic violation detection model, the spatial domain update rules are structured to comprehensively integrate a vehicle's inherent features, the interactions between vehicles, and the dynamic patterns of vehicle behavior. Consequently, each iteration in the model updates the node's feature representation, not solely based on the node's current state but also by incorporating contextual information and temporal variations, thus enriching and enhancing the accuracy of the feature representation. The expression governing the model's spatial domain update rule is formulated as follows:

$$d_{OUT}(c_u) = \sum_{k=1}^{B} \left(\begin{matrix} S_{LO(uk)}Z^{(m)}(c_k)Q_m^{(m)} \\ +S_{PH(uk)}Z^{(m)}(c_k)Q_o^{(m)} \end{matrix} \right. \\ \left. +S_{ED(uk)}Z^{(m)}(c_k)Q_r^{(m)} +V_{j(uk)}Z^{(m)}(c_k)Q_v^{(m)} \right) \tag{19}$$

The traffic violation detection model established in this study represents a deep learning framework that effectively employs GCNs and temporal convolution. Initially, the model processes preprocessed traffic monitoring image data, which encompasses spatial location information of vehicles and their dynamic attributes evolving over time, providing the model with comprehensive and original input data. Following the input layer, a graph convolutional operator is deployed to

extract features related to the vehicle skeleton. By applying the previously mentioned relational matrices, this operator proficiently captures both the direct and indirect interactions among vehicles within the spatial domain, in addition to the vehicles' own characteristics, thereby furnishing enriched spatial features for subsequent temporal analysis. The methodology further incorporates the application of 1D convolutional kernels along the temporal axis to the skeletal data, a critical step for extracting dynamic information over time. This aspect is crucial in interpreting the continuity and evolving patterns of vehicle behavior. A residual connection structure is subsequently implemented within the model. These residual connections are pivotal in mitigating the vanishing gradient issue common in deep networks, thus facilitating the learning of more profound features without compromising training performance. In this context, the residual connection structure also enables the model to account for varying spans of vehicle behavior during temporal analysis, thereby capturing more intricate sequential features. To adjust the temporal frame size of the input features, the methodology integrates temporal feature pooling. This process effectively reduces the temporal dimensionality of the data, aiding in decreasing overall model complexity and computational demands while preserving essential temporal information. In the final segment of the network, convolutional layers are utilized to modify the channel counts of feature maps, further abstracting and compressing these features to provide high-level inputs for the concluding classification or regression tasks. Figure 5 illustrates the network unit structure of the traffic violation detection model.
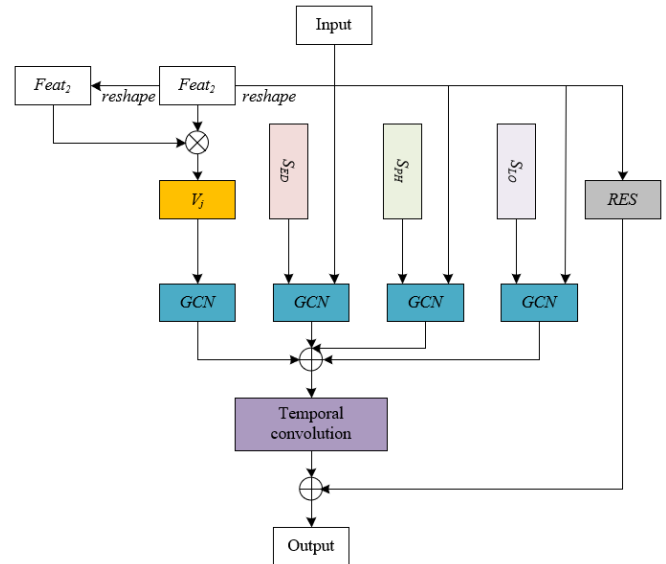
Figure 5. Network unit structure of the traffic violation detection model

The GCNs traditionally employed in traffic violation detection typically utilize adjacency matrices that are static in nature. However, these fixed matrices might not adequately represent the evolving interactions and dependencies between vehicles over time. To address this limitation, the integration of a spatial attention mechanism within the GCN framework has been proposed. This mechanism dynamically adjusts the relational weights between nodes, informed by the learned features from the input data. Essentially, the model autonomously discerns which vehicle interactions are most pertinent for detecting traffic violations. Input features are

represented by $Z_{IN} \in E^{V \times Y \times B}$ and output features by $Z_{OUT} \in E^{V \times Y \times B}$. The series of vectors, $D_{IN} = \{z_1, z_2, ..., z_B\}$ and $D_{OUT} = \{ct_1, t_2, ..., t_B\}$, with individual elements z_u and t_k residing in $E^{V \times Y}$, encapsulate the data. The linear embeddings are defined as $\phi(z_u) = Q_\phi z_u$, $\theta(x_i) = Q_\theta z_k$, and $h(z_u) = Q_h z_k$. The spatial attention mechanism is characterized by a similarity function denoted by d:

$$t_u = \sum_k \text{softmax}\left(d\left(\varphi(z_u)^Y, \theta(z_k)\right)\right) h(z_k) \qquad (20)$$

This similarity scalar, produced by d between feature vectors at positions u and k, is defined by the equation:

$$d(i,c) = 1 \Big/ \arccos\left(\frac{ic}{|i| \cdot |c|}\right) \qquad (21)$$

Additionally, the attention module incorporates a residual connection, enhancing the description of similarity. It also modifies the internal connections within matrix A, resulting in a new matrix $A^\wedge$, defined as $A^\wedge = A\text{-}1$. If the shapes of input and output features D_{IN} and D_{OUT} are transformed into $D_{IN} \in E^{VY \times B}$ and $D_{OUT} \in E^{VY \times B}$, respectively, the matrix form of the spatial attention unit is articulated as:

$$D_{OUT} = \hat{A} Q_h D_{IN} \qquad (22)$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

The evaluation of the license plate recognition model across various scenarios, as delineated in Table 1, reveals significant findings. Positioning accuracy was observed to be consistently high across all scenarios, with values ranging from 93.8% to 96.8%. This high degree of accuracy reflects the model's efficacy in accurately locating license plates within diverse environments. The false detection rate for vehicle plates remained low across scenarios, ranging between 1.3% and 2.4%, indicating the model's proficiency in minimizing false positive identifications of non-license plate areas as license plates. Variability was noted in the vehicle detection rates, with a maximum of 94.5% in Scenario 4 and a minimum of 77.8% in Scenario 2. Despite this variation, these rates underscore the model's general reliability in detecting vehicles. The character recognition rate and segmentation accuracy further exemplify the model's robust performance. Segmentation accuracy, a critical factor in license plate recognition, was notably high in all scenarios, with the lowest rate being 95.7%. Recognition speed exhibited relative consistency across different scenarios, ranging from 23 to 24 seconds, suggesting the model's ability to process diverse scenes at a stable rate. I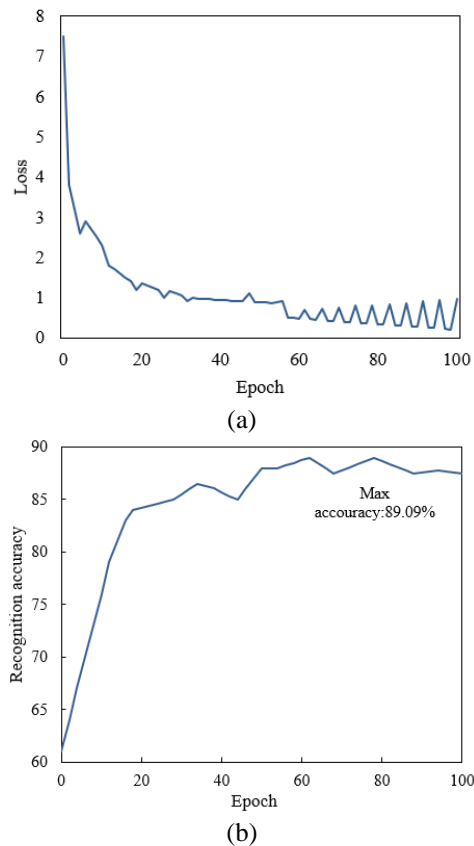n terms of overall recognition rates, the model demonstrated commendable performance in all tested scenarios, with rates ranging from 81.4% to 88.9%. These rates affirm the model's effectiveness and reliability in practical applications. The test results collectively indicate that the proposed license plate recognition model excels in various crucial aspects: positioning accuracy, low false detection rates, reliable vehicle detection, and exceptional character recognition capabilities across different scenarios. Notably, the enhancements in recognizing Chinese characters substantially contribute to the model's increased accuracy. While maintaining a rapid recognition speed, the overall recognition rate of the model is consistently high, underscoring the model's efficacy and suitability for practical license plate recognition applications.

Figure 6 presents the training error trajectory of the license plate recognition model over successive epochs. A declining trend in training error was observed as the number of training iterations increased, indicative of the model's progressive adaptation and improved proficiency in license plate recognition tasks. In the initial phase of training (Epoch 0 to 20), a rapid descent in training error from 7.5 to 2.3 was noted, signifying the model's swift acquisition of relevant features from an initially random state and the commencement of its basic recognition capabilities. Subsequent training between 20 to 60 epochs saw a continued reduction in error, though at a decelerated pace, reflecting the model's refinement in feature learning and parameter optimization, culminating in a training error reduction to approximately 1.2. From 60 to 100 epochs, training error further diminished, stabilizing around 0.2, suggesting the model's approach towards an optimized state. This data underscores the efficacy of the proposed license plate recognition methodology throughout the training process. The consistent reduction in training error signifies the model's effective extraction and optimization of distinguishing features from the training data, enhancing recognition accuracy. This efficiency not only validates the methodology but also demonstrates the model's robust convergence properties, ensuring reliable performance for practical application.

The license plate recognition model depicted in Figure 6 illustrates an overall ascendant trend in recognition accuracy with increased training iterations. This progression evidences the model's continuous adaptation and enhancement in recognition capabilities throughout the training. The model's recognition accuracy notably improved from 61% to 89%. Although minor fluctuations and a slight decrease in accuracy were observed in later training stages, these variations did not detract from the overall effectiveness of the proposed recognition method. Such fluctuations, typical in learning processes, can be mitigated through appropriate training strategies like early stopping or adaptive learning rate adjustments to avoid overfitting while maintaining model performance.

Table 1. License plate recognition model test results in different scenarios

Scenario Number	Positioning Accuracy (%)	False Detection Rate of Vehicle Plates (%)	Vehicle Detection Rate (%)	Character Recognition Rate (%)	Segmentation Accuracy (%)	Recognition Speed (sec)	Overall Recognition Rate (%)
1	96.8	1.3	91.2	93.1	99.3	23	88.9
2	94.2	2.4	77.8	96.7	98.2	24	82.3
3	93.8	2.1	81.2	91.2	95.7	24	81.4
4	95.6	1.9	94.5	84.5	96.3	23.6	83.6

Figure 6. Training loss value and recognition accuracy of the license plate recognition model

Table 2 provides a comparative analysis of performance metrics for different traffic violation recognition models, including Multilayer Perceptron-Convolutional Neural Network (MLP-CNN), Directional Graph Convolutional Network (DGCN), and the model proposed in this study. The proposed model exhibits superior performance with an accuracy of 93.5%, surpassing MLP-CNN's 92.3% and DGCN's 91.8%. This indicates the proposed model's enhanced ability in accurately identifying traffic violations. The recall rate for the proposed model is particularly noteworthy, achieving 95.8%, significantly higher than MLP-CNN's 88.9% and DGCN's 82.3%, indicating fewer missed traffic violations and more accurate identification of true positives. In terms of mean average precision (mAP), the proposed model slightly outperforms MLP-CNN (92.4%) and DGCN (92.6%) with a score of 93.4%, demonstrating superior robustness and recognition accuracy. The model size of the proposed model, at only 12.8M, is considerably smaller than MLP-CNN's 123M and DGCN's 47.8M, indicating a more compact model with reduced storage requirements. Regarding computational complexity, the proposed model's Flops are at 27.5G, higher than DGCN's 15.9G but significantly lower than MLP-CNN's 144.7G, suggesting higher computational efficiency while maintaining robust performance. The processing speed of the proposed model, measured at 67 frames per second (FPS), is substantially higher than MLP-CNN's 32 FPS and DGCN's 41 FPS, demonstrating its suitability for real-time traffic violation detection tasks. The model's superior performance across accuracy, recall, and computational efficiency metrics confirms its effectiveness and suitability for practical traffic violation detection applications.

The accuracy of the traffic violation recognition model proposed in this study, alongside the DGCN and MLP-CNN models, is depicted in Figure 7 across various training epochs. It was observed that the proposed model consistently outperformed the DGCN and MLP-CNN models in all stages of training, particularly in the latter phases. This enhanced accuracy and stability highlight the model's superiority in recognizing traffic violations, a critical aspect for practical deployment in traffic monitoring systems.

The precision-recall (PR) curve, an evaluative metric for classifiers, is presented in Figure 8. This curve illustrates the relationship between precision and recall at varying threshold levels. For the model in this study, a consistent high precision around 0.96 was maintained as the threshold increased from 0 to 0.8. However, beyond the 0.8 threshold, a decline in precision was noted, reaching zero at the threshold of 1. This pattern suggests that while the model sustains high precision in areas of high recall, it encounters challenges in maintaining precision at higher thresholds. Compared to the MLP-CNN and DGCN models, the proposed model demonstrated a more gradual decrease in precision at elevated thresholds, indicative of a more robust performance in balancing precision and recall. Notably, at thresholds approaching 1, the proposed model's reduced decline in precision reflects its ability to maintain a higher recall rate, thus minimizing the possibility of missed detections.

The ablation study, as documented in Table 3, was conducted to quantitatively evaluate the impact of introducing edge connections and spatial attention mechanisms on the performance of the traffic violation recognition model. The baseline model, prior to the integration of edge connections, exhibited an omission rate of 5.3%, a false detection rate of 21.5%, a recognition accuracy of 78.9%, and operated at a speed of 67 FPS. This baseline serves as a reference point, illustrating the model's performance without the benefits conferred by edge connections. The incorporation of edge connections, while excluding the spatial attention mechanism, led to notable improvements: the omission rate decreased to 3.2%, the false detection rate reduced to 16.2%, and recognition accuracy rose to 83.6%, albeit with a slight reduction in speed to 62 FPS. This enhancement underscores the contribution of edge connections to the model's performance, despite a marginal compromise in recognition speed.

Upon the integration of both edge connections and spatial attention mechanisms, the model's performance significantly advanced. The omission rate was substantially lowered to 1.2%, the false detection rate decreased to 5.3%, and recognition accuracy improved markedly to 93.6%. However, this enhancement in accuracy and detection efficacy came with a further reduction in recognition speed, now at 54 FPS. The integration of these advanced components markedly improved the model's accuracy, substantially reducing the omission and false detection rates, and achieving a significant leap in recognition accuracy. These results affirm the effectiveness of the proposed approach in enhancing recognition precision. While the reduction in speed might be perceived as a limitation, it is generally acceptable in scenarios where high accuracy is prioritized. Furthermore, the observed decrease in recognition speed is likely attributable to the increased computational complexity introduced by the advanced model components. Future iterations of this model could explore strategies such as hardware optimization or model compression to strike a more optimal balance between accuracy and processing speed, tailoring it to the specific requirements of various application contexts.

Table 2. Performance comparison of traffic violation recognition models

Model / Performance metrics	MLP-CNN	DGCN	The Proposed Model in This Study
Accuracy (%)	92.3	91.8	93.5
Recall (%)	88.9	82.3	95.8
mAP@0.5-0.95	92.4	92.6	93.4
Model size (M)	123	47.8	12.8
Flops(G)	144.7	15.9	27.5
FPS	32	41	67

Table 3. Ablation study comparison

Algorithm Model	Omission Rate (%)	False Detection Rate (%)	Recognition Accuracy (%)	Recognition Speed (FPS)
Before introducing edge connections	5.3	21.5	78.9	67
Before introducing spatial attention mechanism	3.2	16.2	83.6	62
The proposed model	1.2	5.3	93.6	54
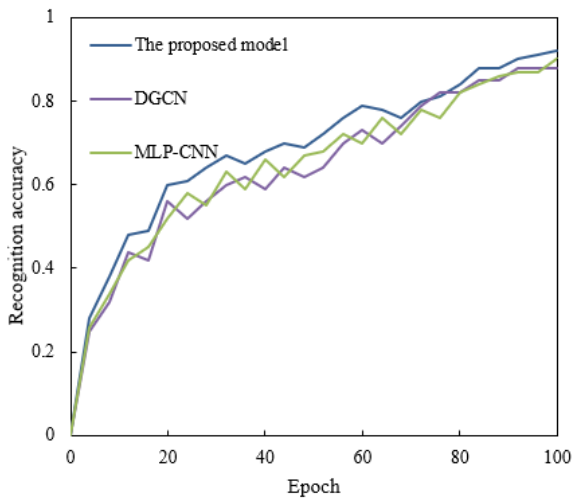
Figure 7. Accuracy of the traffic violation recognition model
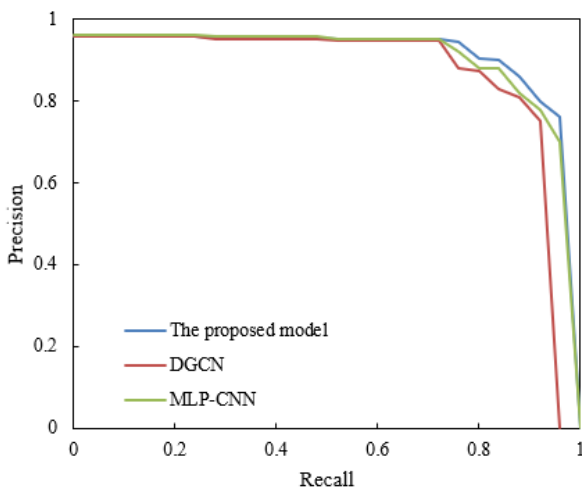
Figure 8. PR curves of traffic violation recognition models

5. CONCLUSIONS

This paper initially addressed the enhancement of license plate recognition technology based on template matching, specifically focusing on the augmented recognition efficiency for English letters and numerals. Concurrently, significant advancements were made in neural network algorithms, notably in elevating the accuracy of recognizing Chinese characters. This aspect is particularly pivotal for the recognition of Chinese license plates, considering the inherent diversity and complexity of Chinese characters which pose substantial challenges in recognition. A novel framework for the detection of traffic violations has been proposed, integrating innovative edge connections into GCNs and complementing these with spatial attention modules. This framework aims to enrich the model's comprehension of intricate traffic scenarios while expediting the processing speed to fulfill the requirements of real-time monitoring.

The experimental results have substantiated that the incorporation of edge connections and spatial attention mechanisms substantially enhances the efficacy of both license plate recognition and traffic violation detection. The outcomes of the ablation study underscore the criticality of these components in boosting the accuracy of the model. Post iterative training sessions, the model for license plate recognition exhibited a consistent trend of improvement, manifesting marked enhancements in terms of omission rate, false detection rate, and overall recognition accuracy. In the realm of traffic violation detection, the proposed model demonstrated superiority over comparative MLP-CNN and DGCN models across essential performance metrics, including accuracy, recall, and mAP. Additionally, the model displayed commendable computational efficiency, characterized by a reduced model size and an elevated FPS, rendering it apt for deployment in real-time applications.

In summary, the methodologies proposed in this paper have evidenced high accuracy and practicality in both domains of license plate recognition and traffic violation detection. They exhibit particular proficiency in processing complex Chinese characters and deciphering multifaceted traffic scenarios. These innovations are of significant importance for the actual implementation and application of intelligent traffic systems, offering more robust and efficient technological support for traffic regulation and management.

REFERENCES

[1] Hu, C.H., Liu, Y., Xu, L.T., Jing, X.Y., Lu, X.B., Yang, W.K., Liu, P. (2023). Joint image-to-image translation for traffic monitoring driver face image enhancement. IEEE Transactions on Intelligent Transportation Systems, 24(8): 7961-7973. https://doi.org/10.1109/TITS.2023.3258634

[2] Kazmi, S.Q., Singh, M.K., Pal, S. (2021). Traffic monitoring system in smart cities using image processing. In Intelligent Manufacturing and Energy Sustainability: Proceedings of ICIMES 2020, Springer, Singapore, pp. 397-405. https://doi.org/10.1007/978-981-33-4443-3_38

[3] Wang, Y.B., Zheng, J., Lü, Z.W., Yan, Y., Yuan, Y.J. (2019). Traffic monitoring image dehazing algorithm based on wavelength related physical imaging model. Acta Photonica Sinica, 48(9): 1902008. https://doi.org/10.3788/gzxb20194809.0910004

[4] Krishnamoorthy, R., Manickam, S. (2018). Automated traffic monitoring using image vision. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, pp. 741-745. https://doi.org/10.1109/ICICCT.2018.8473086

[5] Liu, S.C., Wu, P.D., Zhao, Z.J., Li, C.M. (2020). Image semantic segmentation and stitching method of traffic monitoring video. Acta Geodaetica et Cartographica Sinica, 49(4): 522–532. https://doi.org/10.11947/J.AGCS.2020.20190224

[6] Kilic, I., Aydin, G. (2022). Traffic lights detection and recognition with new benchmark datasets using deep learning and tensorflow object detection API. Traitement du Signal, 39(5): 1673-1683. https://doi.org/10.18280/ts.390525

[7] Nyati, E., Mahlalela, J.S. (2023). Advanced Vehicle Detection and License Plate Recognition via the Kanade-Lucas-Tomasi Technique. Mechatronics and Intelligent Transportation Systems, 2(4): 191-200. https://doi.org/10.56578/mits020401

[8] Wang, TM., Shen, H.W., Xue, Y.J., Hu, Z.K. (2020). A traffic signal recognition algorithm based on self-paced learning and deep learning. Ingénierie des Systèmes d'Information, 25(2): 239-244. https://doi.org/10.18280/isi.250211

[9] Ahmed, S., Ali, A., Naser, E. (2023). Tesseract OpenCV versus CNN: A comparative study on the recognition of unified modern Iraqi license plates. Revue d'Intelligence Artificielle, 37(5): 1331-1339. https://doi.org/10.18280/ria.370526

[10] Huang, H., Li, Z. (2021). FAFNet: A false alarm filter algorithm for license plate detection based on deep neural network. Traitement du Signal, 38(5): 1495-1501. https://doi.org/10.18280/ts.380525

[11] Adebayo, A.S., Olusola, J., Kazeem, R.A., Ikumapayi, O.M., Okokpujie, I.P., Uchegbu, I.D. (2022). Effective two-lane traffic management at the University of Ibadan, Nigeria main gate using multiple vehicle recognition systems. International Journal of Safety and Security Engineering, 12(6): 729-735. https://doi.org/10.18280/ijsse.120609

[12] Wang, L., Li, K. (2022). Design of license plate recognition system based on image processing. In Proceedings-2022 Asia Conference on Algorithms, Computing and Machine Learning, CACML 2022, Hangzhou, China, pp. 316-322. https://doi.org/10.1109/CACML55074.2022.00060

[13] Joshi, D., Mohd, N. (2023). Techniques used in automatic number plate recognition. In 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-6. https://doi.org/10.1109/INCET57972.2023.10170372

[14] Valliyammai, C., Sridharan, J., Ramanathan, A. (2020). Automation of traffic violation detection and penalization. In Advanced Computing and Intelligent Engineering - Proceedings of ICACIE 2018, pp. 235-247. https://doi.org/10.1007/978-981-15-1483-8_21

[15] Alkan, B., Balci, B., Elihos, A., Artan, Y. (2019). Driver cell phone usage violation detection using license plate recognition camera images. In VEHITS 2019- Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems, Heraklion, Crete, Greece, pp. 468-474. https://doi.org/10.5220/0007725804680474

[16] Jose, J.A.C., Billones, C.D., Brillantes, A.K.M., Billones, R.K.C., Sybingco, E., Dadios, E.P., Fillone, A.M., Gan Lim, L.A. (2021). Artificial intelligence software application for contactless traffic violation apprehension in the Philippines. Journal of Advanced Computational Intelligence and Intelligent Informatics, 25(4): 410-415. https://doi.org/10.20965/jaciii.2021.p0410

[17] Shvai, N., Hasnat, A., Nakib, A. (2023). Multiple auxiliary classifiers GAN for controllable image generation: Application to license plate recognition. IET Intelligent Transport Systems, 17(1): 243-254. https://doi.org/10.1049/itr2.12251

[18] Zhao, H. (2020). Application of license plate image recognition technology in intelligent parking lot. In ACM International Conference Proceeding Series, pp. 75-79. https://doi.org/10.1145/3444370.3444551

[19] Abdelaziz, A. H., Chan, Y. K., Koo, V.C. (2021). Enhancement for license plate recognition using image super resolution technique. In Proceedings of the 3rd International Conference on Electrical, Communication and Computer Engineering, ICECCE 2021, Kuala Lumpur, Malaysia, pp. 1-4. https://doi.org/10.1109/ICECCE52056.2021.9514106

[20] Huang, J. (2020). Research on license plate image segmentation and intelligent character recognition. International Journal of Pattern Recognition and Artificial Intelligence, 34(6): 2050014. https://doi.org/10.1142/S0218001420500147

[21] Kaur, P., Kumar, Y., Gupta, S. (2022). Artificial intelligence techniques for the recognition of multi-plate multi-vehicle tracking systems: A systematic review. Archives of Computational Methods in Engineering, 29(7): 4897-4914. https://doi.org/10.1007/s11831-022-09753-4

[22] Alkalai, M., Lawgali, A. (2020). Image-preprocessing and segmentation techniques for vehicle-plate recognition. In Proceedings of the 4th International Conference on Image Processing, Applications and Systems, IPAS 2020, Genova, Italy, pp. 40-45. https://doi.org/10.1109/IPAS50080.2020.9357430

[23] Shikiji, Y., Watari, K., Tsudaka, K., Wada, T., Okada, H. (2015). Novel vehicle information acquisition method using vehicle code for automotive infrared laser radar. In 2014 Australasian Telecommunication Networks and Applications Conference, ATNAC 2014, Southbank,

VIC, Australia, pp. 52-57. https://doi.org/10.1109/ATNAC.2014.7020873

[24] Garcia, M.R.T., Bandala, A.A., Dadios, E.P. (2021). Motorcycle apprehension using deep learning and K-nearest neighbor algorithm. In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, pp. 1-6. https://doi.org/10.1109/I2CT51068.2021.9417918