



# Application of High-Definition Imaging Technology in Traffic Violation Monitoring and Judicial Proceedings

Zijun Hu<sup>\*</sup>, Jia Li<sup></sup>

College of Humanities and Social Sciences, Hebei Agricultural University, Baoding 071001, China

Corresponding Author Email: [huzijun\\_hebei@outlook.com](mailto:huzijun_hebei@outlook.com)

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420234>

## ABSTRACT

**Received:** 9 November 2024

**Revised:** 27 February 2025

**Accepted:** 12 March 2025

**Available online:** 30 April 2025

### Keywords:

*high-definition (HD) imaging technology, traffic violation monitoring, multi-object tracking, DeepSORT algorithm, judicial evidence chain, Discrete Wavelet Transform-Singular Value Decomposition (DWT-SVD)*

With the rapid growth of urbanization and the increasing number of motor vehicles, traffic violations have become more frequent, posing severe challenges to road safety and urban management. The application of high-definition (HD) imaging technology in traffic violation monitoring enables more accurate and efficient detection and processing of violations, significantly enhancing the intelligence level of traffic management and the effectiveness of judicial evidence collection. While existing research has made progress in developing and optimizing traffic violation monitoring systems, there remain significant shortcomings in multi-object tracking and the construction of judicial evidence chains. Current multi-object tracking algorithms are susceptible to occlusion and target loss in complex traffic environments, and existing judicial evidence chain construction methods heavily rely on single-image processing techniques, which limits the reliability and completeness of the evidence. This study focuses on two key aspects: (1) multi-object tracking in HD traffic violation monitoring images based on an improved DeepSORT algorithm and (2) judicial evidence chain construction using the Discrete Wavelet Transform-Singular Value Decomposition (DWT-SVD) technique. The enhanced DeepSORT algorithm integrates deep learning with traditional tracking methods, significantly improving tracking accuracy and robustness in complex traffic conditions. Meanwhile, the DWT-SVD technique enables multi-level analysis and evidence extraction, strengthening the comprehensiveness and reliability of judicial evidence. This research not only advances the theoretical understanding of image processing in traffic violation monitoring and judicial applications but also provides valuable insights for practical implementation and broader adoption.

## 1. INTRODUCTION

With the continuous acceleration of urbanization and the sharp increase in the number of motor vehicles [1, 2], the frequent occurrence of traffic violations has brought great challenges to road traffic safety and urban management [3, 4]. In order to effectively curb traffic violations, the application of HD imaging technology in traffic violation monitoring has become increasingly important [5, 6]. By combining HD cameras with intelligent analysis systems [7, 8], traffic management departments can capture and process traffic violations more accurately and efficiently, thereby improving the scientific and intelligent level of road traffic management.

HD imaging technology not only improves the accuracy and efficiency of traffic violation monitoring [6, 9] but also plays an important role in judicial evidence collection and violation determination [10]. Through HD image recording and analysis of traffic violations, accurate and reliable evidence can be provided to judicial departments, greatly enhancing the accuracy of violation determination and the fairness of judicial decisions. At the same time, the application of this technology also helps to enhance public legal awareness and traffic safety awareness [11, 12], thereby reducing the occurrence of traffic

violations and improving the level of urban traffic management and road safety.

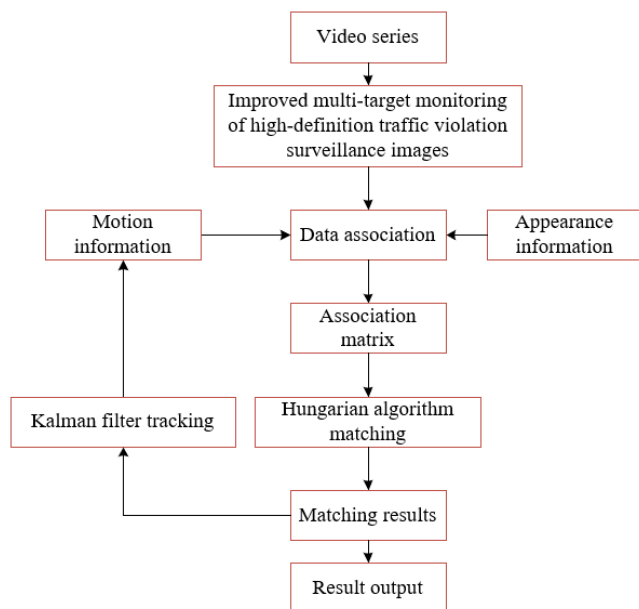
Although a large number of studies have focused on the development and optimization of traffic violation monitoring systems [13-15], the current technical solutions still have certain deficiencies in multi-object tracking and judicial evidence chain construction [16]. Existing multi-object tracking algorithms, such as the traditional Kalman filter and simple deep learning methods [17], often perform poorly in complex traffic environments and are susceptible to occlusion and target loss [18]. In addition, most existing judicial evidence chain construction methods rely on single-image processing technology and lack integrated approaches combining multiple techniques, which limits the reliability and completeness of evidence.

This study addresses the above issues and proposes two main research topics: (1) multi-object tracking in HD traffic violation monitoring images based on an improved DeepSORT algorithm and (2) judicial evidence chain construction for HD traffic violation monitoring images based on DWT-SVD. In terms of multi-object tracking, the improved DeepSORT algorithm enhances tracking accuracy and robustness in complex traffic environments by integrating

deep learning with traditional tracking methods. In terms of judicial evidence chain construction, DWT and SVD techniques are utilized to achieve multi-level analysis and evidence extraction of HD monitoring images, thereby enhancing the comprehensiveness and reliability of the evidence. This study not only enriches the theoretical research on image processing technology in traffic violation monitoring and judicial applications but also provides important reference value and practical significance for real-world applications.

## 2. MULTI-TARGET TRACKING OF TRAFFIC VIOLATION HD SURVEILLANCE IMAGES BASED ON IMPROVED DEEPSORT ALGORITHM

In the study of multi-target tracking of traffic violation HD surveillance images, it is first necessary to extract features specific to traffic scenes. HD surveillance images often contain complex backgrounds and diverse targets. Therefore, in the feature extraction stage, multiple visual attributes are comprehensively utilized to capture various characteristics of the targets. Specifically, grayscale histograms, raw grayscale values, and haar-like features are used to describe the grayscale features of the targets; color histograms are used to characterize the color features of the targets; and boundary features are extracted based on image edge changes. In addition, Scale Feature Transformation (SFT) and Histogram of Oriented Gradients (HOG) are employed to describe the gradient features of the targets, better handling the effects of complex backgrounds and lighting variations. The comprehensive utilization of these features effectively improves the accuracy of target recognition and tracking in traffic violations.



**Figure 1.** Multi-target tracking algorithm process of traffic violation HD surveillance images

After completing feature extraction, this paper proposes a multi-target tracking process design based on the improved DeepSORT algorithm. The improved DeepSORT algorithm combines the powerful feature extraction capability of deep learning with the efficiency of traditional tracking methods. By comparing the confidence levels of predicted bounding

boxes, the prediction with the highest confidence is selected and output to the detector, thereby solving the problem of re-tracking targets after occlusion. To achieve continuous tracking of targets, energy minimization and probability-based data association techniques are adopted, and appearance models and motion models are combined to predict the position of targets in the next frame. The Hungarian algorithm is used for optimal matching, effectively updating the target's position information and ensuring the continuity and accuracy of tracking. Moreover, to cope with noise interference in complex traffic environments, events are analyzed, and noise interference is denoised to obtain continuous motion trajectories, ensuring tracking stability and reliability. The algorithm process is shown in Figure 1.

In the proposed multi-target tracking process design for traffic violation HD surveillance images, the fundamental idea of introducing the Hungarian algorithm is to effectively solve the target matching and tracking problem, thereby achieving efficient and accurate multi-target tracking. When introducing the Hungarian algorithm in the context of traffic violation HD surveillance images, this paper mainly reflects the following four fundamental ideas:

(1) Constructing a Bipartite Graph Model: In multi-target tracking of traffic violations, constructing a bipartite graph is a critical step. Here, the left vertices of the graph represent all detected targets in the current frame, while the right vertices represent targets tracked in the previous frame. The edge weights represent the matching cost or similarity between targets detected in the current frame and targets tracked in the previous frame. The matching cost can be based on various features, such as position, speed, and appearance features. In this way, the multi-target tracking problem is formalized as a graph matching problem, facilitating the use of the Hungarian algorithm for solving it.

(2) Initializing the Matching Relationship: After constructing the bipartite graph, the initial matching relationship is set to an all-zero state, meaning that none of the detected targets in the current frame are assigned to any tracked targets from the previous frame. In practical operations, this means that initially, each detected target is not matched with any tracked target. This approach prepares for subsequent maximum matching searches and matching improvements.

(3) Finding the Maximum Matching: After setting the initial matching relationship, the path-finding mechanism of the Hungarian algorithm is used to find a maximum matching. The goal of maximum matching is to find a set of matchings that minimizes the total matching cost. Specifically, by calculating the matching cost between each detected target in the current frame and each tracked target in the previous frame, the matching relationship that results in the minimum total cost is determined.

(4) Adjusting and Optimizing the Matching: If the obtained maximum matching does not meet certain conditions, an augmenting path is used to improve the matching relationship. An augmenting path is an alternating path that, by finding unmatched vertices in the graph and adjusting the current matching relationship, can further reduce the total matching cost or improve matching accuracy. This process is particularly important in traffic violation surveillance scenarios because real-world environments may involve occlusions and target loss, requiring continuous adjustment and optimization of matching relationships to ensure continuity and accuracy in target tracking.

In the design of the multi-target tracking process for traffic

violation HD surveillance images, this paper introduces the YOLOv8 framework for two main reasons. First, traffic violation surveillance scenarios typically have extremely high real-time requirements. Traditional object detection methods often require long processing times, whereas YOLOv8 treats object detection as an end-to-end regression problem, allowing it to simultaneously predict object categories and bounding boxes within a single neural network, thereby achieving very high detection speed. Second, the types of targets in traffic violation surveillance scenarios are diverse and complex, including different types of vehicles, pedestrians, bicycles, etc., and target sizes and shapes vary significantly. YOLOv8 enhances the model's feature extraction capability by using the C2f module as the backbone, enabling it to better handle object detection tasks in different scales and complex backgrounds. Additionally, YOLOv8 introduces multi-scale training and testing strategies, making it adaptable to object detection requirements of different sizes, further improving its applicability and detection performance in traffic violation surveillance scenarios.

In the design of the detection head, YOLOv8 adopts a combination of an anchor-free mechanism and a Decoupled-head structure. Traditional anchor-based mechanisms often

require meticulous anchor design and incur high computational costs when dealing with diverse traffic targets. In contrast, the anchor-free mechanism directly predicts object center points and bounding boxes, simplifying model design and training while reducing computational complexity. The Decoupled-head structure separates classification and regression tasks, optimizing detection accuracy and robustness. This design provides more precise and stable detection results for complex object detection tasks in traffic violation surveillance scenarios.

Regarding the loss function, YOLOv8 employs a combination of classification Binary Cross Entropy (BCE), regression Complete Intersection over Union (CIoU), and Varifocal Loss (VFL) to further improve detection accuracy. Classification BCE effectively distinguishes different types of traffic participants when handling multi-class objects. Regression CIoU loss function considers not only the overlap of object bounding boxes but also the center distance and aspect ratio difference, which is particularly important for accurately locating object bounding boxes in traffic scenes. VFL dynamically adjusts the weights of positive and negative samples, further optimizing model detection performance. Figure 2 presents the YOLOv8 network architecture diagram.

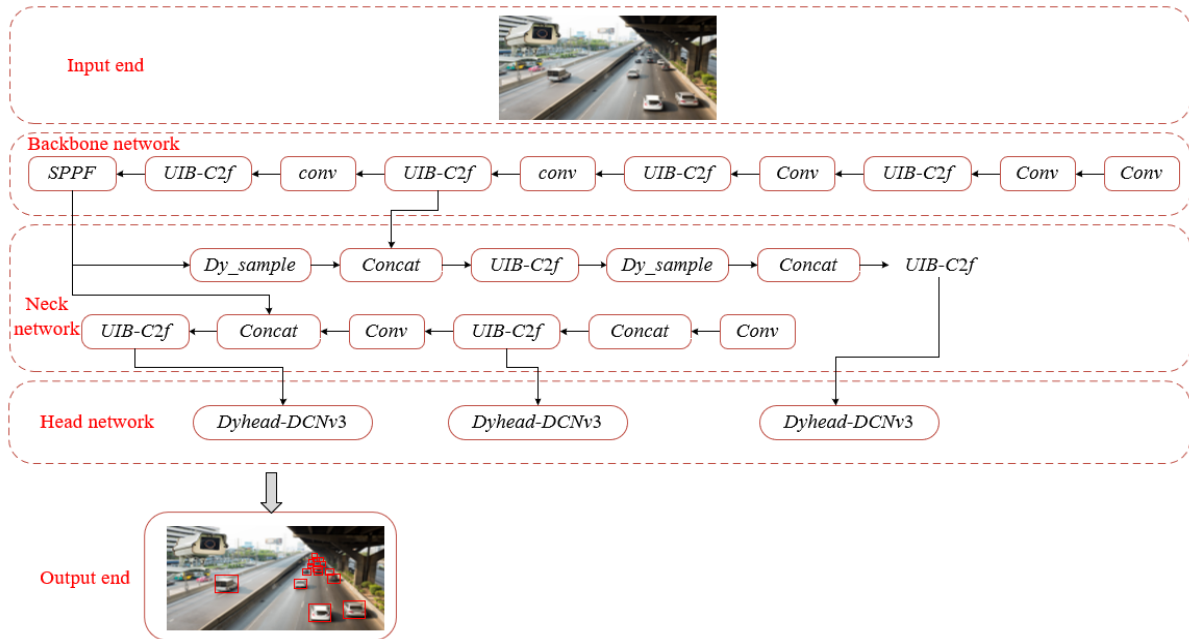


Figure 2. YOLOv8 network architecture diagram

YOLOv8 also introduces the Task-Aligned Assigner mechanism in the box matching strategy. This mechanism dynamically adjusts the allocation ratio of positive and negative samples and uses classification and regression scores as weights to select positive samples. This dynamic matching strategy can adaptively adjust according to real-time detection results, ensuring that suitable positive samples can be selected for training and detection in different traffic scenarios, thereby improving the model's generalization ability and detection accuracy. Suppose the allocation ratio of samples is represented by  $s$ , the predicted score corresponding to the annotated category is represented by  $t$ , the intersection over union (IoU) of the predicted box and the ground truth box is represented by  $w$ , and the weight hyperparameters are represented by  $\beta$  and  $\alpha$ . The allocation ratio formula of the samples is:

$$s = t^\beta + w^\alpha \quad (1)$$

Suppose the probability that the predicted sample belongs to class 1 is represented by  $o$ , and the modulation factor is represented by  $\varepsilon$ . The global classification loss  $N(o, w)$  is calculated as follows:

$$N(o, w) = \begin{cases} -w(w \cdot \log o + (1 - w) \cdot \log(1 - o)), & w > 0 \\ -\beta \cdot o^\varepsilon \cdot \log(1 - o), & w = 0 \end{cases} \quad (2)$$

The multi-target tracking task in traffic violation scenarios presents the following challenges: high-speed target movement, frequent occlusions, target appearance diversity, and the need for real-time processing. In the design of the multi-target tracking process for traffic violation HD

surveillance images, this paper introduces the DeepSORT algorithm. The DeepSORT algorithm generates initial detection results based on YOLOv8. Targets in traffic violation surveillance scenarios typically include various types of vehicles and pedestrians, which may exhibit significant appearance differences at different times and locations. Figure 3 shows the DeepSORT process schematic diagram. In specific implementation steps, the DeepSORT algorithm mainly includes the following six aspects:

(1) Initialize trajectories and Kalman filter: Based on the initial detection results of each frame, a trajectory is generated for each detected target, and a Kalman filter is initialized for each trajectory. The Kalman filter is used to predict the position of the target in the next frame, taking into account the target's motion parameters such as velocity and acceleration. This initialization process ensures that each target has a unique identifier, facilitating subsequent tracking.

(2) Predict target position: The Kalman filter is used to predict each trajectory, generating the predicted bounding box for the target in the next frame. This step considers the motion state of the target and can, to some extent, compensate for rapid target movement, making the tracking process more stable and continuous.

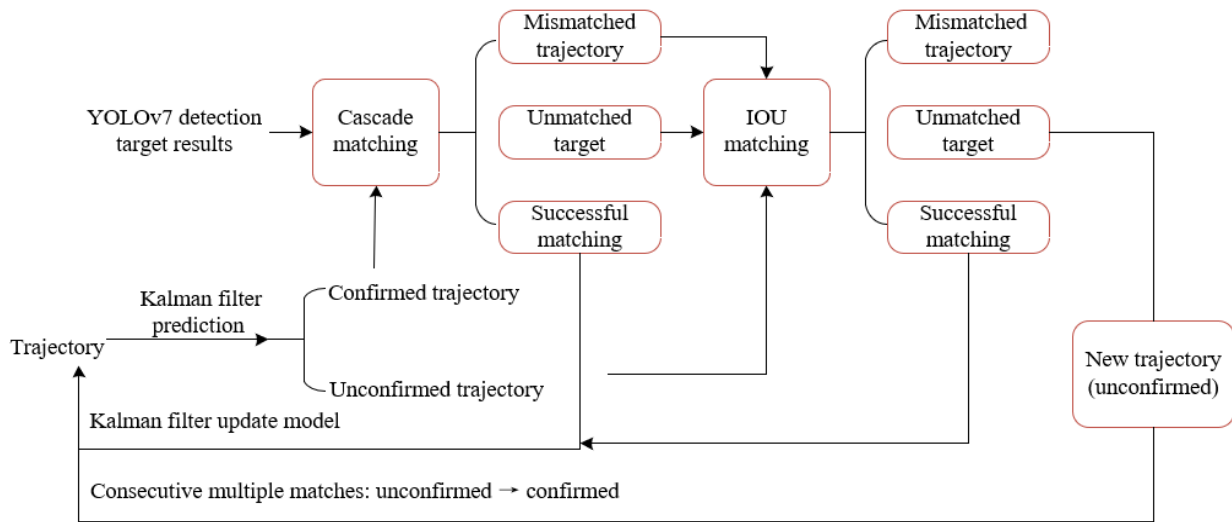
(3) Data association and matching: By comparing the detection results of the current frame with the predicted results of the previous frame, the IoU between them is calculated, and a cost matrix is constructed based on this. This cost matrix reflects the matching relationship between the targets detected in the current frame and those predicted in the previous frame.

In traffic violation surveillance scenarios, targets often experience occlusions and transformations. The DeepSORT algorithm can effectively associate targets across different frames through this approach.

(4) Linear matching using the Hungarian algorithm: The constructed cost matrix is input into the Hungarian algorithm to perform optimal matching. The Hungarian algorithm is a classical combinatorial optimization algorithm that can find the optimal matching scheme in multi-target tracking, ensuring that each target is correctly tracked.

(5) Handling unmatched targets and trajectories: Unmatched detection results and trajectories are processed separately. For unmatched detection results, they may be newly appearing targets, requiring the initialization of new trajectories and Kalman filters. For unmatched trajectories, their status needs to be updated to determine whether they have truly disappeared or are only temporarily occluded. In traffic violation surveillance scenarios, this step is particularly critical because the frequent appearance and disappearance of targets are common phenomena.

(6) Update trajectory information and perform cascade matching: After processing each frame, the Kalman filter is used to update the confirmed trajectories, and cascade matching is performed on the detection results based on appearance features and motion information. Here, the DeepSORT algorithm introduces a deep feature extraction network to further improve the accuracy and robustness of target matching.



**Figure 3.** DeepSORT process schematic diagram

The multi-target tracking task in traffic violation scenarios has certain particularities and complexities. The monitored images contain a wide variety of targets, including various types of vehicles and pedestrians, with significant differences in their sizes and shapes. In addition, the problems of high-speed movement within the monitoring range and frequent occlusions impose high requirements on the real-time performance and accuracy of detection and tracking. Although YOLOv8 has high detection speed and good detection accuracy, it still has some shortcomings in detecting large targets. To improve the overall efficiency and accuracy of the system, we consider optimizing the DeepSORT algorithm, particularly by improving its target detection component. This paper introduces an improved K-Means nearest neighbor

algorithm to process YOLOv8 bounding boxes to address the current issues of detection speed and accuracy.

The traditional K-Means algorithm uses Euclidean distance as the similarity measurement criterion during clustering. However, Euclidean distance is susceptible to interference when processing large bounding boxes, leading to suboptimal clustering results. Suppose the distance between a certain bounding box  $y$  and the cluster center  $z$  is represented by  $f(y,z)$ , and the IoU between the bounding box and the cluster center is represented by  $U(y,z)$ , then the Euclidean distance measurement formula is:

$$f(y,z) = 1 - U(y,z) \quad (3)$$

By improving the K-Means nearest neighbor clustering method, the matching between anchor boxes and actual bounding boxes is enhanced, thereby improving the IoU value. The specific optimization steps are as follows:

(1) Data preprocessing stage: A large amount of target bounding box data from traffic violation HD surveillance images is collected. These data contain the actual bounding box positions and category information of the targets, denoted as  $(a_k, b_k, q_k, g_k)$ , where  $(a_k, b_k)$  represents the center coordinates of the bounding box,  $q_k$  and  $g_k$  represent the width and height of the bounding box, respectively, and  $V$  is the number of target bounding boxes.

(2) Initializing  $J$  cluster centers: Each cluster center contains a width and height  $(Q_k, G_k)$ . The initialization of these cluster centers is based on a preliminary analysis of the collected data to ensure that the cluster centers can represent the characteristics of different target categories. The purpose of this step is to determine characteristic anchor boxes for different target categories through clustering, thereby optimizing the detection process.

(3) Clustering process: The improved distance formula is used to calculate the distance between each actual target bounding box and each cluster center. Through this method, each target bounding box is assigned to the cluster center with the smallest distance, thereby improving the accuracy of clustering. Suppose the distance between two bounding boxes is represented by  $F_1$ , then the calculation formula is:

$$F_1 = 1 - U((a_k, b_k, q_k, g_k), (a_k, b_k, Q_k, G_k)), \quad (4)$$

$$k \in \{1, 2, \dots, V\}, u \in \{1, 2, \dots, J\}$$

(4) Recalculating cluster centers during each iteration of the clustering process: During the update process, the specific characteristics of the targets within each cluster are considered, and the cluster center positions are updated using a weighted average method. This step ensures that the cluster centers can dynamically adapt to data changes, thereby continuously improving the matching between anchor boxes and actual bounding boxes. Suppose the updated width of the  $u$ -th cluster center is represented by  $q'_u$ , the number of actual target bounding boxes in the  $u$ -th cluster is represented by  $V_u$ , the updated height of the  $u$ -th cluster center is represented by  $G'_u$ , the width assigned to the  $u$ -th cluster center is represented by  $q_u$ , and the height of all bounding boxes assigned to the  $u$ -th cluster center is represented by  $g_u$ , then the updated width and height of the  $u$ -th cluster center are:

$$Q'_u = 1 / (V_u \sum q_u), G'_u = 1 / (V_u \sum g_u) \quad (5)$$

(5) When the number of iterations reaches the preset value or the cluster center changes are no longer significant, the clustering process ends. At this point, optimized anchor boxes are obtained, which can still match actual bounding boxes well under low-light conditions. For practical applications in multi-target tracking in traffic violation HD surveillance images, we improve the association mechanism of the DeepSORT algorithm. Specifically, we introduce a dynamic allocation mechanism that adjusts the weights of appearance features and motion information in real time. This improvement takes into account the detection instability caused by lighting changes. Through dynamic adjustment, more accurate associations of detected bounding boxes can be achieved under low-light

conditions. To this end, we define a matching variable  $s(l)$  to evaluate the matching degree between the currently detected bounding box  $l$  and the reference bounding box. This variable comprehensively considers the IoU value of the bounding box, appearance features, and motion information, and dynamically adjusts the weight of each part according to the actual lighting conditions. Suppose the areas of the target and occlusion detection bounding boxes are represented by  $T_x$  and  $T_y$ , and the currently detected bounding box number is represented by  $l$ , then the matching variable  $s(l)$  is calculated as follows:

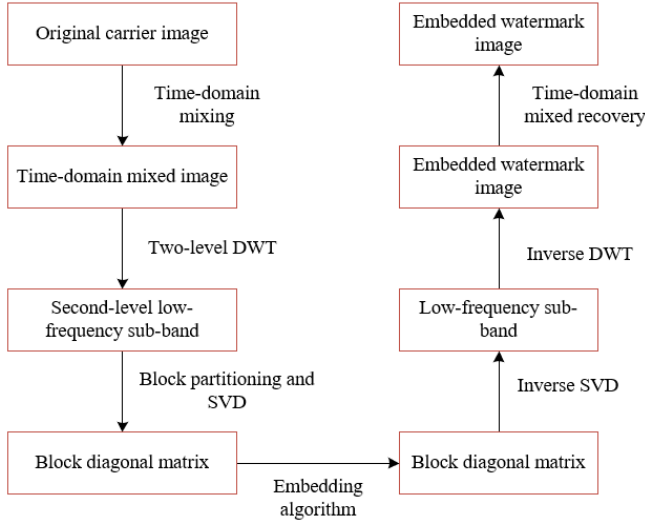
$$s(l) = \begin{cases} 0, & \frac{T_x \cap T_y}{T_y} < \frac{1}{5} \\ \text{MIN} \left\{ 1, 3 \times \frac{T_x \cap T_y}{T_y} \right\}, & \frac{T_x \cap T_y}{T_y} \geq \frac{1}{5} \end{cases} \quad (6)$$

### 3. CONSTRUCTION OF JUDICIAL EVIDENCE CHAIN FOR HD TRAFFIC VIOLATION SURVEILLANCE IMAGES BASED ON DWT-SVD

After completing the multi-object tracking study of HD traffic violation surveillance images, this paper further conducts research on the construction of a judicial evidence chain for HD traffic violation surveillance images. From a legal perspective, the collection of evidence for traffic violations must have a high degree of authenticity and be tamper-proof to ensure its legality and validity in judicial procedures. As an intuitive form of evidence, the completeness, authenticity, and verifiability of HD surveillance images are key factors for their acceptance in court. Specifically, this paper introduces the DWT-SVD algorithm, which decomposes the image into different frequency subbands through wavelet decomposition and then embeds watermark information into these subbands. Since SVD has strong anti-attack capability and robustness, this method ensures that even if the image undergoes common operations such as compression, cropping, or rotation, the embedded watermark information can still be effectively extracted and verified. This technical approach provides strong evidence assurance, ensuring that the image maintains its legality and validity in judicial proceedings.

To further achieve the construction of the judicial evidence chain, it is necessary to establish a systematic evidence management and tracking mechanism based on image watermark embedding. The specific approach includes the following: First, by embedding a digital watermark containing information such as time, location, and device ID into HD surveillance images, each frame of the image is ensured to have a unique identifier. This information can be automatically generated and embedded at the time of image acquisition to ensure the integrity of the original data. Second, an evidence management system should be established to record and manage the entire process of collecting, transmitting, storing, and using each traffic violation surveillance image. This system should have strong logging capabilities to track every operation and access, ensuring that each step is traceable. Finally, in judicial trials, the authenticity of images is verified by extracting the watermark information embedded in the images, and the entire evidence chain is systematically presented to the court to support judicial decisions. Figure 4 illustrates the construction process of the judicial evidence chain for HD traffic violation surveillance images.





**Figure 4.** Judicial evidence chain construction process for HD traffic violation surveillance images

The optimal signal for HD traffic violation surveillance images should possess the following characteristics: good randomness, long period, high linear complexity, and balance. These characteristics ensure that the signal, when embedded in the surveillance image, can maintain image quality while providing a highly reliable verification mechanism. Specifically, good randomness prevents watermark information from being easily predicted or tampered with; a long period ensures the reusability and durability of the signal; high linear complexity enhances the signal's resistance to attacks; and balance ensures that the embedded signal does not cause significant visual impact on the image. These characteristics work together to ensure that the embedded watermark signal can still be effectively extracted and verified under various common attacks, thereby safeguarding the integrity and authenticity of the image.

The optimal binary array is a typical form of signal. In traffic violation surveillance, to ensure that image signals can serve as effective judicial evidence, the array design of the transmitting and receiving ends must meet specific application site requirements. Even if their specific array structures are different, as long as certain conditions are met, these array pairs can still achieve optimal performance. In an array pair, the cross-correlation function of the two arrays is defined as the autocorrelation function. This method allows for the precise design and optimization of various forms of new optimal HD traffic violation surveillance image signals. In practical applications, this means that attention must not only be paid to the performance of a single array but also to the cooperative operation between the transmitting and receiving arrays to ensure complete transmission and accurate reception of signals. Specifically, let  $A = [a(t_1, t_2, \dots, t_v)]$  and  $B = [b(t_1, t_2, \dots, t_v)]$  be two  $v$ -dimensional arrays of order  $V_1 \times V_2 \times \dots \times V_v$ , where  $0 \leq t_u \leq V_u - 1$  and  $1 \leq u \leq v$ . The cyclic cross-correlation function of arrays  $A$  and  $B$  is given by:

$$E_{A,B}(\pi_1, \pi_2, \dots, \pi_v) = \sum_{a_1=0}^{V_1-1} \sum_{a_2=0}^{V_2-1} \dots \sum_{a_v=0}^{V_v-1} a(t_1, t_2, \dots, t_v) \cdot b^*(t_1 + \pi_1, \dots, t_v + \pi_v) \quad (7)$$

In the above equation,  $t_u + \pi_u = (t_u + \pi_u) \text{ MOD } V_u$ . If  $A=B$ , then  $E_{A,B}(\pi_1, \pi_2, \dots, \pi_v)$  is called the cyclic autocorrelation function

of array  $A$ , i.e., the autocorrelation function of array  $A$ . If this function satisfies the following equation, then array  $A$  is called an optimal  $v$ -dimensional binary array of order  $V_1 \times V_2 \times \dots \times V_v$ :

$$E_{A,A}(\pi_1, \pi_2, \dots, \pi_v) = \begin{cases} R \neq 0, (\pi_1, \pi_2, \dots, \pi_v) = (0, 0, \dots, 0) \\ 0, (\pi_1, \pi_2, \dots, \pi_v) = (0, 0, \dots, 0) \end{cases} \quad (8)$$

Furthermore, let  $A = [a(t_1, t_2, \dots, t_v)]$  and  $B = [b(t_1, t_2, \dots, t_v)]$  be two  $v$ -dimensional arrays of order  $V_1 \times V_2 \times \dots \times V_v$ , where  $0 \leq t_u \leq V_u - 1$  and  $1 \leq u \leq v$ . If they form  $v$ -dimensional array pair, denoted as  $(A, B)$ , then  $N = V_1 V_2 \dots V_v$  is called the volume of the array pair. If the elements of  $A$  and  $B$  take values of  $\pm 1$ , then  $(A, B)$  is called a  $v$ -dimensional binary array pair. A one-dimensional array pair is called a sequence pair. Let  $A$  and  $B$  be two matrices of order  $L \times V$ , where the elements of  $A$  and  $B$  are  $\pm 1$ . If they satisfy the equation  $AB^T = zU_L$ , where  $U_L$  is the identity matrix of order  $L$ , and  $B^T$  is the transpose of matrix  $B$ , then  $(A, B)$  is called a binary orthogonal matrix pair.

DWT, as a multi-resolution analysis tool, is selected as the basis for watermark embedding due to its conformity to human visual characteristics and good frequency domain decomposition capability. Through DWT, an image can be decomposed into sub-images of different frequency bands, achieving fine analysis of image content. A one-level wavelet decomposition divides the image into one low-frequency region and three high-frequency regions. The low-frequency region contains the main energy and important features of the image, while the high-frequency regions mainly contain the image's edge and texture details. A further two-level wavelet decomposition decomposes the low-frequency region of the one-level decomposition again, obtaining a new low-frequency region and multiple high-frequency regions. Embedding the watermark in the low-frequency region can enhance the robustness of the watermark, ensuring that it can still be effectively extracted and verified after various common image processing operations, although this may affect the watermark's invisibility. Based on these considerations, this paper adopts the Haar wavelet basis for wavelet decomposition, as it is simple, effective, and suitable for image watermark embedding.

Let  $X \in R^{l \times v}$ , where  $R^{l \times v}$  is the  $l \times v$  real field. Assume that the orthogonal matrix is represented by  $I$  and  $N$ , and a matrix with all off-diagonal elements being 0 is represented by  $T = \text{diag}(\delta_1, \delta_2, \dots, \delta_e)$ , where  $e = \min(l, v)$ . Then the SVD of matrix  $X$  is as follows:

$$X = ITN^S \quad (9)$$

Let  $\delta_u$  be the singular value of the matrix, and its diagonal elements satisfy:

$$\delta_1 \geq \delta_2 \geq \dots \geq \delta_e \geq 0 \quad (10)$$

Below are the detailed steps for watermark embedding based on the DWT-SVD method:

(1) The Lena host image  $x_0$  is divided into blocks of size  $(v \times v)$ , with each block having a size of  $(u, k = l/v)$ . Then, each block is multiplied by the signal  $L_x$ , achieving the time-domain mixing of the host image. This step ensures the integration of the host image and the traffic monitoring image signal, forming a time-domain mixed image of the sub-blocks, and combining them into a new image  $Z$  of size  $L \times L$ .

(2) The Haar wavelet basis is applied to perform a two-level

DWT on the combined image  $Z$ , resulting in six high-frequency sub-bands and one low-frequency sub-band  $X_2$ , with the size of the low-frequency sub-band being  $L/4 \times L/4$ . This step utilizes the multi-resolution analysis advantage of the wavelet transform, decomposing the image into different frequency components, which provides the foundation for subsequent watermark embedding.

(3) Low-frequency sub-band blocking: Since the watermark image selected is a binary image of size  $l \times l$ , the low-frequency sub-band  $X_2$  is divided into blocks of size  $Y = \text{round}(L/4l)$ , resulting in  $l \times l$  sub-blocks  $Y_{uk}$ , where  $u, k = 1, 2, \dots, l$ . This step ensures that the block division of the low-frequency sub-band matches the size of the watermark image, preparing for the next step of SVD.

(4) Perform SVD on each sub-block  $Y_{uk}$  to obtain three matrices:  $I$ ,  $T$ , and  $N$ . The diagonal elements of each matrix  $T$  are sorted in descending order to ensure that the first diagonal element  $\delta_{uk}$  of each  $T$  matrix is the maximum value. The use of SVD is to extract and process important image features, providing a stable carrier for embedding the watermark information.

(5) Watermark embedding: Set the embedding strength factor  $w$ , and let  $c = \text{MOD}(\delta_{uk}, w)$ . According to the set rules, embed the watermark into the first element  $\delta_{uk}$  of each  $T$  matrix. The specific rule is determined based on actual requirements and algorithm design. By adjusting the embedding strength factor, the robustness and concealment of the watermark can be effectively controlled. Let MARK be the watermark signal to be embedded, then:

$$\delta'_{uk} = \begin{cases} \delta_{uk} - c - w/4, \text{MARK} = 1 \text{ and } c < w/4 \\ \delta_{uk} - c + 3w/4, \text{MARK} = 1 \text{ and } c \geq w/4 \\ \delta_{uk} - c + w/4, \text{MARK} = 0 \text{ and } c < 3w/4 \\ \delta_{uk} - c + 5w/4, \text{MARK} = 0 \text{ and } c \geq 3w/4 \end{cases} \quad (11)$$

(6) Inverse SVD and Inverse Wavelet Transform: Perform the inverse operation on the matrix  $T$  formed by the embedded watermark  $\delta'_{uk}$ , i.e.,  $Y'_{uk} = ITN^T$ . Then, recombine the sub-blocks, and perform an inverse wavelet transform on the combined matrix to obtain the watermark-embedded mixed image  $x_1$ . This step converts the image with the embedded watermark back to the time-domain form, ensuring the integrity and visibility of the image.

(7) Time-domain mixing recovery: Divide the mixed image  $x_1$  into blocks according to the size of the optimal traffic violation HD surveillance image signal  $L_x$ , and multiply each block by  $L_x$  again to achieve the time-domain mixing recovery of the host image, resulting in the final mixed image  $Z_2$  with the embedded watermark. This step ensures that the image with the embedded watermark is integrated with the original monitoring signal, forming an image evidence with legal validity.

Below is the detailed description of the watermark extraction process based on the DWT-SVD method:

(1) First, divide the potentially attacked watermark-embedded image  $L_1$  (of size  $L \times L$ ) into blocks according to the size of the optimal traffic violation HD surveillance image signal  $L_x$  (of size  $v \times v$ ). Assume that the size of each sub-block is  $L/v \times L/v$ . Then, multiply each sub-block by the signal  $L_x$  for time-domain mixing. This step ensures that the local features of the embedded watermark image remain consistent with the original image through integration with the original monitoring signal, forming a time-domain mixed image of the

sub-blocks and recombining them into a new image  $Z_1$  of size  $L \times L$ . This processing method guarantees that even if the image is attacked, its local features maintain a certain internal relationship.

(2) Next, apply the Haar wavelet basis to perform a two-level DWT on the combined image  $Z_1$ , resulting in multiple high-frequency sub-bands and one low-frequency sub-band  $X_4$ . The size of the low-frequency sub-band is  $L/4 \times L/4$ . Divide the low-frequency sub-band  $X_4$  into blocks of size  $Y_{uk} = \text{round}(L/4l)$ , resulting in  $l \times l$  sub-blocks  $Y_{uk}$ . Then, perform SVD on each sub-block  $Y_{uk}$ , i.e.,  $Y_{uk} = ITN^T$ , obtaining the singular value matrix  $T$ . This step utilizes the multi-resolution analysis advantage of the wavelet transform, decomposing the image into different frequency components and extracting the image's important features through SVD, providing a stable carrier for the subsequent watermark extraction.

(3) In the singular value matrix  $T$  of each sub-block, find the maximum value  $\delta_{uk}$  on the diagonal. According to the set embedding strength factor  $w$ , calculate  $c = \text{MOD}(\delta_{uk}, w)$ . Based on the pre-established rules, organize the extracted information into the watermark image. This process, through precise calculation and processing of the singular values, can effectively extract the embedded watermark information from the attacked image, thereby verifying whether the image has been tampered with. This step is particularly crucial because it directly determines the accuracy and robustness of the watermark extraction. Specifically, let the extracted watermark information be represented by  $q$ , and the watermark extraction rule is:

$$q(u, k) = \begin{cases} 1, c \geq w/2 \\ 0, \text{others} \end{cases} \quad (12)$$

(4) Perform quality evaluation for both the watermark-embedded image and the extracted watermark image. The watermark-embedded image is objectively evaluated using PSNR to reflect the algorithm's performance in terms of invisibility. A higher PSNR value indicates better image quality after watermark embedding, with the watermark being harder to detect. The extracted watermark image is objectively evaluated using NC, which is the cross-correlation coefficient between the extracted watermark and the original watermark. The closer the NC value is to 1, the higher the similarity between the extracted watermark and the original watermark, and the stronger the robustness of the watermark. These two-evaluation metrics comprehensively verify the effectiveness of the watermark algorithm and the image's integrity. The PSNR calculation formula is:

$$PSNR = 10 \log_{10} \frac{LVMAX[ MAX(U^2) ]}{\sum_u \sum_k [U(u, k) - U'(u, k)]^2} \quad (13)$$

The NC calculation formula for the cross-correlation coefficient between the extracted watermark and the original watermark is as follows:

$$VZ = \frac{\sum_u \sum_k Q_1(u, k) Q_2(u, k)}{\sum_u \sum_k Q_1(u, k)^2} \quad (14)$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

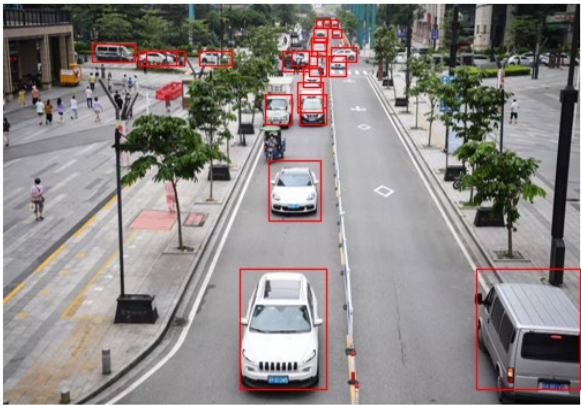
By performing detection and tracking on videos from the test set, Figure 5 demonstrates the multi-target detection effect of the improved DeepSORT algorithm in a complex traffic environment. Specifically, the experimental results show that the improved algorithm can accurately identify and detect multiple traffic targets in the video, significantly improving detection accuracy. This result validates the effectiveness and applicability of the improved algorithm in handling traffic violation HD surveillance images, especially when dealing with multiple targets such as vehicles and pedestrians, while still maintaining efficient detection performance. Further analysis of the results in Figure 6 shows that under a tracking rate of 15.5 frames/s, the improved DeepSORT algorithm achieves rapid tracking of traffic targets. This not only indicates that the algorithm has high application potential in terms of real-time performance but also demonstrates its robustness in complex dynamic environments.

As shown in Figure 7, the results of the ablation experiment demonstrate that, in terms of tracking performance for different types of targets, the proposed method exhibits significant advantages compared to the other three methods. Taking motor vehicles, non-motor vehicles, and pedestrians as examples, the P-R curve of the proposed method shows

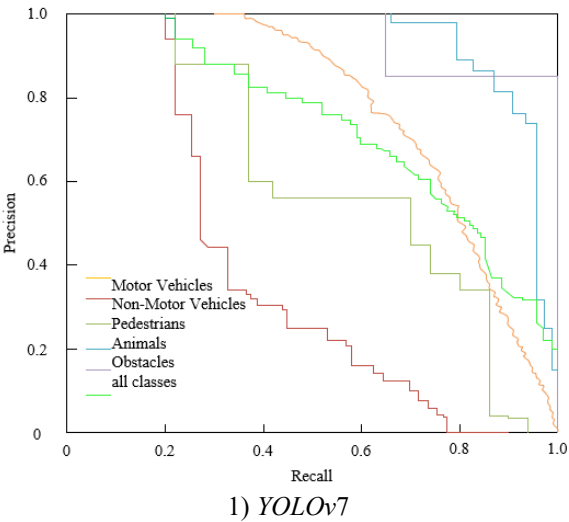
generally higher precision at the same recall rate. Especially when the recall rate approaches 1.0, the precision remains high, indicating stronger tracking stability across various targets. In contrast, methods such as YOLOv7 and YOLOv8 show a more pronounced drop in precision as recall increases. While the YOLOv8 + traditional K-Means method achieves some optimization, it still fails to balance tracking precision and recall for complex targets, falling short of the proposed method. The experimental results show that the proposed method effectively optimizes the multi-object tracking process by introducing an improved K-Means neighbor algorithm for YOLOv8 bounding box handling, in combination with the improved DeepSORT algorithm. The overall performance in the ablation experiment suggests that this improvement strategy significantly enhances tracking performance across all categories, with a larger area under the P-R curve, indicating a better balance between precision and recall. In contrast, other comparison methods are limited by the traditional K-Means' shortcomings and the base algorithm's inadequacies in feature association, which makes precise tracking difficult in complex multi-object scenarios. This further confirms the effectiveness of the improvements in bounding box handling and tracking algorithm enhancements in the proposed method.



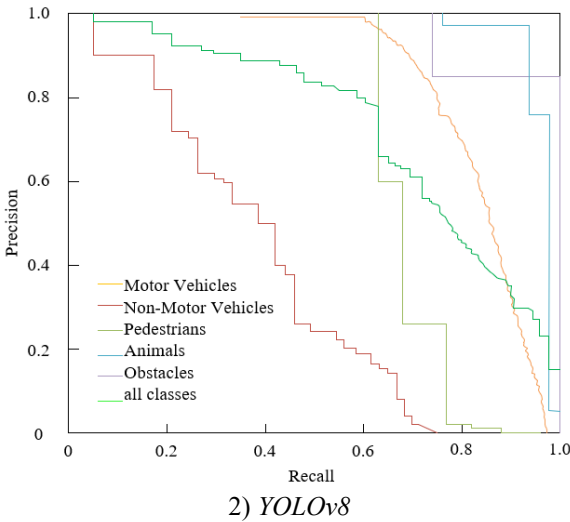
Figure 5. Video detection



(The numbers in the tracking boxes represent the IDs returned by DeepSORT)

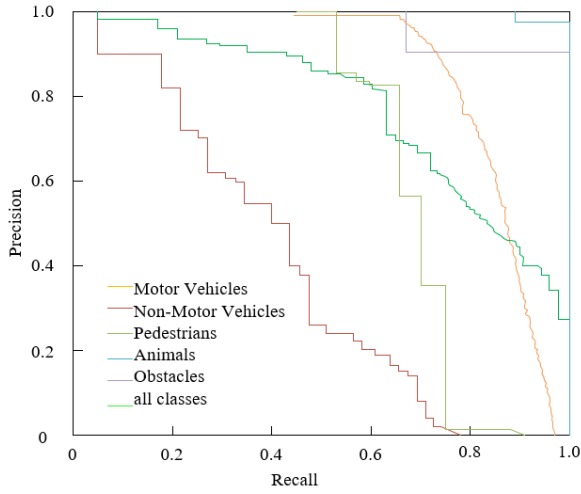


1) YOLOv7

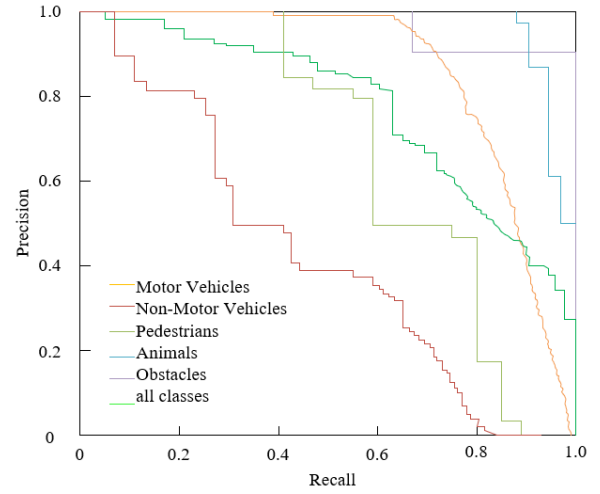


2) YOLOv8



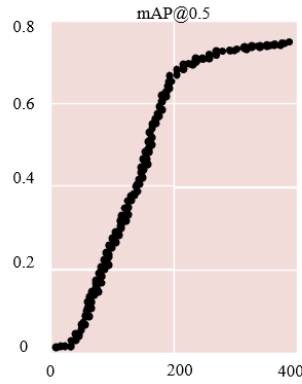


3) YOLOv8+ traditional K-Means

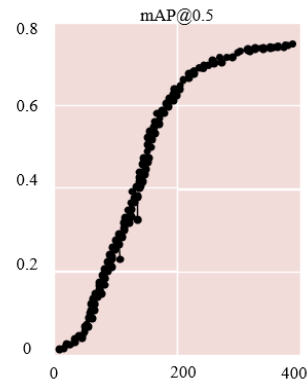


4) The Proposed Method

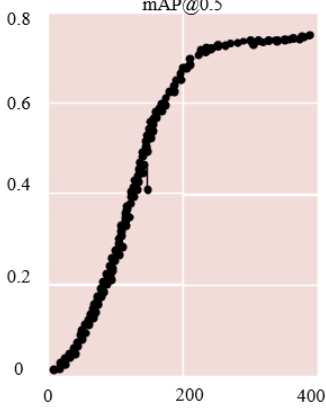
**Figure 7.** Ablation experiment results of the multi-object tracking method for traffic violation HD surveillance images (P-R Curve)



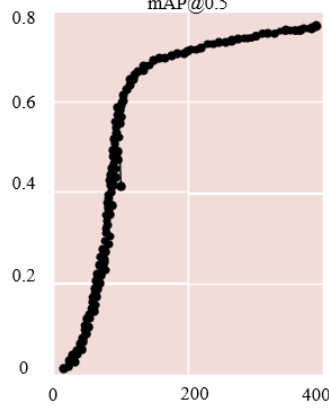
1) YOLOv7



2) YOLOv8



3) YOLOv8+ traditional K-Means



4) The Proposed Method

**Figure 8.** Training process of the multi-object tracking method for traffic violation HD surveillance images

As shown in Figure 8, the mAP@0.5 and mAP@0.8 curves of the training process demonstrate significant training advantages for the proposed method. In terms of convergence speed, the curve of the proposed method has a steeper upward slope, allowing it to quickly approach high mAP values in fewer iterations. In contrast, the curves for YOLOv7 and YOLOv8 rise more gradually, and even after 400 iterations, the final mAP values are still noticeably lower than those of the proposed method. Even YOLOv8 + traditional K-Means shows less improvement during training, highlighting the dual advantages of the proposed method in training efficiency and

performance. Both mAP@0.5 and mAP@0.8 metrics show that the proposed method achieves higher precision in a shorter training cycle, validating the optimization effect of the improvement strategy on model training. A deeper analysis of the ablation experiment results shows that the advantages of the proposed method stem from the synergistic effect of the improved modules. On one hand, the improved K-Means neighbor algorithm optimizes the quality of YOLOv8 bounding box generation, making the detection boxes more closely fit the target features and providing more precise input for subsequent tracking. On the other hand, the improved

DeepSORT algorithm enhances feature association and trajectory management in multi-object tracking, enabling the model to learn target motion patterns and appearance features more efficiently during training. In comparison, the base frameworks of YOLOv7 and YOLOv8 lack targeted optimization, and the limitations of traditional K-Means also restrict the improvement potential of the YOLOv8 +

traditional K-Means method. The dual improvements in bounding box handling and tracking algorithms of the proposed method significantly enhance both model training convergence speed and final performance, fully validating the effectiveness and complementarity of the improvements in the ablation experiment.

Table 1. PSNR and NC values of different watermarking techniques

	PSNR		NC	
	Traditional	Proposed	Traditional	Proposed
White Noise	45.2356	41.2356	1.1215	1.1021
Gaussian Lowpass	36.2584	35.2016	0.9856	1.021
JPEG Compression	41.2036	37.5269	0.9562	0.9752
Shear	37.2356	35.2016	0.9874	1.0012
Rotation	12.3205	12.3256	0.9236	0.9236
White Noise	12.2658	12.0236	0.9125	0.9125

Table 2. Correlation coefficients of extracted watermarks and original watermarked images under different optimal image signal forms

Best Image Signal Form		4th Order Hadamard Matrix	8th Order Hadamard Matrix	16th Order Hadamard Matrix	Best Binary Sequence Constructed 4th Matrix
White Noise	NC	0.9512	0.8845	0.9521	0.9541
Gaussian Lowpass	NC	0.9536	0.9326	0.8569	0.9452
JPEG Compression	NC	0.9684	0.8124	0.9841	0.9786
Shear	NC	0.9235	0.9236	0.9236	0.9235
Rotation	NC	0.7152	0.6659	0.9125	0.8562

Table 1 shows the comparison of Peak Signal-to-Noise Ratio (PSNR) and Normalized Correlation (NC) values for different watermarking techniques under various attack conditions, comparing traditional watermarking techniques with the proposed DWT-SVD-based watermarking method. The data indicates that although traditional watermarking techniques achieve higher PSNR values under white noise interference (45.2356) compared to the proposed method (41.2356), the NC values of the proposed method are generally higher under other attack conditions, especially under Gaussian low-pass (1.021) and JPEG compression (0.9752), suggesting superior watermark integrity and robustness. In shear and rotation attacks, the NC values of the proposed method are 1.0012 and 0.9236, respectively, showing better resistance to attacks. Even with slightly lower PSNR, the proposed method still effectively maintains watermark correlation and detectability. The experimental results confirm that the DWT-SVD-based method exhibits higher robustness and reliability in constructing the judicial evidence chain for HD surveillance images, ensuring that more comprehensive and reliable evidence can be extracted for traffic violation handling.

Table 2 displays the correlation coefficients (NC values) between the extracted watermark and the original watermarked image under different optimal image signal forms. From the data, it can be seen that under white noise attack, the best binary sequence constructed 4th matrix performs the best with a correlation coefficient of 0.9541, while the 16th-order Hadamard matrix also demonstrates high correlation (0.9521). Under Gaussian low-pass filter attack, the 4th-order Hadamard matrix (0.9536) and the best binary sequence constructed 4th matrix (0.9452) show better robustness. Under JPEG compression attack, the 16th-order Hadamard matrix achieves the highest correlation coefficient (0.9841), followed by the best binary sequence constructed 4th

matrix (0.9786), indicating superior performance under compression. In shear attack, all matrices show relatively consistent correlation coefficients (around 0.9235), demonstrating uniform resilience. However, under rotation attack, the 16th-order Hadamard matrix performs well with a correlation coefficient of 0.9125, while the 8th-order Hadamard matrix shows the lowest correlation coefficient (0.6659). The analysis suggests that different optimal image signal forms exhibit varying performances under different attack types, emphasizing the importance of selecting appropriate signal forms to enhance watermark robustness. The best binary sequence constructed 4th matrix performs excellently under multiple attack conditions, especially white noise and Gaussian low-pass, proving its superiority in maintaining watermark integrity. Meanwhile, the 16th-order Hadamard matrix excels in JPEG compression and rotation attacks, showcasing its advantage in handling specific attacks.

5. CONCLUSION

This paper studied the application of HD image technology in traffic violation monitoring and judicial processing, proposing two main research contributions: first, an improved DeepSORT algorithm for multi-object tracking in traffic violation HD surveillance images, and second, a judicial evidence chain construction based on DWT-SVD. In terms of multi-object tracking, the improved DeepSORT algorithm, combining deep learning and traditional tracking methods, significantly enhances tracking accuracy and robustness in complex traffic environments, addressing the tracking loss and misjudgment problems encountered by traditional methods. For judicial evidence chain construction, the use of DWT and SVD technologies allows for multi-layer analysis and evidence extraction from HD surveillance images, greatly

enhancing evidence comprehensiveness and reliability, enabling efficient and accurate evidence collection and identification for traffic violations.

Based on the research findings and results, the following conclusions can be drawn: through the improved DeepSORT algorithm and DWT-SVD technology, this paper effectively improves the performance of traffic violation monitoring systems and the reliability of judicial evidence, with significant practical application value. However, there are some limitations in the research process. First, the improved DeepSORT algorithm may still face challenges when handling extremely complex traffic environments. Second, the DWT-SVD technology requires substantial computational resources and processing time for evidence chain construction. Future research directions may include further optimization of the DeepSORT algorithm to handle more complex environments and scenarios while reducing computational resource consumption. Additionally, more efficient image processing and watermark extraction techniques can be explored to enhance the speed and precision of evidence chain construction, providing more comprehensive and efficient technical support for traffic violation monitoring and judicial processing.

## ACKNOWLEDGMENT

This paper was supported by Hebei Provincial Social Science Fund (Grant No.: HB23FX014).

## REFERENCES

- [1] Kilsztajn, S., Silva, C.R.L.D., Silva, D.F.D., Michelin, A.D.C., Carvalho, A.R.D., Ferraz, I.L.B. (2001). Mortality rate associated to traffic accidents and registered motor vehicles. *Revista de Saude Publica*, 35: 262-268. <https://doi.org/10.1590/S0034-89102001000300008>
- [2] Apelbaum, J., Li, Z., Hensher, D.A. (2011). A correction framework for improving the robustness of motor vehicle registration data: An Australian application. *Transportation Research Part D: Transport and Environment*, 16(7): 562-570. <https://doi.org/10.1016/j.trd.2011.06.004>
- [3] Adi, K., Widodo, C.E., Widodo, A.P., Masykur, F. (2024). Traffic violation detection system on two-wheel vehicles using convolutional neural network method. *TEM Journal*, 13(1): 531-536.
- [4] Alver, Y., Demirel, M.C., Mutlu, M.M. (2014). Interaction between socio-demographic characteristics: Traffic rule violations and traffic crash history for young drivers. *Accident Analysis & Prevention*, 72: 95-104. <https://doi.org/10.1016/j.aap.2014.06.015>
- [5] Yang, H.D., Guo, R.Y. (2023). Joint solution for temporal-spatial synchronization of multi-view videos and pedestrian matching in crowd scenes. *Traitement du Signal*, 40(5): 1807-1820. <https://doi.org/10.18280/ts.400503>
- [6] Fernández-Caballero, A., Gomez, F.J., Lopez-Lopez, J. (2008). Road-traffic monitoring by knowledge-driven static and dynamic image analysis. *Expert Systems with Applications*, 35(3): 701-719. <https://doi.org/10.1016/j.eswa.2007.07.017>
- [7] Scotti, G., Marcenaro, L., Coelho, C., Selvaggi, F., Regazzoni, C.S. (2005). Dual camera intelligent sensor for high definition 360 degrees surveillance. *IEE Proceedings-Vision, Image and Signal Processing*, 152(2): 250-257. <https://doi.org/10.1049/ip-vis:20041302>
- [8] Kumari, S., Choudhary, M., Kumari, K., Kumar, V., Chowdhury, A., Chaulya, S.K., Mandal, S.K. (2022). Intelligent driving system at opencast mines during foggy weather. *International Journal of Mining, Reclamation and Environment*, 36(3): 196-217. <https://doi.org/10.1080/17480930.2021.2009724>
- [9] Kato, J., Watanabe, T., Joga, S., Rittscher, J., Blake, A. (2002). An HMM-based segmentation method for traffic monitoring movies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9): 1291-1296. <https://doi.org/10.1109/TPAMI.2002.1033221>
- [10] Cohn, E.G., Kakar, S., Perkins, C., Steinbach, R., Edwards, P. (2020). Red light camera interventions for reducing traffic violations and traffic crashes: A systematic review. *Campbell Systematic Reviews*, 16(2): e1091. <https://doi.org/10.1002/cl2.1091>
- [11] Bolz, K., Nowacki, G. (2024). Unmanned Aerial Vehicles in the road safety system. *Roads and Bridges-Drogi i Mosty*, 23(1): 73-98. <https://doi.org/10.7409/rabdim.024.004>
- [12] Borisova, S.E. (2021). Particularities of the informational and psychological effect of internet TV news videos on the legal awareness and behavior of road users. *Психология И Право*, 11(4): 78-89.
- [13] Singh, T., Rajput, V., Satakshi, Prasad, U., Kumar, M. (2023). Real-time traffic light violations using distributed streaming. *The Journal of Supercomputing*, 79(7): 7533-7559. <https://doi.org/10.1007/s11227-022-04977-4>
- [14] Kan, H.Y., Li, C., Wang, Z.Q. (2024). Enhancing Urban Traffic Management through YOLOv5 and DeepSORT Algorithms within Digital Twin Frameworks. *Mechatronics and Intelligent Transportation Systems*, 3(1): 39-54. <https://doi.org/10.56578/mits030104>
- [15] Kotapati, G., Ali, M.A., Vatambeti, R. (2023). Deep learning-enhanced hybrid fruit fly optimization for intelligent traffic control in smart urban communities. *Mechatron. Mechatronics and Intelligent Transportation Systems*, 2(2): 89-101. <https://doi.org/10.56578/mits020204>
- [16] Tankebe, J., Boakye, K.E., Amagnya, M.A. (2020). Traffic violations and cooperative intentions among drivers: The role of corruption and fairness. *Policing and Society*, 30(9): 1081-1096. <https://doi.org/10.1080/10439463.2019.1636795>
- [17] Khalkhali, M.B., Vahedian, A., Yazdi, H.S. (2019). Multi-target state estimation using interactive Kalman filter for multi-vehicle tracking. *IEEE Transactions on Intelligent Transportation Systems*, 21(3): 1131-1144. <https://doi.org/10.1109/TITS.2019.2902664>
- [18] Joshi, V.D., Sharma, M., Alsaud, H. (2024). Solving a multi-choice solid fractional multi objective transportation problem: Involving the Newton divided difference interpolation approach. *AIMS Mathematics*, 9(6): 16031-16060. <https://doi.org/10.3934/math.2024777>