IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 May 2024, accepted 28 June 2024, date of publication 8 July 2024, date of current version 20 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424812

RESEARCH ARTICLE

Single-Frame Difference-Based Image Fusion Glare-Resistant Detection System in Green Energy Vehicles

XIANG LYU[1], **NAN WANG**[1], **AND JIA GAO**[2]
[1]School of Information Engineering, Xinyang Agriculture and Forestry University, Xinyang 464000, China
[2]Graduate School of Business (GSB), SEGi University, Kota Damansara, Petaling Jaya 47810, Malaysia

Corresponding author: Xiang Lyu (lvxiang1228@126.com)

This work was supported in part by the Scientific and Technological Project of Henan Province ''Tea Leaf Disease Recognition Based on Adaptive Meta-Transfer Learning and Small Samples'' under Grant 222102210300.

ABSTRACT Green energy vehicles often use technologies that reduce lower inherent noise. However, adverse weather condition and low visibility at night can cause a glare effect from the headlights of oncoming cars. This poses a major threat to traffic safety. In order to solve this problem, this study initially adopts single frame difference for video frame selection, which reduces the computational load of image processing pipeline. Then, combined with visible and infrared images, this paper uses non-downsampled contourlet transform to achieve glare elimination. Finally, an improved convolutional network is used to detect pedestrians in anti-glare images, and volumetric Kalman filter algorithm is used to track pedestrians. Through these operations, the research establishes a Single-Frame Difference-Based Image Fusion Glare-Resistant Detection System applicable to green energy vehicles. The experimental analysis shows that the designed system can eliminate glare more than 80%, and the pedestrian detection accuracy reaches 95.44%. The constructed system aids green energy vehicles in accurately perceiving their surroundings during nighttime driving, ensuring safe travel.

INDEX TERMS Green energy vehicles, single-frame difference, image fusion, glare-free, pedestrian detection.

I. INTRODUCTION

With increasing environmental awareness and considerations of energy scarcity, the development of new energy vehicles has gained significant momentum. However, this progress brings about a series of issues, one of which is the safety of nighttime driving for these vehicles [1]. The electric technology utilized in new energy vehicles often results in reduced vehicle noise, coupled with a relatively lower number of vehicles on the road, causing drivers to have weaker awareness of approaching vehicles during nighttime driving. Additionally, fog, rain, together with other adverse climate situations further reduce visibility, posing greater safety hazards to drivers [2]. During these times, the front and rear lights of vehicles become crucial references for drivers to identify and judge the presence of other vehicles at night. Moreover, excessively bright headlights from oncoming vehicles can obstruct the driver's line of sight. Existing research on anti-glare technology for night vision is not sufficiently in-depth, and the complexity of fusion algorithms leads to challenges in accurately detecting pedestrians during nighttime driving [3]. In recent years, with the rapid development of image processing technology, the application of image fusion and deep learning in the field of night driving safety has been widely concerned. As an effective image fusion method, weighted wavelet visual perception fusion can improve image clarity and information richness by extracting feature information of different image modes and fusing them effectively [4], [5]. Similarly, the Retinex inspired color correction and detail saving fusion method can correct colors while maintaining image detail, further improving the visual quality of night images [6]. In the field of deep learning,

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Anisetti.

models such as CVANet (cascaded visual attention Network) are widely used in object detection and recognition tasks [7], [8]. By training a large amount of data, these networks can automatically learn and extract key features in images, thus achieving accurate detection of targets. To address these issues, this research constructs a Single-Frame Difference-Based Image Fusion Glare-Resistant Detection System based on image fusion and pedestrian detection technologies. The innovation lies in the improvement of the pedestrian detection network model using spatial pyramid pooling structures and dilated convolutions, effectively enhancing pedestrian detection accuracy. The research is divided into four parts: the literature review, analyzing the current domestic and international status of anti-glare and image processing technologies and proposing research directions; the methodology, providing a detailed description of the system construction; the experimental analysis, examining the system's performance through a series of experiments; and the conclusion, summarizing the research content based on experimental results and proposing future research prospects.

II. RELATED WORKS

To address the issue of glare during nighttime driving, intelligent technologies are being explored both domestically and internationally to mitigate the impact of glare on the driver's vision. Praveen et al. undertook the task of constructing a mobile system using state-of-the-art algorithms to detect pedestrian zones in adverse lighting conditions. The approach involved the utilization of numerous object detectors for the detection of human bodies. This method was explored to facilitate safer nighttime driving under challenging lighting conditions [9]. Liu et al. tackled the inherent limitations of a single sensor in adverse weather conditions by employing a sensor fusion approach that combines radar and camera information for anti-glare nighttime driving. The method involved matching the Mahalanobis distance with the observed values of the target sequence and employed a joint probability function for data fusion. Experimental results demonstrated that this method enhances the robustness of the environmental perception system [10]. Murugan and Sathyabama compared the performance of three different target detection models to enhance the safety of nighttime driving. They found that YOLOv4 exhibited higher detection accuracy. Additionally, they introduced low-pass and unsharp filters into the model to reduce noise and improve image clarity [11]. Duan and Xia addressed the challenges of traditional night vision imaging technologies, which are constrained by infrared focal plane arrays and difficulties in nighttime imaging due to strong light interference. They proposed a color night vision imaging solution that does not rely on infrared focal plane arrays. Through the means of accessing the correlation function of two light fields, they reconstructed two infrared night vision images. Experiment outcomes indicated that this method possess the capability to produce high-quality color night vision images [12]. Ashiba and Ashiba introduced three new infrared night vision image enhancement methods to achieve anti-glare detection at night. Through comparative experiments, they found that an image enhancement method combining gamma correction and histogram matching exhibited superior performance. Before enhancing the images, they applied the optimal global threshold segmentation method to segment infrared images, and finally, the images underwent sharpening through a high-pass filter [13].

Currently, image processing technology has gradually matured, and scholars have conducted in-depth research on the improvement of image processing algorithms. Rasti et al. have explored image processing techniques for extracting key grain crop growth indicators from proximal images, outlining the relevant basic image processing technologies. Through performance reviews and comparisons, they revealed limitations in the image processing technology for grain monitoring [14]. Khairandish et al. considering the superiority of convolutional neural networks in benign and malignant tumor classification, combined support vector machines with convolutional neural networks to create a hybrid model for tumor image classification. Experiment outcomes provided the evidence that the hybrid model achieved a recognition accuracy of 98.45% [15]. Z. Tang et al. proposed a new image segmentation method to solve the problem that the image segmentation model based on deep learning requires a large amount of data and the model generalization ability is poor. This method combines deformable model with medical deformer neural network to segment medical image. The results show that the segmentation accuracy of this method reaches about 90% [16]. Wang et al. found that spatial resolution transcriptomics (SRT), while capable of providing gene expression close to or better than single-cell resolution, was affected by high noise levels in expression data. To address this, they used image processing techniques to correlate matched positions and imaging data, accurately expressing SRT gene expression. Experimental results showed that this method could eliminate missing data in single-cell RNA sequencing [17]. Monga et al. reviewed algorithms about processing image and signal. This team extensively introduced algorithm deployment techniques in multi-fields, including imaging recognition and speech processing. By giving previous work review and analysis, they revealed the connection between iterative algorithms and neural networks and introduced the latest theoretical achievements [18].

Combining the literature mentioned above, it is evident that current research on night vision anti-glare technology is not sufficiently deep. The complexity of fusion algorithms has led to difficulties in accurately detecting pedestrians during nighttime driving. Furthermore, there is still a significant gap in the quality of fused images compared to the original images. Therefore, a new anti-glare detection system based on image processing technology and pedestrian detection technology has been developed to address these issues and improve the safety of nighttime driving for vehicles.

In order to comprehensively analyze the research status and compare the existing research with this research work, the research summarized the application technology,

TABLE 1. Summary and comparative analysis of the study status.

Research work	Technology	Performance	Superiority	Inferior strength	In contrast to this study
A. Praveen et al.[9]	Object detector	Improved pedestrian detection capability under harsh lighting conditions	Powerful pedestrian detection function	High complexity, there may be misdetection	This study focuses more on the fusion of anti-halo processing and pedestrian detection
Z. Liu et al.[10]	Radar and camera information fusion	Improve the robustness of environmental sensing systems	Multi-sensor information fusion	The system cost may be relatively high	This study is even less costly
R. A. Murugan et al.[11]	YOLOv4 Target detection model	Higher detection accuracy, reduce noise and improve image sharpness	Efficient pedestrian detection with image enhancement	The image processing algorithm may be more complicated	This study combines pedestrian detection with anti-halo techniques
D. Duan et al.[12]	Colour night vision imaging protocol	No infrared focal plane arrays are required to obtain high-quality color night vision images	Innovative night-vision imaging methods	May be sensitive to a particular light source	Research focuses on real-time and antihalo performance
M. O. Khairandish et al.[15]	The SVM is combined with a convolutional neural network	Tumor image classification with a high accuracy rate	Hybrid model improves the classification performance	Training may require a lot of data	The model structure of this study is much simpler
V. Monga et al.[18]	Algorithm expansion technology	Revealing the connection between iterative algorithms and neural networks	In-depth theoretical analysis and algorithm research	Strong theoretical, practical application may be limited	This study focuses more attention on the practical application and simplifies the processing process

performance, advantages and disadvantages of each research work, as shown in Table 1.

As can be seen from the above table, the research on anti-halo light during night driving and image processing technology has made some progress, but there are still some problems. Most of the research work only focuses on the improvement or application of a single technology, but this study integrates image processing technology and pedestrian detection technology, and introduces video single frame difference to build a more comprehensive and practical anti-halo detection system. In addition, this study also improves the pedestrian detection network model by introducing space pyramid pool structure and cavity convolution to improve the performance and accuracy of the system. Therefore, this study has higher practical value and innovation in solving the problem of halo in night driving.

III. IMAGE FUSION-BASED GLARE-RESISTANT DETECTION SYSTEM

With the continuous development of society, automobiles have brought a lot of convenience to people's daily travel. However, safety issues persist during car travel [19], [20], [21]. To achieve pedestrian detection during night-time driving, a new glare-resistant detection system was constructed based on technologies such as single-frame differences, image fusion, and deep learning.

A. VIDEO FRAME SELECTION AND IMAGE PRE-PROCESSING

With the purpose to reinforce the real-time performance of the glare-resistant detection system, a novel frame selection method was devised based on single-frame differences, enabling the system to promptly detect pedestrians from video frames. Due to the high algorithmic complexity of night vision glare-resistant pedestrian detection systems, intermittent cyclic frame selection using inter-frame differencing was adopted for image fusion within the video. The process of selecting video frames is illustrated in Figure 1.

As shown in Figure 1, in the video frame sequence, the first frame is taken as the reference frame and the second frame as the optional fusion video frame. The frame number of the selected fusion video is then offset from the frame number of the identified reference frame, and the resulting difference is determined to be the number of frames between the two. After determining the number of frames separated by two frames, the study determines the threshold of the maximum interval frame according to the video frame rate and visual residual time. When the number of frames separated by two frames exceeds the threshold, the selected fusion video frame is retained and used as a new reference frame. Then start the difference operation again. When the number of frames separated by two frames is less than the threshold, the cosine Angle of the feature vector between the selected fused video frame and the reference frame is calculated. The threshold of cosine Angle is determined according to the frame removal rate and the correlation between video frames. Finally, the cosine Angle is compared with the threshold value. If it is smaller than the threshold value, the video frame closest to but not exceeding the threshold is retained, and the current fusion frame is positioned as a new reference frame. Otherwise, the reference frame is left unchanged, and the next frame of the fusion frame is determined as the new fusion frame to be selected. The RGB histogram retains which color channel characteristics, while the RGB three-dimensional vector can also be transformed into a one-dimensional representation.

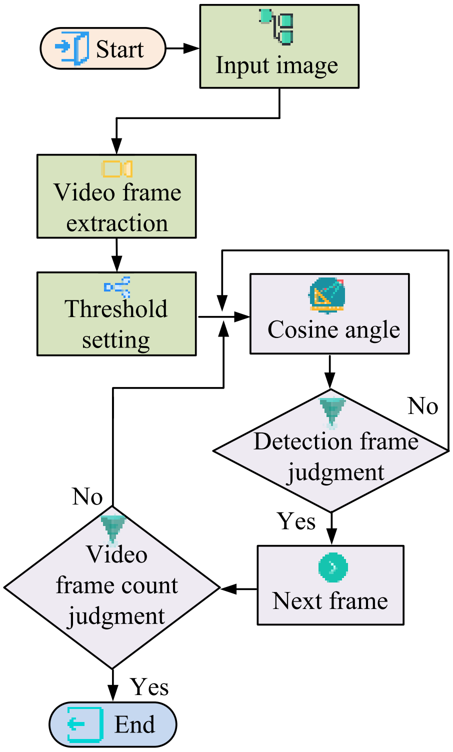


FIGURE 1. Video frame selection process.

For this purpose, the research selected the RGB histogram to express image features, obtaining the RGB histogram of visible light glare images. Image pixel values were combined based on the magnitude of pixel values in each channel, followed by normalization and mapping to integers within 64. The values of the mapped pixel points are shown as in Equation (1).

$$index_i = [B_i/64] \times 4^2 + [G_i/64] \times 4^1 + [R_i/64] \times 4^0, 1 \leq i \leq N \tag{1}$$

$index_i$ represents the mapped value of the i-th pixel, R_i, G_i and B_i refer to the pixel values of the i-th pixel in the image, and N is the symbol of the total number of pixels. $([B_i/64], [G_i/64], [R_i/64])$ is a three-digit base-four number representing four color zones. The calculation of inter-frame content differences is performed using cosine angle, specifically as depicted in Equation (2).

$$\theta = \arccos\left(\frac{R \cdot C}{\|R\| \cdot \|C\|}\right) = \arccos\left(\frac{\sum_{i=0}^{63} r_i c_i}{\sqrt{\sum_{i=0}^{63} (r_i)^2} \times \sqrt{\sum_{i=0}^{63} (c_i)^2}}\right) \tag{2}$$

In Equation (2), θ denotes the cosine angle, R stands for the reference frame feature vector set, with r_i as vectors within the set, and C represents the frame feature vector set to be selected, containing vectors c_i. The threshold for inter-frame content differences ensures the reduction of processed video frames as much as possible while observing consecutive frames, thereby further enhancing the real-time capability of the glare-resistant detection system [22], [23], [24]. To select an appropriate difference threshold, the study represented the correlation between selected video frames using nonlinear correlation information entropy. The nonlinear correlation information entropy, measuring the quantitative correlation between every two adjacent frames among these k video frames, is illustrated in Equation (3).

$$IE = \frac{\sum_{i=1}^{K-1} NCC\left(x_i, x_{i+1}\right)}{K-1} \tag{3}$$

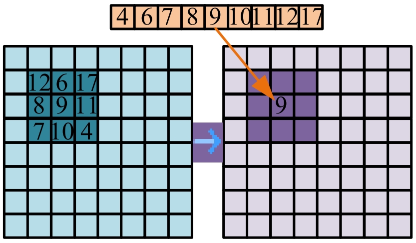
Here, $\mathrm{NCC}\left(\mathrm{X_i}, X_{i+1}\right)$ denotes the nonlinear correlation coefficient between the $\mathrm{X_i}$ th frame image and the $\mathrm{X_{i+1}}$ th frame image. The computation method for this nonlinear correlation coefficient is demonstrated in Equation (4).

$$NCC\left(x_i, x_{i+1}\right) = 2 + \sum_{i=1}^{b^2} \frac{n_{ij}}{N} \log_b \frac{n_{ij}}{N} \tag{4}$$

In Equation (4), b represents the grid number of the network, with a value of 8. n_{ij} is the symbol of the samples number, which are distributed in the ij-order grid, and N is the symbol of the total sample number. Through experimental verification, the study sets the inter-frame content difference threshold to 1.8. To prevent small changes in inter-frame content or algorithm errors and address issues with delayed frame updates, the study introduces the maximum inter-frame threshold, calculated as shown in Equation (5).

$$N' = f \times T \tag{5}$$

In Equation (5), f is the frame rate and T is the maximum detection time threshold. As the optical signal travels from the human eye to the brain, the time for visual persistence to disappear is 0.2 seconds. To secure the video observation continuity, the study selects the maximum detection time threshold as 0.2 seconds. In glare scenes, visible light images may suffer from glare, making it challenging to observe details in dark areas, while infrared images may lack clear edge contours and exhibit monotonous colors. To address this, the study combines these two types of images to create a new anti-glare method. Before performing anti-glare processing, image preprocessing is necessary. When visible light and infrared cameras capture the same scene, differences in sensor types, positions, and external environmental factors result in spatial differences in the captured image information [25], [26], [27]. Therefore, the study employs feature point registration to eliminate spatial differences between the two types of images. Visible light images and infrared images are enhanced using Multi-Scale Retinex with Color Restoration and Multi-Scale Retinex, respectively. To accomplish the reduction of the noise impact on image information, the study utilizes median filtering to eliminate image noise. The principles of the median filtering algorithm and the two types of images selected in the study are illustrated in Figure 2.



(a) Principle of median filtering algorithm

Visible light image Infrared image

(b) Two types of images

FIGURE 2. **The principle and research of median filtering algorithm and the selection of image types.**

B. IMAGE FLARE REMOVAL BASED ON IMAGE FUSION

After obtaining video frame images and preprocessing them, the study aims to obtain images that are free from flares and have clear background contours by utilizing the fusion of infrared light images and visible light images. To enhance the computational speed of the fused image and prevent color information from participating, the study transforms RGB images into the Hue, Saturation, Value (HSV) color space. The transformation process is illustrated in Figure 3.

$$H=\begin{cases}0^\circ, \max=\min\\ 60^\circ\dfrac{(G-B)}{(\max-\min)}, \max=R\,\&\,G\geq B\\ 360^\circ+60^\circ\dfrac{(G-B)}{(\max-\min)}, \max=R\,\&\,G<B\\ 120^\circ+60^\circ\dfrac{(B-R)}{(\max-\min)}, \max=G\\ 240^\circ+60^\circ\dfrac{(R-B)}{(\max-\min)}, \max=B\end{cases}$$

$$S=\begin{cases}\dfrac{\max-\min}{\max}, \max\neq 0\\ 0, \max=0\end{cases} \qquad V=\max$$

FIGURE 3. **Conversion process from RGB.**

In the HSV color space, the high-brightness flare information is mainly concentrated in the V component [28], [29], [30]. Therefore, the study focuses on manipulating the V component. Firstly, the normalized value of the H component is calculated using the product method, as specified in Equation (6).

$$H_i=\left|\frac{H}{60}\right| \tag{6}$$

In Equation (6), H_i represents the normalized value of the H component. Then, based on the normalized quantity obtained from the above Equation, its corresponding frequency is calculated, as shown in Equation (7).

$$f'=\frac{H}{60}-H_i \tag{7}$$

In Equation (7), f' represents the frequency corresponding to H_i. To output an RGB image, the study introduces three intermediate components representing colors, as specified in Equation (8).

$$\begin{cases}p=V\times(1-S)\\ q=V\times\left(1-f'\times S\right)\\ t=V\times\left[1-\left(1-f'\times S\right)\right]\end{cases} \tag{8}$$

In Equation (8), V is the V component in the HSV color space, and p, q, t are three intermediate variables representing colors. Finally, using the intermediate variables as parameters, the study combines them with the value of H_i to obtain the corresponding color vectors for each color. To avoid spectral confusion in the flare image, and to describe the high-frequency subbands and texture details in the image more flexibly, the study adopts the Non-subsampled Shearlet Transform (NSST) algorithm to decompose the night vision flare image. The process of decomposing the NSST algorithm for the image is illustrated in Figure 4.

As shown in Figure 4, the NSST method decomposed the V component of visible light and infrared source images for k times by Nonsubsampled Pyramid to obtain high and low frequencies, and then used SF(Shearlet Filter) to filter the obtained high frequencies to obtain high-frequency subbands with different directions and scales. The decomposition process of the visible light V component and the high-frequency subbands and low-frequency subbands in the infrared image is specifically shown in Equation (9).

$$\begin{cases}L_k\left(i,j\right)=\displaystyle\sum_{i=0}^{p'}\sum_{j=0}^{q'}C_{k-1}\left(p',q'\right)h\left(2p'-i\right)h\left(2q'-j\right)\\ H_{k,l}\left(i,j\right)\\ =\displaystyle\sum_{i=0}^{p'}\sum_{j=0}^{q'}C_{k-1}\left(p',q'\right)h\left(2p'-i\right)h\left(2q'-j\right)g\left(2p'-i\right)g\left(2q'-j\right)\end{cases} \tag{9}$$

In Equation (9), $L_k\left(i,j\right)$ and $H_k\left(i,j\right)$ represent the decomposed low-frequency subband and high-frequency subband, respectively. $h\left(\cdot\right)$ is the non-subsampled pyramid filter, $g\left(\cdot\right)$ is the non-subsampled shearlet wavelet filter, $\left(p',q'\right)$ is the position of the source image, (i,j) is the position of the decomposed image, k is the decomposition scale, and

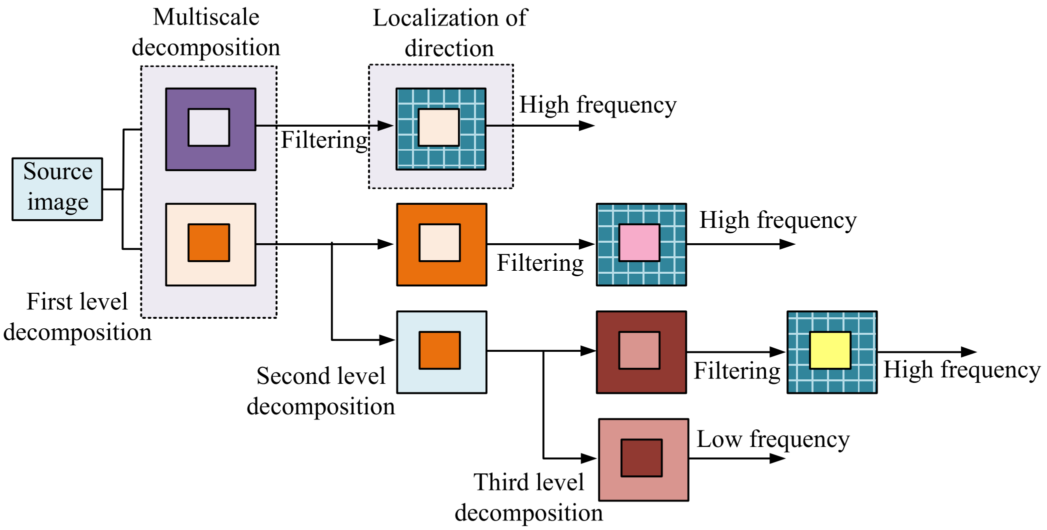


FIGURE 4. **Process of NSST algorithm for decomposing halo images.**

$C_{k-1}(p',q')$ is the image to be decomposed. To obtain the V component after eliminating flares, the study uses NSST to reconstruct the anti-flare image of the fused high and low frequencies. The reconstruction method for the fused high and low frequencies is shown in Equation (10).

$$\begin{aligned} C'_{k-1}(p',q') &= \sum_{p'=0}^{i}\sum_{q'=0}^{j} L_k^{FU}(i,j)\,h(2p'-i)\,h(2q'-j) \\ &\quad + \sum_{p'=0}^{i}\sum_{q'=0}^{j} H_{k,l}^{FU}(i,j)\,h(2p'-i)\,h(2q'-j) \\ &\quad \times g(2p'-i)\,g(2q'-j) \end{aligned} \tag{10}$$

In Equation (10), $C'_{k-1}(p',q')$ represents the obtained new luminance component, L_k^{FU} denotes the low-frequency fusion coefficient, and $H_{k,l}^{FU}(i,j)$ stands for the high-frequency fusion coefficient. It is evident that the high luminance information of visible light primarily resides in its low frequencies. Therefore, when performing low-frequency fusion, it is necessary to eliminate high luminance information. Hence, the research explores a fusion rule where the weighting of the infrared low-frequency component adjusts automatically with the high luminance information using the V component.

$$\eta(i,j) = \frac{1}{2\pi}\tan^{-1} 2\left(L_k^{VI}(i,j) - m\right) + n \tag{11}$$

In Equation (11), $L_k^{VI}(i,j)$ represents the low-frequency subband of the V component, m is the symbol of the critical value at the boundary between halo and cloud boundary of the V component, and n represents the symbol of the infrared low-frequency coefficient weight at the critical value. After multiple calculations and comparisons, the research sets m and n values to 3 and 0.75, respectively. Details, contours, edges, and other information in the image are mainly contained in the high-frequency subbands. Moreover, there are significant differences between the high-frequency components obtained from the V component and the infrared image. To obtain more edge information and clearer contours, the research adopts a strategy based on regional contrast to fuse high-frequency coefficients. Before fusing the high-frequency coefficients, it is necessary to calculate the image contrast, as shown in Equation (12).

$$\begin{cases} Z_{k,l}(i,j) = SML\left(H_{k,l}(i,j)\right)\dfrac{\left|H_{k,l}(i,j)\right|}{\bar{L}_k(i,j)} \\ \bar{L}_k(i,j) = \dfrac{1}{M\times N}\displaystyle\sum_{r=-\left(\frac{M}{2}-1\right)}^{\frac{M}{2}-1}\;\sum_{c=-\left(\frac{N}{2}-1\right)}^{\frac{N}{2}-1} L_k(i+r,j+c) \end{cases} \tag{12}$$

In Equation (12), $M \times N$ is the region size, $Z_{k,l}(i,j)$ is the contrast corresponding to the coefficient of the image in the lth direction subband at the kth scale, and $\bar{L}_k(i,j)$ is the regional mean of the low-frequency subband. The high-frequency fusion coefficients are obtained based on the image contrast, as shown in Equation (13).

$$H_{k,l}^{FU}(i,j) = \begin{cases} H_{k,l}^{VI}(i,j), & Z_{k,l}^{VI}(i,j) > Z_{k,l}^{IR}(i,j) \\ H_{k,l}^{IR}(i,j), & Z_{k,l}^{VI}(i,j) \le Z_{k,l}^{IR}(i,j) \end{cases} \tag{13}$$

In Equation (13), $H_{k,l}^{VI}$ and $H_{k,l}^{IR}$ are the high-frequency subbands of the V component and the infrared image, respectively. $H_{k,l}^{FU}(i,j)$ is the coefficient of the subband in the lth direction at the kth scale for the fused image, and $Z_{k,l}^{VI}(i,j)$ and $Z_{k,l}^{IR}(i,j)$ are the image contrasts of the V component and the infrared image, respectively. The implementation process of the anti-halo fusion method for fusing visible light and infrared images is illustrated in Figure 5.

As shown in Figure 5, three components of the image can be obtained by converting the RGB image into the HSV color space, namely, brightness component V, hue component H and saturation component S. Next, the brightness component and infrared image in the original image are processed,

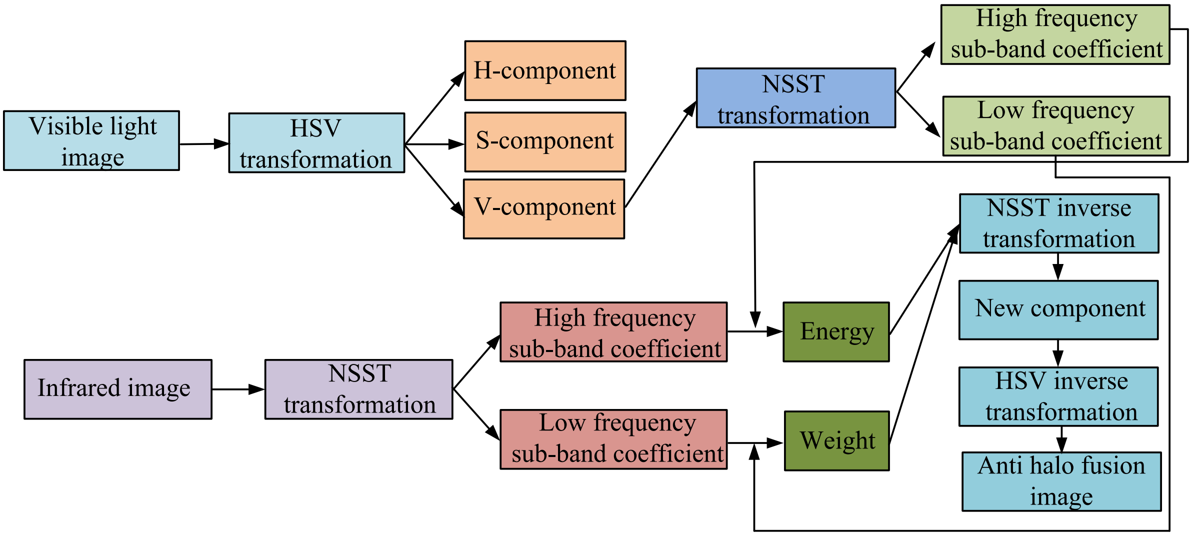


FIGURE 5. The realization process of the anti-halo method of visible light image and infrared image fusion.

in which H and S are not involved, so as to reduce the running time of the fusion algorithm and improve the speed of image fusion. At the same time, the brightness component is separated from other components to avoid color confusion. After the fusion brightness component is obtained, the desired anti-halo image can be obtained after the inverse transformation of HSV and NSST.

As mentioned above, the image fusion model mainly includes the following key components: image preprocessing, image decomposition, coefficient fusion, image reconstruction and result optimization. These components work together to achieve high-quality image fusion results. Image preprocessing is the first step in the fusion process, which involves converting the input visible and infrared images into a format or space suitable for fusion processing. It is converted to HSV color space to separate out the brightness component V, hue component H and saturation component S. Image decomposition is an important step in the fusion process. The preprocessed image is decomposed into multi-scale and multi-direction subband coefficients. The study uses NSST to realize image decomposition. Coefficient fusion is the core of the fusion process. The low frequency fusion coefficient and high frequency fusion coefficient are calculated to ensure that the fusion image retains the details of the infrared image. After the coefficients are fused, the next step is image reconstruction. In this step, by combining the fused brightness component with the original hue component and saturation component, the final anti-halo image is obtained through the inverse transformation of HSV and RGB. Results The image enhancement and denoising operation were used to further adjust and optimize the fused image, so as to improve the visual effect and practicality of the image.

C. ROAD PEDESTRIAN DETECTION BASED ON GLARE-RESISTANT FUSED IMAGES

Due to the fact that pedestrians detected during nighttime driving are situated within a glare-resistant background, despite undergoing glare elimination operations based on image fusion, there still exists a certain gap in image clarity compared to daytime images. Therefore, the study employs the YOLOv4 object detection algorithm to detect pedestrians. To achieve pedestrian detection, the research introduces dilated convolutions into the YOLOv4 backbone network to enhance semantic features of pedestrians in the glare-resistant night vision background. Subsequently, a spatial pyramid pooling structure is employed to learn features of pedestrians of different sizes, reducing false negatives. To simplify pedestrian detection, the study reduces the number of feature fusion layers in the network and utilizes clustering algorithms to enhance model performance. The pedestrian detection network structure designed in the study is illustrated in Figure 6.

To enable the backbone network to extract deeper-level feature information, the study incorporates dilated convolutions into the pedestrian detection network to increase the network's receptive field. In the process of continuous dilation rate variation, varying numbers of zeros need to be added to the convolution kernel. The relationship between the receptive field and the convolution kernel is expressed in Equation (14).

$$x_n = x_k + (x_k - 1) \times (D_r - 1) \tag{14}$$

In Equation (14), x_n is the symbol of the dilated convolution kernel size, x_n is the original convolution kernel size, and D_r is the dilation rate. The calculation method for the receptive field is shown in Equation (15).

$$y_m = y_{m-1} + \left[(x_m - 1) \times \left[\prod_{i=1}^{m-1} s_i \right] \right] \tag{15}$$

In Equation (15), y_m is the receptive field of each point in the m th layer, y_{m-1} is the receptive field of each point in the $m-1$ th layer, x_m is the size of the dilated convolution kernel in the m th layer, and s_i is the stride of the ith layer convolution. To enhance the performance of the feature extraction network in obtaining more features during the extraction

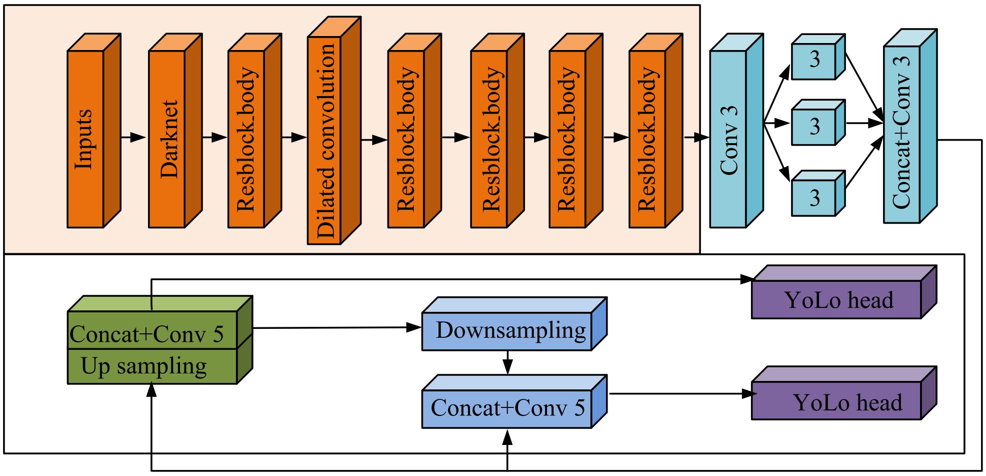


FIGURE 6. Structure of the pedestrian detection network in the study design.

of the pedestrian target, the study directly fuses features from different layers, revealing a more comprehensive set of broader features. The spatial pyramid pooling structure uses parallel convolution layers with different dilation rates to capture target objects and backgrounds, and then concatenates the obtained results. In this structure, dilated convolutions are employed to expand the receptive field without affecting resolution. The size calculation of the feature map is provided in Equation (16).

$$output\_size = \frac{n' + 2p'' - d \times \left(f'' - 1\right) - 1}{s} \tag{16}$$

In Equation (16), n' represents the size of the input feature map, p'' is the zero-padding amount, d is the dilation rate, s is the stride, and f'' is the convolutional kernel size. As the number of convolutional layers increases, the extracted feature information in the network becomes richer, leading to an increase in computational complexity. To balance computational efficiency and detection performance, the study investigates reducing the number of feature layers in YOLOv4 to 2 and clustering the anchor boxes. The study employs the K-means clustering algorithm to generate anchor boxes. However, traditional K-means clustering algorithms have random initial cluster centers, leading to local optima and an inability to obtain optimal candidate box sizes. To address this, the study improves the K-means algorithm by initially selecting a sample from an anti-glare image dataset to work as the initial cluster center. The distance between each sample point in the anti-glare dataset is calculated. The probability of each sample point being selected as the next cluster center is then computed as shown in Equation (17).

$$P\left(x\right) = \frac{D\left(x\right)}{\sum D\left(x\right)^2} \tag{17}$$

In Equation (17), $D\left(x\right)$ is the distance between the sample point and the existing cluster center, and $P\left(x\right)$ is the probability of being selected as the next cluster center. The improved K-means algorithm results in modified anchor box sizes of (31,129), (37,138), (40,153), (42,215), (44,169), and (48,193). The study adopts the Mish function as the activation function and utilizes the CIOU_loss loss function to calculate the error between predicted and actual bounding boxes in a pedestrian detection network. Model optimization is achieved through training. Due to the dynamic nature of green energy vehicles driving at night and pedestrians walking, ensuring safe driving requires further tracking of detected pedestrians after implementing night vision anti-glare pedestrian detection. Therefore, the study combines Kernelized Correlation Filters (KCF) with Closed Loop (CL) algorithms for pedestrian tracking. This method predicts pedestrian positions by jointly using color and spatial features. The proposed pedestrian tracking method's workflow is illustrated in Figure 7.

In summary, the research initially collects pedestrian detection datasets from in-car cameras and selects video frames based on single-frame differences. After preprocessing the obtained image frames, the study employs the HSV and NSST algorithms to fuse and eliminate glare in visible light and infrared images. Finally, night vision pedestrian detection and tracking are performed using the CL algorithm, YOLOv4, and KCF.

IV. PERFORMANCE AND APPLICATION EFFECT ANALYSIS OF GLARE-RESISTANT DETECTION SYSTEM

A series of experiments are designed to test the performance of the anti-halo detection system. The data set selected in the experiment is the video frame data set collected by the vehicle camera. The dataset contains videos of both suburban roads and urban arterial roads. Slow video and Fast video in the two scenarios were used as experimental data sets. A Slow video on a road in an urban area is a slow video sequence with a frame count of 300, a playback duration of 12s, a playback frame rate of 25 frames /s, and a frame interval threshold of 5. Fast vedio is a motion sequence of fast video. The frame number of the video sequence is 86, the playback time is 3.44s, the playback frame rate is 25 frames /s, and the maximum

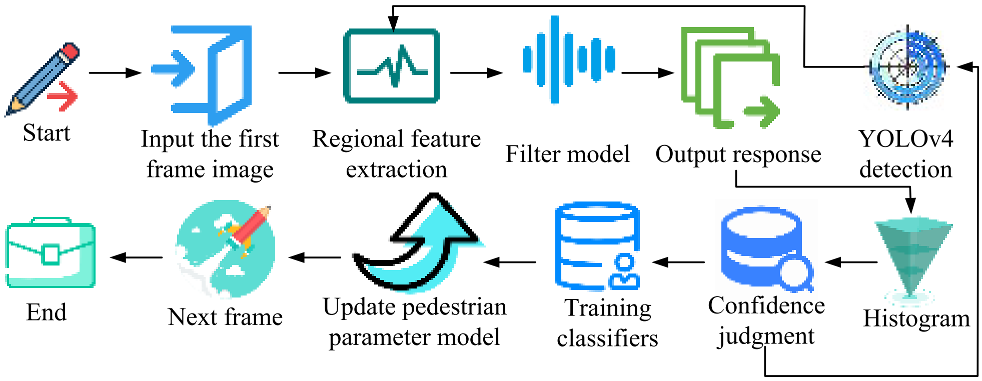


FIGURE 7. The flow of pedestrian tracking methods for the study design.

frame interval threshold N is set to 5. In suburban scenarios, the number of Slow video sequences is 371, the playback duration is 14.84s, the playback frame rate is 25 frames /s, and the maximum frame interval threshold N is set to 5. Fast video indicates a fast video sequence. The frame rate of the video is 25 frames /s, the maximum frame interval threshold N is 5, the playback duration is 8s, and the number of video frames is 200.

These data features mainly include the appearance characteristics, posture characteristics, movement characteristics and scene characteristics of pedestrians. Through the effective extraction and utilization of these features, the accuracy and reliability of pedestrian detection can be improved. Because of the video captured by the car camera, many frames may be redundant. Therefore, the study adopts a video frame selection method based on single frame difference. By calculating the difference between successive frames, representative frames are selected for subsequent processing, thus improving the quality and efficiency of data. Then, we study the image preprocessing of the selected video frames. Denoising, enhancement and normalization. Due to the diversity of pedestrian walking posture, dress and other factors in the actual scene, the samples in the data set may not cover all cases completely, which may lead to poor performance of the model in some special cases. Secondly, because the collection and annotation of data sets require a lot of manpower and material resources, the scale of data sets may be limited to some extent, which may affect the generalization ability and performance of the model. To overcome these limitations, the study may consider using data enhancement techniques to augment the data set.

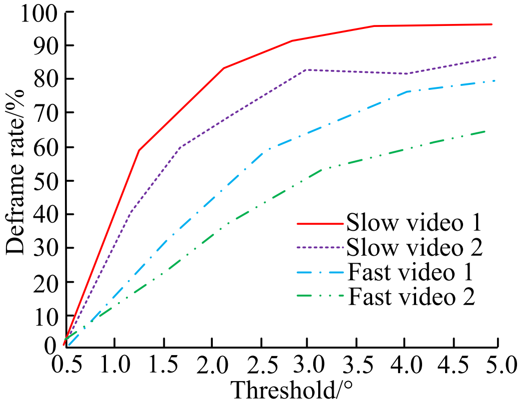
In order to ensure the validity of the data and achieve the best results in the process of experimental verification and model training, a series of data preprocessing is carried out. In the process of processing, the feature point registration method is used to eliminate the spatial difference between the two types of image information. For visible image and infrared image, Multi-Scale Retinex with Color Restoration and Multi-Scale Retinex were used to enhance the image, respectively. In order to reduce the influence of noise on image information, the median filter is used to eliminate image noise.

To gauge the effectiveness of metrics in balancing frame interval detection and visual outcomes, the study investigated the relationship between the optimal frame difference threshold, frame drop rate, and nonlinear correlation information entropy through experiments. Two fast-playback videos and two slow-playback videos were selected for the experiments, and the variations in frame drop rates and information entropy for both types of videos with changing thresholds were illustrated in Figure 8.

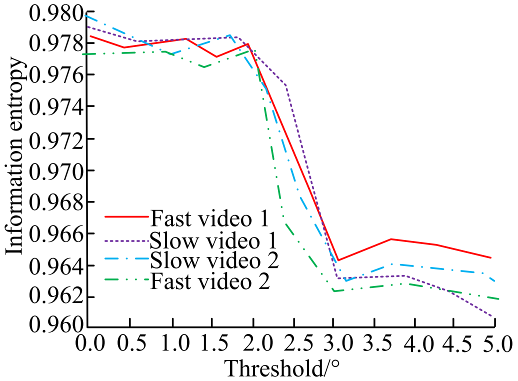
As depicted in Figure 8(a), the frame drop rates for all four sets of video sequences continuously increased with the rising threshold. The increase in slow video sequences 1 and 2 was notably higher compared to the increment in the fast video group. Figure 8(b) indicates that as the threshold increased, the size of information entropy gradually decreased, with a significant drop occurring between the threshold values of 1.5 and 2.5. Figure 8(c) reveals that at a threshold of 2.2, there was a substantial variation in information entropy values. Hence, in threshold selection, it's crucial to consider values below 2.2. Combining the findings of frame drop rate variations, a threshold value of 1.8 yielded a higher number of interval video frames and a stronger correlation among video sequences.

To validate the efficacy of the proposed single-frame difference-based frame selection method, the study conducted experiments using infrared videos. Experiments were performed in the urban area of A city on a ring road scene, obtaining motion sequences from slow-playback videos. The video sequence comprised 365 frames with a playback duration of 14.72 seconds and a frame rate of 25 frames/s. The maximum inter-frame threshold was set to 5 based on the playback frame rate. Employing the frame selection method established in the study, the results for the initial 50 frames are presented in Figure 9.

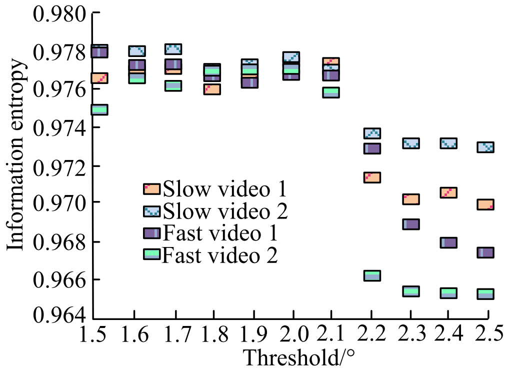
From Figure 9, it is evident that while the original video sequence consisted of 365 frames, the frame selection method based on inter-frame differences required only 159 frames, reducing the processing load. Among these, 133 frames met



(a) Deframe rate



(b) Information entropy 1



(c) Information entropy 2

FIGURE 8. **Deframe rate and information entropy changes of two types of videos.**

the conditions of a time interval not exceeding 0.2 seconds, with minimal inter-frame content difference and closest to the threshold condition. This indicates that in slow-playback videos, the majority of video frames exhibit minimal content change between consecutive frames. The collective information in Figure 9 demonstrates that the devised frame selection method in this study can minimize workload while ensuring real-time processing.

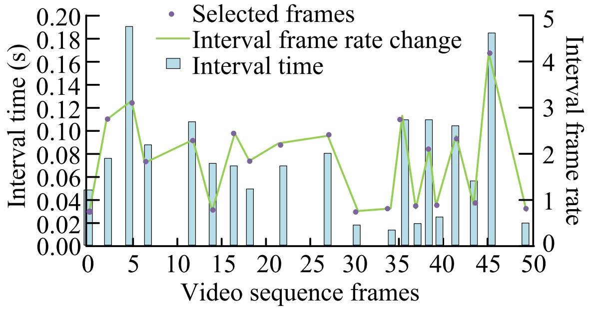


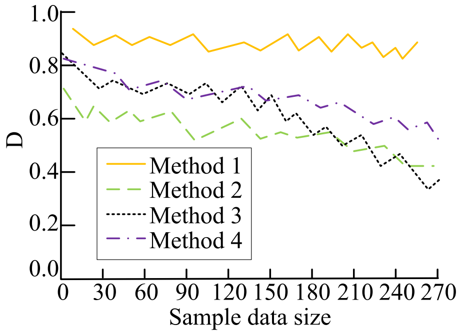
FIGURE 9. **Screening effect of the frame screening method designed in the study.**

For accomplishing the evaluation of the effectiveness of the glare-resistant method employed by the research institute, the study compared existing glare-resistant methods with the glare-resistant method designed in the study (Method 1). The comparison methods include anti-halo method in [31] (method 2), anti-halo method in [32] (method 3), and anti-halo method in [33] (method 4). The comparison of methods involved evaluating the glare elimination effects using parameters such as Glare Elimination Degree (D), Average Gradient (AG), Spatial Frequency (SF), and Root Mean Square Error (RMS). RMSVI and RMSIR fused the differences between the image and the original infrared and visible light images, respectively. The stability of the method was also studied by increasing the number of contrast images. The specific comparison results are shown in Figure 10.

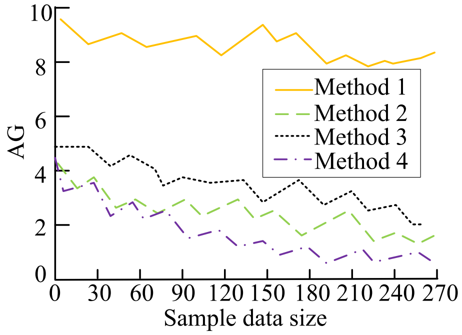
As shown in Figure 10 (a), Method 1 exhibited a 25.44%, 11.54%, and 10.44% increase in the D index compared to the other three methods. This indicates that Method 1 achieved more thorough glare elimination in the images. In Figure 10 (b), Method 1's AG value was 9.966, significantly higher than the other three methods. This suggests that the glare-resistant fused images obtained by Method 1 had more distinct object contours, making it more conducive to observing the surrounding conditions for green energy vehicles at night. Figure 10 (c) shows that Method 1's SF value was 27.14, more than 30% higher than the other three methods. This indicates that Method 1 could reflect the spatial variations in pixel grayscale in the obtained fused images. In Figure 10 (d), both RMS values of Method 1 were significantly lower than the other three methods, indicating that the fused images reduced the differences with the original images while maintaining image clarity.

To further evaluate the performance of the designed glare-resistant method, the study introduced Peak Signal-to-Noise Ratio (PS), Image Sharpness (FD), and the mean value μ of non-glare region selection. The indicator data for the four glare-resistant methods in two different video image sets are shown in Table 2.

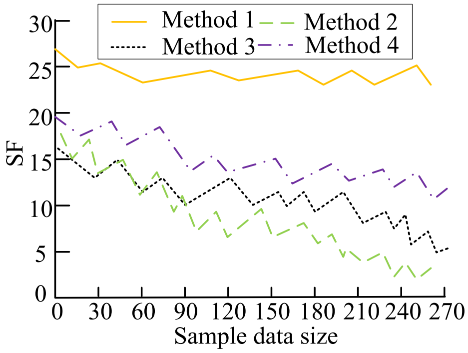
From Table 2, Method 1's PS value for infrared images was 28.455, a 17.8%, 5.1%, and 15.4% increase compared to Methods 2, 3, and 4, respectively. The visible light image's PS value for Method 1 was 40.858, significantly higher than the other three methods. This indicates fewer interference signals between the fused images and the original images. Method 1's FD value was significantly higher by 65.45%, 60.81%, and 63.88% compared to the other three methods, indicating an enhancement in visual features, color



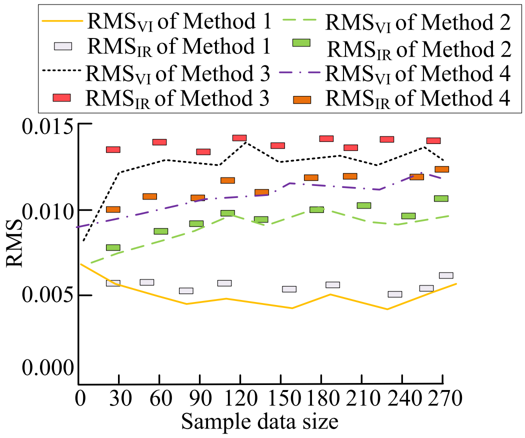
(a) D Value



(b) AG Value

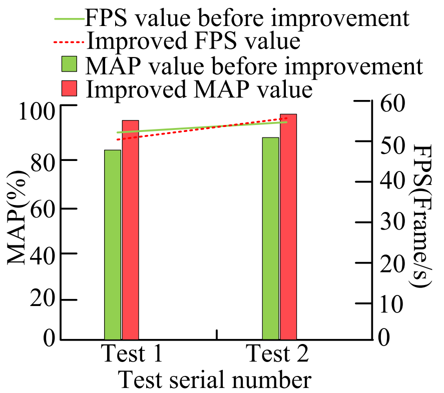


(c) SF Value

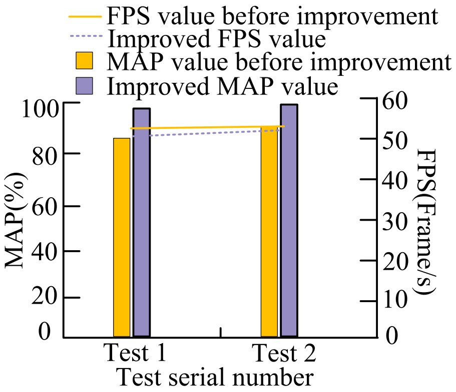


(d) EMS Value

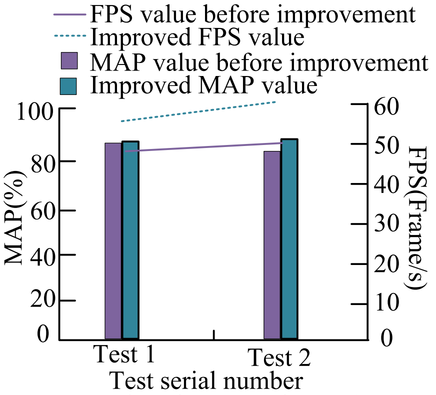
FIGURE 10. Comparison of anti halo performance indicators using different methods.



(a) Before and after the improvement of the backbone network



(b) Comparison of feature enhancement networks before and after improvement



(c) Comparison before and after feature fusion improvement

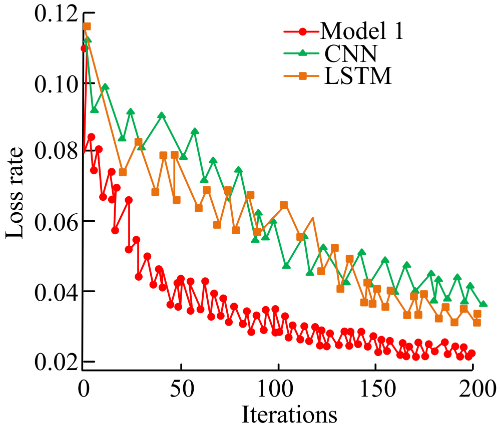
FIGURE 11. Comparison of performance indicators before and after improvement of pedestrian detection network.

texture, and richness of detail information. Moreover, Method 1's μ was the highest among the four methods. In summary, the glare-resistant method designed by the research institute effectively addresses the glare issues in nighttime video images.

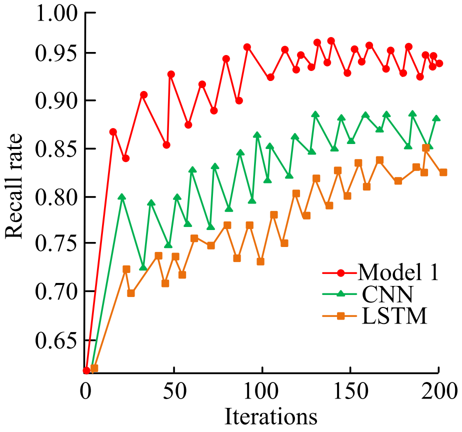
To assess the rationality of the study's impact on the YOLOv4 pedestrian detection network, the research conducted experiments by comparing the performance indicators of different network layers before and after the improvement. The compared indicators included Mean Average Precision

TABLE 2. Indicators of four anti halo methods in two different video image sets.

Project	Image Set 1				Image Set 2			
	PSVI	PSIR	μ	FD	PSVI	PSIR	μ	FD
Method 1	40.588	28.454	95.842	12.844	41.128	28.456	94.987	12.183
Method 2	33.520	24.855	65.083	8.421	34.921	24.753	68.453	8.811
Method 3	38.612	26.141	70.841	8.055	38.751	26.531	71.854	8.015
Method 4	35.171	25.084	68.155	8.155	35.641	25.184	68.344	8.151



(a) Loss



(b) Recall

FIGURE 12. Comparison of training situations of different models.

(MAP) and Frames Per Second (FPS). The comparison results are shown in Figure 11.

According to Figure 11(a), it is evident that adding dilated convolutions to the backbone network effectively enhances the feature extraction capability. After incorporating dilated convolutions, the MAP value of the network increased to 93.84%, indicating a significant improvement. As shown in Figure 11(b), adding dilated convolutions to the feature enhancement network effectively enhances pedestrian features in glare-resistant night vision, enlarging the receptive field and further increasing the MAP value of the detection results. Figure 11(c) reveals that removing a feature fusion layer effectively improves the detection frame rate, leading to a further enhancement in the pedestrian detection algorithm's speed, with the FPS value increasing from 50 to 60, a 20% improvement.

To assess the training effectiveness of the pedestrian detection model (Model 1) used in the study, it was simultaneously trained with convolutional neural networks (CNN) and long short-term memory networks (LSTM). The training results are depicted in Figure 12.

Figure 12(a) shows that, after 200 iterations, the loss values of pedestrian detection boxes for Model 1 gradually stabilized, and compared to CNN and LSTM, Model 1 converged earlier. As depicted in Figure 12(b), at 134 iterations, the recall rate of pedestrian detection for Model 1 stabilized, and Model 1 required significantly fewer iterations compared to CNN and LSTM, achieving a higher converged recall rate. Overall, the content in Figure 12 indicates that Model 1 exhibits good convergence.

In this study, YOLOv4 algorithm was used for pedestrian detection. In order to verify the correctness of this method, it was compared with a higher version of the algorithm, highlighting the advantages of the YOLOv4 algorithm. The comparison indexes include: detection accuracy, detection speed, model complexity, and resource consumption. The specific comparison results are shown in Table 3.

As can be seen from Table 3, the detection accuracy of YOLOv4 reached 93.84%, which was slightly lower than 94.21% of YOLOv5, but higher than other methods such as YOLOX, Faster R-CNN and SSD. In terms of detection speed, the FPS value of YOLOv4 is 60, which is slightly lower than 65 of YOLOX, but significantly higher than 20 of Faster R-CNN, showing its advantages in real-time. In terms of model complexity, YOLOv4, like most object detection algorithms, has O(n) complexity. In terms of resource consumption, YOLOv4 has a medium level of resource consumption. Based on the above analysis, although YOLOv5 is slightly better than YOLOv4 in detection accuracy, it is relatively poor in detection speed and resource consumption. Although YOLOX has a slight advantage in detection speed, its detection accuracy is slightly lower than YOLOv4. Considering the comprehensive requirements of detection accuracy, speed and resource consumption in practical applications, it is reasonable to use YOLOv4 algorithm for pedestrian detection.

To evaluate the effectiveness of pedestrian detection, the study compared Model 1 with CNN, LSTM, and other advanced pedestrian detection models. The results are shown in Figure 13, with the compared models including the pedestrian detection model in [34] (model 2), the pedestrian detection model in [35] (model 3), and the pedestrian detection model in [36] (model 4).

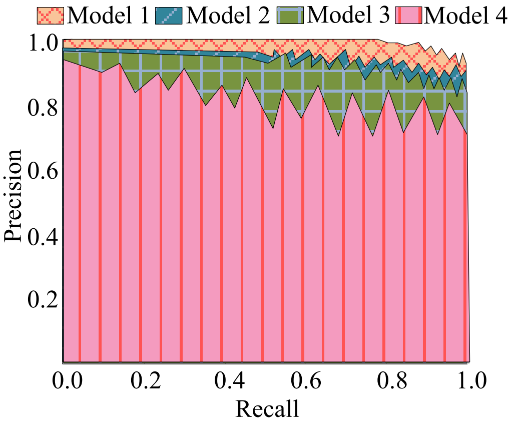
TABLE 3. Comparison of YOLOv4 and Other advanced versions for pedestrian detection.

Method	Detection Accuracy(%)	Detection Speed (FPS)	Model Complexity	Resource Consumption
YOLOv4	93.84	60	O(n)	Moderate
YOLOv5	94.21	55	O(n)	High
YOLOX	93.58	65	O(n)	Low
Faster R-CNN	92.45	20	O(n^2)	High
SSD	91.89	45	O(n)	Moderate

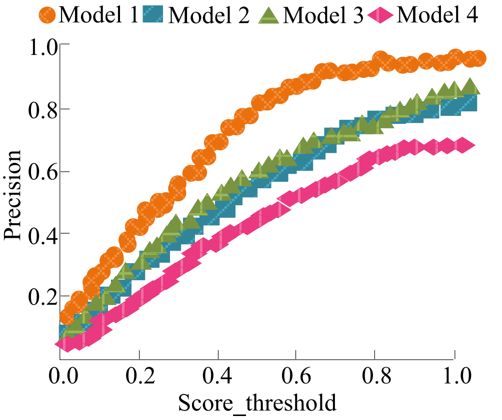
TABLE 4. A questionnaire survey on the user experience of 35 coaches.

Option	Can clearly distinguish road conditions at night		Can effectively detect pedestrians		The system responds quickly		Satisfied with the system operation experience	
	Select the number of people	Proportion (%)	Select the number of people	Proportion (%)	Select the number of people	Proportion (%)	Select the number of people	Proportion (%)
A	31	88.58	30	85.71	29	85.71	30	82.86
B	2	5.71	3	8.57	4	11.43	3	8.57
C	2	5.71	2	5.71	2	5.71	1	2.86
D	0	0.00	0	0.00	1	2.86	1	2.86

Note: A indicates strong approval; B represents recognition; C represents general; D means not approving.



(a) AP



(b) Accuracy

FIGURE 13. Comparison of trust detection performance of four models.

Figure 13(a) reveals that, with an increase in recall rate, Model 1 maintains a consistently high AP value, outperforming Model 2 and Model 3. Model 4 exhibits more significant fluctuations. This indicates that Model 1 has a higher learning level. Figure 13(b) shows that Model 1 achieves an average accuracy of 95.44%, significantly superior to the other three models. Thus, the pedestrian detection model designed in the study can achieve high-precision pedestrian detection.

To assess the practical application of the glare-resistant detection system based on single-frame difference image fusion, the study deployed it in a driving school. Daytime, instructors often need to drive home at night. The study involved 35 instructors from driving school A, who used the system for a month. Afterward, a survey was conducted, and the results are provided in Table 4.

Table 4 indicates that 94.29% of instructors believe the system provides a clear view of the road during nighttime driving, effectively alleviating visibility issues caused by glare. Furthermore, 94.28% of instructors think the system can detect pedestrian positions in real-time, aiding in decision-making. About 97.14% of instructors appreciate the system's fast response time, and 91.43% of drivers express satisfaction with the system. In conclusion, the glare-resistant detection system constructed in the study effectively facilitates safe driving on nighttime roads.

To further test the accuracy of the technology used in the constructed model, an ablation experiment was designed to analyze the effects of different components on pedestrian detection performance. The specific results are shown in Table 5.

From the ablation experiment results in Table 5, it can be observed that the full model (Model 1) incorporating all components achieves the highest precision, recall, and average precision (AP) compared to the models with individual components removed. Specifically, removing the single-frame difference image fusion component leads to a decrease in precision, recall, and AP, indicating its effectiveness in improving the detection accuracy. Similarly, excluding the YOLOv4 component results in a significant drop in all evaluation metrics, highlighting the importance of the chosen detection algorithm. Additionally, disabling the nighttime enhancement component also results in a reduction in

TABLE 5. **Ablation experiment results.**

Component	Precision(%)	Recall(%)	Average Precision(%)
Full Model	95.44	92.31	94.13
Without Single-frame Difference Image Fusion	92.16	89.43	90.28
Without YOLOv4	88.74	85.67	86.45
Without Single-frame Difference Image Fusion	91.32	88.29	89.56
Without Nighttime Enhancement	90.58	87.41	88.74

performance, demonstrating its contribution to enhancing the visibility of pedestrians in nighttime conditions.

V. CONCLUSION

The development of glare-resistant technology in automobiles was a rather intricate task. However, successfully achieving glare resistance could significantly bolster nighttime traffic safety. To this end, research focused on employing image fusion techniques that combined visible light images with infrared images. This fusion utilized HSV color space conversion and the NSST algorithm to eliminate image glare. Simultaneously, an enhanced YOLOv4 was employed for pedestrian detection. This endeavor led to the creation of a new system for detecting glare resistance in nighttime driving images through image fusion. Experimental analysis revealed that the original video sequence comprised 365 frames. By employing a frame difference-based frame selection method, only 159 frames needed fusion, significantly reducing the number of processed video frames. This demonstrated the efficacy of the proposed single-frame difference-based video frame selection method in reducing computational complexity. Method 1 exhibited a 25.44%, 11.54%, and 10.44% improvement in the D index compared to the other three methods. This indicated that Method 1 was more effective in eliminating glare from the images. Furthermore, the addition of dilated convolutions in the backbone network substantially enhanced feature extraction capability. With the inclusion of dilated convolutions, the network's MAP increased to 93.84%. This showcased that the improvements made to the pedestrian detection network significantly enhanced its performance. Practical application revealed that over 90% of the participants were satisfied with the functionality of the developed system, including its glare-resistant effect and pedestrian detection capability. This indicated that the designed model was proficient in achieving effective glare resistance and pedestrian detection. The current tracking algorithm utilized in the research is rather complex. Future studies could enhance algorithm performance by optimizing parameter quantities in the optimizer.

REFERENCES

[1] Y. Teng, Y. Cao, M. Liu, F. R. Yu, and V. C. M. Leung, "Efficient blockchain-enabled large scale parked vehicular computing with green energy supply," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9423–9436, Sep. 2021, doi: 10.1109/TVT.2021.3099306.

[2] C. P. Guzmán, N. B. Arias, J. F. Franco, J. Soares, Z. Vale, and R. Romero, "Boosting the usage of green energy for EV charging in smart buildings managed by an aggregator through a novel renewable usage index," *IEEE Access*, vol. 9, pp. 105357–105368, 2021, doi: 10.1109/ACCESS.2021.3099426.

[3] L. Fu, H. Yu, F. Juefei-Xu, J. Li, Q. Guo, and S. Wang, "Let there be light: Improved traffic surveillance via detail preserving night-to-day transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8217–8226, Dec. 2022, doi: 10.1109/TCSVT.2021.3081999.

[4] Q. Guo, F. Yang, and H. Wang, "Night vision anti-halation algorithm based on different-source image fusion combining visual saliency with YUV-FNSCT," *Electronics*, vol. 12, no. 6, p. 1303, Mar. 2023, doi: 10.3390/electronics12061303.

[5] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari, "Machine learning methodology for identifying vehicles using image processing," *Artif. Intell. Appl.*, vol. 1, no. 3, pp. 170–178, Apr. 2023, doi: 10.47852/bonviewAIA3202833.

[6] W. Zhang, L. Zhou, P. Zhuang, G. Li, X. Pan, W. Zhao, and C. Li, "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2469–2483, Apr. 2024, doi: 10.1109/TCSVT.2023.3299314.

[7] W. Zhang, L. Dong, and W. Xu, "Retinex-inspired color correction and detail preserved fusion for underwater image enhancement," *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106585, doi: 10.1016/j.compag.2021.106585.

[8] W. Zhang, W. Zhao, J. Li, P. Zhuang, H. Sun, Y. Xu, and C. Li, "CVANet: Cascaded visual attention network for single image super-resolution," *Neural Netw.*, vol. 170, pp. 622–634, Feb. 2024, doi: 10.1016/j.neunet.2023.11.049.

[9] A. Praveen, M. G. Shweta, C. SharonPriyansh, and SriSaiRishitha, "Embedded night-vision system for pedestrian detection using AdaBoosta machine learning meta-algorithm," *Ann. Romanian Soc. Cell Biol.*, vol. 25, no. 6, pp. 7825–7830, Mar. 2021. [Online]. Available: http://annalsofrscb.ro/index.php/journal/article/view/6973

[10] Z. Liu, Y. Cai, H. Wang, L. Chen, H. Gao, Y. Jia, and Y. Li, "Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6640–6653, Jul. 2022, doi: 10.1109/TITS.2021.3059674.

[11] R. A. Murugan and B. Sathyabama, "Object detection for night surveillance using Ssan dataset based modified YOLO algorithm in wireless communication," *Wireless Pers. Commun.*, vol. 128, no. 3, pp. 1813–1826, Sep. 2023, doi: 10.1007/s11277-022-10020-9.

[12] D. Duan and Y. Xia, "Pseudo color night vision correlated imaging without an infrared focal plane array," *Opt. Exp.*, vol. 29, no. 4, pp. 4978–4985, Oct. 2021, doi: 10.1364/OE.413163.

[13] H. I. Ashiba and M. I. Ashiba, "Super-efficient enhancement algorithm for infrared night vision imaging system," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 9721–9747, Mar. 2021, doi: 10.1007/s11042-020-09928-w.

[14] S. Rasti, C. J. Bleakley, N. M. Holden, R. Whetton, D. Langton, and G. O. Hare, "A survey of high resolution image processing techniques for cereal crop growth monitoring," *Inf. Process. Agricult.*, vol. 9, no. 2, pp. 300–315, Jun. 2022, doi: 10.1016/j.inpa.2021.02.005.

[15] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images," *IRBM*, vol. 43, no. 4, pp. 290–299, Aug. 2022, doi: 10.1016/j.irbm.2021.06.003.

[16] Z. Tang, J. Duan, Y. Sun, Y. Zeng, Y. Zhang, and X. Yao, "A combined deformable model and medical transformer algorithm for medical image segmentation," *Med. Biol. Eng. Comput.*, vol. 61, no. 1, pp. 129–137, Jan. 2023, doi: 10.1007/s11517-022-02702-0.

[17] Y. Wang, B. Song, S. Wang, M. Chen, Y. Xie, G. Xiao, L. Wang, and T. Wang, "Sprod for de-noising spatially resolved transcriptomics data based on position and image information," *Nature Methods*, vol. 19, no. 8, pp. 950–958, Aug. 2022, doi: 10.1038/s41592-022-01560-w.

[18] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021, doi: 10.1109/MSP.2020.3016905.

[19] L. Hu, H. Li, P. Yi, J. Huang, M. Lin, and H. Wang, "Investigation on AEB key parameters for improving car to two-wheeler collision safety using in-depth traffic accident data," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 113–124, Jan. 2023, doi: 10.1109/TVT.2022.3199969.

[20] M. L. Cummings and B. Bauchwitz, "Safety implications of variability in autonomous driving assist alerting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12039–12049, Aug. 2022, doi: 10.1109/TITS.2021.3109555.

[21] L. Zhu, F. Tao, Z. Fu, H. Sun, B. Ji, and Q. Chen, "Multiobjective optimization of safety, comfort, fuel economy, and power sources durability for FCHEV in car-following scenarios," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 1, pp. 1797–1808, Mar. 2023, doi: 10.1109/TTE.2022.3193806.

[22] S. Zhang, C. Wang, Y. Jin, J. Wu, Z. Qian, M. Xiao, and S. Lu, "Adaptive configuration selection and bandwidth allocation for edge-based video analytics," *IEEE/ACM Trans. Netw.*, vol. 30, no. 1, pp. 285–298, Feb. 2022, doi: 10.1109/TNET.2021.3106937.

[23] Z. Li, Y. Li, B. Tan, S. Ding, and S. Xie, "Structured sparse coding with the group log-regularizer for key frame extraction," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 10, pp. 1818–1830, Oct. 2022, doi: 10.1109/JAS.2022.105602.

[24] Y. Zhang, H. Liu, Y. Yang, X. Fan, S. Kwong, and C. C. J. Kuo, "Deep learning based just noticeable difference and perceptual quality prediction models for compressed video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1197–1212, Mar. 2022, doi: 10.1109/TCSVT.2021.3076224.

[25] X. Li, H. Shao, S. Lu, J. Xiang, and B. Cai, "Highly efficient fault diagnosis of rotating machinery under time-varying speeds using LSISMM and small infrared thermal images," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 12, pp. 7328–7340, Dec. 2022, doi: 10.1109/TSMC.2022.3151185.

[26] H. Shao, W. Li, B. Cai, J. Wan, Y. Xiao, and S. Yan, "Dual-threshold attention-guided GAN and limited infrared thermal images for rotating machinery fault diagnosis under speed fluctuation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 9, pp. 9933–9942, Sep. 2023, doi: 10.1109/TII.2022.3232766.

[27] K. Song, J. Wang, Y. Bao, L. Huang, and Y. Yan, "A novel visible-depth-thermal image dataset of salient object detection for robotic visual perception," *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 3, pp. 1558–1569, Jun. 2023, doi: 10.1109/TMECH.2022.3215909.

[28] L. Chen, S. Lin, X. Lu, D. Cao, H. Wu, C. Guo, C. Liu, and F.-Y. Wang, "Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3234–3246, Jun. 2021, doi: 10.1109/TITS.2020.2993926.

[29] K. Dasgupta, A. Das, S. Das, U. Bhattacharya, and S. Yogamani, "Spatio-contextual deep network-based multimodal pedestrian detection for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15940–15950, Sep. 2022, doi: 10.1109/TITS.2022.3146575.

[30] Y. He, C. Zhu, and X.-C. Yin, "Occluded pedestrian detection via distribution-based mutual-supervised feature learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10514–10529, Aug. 2022, doi: 10.1109/TITS.2021.3094800.

[31] Q. Guo, J. Liang, and H. Wang, "Night vision anti-halation algorithm of different-source image fusion based on low-frequency sequence generation," *Mathematics*, vol. 11, no. 10, pp. 2237–2239, May 2023, doi: 10.3390/math11102237.

[32] F. Luo, Y. Li, G. Zeng, P. Peng, G. Wang, and Y. Li, "Thermal infrared image colorization for nighttime driving scenes with top-down guided attention," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15808–15823, Sep. 2022, doi: 10.1109/TITS.2022.3145476.

[33] Y. Lang, Y. Qian, F. Shi, H. Cheng, and X. Kong, "An uncooled infrared and ICMOS image fusion system based on FPGA," *Proc. SPIE*, vol. 11763, no. 2, pp. 2651–2659, Jul. 2021, doi: 10.1117/12.2588329.

[34] H. Xu, M. Guo, N. Nedjah, J. Zhang, and P. Li, "Vehicle and pedestrian detection algorithm based on lightweight YOLOv3-promote and semi-precision acceleration," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19760–19771, Oct. 2022, doi: 10.1109/TITS.2021.3137253.

[35] A. Abdelmutalab and C. Wang, "Pedestrian detection using MB-CSP model and boosted identity aware non-maximum suppression," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24454–24463, Dec. 2022, doi: 10.1109/TITS.2022.3196854.

[36] B. Bakker, B. Zablocki, A. Baker, V. Riethmeister, B. Marx, G. Iyer, A. Anund, and C. Ahlström, "A multi-stage, multi-feature machine learning approach to detect driver sleepiness in naturalistic road driving conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4791–4800, May 2022, doi: 10.1109/TITS.2021.3090272.

XIANG LYU was born in Henan, China, in 1986. He received the B.S. degree in computer science and technology from Nanyang Normal University, Nanyang, China, in 2010, and the M.S. degree in computer technology engineering from Zhengzhou University, Zhengzhou, China, in 2013. He is currently pursuing the Ph.D. degree in management with SEGi University, Malaysia.

From 2013 to 2015, he was a Software Development Management Engineer with China Railway Engineering Equipment Group. Since 2015, he has been a Lecturer with the School of Information Engineering, Xinyang Agriculture and Forestry University, Xinyang, China. He is the author of one book, more than ten articles, five applied invention patents, and more than 20 software copyrights. His research interests include graphic image recognition, algorithm analysis, and smart cities.

NAN WANG was born in Henan, China, in 1993. She received the B.S. degree in art and design from Xinyang Normal University, Xinyang, China, in 2015, and the M.S. degree in design from Jilin University, Changchun, China, in 2018.

Since 2019, she has been a Teaching Assistant with the School of Information Engineering, Xinyang Agriculture and Forestry University, Xinyang. She has published a total of seven articles, including one core article. Her research interests include graphic and image processing, UI design, and interface design.

JIA GAO was born in Yunnan, China, in 1987. He received the B.S. degree in management from Yunnan Normal University, China, in 2011, and the M.S. degree in business administration from the Panyapiwat Institute of Management, Bangkok, Thailand, in 2021. He is currently pursuing the Ph.D. degree in management with SEGi University, Malaysia.

From 2011 to 2019, he held senior managerial positions with China Eastern Airlines. Since 2019, he has been an Assistant Lecturer with Yunnan Technology and Business University. He has published over six research articles. His research interests include tourism, sustainability, smart cities, education management, and marketing.

• • •