

Road Image Enhancement and Real-Time Vehicle Detection at Night Based on Convolutional Neural Network

Tao Dong*

School of Information Engineering, Liaodong University, Dandong 118000, China

In the era of artificial intelligence, machine vision technology has been widely used in security, traffic and other contexts, but the image taken at night by machine vision technology has a low detection effect. Therefore, this paper proposes a nighttime road image enhancement and real-time vehicle detection model based on a convolutional neural network (CNN). For the model, the Retinex theory is applied to build an end-to-end image enhancement structure, and inputs the enhanced image as data into a D-CNN recognition model, which optimizes the Softmax classifier. This approach greatly improves the enhanced image recognition effect. The experimental results show that with the proposed image enhancement model, the MSE value of the proposed R-ETELIES algorithm is about 500, the performance is the best among the comparison algorithms, and the processing speed is less than 0.1s when processing high-resolution images. Among the recognition and detection models, the recognition rate of the D-CNN model reaches 96%, while the false detection rate is only 5%. The model constructed in this study still has good detection performance in a low-light environment, which has an important influence on the construction of a smart city.

Keywords: convolutional neural network; low light image; image enhancement; linear discriminant analysis; inspection of vehicle

INTRODUCTION

Despite the development of machine vision technology, the main problem today is the image captured in a low-light environment. When the scene is clearly visible, if there is insufficient light, the captured image is dim and fuzzy, has poor contrast and other problems, and subsequently has a serious impact on the field of image recognition. Therefore, this study explores image enhancement and recognition technology. Common image enhancement techniques include Global Light Awareness and Detail Retention Network, GLADNet) and Multi-Scale Retinex(MSR) [1]. However, GLADNet has certain limitations in terms of image enhancement. The MSR-net network cannot eliminate noise, and image restoration is often accompanied by the block effect [2]. Therefore, an

end-to-end image enhancement structure is constructed based on the Retinex theory. In the traditional image detection model, the classification effect of the Softmax classifier is poor for the sample data as it has incorrect classification or large noise, which causes the loss rate curve of CNN to fluctuate greatly [3]. Therefore, Linear Discriminant Analysis (LDA) is introduced to optimize the Softmax classifier. The aim of this study is to further optimize the enhancement technology applied to solve the problem of low-light image and improve the rate of image recognition.

1. RELATED WORK

At present, machine vision technology is being widely used in the construction of smart cities. In regard to this technology,

*Email of corresponding author: dongtao202200@yeah.net

many scholars are focusing on CNN. Wang K. N. et al. proposed a multi-modal method in their study of facial expression recognition, and adopted the emotional arousal - valence model and CNN. This method uses CNN to identify modal data and build the relevant model. After a lot of learning of modal and affective data, experiments were carried out, with the experimental results showing that the accuracy rate of the recognition system reached 67.8% in titer recognition and 77.0% in arousal recognition [4]. Garland J's team used CNN to identify myocardial infarction in tissue sections, and used 150 myocardial images for training, comprising normal myocardial infarction, acute myocardial infarction and old myocardial infarction images. The experimental results indicated that the recognition accuracy of the CNN model for the three kinds of myocardium reached 95%, which promoted the research of forensic medicine and postmortem pathology on the central nervous system [5]. Qin S. and other researchers developed a graphical convolutional neural network, which encodes the structure of surfactants in the form of molecular diagrams and uses Critical Micelle Concentration(CMC) data for training. According to the analysis of experimental results, the prediction accuracy of CMCs has been greatly improved, and the physical rules related to the influence of surfactant molecules on CMCs are reflected [6]. Patil S. P. and other researchers propose to improve the efficiency of image and video forgery detection by using hybrid cellular neural networks, which are incremental and enhanced during the learning phase and finally detected by a hybrid CNN model. Experimental results show that the detection accuracy of this method for different types of forged videos reaches 98%, which solves the problems of the original CNN detection model [7]. Pan D. et al. proposed an early diagnosis model for AD that combined 3D convolutional neural network with genetic algorithm. The model trained a classifier for each ROI, combined the classifiers with a genetic algorithm, and diagnosed patients using a combination of classifiers. Experimental results showed that the accuracy rate of early diagnosis of AD reached 88.6%, and 88.1% of Mild Cognitive Impairment(MCI) patients would be converted to AD [8].

In machine vision, image enhancement and real-time vehicle detection technology has also attracted much interest. Gassenmaier S. et al. studied abdominal magnetic resonance imaging by combining image enhancement technology with new iterative denoising technology. Two doctors conducted experiments to verify the proposed method, and the results showed that the method provided better image quality and diagnostic reliability [9]. Li M. and other researchers combined different types of networks to build a multi-module feature extraction model, and used Retinex theory to extract the reflection map of the image so as to preserve image details and remove noise as much as possible while improving the contrast. Experiments show that this method has good performance in terms of both the quantitative and qualitative features of images [10]. Yang H. H. and other researchers proposed an image enhancement model based on deep learning architecture which uses multi-scale grid convolution network to learn different types of information. A comprehensive experiment was carried out for the benchmark image and the underwater image, and the experiment proved the effectiveness of the model, which had a good enhancement

effect on the underwater image and improved its quality [11]. Chen G.'s team studied vehicle detection by using the learning method of pseudo image input combined with 2D Lidar. Experimental results show that this method has good performance in terms of accuracy. The mixed learning approach adopted in this method provides superior performance of a lightweight model and solves the problems associated with unstable and expensive traditional detection [12]. Han Z.'s team proposed an M2R-Net accuracy detector, which used a multi-layer feature fusion strategy to extract a feature map, adopted a balanced sampling strategy and global balance loss function to address the imbalance between space and target, and finally used a rotary anchor to predict *moving* vehicles. Through experimental analysis, this method has advanced detection performance and enriches the methods of daily traffic monitoring [13]. Shu M. and other scholars proposed the E-GMMTFD model to detect moving vehicles. In the proposed method, a local contrast enhancement method was adopted to improve the contrast between the target and the background, and GMM and TFD were used to detect moving vehicles. Experiments proved that this method can effectively detect moving vehicles in satellite videos and provide technical support to relieve urban traffic congestion [14].

In summary, image enhancement technology can improve the accuracy of image recognition, and has been applied in vehicle detection. However, there is still room for improvement in the recognition rate of the traditional CNN detection model. Therefore, this study summarized the above methods and constructed a composite vehicle detection model, which optimized the image enhancement technology and recognition and classification technology, aiming at improving the accuracy of detecting vehicles at night.

2. NIGHTTIME ROAD IMAGE ENHANCEMENT AND VEHICLE REAL-TIME DETECTION MODEL CONSTRUCTION BASED ON CONVOLUTIONAL NEURAL NETWORK

2.1 Low light Road Image Enhancement Based on Convolutional Neural Network

In the construction of a smart city, in order to accurately and efficiently identify low-light road images at night, image enhancement processing is needed. The traditional low illuminance image (LII) enhancement algorithm can improve the local image to a certain extent, but its robustness is weak [15–17]. The physical model method can enhance the image, but it has some requirements in terms of the applicable scene and has several limitations. LII based on CNN uses a neural network structure and a large amount of data for training, in order for LII to have image-enhancement capabilities. The common enhancement methods of LII are GLADNet and MSR-net. GLADNet takes LII as the input item, uses the codec network to estimate the reflectivity map, and finally

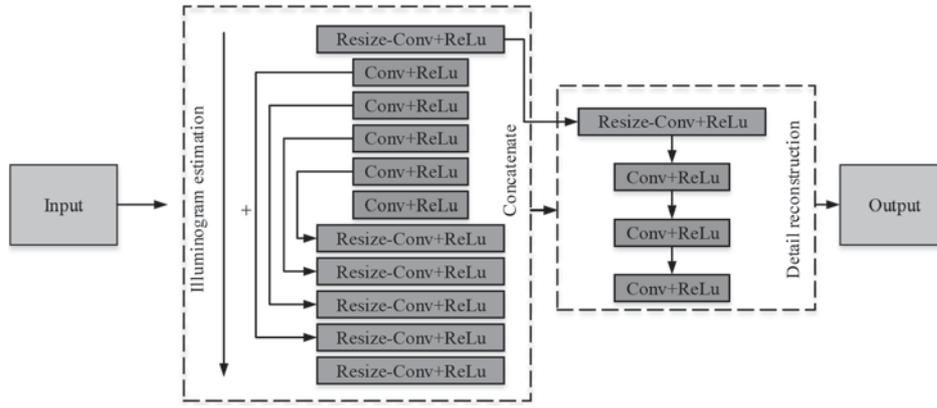


Figure 1 GLADNet Network structure diagram.

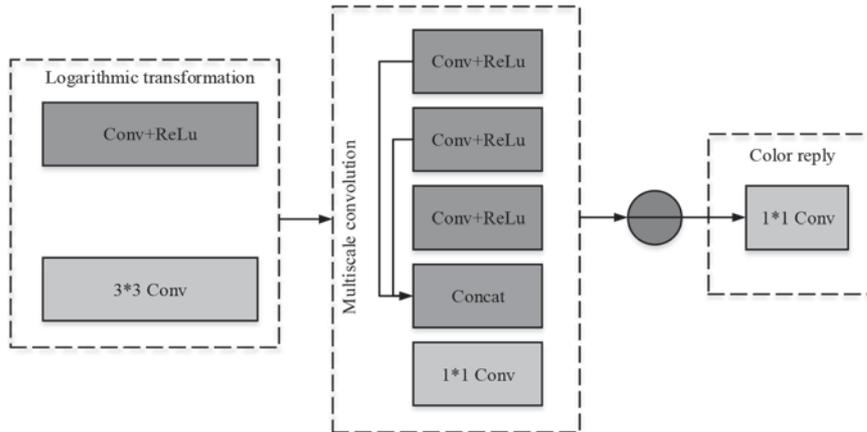


Figure 2 MSR-net model structure.

uses the full convolution layer to restore the image details, thus obtaining the normal light image. The network structure of GLADNet is shown in Figure 1.

As shown in Figure 1, the GLADNet network structure is divided into two main parts: light map estimation and detail restoration. In the light map estimation part, the image is uniformly scaled to facilitate subsequent image processing, and then the scaled image is encoded until the size of the feature map is consistent with the size of the convolution kernel. Finally, the decoder makes the model learn the residual forcibly through the method of residual connection, and speeds up the training process. The network structure minimizes the loss between the enhanced image and the original image; the loss function of the network is shown in Formula (1).

$$Loss(X, Y) = \frac{\sum_{i=1}^N \|F(X_i, \Theta) - Y_i\|}{N} \quad (1)$$

In Formula (1), $F(X_i, \Theta)$ represents the enhanced image; Y represents the original image; N represents the number of samples involved in training; $\| \cdot \|$ is the norm of the loss function. MSR-net is based on the structure constructed by MSR theory, and the MSR algorithm is a branch of the Retinex theory. Its derivation process is shown in Formula (2).

$$\begin{cases} I(x, y) = R(x, y) \cdot L(x, y) \\ r(x, y) = \log(R(x, y)) = \log(I(x, y)) - \log(L(x, y)) \end{cases} \quad (2)$$

In Formula (2), $I(x, y)$ represents the image; $R(x, y)$ represents the reflection estimation of the image; $L(x, y)$ represents the brightness estimate of prominence; $r(x, y)$ represents the original reflection component. By solving the formula to build MSR-net, image conversion into the logarithmic domain can be carried out by the convolution operation. The MSR-net model structure is shown in Figure 2.

In Figure 2, due to the color distortion in the image processed by MSR, a 1×1 convolution kernel is added at the end of the model structure to eliminate the overall gray problem of the image. The above GLADNet intelligence enhances the LII of some scenes that have poor robustness and color bias in image color restoration. However, the MSR-net network has no anti-noise ability, and image restoration is often accompanied by the block effect. Therefore, an End-to-End Low Illuminance Image Enhancement Structure based on Retinex (R-ETELIIES) is proposed in this study. Its structure is shown in Figure 3.

Applying the Retinex theory, the image is decomposed by CNN, and the reflectance map and the illumination map are obtained. In the CNN model, first, odd convolution check images are used for shallow feature extraction, whereby the convolution kernel size types are divided into three types: 3×3 , 5×5 and 7×7 . Three different convolution kernels were used

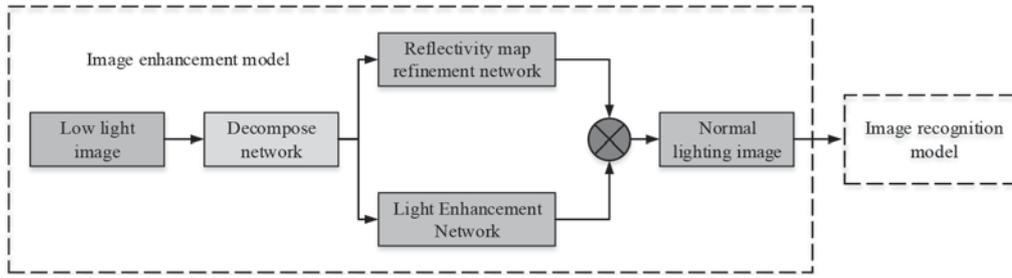


Figure 3 End-to-end low illumination image enhancement network architecture.

to extract different feature information, and a Depthconcat layer was used to combine them; that is, multi-feature extraction technology was used. Then, two 3*3 convolution checks were used to refine the extracted features, and the reflectance map and illumination map were obtained. Finally, a convolution kernel of the same size was used for output. In the shallow feature extraction and convolution operation, the appropriate activation function has an important influence on the convolution process. In the study of the first layer of convolution, ReLu activation function is used, expressed in Formula (3).

$$\text{ReLu}(x) = \max(x, 0) = \begin{cases} 0 \\ x \end{cases} \quad (3)$$

When the value of x is greater than 0, the ReLu function is x itself. If the value of x is not greater than 0, the ReLu function is 0. In the middle layer of the network structure, the number of convolution cores is large, and the data needs to be standardized after the convolution operation. In this case, PReLU is used as the activation function. The expression of PReLU activation function is shown in Formula (4).

$$\text{PReLU}(x_i) = \begin{cases} x_i \\ a_i x_i \end{cases} \quad (4)$$

In Formula (4), a_i represents the constant of channel function i . When the value of x_i is greater than 0, the value of PReLU is x_i . When x_i is not greater than 0, the value of PReLU is $a_i x_i$. If a_i is 0, the PReLU function becomes ReLu function. In the last layer of convolution, the Sigmoid activation function was used to normalize the output data, among which three channels were the reflectivity map data and the light map data. In the same figure, regardless of whether the image is captured under normal light or low light, the reflectivity graph is the same. Therefore, the reflectivity non-deformation loss function can be designated as L_r . The loss function of light pattern is L_l ; Image decomposition and fusion are opposite processes. In order to obtain corresponding and correct images, a reconstruction loss function is constructed, which is set as L_s . The loss function expression of the decomposed network in the model is shown in Formula (5).

$$\begin{cases} L = L_s + \lambda_r L_r + \lambda_l L_l \\ L_s = \sum_{i=low,normal} \sum_{j=low,normal} \|R_i \cdot I_j - S_j\|_2 \\ L_r = \|R_{low} - R_{normal}\|_2 \\ L_l = \sum_{i=low,normal} \|\nabla L_i \cdot e^{-\lambda \nabla R_i}\|_2 \end{cases} \quad (5)$$

In Formula (5), λ_r represents the constancy coefficient of the equilibrium reflectance map, and its value is 0.01. λ_l represents the local smoothness characteristic coefficient of the light image, and its value is 0.1. $e^{-\lambda \nabla R_i}$ represents the constraint of smoothness; ∇ is the gradient in the function. The loss function in the light map enhancement network is shown in Formula (6).

$$\begin{cases} L_A = L_1 + \lambda_1 L_2 + \lambda_2 L_3 \\ L_1 = \sum_{i=low,normal} \|R_i \cdot L_{low} - S_{normal}\|_2 \\ L_2 = \sum_{i=low,normal} \|\nabla L_{low} \cdot e^{-\lambda \nabla R_i}\|_2 \\ L_3 = \|L_{low} - L_{normal}\|_2 \end{cases} \quad (6)$$

In Formula (6), L_A represents the total loss function of the illumination image enhancement network; L_1 represents the reconstruction loss function, and its value is 0.1. L_2 represents the loss function of the local uniformity of the illumination image and the design of the structure perception characteristics, and its value is 0.01. L_3 represents the loss function that makes the normal light image consistent with the enhanced light image. The image captured in the night scene will be affected by the surrounding light, so that the color of the image will be biased towards the ambient light, resulting in color distortion. At the same time, when the low-light image is decomposed, artificial constraints will be imposed, and the details and noise in the light image will be decomposed into the reflectivity image. In order to solve these two problems, the residual learning strategy is used in the refinement process. Its loss function is shown in Formula (7).

$$\begin{cases} P = R_{normal} - R_{low} \\ L_B = \|P - P'\|_2 \end{cases} \quad (7)$$

In Formula (7), P represents the true residual; R_{normal} represents the reflectance of normal light; R_{low} represents low light reflectance map; P' represents the residual of network learning, and the model is finally fused to obtain low-light enhanced images.

2.2 Construction of Vehicle Identification Model Based on D-CNN

As Figure 3 shows, after the low-light image is enhanced, its data is transmitted to the recognition model for recognition. During the training of the conventional CNN recognition

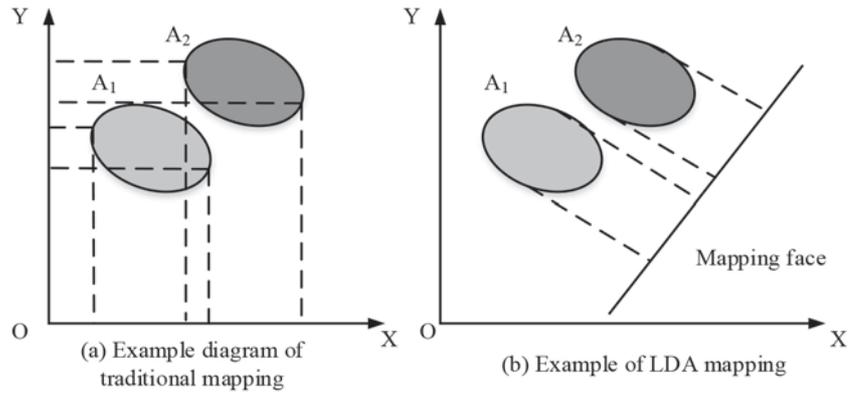


Figure 4 Example of LDA mapping.

model, the loss rate curve will fluctuate greatly. Therefore, in the accuracy test, the correct rate curve will also fluctuate greatly. To solve this problem, Linear Discriminant Analysis (LDA) is introduced into the CNN model to obtain a depth convolution neural network (D-CNN). The Softmax classifier of the model was optimized. The core idea of LDA is to transform a high-dimensional sample feature into a vector space by using linear or nonlinear mapping, and then classify the samples in the vector space [18–20]. In the transformation process, the basic principle should follow that the transformed low sample features have the maximum inter-class distance and the minimum intra-class distance [21–22]. Now assume that there is a spatial dimension, in which there are m samples, expressed as $\{x_1, x_2, \dots, x_m\}$. The coefficients in the LDA algorithm are now defined; the expression is shown as Formula (8).

$$\begin{cases} u_i = \frac{\sum_{x \in \text{class } i} x}{n_i} \\ u = \frac{\sum_{i=1}^m x_i}{m} \end{cases} \quad (8)$$

In Formula (8), u_i represents the mean value of i samples; u is the mean of all samples; n_i represents the dimensions of the sample. According to the definition of inter-class distance matrix and intra-class distance matrix, the expression of these matrices is shown in Formula (9).

$$\begin{cases} S_b = \sum_{i=1}^c n_i (u_i - u)^{T+1} \\ S_w = \sum_{i=1}^c \sum_{x_k \in \text{class } i} (u_i - x_k)^{T+1} \end{cases} \quad (9)$$

In Formula (9), S_b represents the inter-class dispersion matrix; S_w represents the dispersion matrix within the class; c indicates the sample type. Formula (9) is essentially a covariance matrix, representing the relationship between samples and total samples. The diagonal function of the matrix represents the dispersion degree of the sample, while the non-diagonal elements represent the correlation degree or redundancy degree between the sample and the overall sample. Therefore, Formula (9) calculates the sum of the covariance matrix of the sample corresponding to the population, and describes the degree of discrete redundancy between the sample class and the population from a macro perspective. By introducing Fisher's discrimination criterion

into the dispersion formula, the model cost function can be obtained; its expression is shown in Formula (10).

$$J_{fisher}(\varphi) = (\varphi^T S_b \varphi) / (\varphi^T S_w \varphi) \quad (10)$$

In Formula (10), φ represents any column vector. When the value of $J_{fisher}(\varphi)$ is the maximum, the mapping vector of the model is φ , which means that the intra-class dispersion of the sample is kept at the lowest level, while the inter-class dispersion is kept at the maximum level. Figure 4 depicts an LDA mapping diagram.

In Figure 4 (a), the mapping between the two sample sets will result in a large number of overlaps, resulting in poor classification. In Figure 4 (b), the LDA algorithm is used to find a new mapping surface, and there is no intersection in the mapping, thus improving the classification effect. When solving the LDA algorithm, the concept of middle distance is introduced into Formula (8), then the inter-class distance matrices are and intra-class distance matrices are shown in Formula (11).

$$\begin{cases} D_1 = (x_i - u_i)^{T+1} \\ D_2 = \sum_{j=1}^k (x_i - u_j)^{T+1} \end{cases} \quad (11)$$

In Formula (11), D_1 is the mean distance between the sample and the category; D_2 is the sum of the mean distances from the sample to the classes. Convert Formula (11) into Softmax function to obtain Formula (12).

$$p(y^{(1)} = j | x^{(i)}; \theta) = (\theta_j^T (x^{(i)} - u_j)^{T+1}) / \left(\sum_{l=1}^k \theta_l^T (x^{(i)} - u_l)^{T+1} \right) \quad (12)$$

In Formula (12), θ represents the model parameters and p represents the probability of occurrence. In Formula (12), when the mean distance between samples and a certain class is greater than the sum of the mean distance between samples and all classes, and when the value of p is the minimum, this indicates that the distance between samples and this class is the minimum, indicating that the samples belong to this type. The cost function in the new algorithm is changed to Formula (13).

$$J_\theta = \frac{\left[\sum_{i=1}^m \sum_{j=1}^k \{y^{(1)} = j\} (\theta_j^T (x^{(i)} - u_j)^{T+1}) / \left(\sum_{l=1}^k \theta_l^T (x^{(i)} - u_l)^{T+1} \right) \right]}{m} \quad (13)$$

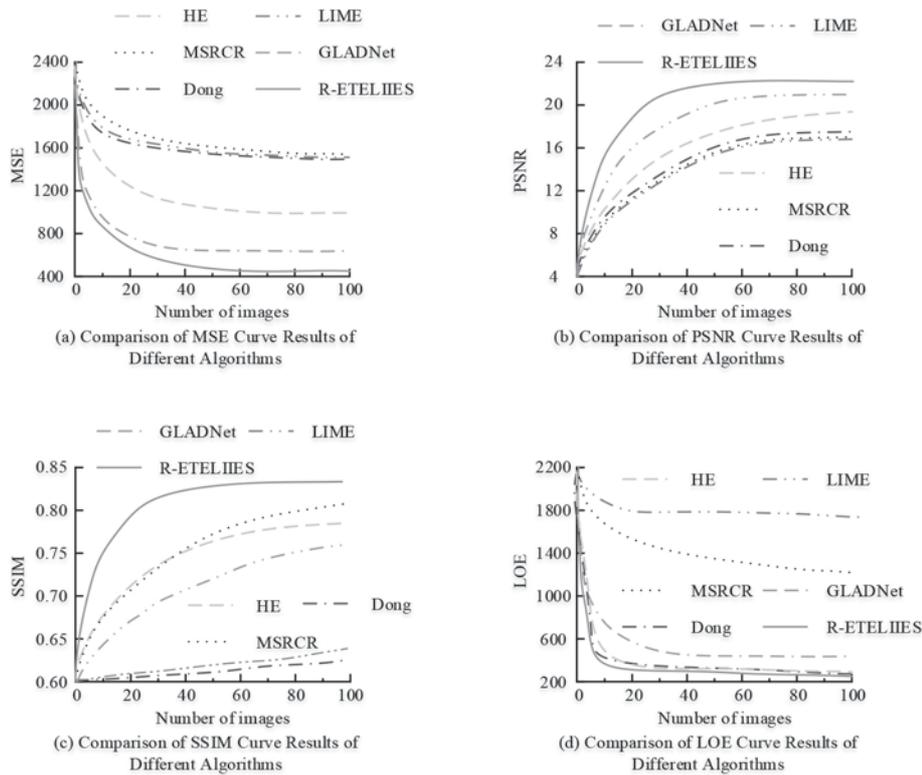


Figure 5 Comparison of results for different algorithms in each evaluation index.

By differentiating Formula (13), the descending gradient formula of model parameters can be obtained; its expression is shown in Formula (14).

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left\{ \frac{D_j}{\sum_{l=1}^k D_l \theta_l^T} \left[1 \left\{ y^{(1)} = j \right\} - p(y^{(1)} = j | x^{(1)}; \theta) \right] \right\} + \lambda \theta_j \quad (14)$$

If Formula (14) is introduced into the gradient algorithm, the time cost function can be minimized, so as to complete the improvement effect of the classifier.

3. NIGHT ROAD IMAGE ENHANCEMENT AND VEHICLE REAL-TIME DETECTION PERFORMANCE ANALYSIS BASED ON CONVOLUTIONAL NEURAL NETWORK

3.1 Performance Analysis of Low-Light Road Image Enhancement Based on Convolutional Neural Network

The evaluation of the image quality produced by the image enhancement model is an important step. In the experiment, HE, MSRCR, Dong, LIME, GLADNet and R-ETELIIES algorithms were used for comparison. Its performance indicators are Peak Signal to Noise Ratio (PSNR), Structural similarity

(SSIM), Mean square error, MSE) and lightness order error (LOE). The performance of the algorithm was evaluated by the Natural Image Quality Evaluator (NIQE) and Visual Information Fidelity (VIF) scores for different data sets.

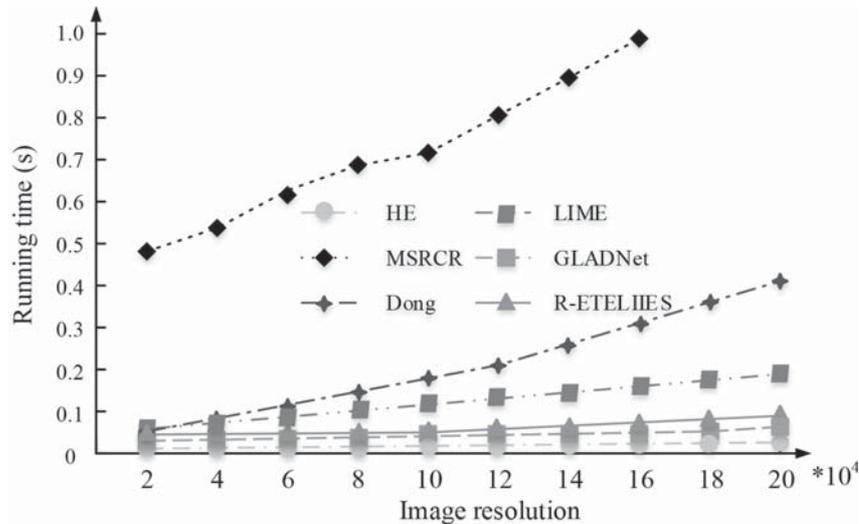
The results for different algorithms in each evaluation index are shown in Figure 5. Figure 5 (a) is the comparison result of MSE value of the algorithm. MSE represents the distance between images. The smaller the value, the smaller the gap between the image and the normal light image. Compared with other algorithms, the R-ETELIIES algorithm proposed in this study has the lowest MSE value of about 500. Figure 5 (b) shows the comparison result of the algorithm's PSNR value. The PSNR indicates the extent of the distortion between the enhanced image and the normal illumination image. The smaller is the degree of distortion, the larger is the PSNR value. The PSNR value of the proposed algorithm is about 22.8dB, which is the highest compared with that of other algorithms. Figure 5 (c) shows the comparison result for the algorithm's SSIM value. The SSIM represents the integrity of the image structure information. The SSIM value of the proposed algorithm is about 0.82, which is higher than that of all other compared algorithms. Figure 5 (d) is the comparison result for LOE values of the algorithm. The LOE value shows the extent to which the naturalness of the image has been preserved. The smaller the value, the better is the brightness sequence of the image, and the greater the naturalness of the image. The LOE value of the proposed algorithm is about 289, which is the lowest LOE value compared with that of other algorithms. Therefore, the R-ETELIIES algorithm proposed in this study has good image enhancement performance, and the enhanced image is the closest to the normal illumination image.

Table 1 Comparison results of NIQE scores of different algorithms in the public dataset.

Data set	HE	MSRCR	SRIE	Dong	LIME	GLADNet	R-ETELIIES
DICM	3.5684	3.2380	3.1783	3.9203	3.3878	3.0035	2.6904
Fusion	3.5345	3.3742	3.4251	3.4856	3.6552	3.3170	3.1897
MEF	4.2657	3.9503	4.0692	4.5841	4.3355	3.5788	3.3135

Table 2 Comparison results of VIF scores of different algorithms in the public dataset.

Data set	HE	MSRCR	SRIE	Dong	LIME	GLADNet	R-ETELIIES
DICM	0.4617	0.2659	0.4917	0.3992	0.1509	0.6390	0.7393
Fusion	0.7010	0.4790	0.6054	0.4000	0.2604	0.7416	0.9795
MEF	0.2698	0.1958	0.4272	0.2616	0.1371	0.4551	0.6861

**Figure 6** Processing speed of images with different resolutions in different algorithms.

The distance between the feature model parameters of the resulting image and the trained model parameters is calculated, and the NIQE is obtained by measuring the distance. The NIQE scores of each algorithm for different data sets are shown in Table 1. In Table 1, the R-ETELIIES algorithm proposed in the study has the lowest score of 2.6904 in the DICM data set compared with other algorithms. In the Fusion data set, the NIQE score of other algorithms is all greater than 3.2 points, while the score of the algorithm proposed in this study is less than 3.2, and its score is 3.1897. In the MEF data set, the NIQE score of R-ETELIIES algorithm is 3.3135 points, and the scores of other algorithms are all higher than that of R-ETELIIES algorithm. Therefore, according to the analysis of QINE index, in different data sets, the quality of the image generated by the training model of the R-ETELIIES algorithm is better.

The VIF score results for each algorithm used on different data sets are shown in Table 2. The higher VIF score, the better is the image quality. In Table 2, the R-ETELIIES algorithm proposed in the study achieved a score of 0.7393 for the DICM data set, the highest among all other compared algorithms. For the Fusion data set, the VIF scores of other algorithms are all less than 0.9. Only the algorithm proposed in this study achieved a score greater than 0.9, obtaining 0.9795. For the MEF data set, the VIF score of the proposed R-ETELIIES algorithm is 0.6861, which is higher than the scores obtained

by all the other algorithms. Therefore, the analysis of VIF index shows that the quality of the image generated by the training model in the R-ETELIIES algorithm is better in all the different data sets.

The image processing speed of different algorithms is also an important criterion for the performance of the algorithm. Therefore, Figure 6 shows the comparison of the processing speed results of different algorithms with different resolutions. In Figure 6, the MSRCR algorithm takes a long time to process the image with a resolution of 160,000. The HE algorithm is the fastest in terms of processing speed, but the image processing effect is poor. The processing speed of the Dong, GLADNet and R-ETELIIES algorithms is positively correlated with the image size, but the processing speed of the Dong algorithm is slower. The LIME algorithm over-intensifies the locally normally-lit areas. Comprehensive analysis results indicate that the R-ETELIIES algorithm can still meet the real-time requirements of processing large images, and can better retain the original color of the image.

3.2 Effect Analysis of Vehicle Recognition Model Based on D-CNN

After the performance of an image enhancement algorithm is established, the verification of its recognition and

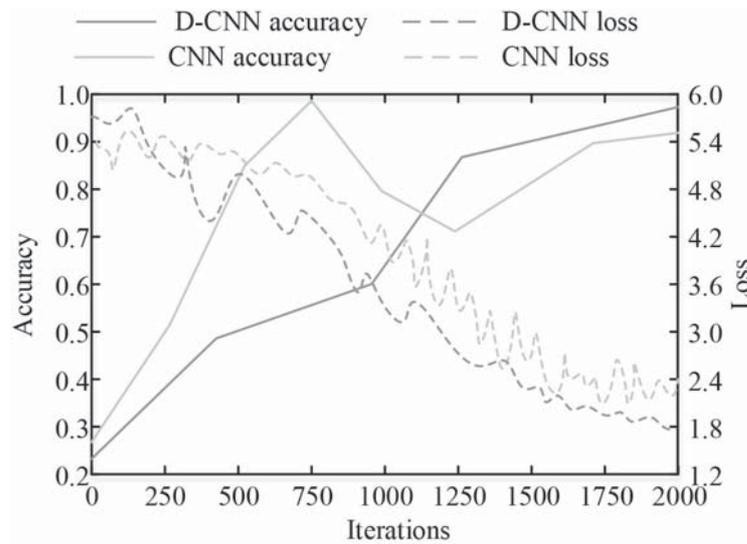


Figure 7 Comparison of Loss and Accuracy Curves between D-CNN and CNN.

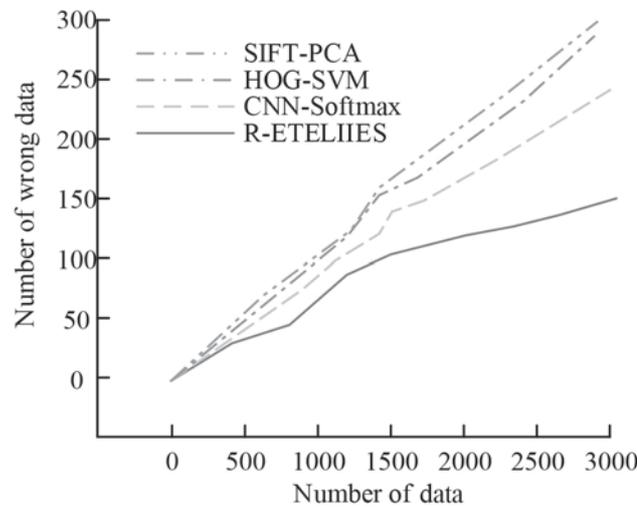


Figure 8 Error detection curves of different algorithms.

classification performance is also important. The Softmax classifier of the CNN model is replaced by a new algorithm, and the model is trained by a new connection. In order to reduce the fluctuation of the Loss rate curve during CNN training, the model performance was determined by the loss curve and the test accuracy curve. When different algorithms are compared, training time, test time and accuracy rate are used as the criteria for performance.

As seen in Figure 7, after 2000 iterations of the D-CNN model, the test accuracy reaches about 96% and the loss rate is about 0.18. After 2000 iterations, the test accuracy of the CNN model is about 91%, and the loss rate is about 0.24. Compared with CNN model, the test accuracy of the D-CNN model is increased by about 5%, and the loss rate is decreased by about 0.06. According to the local analysis, the prediction accuracy curve fluctuates greatly when the CNN model iterates 800 times. When the number of iterations is 1000, the loss rate curve of the CNN model also fluctuates greatly. The curve of the D-CNN model is relatively smooth and stable. Therefore, the D-CNN algorithm is more stable according to the indexes for test accuracy and training loss rate.

The CNN-Softmax model, SIFT-PCA model, HOG-SVM model and D-CNN model were selected for comparison. The vehicle false detection results obtained by the four algorithms are shown in Figure 8. In Figure 8, the false detection rate of the four algorithms is positively correlated with the number of samples. The greater the number of samples, the higher is the false detection rate. Among the models, the Sift-PCA has the highest false detection rate. When the number of samples is 3000, the false detection rate of this model reaches about 10%. When the number of detected samples is 3000, the false detection rate of the HOG-Based model is about 9.5%. The error detection rate of the CNN-Softmax model is about 8%. The false detection rate of the D-CNN model is about 5%. Compared with the other three models, the D-CNN model has the lowest false detection rate. The results show that the D-CNN model can still maintain a high detection rate when there is a large amount of data.

The comparison results for training time, test time and accuracy of each algorithm on the test set are shown in Figure 9. In Figure 9 (a), CNN-Softmax has the shortest training time of 2.4min. The training time of D-CNN was

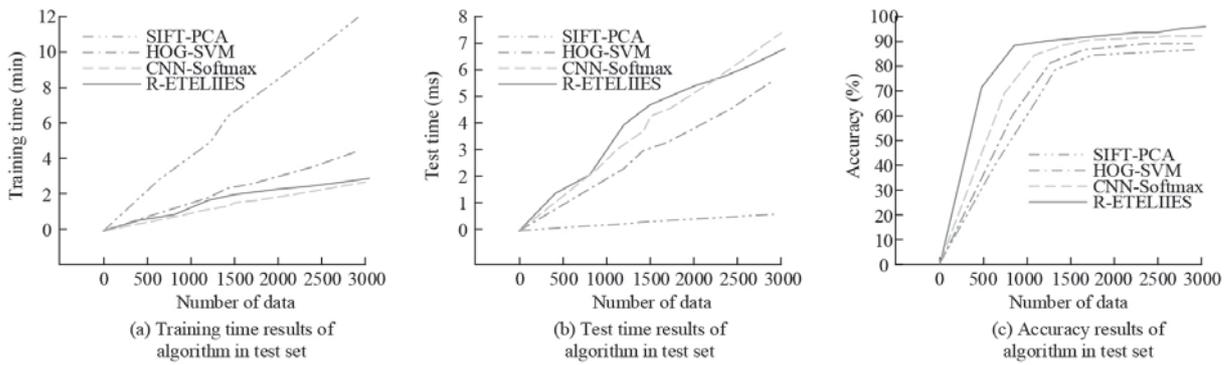


Figure 9 Comparison results of prediction time and accuracy of each algorithm.

0.2min longer than that of CNN_Softmax, and the training time of the D-CNN model was 2.6min. As shown in Figure 9 (b), the test time of SIFT-PCA is the fastest, which is 0.5ms. The test time of the D-CNN model is 6.9 ms. Figure 9 (c) indicates the recognition and classification accuracy of the D-CNN model reaches 95.8%. The accuracy rate of CNN-Softmax is 92.7%; The accuracy of SIFT-PCA model and HOG-SVM model is 88.6% and 89.4%, respectively. Comprehensive analysis shows that the HOG-SVM model has a good realization in the performance of training time and test time, but the accuracy is low. The SIFT-PCA model has excellent performance in terms of test time, but poor performance in training time and accuracy. Compared with the CNN_Softmax model, the D-CNN model shows excellent accuracy when the time performance gap is not large. Therefore, the D-CNN model has the best overall performance.

4. CONCLUSION

At night, due to the low amount of light, the photosensitive devices on the camera cannot adequately capture the scene, resulting in blurred images with low contrast, which greatly limits the application and development of the imaging system. Therefore, this study proposes an image enhancement under low light conditions and a vehicle detection method based on a convolutional neural network. The study adopts the end-to-end low-light image enhancement method. In the vehicle detection model, the D-CNN algorithm is used to optimize the Softmax classifier. The experimental results show that the MSE and LOE values of the R-ETELIIES algorithm are about 500 and 298 respectively, and the PSNR and SSIM values are about 22.8dB and 0.82, respectively. R-ETELIIES algorithm has better image enhancement performance. In the recognition model, the D-CNN model achieves 96% accuracy, the loss rate is about 0.18, and the false detection rate is only 5%. Moreover, compared with other algorithms, the D-CNN has better overall performance. The research has achieved good results in the experiment, but there are still some shortcomings. In an environment that has very weak light, the image enhancement needs to be further improved, and the system running time also has room for improvement. Further research will be carried out to address these shortcomings and further improve the system's performance.

FUNDINGS

The research is supported by: Natural Science Foundation of Liaoning Province in the year of 2022 (Joint Open Fund from Key Science and Technology Innovation Base): Research on Road Image Enhancement and Vehicle Driving Area Detection under Adverse Environment (Fund No.: 2022-KF-12-13); University Basic Scientific Research Project of Liaoning Provincial Department of Education in the year of 2023: Research on lightweight Road Environment perception and Driving area Detection (Fund No.: JYTMS20230709).

REFERENCES

- Panetta K., Kezebou L., Oludare V., Sos A. (2022). Comprehensive Underwater Object Tracking Benchmark Dataset and Underwater Image Enhancement with GAN. *IEEE Journal of Oceanic Engineering*, 47(1):59–75.
- Zheng H., Shao Q., Chen J., Shan Y., Qin X., Ma J., Xu X. (2022). LIC color texture enhancement algorithm for ocean vector field data based on HSV color mapping and cumulative distribution function. *Acta Oceanologica Sinica*, 41(10):171–180.
- Li Z., Liu H., Zhao J., Bi T., Yang Q. (2022). A Power System Disturbance Classification Method Robust to PMU Data Quality Issues. *IEEE Transactions on Industrial Informatics*, (18–1).
- Wang K.N., Xie H., Zou D, Chou K.L. (2021). Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Information Fusion*, 77(5): 107–117.
- Garland J., Hu M., Duffy M., Kesha K., Glenn C., Morrow P., Stables S., Ondruschka B., Broi U., Tse R. (2021). Classifying Microscopic Acute and Old Myocardial Infarction Using Convolutional Neural Networks. *American Journal of Forensic Medicine & Pathology*, 42(3):230–234.
- Qin S., Lehn R.V., Zavala V., Jin T. (2021). Predicting Critical Micelle Concentrations for Surfactants using Graph Convolutional Neural Networks. *The Journal of Physical Chemistry B*, 125(37):10610–10620.
- Patil S.P., Jariwala K.N. (2021). Improving the Efficiency of Image and Video Forgery Detection Using Hybrid Convolutional Neural Networks. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 29(1):101–117.
- Pan D., Zou C., Rong H., Zeng A. (2021). [Early diagnosis of Alzheimer's disease based on three-dimensional convolutional neural networks ensemble model combined with genetic

- algorithm]. *Sheng wu yi xue gong cheng xue za zhi = Journal of Biomedical Engineering = Shengwu yixue gongchengxue zazhi*, 38(1):47–55.
9. Gassenmaier S., Afat S., Nickel D., Kannengiesser S., Herrmann J., Hoffmann R., Othman A. (2021). <Image> Application of a Novel Iterative Denoising and Image Enhancement Technique in T1-Weighted Pre-contrast and Postcontrast Gradient Echo Imaging of the Abdomen: Improvement of Image Quality and Diagnostic Confidence. *Investigative Radiology*, 56(5): 328–334.
 10. Li M., Zhou D., Nie R., Xie S., Liu Y. (2021). AMBCR: Low-light image enhancement via attention guided multi-branch construction and Retinex theory. *IET Image Processing*, 15(9):2020–2038.
 11. Yang H.H., Huang K.C., Chen W.T. (2021). LAFFNet: A Lightweight Adaptive Feature Fusion Network for Underwater Image Enhancement. *IET Image Processing*, 15(3): 774–785.
 12. Chen G., Wang F., Qu S., Chen K., Yu J., Liu X., Xiong L., Knoll A. (2021). Pseudo-Image and Sparse Points: Vehicle Detection with 2D LiDAR Revisited by Deep Learning-Based Methods. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7699–7711.
 13. Han Z., Wang C., Fu Q. (2021). M 2 R-Net: deep network for arbitrary oriented vehicle detection in MiniSAR images. *Engineering Computations*, 38(7):2969–2995.
 14. Shu M., Zhong Y., Lv P. (2021). Small moving vehicle detection via local enhancement fusion for satellite video. *International Journal of Remote Sensing*, 42(19/20):7189–7214.
 15. Choi S.W., Park J.H., Kim J.H., Kim Y.H., Song P.K., Shin M.H., Kwon J.D. (2022). Evaluation method of the light-trapping structure for a transparent thin-film silicon solar cell with low-illuminance condition. *Solar Energy*, 231: 1107–1114.
 16. Wang R., Jiang B., Yang C., Li Q., Zhang B. (2022). MAGAN: Unsupervised Low-Light Image Enhancement Guided by Mixed-Attention. *Big Data Mining and Analytics*, 5(2):110–119.
 17. Xie Q., Li D., Yu Z., Zhou J., Wang J. (2020). Detecting Trees in Street Images via Deep Learning with Attention Module. *IEEE Transactions on Instrumentation and Measurement*, 69(8):5395–5406.
 18. Rzecki K., Baran M. (2022). Application of Elastic Shape Analysis to User Authentication and Identification. *IEEE Transactions on Emerging Topics in Computing*, 10(2):1157–1165.
 19. Sessanna R., Iavorivska L., Kelleher C. (2022). Applying multispectral UAV imagery to delineate in and near stream cover along a small urban stream. *River Research and Applications*, 38(4):717–726.
 20. Nguyen A., Nguyen C.L., Gharehbaghi V., Perera R., Brown J., Yu Y., Kalbkhani H. (2022). A computationally efficient crack detection approach based on deep learning assisted by Stockwell transform and linear discriminant analysis. *Structures*, 45:1962–1970.
 21. Lee C.K.H., Wu K.Y.K. (2023) Making autonomous vehicle systems human-like: lessons learned from accident experiences in traffic. *Enterprise Information Systems*, 17(6): 1998641–1998641.
 22. Yadav K., Kariri E., Alotaibi S.D., Viriyasitavat W., Dhiman G., Kaur A. (2023) Privacy protection against attack scenario of federated learning using internet of things. *Enterprise Information Systems*, 17(9): 2101025–2101025.