2025 11th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2025)

Computer Vision System Optimization and Real-time Recognition Technology Based on Deep Learning

Chen Su

Guilin Institute of Information Technology, Guilin, 541004, China

Keywords: Deep Learning; Computer Vision System; Real-Time Identification; Algorithm Optimization; Pattern Plan

Abstract: This article introduces the core position of computer vision technology and the vital importance of DL (Deep Learning) in this field, and expounds the research background and importance of this study. Its goal is to deeply study the application of DL in computer vision system optimization and real-time identification technology, aiming at improving the performance of the system to meet the strict requirements of accuracy and real-time in practical applications. In this p·aper, a neural network architecture is innovatively constructed. Through careful design of model hierarchy, convolution kernel configuration and activation function, the representation and generalization ability of the model are enhanced. Subsequently, this article realizes the algorithm model, and optimizes it by means of parallel processing and model pruning, which effectively reduces the computational complexity and storage requirements. In order to test the effect of the algorithm, the experimental scheme is carefully designed, and the algorithm is comprehensively evaluated through comparative experiments. Experimental data show that the optimization algorithm proposed in this article has achieved significant improvement in accuracy, processing speed and storage occupation, which fully demonstrates the reliability of the algorithm model.

1. Introduction

As an important branch of artificial intelligence, computer vision technology is full of exploration and innovation [1]. It started from simple digital image processing, and gradually developed into the application of feature extraction and machine learning algorithm, and the technology became more and more mature [2]. The rise of DL has brought unprecedented opportunities for progress in the field of computer vision. DL technology imitates the operation mechanism of human brain neural network, can learn and extract advanced features from images independently, and significantly improves the performance of computer vision tasks [3]. At present, DL has become the key technology of computer vision, and promoted the rapid development of image classification, object detection, semantic segmentation and other tasks [4].

Real-time recognition technology is an important aspect of computer vision application, which shows great application potential and value in many industries [5]. In the automatic driving system, real-time identification technology can accurately identify road signs, pedestrians and vehicles, and provide key decision support for the system [6]. In the field of security monitoring, it can quickly identify abnormal behaviors or suspicious individuals and give an alarm in real time [7]. Therefore, the development of real-time identification technology has far-reaching significance for promoting the intelligent transformation of various industries.

The purpose of this study is to deeply discuss the optimization and real-time identification technology of computer vision system based on DL, aiming at improving the accuracy and efficiency of real-time identification technology through the innovation of algorithm model and system optimization. This helps to promote the continuous progress of computer vision technology, and provides a solid technical support for the wide implementation of intelligent applications. The research focus of this article includes the innovation of algorithm model, the optimization of system performance and the improvement of real-time identification technology. The specific research objectives are: to improve the accuracy of real-time identification system and ensure that the target can still be accurately identified in complex environment; Reduce the calculation delay and meet

  DOI: 10.25236/icmmct.2025.033

the requirements of real-time processing; Enhance the robustness of the system and improve the resistance to noise, occlusion and other interference factors. By achieving these goals, this article will provide a new research path for DL in the field of computer vision system optimization and real-time recognition technology, and help the sustainable development of this field.

2. DL-based computer vision system optimization algorithm

In the optimization process of computer vision system based on DL, the design of algorithm model undoubtedly plays a core role. An excellent algorithm model can not only accurately identify and understand the information in the image, but also pursue the maximization of processing efficiency while ensuring high recognition accuracy, and strive to find the best balance between computational complexity and performance [8]. Based on this, when constructing the model framework, the article chooses the multi-layer CNN (Convolutional Neural Network) structure as the cornerstone. This hierarchical design allows the model to go deep into the image layer by layer, from low-level edge features to high-level abstract concepts, and gradually extract the key information in the image. Each layer of convolution is like a fine analysis of the image, which enables the model to capture the details and structural features of the image more comprehensively, so as to enhance its understanding of complex scenes. Convolution layer formula:

$$o_{i,j}^{l} = \sigma\left(\sum_{m}\sum_{n} I_{i+m,j+n}^{l-1} \cdot K_{m,n}^{l} + b^{l}\right) \tag{1}$$

Where $o_{i,j}^{l}$ is the output characteristic diagram of the l layer convolution layer at the position (i, j). $I_{i,j}^{l-1}$ is the value of the input feature map of the $l-1$ layer at the position (i, j). $K_{m,n}^{l}$ is the weight of l layer convolution kernel at position (m,n). b^{l} is the bias term of the first l layer. m,n is the index of convolution kernel.

In the design of convolution kernel, this article has made careful consideration. The size and number of convolution kernels directly affect the feature extraction ability and computational efficiency of the model. Although too large convolution kernel can capture a wider image area, it will also bring about a sharp increase in computation. Therefore, after many experiments and optimization, we chose the convolution kernel configuration: convolution kernel size -55, which can effectively extract features without excessively increasing the computational burden. Number of convolution kernels -64. In this way, the model can ensure the full extraction of features while maintaining efficient operation.

For an input image y^{l-1} with size $M_x^{l-1} \times M_y^{l-1}$ and a convolution kernel with size $M_x^{l} \times M_y^{l}$, the size of the output image y^{l} is:

$$M_x^{l} = M_x^{l-1} - K_x^{l} + 1 \tag{2}$$

$$M_y^{l} = M_y^{l-1} - K_y^{l} + 1 \tag{3}$$

The pooling layer divides the output of the upper layer into l rectangular areas with $k_x \times k_y$ size, and pools (downsamples) the maximum value of each sub-area to produce translation invariance. Namely:

$$y_j^{l+1} = \frac{1}{M}\sum_{i=1}^{M}\left|y_{ij}^{l}\right| \tag{4}$$

The classification layer is responsible for transforming the features extracted from the front end of the network into specific classification results. After a series of feature extraction of convolution layer, the network comes to the last convolution layer, where the output contains rich spatial feature information. In order to further concise these features, the classification layer will first reduce the output of the last convolution layer to the size of one pixel through pooling operation. The last layer is the fully connected layer. Each neuron in this layer corresponds to a specific classification. The output of each neuron can be understood as the possibility that the image belongs to the corresponding classification. This design enables the network to directly output the probability of each classification, which provides an intuitive basis for subsequent decision-making. In the last classification function, Parzen classifier is adopted. Parzen classifier is a nonparametric estimation method, which can evaluate the possibility that each input image belongs to various types through kernel density estimation. The output of each neuron is based on the input feature vector, and the probability of belonging to this category is calculated by the kernel function of Parzen classifier. Parzen classifier and CNN's image recognition framework are shown in Figure 1:

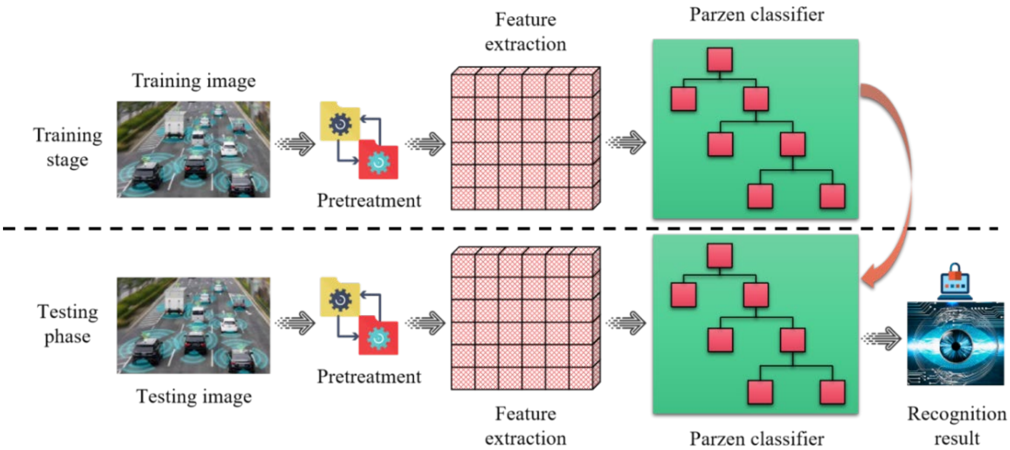

Figure 1 Parzen classifier and CNN's image recognition framework

The choice of activation function is also very important. In this article, the activation function - Sigmoid with strong nonlinear mapping ability is selected:

$$\sigma(z)=\frac{1}{1+e^{-z}} \tag{5}$$

Sigmoid function compresses the input between 0 and 1, which is suitable for binary classification problems. It can introduce necessary nonlinear characteristics into the model, so that the model can learn and express more complex image features. This nonlinear mapping ability greatly enhances the expression ability of the model and makes it better adapt to various complex visual tasks.

On the definition of loss function, the article has customized the design according to the requirements of specific tasks. The loss function is the compass of model training, which guides the model to search and optimize constantly in the parameter space to minimize the difference between the predicted results and the actual labels. Using loss function to train network;

$$L=-\sum_{t=1}^{T}\sum_{i=1}^{N} y_{t,i}\log\left(\hat{y}_{t,i}\right) \tag{6}$$

Where: T is the total time steps of the sequence. N is the number of categories. $y_{t,i}$ is the real label for class i at time step t. $\hat{y}_{t,i}$ is the prediction probability. Through the carefully designed loss function, the training process of the model can be effectively supervised, and the recognition accuracy can be continuously improved, and finally a satisfactory performance level can be achieved.

3. Experimental verification and result discussion

In order to test the effect of DL-based computer vision system optimization algorithm, this section has carried out a series of well-designed experiments. The experiment selects a widely representative data set, which contains images under various scenes and lighting conditions. Then, this section builds an experimental environment, including hardware equipment and software framework, to ensure the smooth progress of the experimental process. In terms of assessment index,this article adopts several indexes such as accuracy, processing speed and memory occupation to comprehensively evaluate the performance of the algorithm. The data set was preprocessed before the experiment. Pretreatment includes image cleaning, labeling and division to ensure the quality and consistency of data. At the same time, the experimental parameters are carefully set, including learning rate, batch size and training rounds, so as to ensure the repeatability and comparability of the experiment. During the experiment, the model was trained and tested in strict accordance with the established settings. In the training stage, cross-validation technology is adopted, and the data set is divided into training set and validation set. By constantly adjusting parameters and optimizing algorithms, the performance of the model on the validation set is gradually improved. In the testing stage, an independent test set is used to evaluate the generalization ability of the model, which ensures the reliability of the results. Figure 2 shows the performance of the optimization algorithm in accuracy.

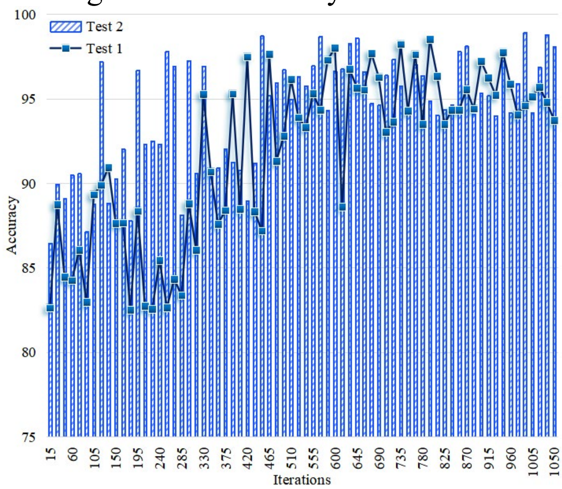

Figure 2 Performance of optimization algorithm in accuracy

Figure 3 shows the performance of the optimization algorithm in terms of processing speed.

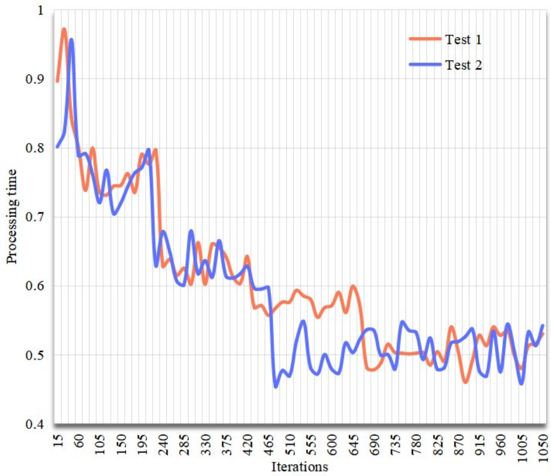

Figure 3 Performance of optimization algorithm in processing speed

With the training of the algorithm, the memory occupation is shown in Figure 4.

The above experimental results show that the optimization algorithm proposed in this article has achieved significant improvement in accuracy, processing speed and memory occupation. Specifically, compared with the classical computer vision algorithm, the accuracy of the algorithm proposed in this article is improved by about 12 percentage points, the F1 score is increased by

14.5%, the processing speed is accelerated by 1.5 times, and the memory occupation is also greatly reduced. This makes the algorithm more suitable for real-time application scenarios.

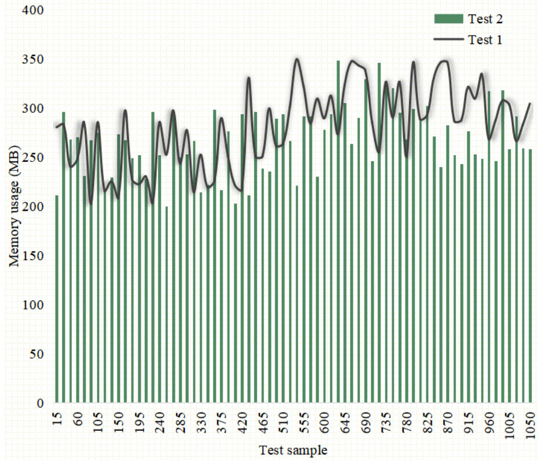

Figure 4 Memory occupancy

4. Conclusions

This study deeply discusses the application of DL in computer vision system optimization and real-time recognition technology. The research focuses on the innovative design, implementation and optimization of the algorithm model, and its effectiveness is verified by experiments. In the design stage of the algorithm model, a novel neural network architecture is proposed. By carefully adjusting the model level, convolution kernel configuration and activation function, the representation and generalization ability of the model are greatly improved. The experimental verification part fully proves the excellent performance of the proposed algorithm in accuracy, processing speed and memory usage through comparative experiments.

The achievement of this study lies not only in the innovation of algorithm model, but also in the effectiveness of these innovations verified by experiments. The experimental results clearly show that the proposed algorithm has achieved remarkable progress in many performance indexes, which provides strong evidence for the improvement of computer vision system and the practical application of real-time recognition technology. In the future, with the continuous upgrading of hardware facilities and the continuous improvement of algorithm models, the accuracy and processing speed of real-time identification technology will be further improved to meet the needs of a wider range of application scenarios. DL-based computer vision system optimization and real-time identification technology will bring more convenience and new discoveries to people's life and work in the near future.

References

[1] Huang Y J, Gao L, Yang T. Research on Image Recognition Algorithm Based on Multi-scale Features and Meta-learning [J]. Computer Applications and Software, 2024, 41(8): 203-209.

[2] Wan C Z, Ji X H, Yang M, et al. Mineral Image Recognition Based on Progressive Multi-granularity Training Deep Learning [J]. Earth Science Frontiers, 2024, 31(4): 112-118.

[3] Dong T, Yang B H. Target Enhancement Recognition of Visual Sensing Images Based on Deep Learning with Multiple Datasets [J]. Chinese Journal of Sensors and Actuators, 2024, 37(1): 64-70.

[4] Li Y H, Xu X H, Zhu H H, et al. Rock Fracture Identification, Characterization, and Software Development Based on Computer Vision [J]. Chinese Journal of Rock and Soil Mechanics, 2024, 46(3): 459-469.

[5] Chen L, Zhang J L, Peng H, et al. Few-shot Image Recognition with Multi-scale Attention and Domain Adaptation [J]. Opto-Electronic Engineering, 2023, 50(4): 60-73.

[6] Kumar A, Shukla S K, Yadav P R K. A Deep Learning and Powerful Computational Framework for Brain Cancer MRI Image Recognition [J]. Journal of The Institution of Engineers (India), Series B. Electrical Engineering, Electronics and Telecommunication Engineering, Computer Engineering, 2024, 105(1): 1-18.

[7] Wang Z, Lei M, Wang J, et al. Unsupervised deep learning-based ground penetrating radar image translation for internal defect recognition of underground engineering structures [J]. Structural Health Monitoring, 2024, 23(2): 649-670.

[8] Dourado C M J M, Silva S P P D, Nobrega R V M D, et all. An Open IoHT-Based Deep Learning Framework for Online Medical Image Recognition [J]. IEEE Journal on Selected Areas in Communications, 2021, 39(2): 541-548.