Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 1, 2024

Application and Optimization of Lightweight Convolutional Neural Network in Target Detection

Liangyuan Li

Affiliated Middle School of Henan Normal University, No. 85 Jianshe East Road, Muye District, Xinxiang City, Henan Province, China

Abstract: Due to its low computational complexity, minimal storage needs, and high real-time performance, lightweight convolutional neural networks (LCNN) have garnered significant attention amidst the swift advancement of deep learning technology. By automatically learning high-level feature representation, LCNN can capture key information in images more effectively. This paper aims to discuss the application and optimization of LCNN in target detection task. This paper discusses the key strategies such as network structure optimization, training method optimization and post-processing optimization to improve the performance of LCNN in target detection. The experimental results show that compared with the traditional SIFT-based target detection algorithm, LCNN has obvious advantages in detection success rate, time consumption and adaptability to different scenes. In addition, the lightweight design of LCNN makes it easier to deploy on equipment with limited resources. LCNN shows strong performance and wide application prospect in target detection, which is expected to provide new impetus for the innovation of target detection technology.

Keywords: Lightweight; Convolutional neural network; Target detection; Optimization strategy.

1. Introduction

Today, with the rapid development of information technology, computer vision, as an important branch in the field of artificial intelligence, is promoting industrial upgrading at an unprecedented speed. As one of the core tasks of computer vision, target detection aims to accurately identify and locate the interested target object from complex images or video sequences [1]. This technology has shown great application potential in many fields such as automatic driving, intelligent monitoring, medical image analysis and so on. Traditional target detection methods mainly rely on manually designed features and classifiers, such as HOG+SVM, Haar+AdaBoost, etc [2]. Although these methods can achieve certain results in specific scenes, their generalization ability and robustness are obviously insufficient when dealing with complex and changeable actual environments [3]. By automatically learning high-level feature representation, CNN can capture the key information in the image more effectively, thus achieving high-precision target detection [4]. However, high performance is often accompanied by high computing cost and storage requirements, which limits the wide application of deep learning model in resource-constrained devices.

The appearance of LCNN makes it possible to solve this contradiction. Lightweight network can significantly reduce the parameters and computational complexity of the model while maintaining or improving the detection accuracy through well-designed network structure and optimization strategy [5]. How to balance the detection speed and accuracy with limited computing resources has become a key issue in current research. This paper aims at systematically discussing the application and optimization strategy of LCNN in target detection.

In order to verify the effectiveness of the above research contents, this paper will use standard data sets and evaluation indicators to evaluate the performance of lightweight networks and optimization strategies. Through the comparative analysis of experimental results, the advantages and disadvantages of lightweight network in target detection and the actual effect of optimization strategy are discussed.

2. Related Theoretical Basis

2.1. Basic principle of convolutional neural network

Convolutional neural network (CNN) is a deep learning model with grid topology (such as images and voice signals) specially used for processing data. Its core idea lies in local perception, weight sharing and pooling operation [6]. The traditional fully connected network ignores the spatial structure information of the image when processing the image, while CNN only pays attention to the connection between each neuron and a local area in the input image through local connection. In the same layer, multiple neurons share the same set of weights, which means that in feature extraction, no matter where the feature is located in the image, the same convolution kernel is used for operation. Pooling layer (such as maximum pooling and average pooling) enhances the robustness of the model to input changes by downsampling the feature map.

2.2. Design principle and typical model of lightweight network

The core of lightweight network design is to achieve efficient implementation of the model by reducing the parameters and computational complexity, while striving to maintain or even improve the original performance level [7]. This design follows several key principles: firstly, the complexity of the model is effectively reduced by using deep separable convolution, network pruning, weight quantization and other technical means; Secondly, an efficient network structure is adopted to enhance the circulation of information and the reuse efficiency of features; Finally, in the process of pursuing lightweight, the precision and running speed of the model are carefully balanced.

Under this design concept, many typical lightweight models have emerged. MobileNetV1 creatively proposed the depth separable convolution, which skillfully decomposed the standard convolution into depth convolution and point-by-point convolution. Subsequently, MobileNetV2 further innovated on this basis, and introduced the inverted residual structure and linear bottleneck layer, which significantly improved the performance. Another important series is ShuffleNet, which strengthens the information exchange between different channels by introducing channel shuffling operation. In addition, EfficientNet series is based on a compound scaling method, which can maximize the performance under the condition of limited resources by adjusting the depth, width and resolution of the network at the same time.

2.3. Overview of target detection algorithms

Traditional target detection methods mainly rely on manually designed features and classifiers. For example, the pedestrian detection method using HOG features combined with SVM classifier and the application of AdaBoost classifier based on Haar features in face detection [8]. With the rise of deep learning, the target detection algorithm has ushered in a new breakthrough. Among them, R-CNN series algorithm is the first algorithm to apply CNN to target detection. By extracting candidate regions and using CNN for feature extraction and classification, R-CNN has achieved high detection accuracy. Subsequently, Fast R-CNN, fast r-CNN and other algorithms further optimized the detection process and speed, making the target detection method based on deep learning more efficient in practical application.

In addition to R-CNN series, YOLO series algorithms have also received extensive attention. YOLO is a single-stage detection algorithm, which transforms the target detection task into a single forward propagation problem, thus realizing real-time detection. With the introduction of YOLOv3, YOLOv4 and other subsequent versions, the detection accuracy and speed of YOLO series algorithms have been further improved by introducing more advanced network structure and optimization strategy.

In addition, SSD(Single Shot MultiBox Detector) is also an important single-phase detection algorithm. It realizes high-precision target detection by predicting the position and category of targets on feature maps of different scales. The design of SSD algorithm enables it to perform well in complex scenes, which further promotes the development of target detection technology.

3. Application and Optimization of LCNN

The application of lightweight network in the field of target detection aims to achieve a key balance by finely optimizing the network structure and parameter configuration, that is, to greatly reduce the computational complexity and storage requirements while maintaining a high level of detection accuracy. In view of the particularity of target detection task, designers carefully construct an efficient network architecture, aiming at capturing key image features with less computing resources. By generating feature maps at different levels of the network and predicting targets on these feature maps respectively, the detection ability of the algorithm for various sizes of targets can be significantly improved, and the comprehensiveness and accuracy of detection can be ensured. Drawing lessons from advanced algorithms such as Faster R-CNN, a group of anchor frames with different scales and aspect ratios are preset. Then, through fine regression adjustment, the detection frame fits the actual target contour better. After the initial detection frame is obtained, efficient post-processing means such as non-maximum suppression (NMS) can effectively eliminate duplicate and redundant detection frames, and ensure that the output results are concise and accurate. Figure 1 shows the basic structure of LCNN.

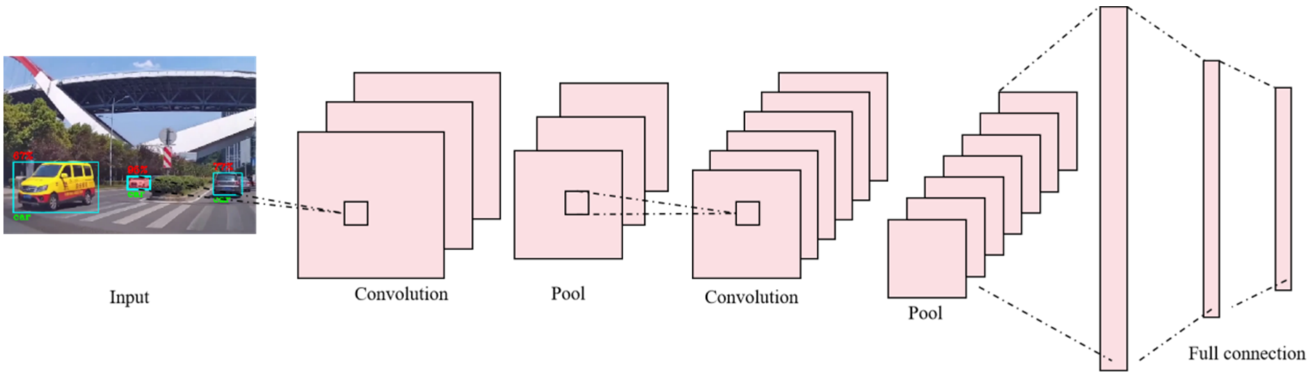

Figure 1. LCNN

In CNN architecture, neurons serve as the fundamental processing units tasked with computation. For those neurons possessing an x_1, x_2, x_3 input dimension, their output is determined through the interaction of a designated convolution kernel (or filter) with the input data.

$$Y_{w,b}(x) = f(W^T x) = f\left(\sum_{i=1}^{3} W_i x_i + b_i\right) \tag{1}$$

In this context, W_i denotes the weight value assigned to each node, while b_i signifies the bias constant. Neurons have the capability to interconnect, thereby constituting a neural network.

Aiming at the optimization of network structure, the introduction of deep separable convolution can significantly reduce the amount of calculation and parameters. At the same time, feature fusion and enhancement strategies, such as splicing different levels of feature maps or introducing attention mechanism, further improve the feature expression ability of lightweight networks. Dynamically adjust the network structure according to the complexity of the input image and the change of the target scale, and realize the reasonable allocation of computing resources.

This study employs the Adam optimization algorithm for updating network parameters, with the following update rules:

$$\upsilon_t = \mu \cdot \upsilon_{t-1} - \eta \cdot \nabla_\theta J(\theta) \tag{2}$$

$$\theta_{t+1} = \theta_t + \upsilon_t \tag{3}$$

In this context, υ_t represents the speed at time step t, μ denotes the momentum coefficient, η signifies the learning rate, $\nabla_\theta J(\theta)$ indicates the gradient of the loss function $J(\theta)$ concerning model parameter θ, and θ_t stands for the model parameter at time step t.

To enhance generalization and mitigate overfitting, this study employs L2 regularization:

$$loss = -\sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2 \tag{4}$$

Here, y denotes the expected value, $\hat{y}$ the predicted value, λ represents L2 regularization, and θ signifies the parameter set of the CNN. Selecting the suitable regularization coefficient λ ensures model performance while mitigating the risk of over-fitting.

In terms of training methods, data enhancement and sample balance techniques increase the diversity of training samples and improve the generalization ability of the model. Knowledge distillation and transfer learning use the knowledge of large model to guide the training of lightweight network. Joint training and multi-task learning realize knowledge sharing among different tasks by sharing network layer or parameters.

4. Experimental Results and Analysis

Figure 2 shows the comparison between LCNN and SIFT-based target detection algorithm in detection success rate. The detection success rate of LCNN on all kinds of targets is significantly higher than that of SIFT algorithm. Through deep learning and large-scale data training, LCNN can capture more abundant target features and maintain stable detection performance under various illumination, occlusion and angle changes.

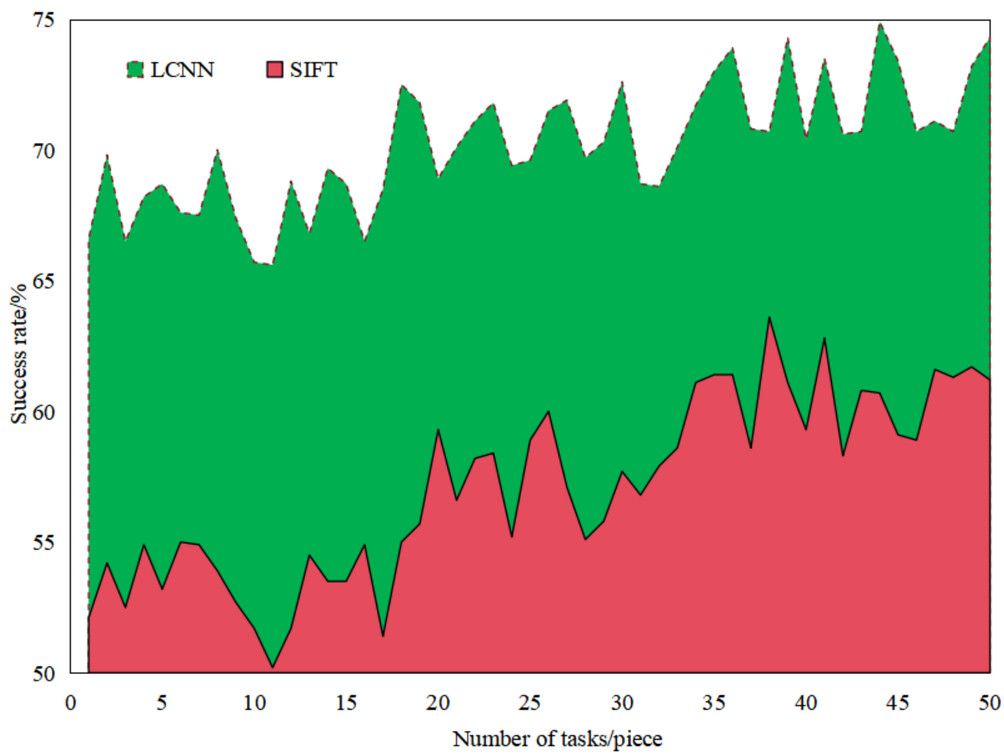

Figure 2. Comparison of detection success rate

Figure 3 further reveals the advantages of LCNN in time loss of target detection. Compared with SIFT-based algorithm, LCNN achieves a significant improvement in detection speed. In practical application, this means that LCNN can complete the target detection task in a shorter time and meet the requirements of real-time scenes. However, SIFT algorithm needs to calculate a large number of feature points and descriptors, which leads to a large time loss and is difficult to meet the real-time requirements.

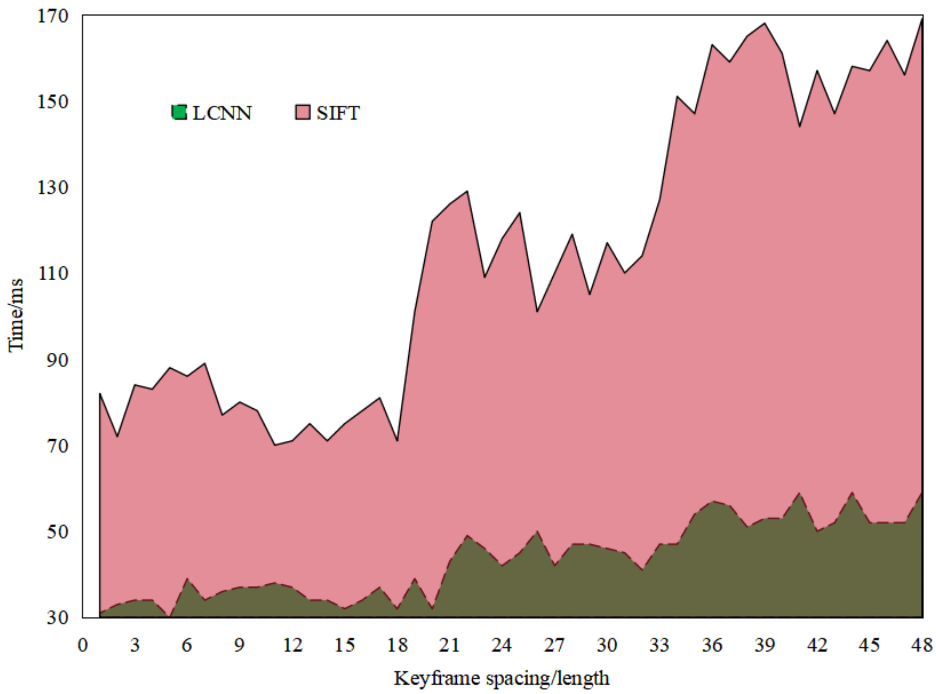

Figure 3. Comparison of Time Loss

In addition to detecting the success rate and time loss, the performance of LCNN in different scenarios is deeply analyzed. The experimental results show that LCNN performs well in various scenes, and can detect targets accurately and quickly, whether indoors or outdoors, day or night.

5. Conclusions

Based on the in-depth study of LCNN in target detection, this paper discusses its design ideas, optimization strategies and performance analysis. The research shows that LCNN significantly reduces the computational complexity and storage requirements by optimizing the network structure and parameters, while maintaining high detection accuracy and real-time. In terms of design ideas, LCNN adopts strategies such as deep separable convolution, feature fusion and enhancement, which improves the ability of feature extraction and adaptability to complex scenes. By optimizing training methods and post-processing means, the performance of LCNN is further improved. The results show that LCNN is superior to the traditional SIFT-based target detection algorithm in detection success rate, time loss and adaptability to different scenes. The lightweight design of LCNN makes it easier to deploy on equipment with limited resources.

To sum up, LCNN shows significant advantages in target detection. In the future, with the continuous progress of deep learning technology and the emergence of new hardware platforms, LCNN is expected to play an important role in more fields.

References

[1] Payghode V, Goyal A, Dubey I A K. Object detection and activity recognition in video surveillance using neural networks[J]. International journal of web information systems, 2023, 19(3/4):123-138.

[2] Tang J, Zhou H, Wang T, et al. Cascaded foreign object detection in manufacturing processes using convolutional neural networks and synthetic data generation methodology[J]. Journal of Intelligent Manufacturing, 2023, 34(7):2925-2941.

[3] Boisclair J, Kelouwani S, Ayevide, Follivi KloutseAmamou, AliAlam, Muhammad ZeshanAgbossou, Kodjo. Attention transfer from human to neural networks for road object detection in winter[J]. IET image processing, 2022, 16(13): 3544-3556.

[4] Ding X, Li B, Wang J. Geometric property-based convolutional neural network for indoor object detection:[J]. International Journal of Advanced Robotic Systems, 2021, 18(1):261-318.

[5] Zhao Z, Huang Z, Chai X, et al. Depth enhanced cross-modal cascaded network for RGB-D salient object detection[J]. Neural Processing Letters, 2023, 55(1):361-384.

[6] Minor E, Howard S, Green A, et al. End-to-end machine learning for experimental physics:using simulated data to train a neural network for object detection in video microscopy. [J]. Soft matter, 2020, 16(7):1751-1759.

[7] Ye T, Zhang X, Zhang Y, et al. Railway traffic object detection using differential feature fusion convolution neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(3):1375-1387.

[8] Guo H, Bai H, Zhou Y, et al. DF-SSD:a deep convolutional neural network-based embedded lightweight object detection framework for remote sensing imagery[J]. Journal of Applied Remote Sensing, 2020, 14(1):014521-014521.