Computing, Performance and Communication Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/cpcs.2025.090110
ISSN 2371-8889 Vol. 9 Num. 1

Design and Implementation of a Fruit Image Recognition System Based on CNN

Ranning Deng

School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China
2840820473@qq.com

Keywords: Convolutional Neural Network; Fruit Image Recognition; Lightweight Model; Activation Function; Image Classification

Abstract: Fruit classification and recognition have significant application value in fields such as agricultural sorting and fresh food retail. To address the poor robustness of traditional manual feature extraction methods, this paper designs and implements a fruit image recognition system based on lightweight convolutional neural networks (CNNs). The system adopts 20 common types of fruit images from the Fruit-360 dataset. After preprocessing, an improved LeNet-5 model and a simplified AlexNet model are constructed, with a comparative analysis of the performance differences between ReLU and Sigmoid activation functions. Experimental results show that the simplified AlexNet combined with the ReLU activation function achieves the optimal performance, with an average test set accuracy of 96.87%. It also features fast training convergence, making it suitable for real-time recognition requirements in low-computing-power scenarios.

1. Introduction

1.1 Research Background and Significance

With the rapid development of agricultural intelligence and new retail, the demand for automated fruit classification has become increasingly urgent. Traditional manual sorting is inefficient and error-prone, while conventional machine learning methods relying on manual feature extraction lack robustness in complex scenarios. Convolutional Neural Networks (CNNs) can automatically extract multi-level image features without human intervention, demonstrating excellent performance in image classification tasks. Applying lightweight CNNs to fruit recognition enables efficient and low-cost automated classification, which holds important practical significance for improving industry operational efficiency.

1.2 Research Status at Home and Abroad

Foreign scholars initiated relevant research earlier. Hussain et al. achieved a recognition accuracy of 92.5% on the Fruit-360 dataset based on an improved LeNet-5, but did not involve a comparison of activation functions. Domestic research mostly focuses on the application of single models. Li Ming et al. designed a three-layer CNN with an accuracy of 93.1% for 6 types of fruits, but it lacks generalization ability and lightweight design. Similar to other domestic applications of CNNs, such as wood species recognition [5] and citrus yield estimation [6], existing fruit recognition studies also insufficiently explore lightweight optimization and the impact of key components like activation functions. Existing studies insufficiently explore lightweight model optimization and the impact of activation functions, making it difficult to meet the requirements of low-computing-power deployment.

1.3 Research Content and Structure

The core research content of this paper includes selecting and preprocessing the Fruit-360 dataset, constructing two lightweight CNN models, and comparing the training effects and recognition performance of different activation functions. The paper is divided into 5 chapters: the introduction elaborates on the research background and content; related technologies introduce CNN fundamentals and activation functions; system design explains data preprocessing and model structure; the experimental part presents the design and results; and the conclusion and outlook summarize the achievements and improvement directions.

2. Related Technical Foundations

2.1 Basic Principles of Convolutional Neural Networks

CNN is a deep learning model for processing image data, realizing feature extraction and classification through convolutional layers, pooling layers, and fully connected layers. As emphasized in the theory of very deep convolutional networks [2], the hierarchical structure of CNNs enables the step-by-step extraction of low-level (e.g., edges) to high-level (e.g., semantic) features. Specifically, convolutional layers extract local features using convolution kernels, pooling layers reduce the number of parameters through downsampling, and fully connected layers integrate features and output classification results—this feature extraction mechanism is consistent with the rich feature hierarchy framework proposed by Girshick et al. [7]. Its core advantages lie in weight sharing and automatic feature learning, which effectively reduce model complexity and improve recognition accuracy.

2.2 Lightweight CNN Models

LeNet-5: A classic lightweight model consisting of 2 convolutional layers, 2 pooling layers, and 2 fully connected layers. It has a small number of parameters and high training efficiency, suitable for small-scale image classification tasks.

Simplified AlexNet: Optimized from the original AlexNet, it reduces the number of convolution kernels and neurons in fully connected layers, and removes the LRN layer. It maintains feature extraction capability while reducing computational complexity.

2.3 Activation Functions

ReLU: Defined as $f(x) = \max(0, x)$, it is simple to compute, converges quickly, and can alleviate gradient vanishing. It is currently the most widely used activation function.

Sigmoid: Defined as $f(x) = \frac{1}{1+e^{-x}}$, its output range is (0, 1). However, it is prone to gradient vanishing when the input absolute value is too large, with high computational complexity.

3. Overall System Design

3.1 Dataset Selection and Preprocessing

20 common types of fruits from the Fruit-360 dataset are selected, totaling 14,200 images. Among them, the training set contains 10,000 images and the test set 4,200 images, with balanced data distribution across all categories. Preprocessing steps include: resizing images to 64×64×3; normalizing pixel values to [0, 1] to accelerate convergence; performing data augmentation through random flipping, rotation, and brightness adjustment to mitigate overfitting; converting data format to NumPy arrays; and using one-hot encoding for labels.

3.2 Model Structure Design

1) Improved LeNet-5

To adapt to the feature extraction needs of RGB color fruit images, structural optimization is performed on the classic LeNet-5 model. The model adopts a classic architecture of alternating "convolution-pooling" stacks followed by fully connected layers. The specific parameter design is as follows:

Input Layer: Receives preprocessed 64×64×3 RGB image data, providing raw input for subsequent feature extraction.

Convolutional Layer 1: Configures six 5×5 convolution kernels with a stride of 1 and Same padding, ensuring the output feature map size is consistent with the input to fully retain image edge features.

Pooling Layer 1: Uses 2×2 max-pooling to implement downsampling by retaining local maximum feature values, reducing computational complexity while enhancing feature robustness.

Convolutional Layer 2: Expands the number of convolution kernels to 16, maintaining a 5×5 kernel size and a stride of 1 with Valid padding, further extracting deep texture and shape features of images.

Pooling Layer 2: Continues the 2×2 max-pooling configuration to further compress feature map dimensions and reduce model parameters.

Fully Connected Layer 1: Flattens the multi-dimensional feature map output from the pooling layer into a one-dimensional vector, then connects to 120 neurons to achieve global integration of local features.

Fully Connected Layer 2: Sets 84 neurons to deepen the nonlinear mapping capability of features, providing more discriminative feature expressions for classification tasks.

Output Layer: For 20 fruit categories, 20 neurons are configured with the Softmax activation function to convert outputs into category probability distributions and complete classification prediction.

2) Simplified AlexNet

Based on the original AlexNet architecture, lightweight improvements are made by streamlining network parameters and removing redundant structures. This ensures feature extraction capability while reducing computational complexity. The specific design is as follows:

Input Layer: Similarly receives 64×64×3 RGB image data, maintaining input consistency with the improved LeNet-5 for experimental comparison.

Convolutional Layer 1: Configures 16 11×11 convolution kernels with a stride of 4, quickly extracting low-level edge and contour features of images, balancing feature extraction range and computational efficiency.

Pooling Layer 1: Uses 3×3 max-pooling with a stride of 2, effectively compressing feature map size and suppressing overfitting.

Convolutional Layer 2: Increases the number of convolution kernels to 32, reduces the kernel size to 5×5, and adopts Same padding to refine feature expressions and capture local detail features of fruits.

Pooling Layer 2: Maintains 3×3 max-pooling with a stride of 2 to further streamline feature dimensions.

Convolutional Layers 3-5: Convolutional Layers 3-5: Sequentially set 64, 64, and 32 3×3 convolution kernels, all using Same padding. Referring to the design concept of "going deeper with convolutions" [1],through multi-layer convolution stacking, the deep convolutional neural network (DCNN) for fruit recognition deepens its feature abstraction capability and gradually extracts high-level semantic features of fruits.

Pooling Layer 3: Continues the 3×3 max-pooling with a stride of 2 configuration to complete the final feature downsampling.

Fully Connected Layer 1: Connects to 256 neurons to perform global integration and nonlinear transformation of deep convolutional features.

Fully Connected Layer 2: Sets 120 neurons to optimize the discriminability of feature expressions, laying the foundation for classification output.

Output Layer: Consistent with the improved LeNet-5, 20 neurons and the Softmax activation function are configured to realize probability prediction for 20 fruit categories.

4. Experimental Design and Result Analysis

4.1 Experimental Environment and Parameters

Experimental Environment: CPU: Intel Core i7-10700; GPU: NVIDIA RTX 3060; Framework: TensorFlow 2.8.

Training Parameters: Batch size = 32; Learning rate = 0.001; Number of iterations = 50; Loss function = Cross-entropy loss; Optimizer = Adam; Regularization = Dropout (probability = 0.5).

4.2 Experimental Results and Analysis

The test set performance indicators of the two models combined with different activation functions are shown in the following table 1:

Table 1 Performance Comparison of Different Models and Activation Functions

Model	Activation Function	Accuracy (%)	Recall (%)	F1 Score (%)
Improved LeNet-5	ReLU	91.53	90.87	90.20
Improved LeNet-5	Sigmoid	83.21	82.54	82.87
Simplified AlexNet	ReLU	96.87	96.52	96.69
Simplified AlexNet	Sigmoid	88.55	87.93	88.24

Impact of Activation Functions: For the same model, the ReLU activation function outperforms Sigmoid in accuracy, recall, and F1 score, with fewer convergence iterations. This is because ReLU avoids the gradient vanishing problem and achieves higher training efficiency, while Sigmoid exhibits significant gradient decay in deep networks, affecting parameter optimization.

Impact of Model Structure: The simplified AlexNet performs better than the improved LeNet-5. It contains more convolutional layers, enabling extraction of more complex fruit features. The deep network’s stronger feature abstraction capability is suitable for multi-category fruit recognition tasks.

Differences in Category Recognition: The simplified AlexNet + ReLU model achieves an accuracy of over 98% for fruits with large morphological differences (e.g., bananas, pineapples), and approximately 94% for similar categories (e.g., red apples, green apples). This is mainly due to the high feature overlap in color and shape between similar fruits.

5. Conclusion and Outlook

This paper designs a fruit image recognition system based on lightweight CNNs. By comparing the performance of the improved LeNet-5 and simplified AlexNet models, as well as the ReLU and Sigmoid activation functions, the optimal effect of the simplified AlexNet combined with ReLU is verified. The system achieves a test set accuracy of 96.87%, featuring a simple structure and easy deployment, which can meet the needs of low-computing-power scenarios.

Research Limitations: The impact of practical scenarios such as complex backgrounds and occlusions is not considered, and the model’s generalization ability needs to be improved. Only two activation functions are compared, without involving other optimization algorithms.

Future Directions: Expand the dataset to include samples from complex scenarios; introduce transfer learning to further improve model performance. Pan & Yang [3] systematically summarized the theoretical basis of transfer learning, pointing out its advantage in leveraging pre-trained model knowledge for new tasks, and Yosinski et al. [4] verified that features in deep neural networks have strong transferability across similar image recognition tasks—these findings provide theoretical support for applying transfer learning to fruit image recognition; explore more efficient lightweight network structures to optimize the balance between recognition speed and accuracy.

References

[1] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[2] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[3] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359.

[4] Yosinski, Jason, et al. "How transferable are features in deep neural networks?" Advances in neural information processing systems 27 (2014).

[5] Miao Yujie, Zhu Shiping, Pu Jing, Li Junxian, Ma Lingkai, Huang Hua. Recognition of Furniture Wood Image Species Based on Convolutional Neural Networks [J]. Scientia Silvae Sinicae, 2023, 59 (8): 133-140.

[6] Xiaoyue Qin, Ruwei Huang, and Bei Hua. "Research and implementation of yield recognition of Citrus reticulata based on target detection." Journal of physics: conference series. Vol. 1820. No. 1. IOP Publishing, 2021.

[7] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.