Fruit Image Classification using the Inception-V3 Deep Learning Model

Samuel Enobong Sunday
Faculty of Electronic Information Engineering
Huaiyin Institute of technology
Huai'an City, China
sundaysamuel1706@gmail.com

Rendong Ji*
Faculty of Electronic Information Engineering
Huaiyin Institute of technology
Huai'an City, China
jrdgxy@163.com

Ahmed N Abdalla
Faculty of Electronic Information Engineering
Huaiyin Institute of technology
Huai'an City, China
line 5: dramaidecn@gmail.com

Haiyi Bian
Faculty of Electronic Information Engineering
Huaiyin Institute of technology
Huai'an City, China

2023 International Conference on the Cognitive Computing and Complex Data (ICCD) | 979-8-3503-0629-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCD59681.2023.10420760

***Abstract*—Fruit image recognition plays a vital role in the realm of deep learning, with applications extending to smart agriculture and harvesting robots. However, traditional image classification algorithms often suffer from limited generalization ability and low accuracy. In this research, we address these challenges by proposing a novel fruit image classification algorithm that leverages deep learning and transfer learning techniques. Specifically, we employ a modified version of the Inception-V3 model for feature extraction from fruit images and subsequently utilize a deep learning classifier to classify these extracted features. Additionally, we incorporate transfer learning to enhance the training process. Through comprehensive testing, our proposed algorithm demonstrates significantly higher recognition accuracy compared to conventional fruit classification approaches, marking a significant advancement in the field of fruit image classification using deep learning techniques.**

Keywords—Image recognition; Deep learning; Transfer learning

I. INTRODUCTION

Image recognition has emerged as a crucial application in deep learning, encompassing a diverse range of tasks, such as object detection, image segmentation, and classification. Fruit image classification, within the realm of image recognition, has significant applications in agriculture, enabling farmers to efficiently identify and manage different types of fruits in their fields. Additionally, the deployment of fruit image classification algorithms in harvesting robots can optimize fruit-picking processes, leading to increased productivity and reduced labor costs.

In recent years, with the rapid development of artificial intelligence and computer vision technology, image recognition, as one of the important applications in deep learning technology, has received more and more attention. At present, people's research on fruit image recognition mainly focuses on fruit quality grading, maturity recognition, and other aspects. However, there is less research on intelligent agriculture and multi class fruit classification and recognition of picking robots [1], and most of them use traditional machine learning methods for classification and recognition. Traditional machine learning methods, such as HOG (Histogram of Orientated Gradient) and SIFT (Scale Invariant Feature Transform), extract the features of fruit images, and input the extracted features to the classifier to achieve the classification and recognition of fruits. The features extracted by this method are essentially artificially set features. Although they are more effective for specific recognition tasks with small data scales, their overall generalization ability is poor and has certain limitations [2]. Compared with the traditional methods mentioned above, deep learning methods have the advantage that image features do not need to be manually set, but rather extract more accurate features through deep network structures. Convolutional Neural Network (CNN), a popular type of deep neural network in Deep Learning [3], eliminates the need for manual feature extraction, unlike traditional feature extraction algorithms such as SIFT, LBP, etc. This feature extraction automation has contributed to the widespread adoption of CNNs.

This paper presents a promising contribution to the field of fruit image classification by leveraging the Inception-V3 deep learning model and transfer learning to attain precise recognition and classification of diverse fruits. By addressing the limitations of traditional algorithms, we anticipate significant progress in the application of deep learning for fruit image recognition, ultimately benefiting the agricultural sector and automation technologies.

II. REALTED WORK

A. Inception-V3 Model

In recent years, Google Corporation in the United States has developed numerous image classification models, including QuocNet, AlexNet, Inception (GoogleNet), and BN-Inception V2, among others. Inception-V3, for example, was employed to train the ImageNet Large-Scale Visual Recognition Challenge dataset [4], a standard task in the field of computer vision. In this task, the model categorizes the entire image set into 1000 classes such as zebras, Dalmatians, and dishwashers. In order to compare and evaluate the performance of the model, the error rate of the top five classification results in the model's predictions, excluding the correct category, is checked, known as the "t-5 error rate". On the 2012 validation dataset, AlexNet achieved a top-5 error rate of 15.3%, BN-Inception V2 had an error rate of 6.66%, and Inception V3 demonstrated an error rate of 3.46% [5]. It can be seen that the Inception-V3 model has better performance in image classification.

Authorized licensed use limited to: R V College of Engineering. Downloaded on July 03,2025 at 19:00:39 UTC from IEEE Xplore. Restrictions apply.

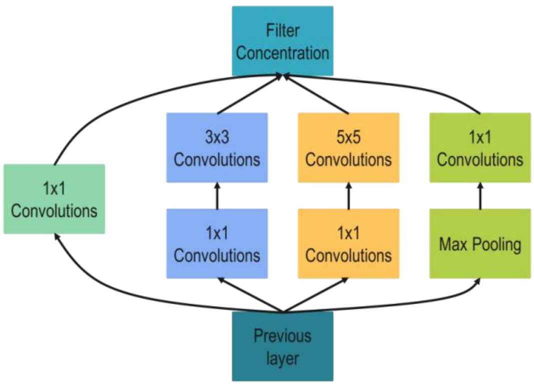

Fig. 1. Inception module structure

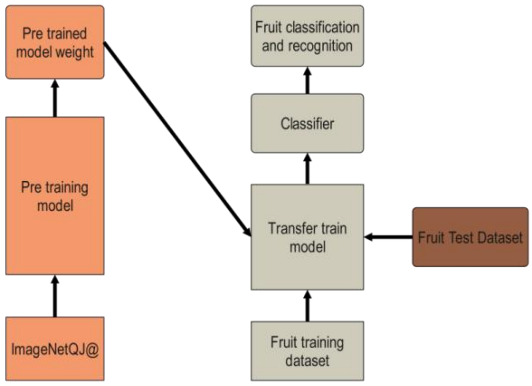

Fig. 2. Fruit Classification Algorithm

B. Deep learning and Transfer learning

The concept of deep learning originates from the research of artificial neural networks, which is a method of Feature learning based on data in machine learning. The Multilayer perceptron with multiple hidden layers is a kind of deep learning structure. The biggest advantage of deep learning lies in using unsupervised or semi supervised feature learning and efficient hierarchical feature extraction algorithms to replace manual feature acquisition. In the field of deep learning, Transfer learning is a learning method that retrains the pre trained model and applies it to other tasks. Usually, these pre trained models have consumed significant time and computational resources during development [6]. The Inception-V3 model to carry out Transfer learning, the advantages of this model in image classification will be brought into play in fruit image recognition, so that the classification and recognition of fruit images will be faster and more accurate. By leveraging transfer learning, one can selectively freeze particular layers to prevent them from modifying their weights and biases during the training process. This technique becomes especially beneficial when working with pre-trained models that have already been trained on vast amounts of data. Consequently, this approach saves significant computational time, as it only requires training a portion of the model rather than the entire architecture [7].

III. Proposed Method

The designed fruit classification algorithm architecture is shown in Fig. 2. Under the deep learning framework, use the Inception-V3 model developed by Google to conduct Transfer learning. In Fig. 1., the CNN model trained by a large number of ImageNet data sets is taken as the pretraining model, which will retain the overall structure before the classifier in the Inception-V3 model, define a new Softmax classifier in the model, and then use the fruit data set for further Transfer learning training to obtain a new network model for the classification of fruit images.

In the Inception-V3 model, images of size 299x299 are taken as input, and after processing through the convolutional layers, they produce image feature vectors $x^{(i)}$ of size 1x2048. In this paper, the image feature vector $x^{(i)}$ and its corresponding image label value are used as input for the Softmax classifier, represented as $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ with corresponding labels $\{y^{(i)}\} \in \{1, 2, ..., k\}$, where m is the number of samples, k is the total number of fruit categories, and its value is 6. For a given input sample x(i), the probability value P of it being assigned to each class can be represented as follows:

$$P\left(y^{(i)} = j | x^{(i)}; \theta\right) = \frac{\exp\left(\theta_j^T x^{(i)}\right)}{\sum_{j=1}^{n} \exp\left(\theta_j^T x^{(i)}\right)}, j = 1,2,\ldots,k \qquad (1)$$

In the equation, $\frac{1}{\sum_{j=0}^{n} \exp(\theta_j x^{(i)})}$ is used to normalize the probability values, ensuring that the sum of all probabilities is equal to 1.

In the equation, $\theta_k (k_\in (1{,}2048)$, k is an integer) represents the node weight parameters of the Softmax classifier in the model. For simplicity, the symbol θ is also used to represent all model parameters. The matrix θ is obtained by arranging θ1, θ2, ..., θk in rows, resulting in a size of 2048x6. θ can be represented as $\theta = [\theta_1^T, \theta_2^T, \ldots, \theta_k^T]^T$.

The cost function of Softmax involves summing up the k probability values corresponding to class labels and can be represented as follows:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\left\{y^{(i)} = j\right\} \ln\left(P\left(y^{(i)} = j | x^{(i)}; \theta\right)\right)\right] \qquad (2)$$

where 1{•} represents the indicator function, which takes a value of 1 when the condition inside the curly braces is true.

During training, the gradient descent method is used to continuously adjust the training model parameters θ, aiming to minimize the cost function J(θ) and ensure convergence to the global optimal solution.

IV. Result and discussion

The Inception-V3 model consists of a substantial 96 convolutional layers, demanding significant computational resources for its pretraining. This makes its implementation on a standard computer challenging. Consequently, the approach of Transfer learning is employed to train the model for fruit classification and recognition. The training process of Transfer learning mainly includes the following steps:

Authorized licensed use limited to: R V College of Engineering. Downloaded on July 03,2025 at 19:00:39 UTC from IEEE Xplore. Restrictions apply.

data set establishment, model preparation, classifier design and model transfer training. First, the trained Inception-V3 model provided by Google as the pretraining model for this Transfer learning. The model file is named inception-V3.ckpt. Before training, set training related parameters. After model feature extraction, the number of output nodes is 2048, and the Learning rate is 0.01, and set the batch size for gradient descent to 100 and the number of classifications to 6. The Inception-V3 model consists of 11 Inception modules, and a typical Inception structure is shown in Fig. 1. This structure combines different convolutional layers in parallel, using all filters of different sizes, and then concatenates the resulting matrix. The Inception-V3 model has a total of 96 convolutional layers, and the pre-training of the model requires a large amount of computation, making it difficult to implement on a regular computer [5]. Therefore, Transfer learning is used to train the fruit classification and recognition model.

From training data can be viewed on the web interface the values under different iterations will be compared, as shown in Fig. 4. When the number of iterations is 1000, the loss is 0.18; When the number of iterations is 3000, the loss is 0.13; When the number of iterations is greater than 3000, the loss is 0.11.

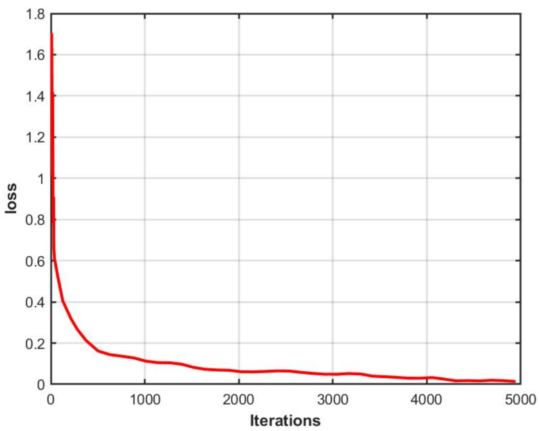

Fig. 3. Shows the loss at different iterations.

The variation curve of accuracy during the training process is shown in Fig. 5. The curve has been smoothed and filtered. During the transfer training on the fruit dataset, the training accuracy rapidly increases to 95% at around 1000 iterations. After 4000 iterations, the training accuracy gradually improves to 98.5%.

The recognition performance of the Inception-V3 model with transfer learning is compared with the results obtained using traditional machine learning methods [8]. The recognition accuracy is shown in Table I.

TABLE I. FRUIT RECOGNITION ACCURACY FOR DIFFERENT METHODS.

Method	Accuracy	Methods	Accuracy
BP	83.9	SIFT + SVM	90.2
SVM	81.1	Algorithm in this paper	97.7
PCA	78.4		

Table I, compares the different methods that were used for fruit recognition. These methods include:

- BP: Which refers to Backpropagation, which is a common training algorithm for neural networks.
- SIFT + SVM: This method involves using Scale-Invariant Feature Transform (SIFT) for feature extraction and Support Vector Machine (SVM) as the classification algorithm.
- PCA: Refers to Principal Component Analysis (PCA) and can be used as a preprocessing step to extract the most informative features (principal components) from the original images.
- Algorithm in this paper: Refers to the proposed method described in the paper, which is based on deep learning and transfer learning with the Inception-V3 model.

The accuracy column displays the recognition accuracy achieved by each method. This represents the proportion of correctly classified instances in the total number of instances. It is usually expressed as a percentage.

The purpose of this table is to demonstrate how the proposed method (Algorithm in this paper) compares to established methods (BP, SIFT + SVM) in terms of accuracy. The higher the accuracy, the better the method is at correctly classifying fruit images. From Table I, the proposed method utilizing deep learning and transfer learning with the Inception-V3 model outperforms the other methods in terms of accuracy, achieving a recognition accuracy of 97.7%. This indicates that the proposed method has the potential to provide more accurate fruit image classification compared to traditional methods like Backpropagation and SIFT + SVM.

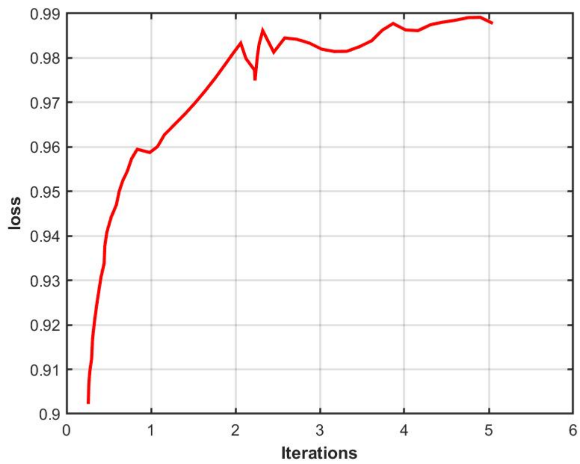

Fig. 4. The variation of accuracy during the training process.

In Table I, it can be seen that the method utilizing transfer learning with the Inception-V3 model based on CNN significantly improves the recognition accuracy compared to traditional methods. Traditional methods require rich manual experience during the feature extraction process, leading to considerable uncertainty in feature extraction. Additionally, traditional methods involve complex parameter tuning, which greatly increases the training time. By employing transfer learning with the Inception-V3 model, parameter fine-tuning can be performed within a well-established classification model, resulting in a rapid improvement in classification recognition accuracy.

V. CONCLUSION

In this paper, we proposed a novel fruit image classification method that leverages transfer learning with the Inception-V3 model of Convolutional Neural Networks

Authorized licensed use limited to: R V College of Engineering. Downloaded on July 03,2025 at 19:00:39 UTC from IEEE Xplore. Restrictions apply.

(CNN). By transferring the pre-trained CNN model to a smaller target set, we retained the original convolution layer structure and introduced a new Softmax classifier for data classification. The results demonstrated that utilizing the Inception-V3 model through transfer learning significantly improved the accuracy of fruit recognition compared to traditional fruit classification algorithms. The designed real-time fruit classification algorithm can be used for real-time recognition of six types of fruits, including apples, bananas, kiwifruit, mangoes, citrus, and pears, and has good application prospects in smart agriculture and other fields. This paper have some limitation because it focuses on six types of fruits. While the chosen fruits are commonly encountered, the model's effectiveness might decrease when faced with new or less common fruit categories. Expanding the dataset to include a broader range of fruit types would improve the model's robustness. The training and testing sets in this article both use fruit images with similar light intensity. However, in practical applications, images are often affected by factors such as lighting conditions, rotations, and occlusions. Therefore, developing techniques that make the model robust to changing environments, such as outdoor lighting conditions, would enhance its practical usability. Another way of improving the method would be designing systems that can learn new fruit categories over time without forgetting previously learned ones would be beneficial in practical scenarios where the types of fruits may expand. The method is also computationally expensive. Further research can be conducted to address these influencing factors in fruit classification and recognition.

ACKNOWLEDGMENT

Author grateful to Huayin Institute of Technology for supporting this project.

REFERENCES

[1] K. Javed Mannan, A. U. Rehman, H. A. Babri, and S. K. Noon, "Classification of degraded traffic signs using flexible mixture model and transfer learning," IEEE Access, vol. 7, pp. 148800-148813, 2019.

[2] Q. Jiang, D. Tan, Y. Li, S. Ji, C. Cai, and Q. Zheng, "Object detection and classification of metal polishing shaft surface defects based on convolutional neural network deep learning," Applied Sciences, vol. 10, no. 1, pp. 87, 2019.

[3] H. Zhu, Q. Liu, Y. Qi, X. Huang, F. Jiang, and S. Zhang, "Plant identification based on very deep convolutional neural networks," Multimedia Tools and Applications, vol. 77, pp. 29779-29797, 2018.

[4] Z. Jiang, Z. Dong, W. Jiang, and Y. Yang, "Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning," Computers and Electronics in Agriculture, vol. 186, p. 106184, 2021.

[5] A. Benhamida, A. R. Várkonyi-Kóczy, and M. Kozlovszky, "Traffic Signs Recognition in a mobile-based application using TensorFlow and Transfer Learning techniques," in 2020 IEEE 15th International Conference on System of Systems Engineering (SoSE), pp. 000537-000542, IEEE, 2020.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[7] M. Mishra, J. A. Latorre, J. Pool, D. Stosic, D. Stosic, G. Venkatesh, C. Yu, and P. Micikevicius, "Accelerating sparse deep neural networks," arXiv preprint arXiv:2104.08378, 2021.

[8] X. Chen, G. Zhou, A. Chen, L. Pu, and W. Chen, "The fruit classification algorithm based on the multi-optimization convolutional neural network," Multimedia Tools and Applications, vol. 80, pp. 11313-11330, 2021.

[9] R. Dandavate and V. Patodkar, "CNN and data augmentation based fruit classification model," in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 784-787, IEEE, 2020.

[10] K. Kang, C. Xie, C. He, M. Yi, M. Gu, Z. Chen, K. Zhou, and H. Wu, "Learning efficient illumination multiplexing for joint capture of reflectance and shape," IEEE Transactions on Visualization and Computer Graphics, vol. 38, no. 6, pp. 165-1, 2019.

[11] R. Suruliandi, J. C. Kavitha, and D. Nagarajan, "An empirical evaluation of recent texture features for the classification of natural images," International Journal of Computers and Applications, vol. 42, no. 2, pp. 164-173, 2020.

[12] S. Chakraborty, S. H. Shaikh, A. Chakrabarti, and R. Ghosh, "Quantum image edge extraction based on classical robinson operator," Multimedia Tools and Applications, vol. 81, no. 23, pp. 33459-33481, 2022.

[13] M. D. Ansari and S. P. Ghrera, "Intuitionistic fuzzy local binary pattern for features extraction," International Journal of Information and Communication Technology, vol. 13, no. 1, pp. 83-98, 2018.

[14] M. D. Ansari and S. P. Ghrera, "Feature extraction method for digital images based on intuitionistic fuzzy local binary pattern," in 2016 International Conference System Modeling & Advancement in Research Trends (SMART), pp. 345-349, IEEE, 2016.

Authorized licensed use limited to: R V College of Engineering. Downloaded on July 03,2025 at 19:00:39 UTC from IEEE Xplore. Restrictions apply.