Fruit Recognition Using Deep Learning

Mu Niu 12112004 | Zihao Chu 12111409 | Zhonghang Zheng 12110419 | Lishengqi Jiang 12110727 | Zhonghao Luo 12112315 | Chaoling Tan 12112009

***Abstract*—In this study, we utilized a high-quality image dataset containing a variety of fruits and vegetables. Through machine learning, especially the convolutional neural network (CNN) algorithm, we trained a neural network model to accurately distinguish different types of fruits and extract the relevant features of fruits. During the training process, we analyzed the training results and the content of the dataset. We found that the dataset had excessive redundancy. To solve this problem, we independently collected a new dataset and used the same training method, combined with a visual camera, to achieve accurate identification of a specific type of fruit. The camera recognition results were highly consistent with the training results. Apart from that, this failed experience of choosing datasets inspired us to seek a new way of judging the quality of a dataset without the process of machine learning. At last, we were able to establish a systematic approach to evaluate the quality of a dataset by Facets and structural similarity index (SSIM).**

I. Introduction

In this experiment, we use the convolutional neural network (CNN) model implemented by TensorFlow and Python, combined with a self-built dataset and a camera, to realize the identification of certain fruit species. Additionally, we have established a systematic approach to evaluate the quality of fruit datasets.

When using a widely adopted dataset, we achieved a high training set success rate of 98 percent and a high test set success rate of 96 percents. However, when migrating this model to images from another dataset, we observed a high error rate. This prompted us to closely examine the sample data of the original dataset.

One of the main objectives of this project is to construct a robust fruit identification database and develop a method that can accurately identify various types of fruits. It is crucial for this method to possess a certain level of transferability, allowing it to identify fruit images that are not part of the original training dataset.

Another important purpose is to establish a systematic approach to evaluate the quality of a dataset. First, we utilized Facets to extract the main features of the images in the dataset and conducted a comprehensive analysis of these features. This enabled us to make an initial assessment of the quality of the dataset. Then we use a standard dataset and compare the image color and texture similarity as a benchmark to judge the redundancy and overall quality of the dataset.

In future work, we will explore larger datasets, try more kinds and larger quantities of data, and attempt to modify the network structure to enhance the fruit recognition system. These modifications aim to improve the model's ability to extract relevant features from fruit images, thereby obtaining more accurate classification results.

A. Background

The fruit industry has emerged as the third largest industry, following grain and vegetables, in terms of economic significance. While the rapid development of the fruit industry has brought substantial economic benefits, it has also presented us with a range of challenges, with fruit classification being one of them. To achieve high-precision fruit classification, it is essential to have not only a robust feature extraction method but also a high-quality and comprehensive database. Both components are crucial for attaining accurate and reliable fruit classification results.

B. Related Work

This project mainly refers to the paper Fruit recognition from images using deep learning, which creates a dataset of fruit and vegetable images called Fruit-360. The dataset contains 90483 images covering 131 different fruit and vegetable categories, each with a resolution of 100x100 pixels and a uniform white background. The authors divided the dataset into a training set and a test set in a ratio of approximately 3:1, where the training set contains 67,692 images, the test set contains 22,688 images, and there are also 103 images that contain multiple fruits.The Fruit-360 dataset was created with the aim of providing a high-quality training and testing resource for deep learning models to better perform fruit and vegetable recognition tasks.

The paper uses a Convolutional Neural Network (CNN), which has had great success in image recognition and classification tasks, including the use of a Lambda layer to convert an RGB image to HSV color space while adding grayscale colors to enhance the model's robustness to color variations; the use of multiple convolutional layers to extract features, which makes the network deeper and capable of learning more complex feature representations; and the use of a ReLU activation layer to introduce nonlinearity so that the model can learn and simulate more complex function mappings; a pooling layer to reduce the feature dimensions, the number of parameters and computational complexity; a spreading layer and a fully connected layer to combine the local features extracted by the previous convolutional and pooling layers for the final classification task; and a Dropout layer to reduce the overfitting and to improve the model's ability to generalize over unknown data. Finally, the Softmax layer is used to convert the output of the model into a probability distribution to accomplish the multi-class classification task. With this architecture, the model is able to train on a large amount of image data and learn to recognize and classify different fruits.

The TensorFlow framework is used in the paper for the implementation, training and testing of the deep learning model, and the performance of the network is demonstrated through a series of numerical experiments. The model using the dataset of the paper was able to achieve more than 99.99 percent accuracy for the test set and more than 98percent accuracy for the test set after 25 epochs of training, and the results are shown in the following figures. From the results given in the previous paper and the results of our local deployment, after 25 epochs of training using the paper's dataset and model, the accuracy of the test set is able to reach more than 99.99 percent and the accuracy of the test set is able to reach more than 96percent. [Fig 3. Results From us] However, since the fruit samples of this Fruit-360 dataset are somewhat different from the color and shape of the actual fruits, the quality of this dataset is not high in this aspect, and there is a possibility of overfitting of the model. In our real scenario test, its generalization ability is insufficient, and the accuracy of the real scenario is much lower than the accuracy of the test set, and there is a certain degree of misclassification.

C. Introduction of Our Project

We were inspired by these papers and used it as a reference for our subsequent fruit recognition project. Our project uses the same convolutional neural network architecture for training and testing, in order to side-by-side compare how well our approach solves the problems of the previous paper. The Fruits-360 dataset (2020) Horea Muresan [Source code](https://github.com/Horea94/Fruit-Images-Dataset) mentioned in the paper, while academically valuable, is not sufficiently generalized in our real-world application scenario. In order to solve this problem, we took and created a dataset on our own that is more in line with the practical application requirements. We collected image data on three types of fruits, apple, orange and lychee, respectively. Our dataset contains images of multiple fruits of different shapes and sizes under different lighting conditions and poses to improve the generalization ability and robustness of the model. Unlike the work in the paper, our project not only focuses on the recognition of static images, but also implements the function of calling the camera for real-time fruit recognition. We performed rigorous accuracy validation on the test set to ensure the performance of the model in real applications. In addition, we have established a set of criteria for evaluating the dataset, which consists of comparing the color and texture of the images, aiming to systematically assess the quality and applicability of the dataset. Based on the above dataset evaluation criteria, we compare Fruits-360 with datasets collected online in order to measure the quality of each dataset.

In future work, we will explore larger datasets, try more kinds and larger quantities of data, in addition, we will try to modify the network structure to enhance the fruit recognition system. This involves experimentally adding or removing layers, exploring different architectures, and considering pre-trained models. These modifications are designed to improve the model's ability to extract relevant features from fruit images to obtain more accurate classification results.

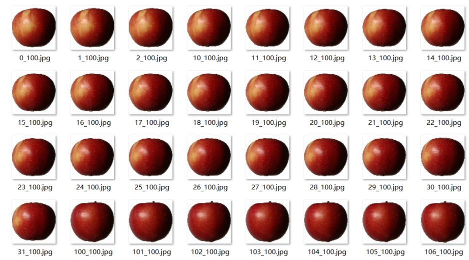

Fig. 1. Fruit-360 Data Set

Nr.	Configuration			Accuracy on training set	Accuracy on test set
1	Convolutional	5 x 5	16	100%	98.66%
	Convolutional	5 x 5	32		
	Convolutional	5 x 5	64		
	Convolutional	5 x 5	128		
	Fully connected	-	1024		
	Fully connected	-	256		

Fig. 2. Results From The Paper

Train: accuracy = 0.999941 ; loss_v = 0.000266
Test: accuracy = 0.967824 ; loss_v = 0.204051

Fig. 3. Results From us

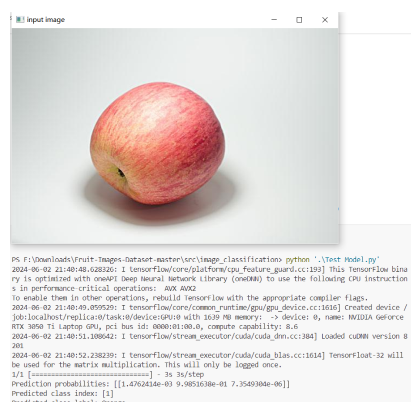

Fig. 4. Result of Original Dataset

II. Method

A. Fruit Classification Task

In this project, we follow the same convolutional neural network architecture as the Fruit-360 dataset. However, before building our own fruit image dataset, we conducted a series of experiments to explore the impact of dataset size on model performance. For this purpose, we systematically adjusted the size of the training set and recorded the accuracy and loss values of the model on the training and test sets for different dataset sizes. From the above experiments, we plotted the graphs of accuracy and loss values for the training and test sets.

The experimental graphs show that when the average number of training images per fruit category is lower than 150, the model's accuracy on both training and test sets shows a significant improvement with the increase in training set size, while the loss values are significantly reduced; whereas, when the average number of training set images used for each fruit category is greater than 150, the increase in the number of datasets has a very small optimization on the accuracy and loss values. Based on the above findings, we concluded that too large a dataset size is redundant and unnecessary, and decided to select 150 images for each fruit category to construct the training set, and 50 images as the test set, in order to use less data for training and still achieve a good test accuracy.

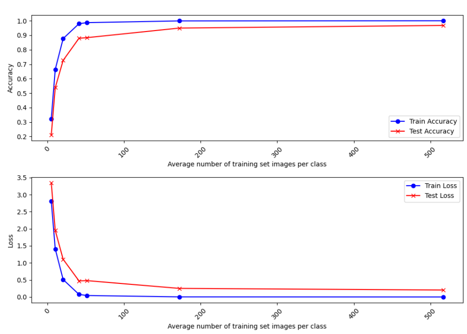

Fig. 5. Accuracy and loss graphs for training and test sets

B. Dataset Evaluation Task

In the human visual system, color and texture are the two most critical visual cues for recognizing and classifying fruits. Based on this knowledge, we consider color and texture to be the two core dimensions for evaluating the quality of a dataset. The dataset evaluation criteria we set are mainly centered on these two main elements, and we believe that a high-quality dataset should satisfy the authenticity and consistency of color and texture: Color: A high-quality dataset should contain images that accurately reflect the color of fruits in the real world. This involves not only the natural rendering of colors under different lighting conditions, but also the color variations of different individuals within the same fruit category, and the evolution of colors at different stages of ripeness. Textural aspects: the images in the dataset should also be able to realistically show the textural details of the fruit. This includes the clarity of the texture under different lighting conditions, the texture variation of different individuals in the same category, and the consistency of the texture when the same fruit is viewed from different angles. Combining these two dimensions, we believe that a good dataset should have the following characteristics:

Color accuracy: the colors of the fruits in the images should match the colors in the real environment, ensuring that the model can correctly identify the colors under variable lighting conditions.

Texture clarity: the images should provide enough detail to allow the model to capture the unique texture characteristics of the fruit surface.

Color and texture consistency: Fruits of the same category should exhibit consistent color and texture features across images, helping the model to establish stable recognition patterns.

Sample diversity: The dataset should contain fruit images with a variety of color and texture features to support the model to learn a wider range of more discriminative features.

Feature balance: color and texture samples of various types of fruits in the dataset should be balanced to avoid over- or under-representation of certain features.

So firstly, we use Facets to analyze the features of pictures in the datasets. Facets is an open-source visualization tool developed by Google that provides a powerful and intuitive way to understand and analyze datasets. By using Facets, we conducted feature extraction and analysis on several characteristics of the images within the dataset, allowing us to make an initial assessment of the dataset's quality. In our analysis, we extract several key features from the images within the dataset. Specifically, we calculated the mean color, mean hue, mean saturation, and mean value of all 32x32 pixel patches within each image.

The mean color represents the average BGR (Blue-Green-Red) color values of the 32x32 pixel patches in the image. It provides insights into the overall color composition and distribution within the image.

The mean hue measures the average hue value of the 32x32 pixel patches, indicating the dominant color tones present in the image. This feature helps to identify the primary colors or color schemes within the dataset.

The mean saturation represents the average saturation value of the 32x32 pixel patches, reflecting the intensity or purity of the colors in the image. It assists in understanding the vividness or desaturation of the colors in the dataset.

Lastly, the mean value denotes the average brightness value of the 32x32 pixel patches, which indicates the overall lightness or darkness of the pixels in the image. This feature aids in assessing the relative brightness levels within the dataset.

In addition to utilizing Facets, we also employed computer vision techniques to develop a system capable of assessing the similarity between images within a dataset. This system outputs two values representing texture and color similarity, respectively. The closer these values are to those obtained from a standard dataset, the higher the degree to which the dataset can be considered of high quality.

III. Result

A. Fruit Identification Results of the Self-Collected Dataset

Through machine learning techniques, we conducted processing on a dataset consisting of three categories and 600 pictures. The training process yielded an accuracy of 0.7496 on the training set, indicating that our model effectively learned the patterns and characteristics of the fruit images within the dataset. Additionally, the test set accuracy of 0.6552 suggests that our model performed reasonably well on unseen data, demonstrating its generalization ability.

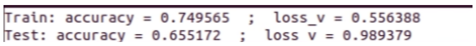

Fig. 6. Result of Deep Learning

Nr.	Configuration			Accuracy on training set	Accuracy on test set
1	Convolutional	5 x 5	16	100%	98.66%
	Convolutional	5 x 5	32		
	Convolutional	5 x 5	64		
	Convolutional	5 x 5	128		
	Fully connected	-	1024		
	Fully connected	-	256		

Fig. 7. Result of Original Dataset

B. Model Test

Based on our trained model, we conducted testing to evaluate its performance. Initially, we fed a set of self-captured images into the model. These images had relatively uncomplicated backgrounds, totaling 11 pictures. Subsequently, we utilized the model for identification purposes, and remarkably, each picture was accurately recognized.

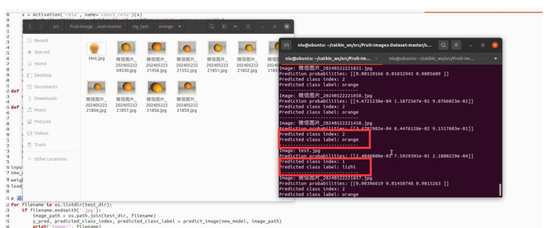

Fig. 8. Test of training results

C. Camera Physical Recognition

Visual recognition plays a vital role in product identification, offering the ability to efficiently process large volumes of data and significantly enhance processing efficiency. It has become widely utilized in machine learning applications within commercial settings.

In our study, we successfully verified the reliability of our training results by combining them with visual learning techniques. This approach allowed us to demonstrate the transferability of our dataset, showcasing its effectiveness in diverse scenarios.

By leveraging the capabilities of computer cameras, we were able to capture images of fruits directly into our model, enabling efficient detection and classification. Taking apple and lychee as examples, although some errors occurred during the detection process, the overall accuracy was found to be close to the accuracy achieved on the training set (approximately 0.7496). This result further substantiates the effectiveness of our customized identification dataset, which exhibits strong transferability and outperforms the initial dataset through optimization efforts.

The successful integration of visual learning techniques and the utilization of real-time camera input demonstrate the practical application of our model in fruit detection and classification tasks. This achievement highlights the potential of our approach to enhance the efficiency and accuracy of fruit identification processes in real-world scenarios.

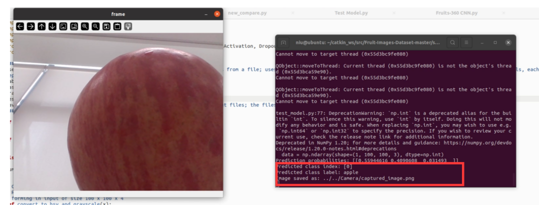

Fig. 9. The Camera Recognizes the Apple

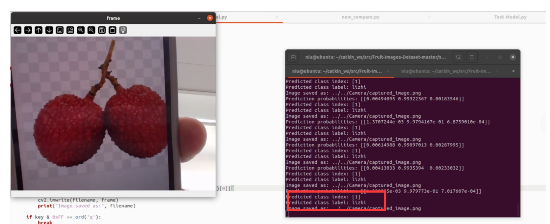

Fig. 10. The Camera Recognizes the Lychee

D. Dataset Evaluation

1) Facets: During our initial analysis using facets, we focused on extracting key features from the images. Specifically, we calculated the mean color, mean hue, mean saturation, and mean value of all 32x32 pixel patches within each image. These features provided us with essential information on the color distribution, color tones, and overall brightness levels present in the dataset.

Fig. 11. Facets Overview

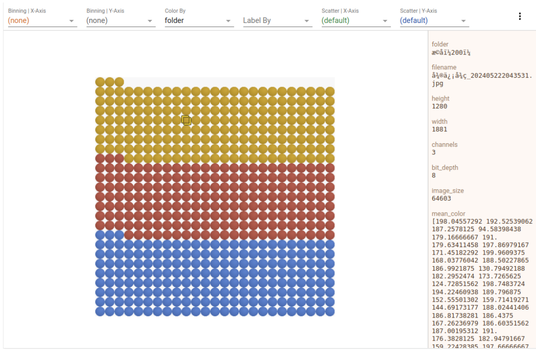

Fig. 12. Facets Deep

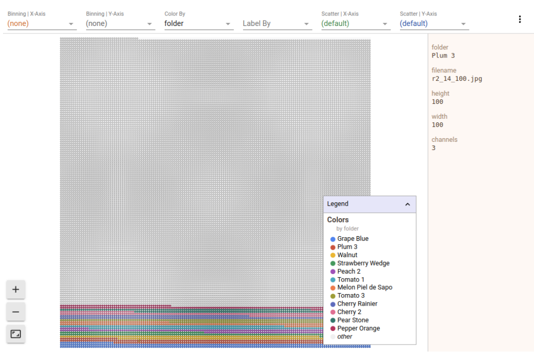

Fig. 13. Facets Deep of Fruit 360

By comparing the training dataset information obtained from Facets with the data information from a standard dataset, we can make an initial assessment of the quality of a dataset. One of the advantages of this approach is its low computational cost and high efficiency.

2) Index of Similarity: The evaluation of our dataset primarily focuses on analyzing similarity, particularly regarding color and texture. Color and texture are crucial visual and tactile cues that aid in fruit recognition and differentiation. Color is typically the initial feature we notice when observing fruit, providing essential visual information about their diversity. Texture, on the other hand, provides tactile details concerning the surface and internal consistency of the fruit. By comparing the colors and textures of different fruits, we can swiftly assess their similarities and distinctions. Similar fruits may exhibit variations in color and texture, aiding in their differentiation. For instance, apples and tomatoes, despite both being round fruits, differ in color (ripe apples are red, ripe tomatoes are also red) and texture (apples are smooth and waxy, while tomatoes are smooth and slightly soft), facilitating their identification.

To evaluate the color similarity and texture similarity within our dataset, we employed methods such as comparing image color histogram correlation and utilizing the structural similarity index (SSIM) for assessing structural similarity. By applying weighted processing, we obtained measurements for color similarity and texture similarity within the dataset.

We discovered a dataset on Kaggle that contains a variety of apples and possesses sufficient data with simple backgrounds. This dataset was utilized as a benchmark or standard dataset for comparison. By inputting our original dataset into the model, we observed that the dataset exhibited high texture similarity, approximately 30 percent higher than the benchmark dataset. This indicates that the dataset contains redundant data and necessitates the inclusion of additional apple types to expand its diversity.

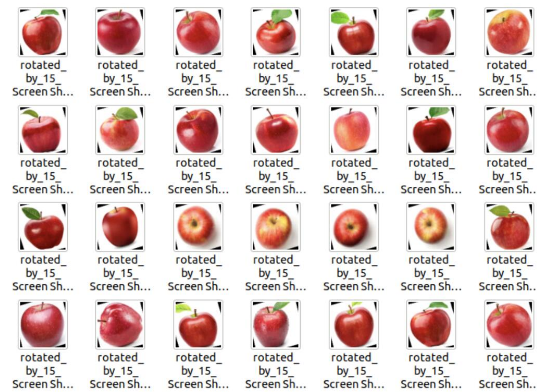

Fig. 14. Standard dataset

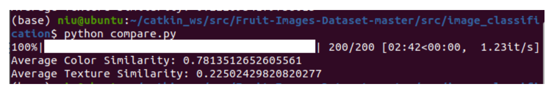

Fig. 15. Similarity of the standard datasets

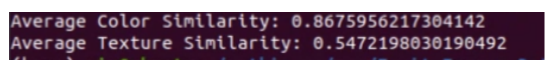

Fig. 16. Similarity of the original datasets

IV. Future Work

A. Expansion of dataset

Although we have optimized the original dataset, the existing dataset still has drawbacks, We lacked more diverse samples, such as the picture of other kinds of apples, photos of apples of different maturity. In the future work we aim to collect a larger and more diverse dataset that includes a broader range of fruit species and variations in environmental conditions, such as lighting and background complexities. This expansion will help improve the model's ability to generalize across different scenarios and reduce the likelihood of overfitting.

B. Network Architecture Optimization

Experimenting with different CNN architectures, such as ResNet, Inception, or EfficientNet, could provide insights into how architectural variations impact performance. Additionally, we plan to explore the use of transfer learning with pre-trained models to leverage existing knowledge and enhance feature extraction capabilities.

Fig. 17. Self-collected datasets

C. Expand the evaluation indicators and tools

Further refinement of the dataset evaluation criteria using additional metrics and tools beyond Facets and SSIM can provide a more comprehensive assessment of dataset quality. Incorporating measures for dataset bias, imbalance, and representativeness will be crucial for ensuring high-quality data.

V. Conclusion

In this study, we addressed the challenge of fruit classification using convolutional neural networks (CNN) by leveraging a high-quality image dataset and self-built datasets to improve model accuracy and generalization. Initially, we employed the widely recognized Fruit-360 dataset and achieved impressive accuracy rates of 98 percent on the training set and 96 percent on the test set. However, we observed significant performance degradation when applying the model to real-world images, indicating insufficient generalization and potential overfitting due to dataset redundancy.

To mitigate these issues, we independently collected a new dataset comprising images of apples, oranges, and lychees under various conditions to enhance diversity and realism. By training our CNN model with this dataset, we observed improved accuracy and robustness, with training and test set accuracies of 74.96 percent and 65.52 percent, respectively. Furthermore, the real-time fruit recognition using a visual camera demonstrated consistency with our training results, validating the practical applicability of our approach.

We also proposed a systematic method for evaluating dataset quality using Facets and the Structural Similarity Index (SSIM). By analyzing key image features such as color and texture, we identified and addressed redundancy in the original dataset. Our findings emphasized the importance of dataset diversity and realism in achieving reliable fruit classification.

In future work, we plan to explore larger and more varied datasets, refine network architectures, and experiment with pre-trained models to further enhance the performance and generalization of our fruit recognition system. Our research underscores the critical role of high-quality datasets in machine learning and provides a framework for their evaluation and improvement.

References

[1] Bargoti, S., and Underwood, J. Deep fruit detection in orchards. In2017 IEEE International Conference on Robotics and Automation (ICRA)(May 2017)

[2] Cheng, H., Damerow, L., Sun, Y., and Blanke, M. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. Journal of Imaging 3, 1 (2017).

[3] Li, D., Zhao, H., Zhao, X., Gao, Q., and Xu, L. Cucumber detection based on texture and color in greenhouse. International Journal of Pattern Recognition and Artificial Intelligence 31 (01 2017).

[4] Rahnemoonfar, M., and Sheppard, C. Deep count: Fruit counting based on deep simulated learning. Sensors 17, 4 (2017).

[5] Xiong, J., Liu, Z., Lin, R., Bu, R., He, Z., Yang, Z., and Liang, C. Green grape detection and picking-point calculation in a night-time natural environment using a charge-coupled device (ccd) vision sensor with artificial illumination. Sensors 18, 4 (2018).

[6] H. Muresan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Informatica*, vol. 10, no. 1, pp. 26–42, 2018.