Classification of Fruit Images as Fresh and Rotten Using Convolutional Neural Networks

Tuğçe Göksu
Computer Engineering Department
Fatih Sultan Mehmet Vakıf University
İstanbul, TURKEY
tkocak@fsm.edu.tr

Zeliha Kaya
Computer Engineering Department
Fatih Sultan Mehmet Vakıf University
İstanbul, TURKEY
zkaya@fsm.edu.tr

Shaaban SAHMOUD
Computer Engineering Department
Fatih Sultan Mehmet Vakıf University
İstanbul, TURKEY
ssahmoud@fsm.edu.tr

***Abstract*— Many fruits are produced all over the world, and the fruits produced are sent abroad and sold in a relatively short time in many countries. During the period between collection and sale, the fruits may undergo various types of spoilage. In many cases, the fruits go through a number of fabrication processes for sale. If the fruits comply with certain standards after passing the quality tests at the factory, for example, if the fruits have not started to rot, shipment can begin. In this study, we consider the problem of detecting any type of rot/mold as a binary classification problem. To train and test different convolutional neuralnetwork models, we have collected a dataset using images of rotting /spoiled and fresh fruits. Different deep-learning models have been implemented and tested to solve this problem. In this study, the ResNet50 network architecture with real-time data augmentation obtained the best results. Our experimental results show that the ResNet50 model can solve this problem with an accuracy of around 90%.**

Keywords – Image classification, ResNet50, Convolutional Neural Networks

I. Introduction

Image classification is one of the main topics in the field of computer vision and is widely used. With the spread of deep learning, significant improvements in accuracy in image classification have been achieved and it has spread to a wide range of applications such as object detection, face recognition, and medical imaging.

In the food industry, the freshness of fruits and vegetables is an important feature of the products. If rotten fruit is not separated from other fruit, it can damage fresh fruit and affect productivity. Traditionally, fruits and vegetables mustbe visually inspected by trained personnel for quality assessment as products or crops. However, manual classification requires specific knowledge of many characteristics of fruits and vegetables. Also, maintaining consistency in manual classification is a common challenge.

Recently, deep convolutional neural networks (CNNs) have demonstrated remarkable capabilities by achieving the highest levels of performance in image classification tasks across various domains. Therefore, we have utilized different deep-learning algorithms for solving the freshness binary classification problem. In this study, a data setwas gathered by collecting fruit and vegetable images from the internet. Our results show that the ResNet50 architecture, whichis considered one of the most popular CNNs can achieve very good results. ResNet50 [1] architecture is one of the best-performing networks among convolutional neural networks in image classification tasks. The ResNet architecture has achieved high performance on the ImageNet dataset and has been used as a benchmark model for image classification tasks in many subsequent articles. ResNet50 was later used as a feature extractor for semantic segmentation [2]. In a study conducted in 2018 [3], ResNet50 was used as a feature extractor for object detection, and a one-step optimization method was proposed to improve the performance. By reviewing the literature, many studies can be found for discussing ResNet50 applications in different fields[2][3].

The organization of the paper is as follows. In Section 2, the recent fruit classification studies in the literature using CNN models are summarized. In the thirdsection, the features of the used dataset in the study are described. The proposed architecture of the artificial neural network for the considered classification problem is explained in Section IV. The obtained results and discussions are given in Section V. Finally, Section VI presents the conclusion of this research paper and gives some future work recommendations.

II. Related Work

Classifying fruits according to whether they are fresh or rotten is one way to make agriculture easier. However, this process is quite challenging. The similarity of the fruits can affect the human power-based performance classification. In this section, we focused on the research related to the classification of various fruits and determining whether the fruits used are fresh or rotten using CNN networks.

A CNN model was proposed to prevent the spread of decay in [4]. The proposed model classifies fresh and rotten fruit from input fruit images. Three types of fruit were considered in this study which are apples, bananas, and oranges. A Convolutional Neural Network (CNN) was used to extract features from input images of fruits, and then the Softmax function was applied to separate these images into two classes: fresh and rotten fruits. [4]. A CNN-based model focused on creating a transfer learning approach was introduced in [5]. VGG16, VGG19, MobileNet, and Xceptions learning models were used in this study and accuracies were compared with the proposed CNN model. 97.82% accuracy value was obtained for the proposed CNN model. Using the same dataset, another study [5] obtained an accuracy value of 98.89% using the resnet50 model. The time required to classify the single fruit image is approximately 0.2 seconds.

Again, using the same data set, the authors in [6] obtained the highest training accuracy value in the MobileNet V2 model. This value is 99.46%. Unlike the other papers, 5658 images belonging to 10 classes were classified as fresh and rotten in [7]. In the proposed classification system, five CNN models were studied. The best result was obtained with the Inception V3 model with an accuracy value of 97.34%.

Based on the difficulty and complexity of the fruit classification problem, for example, some apple and peach varieties are very similar to each other. In [8], the authors proposed a convolutional neural network-based fruit auto-recognition and classification method. First, a two-color fruit image dataset (general dataset and self-made dataset) was obtained. The publicly accessible datasets featured fruit images captured against a plain backdrop, whereas the self-created dataset included fruit images taken in intricate surroundings. Subsequently, numerous research experiments were conducted by fine-tuning various parameters using a convolutional neural network. As a result, the highest average classification accuracy achieved across the entire dataset reached 99.8%. The classification accuracy of the dataset created by adopting appropriate data development techniques was increased from the original 90.2% to 98.9%.

In addition, a fruit classification framework was proposed with a 6-layer CNN network [9]. This network comprises convolution layers, pooling layers, and fully connected layers. The dataset includes 9 fruit species, each with 200 images. By comparing the CNN network designed in the study with the advanced model (V-SVM and GA), the accuracy value was 91.44%. The authors suggested enhancing the parameters of the proposed CNN by adjusting them using advanced algorithms to further improve the classification performance.

By increasing the diversity of the data set, fresh and rotten classification was made with CNN networks using fruits and vegetables [10]. The network under consideration is comprised of four convolutional layers, and the accuracy for classifying ranged from 97.74% to 99.92%.

According to the presented related works, the problem of rotten detection still needs much improvement from different perspectives. One of the challenges of the current research work is that the variety of fruit in most existing studies is very limited and normally includes two to five types of fruits or vegetables. Moreover, there are still many new CNN network models that have not been tested or analyzed.

III. DATASET

Our dataset consists of a wide variety of fruit images collected from the Internet. We classify the fruits as fresh and rotten according to their robustness. Sample images from both classes are given in Figure I.

The images in the dataset are RGB colored with different dimensions. We resize all images to be 224 x 224 dimensions. In our dataset, 524 different images are collected. Although this number is not enough, by using augmentation techniques it can become convenient for CNN architectures. In our dataset, 55% of images belong to fresh fruit images and 45% of images belong to rotten fruit images. There are 32 different kinds of fruit like strawberry, pomegranate, and quince.

Since the collected images were not sufficient for training the convolutional neural network, a data duplication strategy was applied. The ImageDataGenerator method in Python's Keras library is used for this process. ImageDataGenerator is a generator for images and creates image data as chunks by performing real-time data augmentation. This function can increase the number of images by applying a number of data duplication techniques such as random rotations, pans, zooms, and flips to the original images. This allows the model to be trained on a larger dataset without having to manually retrieve and label more data. Output images that produced by the manufacturer have the same output dimensions as input images. This method does not increase the number of images in the dataset, it simply creates new ones by applying data augmentation techniques to images already existing in the dataset. In this way, the model can see different image variations of the same image and can generalize better.

The parameter values used in the ImageDataGenerator method include a rotation range of 40, a width shift range equal to 0.2, a height shift range equal to 0.2, and a shear range set to 0.2. The zoom range is also set to 0.2 with a channel shift range equal to 10.

Fresh	**Rotten**
	
	
	

Fig.1. Sample images from the collected dataset.

IV. PROPOSED METHOD

The ResNet50 architecture used in our study is a variant of the ResNet architecture characterized by residual connections. In a traditional CNN, the deeper the network, the harder it is to train due to the vanishing gradient problem. The ResNet architecture helps with this problem by providing residual connections that allow information to bypass multiple layers and reach deeper layers more easily.

The ResNet50 architecture consists of 50 layers, including convolutional layers, normalization layers, and ReLU activation layers. The ResNet50 architecture consists of multiple convolutional blocks, each of which includes either convolutional layers or identity blocks, and within each of these blocks, there are three convolutional layers. In total, ResNet50 has more than 23 million parameters that can be trained. The architecture is divided into 6 stages, each stage consisting of several residual blocks.

- The first stage is a convolutional layer with 7x7 kernels, 2 stride values, and 64 filters.
- Next comes the maximum pooling layer with a step size of 2.
- The second stage consists of 3 residual blocks, each containing 3 convolutional so giving us 9 layers in this step. There are 1x1 64 kernels, 3x3 64 kernels, and at the end a 1x1 256 kernels.
- The third stage consists of 4 residual blocks, each containing 3 convolutional layers so giving us 12 layers in this step. There are 1x1 128 kernels, 3x3 128 kernels, and at the end a 1x1 512 kernel.
- The fourth stage consists of 6 residual blocks, each containing 3 convolutional layers so giving us a total of 18 layers. There is a 1x1 256 kernel, a 3x3 256 kernel, and finally 1x1 1024 kernels.
- The fifth stage consists of 3 residual blocks, each containing 3 convolutional layers so giving us a total of 9 layers. There are 1x1 512 kernels, 3x3 512 kernels, and finally, 1x1 2048 kernels.
- The final stage is a fully connected layer with a global average pooling layer and a softmax activation for classification.

The ResNet50 architecture, model shown in Figure II also uses short-cut connections that allow the input to be added directly to the output of a convolutional block; this makes the learning process more efficient, and the network can better learn low-level features as information flows between layers.

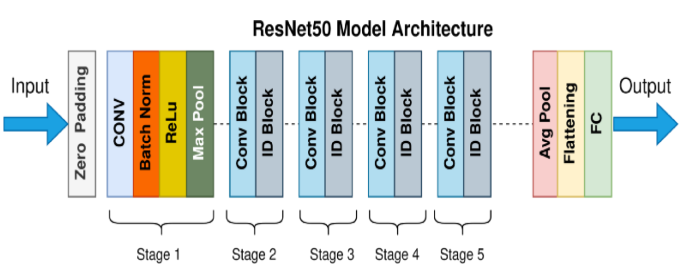

Fig.2. Architecture of ResNet50.

V. RESULTS

The dataset is divided into two parts 80% for training and 20% for testing. To evaluate the performance of the different considered CNN models, four performance metrics are used: accuracy, f1-score, precision, and recall. In addition, five different CNN models are considered in our comparative study which are ResNet50, DenseNet, MobileNet, Inception, and VGG16. These models represent the most common and popular architectures that are successfully used to solve many image processing and computer vision problems. The results of the testing process are given in Table I. A brief description of the metrics is given below.

Accuracy represents how many correctly classified data samples are in the total number of data samples.

$$accuracy = \frac{TN + TP}{TN + FN + TP + FP}$$

Precision represents the proportion of positive predictions that are correctly identified as belonging to the positive class.

$$precision = \frac{TP}{TP + FP}$$

The recall represents the proportion of actual positive instances correctly identified among all the positive instances present in the dataset.

$$recall = \frac{TP}{TP + FN}$$

The F1 Score value represents the harmonic average of Precision and Recall values.

$$F1\,Score = 2 * \frac{precision * recall}{precision + recall}$$

All models were trained using 30 epochs with a batch size equal to 8. The accuracy value for the best model which is ResNet50 is 98% in the training phase and 90.8% in the testing phase.

TABLE I COMPARISON OF THE RESULTS OF DIFFERENT CNN MODELS USING FOUR PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1 score
ResNet50 [1]	90.8%	84.0%	87.0%	85.0%
DenseNet [11]	87.3%	87.0%	87.0%	87.0%
MobileNetV2 [12]	88.5%	89.0%	89.0%	89.0%
InceptionV3 [13]	87.3%	85.0%	88.0%	86.0%
VGG16 [14]	80.4%	83.0%	83.0%	83.0%

We used five different convolutional neural network (CNN) models, namely VGG16, InceptionV3, DenseNet, MobileNetV2, and Resnet50, to perform the classification task between fresh and rotten fruits. Our suggested ResNet50 model obtained the highest accuracy with 90.80%. Thanks to the ResNet50 architecture, it increases accuracy by using residual blocks. It constantly tries to learn new features while decreasing the effect of any overfitting because of the short-used links between layers.

[1]We have released the collected dataset and trained models in: https://github.com/TugceKocak/Fruit-image-classification-as-as- fresh-and-rotten.

The second-best accuracy result was obtained from the MobileNetV2 model which commonly gets acceptable results in many real-world problems. The lowest accuracy value was obtained by the VGG16 model. This is an expected situation since the VGG16 is an old and slow model and sometimes undergoes too many overfitting scenarios.

Fig.3. Two examples of falsely classified fruit images.

Figure III shows two images that our model could not predict correctly. In the first image, there is one rotten cherry in the cherry image on the left which introduces some difficulties to the algorithm since there is one good and one rotten fruit. At the same time, rotten fruits sometimes become like other types of fruit. In this image, the bright and vivid colors of the image also help in misleading the model and classify the image as fresh. The malt plum image on the right contains fresh fruits. However, the background blackness caused the model to make an incorrect prediction and classify it as rotten.

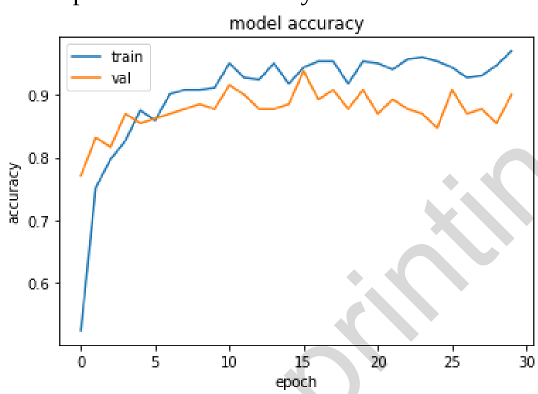

Fig.4. A plot to show the training and validation accuracies during training iterations.

Figure IV shows the accuracy values of training and validation depending on the number of epochs. The curves in the graph converge to 1. As shown in the figure, the ResNet50 is trained properly without overfitting.

VI. CONCLUSION

Classification of fresh and rotten fruits is an important task for agricultural quality and human health. In our study, we used different CNN networks to classify fruits as fresh and rotten. We solved this problem using VGG16, DenseNet MobileNetV2, ResNet50, and InceptionV3 models. The accuracy values of the models increased when different hyperparameters, namely batch size, number of epochs, optimizer, and learning speed were improved. The highest accuracy value was obtained from the ResNet50 model. The network could be trained even though the data set had high diversity and low image. According to these results, the data set can be classified as fresh and rotten fruits with 90.80% accuracy. Thus, manpower can be reduced, and the classification process can be automated.

In our future work, we plan to enhance the classification performance by fine-tuning the parameters of our CNN models using advanced algorithms. Furthermore, we intend to validate our approach using more extensive datasets through additional testing. The variety of fruits in the data set we used is high and we plan to increase them to represent all types. Moreover, the best color model for fruit image classification will be analyzed as done in [15] to check if it can enhance performance or not. Handling the images directly or after segmenting [16] is another approach that needs to be examined and tested [17].

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016

[2] L. Chen, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence 40.4 (2017): 834-848.

[3] S. Zhang, et al. "Single-shot refinement neural network for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018

[4] S. S. S. Palakodati, V. R.R. Chirra, D. Yakobu, and S. Bulla, "Fresh and Rotten Fruits Classification Using CNN and Transfer Learning." Rev. d'Intelligence Artif. 34.5 (2020): 617-622.

[5] Foong, C. C., Meng, G. K., & Tze, L. L."Convolutional Neural Network based Rotten Fruit Detection using ResNet50." 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC). IEEE, 2021.

[6] Chakraborty, S., Shamrat, F. J. M., Billah, M. M., Al Jubair, M., Alauddin, M., & Ranjan, R. (2021, June). Implementation of deep learning methods to identify rotten fruits. In 2021 5th international conference on trends in electronics and informatics (ICOEI) (pp. 1207-1212). IEEE.

[7] M. Miah, T. Tasnuva, M. Islam, M. Keya, M. Rahman, and S. Hossain, An advanced method of identification fresh and rotten fruits using different convolutional neural networks. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE..

[8] L. Wu, et al. "Fruit classification using convolutional neural network via adjust parameter and data enhancement." 2020 12th International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2020.

[9] S. Lu, "Fruit classification based on six-layer convolutional neural network." 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). IEEE, 2018.

[10] S. Jana, P. Ranjan Parekh, and S. Bijan, "Detection of Rotten Fruits and Vegetables Using Deep Learning." Computer Vision and Machine Learning in Agriculture (2021): 31-49.

[11] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

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

[15] S. Sahmoud, and H. Fathee, Fast iris segmentation algorithm for visible wavelength images based on multi-color space. In Advanced Concepts for Intelligent Vision Systems: 20th International Conference, ACIVS 2020, Auckland, New Zealand, February 10–14, 2020, Proceedings 20 (pp. 239-250). Springer International Publishing.

[16] S. Sahmoud, Enhancing iris recognition. Islamic University, Computer Engineering Department, M. Sc. Thesis. 2011.

[17] M. Hoai, Z. Z. Lan, and F. De la Torre. Joint segmentation and classification of human actions in video. In CVPR 2011 (pp. 3265-3272). IEEE.