IOP Conf. Series: Journal of Physics: Conf. Series **1327** (2019) 012050 doi:10.1088/1742-6596/1327/1/012050

A fruits recognition system based on a modern deep learning technique

Dang Thi Phuong Chung[1] and Dinh Van Tai [2]

[1] Postgraduate Student, National Research Tomsk State University, Tomsk, Russia
[2] Postgraduate Student, National Research Tomsk Polytechnic University, Tomsk, Russia

dangthiphuongchung2018tsu@gmail.com

Abstract. The popular technology used in this innovative era is Computer vision for fruit recognition. Compared to other machine learning (ML) algorithms, deep neural networks (DNN) provide promising results to identify fruits in images. Currently, to identify fruits, different DNN-based classification algorithms are used. However, the issue in recognizing fruits has yet to be addressed due to similarities in size, shape and other features. This paper briefly discusses the use of deep learning (DL) for recognizing fruits and its other applications. The paper will also provide a concise explanation of convolution neural networks (CNNs) and the EfficientNet architecture to recognize fruit using the Fruit 360 dataset. The results show that the proposed model is 95% more accurate.

1. Introduction

Considering the rapid advancement of the human race, a significant concern is given to the foods that we consume. Different techniques have been used over the past years for fruit recognition using Computer vision technology. One of the most notable applications is the use of DNN to identify, classify, and differentiate between different kinds of fruits from a dataset of images show that they outperform other algorithms. Moreover, DNNs are used in a wide range of application and provide optimal solutions for the problems encountered in multiple domains such as image analysis, speech recognition, forecasting, prediction, large dataset analysis, and marketing [1].

Currently, the most common artificial neural network (ANN) type used across multiple domains is the CNN. CNNs are used for classifying 2-D input images and recognizing the objects based on pooling and convolution layers. ANN architecture consists of three layers, starting with the input layer, followed by the hidden layer and ending with the output layer. Each layer is composed of multiple neurons. The input to each of the next layer's neurons consists of the summation of the output of the neurons from the previous layer. The output is compared with the target values based on the cost function. This is important because accurately recognizing fruits is of paramount importance in the yield mapping field. In this paper, an optimal scheme is introduced for differentiating between a variety of fruits using a dataset, which is accessible and simulates real-time prediction using EfficientNet.

The structure of the paper is as follows: section 1 will deal with the applications of DL in identifying fruits. Section 2 will elaborate on the topic of CNNs. The architecture of EfficientNet will be highlighted in more detail in the following paragraph. Section 3 will explain the reason for choosing the dataset.

Section 4 presents the prediction outcome following the results and will discuss the potential for future improvements.

2. Deep learning algorithms

Deep Learning is the sub-field of Machine Learning, which is the sub-field of Artificial Intelligence. It is a collection of techniques that model high-level abstractions in data. In deep learning, a computer-based statistical model understands and learns from pictures, sound, or text to conduct analysis. These models can attain state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers in term of accuracy [2].

While the concept of deep learning was first put forward back in the 1980s, the idea subsequently became popular because of two reasons: it needs a huge amount of labeled data and substantial computing power. The number of deep learning applications has been experiencing research growth in the last decade, including natural language processing, image classification, and information retrieval, etc. The deep learning term could be divided into two parts and understand them individually: deep and learning. Learning is about taking previous understanding and information and creating an inner depiction of the matter that the agent can use to act. Typically, the internal depiction is a compact representation for summarizing the data. The field of Machine Learning offers different functions and techniques for learning automatically from the available information, and this learning from the information is used for forecasting and projections in the future [2].

Artificial Neural Network gets the inspiration from the human's brain system and is the most commonly deployed algorithm in the field of Machine Learning [3]. It consists of integrated processing units named as neurons. ANN is comprised of input, hidden, and output layer. Input layer takes an input, for example, an image and passes it to the hidden layer, and then the output layer gives output - the maximum probability that what object in an image is. We can have multiple hidden layers for more complex functions.

2.1. Convolutional neural networks

For visual analysis, CNN, a neural network from the DNN class, is widely used. CNNs are generally viewed as feed-forward neural networks (FFNN) which can quickly identify, classify, and recognize any features in an image. The first CNN, commonly known as LeNet, was created by Yann LeCun in 1988, a member of the AI research group at Facebook. In CNNs, the network input consists of image pixel values carrying different weights depending on the feature that needs to be extracted as defined in the hidden layer. In an input image, CNNs are also comprised of fully connected layers to recognize different items, despite the pooling and convolution layer.

The convolution operation is performed in a CNN classifier over pixels in an image. It consists of four of the commonly used layers. The first one is the convolution layer, which is tasked with convolving the pixels in an image with a chosen kernel (Harris) to extract or remove different features. The second one is the ReLU layer, which defines an activation function, which can be a sigmoid or any non-linear function. The image is passed several times between the convolution, and ReLU layers, where all the negative pixel are converted to zero and trends and attributes are analyzed in an image. The third layer is known as the Pooling layer, and the main purpose of this layer is to transform the image into the required dimension without blurring it. For that purpose, the pooling layer encompasses different kernels to identify the sharp edges and to detect different contours in an image. The image is then transformed into a 1-D linear matrix. The last layer is the fully connected one, which is used to identify the images and classify them as per the accuracy (confidence value) achieved. A typical architecture of the CNN is shown in Figure 1.

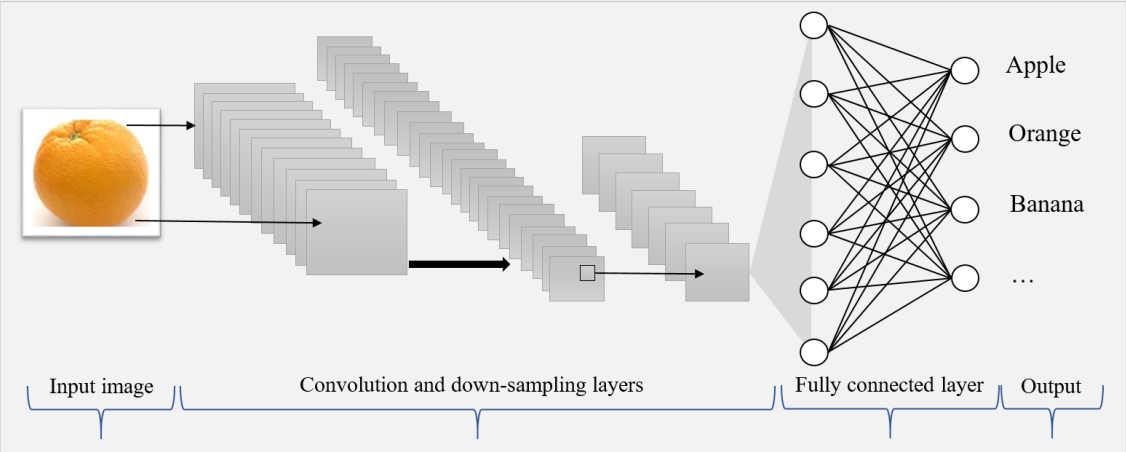

Figure 1. A convolutional neural network.

2.2. EfficientNet

EfficientNet use pre-trained convolution neural networks for conducting image related functions as a base network. These base networks are capable of learning from a wide range of dataset so that more particular models with restricted training data can be created more quickly [4]. Such networks are useful for functions such as classification of images and facial recognition that provides benefits for high-use situations as well as the use of more precise and efficient models.

While the conventional arbitrarily defined scaling, process nevertheless provides functional results. EfficientNet initially conducts a grid search of the base network to determine the relationships between the different scaling dimensions of the network while considering both model size and available computational resources [5]. In most situations, the findings of the initial testing show a higher amount of precision and velocity. The following table demonstrates the summary of the architecture of EfficientNet-B0:

Table 1. EfficientNet-B0 architecture.

Stage *i*	**Operator** *Fi*	**Resolution** *Hi x Wi*	**Channels** *Ci*	**Layers** *Li*
1	Conv3x3	224 x 224	32	1
2	MBConv1, k3x3	112 x 112	16	1
3	MBConv6, k3x3	112 x 112	24	2
4	MBConv6, k3x3	56 x 56	40	2
5	MBConv6, k3x3	28 x 28	80	3
6	MBConv6, k3x3	14 x 14	112	3
7	MBConv6, k3x3	14 x 14	192	4
8	MBConv6, k3x3	7 x 7	320	1
9	Conv1x1 & Pooling & FC	7 x 7	1280	1

3. Dataset

For training and testing, all the pictures were chosen from the fruits 360 dataset, which is publicly available on Kaggle. The dataset contains 77917 different fruits pictures of 103 categories [6]. The fruits pictures were received by registering the fruits while a motor revolves them and then producing frames. A white paper is placed behind the fruits was used as a background. Due to the disparity in the lighting, a flood-fill algorithm was applied to extract the fruit from the background. After removing the

background, all the fruits were resized to 100×100 pixels of standard RGB pictures [6]. From the fruits-360 dataset, we selected 17624 pictures from 25 different categories. We used 13218 images (75%) to create the training set and the rest 4406 images (25%) for testing the model [7]. Table 1 shows the 25 categories of fruits we used for analysis.

Table 2. Using categories of fruits.

Name of fruits	Number of training images	Number of testing images
Apple Golden	492	164
Apple Granny Smith	492	164
Apple Red	492	164
Apricot	492	164
Avocado	426	142
Banana	489	163
Cherry	492	164
Cocos	489	163
Grape Blue	984	328
Grape White	489	163
Grapefruit Pink	489	163
Kumquats	489	166
Lemon	492	164
Limes	489	163
Mandarine	489	163
Mango	489	163
Orange	480	160
Pear	492	164
Pepper Green	444	148
Pepper Yellow	666	222
Pepper Red	666	222
Strawberry	492	164
Tomato	738	246
Pineapple	489	163
Kiwi	468	156

4. Experimental results

In this paper, we have applied EfficientNet-b0 on Fruit Dataset to discover the better classification performance of the network. From Fruits 360 dataset, we have taken 17624 images from 25 different categories: 75 % of the images from these are used for training, and 25 % are used for testing the model.

The network is trained for 35 epochs with a batch size of 20. We compared our model with the present state-of-the-art models and results were exceptional. The accuracy of the proposed model was 95.67 %. The comparison of the proposed model with the conventional models shows that the results of our model are exceptionally good and promising to use in real-world applications. This sort of higher accuracy and precision will work to boost the machine's general efficiency in fruit recognition more appropriately. As a prototype, a program was developed in Python with PyQt library in a Visual Studio environment. The appearance of the program is shown in Figure 2.

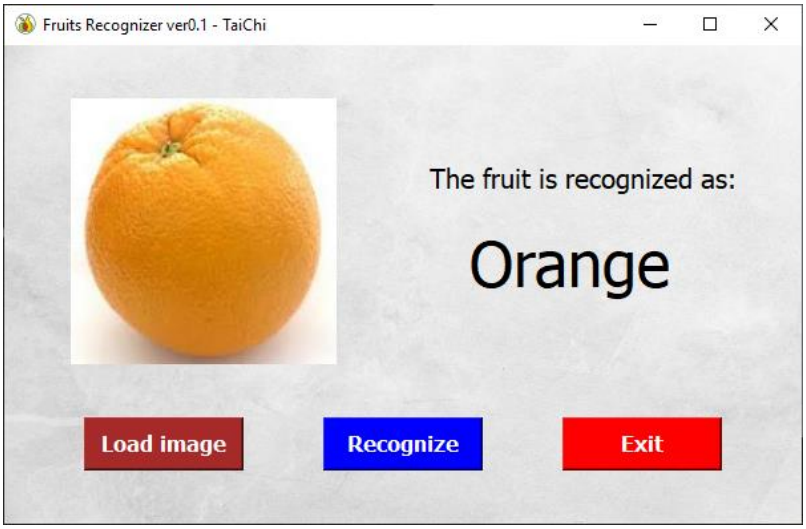

Figure 2. Main window of the program.

5. Conclusion

This paper explores a fruits recognition classifier based on EfficientNet algorithm. The recognition rate has dramatically improved throughout the experiment. Among all the cases, the model achieved the best test accuracy of 98% in case 4 from 11 to 15 epochs and best training accuracy of 96.79% at epoch 13. This type of higher accuracy will cooperate to stimulate the overall performance of the machine more adequately in fruits recognition. In the future, our plan is to improve recognition system by extending its functions to process and recognize more variety of different fruit images.

References

[1] Rocha A, Hauagge D C, Wainer J, Goldenstein S 2010 Automatic fruit and vegetable classification from images *Comput. Electron* **70** 96–104

[2] I Sa, Z Ge, F Dayoub, B Upcroft T. Perez, and C McCool 2016 Deepfruits: A fruit detection system using deep neural networks *Sensors* **16(8)** 1222

[3] L Deng, G Hinton, and B Kingsbury 2013 New types of deep neural network learning for speech recognition and related applications: An overview *IEEE International Conference on Acoustics, Speech and Signal Processing* 8599–8603

[4] Y LeCun and Y Bengio Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks* **3361(10)** 1995

[5] M. Tan and Q. V. Le 2019 EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks *arXiv preprint arXiv:1905*.11946

[6] H. Muresan, M. Oltean 2018 Fruit recognition from images using deep learning, *Proceeding of the Acta Univ. Sapientiae, Informatica* **10(1**) 26–42

[7] Mekhtiyev A D, Yurchenko A V, Bulatbayev F N, Neshina Y G and Alkina A D 2018 Theoretical bases of increase of efficiency of restoration of the worn out hinged joints of mine hoisting machine *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences* **5(431)** 66-75