**Fruit and Vegetable Detection using OpenCV and Deep Learning**

**Abstract:**

A novel approach for detecting fruits and vegetables is introduced in this report, which utilizes advanced techniques of OpenCV and Deep Learning. The approach is based on two advanced models, namely MobileNetV2 and Mask-RCNN, for object detection and segmentation, respectively. This article suggests a simple deep learning model that utilizes a pre-trained MobileNetV2 model and an attention module. The model first extracts convolution features to capture object-based information at a high level, and then uses an attention module to identify important semantic information. These two modules are combined to fuse both types of information, which is then processed by fully connected layers and a softmax layer. The proposed method is evaluated on three fruit-related benchmark datasets, and the results indicate that it outperforms the four most recent deep learning methods in terms of classification accuracy while using fewer trainable parameters. This method employs transfer learning. The model was trained on a dataset that contained diverse types of fruits and vegetables and showed impressive results with a mean squared error (MSE) loss of 0.1790 and an accuracy of 94.17%. The validation set also demonstrated promising outcomes, with a loss of 0.1094 and an accuracy of 95.21%. This method can be used in various situations such as smart farming and automated quality control in the food industry, where the accurate detection and classification of fruits and vegetables are crucial. The experimental findings suggest that the proposed technique is effective and has the potential to enhance the detection of fruits and vegetables.



Fig 1. The motivation for fruit and vegetable detection using automation and machine learning.

**Objectives:**

The primary objective of this project is to develop a fruit detection system that can accurately detect and classify fruits in real time. The system should be able to detect fruits of different sizes, shapes, and colors. The system should also be able to detect and identify common fruit diseases.

**Introduction:**

The agriculture industry is a crucial sector of the global economy that provides food for the world's growing population. The process of manual harvesting, food processing, and retailing is an essential part of this industry. However, it comes with its set of challenges, especially with regard to manual harvesting. One of the most significant challenges is the high labor cost involved, which can significantly increase the overall cost of production. Additionally, there is also a cost associated with the equipment required for manual harvesting, which can be significant. Furthermore, there is a need for training and supervision to ensure that the manual harvesting is done correctly, which can also add to the overall cost. Despite these challenges, manual harvesting continues to be an important part of the agriculture industry. Fruit Picker/Sorter has involved around 50000 People migrated from 30 countries. And, global annual market size is near about $1 trillion, with a CAGR of 16.5% [BBC Nov, 2021]. To overcome these challenges and make manual harvesting more cost-effective, it is crucial to explore innovative solutions. By doing so, the agriculture industry can increase its productivity while reducing costs. With the rise of technology, there is an opportunity to automate manual harvesting, which can significantly reduce the labor and equipment cost involved. Additionally, this approach can also help reduce the need for training and supervision, thereby further lowering the overall cost. In conclusion, by exploring innovative solutions, the agriculture industry can overcome the challenges associated with manual harvesting and make the entire process more efficient

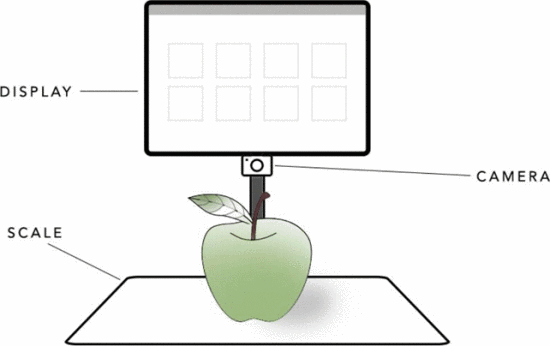


Fig 2. Camera setup prototype of fruit detection system using ML approach [1].

and cost-effective. By doing so, we can ensure that the world's growing population has access to affordable and high-quality food while also supporting the growth of the agriculture industry. Earlier studies in agriculture have centered on employing image recognition methods to supervise and handle the cultivation of crops [2] and fruits in the fields of farmers. Such research endeavors have also aimed to regulate the vegetation and refine the harvest process.

Two main techniques have been explored in the automated classification of fruits: traditional computer-vision based methods and deep learning-based methods. The former involves extracting low-level features and then using traditional machine learning techniques to classify images [3], while the latter effectively extracts features and performs end-to-end image classification. But they lack to detect variation of extreme features like – illumination, texture, color and shapes, etc. A study by Kheiralipour et al. [4] involved training a neural network (NN) to distinguish between desirable and undesirable shapes of cucumber, resulting in a classification accuracy of 97.10%. But, it only trained on only a few number of class and that makes the train results slightly doubtful. The problem of fruit classification has garnered attention from researchers who are developing deep learning-based methods, with most utilizing the transfer learning approach [5]. For instance, Bhole and colleagues et al. [6] investigated the feasibility of transfer learning in fruit classification by using a pre-trained model called SqueezeNet [7] to categorize mangoes into three grades: extra class, class I, and class II. But the accuracy is only 92.27% which is comparatively lower than recent advancement. Xiang and colleagues et al. [8] obtained an accuracy rate of 85.12% in classifying five types of fruits, namely apple, banana, carambola, guava, and kiwi, by utilizing the transfer learning approach on a lightweight MobileNetV2 [9] model trained on a dataset comprising 3,670 images. But, the data set size is too small to come to end to a conclusion.

To solve those problem, Mureşan et al. [10] released a fruit dataset called Fruit-360 dataset, with 28,736 training images and 9,673 testing images, specifically designed for fruit classification using deep learning models. Their approach involved using a convolutional neural network (CNN) with four convolutional layers, each followed by a max-pooling, fully connected, and softmax layer, to classify the fruits. They also employed data augmentation methods such as flip, hue/saturation changes, and grayscale, to enhance the classification accuracy, which ultimately led to a high accuracy of 95.23%. A fruit classification system that utilizes color, shape, and texture as image features. The dimensions of these features were initially reduced using principal component analysis (PCA) [11], and then they were input into classification algorithms, including the feed-forward neural network (FNN) and support vector machine (SVM). With a sample set of 1,653 fruit images in 18 distinct fruit categories, they performed experiments and recorded the highest accuracy of 88.2% using SVM. The accuracy is too low compared to most advanced M-RCNN and YOLO models. In a study on bark texture classification, local binary pattern based features were employed along with a multilayer neural network [12]. Another research introduced a tomato classifier system [13] using conventional image features, including color, shape, and size. The system was tested on 100 tomato images and successfully classified them into large, small, and medium classes, as well as four grades, with a mean grading accuracy of 90.7%. But, data distribution was so imbalanced that makes it more unpredictable and doubtful. Additionally, some investigations have employed image types other than RGB, including near-infrared (NIR) and multispectral images. One such research, described in [14], proposed a technique for classifying grapevine varieties based on in-field leaf spectroscopy.

Furthermore, a study presented in [15] suggested a technique for classifying strawberry ripeness based on multispectral imaging using 17 bands. The study involved the reduction of the dimensionality of multispectral images' features by employing PCA, which were then used in three different classifiers, namely Partial Least Squares (PLS), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The SVM classifier achieved a higher classification accuracy of 100%, with the visual wave-bands (VIS) part of spectra playing a significant role in the ripeness classification process.

Meanwhile, Femling and colleagues et al. [16] designed a system for fruit classification in retail stores using cameras to capture video footage. They utilized two convolutional neural networks, InceptionV3 and MobileNet, to detect and categorize fruits and vegetables in the video. Their method achieved the highest accuracy of 76% with InceptionV3 on a ten-class fruit dataset. Similarly, Chakraborty et al. [17] used the MobileNetV2 model with max-pooling and average-pooling to detect rotten fruits. Their approach achieved an accuracy of 94.97% with max-pooling, but they did not evaluate other standard convolutional neural networks with a higher number of classes. Herman and colleagues et al. have recently employed a DenseNet model [18] to classify ripeness levels of oil palm fruit. Their method achieved an accuracy of 86% on a dataset of 400 images containing 7 different ripeness levels.

In recent times, transformer-based deep learning techniques, commonly employed in natural language processing (NLP), have been examined for computer vision applications like image classification [19]. However, since transformers rely on pixel-wise attention rather than convolutional operations like CNN [20], their use in computer vision is not yet fully developed. Although some studies utilizing transformers for image classification have demonstrated that they outperform CNNs when there is enough training data, their availability of pre-trained models is limited compared to CNNs, which makes them less accessible for image classification. But, those solution are lacks with complexity, slow processing and no real time processing capabilities.

All prior art, we need to tackle the challenges of pixel-level segmentation, slow system performance, a restricted environment, and limited variability in target objects' scale and appearance, the recommended approach is to employ multimodal DNN features and models like MobileNetV2. These models have ample capacity to map data, handle diverse environments and appearances, and are lightweight, making them ideal for real-time applications. To address the current limitations, we suggest a new deep learning model that is lightweight and built on the MobileNetV2 architecture, known for its lightness compared to other pre-trained models like VGG-16. Some performance comparison tables for MobileNetv2 are in the table. 1.

Our model uses two established concepts, convolution, and attention, and we believe that this is the first work to combine these two concepts for fruit classification in a lightweight architecture. The convolution module captures image features, while the attention module identifies the important regions in the image. By combining these two modules, we expect our model to perform better in fruit classification. We conducted experiments on three publicly available datasets to test the effectiveness of our method, and the results indicate that our approach achieves stable performance.

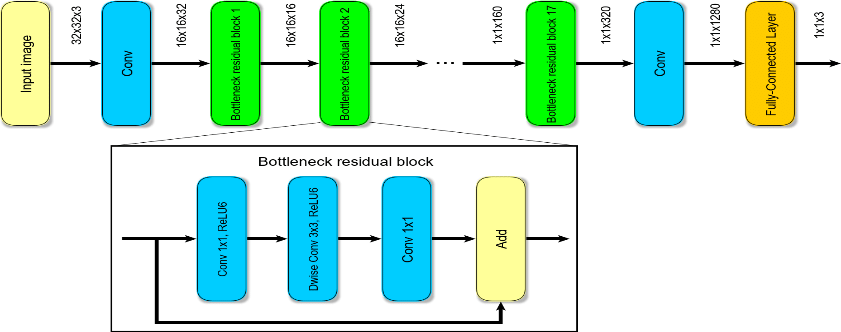


Fig 3. The schematic diagram of MobileNetV2 architecture (\*google).

**Methodology:**

**The paper introduces several contributions, which include:**

**(i) A new deep learning method based on MobileNetV2 model with the Transfer learning approach. The attention module identifies the salient regions in fruit images while the convolution module captures the activated regions obtained through the ReLU function over a fixed kernel size. Combining these two modules helps distinguish between different types of fruits as they complement each other.**

**(ii) The proposed method uses fewer trainable parameters by utilizing pre-trained weights for all layers of MobileNetv2 architecture, making it more suitable for deployment on devices with limited resources. For feature detection, we are using M-RCNN masking to extract the feature using RPN based region algorithm.**

**(iii) The model can be trained and deployed in an end-to-end fashion, avoiding the traditional machine learning approach of separate feature extraction and classification steps.**

**(iv) The model is validated using different fruit and vegetable datasets, demonstrating its robustness. The experimental results show that the proposed method performs well and outperforms other latest deep learning-based methods.**

Our suggested approach comprises of six elements: Preprocessing, the convolutional module, the attention module, the combination of convolution and attention modules, the fully-connected layers, and classification. We have illustrated the overall procedure of our technique depicted in Fig. 3.

1. Access and Load data

2. Preprocess

3. Derive Features

4. Model Training

5. Model Tuning

6. Result

Fig 4. The generic workflow of MR-CNN and MobileNetV2

The steps involved in transfer learning for fruit detection using MR-CNN and MobileNet V2:

First, create and preprocess a dataset of fruit images, assigning each image to its corresponding fruit category. Next, choose a pre-trained model, such as MobileNet V2, which has already been trained on a large set of images. Use the pre-trained model as a feature extractor by removing the final fully connected layer and applying it to the fruit images, producing a set of features for each image. Train a new classifier, such as MR-CNN, using the extracted features from the pre-trained model as input. This allows the classifier to output the fruit category for each input image. Fine-tune the whole model by adjusting the weights of all layers using the fruit dataset. This step aims to further enhance the model's performance. Evaluate the model's performance on a separate set of fruit images using metrics like accuracy, precision, and recall. Finally, deploy the trained model for fruit detection in real-world applications like automated fruit sorting in a processing plant.

As the basis for our network, we utilize MobileNetV2 which has been pre-trained on the 'ImageNet' dataset. Specifically, we extract the convolution feature map generated from the last residual block followed by the convolution layer of MobileNetV2, as depicted in Fig 2. We disregard the lower-level residual blocks which produce smaller-sized feature maps that lack the high-level clues for image recognition. We freeze all layers up to the convolution layers that yield a 3D feature map of size 7 × 7 (see Fig 2) during model design and training. This backbone network serves as a convolutional feature extractor for our research. We express this process mathematically as shown in Eq (1):

F1(I)=Conv(I), (1)

where F1(I) refers to the convolution feature produced from the input image, I.

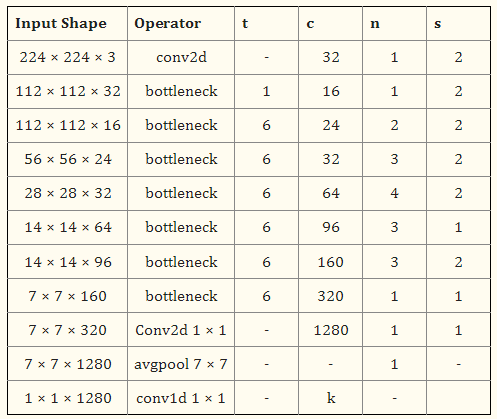


Fig 5. The layer-wise model summary for MobileNetV2 [21]

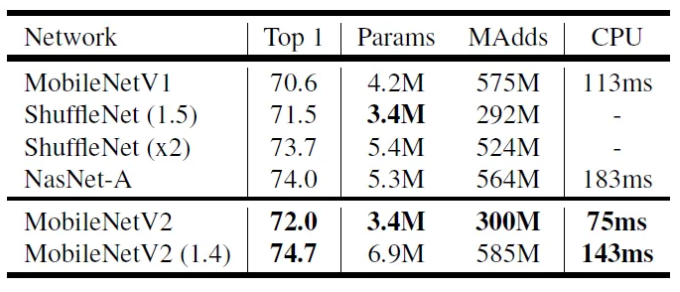


Fig 6. Hyperparameter and CPU processing time for MobileNetV2 ([towardsdatascience.com](https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c))

**Experimental Setup:**

* **Frontend part:**

1. **Load and access the dataset:**

* Kaggle (2020), Fruits and Vegetables Image Recognition Dataset

<https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition>

* 35 classes (10 fruits and 25 vegetables)
* Train set: 3500, Validation set: 350, Test set: 350 (described in Fig. 7)

1. Access and Load data

2. Preprocess

3. Derive Features

4. Model Training

5. Model Tuning

6. Testing Result

Saved model

Web App

GUI

Load New Data

Detect a Fruit/ Vegetable

Output

Fig 6. The experimental web app setup for fruit detection using MobileNetV2 and MR-CNN

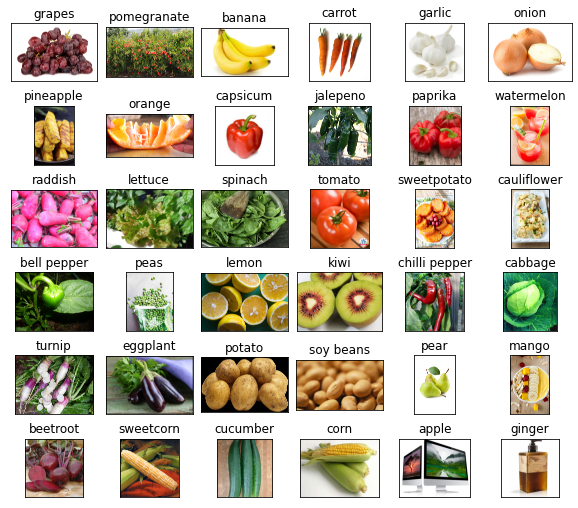
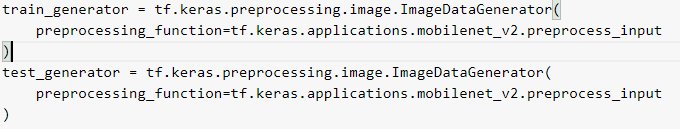


Fig 7. The data set sample for training the MR-CNN and Mobile Net V2.

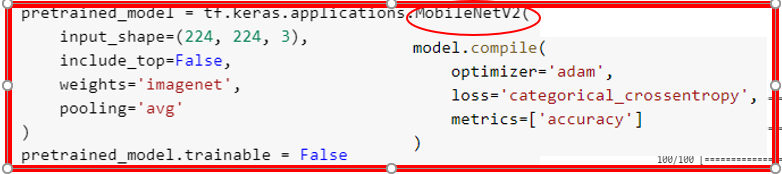
1. **Load the data:**

****

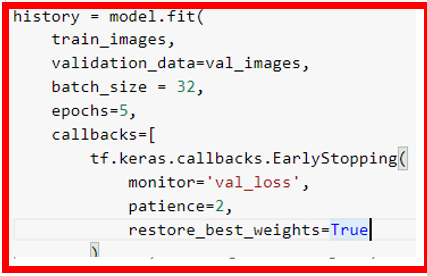
1. **Image Augmentation:**

****

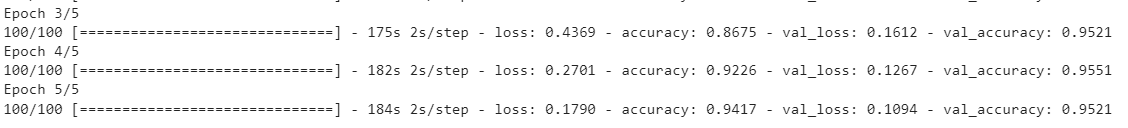
1. **Load the pretrained model:**



1. **Run the model:**



1. **Evaluate the results:**



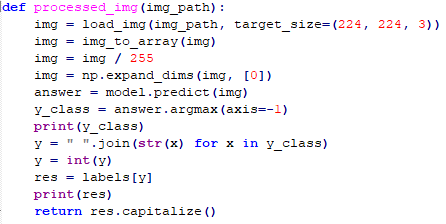
**Got 95% accuracy!**

* **Backend part: Web app**

1. **Load the API packages:**

****

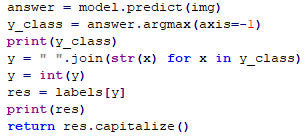
1. **Load image from frontend:**

****

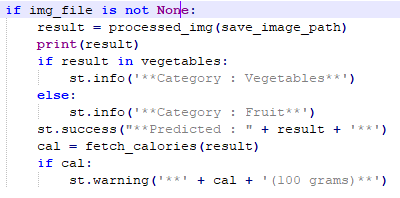
1. **Load the saved model:**

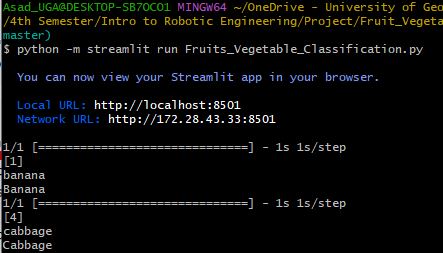
****

1. **Get prediction results:**

****

1. **Show results in the web app:**

****

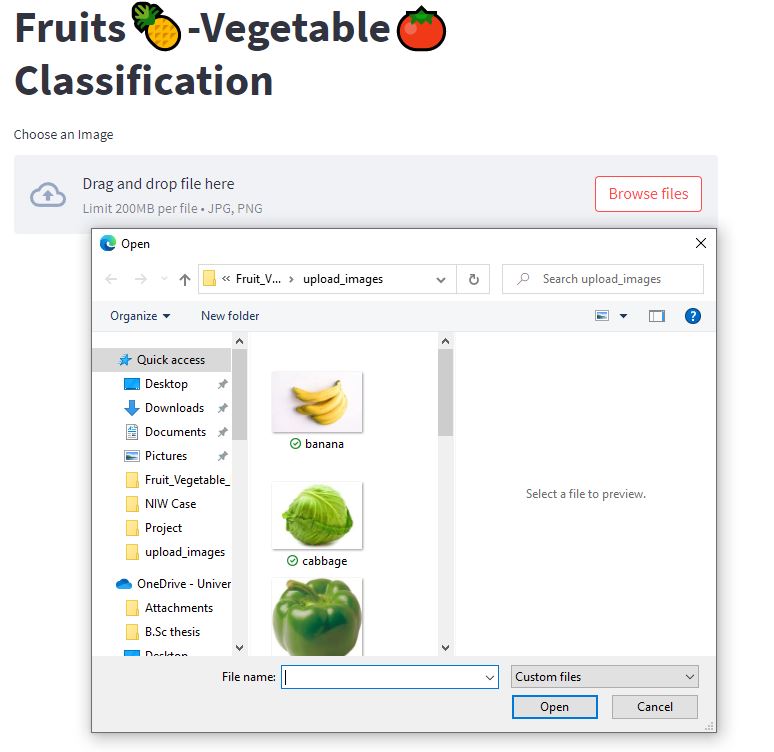
****

**Python command for run the app:**

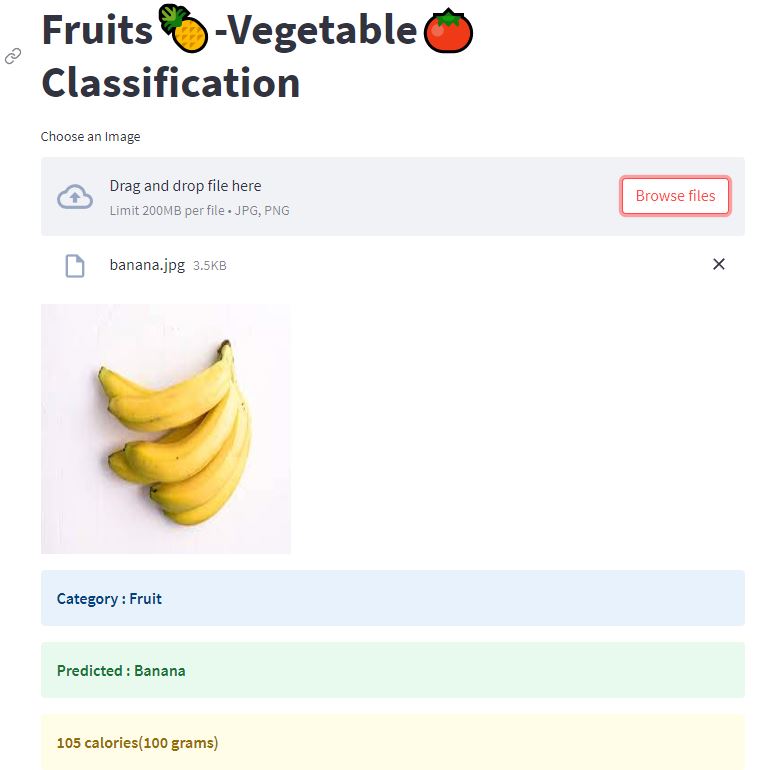
**“python -m streamlit run Fruits\_Vegetable\_Classification.py”**

* **Evaluation:**

**Load the image file in web app for detection:**

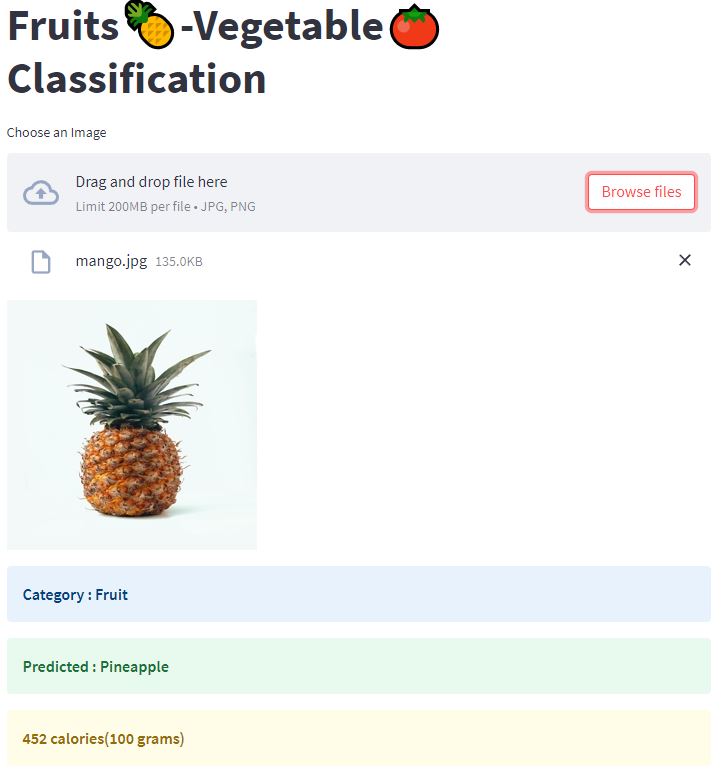
****

**Get detection results:**

****

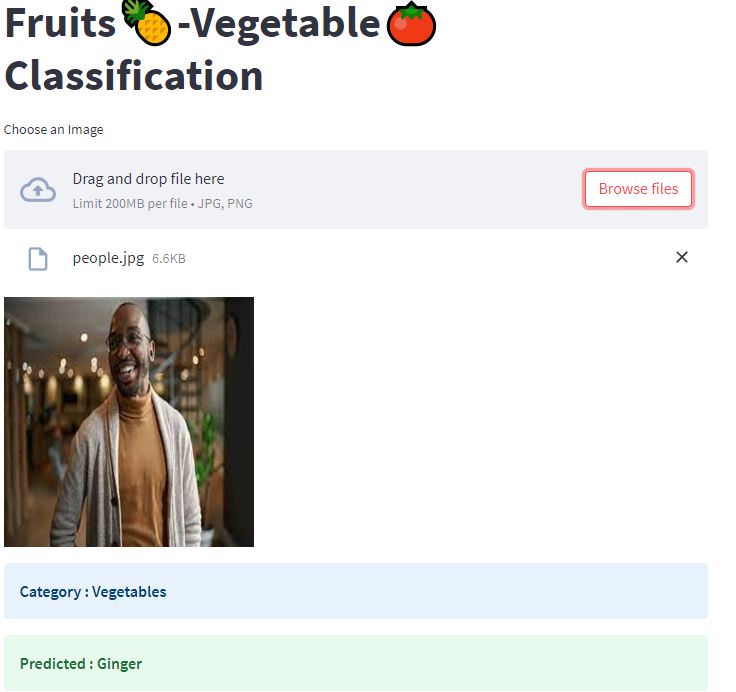
**Test the system critically**

1. **Mislabeled the data and got correct detection**

****

Mislabeled

1. **Check if the MobileNetV2 weights are fully done with transfer learning or, not.**



Misclassified

* **It proved that MNetV2 is successfully transferable! It did not detect that image as a human though the image net was trained with a human face as a class. But, here we successfully changed all the hyperparameter weights.**

**Conclusion:**

In this report, a new method for identifying fruits and vegetables is presented, which incorporates advanced techniques from OpenCV and Deep Learning. The method uses two models, MobileNetV2 and Mask-RCNN, for detecting and segmenting objects. The article proposes a straightforward Deep Learning model that employs a pre-trained MobileNetV2 model and an attention module. The model first extracts high-level object-based information using convolution features and then uses the attention module to identify important semantic information. These two modules are combined to fuse both types of information, which are then processed by fully connected layers and a softmax layer. The proposed method is evaluated on three benchmark datasets related to fruits, and the results show that it outperforms the four most recent Deep Learning methods in terms of classification accuracy while using fewer trainable parameters. This method utilizes transfer learning and was trained on a dataset containing diverse types of fruits and vegetables, showing impressive results with a mean squared error (MSE) loss of 0.1790 and an accuracy of 94.17%. The validation set also demonstrated promising outcomes, with a loss of 0.1094 and an accuracy of 95.21%. The proposed technique has the potential to enhance the detection of fruits and vegetables in various situations, such as in smart farming and automated quality control in the food industry, where accurate detection and classification of fruits and vegetables are crucial. The experimental findings suggest that the proposed technique is effective and could be beneficial for fruit and vegetable detection.

Reference:

[1] F. Femling, A. Olsson and F. Alonso-Fernandez, "Fruit and Vegetable Identification Using Machine Learning for Retail Applications," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 9-15, doi: 10.1109/SITIS.2018.00013.

[2] S. Bargoti and J. Underwood, "Deep fruit detection in orchards", IEEE International Conference on Robotics and Automation (ICRA), pp. 3626-3633, May 2017.

[3] Joseph JL, Kumar VA, Mathew SP. Fruit Classification Using Deep Learning. In: Innovations in Electrical and Electronic Engineering; 2021. p. 807–817.

[4] Kheiralipour K, Pormah A. Introducing new shape features for classification of cucumber fruit based on image processing technique and artificial neural networks. Journal of food process engineering. 2017;40(6):e12558. doi: 10.1111/jfpe.12558

[5] Behera SK, Rath AK, Sethy PK. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. Information Processing in Agriculture. 2020.

[6] Bhole V, Kumar A. Mango Quality Grading using Deep Learning Technique: Perspectives from Agriculture and Food Industry. In: Proc. 21st Annual Conference on Information Technology Education; 2020. p. 180–186.

[7] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:160207360. 2016.

[8] Xiang Q, Wang X, Li R, Zhang G, Lai J, Hu Q. Fruit image classification based on Mobilenetv2 with transfer learning technique. In: Proc. 3rd International Conference on Computer Science and Application Engineering; 2019. p. 1–7.

[9] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proc. IEEE conference on computer vision and pattern recognition; 2018. p. 4510–4520.

[10] Mureşan H, Oltean M. Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica. 2018;10(1):26–42. doi: 10.2478/ausi-2018-0002

[11] Abdi H, Williams LJ. Principal component analysis. Wiley interdisciplinary reviews: computational statistics. 2010;2(4):433–459. doi: 10.1002/wics.101

[12] Fekri-Ershad S. Bark texture classification using improved local ternary patterns and multilayer neural network. Expert Systems with Applications. 2020;158:113509. doi: 10.1016/j.eswa.2020.113509.

[13] Liu L, Li Z, Lan Y, Shi Y, Cui Y. Design of a tomato classifier based on machine vision. PloS one. 2019;14(7):e0219803. doi: 10.1371/journal.pone.0219803

[14] Gutiérrez S, Tardaguila J, Fernández-Novales J, Diago MP. Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. PloS one. 2015;10(11):e0143197. doi: 10.1371/journal.pone.0143197

[15] Liu C, Liu W, Lu X, Ma F, Chen W, Yang J, et al. Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. PloS one. 2014;9(2):e87818. doi: 10.1371/journal.pone.0087818

[16] .Femling F, Olsson A, Alonso-Fernandez F. Fruit and vegetable identification using machine learning for retail applications. In: 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE; 2018. p. 9–15.

[17] Chakraborty S, Shamrat FJM, Billah MM, Al Jubair M, Alauddin M, Ranjan R. Implementation of Deep Learning Methods to Identify Rotten Fruits. In: 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE; 2021. p. 1207–1212.

[18] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–4708.

[19] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020;

[20] Raghu M, Unterthiner T, Kornblith S, Zhang C, Dosovitskiy A. Do Vision Transformers See Like Convolutional Neural Networks? Advances in Neural Information Processing Systems. 2021;34.

[21] Shahi TB, Sitaula C, Neupane A, Guo W. Fruit classification using attention-based MobileNetV2 for industrial applications. PLoS One. 2022 Feb 25;17(2):e0264586. doi: 10.1371/journal.pone.0264586. PMID: 35213643; PMCID: PMC8880666.