Automatic Fruits Classification System Based on Deep Neural Network

Khadija Munir, Arif Iqbal Umar, and Waqas Yousaf
Department of Information Technology, Hazara University, Mansehra, KPK, Pakistan
Email: waqas21140@gmail.com

Abstract

Fruit classification is playing a vital role in robot-based farming. The plucking of fruits and packing is done using robots nowadays. This could only be possible using efficiently trained robots base on machine learning. Different techniques have been developed for fruit classification, but still, there are many gaps, i.e., efficiency and accuracy. In this research work, we are targeting classification accuracy. This paper presented an Automatic Fruit Detection tool with good precision and recalled using deep learning neural networks. It will help in farming, cultivation, and produce sound effects in robotic farming. The aim is to build an accurate, fast and reliable fruit detection system, a vital element of an autonomous agricultural robotic platform; it is a crucial element for fruit yield estimation and automated harvesting. We used the ResNet-50 in the context of transfer learning. Different training choices were defined, i.e., 10% to 80%. Experimental results show that we compete for the prior approaches even on only 10% training. The proposed approach achieves state-of-the-art results compared to prior work with the F1 Score, which considers both precision and recall performances improving from 0.838 to 0.894 and 0.995 of accuracy. In addition to improved accuracy, this approach is also much quicker as compared to recent approaches.

Keywords: Deep convolutional neural network, harvesting robots, agricultural robotics, Image Recognition, Fruit Recognition

Introduction

Introducing [1], the supply of skilled labor in the agricultural industry is the most significant costly issue in the industry. Increment in the cost of provisions, for example, vitality, water system, agrochemicals and so forth, is because of that reason. These are leading agricultural companies and horticultural industries that experience stress with little profit. Following these difficulties, sustenance production is essential to encounter the developing needs of the world's developing population, representing a genuine test.

Figure 1: a). Instances of red and golden apple fruit detected by the robot. b). Instances of yellow and green orange fruit detected by the robot

Mechanical (Robotic) reaping can potentially answer this issue by diminishing work and incrementing the organic product quality. Creating enthusiasm for horticultural robots' utilization to reap products of the soil throughout the previous three years because of these say actualities [2]. The improvement of this stage has numerous issues, for example, work and decision. Be that as it may, the advancement of a precise natural product discovery framework is an imperative advance for robotized automated reaping machines. It is a front acknowledgment framework before the control and obtaining framework. If the fruits are not found or cannot see, they cannot select fruit properly. This transition is difficult due to several factors, but in some cases, the change in illumination, occlusion and fetus may have a visual appearance similar to the background.

Deep Learning

Deep learning can be well defined as a model that attempts to extract higher-level data based on a specific set of algorithms. These models include each specific processing and can be used to extract different linearity [3]. An essential idea of inside and out preparation can be comprehended because it is a comprehensive technique for machine learning in informal learning. The portrayal of perceptions (Images) can be performed in different courses, such as pixel force esteems into vectors called trademark vectors. These perceptions can be speaking to as an arrangement of edges. In terms of simplifying specific field-based learning tasks, such as specific modules, there are safer methods than other methods (for example, facial recognition and facial recognition). In this area, to study expressions of vast amounts of data without noise without labels, like images, studies were carried out to create excellent representations and design models with levels [4]. It is understood that in-depth training is a new trend. In general, data analysis will be presented in the top ten innovative technologies of 2013 [4]. In-depth training can be said that the artificial neural network has been partially completed with high abstraction and better prediction of raw data [5]. Regarding machine learning, top to bottom preparation turns into an excellent tool for artificial vision and picture situations. It is important to note that DCNN offers several percentage points per year in the classification activities of the ImageNet objects. Donohue et al. using DCNN, exceed the most advanced improvements every year. Donohue and his colleagues used the DCNN to classify scene images and classify small grain birds to obtain excellent accuracy values [5]. Razavian et al. using DCNN, The effective and accurate detection of human attributes and the visual recovery of cases have led to the most advanced results

https://doi.org/10.24949/njes.v13i1.501

[6]. Therefore, the use of a deep slope indicates that it is a precise and effective solution, not for other studies but also for the fruits' classification.

Related Work

Numerous scientists have thought about organic product discovery, for example, the record exhibited in [7-12], yet as can be seen from the review [13], to make a quick and dependable framework for identifying natural products. The issue continues. This is because of severe changes in the organic products' presence on the field, including the qualities of shading, shape, size, consistency, and reflectivity. Moreover, in many of these media, the organic product is somewhat dynamic, the lighting conditions and the shadows continuously change. A few reports exhibited in writing think about organic product identification as an issue of picture division (for instance, a natural product on a foundation). Another approach by [10] examined the issue of finding an apple with a yield figure. They built up a framework to distinguish apples in light of shading and a perfect representation of one-of-a-kind reflection. Additional information, for example, the standard size of an apple, has been used to dispense with false positives or portions of areas that may contain distinctive apples. Other heuristics should have based only on rounded areas such as surveying. Bac et al. [11] used a segmentation technique to identify sweet peppers. They utilized a 6-band multispectral camera and utilized different capacities, including crude multispectral information, a standardized contrast file, and entropy-based work attributes. Analyses directed in a firmly controlled greenhouse environment have demonstrated that this approach has prompted accurate output by using segmentation. In any case, the author demonstrated that it isn't sufficiently exact to make maps of solid deterrents.

Hung et al. [12] proposed the utilization of contingent irregular fields for the segmentation of almonds. They proposed a five-year way to deal with segmentation, which investigated capacities utilizing Sparse Auto Encoder (SAE). These highlights have been utilized in the system of the CRF and have been superseded by past work. They achieved noteworthy segmentation execution. However, they did not attain proper object identification. Moreover, they noticed blocking is a significant issue. Automatically, this method can just concern a low level of block.

Yamamoto et al. [9] achieved tomato recognition by using color-based segmentation. So, we have prepared classification classifiers and regression trees (CARTs) using color and shape properties. This created a segmentation chart and the associated features assembled in groups. Every group confirmed as detection and reduced the quantity of error. They trained classifiers that do not use fruit, using random forests in a greenhouse-controlled environment.

In most of the above literature, segmentation techniques depend on pixels for fruit identification and recognition to evaluate the yields [7, 10]. The limited research that has carried out an accurate survey of the fruits has done this on the fruits of a controlled greenhouse environment. Therefore, the problem of finding fruits in challenging conditions remains unsolved.

This shows good performance when testing with selected image data, yet because of the high inconstancy of the target aspect in farming structure, the exemplary sliding strategy is moving toward [14]. When it is distributed in a natural farm configuration, it is difficult to change and deal with the desired (target) object's size and appearance.

In recent years, deep neural networks have advanced the classification and detection of objects [15-17]. The advanced PASCAL-VOC detection system [18] is based on two phases. In the first phase channel, domain offer methods are used, such as selective searches [19] to extract the areas of interest from the images and deliver them to the deep neural network to classify. Although recovery performance is high, this channel has a high cost and cannot be used in real-time robot applications. RPNs (Regional proposal networks) [20-22] solved that issue by integration of deep networks classification for packages and offers, simultaneously providing for object boundaries, classifying them in each position and because parameters of a network are common, performance is significantly improved and suitable for use in robots. In a real outdoor environment, information is often provided in the single sensor mode to find interesting fruits with a wide spectrum of light changes, partial obstruction, and different appearances. This is an example of using a multimodal fruit detection system, as various sorts of sensors can deliver additional evidence on various features of fruit. Previously, DNN was very promising in multimodal systems in other sectors and the automation of agriculture where audio/video is used very well [23, 24]. Control system Development for fruit classification based on a Convolutional neural network has been suggested by Zaw Min Khaing in [14]. The model was evaluated under limited dataset FIDS30 as the fruit image dataset consists of 971 images for 30 different kinds of fruits. The proposed model had accuracy in the classification close to 94%. The model was developed using a limited dataset. So, it cannot meet the actual application situation. Fruit and vegetable classification system using image saliency and convolution neural network has been proposed by [27]. In this method, the model was evaluated under a limited dataset consisting of 12173 images for 26 categories and an accuracy of 95.6%. Out of 12173 images, 80% were used for training and 20% for validation. The number of training samples is small, which is not conducive to the extensibility of an algorithm.

Challenges

The literature analysis shows that the fruits' classification is one of the combustion problems and that the latest technologies are emerging. However, there are problems in analyzing images and fruits when an image of poor quality of fruits or direction and rotation effects are different. On the other hand, to accurately predict the fetus's type, problems arise in analyzing this image. Therefore, to accurately classify the fruits' images in each category, it is necessary to apply an efficient and effective learning model based on the deep.

Proposed Model

Apply emerging deep learning techniques for fruit classification tasks. The proposed work is beliefs on providing a more accurate fruit classification method. This research will use the latest dataset publicly available for comparison and evaluation of this work.

Res-NET-50

ResNet-50 [17] is a convolutional neural network trained on more than a million images from the ImageNet database. The network is 50 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224. This model was released in 2015 by Microsoft Research Asia. The ResNet architecture (with its three realizations ResNet-50, ResNet-101 and ResNet-152), obtained very successful results in the ImageNet and MS-COCO competition. The core idea exploited in these models, residual connections, greatly improves gradient flow, thus allowing training of much deeper models with tens or even hundreds of layers. ImageNet classes are mapped to Wolfram Language Entities through their unique WordNet IDs. This model is trained using the ImageNet Large Scale Visual Recognition Challenge 2012 classification dataset, consisting of 1.2 million training images with 1,000 classes of objects. This model achieves 77% top-1 and 93.3% top-5 accuracy in 1-crop validation, and 78.6% top-1 and 94.3% top-5 accuracy in 10-crop validation on the ILSVRC 2012 dataset. The degradation (of training accuracy) indicates that not all systems are similarly easy to optimize. So, we considered the Res-Net50 shallower architecture and its deeper counterpart that adds more layers to it. It provides a solution by construction to the deeper model: the added layers are identity mapping, and the other layers are copied from the learned shallower model. The existence of this constructed solution indicates that a deeper model should produce no higher training error than its shallower counterpart. However, experiments show that our current solvers on hand cannot find comparably good or better solutions than the constructed solution (or unable to do so in a feasible time). Instead of hoping every few stacked layers directly fit a desired underlying mapping, Res-Net50 explicitly lets these layers fit a residual mapping. Formally, denoting the desired underlying mapping as H(x), Res-Net50 let the stacked nonlinear layers fit another mapping of F(x): = H(x) - x. The original mapping is recast into F(x)+x.

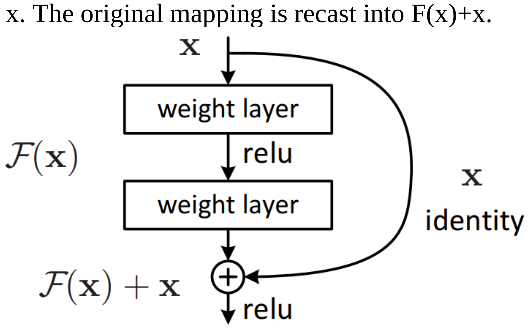

Figure 2. Residual learning: a building block

It is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. We used Res-Net50 because it provides the best solution in the context of accuracy and less training time. There are two kinds of residual connections: The identity shortcuts (x) can be directly used when the input and output are of the same dimensions.

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

When the dimensions change, A) The shortcut still performs identity mapping, with extra zero entries padded with the increased dimension. B) The projection shortcut is used to match the dimension (done by 1*1 conv) using the following formula.

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s\mathbf{x}.$$

The first case adds no extra parameters; the second one adds in W_{s}.

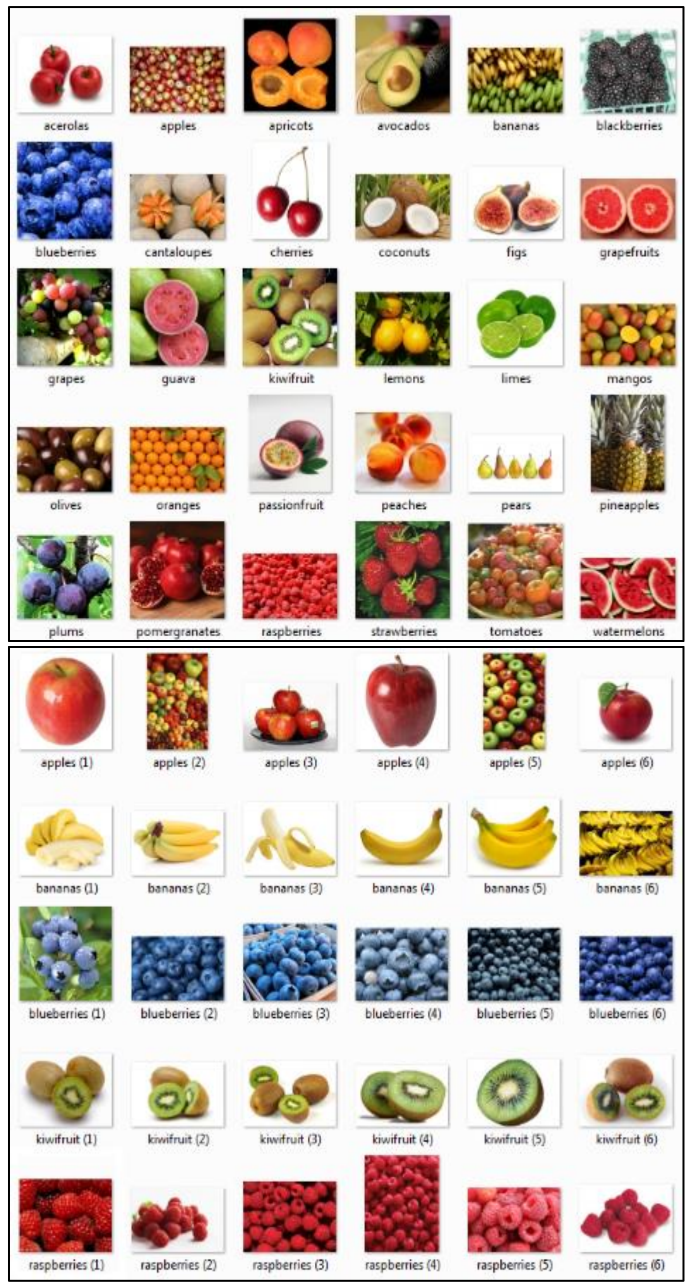

Figure 3: a). Examples of classes in FIDS-30 dataset. b). Examples of within-class varieties in FIDS-30 dataset.

PUBLICLY AVAILABLE IMAGE DATABASES

1) FRUIT-360

It consists of the best feature images comprising of different fruits and different types. Like "Apples, Apricot, Avocado, Banana, Cactus fruit, Carambula, Cherry, Cherry Wax (Yellow, Red, Black), Clementine, Cocos, Dates, Granadilla, Grape (Pink, White, White2), Grapefruit (Pink, White), Guava, Huckleberry, Kiwi, Kaki, Kumquats, Lemon, Lime, Lychee, Mandarine, Mango, Maracuja, Melon Piel de Sapo, Mulberry, Nectarine, Orange, Papaya, Passion fruit, Peach, Pepino, Pear, Pineapple (normal, Mini), Pitahaya Red, Plum, Pomegranate, Quince, Rambutan, Raspberry, Salak, Strawberry, Amarillo, Tangelo, Tomato, Walnut". Fruit 360 dataset consists of 55244 images. These pictures are divided into two categories training and test data. 41322 images are used for the training model, and 13877 images are used for testing purposes. The dataset contains 81 classes of fruit. The size of the image is 100 x 100 pixels [25]. This dataset can be downloaded from

https://www.kaggle.com/moltean/fruits

2) FIDS-30

The fruit image data set consists of 971 images of common fruit. The images are classified into 30 different fruit classes. Every fruit class contains about 32 different images. The fruit images are in the JPEG image format, spreading from a few KB to a few MB in size. The images are very diverse. The images contain many fruits, some contain just a single fruit and others contain dozens. Some of the images contain much noise, such as trees, leaves, plates and other backgrounds. This dataset is publicly available on http://www.vicos.si/Downloads/FIDS30. For performing fruit classification, we are used two datasets, say Fruits-360 and FIDS-360. We have trained these two datasets in varieties of percentages. We first take the same number of images from each class. We used the automatic splitting approach for randomly splitting datasets in trainset and test sets. Both datasets are trained on 10 to 80%. For instance, we first trained 10% of the dataset and evaluated it on 90% to compute evaluation metrics. The same fashion was applied next, i.e., 20% and 80% testing. Lastly, we train 80% of the dataset and 20% is used for testing. For both datasets, we employ such a strategy as shown in table 1.

Results

After implementation, we obtained outputs (results) and examined them properly and then compared them with the existing methods to check the proposed method's accuracy and efficiency. Table 2 shows the accuracy, precision, recall, and F1 Score achieved by the proposed system on different training choices using the FIDS-30 dataset. The proposed model is competing with the state-of-the-art method even on 60% training. Hence the ResNet-50 is a more suitable model as compare to generic CNN models. Table 3 shows accuracy, precision, recall and F1 Score achieved by the proposed system on different training choices using Fruit-360. The proposed model is competing with the state-of-the-art method even on only 10% training. Hence the ResNet-50 is the more suitable model as compare to generic CNN models.

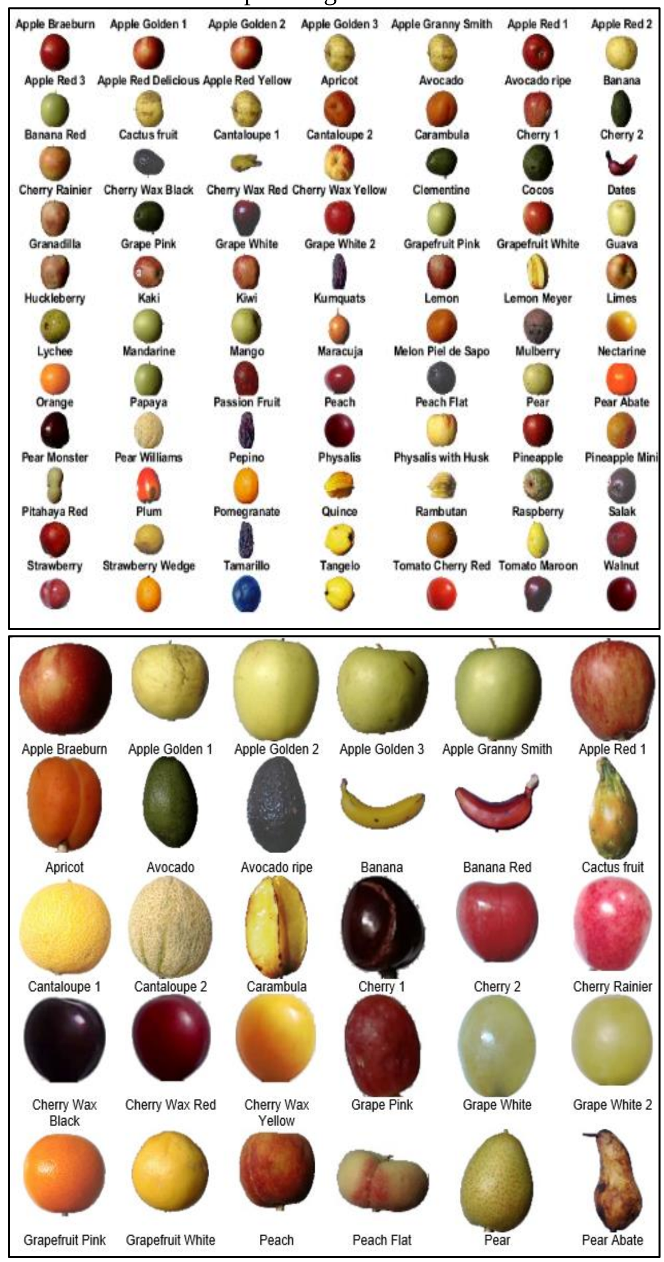

Figure 4: a). Examples of classes in Fruit-360 data set. b). Examples of within-class variety in Fruit-360 dataset

Table 1. Images used for training and testing for FIDS-30 and Fruit-360

Dataset	Classes	Train Set	Test Set	Total
FIDS-30	30	60 (10%)	600 (90%)	660
		120 (20%)	540 (80%)	
		210 (30%)	450 (70%)	
		270 (40%)	390 (60%)	
		330 (50%)	330 (50%)	
		390 (60%)	270 (40%)	
		450 (70%)	210 (30%)	
		540 (80%)	120 (20%)	
Fruit-360	77	3773 (10%)	34265 (90%)	38038
		7623 (20%)	30415 (80%)	
		11396 (30%)	26642 (70%)	
		15246 (40%)	22792 (60%)	
		19019 (50%)	19019 (50%)	
		22792 (60%)	15246 (40%)	
		26642 (70%)	11396 (30%)	
		30415 (80%)	7623 (20%)	

Table 2. Resultant precision, recall, F1 Score and accuracy values for FIDS-30

Technique	Dataset	Train Set	Test Set	Accuracy	Precision	Recall	F1 Score
Sa, I., et al 2016 [26]	FIDS-30	100 (82%)	22 (18%)	0.9	0.57	0.8	0.838
PROPOSED RESNET-50	FIDS-30	60 (10%)	600 (90%)	0.5633	0.6643	0.5633	0.5716
		120 (20%)	540 (80%)	0.7222	0.7522	0.7222	0.7184
		210 (30%)	450 (70%)	0.7666	0.8181	0.7666	0.7731
		270 (40%)	390 (60%)	0.8076	0.8394	0.8076	0.8106
		330 (50%)	330 (50%)	0.8242	0.8384	0.8242	0.8231
		390 (60%)	270 (40%)	0.8777	0.8896	0.8777	0.8750
		450 (70%)	210 (30%)	0.8476	0.8713	0.8476	0.8459
		540 (80%)	120 (20%)	0.8916	0.9213	0.8916	0.8946

Table 3. Resultant precision, recall, F1 Score and accuracy values for Fruit-360

Technique	Dataset	Train Set	Test Set	Accuracy	Precision	Recall	F1 Score
Mureşan, H. 2018 [25]	Fruit-360	75.0000%	25.0000%	96.3%	-	-	-
PROPOSED RESNET-50	Fruit-360	3773 (10%)	34265 (90%)	0.9920	0.9920	0.9930	0.9919
		7623 (20%)	30415 (80%)	0.9950	0.9950	0.9952	0.9950
		11396 (30%)	26642 (70%)	0.9951	0.9951	0.9957	0.9950
		15246 (40%)	22792 (60%)	0.9958	0.9958	0.9964	0.9958
		19019 (50%)	19019 (50%)	0.9959	0.9959	0.9967	0.9958
		22792 (60%)	15246 (40%)	0.9933	0.9933	0.9936	0.9932
		26642 (70%)	11396 (30%)	0.9965	0.9965	0.9972	0.9965
		30415 (80%)	7623 (20%)	0.9952	0.9952	0.9952	0.9952

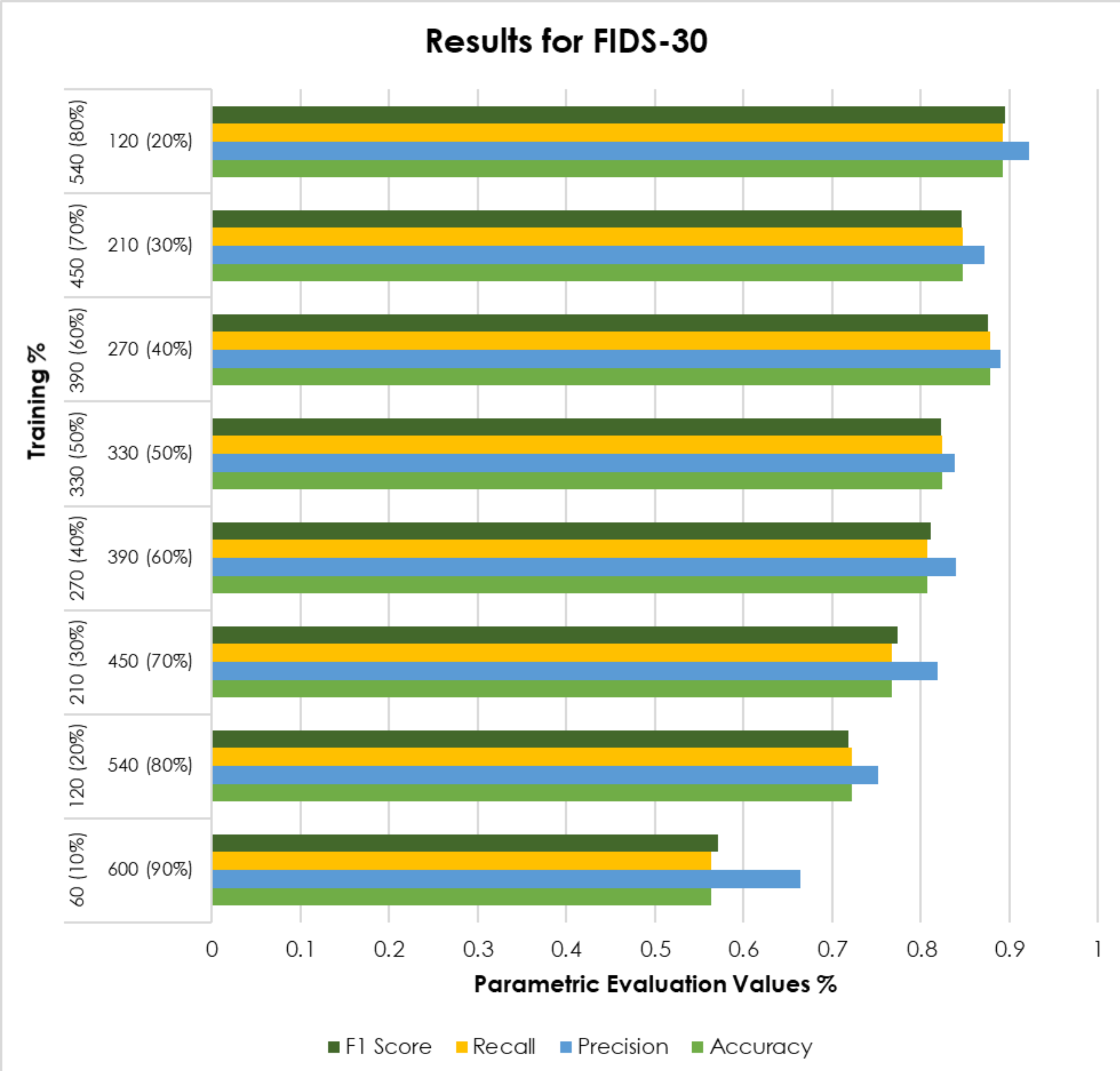

Figure 5: Accuracy, precision, recall and F1 Score on FIDS-30

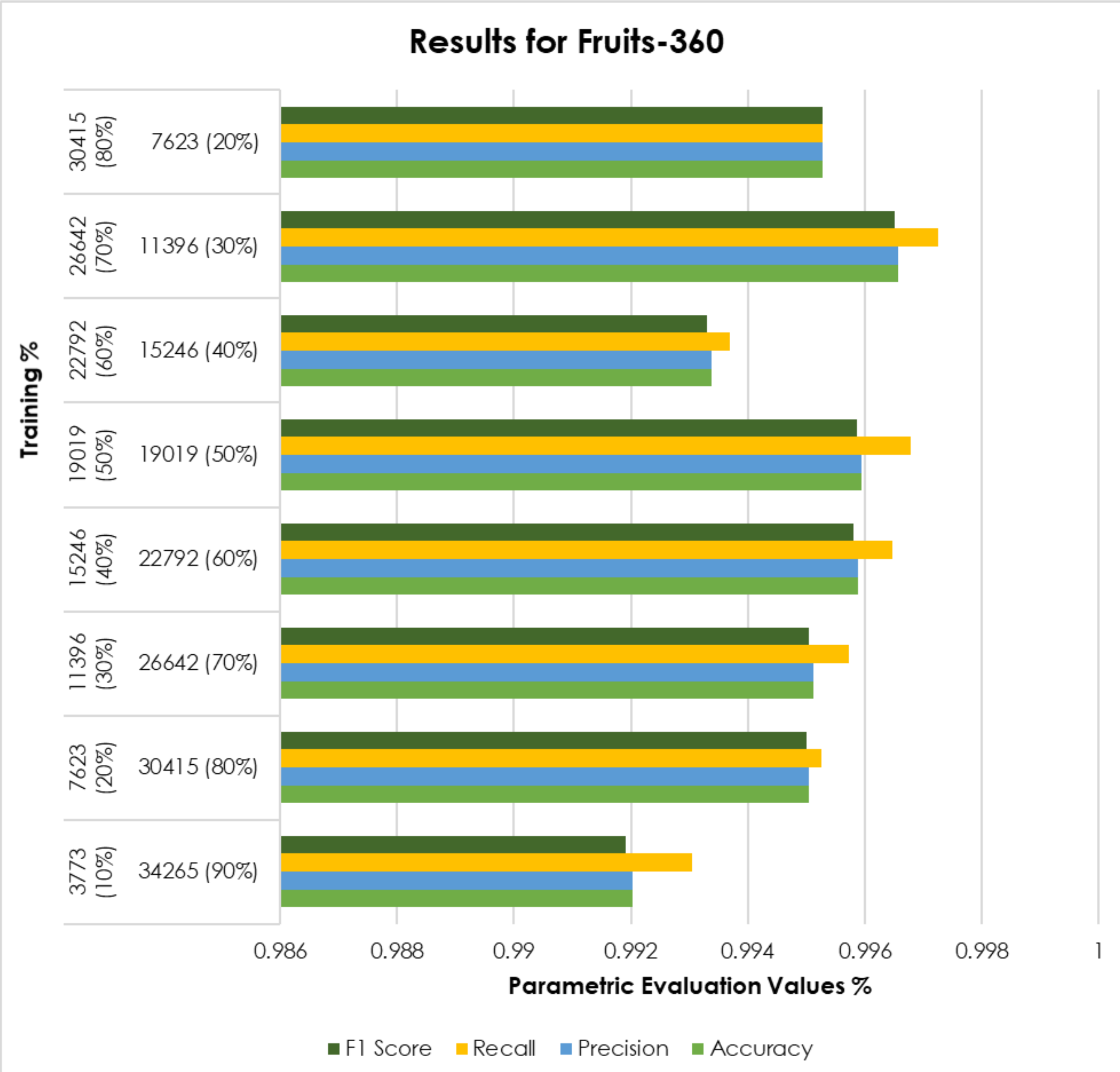

Figure 6: Accuracy, precision, recall and F1 Score on Fruits-360

Conclusion

Robotics farming is growing day today because of the rapidly increasing needs of the human being. Modern farmer wants to grow more crops to increase turnover. This can be possible by using modern farming techniques like robotics base harvesting. Fruit plucking from fruit plants is time-consuming and takes much workforce. This can be done in less time using robots trained for specific object detection and classification. This research work presented an automatic fruit classification system based on a neural network to understand better and detect fruit. For this, we use the ResNet-50 for the classification of fruits. There are two publicly available datasets, i.e., FIDS-30 and Fruits-360. We resized all the images into the exact dimensions of 224×224 to avoid overfitting during the training phase. The main advantage of using ResNet-50 is residual connections are found to improve gradient flow significantly. The ResNet consists of 50 layers that classify the given images of different fruits into specified categories. The proposed system achieved high accuracy, precision, recall and F1 Score even on significantly less training. Thus, competing for the state-of-the-art approaches by scoring 99.5% accuracy.

REFERENCES

1. Valle, H., T. Caboche, and M. Lubulwa, *Australian vegetable growing farms: An economic survey, 2011-12 and 2012-13*. 2014: ABARES Canberra, ACT.
2. Kondo, N., M. Monta, and N. Noguchi, *Agricultural Robots: Mechanisms and Practice*. 2011: Apollo Books.

3. Deng, L. and D. Yu, *Deep learning: methods and applications.* Foundations and Trends® in Signal Processing, 2014. **7**(3–4): p. 197-387.
4. Schmidhuber, J., *Deep learning in neural networks: An overview.* Neural networks, 2015. **61**: p. 85-117.
5. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning. nature 521 (7553): 436.* Google Scholar, 2015.
6. Sharif Razavian, A., et al. *CNN features off-the-shelf: an astounding baseline for recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014.
7. Nuske, S., et al. *Yield estimation in vineyards by visual grape detection*. in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. 2011. IEEE.
8. Nuske, S., et al., *Automated visual yield estimation in vineyards.* Journal of Field Robotics, 2014. **31**(5): p. 837-860.
9. Yamamoto, K., et al., *On plant detection of intact tomato fruits using image analysis and machine learning methods.* Sensors, 2014. **14**(7): p. 12191-12206.
10. Wang, Q., et al. *Automated crop yield estimation for apple orchards*. in *Experimental robotics*. 2013. Springer.
11. Bac, C., J. Hemming, and E. Van Henten, *Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper.* Computers and electronics in agriculture, 2013. **96**: p. 148-162.
12. Hung, C., et al. *Orchard fruit segmentation using multispectral feature learning*. in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. 2013. IEEE.
13. Kapach, K., et al., *Computer vision for fruit harvesting robots–state of the art and challenges ahead.* International Journal of Computational Vision and Robotics, 2012. **3**(1/2): p. 4-34.
14. Song, Y., et al., *Automatic fruit recognition and counting from multiple images.* Biosystems Engineering, 2014. **118**: p. 203-215.
15. Simonyan, K. and A. Zisserman. *Two-stream convolutional networks for action recognition in videos*. in *Advances in neural information processing systems*. 2014.
16. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
17. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
18. Everingham, M., et al., *The pascal visual object classes challenge: A retrospective.* International journal of computer vision, 2015. **111**(1): p. 98-136.
19. Uijlings, J.R., et al., *Selective search for object recognition.* International journal of computer vision, 2013. **104**(2): p. 154-171.
20. Zitnick, C.L. and P. Dollár. *Edge boxes: Locating object proposals from edges*. in *European conference on computer vision*. 2014. Springer.
21. Ren, S., et al. *Faster r-cnn: Towards real-time object detection with region proposal networks*. in *Advances in neural information processing systems*. 2015.
22. He, K., et al. *Spatial pyramid pooling in deep convolutional networks for visual recognition*. in *European conference on computer vision*. 2014. Springer.
23. Ngiam, J., et al. *Multimodal deep learning*. in *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011.
24. Eitel, A., et al. *Multimodal deep learning for robust rgb-d object recognition*. in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. 2015. IEEE.
25. Mureşan, H. and M. Oltean, *Fruit recognition from images using deep learning.* Acta Universitatis Sapientiae, Informatica, 2018. **10**(1): p. 26-42.
26. Sa, I., et al., *Deepfruits: A fruit detection system using deep neural networks.* Sensors, 2016. **16**(8): p. 1222.
27. Kim, J.-y., M. Vogl, and S.-D. Kim. *A code based fruit recognition method via image convertion using multiple features*. in *IT Convergence and Security (ICITCS), 2014 International Conference on*. 2014. IEEE.