Pure-CNN: A Framework for Fruit Images Classification

Asia Kausar[1], Mohsin Sharif[2], JinHyuck Park[3] and Dong Ryeol Shin[4]
Department of Electrical and Computer Engineering
Sungkyunkwan University,
Suwon, South Korea
E-mail: {asiakausar[1], mohsin[2], jjangdol[3] drshin[4]}@skku.edu

***Abstract*—Automation in fields like robot harvesting, farming, health and education require object classification using machine learning and computer vision techniques. Among these fruit classification is a challenging task because of its several varieties and similarity in color, shape, size and texture features. In order to recognize multiple fruits more accurately, we proposed a Pure Convolutional Neural Network (PCNN) with minimum number of parameters. The PCNN consists of 7 convolutional layers, some of them followed with stride. Additionally, to reduce overfitting and taking average of whole feature maps we employed recently developed Global Average Pooling (GAP) layer that verified to be very effective. We demonstrate our classification performance using PCNN on recently introduced fruit-360 dataset. The experimental results of the 55244 color fruit images from the 81 categories, show that the PCNN achieve a classification accuracy of 98.88%.**

***Index Terms*—Convolutional Neural Network, Fruit classification, Object recognition, Multi-class image classification, Deep Learning**

I. Introduction

Fruit classification has gained focus over the last few years, it is important and difficult task in the agriculture industries such as food production, marketing, packaging and education as well. Until the several years, agriculture was labor intensive as finding trained farm labor in the agriculture production is one of the cost demanding factor. Additionally, collecting and sorting of specialty crops such as apple, citrus, cherry, orange and mango are time consuming and tiresome job due to number of varieties of same fruit, e.g., more than 7,000 varieties of apples are grown all over the world as reported in [1]. In such consequence, automation can reduce the labor cost and increase production rapidly.

In early studies, scholars have proposed various computer vision and machine learning based classification approach to overcome the cost of manually recognition of fruits. In their study, they have used color, shape, size and texture features with classification algorithms [2], [3], [4]. As discussed in their work, most of them have used pre-processing or adopted the method of feature extraction combined with classifier. But, most of developed classifiers are not robust for all types of fruits and can make prediction with considerable misclassification.

Convolutional Neural Network (CNN) has become a hot research topic in the field of object detection and image classification. The usage of CNN are driven by the fact that they can extract features from an input image. As compared to traditional classification algorithm, in CNN the image can be directly fed into the network, it avoids pre-processing and feature extraction process. Classical CNN built using three main layers; Convolutional layers, Pooling layers and Fully Connected (FC) layers. CNN got lot of attention after winning competition of ImageNet [5]. Later, different CNN models established by numerous scholars, by varying depth and width of layers and also replacing the layers, discussed in [6], [7], [8], [9]. Such as, to simplify the CNN architecture, for dimension reduction, pooling layers replaced by Conventional layer with increased number of stride [7]. Layers of CNN behaves as object detectors,but using FC layers as classification this ability become vanished. FC layers are not only computationally expensive but also opt to overfitting problem. One possible solution to overcome overfitting is Dropout, concept of dropping connection of some layer during training proposed in [10]. On the other hand, FC layer can be replaced by Global Average Pooling (GAP) layer, which has proved to be effective in [9].

In this paper, We focus on category recognition, specifically the task of conveying a particular class label to an image that surrounds one or more instances of a category of object. In order to classify fruit images, We proposed a fruit recognition approach based on Pure Convolutional Neural Network (PCNN) framework with GAP layer. PCNN consists of partially connected layers, two of them followed with stride 2 and an GAP layer. To train our model we are using newly introduced fruit-360 dataset acquired from kaggle [11], which contains 81 categories of fruits. To increase efficiency and reduce time, proficient GPU was used in our experiment. Despite the apparent simplicity in our approach, for the fruit-360 dataset, our best model achieves average accuracy of 98.88% . Furthermore, to make comparison, classical CNN with dropout and without dropout is being used for classification on same dataset. As result, we found that simple PCNN with GAP architecture outperforms then previous architectures.

The remainder of this paper is organized as follows. In section II, reviews related work on fruit image classification. In Section III, we present the details of Fruite360 dataset used in our experiment. We discuss the details of our proposed PCNN

978-1-7281-1360-9/18/$31.00 ©2018 IEEE
DOI 10.1109/CSCI46756.2018.00082

model in section IV. Section V presents our simulation results. The discussion of the results is given in Section VI.

II. Related Work

A recent review of literature found that, Deep Neural Networks have made considerable progress in image classication and object detection.CNN has been widely applied to pattern recognition problem, such as image classification and object detection [5], [6], [12]. An Faster Region-base CNN approach has used for automatic fruit harvesting and detection from images. The network is trained using RGB and NIR (near infra-red) images, using combination of these images network obtain better performance [13]. Another approach using, Faster R-CNN architecture demonstrate state-of-the-art performance for fruit detection in orchard including mango's, almonds and apples [14]. Many attempts have been made with the purpose of fruit recognition and classification in robot harvesting and farming using Deep learning approach [2], [15].

CNN has been explored by some authors for fruit/vegetable recognition and classification task [16], [17]. We draw inspiration from work presented in [18], where Classic CNN model used to train and test same dataset as we used in our experiment. The CNN architecture in their work consists of multiple convolutional layers, max pooling with strides and FC layers.To further increase accuracy they generate Grayscale images so the total depth (channel) increased from 3 (RGB) to 4 (RGB + Grayscale). Additionally, preprocessing was applied in order to augment data. In order to recognize fruit images, CNN was training over 40,000 iteration with batch of 50 images. The final accuracy achieved on testing set was 96.3%.

III. Dataset: Fruit-360

The fruit-360 dataset contains 81 classes, each class has different category of fruit and is split into three sets; training set contains 41322 images, validation set has 9744 images and 4133 images in testing set,the dataset contains 55244 images in total. The fruits were filmed while rotated by a low speed motor (3rpm) and a short movie of 20 second was recorded. Due to variation in the lightning condition background was not in the uniform condition, so they changed the background of all images into white color. From 81 classes of fruits, only one image acquired from each category shown in figure 2.

All images size were 100×100 pixels having white background, high resolution images are essential for image classification because different fruits has same color and shape while not the same in size. Also, some fruits are in same size, color and shape but belong to different type of same category e.g. apple, mango, tomato's has number of variants. To classify such type of fruits, different variant of the same fruit were sorted as belong to different classes. In contrast despite other general classification datasets, fruit-360 contains only images of fruits, which have large Intra-class similarities due to same features.

Figure 1: Fruit-360: Only one image picked up from each category.

IV. Purposed Model

In this section we will describe our approach towards fruit recognition and classification. PCNN simplify the architecture with minimum number of parameters, defines like this;

{Input, Convolutional layers, Strides, Relu, GAP Layer and softmax}

Convolutional networks perceive images as volumes in three dimension objects as Height, width and depth. Depth here is three layers deep because digital color images have a Red-Green-Blue (RGB) encoding, referred as channel in CNN. Originally, fruit-360 dataset has 100×100 pixels images size so we are using same image size as our input. This input image then fed into the convolutional layer.

Convolutional Layer extract features maps by linear convolutional filters from input image followed by nonlinear activation functions (Relu, Sigmoid, tanh etc.). Then, extracted feature maps pass through another layer, two of them followed with stride 2. In PCNN we are using convolutional layers along with stride for down-sampling. A new set of feature map created by passing the filters over first down-sampling.

Strides used to control how the filter convolve around input volume. As kernel is sliding the input, it uses stride to determine how many position to skip. The amount by which

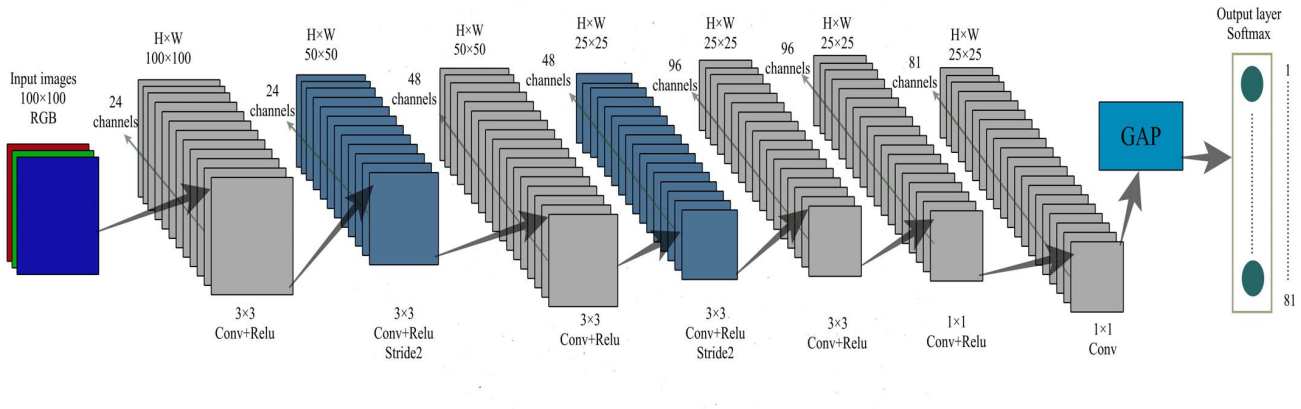

Figure 2: Architecture of Pure Convolutional Neural Network (PCNN) with Global Average Pooling (GAP) Layer.

the filter shifts is the stride. We are using stride 2, which means 2 positions to be skipped.

Linear Rectifying Units (Relu) are used for hidden layer in deep learning. Relu has output 0 if the input is less than 0 and if the input is greater than zero the output will be equal to input. So, it does not reduce the size of network. Result of using Relu activation function is much faster to train large network and also increase nonlinear property.

Global Average Pooling (GAP) used to minimize the number of parameters and to protect model from overfitting. The idea is to reduce filter size by simply taking average of whole feature map first introduced by [13]. Although, GAP is similar to max pooling layer but it perform more extreme type of dimension reduction. Where dimension $h \times w \times d$ is reduced into the dimension size $1 \times 1 \times d$. GAP reduce each feature map by simply taking the average of whole feature maps.

As the last layer, also called classification layer we used softmax layer. It can be used for predicting hundreds and thousands of classes. The output of softmax is equal to the

PCNN + GAP layer		**CNN + FC layers**		**CNN + FC layers + Dropout**	
Layer Type	Dimensions	Layer Type	Dimensions	Layer Type	Dimensions
Convolutional Layer	$24 \times 3 \times 3$	Convolutional Layer	$24 \times 3 \times 3$	Convolutional Layer	$24 \times 3 \times 3$
		Convolutional Layer	$24 \times 3 \times 3$	Convolutional Layer	$24 \times 3 \times 3$
Convolutional Layer Stride 2	$24 \times 3 \times 3$	Max-Pooling layer (2×2)		Max-Pooling layer (2×2)	
				Dropout 0.10	
Convolutional Layer	$48 \times 3 \times 3$	Convolutional Layer	$48 \times 3 \times 3$	Convolutional Layer	$48 \times 3 \times 3$
		Convolutional Layer	$48 \times 3 \times 3$	Convolutional Layer	$48 \times 3 \times 3$
Convolutional Layer Stride 2	$48 \times 3 \times 3$	Max-Pooling layer (2×2)		Max-Pooling layer (2×2)	
				Dropout 0.15	
Convolutional Layer	$96 \times 3 \times 3$	Convolutional Layer	$96 \times 3 \times 3$	Convolutional Layer	$96 \times 3 \times 3$
		Convolutional Layer	$96 \times 3 \times 3$	Convolutional Layer	$96 \times 3 \times 3$
Convolutional Layer	$96 \times 1 \times 1$	Max-Pooling layer (2×2)		Max-Pooling layer (2×2)	
				Dropout 0.20	
Convolutional Layer	$81 \times 1 \times 1$	Flatten Layer		Flatten Layer	
GAP Layer		FC Layer	256	FC Layer	256
Softmax (Activation)	81	Softmax	81	Softmax	81

Table I

Model Description of three diffrent classifiers; Pure Convolutional Neural Network (PCNN) with Global Average Pooling (GAP) layer, CNN along with Fully Connected (FC) layer and CNN with FC layer and Dropout.

total number of fruit classes.

As shown in figure 2, the network contains seven convolutional layers, two of them followed with stride 2 and a GAP layer. The Relu is applied to the output of every convolutional layer.At first ,a batch of 128 images, of 100×100 input size is passed through a convolutional layer of 24 filters of size 3×3. For dimension reduction, Convolutional layer followed with Stride 2 demonstrated in layers 2, passed through same series of layers once again, layers contains 48 kernel of size 3×3 as shown in layer 3 and 4. The output from these layers, then passed into another convolutional layer has 96 kernels of size 3×3. To more simplify the network, the output from that fed into a simple 1×1 convolutional layer from layer 6, before going through another activation. The idea is to reduce filter size from 3×3 to 1×1 by taking the average of whole feature maps first investigated by [13]. In conclusion, the resultant output passing through GAP layer, then directly fed into the softmax and the model weights are updated by back propagation to reduce loss at each iteration.

To achieve good performance, the model was trained on 40 epochs while 95,817 numbers of parameter was used. During testing, a forward pass is applied on an input image and the highest score is predicted to be fruit category.

V. Experiments

In order to create our PCNN network we used Keras with tensor flow. Keras is an open source library design to make experiment faster for Deep Learning. It is not only user friendly and easy to use but it also contains building blocks of Neural Network such as layers, activation functions and optimizers [15].

Furthermore, to make training faster and minimize the time duration, a proficient GPU is being used during experiment. GPU perform more efficiently over CPUs. During experiment we found that training our model on CPU took 6 days while using GPU it takes only 22 mints.

A. Experimental Setup

Intended for experiments on fruit-360, we first train PCNN with GAP layer as explained before. And then, to make comparison, we train Classical CNN with and without dropout, the method is essentially same as that used by [18] with some iterations. Classical CNN contains six convolutional layers, three max pooling layers for dimension reduction along with stride 2 and an FC layer, as demonstrate in second network. Thirds network model is same as second one in term of convolutional layers, max pooling layers and FC layer but Dropout layers are added in this network to avoid overfitting [10]. Model description of these three networks displayed in table I.

B. Classification Results

The experiment shows that CNN model with FC layers does not perform very well. One possible explanation is that the neurons in the FC layers have connections with all activation in the previous layers. Adding FC layers may give us better accuracy for image classification but the complexity and model parameters increased significantly. As observe across table II , Network with FC layer's accuracy is 97.41% and total number of wrongly predicted images are 107. Further,to avoid overfitting, dropout used in third network to drop connection of some layers during training. Although, dropout avoid overfitting and improve accuracy but the number of parameters remains same as CNN without dropout. Average accuracy slightly increased to 97.7% but the wrongly predicted images decreased to 88. As presented , PCNN with GAP not only achieved highest accuracy but also the number of wrongly predicted images are less than other two models.

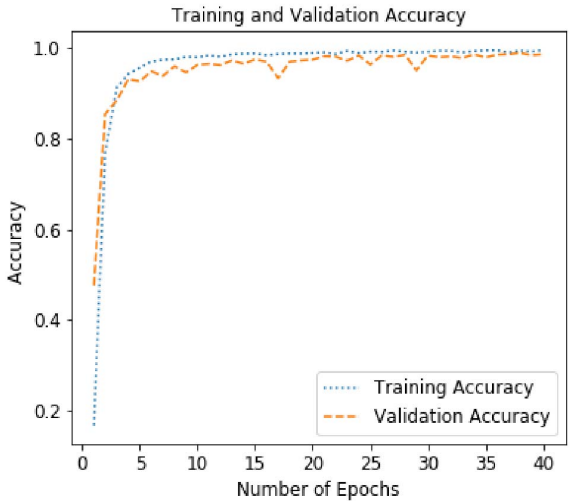

Figure 3: Average Accuracy at each epoch.

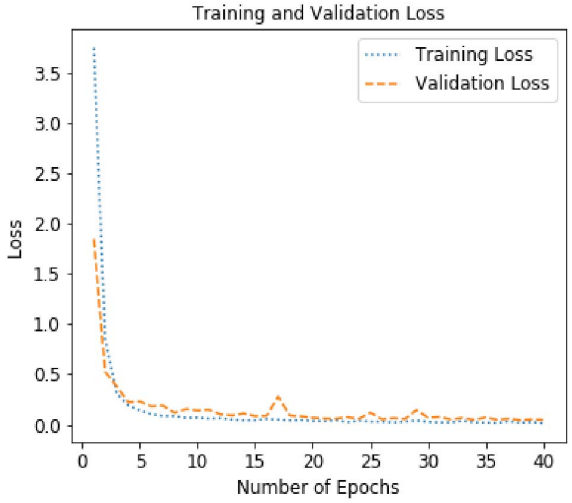

Figure 4: Average loss at each epoch.

Therefore, using GAP layer over FC layers proved to be

	Testing Accuracy	Wrong Predicted	Total no of Parameters	Computation Time
PCNN+GAP	98.88%	46 images	95,817	20.02ms
CNN+FC layer	97.41%	107 images	2,640,361	20.89ms
CNN+FC Layer+ Dropout	97.87%	88 images	2,640,361	21.02ms

Table II
Fruit images classification results of Pure Convolutional Neural Network (PCNN) with Global Average Pooling (GAP) layer and CNN with Fully Connected (FC) layer with and without using Dropout

efficient in many aspects. As can be observed in figure 3 and 4, after 40 epochs we have reached a remarkable accuracy of 98.88%. This result have been achieved without extensive optimization of the network and without using dropout. Using GAP layer have not only proved to give better accuracy for fruit image classification but also prevents whole structure from overfitting.

VI. Conclusion

We presented a novel approach to improve fruit image classification using Pure Convolutional Neural Network (PCNN) with Global Average Pooling (GAP). We have found that using GAP layer, rather than Fully Connected (FC) layer, provides superior performance and overcome overfitting problem. Moreover, we evaluated the classification accuracy on fruit-360 dataset which contains 81 category of fruit images and also variant of same fruit .In most cases, we obtained the highest classification accuracy of 98.88% when using PCNN with GAP layer, which is higher than state of the art methods. PCNN can be successfully trained to classify various types of fruit images. This makes it possible to use same approach for both object recognition and multi-class image classification.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education

References

[1] US Apple Association. http://usapple.org. Last visited on 1 September 2018.

[2] RC Harrell, DC Slaughter, and Phillip D Adsit. A fruit-tracking system for robotic harvesting. *Machine Vision and Applications*, 2(2):69–80, 1989.

[3] Woo Chaw Seng and Seyed Hadi Mirisaee. A new method for fruits recognition system. In *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on*, volume 1, pages 130–134. IEEE, 2009.

[4] Joel Andersson, Eskil Jarlskog, and Richard Wang. Fruit recognition.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[7] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[8] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[9] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[11] Andrew Thompson. Fruits 360 dataset. https://www.kaggle.com/moltean/fruits, 02 2017. Last visited on 28 August 2018.

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

[13] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*, pages 379–387, 2016.

[14] Suchet Bargoti and James Underwood. Deep fruit detection in orchards. *arXiv preprint arXiv:1610.03677*, 2016.

[15] Inkyu Sa, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, and Chris McCool. Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 16(8):1222, 2016.

[16] Lei Hou, QingXiang Wu, Qiyan Sun, Heng Yang, and Pengfei Li. Fruit recognition based on convolution neural network. In *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, pages 18–22. IEEE, 2016.

[17] Niko Sünderhauf, Chris McCool, Ben Upcroft, and Tristan Perez. Fine-grained plant classification using convolutional neural networks for feature extraction. In *CLEF (Working Notes)*, pages 756–762, 2014.

[18] Horea Mureşan and Mihai Oltean. Fruit recognition from images using deep learning. *arXiv preprint arXiv:1712.00580*, 2017.