Mango Image Recognition Challenge: Grade Classification

[1] Chung-Yu Chang (張仲宇), [2] Chiou-Shann Fuh (傅楸善)

[1] Department of Civil Engineering,
National Taiwan University,
Email: r08521516@ntu.edu.tw,
[2] Department of Computer Science and Information Engineering,
National Taiwan University,
Email: fuh@csie.ntu.edu.tw

Abstract

Grade classification of mangos can improve value and make the usage of mangos better. Mangos can be classified into Grade A, Grade B, and Grade C in Taiwan. Due to the shortage in labor force, classifying the grade of mango by artificial intelligence is an ideal solution. In this research, we attend AI CUP 2020 Mango Image Recognition Challenge. To reach higher accuracy, we proposed two methods to help farmers. YOLO v2 was adopted to segment mango images for both methods. Method 1 used a simple CNN structure to classify, while Method 2 is a two-step classification process for a better result. Other different models and preprocess methods were also tried. Detail experimental results of different methods was provided in this paper.

Keywords: *Image recognition, Neural network, Subgrade classification.*

1. Introduction

Mango is one of the most valuable export agricultural products of Taiwan. Although Taiwan's mango is famous for its good flavor and high quality, it faces other countries' price competition. Value improvement of Taiwan's mango is a big issue recently. The most urgent thing is to improve the process technique after collection. Mangos can be classified into three classes, A, B, and C, for export, domestic, and food processing, respectively. However, due to the shortage in labor force, it is hard to classify by manpower. Using artificial intelligence to help classifying is an ideal solution to this issue.

Thanks to the development of computer vision in machine learning, we can detect and classify objects appear in images easily nowadays. However, existing machine learning methods can only provide classification among some common objects like, person, bike, and car. It is seldom to see a classifier can handle fruit categories, not to mention the classification of mango grades. In this research, we proposed a novel classification process that can automatically segment mangos from images, detect mango edges, and classify mango grades.

2. Problem Description

AI CUP 2020 Mango Image Recognition Challenge is hold by Industrial Technology Research Institute, Taiwan. The goal of this challenge is to accurately classify mangos into three Grades, A, B, and C. Table 1 listed the number of images in the challenge dataset. Fig. 1 is an example of the three grades.

As we can see, it is hard to distinguish the grades from first glance. The biggest difference between Grade A and Grade B is the color of their peel. Color distribution of Grade A is smoother than Grade B, also, it is redder on the peel of grade mangos, this means the mango is more delicious and more valuable. The difference between Grade C and other two grades can be distinguished easier. Because there would be some dark spots on the peel of Grade C mangos.

However, to see with human eyes is one thing, to see with computers is another. How to let computers know the differences and classify them are big challenges. We human can identify objects' category by experiences and knowledge, what computers see are some list of numbers.

Moreover, aside from the appearance of mango itself, the backgrounds of released dataset are annoying. Computers cannot know the difference between backgrounds and mangos without learning, it is a challenge to find features among them to segment better.

3. Proposed Method

In this challenge, we proposed two methods to see the performance on mango classification problem. Method 1 simply used convolutional neural network (CNN) to classify grades, and

Method 2 adopted a two-step classification process. We firstly present the concepts of two methods, and then introduce the detail process of them, respectively.

Table 1: Number of images of dataset. Train means training set, and Dev means development set.

	Train	Dev
Grade A	1,792	243
Grade B	2,068	293
Grade C	1,740	264
Total	5,600	800

Fig. 1: Example of three mango grades. Left column is Grade A, center column is Grade B, and right column is Grade C.

3.1 Method 1 Overview

As shown in Fig. 2, Method 1 first segments mangos from raw images using specially trained YOLO v2 [1] model, then segmented images are put into a CNN model to classify the three grades.

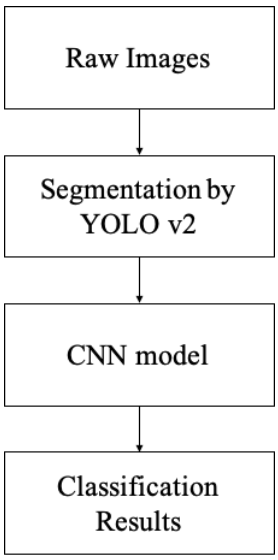

Fig. 2: Flowchart of Method 1.

In this method, we want to know the performance on simply model structure and process. If we can accurately classify grades using simple process, we would not need to spend too much time on playing tricks.

3.2 Image Segmentation

As we mentioned at Section 3, the backgrounds of released images are messy. This situation would make model mistakenly extract features from background, causing serious bias. To prevent this, we adopt Darknet of YOLO v2 [1] to crop mangos from original images.

Due to lack of time, we annotated 450 mango images for the training of segmentation. After the simple training, we can successfully detect mangos in 6,159 out of 6,400 images. Fig. 3 is an example of detection and cropping.

Fig. 3: An Example of our mango cropping.

After the background subtraction, the cropped mango images can be used to preprocess and be used by any type of prediction model. It is a more flexible way to apply the images.

3.3 Convolutional Neural Network Model

The classification model we use in this challenge is a simple CNN structure. Fig. 4 is the structure of our CNN. There are five convolution layers, each is followed by a max pooling layer. Then the features will be flattened and followed by two fully connected layer. The output is the

grade.

3.4 Method 2 Overview

The idea of Method 2 was inspired by the appearance differences of three grades. Since Grade C mangos have significant dark spots on peel, we can firstly separate Grade C from data to get a better classification result. After the separation of Grade C, we can then take advantage of color distribution differences between Grade A and Grade B to classify. This is a two-step method.

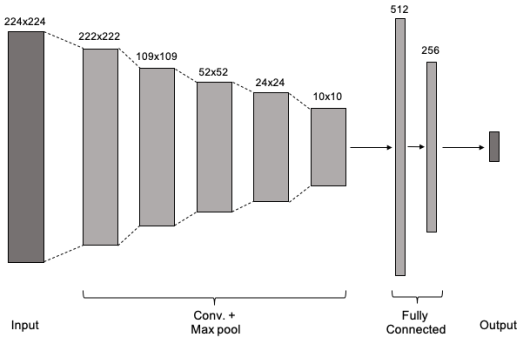

Fig. 4: Our CNN structure.

3.5 Preprocessing of Method 2

First of all, backgrounds were removed. We used traditional edge detection skills to find edge features. Images were turned into grayscale images for better extraction, then convolved with Gaussian blur kernel, and finally we used Canny filter to extract edge.

In Fig. 5, we can see obvious difference between Grade A and Grade C after edge extraction. We can also observe differences of three grades after the edge extraction in Fig. 6. For Grade A and Grade B, there are no obvious spot appear near center of images, this means we can separate Grade C from other two grades first. The same CNN structure as Method 1 was adopted.

Before Step 2 classification, we adopted another preprocess to extract features between Grade A and Grade B. Because Grade A mangos have more red color appear on peel, color difference may be a significant feature to distinguish the two grades. The color of mangos is consisted of red and yellow, if we remove color of blue from RGB channel, we can tell two grades easier. Aa shown in Fig. 7, there are more yellow color appear on peel of Grade B mangos.

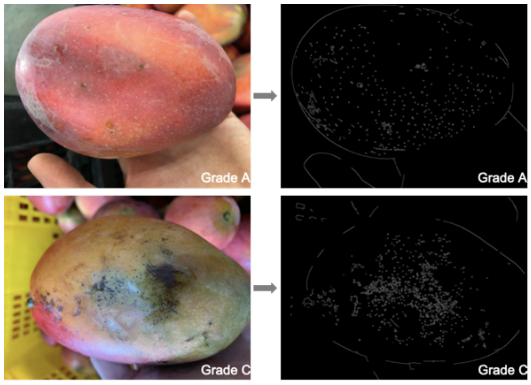

Fig. 5: Original mango images of Grade A and Grade C and their images after edge extraction.

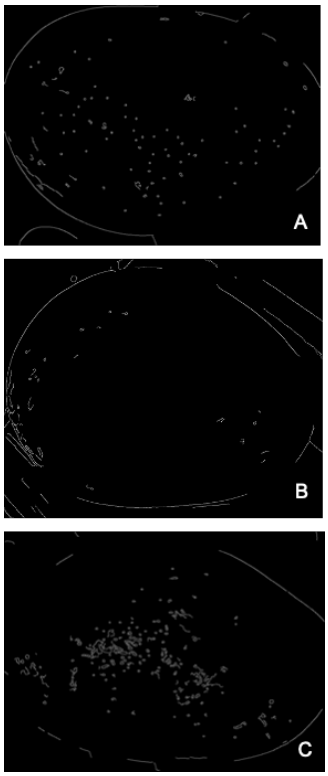

Fig. 6: Result of three grades of mangos after edge extraction. Image color was modified for better representation.

3.6 Classification Process of Method 2

Step 1 of Method 2 is to distinguish Grade C from other two grades. We extract edge features by Gaussian blur followed Canny filter. After preprocess of Step 1, the processed images were put into CNN as mentioned in Model 1. Assume the model can efficiently find Grade C from data, Step 2 can be adopted then. In Step 2, we remove blue color channel from Grade A and Grade B images, put images only with red and yellow color channel into same CNN model. This step can classify Grade A and Grade B. Fig. 8 is the rough flowchart of Method 2.

Fig. 7: Comparison between grades A and B after subtraction of blue channel.

4. Experimental Results

The baseline of this competition had been released by organizer. The baseline of grade classification is ACC = 71.625%. The baseline used four Single Nets to extract features from images, then concatenate features from four nets by decision function. The fusion was fed into support vector machine (SVM) models to classify. The result of our proposed methods should exceed this standard. Of course, the higher the accuracy means that the higher chance to win this competition.

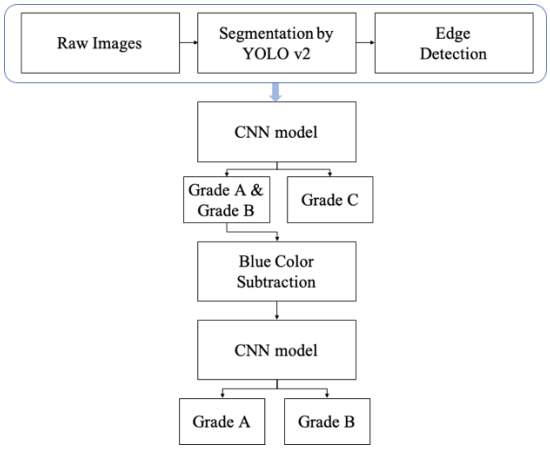

Fig. 8: Flowchart of Method 2.

To see how image augmentation tricks can improve the accuracy of prediction, we also adopted augmentation functions build in TensorFlow. Images will be rescaled, rotated, width shifted, height shifted, sheared, zoomed, and flipped. By doing this, we can get more varied training data with transformation.

Although we think the subtraction of messy background can help model finding really useful features, we still want to know what is the difference between them. Method 1 and Method 2 without cropping were tested.

If a simple structure can reach an acceptable and accurate result, we need not to cost extra time on computing. A simple and efficient model is what we want. To find this, we simplified the CNN structure and predicted by this structure.

Table 2 listed the detail results of different combination, and Fig. 9 is different adjusted Method 1's training and validation accuracy and loss. We want to know whether augmentation and cropping mango's RoI from original images can help improve classification accuracy. So, for each of the two methods, we tested different combinations to make detail comparison. A simpler CNN structure also tested. This simple structure removed one convolution layer, one max pooling layer, and one dense layer.

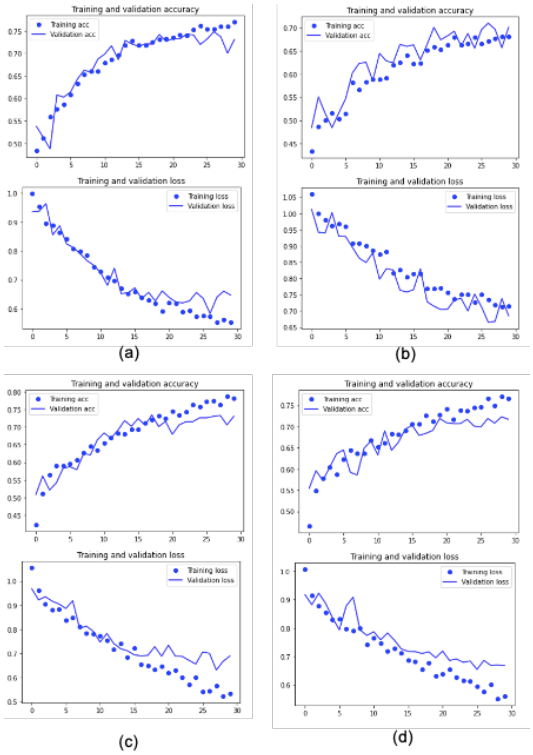

Fig. 9: Training and validation accuracy and loss of different adjusted Method 1. (a) the original Method 1; (b) Method 1 with augmentation; (c) Method 1 without image cropping; and (d) Simpler Method 1.

4.1 Results of Method 1

We can start to analyze the results from Method 1. The accuracy of Method 1 is 0.7304, which is also the best among all testing combinations. Our original proposed method (without image augmentation and with cropping) is better than the baseline. Normally, image augmentation can help a model with small dataset to improve accuracy. In Method 1, the model with image augmentation has a lower accuracy than original one, possibly because image augmentation will rotate and crop randomly to produce more image data. Moreover, this random transform extracts wrong features and confuses

the model. Method 1 without image cropping has a slightly worse validation accuracy. The difference is small, and computational time can be reduced because we do not have to train a model to crop images. The accuracy of simpler structure is 0.7170, which is worse than the base line.

Table 2: Results of different structures. *Aug.* means image augmentation; *Crop* means cropping mangos' RoI from original image; and *Simple* means simpler CNN structure.

	Training Accuracy	Validation Accuracy
Method 1	0.7694	**0.7304**
Method 1 w/ *Aug.*	0.6807	0.7004
Method 1 w/o *Crop*	0.7809	0.7300
Method 1 *Simple*	0.7655	0.7170
Method 2	0.9524/0.8335	0.7605/0.7571
Method 2 w/ *Aug.*	0.7755/0.6932	**0.7749**/0.7643
Method 2 w/o *Crop*	0.7705/0.8591	0.7670/**0.7900**

4.2 Results of Method 2

Method 2 takes much time to train and predict. Compared with Method 1, this method has to preprocess images, i.e., edge detection, and subtraction. After such preprocessing, they have to feed into CNN model, respectively. The results of Method 2 have two parts. First part is accuracy after edge detection, i.e., we separate Grade C from all of mangos first. The second part is accuracy after blue color subtraction, i.e., we separate Grades A and B by their color.

The highest validation accuracy of Method 2 does not appear in our original proposed method. Method 2 with image augmentation has the highest validation accuracy of Step 1, while Method 2 without image cropping has the highest validation accuracy of Step 2. Two parts of validation accuracy in Method 2 with image augmentation and Method 2 without image cropping are both higher than original proposed method, this is different from the results of Method 1. In this method, we observe the most different features of these grades and classify separately. Augmentation here can really augment the important features and help the model to classify them more easily. As for cropping, it is hard to explain why images without cropping have higher accuracy here. It should be more accurate if we remove the irrelevant features.

The training accuracy of original Method 2 are overfitted. The size of this dataset is small (5,600 training images and 800 validation images). When we adopt image augmentation, we can increase the size of our data to prevent overfitting issue. Training accuracy drops from 0.9524/0.8335 to 0.7755/0.6932, and they are closer to validation accuracy. If we remove image cropping from origin Method 2, training accuracy also drops. We infer that more features not focus on mangos themselves are found, these features mislead the model, so the training accuracy drops. This is a good mystery to be researched in the future.

5. Ablation Study

A traditional machine learning model, SVM, was also adopted to compare the results with our proposed methods. Table 3 is the results that we replace CNN model with SVM. Accuracy in Method 2 is slightly better than originally proposed Method 2. The computational time of SVM is much smaller than CNN models.

We compare our CNN model with state-of-the-art CNN models: AlexNet [2], VGG16 [3], InceptionNet [4], and ResNet [5]. Because of the computational time, we only test some of the combinations in Table 4. All of these models use the cropped image as training and validation data.

If we train these state-of-the-art models with 30 epochs, the performance of AlexNet and ResNet are worse than our proposed Method 1, and our Method 1 is simpler than these models. We increase the training epochs of AlexNet, validation accuracy increases and exceeds our proposed Method 1. But the time is valuable, we cannot continuously increase the epoch to get a better result, and it would lead to some overfitted issues. This kind of problems can be discussed in the future. Fig. 10 is an example of overfitted model using InceotionNet.

Table 3: Two methods using SVM as classification model.

Method 1	
	Accuracy
SVM	0.68
Method 2	
	Accuracy
SVM	0.76/0.77

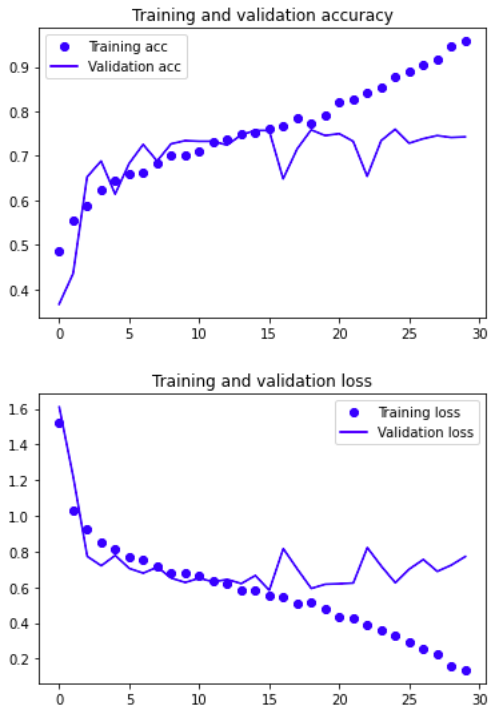

Fig. 10: Training and validation accuracy and loss when we used InceptionNet as CNN structure to classify.

Due to the time issue, we only using AlexNet to test Method 2. The validation accuracy is lower than results listed in Table 2. It is hard to increase the accuracy by training more epochs. As we can see in Fig. 13, the training accuracy is pretty high, whereas the validation accuracy cannot follow up. The model is overfitted, we should adjust the parameter or model to prevent this.

We also tried combining AlexNet and Random Forest to see how is the effect. If we just use original mango images without any preprocessing, we can get a validation accuracy of 65.5%. Here we extract the features from second dense layer of AlexNet. We can get 4,096 features from each image. Then we put these features into Random Forest to train. Unfortunately, the accuracy is 67%, which is kind of low.

Table 4: Methods 1 and 2 with different CNN structures.

Method 1		
Model	Training Accuracy	Validation Accuracy
AlexNet [2] Epoch=30	0.7124	0.6940
VGG16 [3]	0.8272	0.7557
InceptionNet [4]	0.9576	0.7429
ResNet [5]	1.0000	0.6543
AlexNet Epoch=50	0.7934	**0.7651**
Method 2		
Model	Training Accuracy	Validation Accuracy
AlexNet	0.9992/0.8578	0.7605/0.7771

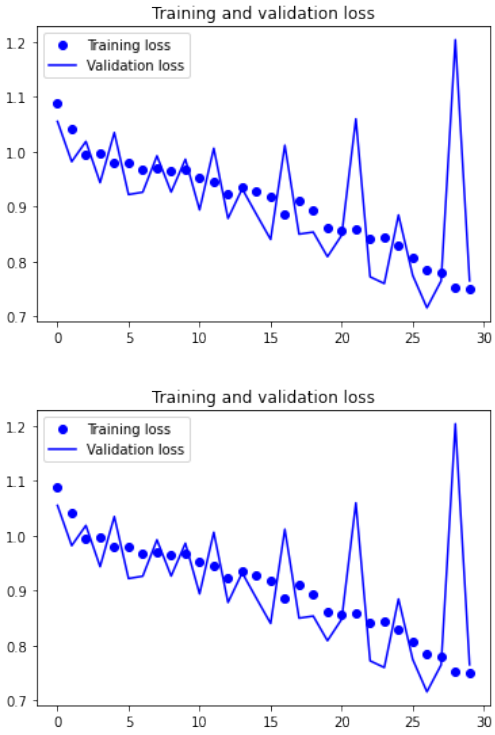

Fig. 11: Accuracy and loss when we used image without preprocessing.

6. Testing Set

The competition organizer released a testing set for preliminary. The testing set contained 1,600 unlabeled images of mangos. Baseline for preliminary was also released. The baseline model used a two-stage classification network. This model was evaluated on the testing set, and got an accuracy of 70.25%.

To get a better accuracy, we combine 3 of models listed above. They are our proposed Method 1, AlexNet, and VGG16. The original image was first cropped by our trained YOLO v2, then predicted by three models to get three different confidence set of three grades. Every model can output three confidence for three grades. We used highest mean confidence as the final prediction result. Fig. 12 is the testing process. We got an accuracy of 75.26% on validation set, while the testing accuracy is 73.68%, which labels can only be known after submitting to competition platform. We also submitted prediction using single model. Accuracy of VGG16 was 69.87%, 72.43% for our Method 1, and 74.81% for AlexNext, which is better than our combination result.

Our team's ranking is 325 among 543 in public leaderboard. Table 5 is the top 10 teams and our team on public leaderboard. Their accuracy can reach up to 83.6%.

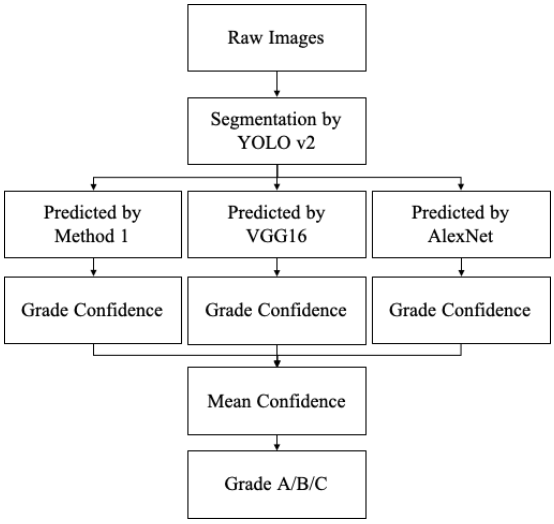

Fig. 12: The testing process.

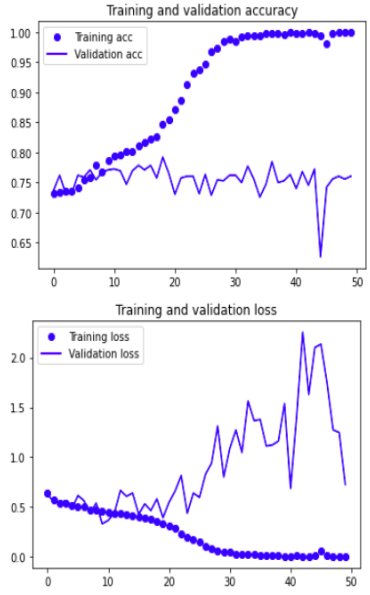

Fig. 13: Training and validation accuracy and loss when we used AlexNet as CNN structure in Method 2 to initially separate Grade C from three grades.

7. Conclusions

Image classification using CNN is a popular choice to solve this kind of problem. In this paper, we proposed two methods to classify three grades of mangos. A YOLO v2 model was trained to crop bounding box of mangos from images with messy background. Method 1 directly puts cropped images into our CNN structure to get classification results. Method 2 first adopts edge detection to distinguish Grade C from all grades, and adopts blue channel subtraction to classify Grade A and Grade B. Image augmentation, simple structure, and image without cropping were also tested to know the difference of different methods. We also adopted SVM and other state-of-the-art CNN structures to classify. Our proposed models are better than the baseline of this competition.

This paper provides a detail experimental results of this kind of subclass classification problem. We adopted different methods and CNN structures to see their result differences.

Subgrade classification of mango is a very different problem aside from classification among fruits. We can easily distinguish fruits from their shape, color, and texture. However, subgrade classification of same kind of mangos is another world. They are all look the same, we can only classify them by tiny clues. More methods should try in the future to get better accuracy.

Table 5: Top 10 public leaderboard teams and ranking of our team.

Ranking	Team	Score	Upload Times
1	刷榜請搜NYKD-54	0.83625	24
2	Beginner	0.83125	25
3	機器學不會	0.829375	23
4	fammy	0.828125	19
5	Xie29	0.828125	19
6	hao123kuo	0.825	13
7	GPU NOT FOUND	0.82375	23
8	打芒果	0.823125	25
9	一起學AI隊	0.821875	23
10	下次初賽不算分先講好嗎？	0.821875	25
...	...	...	...
325	**Taco**	**0.748125**	**5**

References

[1] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 6517-6525.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," In *Neural Information Processing Systems 25 (NIPS 2012)*, Lake Tahoe, NV.

[3] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” In *International Conference on Learning Representations (ICLR)*, San Diego, CA, 2015.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going Deeper With Convolutions,” In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.