20/05/2024

|  |  |
| --- | --- |
| Date | Name |
|  |  |
| 30/6/2024 | Ziheng Liou |

**Food Image classification Using MobileNetV2 with SVM classifier.**

Table of Contents

[Food Image classification MobileNetV2 with SVM classifier. 2](#_Toc176168106)

[(1) Methodology 2](#_Toc176168107)

[MobileNetV2 2](#_Toc176168108)

[Transfer learning using mobileNetV2 2](#_Toc176168109)

[Hybrid approach combining CNN and SVM 3](#_Toc176168110)

[(2) Evaluation 4](#_Toc176168111)

[MobileNetV2 model performance 4](#_Toc176168112)

[Result Summary of all mobilenetV2 6](#_Toc176168113)

[Confusion Matrices Analysis 7](#_Toc176168114)

[(3) Conclusion 10](#_Toc176168115)

[Appendix 11](#_Toc176168116)

# Food Image classification MobileNetV2 with SVM classifier.

## Methodology

### MobileNetV2

MobileNetV2 in Fig \_ is an pre-train model that had been pre-trained on ImageNet, a dataset containing over 14 million images and 1000 classes. It’s has been achieves 72.834% top-1 accuracy and 91.060% top-5 accuracy on ImageNet validation set(<https://github.com/d-li14/mobilenetv2.pytorch>). As MobileNetV2 is Prepared for small dataset and an 10 class with 800 image for training set and 200 image for testing set had been developed.

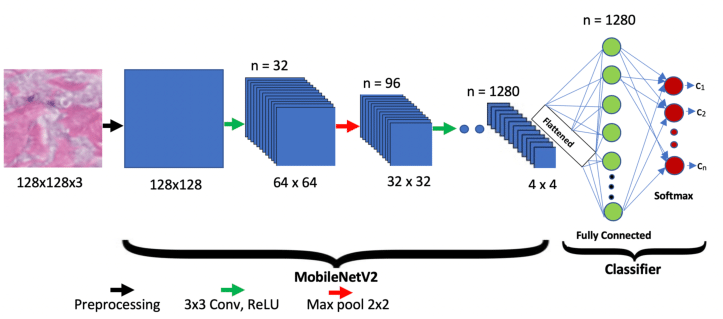


Figure \_ Example MobileNetV2

### Transfer learning using mobileNetV2

The Deep learning model has been developed with an mobileNetV2 as a backbone and by performing transfer learning. Transfer learning leverages pre-trained models on large datasets to improve performance on specific tasks with limited data [1]. The original pre-trained mobileNetV2 has take 17,300 sec so for an wise of time, I had use “mobile\_base.trainable = False” to freezing the model parameters. By freezing the early layers of deep neural networks, it ensures that these fundamental features are preserved and reused. The next try is unfreeze the last 10 layer of model and by freezing early layers and training only the later ones, the model can focus on learning features more specific to the new task. This technique also reduce computational power and time, making it more feasible to train on limited resources and preventing overfitting.

The mobileNetV2 will be add an classification layer with an “GlobalAveragePooling2D”, “Dropout”, “Dense - softmax” in Fig\_. GlobalAveragePooling2D layer performs down sampling by computing the mean of the height and width dimensions of the input to reduces the spatial dimensions of a tensor . The Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time to prevent over fitting and Dense – softmax will act as an classifier to classify the image.

A screenshot of a computer program

Description automatically generated

Figure \_ mobileNetV2 summary with classifier

### Hybrid approach combining CNN and SVM

The study highlights the relative advantages of traditional machine learning (SVM) in small datasets and deep learning (CNN) in larger contexts [4]. For this approve the we are using an small dataset for training and it conclude that SVM will be an better performance classifier then CNN in this situation. The hybrid approach combining CNN and SVM was explored to take advantage of CNN's feature extraction capabilities and SVM's robust classification properties [2].

In the example it is using an CNN model and combination of SVM classifier for handwriting recognition. With this technique working for food image, I had improve it by the method of using an MobileNetV2 with transfer learning leverages. Instead of classifying them with the softmax layer, SVM is used to classify these extracted features [3]. Food image dataset is feed into MobileNetv2, and features are extracted by this network after applying depth wise separable convolution, batch normalization and ReLU as non-linearity is applied with pooling operations throughout the network [3]. I had used the method of this research paper to train my model but I had only extracted from “out\_relu layer”. The extracted features will be training the SVM model with a use of one-vs-all multiclass classification.

## Evaluation

### MobileNetV2 model performance

In this section, we will present and analyse the results of three different approaches for food image classification: MobileNetV2 with CNN classification, MobileNetV2 with SVM classification (all layers frozen), and MobileNetV2 with SVM classification (last 10 layers unfrozen). We will compare these methods in terms of accuracy, F1-score, prediction time, and computational demands. Additionally, we will analyse confusion matrices to identify and discuss any failure cases.

#### MobileNetV2 train and test result

The pre-trained mobileNetV2 has take 17,300 sec for training and had an Test accuracy of 0.178. The model had been train with only 10 epoch

A graph of a graph of a graph

Description automatically generated with medium confidenceThe Graph in Fig \_ had show an overfitting with an high accuracy in training set and low accuracy in validation set.

Figure \_ mobileNetV2 Classification performance

#### MobileNetV2 with Transfer learning train and test result

By using Transfer learning technique, There are 2 testing about how well is the model work. The Fig\_ had shown that the freeze all layer of the model had an good performance in the model, It does not had any overfitting and the training only take 980 sec with 20 epochs. This may because the dataset we had use is an food 101 that are an classical dataset for machine learning. So the image Net (source of data for mobileNetV2 pre-train ) may include the data.

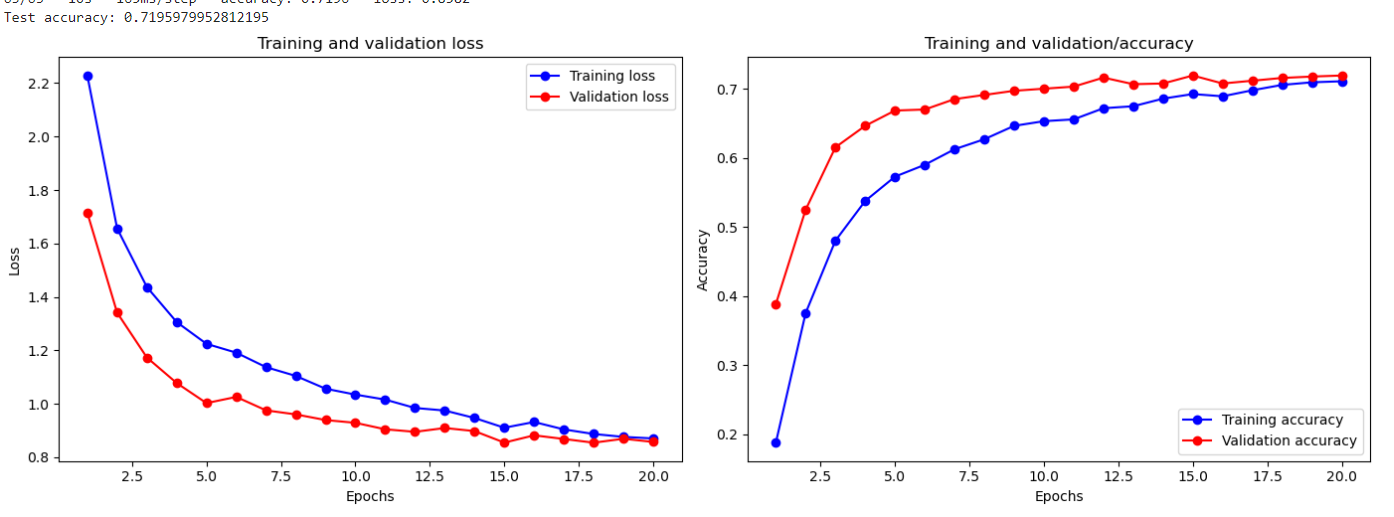


Figure \_ mobileNetV2 Classification performance Freeze all layer

#### MobileNetV2 with Transfer learning(unfreeze last 10 layer )train and test result

The model in Fig\_ has only been freeze the upper layer of the last 10 layer and the purpose of this model is to be tuning the model to have more focused on “food”. The model will be retraining the last 10 layer with the data we provide in the 10 class. This can let model to learn more detail of food and more suitable for food classification task. The model compare to the model before has a longer training time (1160 sec) but has a testing accuracy of 0.701 and from graph in Fig \_ it had been show a over fitting as the lower performance of validation accuracy compared to training and it also unstable in the training.

A graph of a graph and a graph of a graph

Description automatically generated with medium confidence

Figure \_ mobileNetV2 Classification performance Freeze all layer and unfreeze last 10 layer.

For the best result of the model, we will use the mobilenetV2 that had an best performance that is the freeze all layer. The Performance of the model in classification food has been show in Fig \_. The test set prediction time will take 15 sec for prediction for the best model

### Result Summary of all mobilenetV2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Accuracy | F1-Score | Prediction Time(test set) |
| mobilenetV2 - CNN Classification  (Freeze all layer) | 0.719 | 0.716 | 15 Sec |
| mobilenetV2 – SVM  (Freeze all layer) | 0.724 | 0.722 | 1.7 Sec |
| mobilenetV2 – SVM  (unfreeze last 10 layer) | 0.74 | 0.74 | 1.86 Sec |

table \_ Result Summery

### Confusion Matrices Analysis

#### MobileNetV2 - CNN Classification (Frozen Layers)

This confusion matrix (Fig \_) shows the performance of the MobileNetV2 model with all layers frozen and using a CNN classifier. The model achieved an accuracy of 71.9% and an F1-score of 71.6%. While the model performed reasonably well, there were notable misclassifications, particularly between classes 5 and 6, and classes 0 and 8, indicating areas where the model struggled to differentiate.

So, by reducing the problem we may use SVM classifier for the work. By Extract features from the training and test sets and flatten the data, The OneVsRestClassifier for SVM had use 22.15 sec for training and it come with OneVsRestClassifier accuracy 0.72 and OneVsRestClassifier test time 1.7 sec. The prediction time is more closely related to our aim of using the model in device with fast result generation.

A screenshot of a graph

Description automatically generated

Figure \_ model classification of freeze all layer(Confusion Matrix)

#### MobileNetV2 - SVM Classification (Frozen Layers)

This confusion matrix (Fig\_) illustrates the performance of the MobileNetV2 model with all layers frozen but using an SVM classifier. This approach yielded a slightly higher accuracy (72.4%) and F1-score (72.2%) compared to the CNN classifier. The SVM classifier performed better in distinguishing between certain classes, particularly reducing misclassifications in classes 5 and 6.

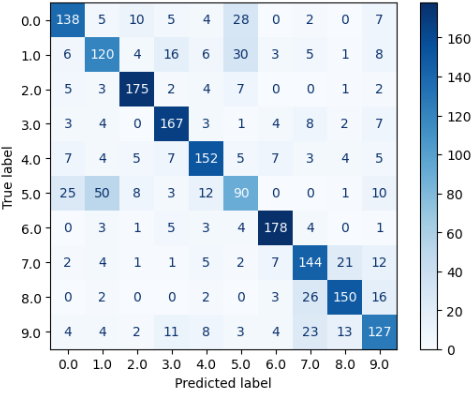


Figure \_ model of freeze all layer and SVM classifier (Confusion Matrix)

#### MobileNetV2 - SVM Classification (unfrozen last 10 Layers)

This confusion matrix (Fig\_) represents the MobileNetV2 model with the last 10 layers unfrozen, combined with an SVM classifier. This configuration achieved the highest accuracy (74.0%) and F1-score (74.0%) among the three methods. The ability to fine-tune the last few layers allowed the model to better adapt to the specific characteristics of the food image dataset, resulting in fewer misclassifications.

A graph of a number

Description automatically generated with medium confidence

Figure \_ model of unfreeze only last 10 layer and SVM classifier (Confusion Matrix)

## Conclusion

For model performance, the results demonstrate that unfreezing the last 10 layers of MobileNetV2 and using an SVM classifier provided the best performance. The fine-tuning allowed the model to learn more specific features of the food images, improving classification accuracy. The comparison between CNN and SVM classifiers highlights the advantage of SVM in this context, particularly when combined with transfer learning. The SVM classifier consistently outperformed the CNN classifier in both frozen and unfrozen layer configurations. The prediction times indicate that the SVM classifier is more efficient than the CNN classifier, even when fine-tuning additional layers. This makes the SVM approach more suitable for real-time or resource-constrained environments as this is a main objective of this project.

Analysing the confusion matrices reveals specific failure cases, such as misclassifications between visually similar food items. For instance, classes 0 and 8, and classes 5 and 6, showed higher rates of confusion. This suggests that further improvements could be made by enhancing the model's ability to distinguish between these specific classes, potentially through additional data augmentation or incorporating more fine-grained feature extraction.

The evaluation clearly indicates that using MobileNetV2 with the last 10 layers unfrozen and an SVM classifier provides the best balance of accuracy, efficiency, and computational demand for the food image classification task. This approach leverages the strengths of both transfer learning and traditional machine learning techniques, resulting in superior performance.

# Appendix

[1]VijayaKumari.G, Priyanka.V, Vishwanath.P, “Food classification using transfer learning technique”, June 2022. Available: <https://www.sciencedirect.com/science/article/pii/S2666285X22000334>

[2] Ahlawat .S , Choudhary.A, “Hybrid CNN-SVM Classifier for Handwritten Digit Recognition”, 16 April 2020. Available: <https://www.sciencedirect.com/science/article/pii/S1877050920307754>

[3] Yadav . S, Chand. S, “Food image recognition based on MobileNetV2

using support vector machine” 22 April 2024. Available: <https://books.aijr.org/index.php/press/catalog/download/114/44/1501-1?inline=1>

[4] Wang.P, Fan.E, Wong.P, “Comparative analysis of image classification algorithms based on traditional machine learning and deep learning”, 1 August 2020 . Available: <https://www.sciencedirect.com/science/article/abs/pii/S0167865520302981>