Capstone 2
Fruit Classifier

Sava Dashev

Problem

Customers buy fresh fruits in grocery stores. We want to automatically recognize type of the fruit by the scale – fruit type.

Data source

The source of data is kaggle fruit-recognition data set.

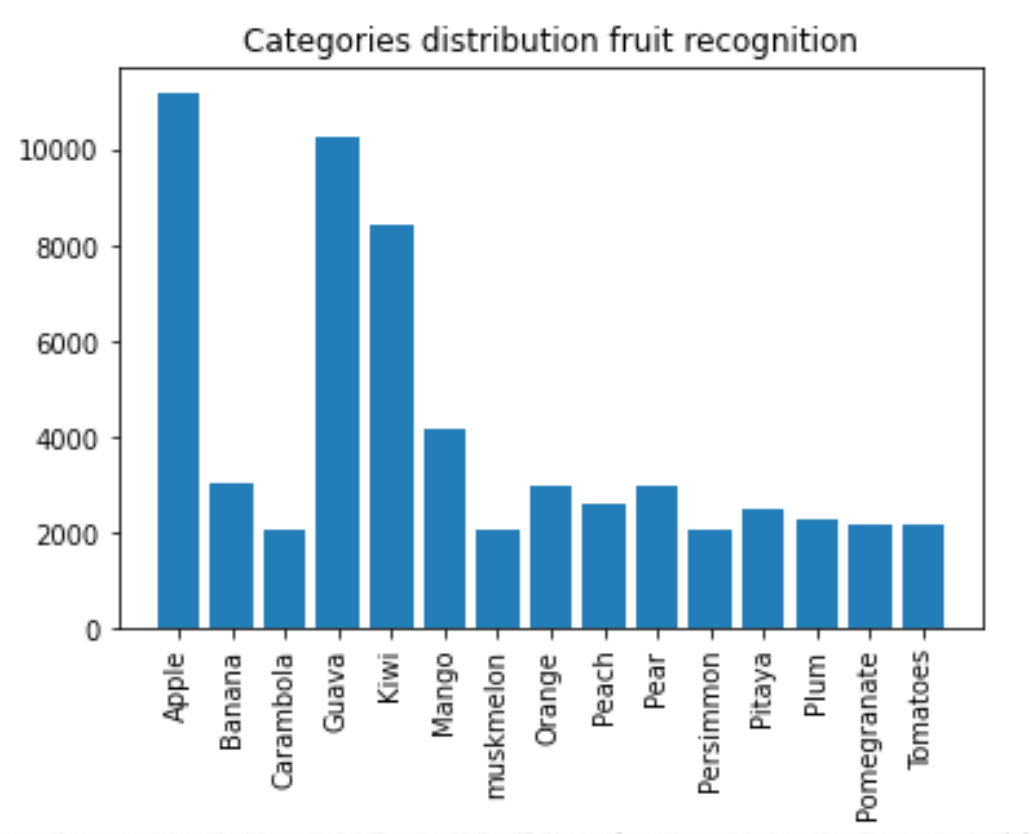
- Size of data is 8 GB.
- Each image is size (200, 200), (320, 258) or (480, 322).
- Only one image is size (320, 240).
- There are 54.

Data - classes

The images represent 15 different fruits.

- 3 classes have more than one sub-folder containing images;
- 9 classes have “resized” or “resized resized” images – 200 by 200 pixels.
-

Data distribution



Data findings

We found 15 type of images.

- Class Apples has the most images.
- Classes Guava and Kiwi has similar, but smaller amount images.
- All other classes have about 2000 to 4000 images each.

Data

We are going to use the images which are the most in the dataset.

We created file with names of images same size.

Further, we mix all names in train test split and saved for machine learning.

Deep learning

Common architecture used for computer vision is convoluted neural network.

- Common ways to transform image is convolution, maxpooling or filtering.

Deep learning

- Convolution is data manipulation on image pixels. In the core is calculating values of each pixel of a feature map by multiplying intensity value of part of the image and a small tensor.
- It is used to learn local patterns.
- At each level, we apply multiple convolutions.

Deep learning

- Maxpooling is a way to change the scale of image. It is applied between convolutions, so the convolution will learn local patterns at different scale.
- We can also use average of all pixels to change the scale.

Deep learning

There are two main approaches to create an architecture.

- We can design model from scratch. In this case, we choose all parameters and structure.
- We can use frozen base and attach new classifier to learn the classes.

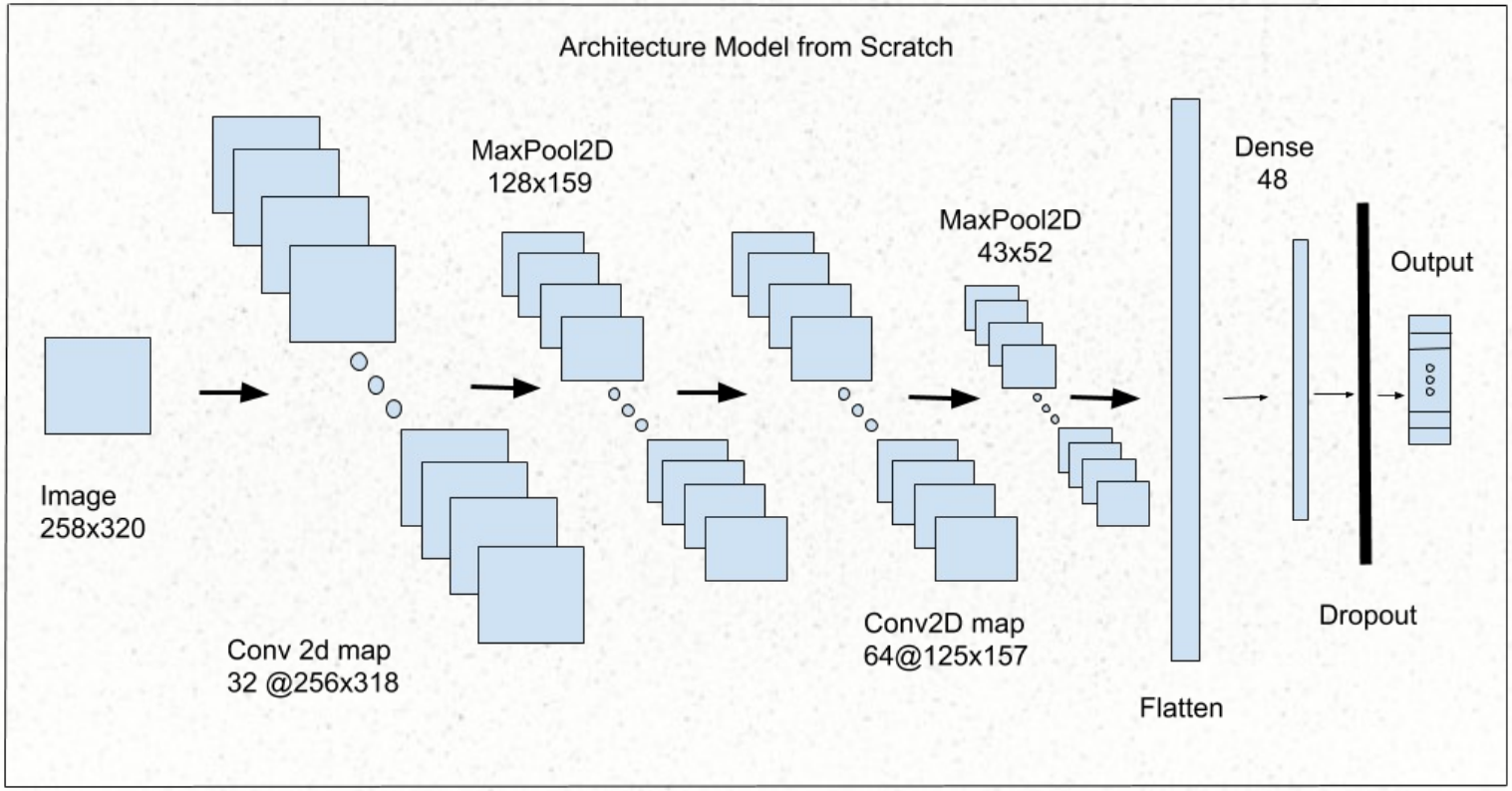
Deep learning

When re-purposing model, we can use already fitted values and only re-train the last layers.

Other option is to train the whole model from scratch, only using the structure.

Our data set is small, we used the first approach.

Model from scratch



Training of the model

Training the model takes three pieces: optimizer, loss function, and metric to measure performance.

- The optimizer is the algorithm to search for values of the connections that will minimize the loss function.
- We used a few optimizer algorithms to fit.

Training of the model - loss

Training the model takes three pieces: optimizer, loss function, and metric to measure performance.

- The loss function for multi-class problems is called cross-entropy.

Training of the model - metrics

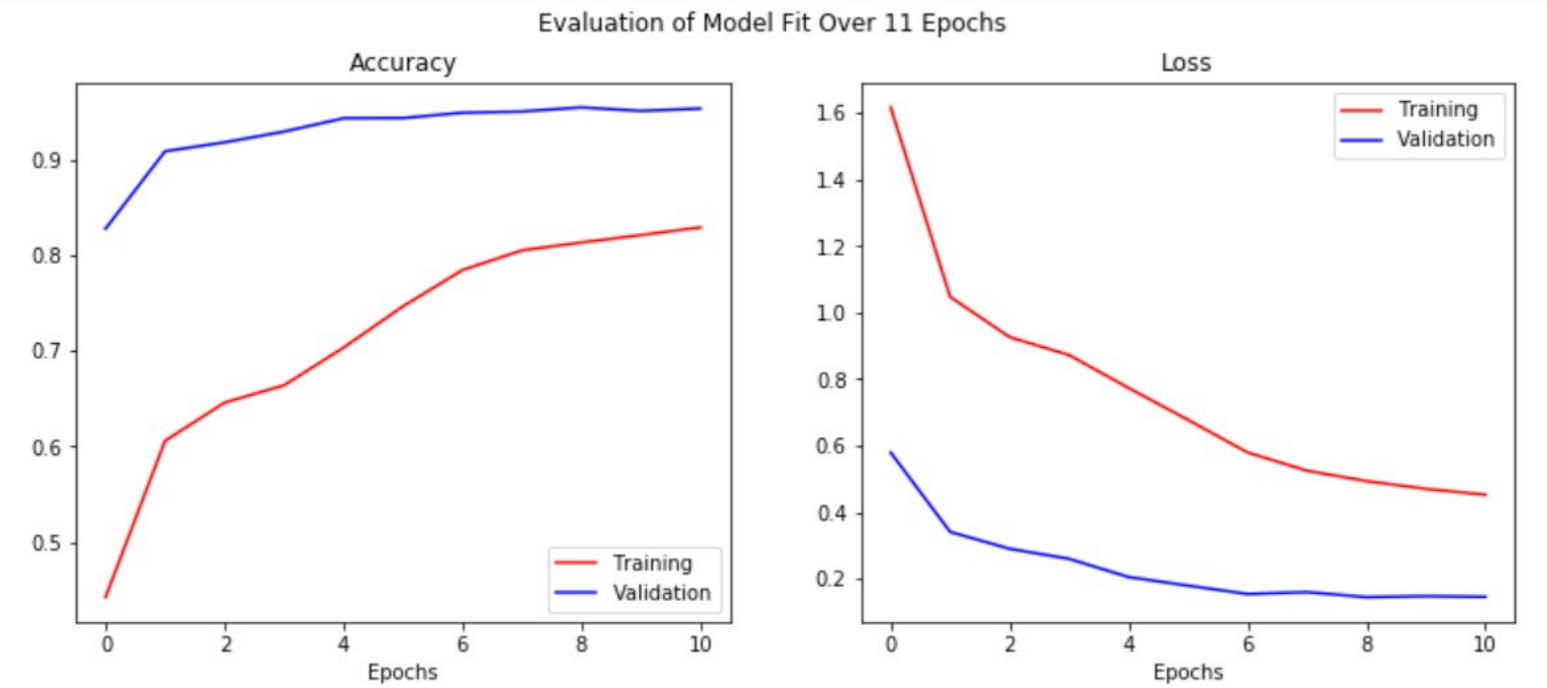
Training the model takes three pieces: optimizer, loss function, and metric to measure performance.

- We measure the accuracy of classifying images.

Base model performance

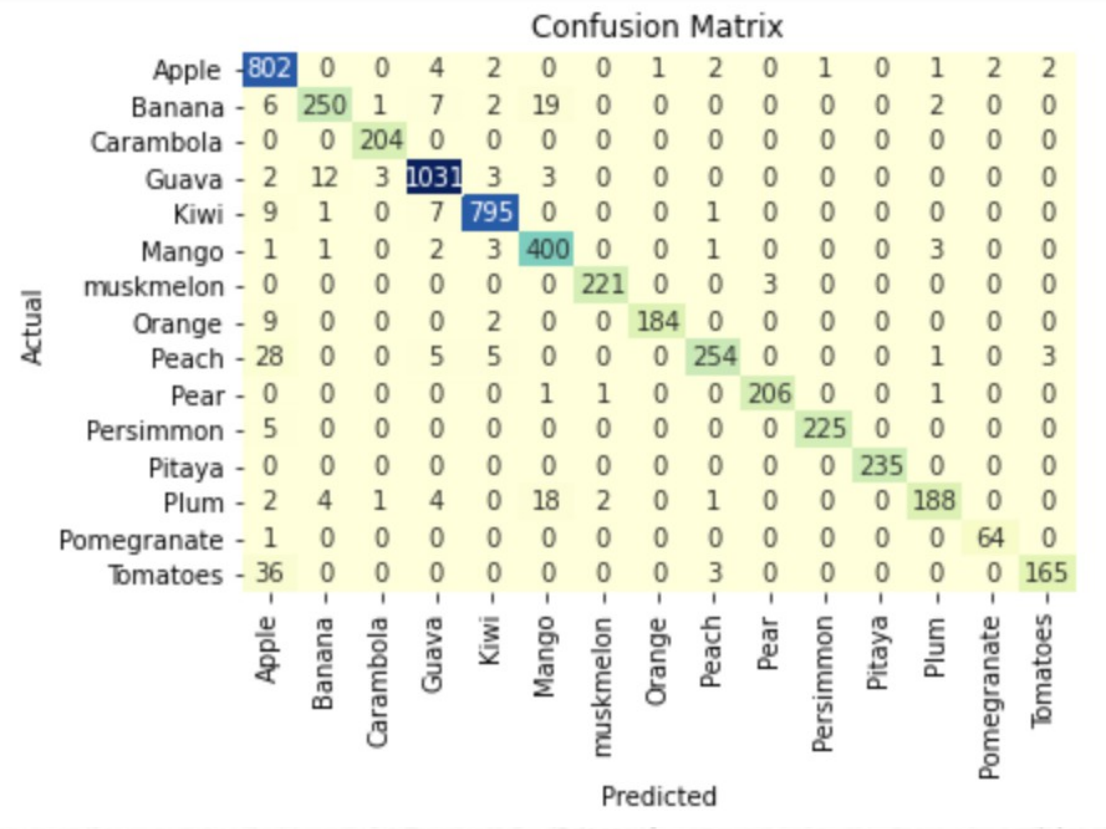
- Base model was trained using 11 epochs
- We used early stopping monitor with patience 2.

Base model performance



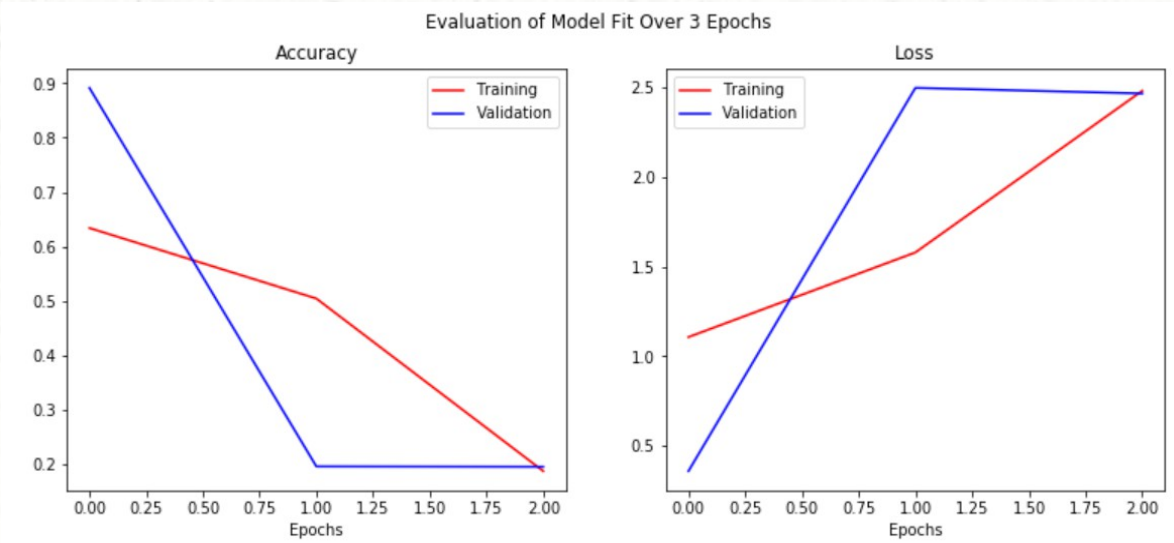

```
history_bsln.csv
Epochs: 11; Final accuracy: 0.829
Epochs: 11; Final val accuracy: 0.953
```

Base model test



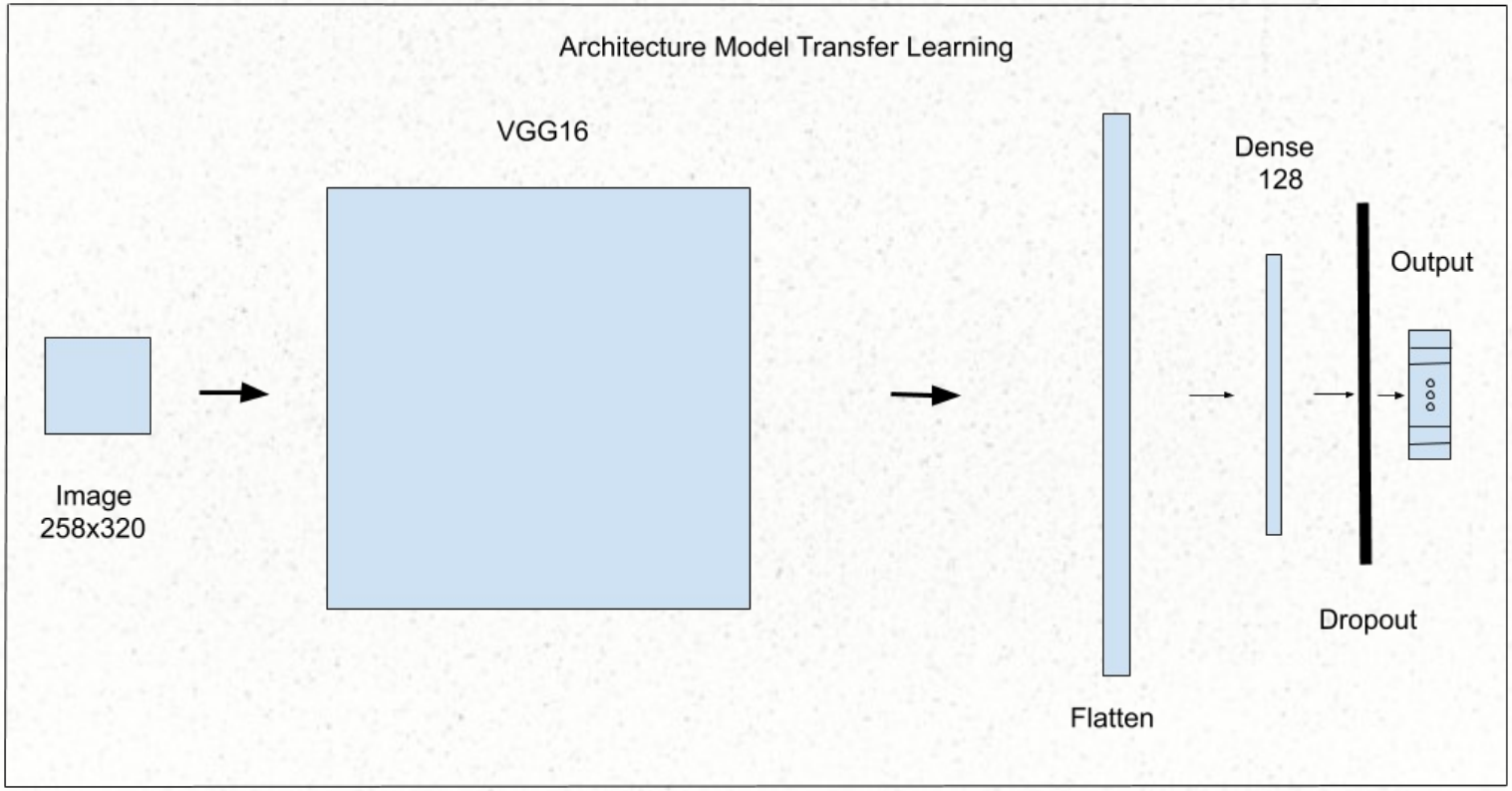
Other optimizers

One of the optimizers did not perform well. Possible reasons – large learning rate or wrong batch size.

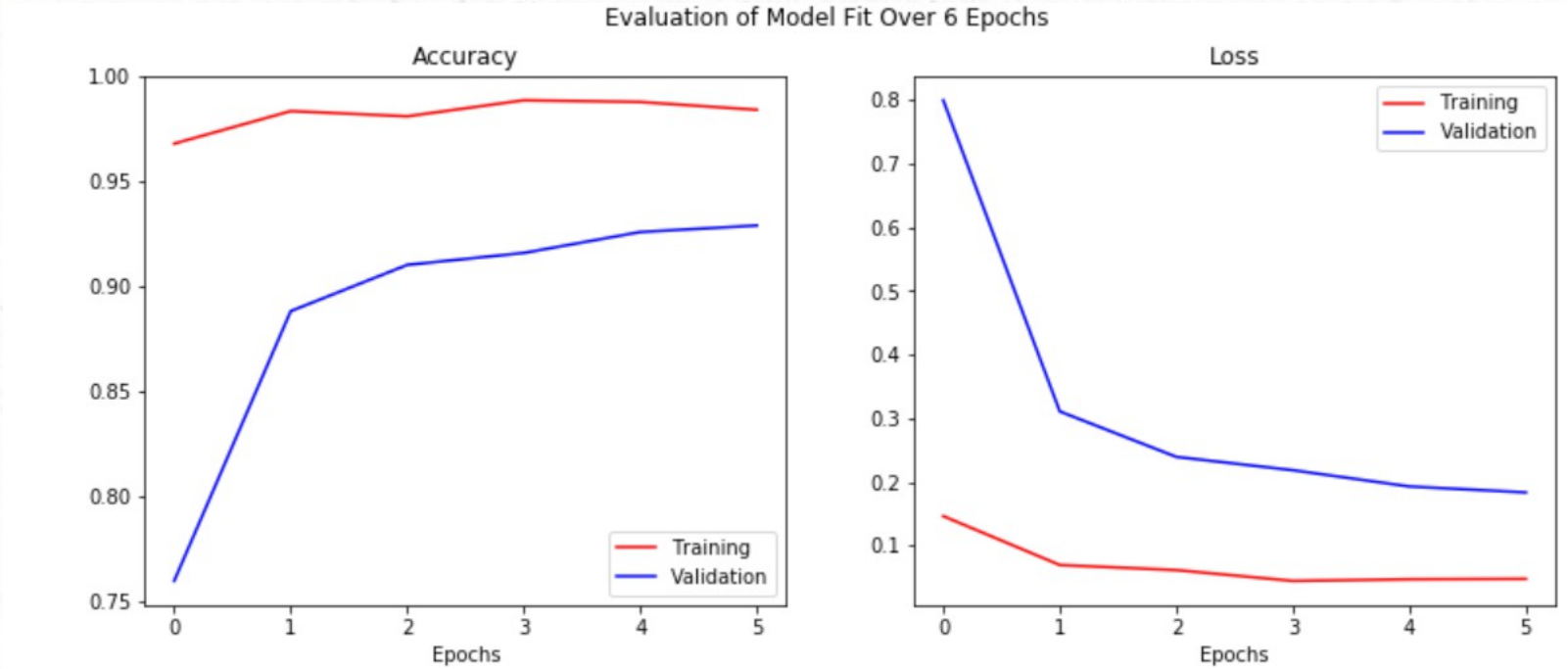



```
history_nadam.csv
Epochs: 3; Final accuracy: 0.187
Epochs: 3; Final val accuracy: 0.195
```

Transfer learning

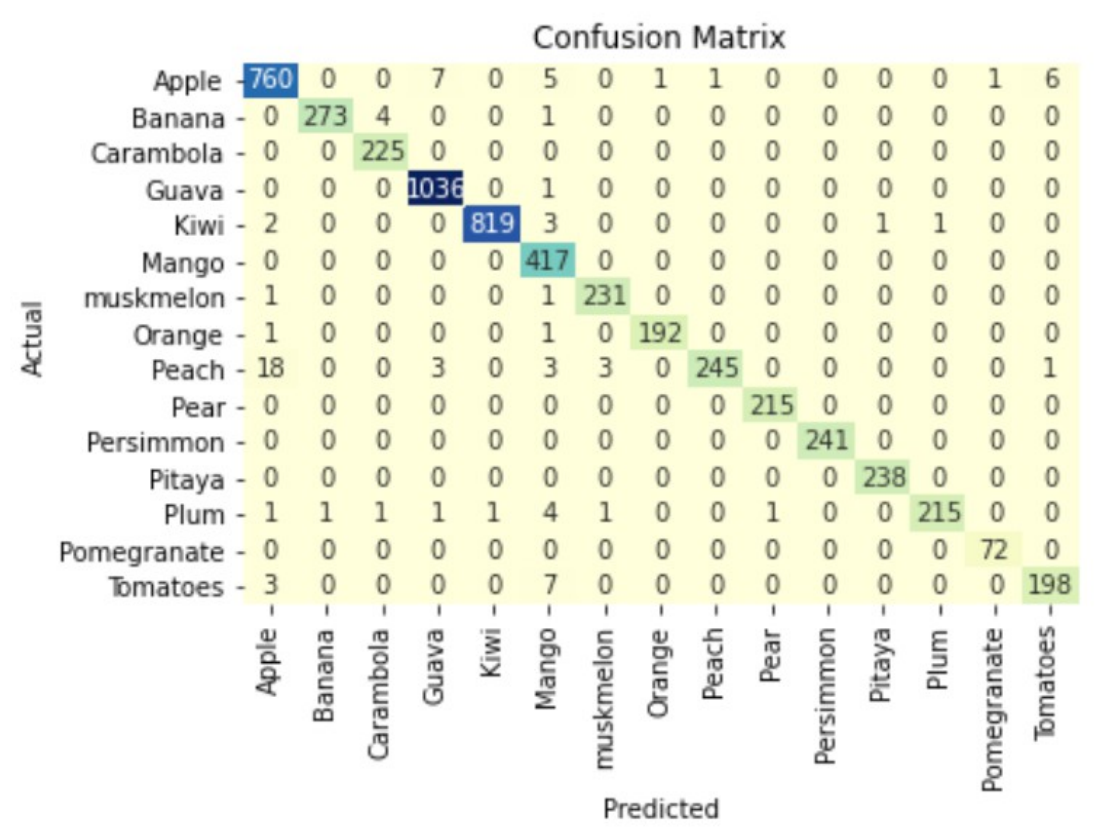


Transfer learning training




```
history_transfer.csv
Epochs: 6; Final accuracy: 0.984
Epochs: 6; Final val accuracy: 0.929
```

Transfer learning - test



Conclusion

Two main approaches were presented here.

- We built classifier from scratch to recognize the name of fruits.
- We used pretrain classifier to recognize the images.

Conclusion

- The pretrained classifier performed better.
- It has higher accuracy
- It was trained in half of the time, even so it has much more complex architecture.

Conclusion

The classifier was trained on about 50000 images. We did not use augmentation and re-scaling.

More images can be collected.

We also can re-train the model, if more types of images are available.