ResNet50

Description

- ResNet50 - 5 upper layers are trainable
- Flatten the extracted feature
- Dense-512 (prediction)
- Activation function - *softmax*
- Loss function - *Categorical Cross Entropy*
- Optimizer - *SGD (Stochastic Gradient Descent)*
- Learning rate - *0.00005*

```python

# base model
restnet = ResNet50(include_top=False, weights='imagenet', input_shape=(IMG_HEIGHT,IMG_WIDTH,3))

# all 170 layers among 175 are non-trainable
# top 5 layers will be trainable
for layer in restnet.layers[:170]:
    layer.trainable = False

base_model = restnet
x = base_model.output


x = Flatten()(x)

# a dense layer for prediction
predictions = Dense(NUM_CLASSES, activation='softmax')(x)

# compile the model
# loss = 'categorical_crossentropy'
# optimizer = SGD
# learning rate = 0.00005
# metrics = 'accuracy'
model_finetuned = Model(inputs=base_model.input, outputs=predictions)
model_finetuned.compile(loss='categorical_crossentropy',
              optimizer=optimizers.SGD(lr=0.00005),
              metrics=['accuracy']
```

Fig 01- Model description. Classifier on top of ResNet50

Model Performance

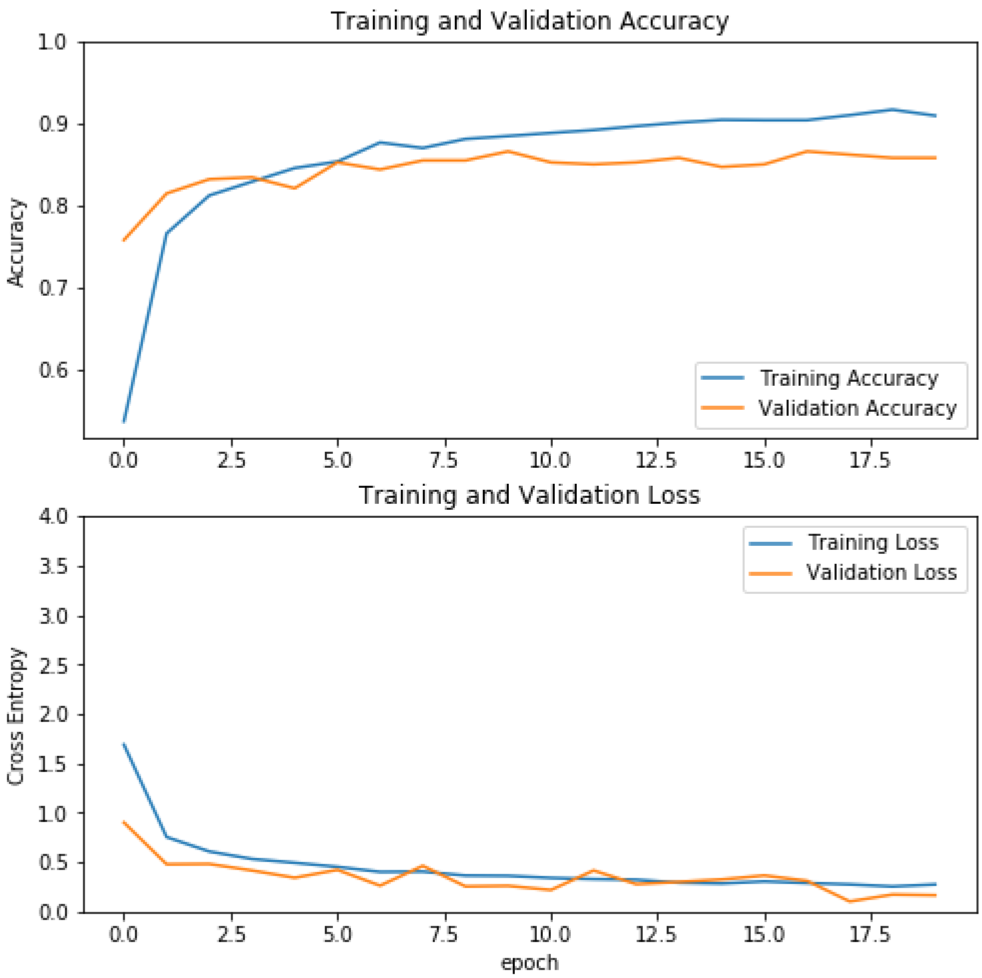

Fig 02- Model training and validation accuracy and losses

Train accuracy - 90%

Train loss - 0.27

Validation accuracy -85%

Validation loss - 0.16

Test accuracy - **86.46% (1367 / 1581)**

Result Analysis

- Total number of classes - 15
- Train Images per-classes
  - Bags = 339
  - Belts = 337
  - Bottomwear = 336
  - Eyewear = 357
  - Flip Flops = 327
  - Fragrance = 339
  - Innerwear = 334
  - Jewellery = 318
  - Lips = 342
  - Sandal = 323
  - Shoes = 336
  - Socks = 344
  - Topwear = 356
  - Wallets = 341
  - Watches = 330
- Validation Images per-classes
  - Bags = 92
  - Belts = 87
  - Bottomwear = 85
  - Eyewear = 73
  - Flip Flops = 84
  - Fragrance = 87
  - Innerwear = 90
  - Jewellery = 85
  - Lips = 83
  - Sandal = 87
  - Shoes = 87
  - Socks = 76
  - Topwear = 76
  - Wallets = 74
  - Watches = 99
- Test Images per-classes
  - Bags = 96
  - Belts = 103
  - Bottomwear = 106
  - Eyewear = 97
  - Flip Flops = 116
  - Fragrance = 101
  - Innerwear = 103
  - Jewellery = 124
  - Lips = 102
  - Sandal = 117
  - Shoes = 104
  - Socks = 107
  - Topwear = 95
  - Wallets = 112
  - Watches = 98
- Wrong prediction per-classes
  - Bags = 18
  - Belts = 2
  - Bottomwear = 13
  - Eyewear = 0
  - Flip Flops = 11
  - Fragrance = 15
  - Innerwear = 14
  - Jewellery = 14
  - Lips = 10
  - Sandal = 50
  - Shoes = 35
  - Socks = 10
  - Topwear = 5
  - Wallets = 10
  - Watches = 7
- Accuracy per-classes (%)
  - Bags = 81.25
  - Belts = 98.05
  - Bottomwear = 87.7
  - Eyewear = 100
  - Flip Flops = 90.51
  - Fragrance = 85.14
  - Innerwear = 86.40
  - Jewellery = 88.70
  - Lips = 90.19
  - Sandal = 57.26
  - Shoes = 66.34
  - Socks = 90.65
  - Topwear = 94.73
  - Wallets = 91.07
  - Watches = 92.85

Model Accuracy Metric

For calculating accuracy, we used the accuracy algorithm that was built in in the model. For accuracy the metric is:

Accuracy = (TP+TN)/(TP+TN+FP+FN)

- True Positive, or TP, are cases with positive labels which have been correctly classified as positive.
- True Negative, or TN, are cases with negative labels which have been correctly classified as negative.
- False Positive, or FP, are cases with negative labels which have been incorrectly classified as positive.
- False Negative, or FN, are cases with positive labels which have been incorrectly classified as negative.

In our case,

- If a T-shirt is correctly classified as a T-shirt, then it is true positive or TP.
- If a Pant is classified as T-shirt, then for T-shirt it is false positive or FP.
- If a Pant is classified as anything other than a T-shirt, then for a T-shirt it is true negative or TN.
- If a T-shirt is classified as anything other than a T-shirt, then for T-shirt it is false negative or FN.

So basically the accuracy we calculated is, the percentage of correctly predicted classes. Mathematically,

Accuracy = correct prediction / total samples.