Assignment No. 6

Aim: Apply transfer learning with pre-trained VGG16 model on assignment 3 and analyze the result.

Objectives:

1. To learn pre-trained models
2. To learn transfer learning

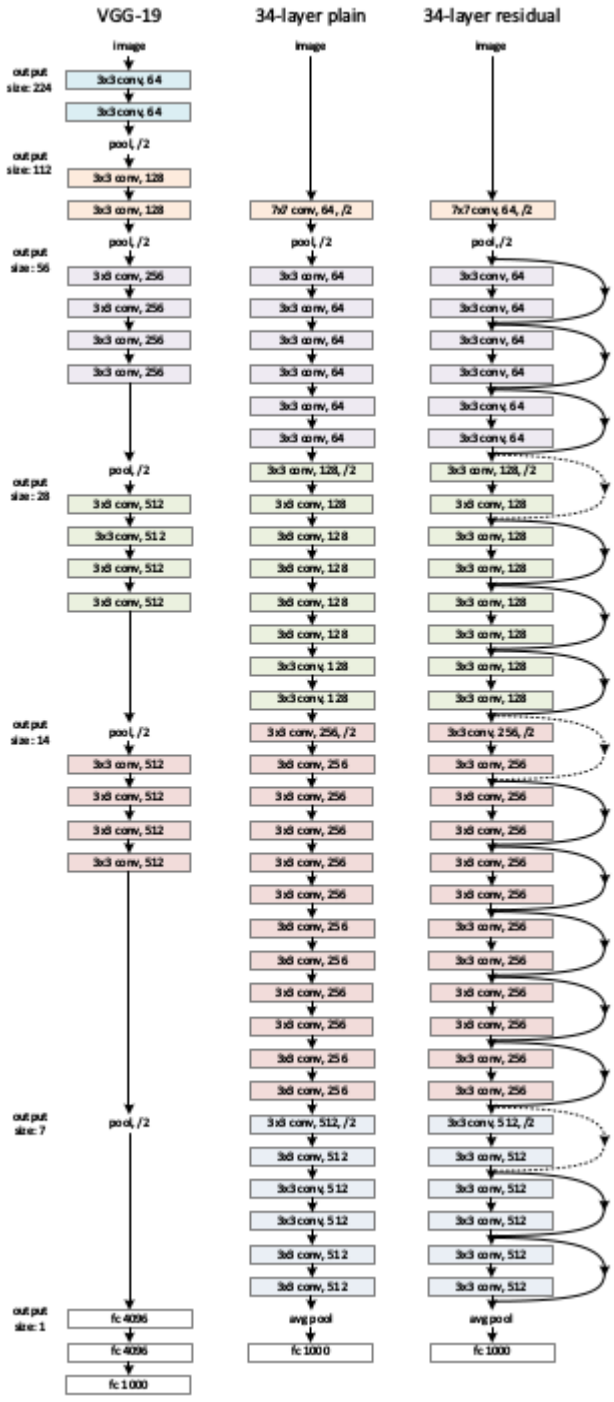
Theory:

Transfer learning

Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although formal ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a reinforcement learning agent.

ResNet 50

1. Use 3*3 filters mostly.
2. Down sampling with CNN layers with stride 2.
3. Global average pooling layer and a 1000-way fully-connected layer with Softmax in the end.



Plain VGG and VGG with Residual Blocks

There are two kinds of residual connections:

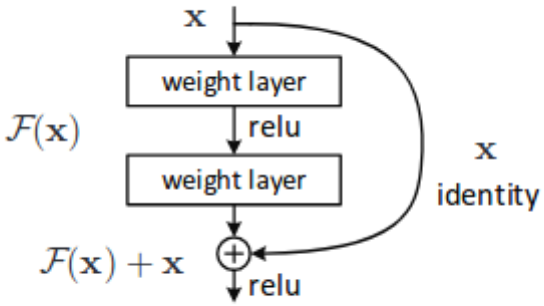


Figure 2. Residual learning: a building block.

Residual block

1. The identity shortcuts (x) can be directly used when the input and output are of the same dimensions.

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}. \qquad (1)$$

Residual block function when input and output dimensions are same

2. When the dimensions change, A) The shortcut still performs identity mapping, with extra zero entries padded with the increased dimension. B) The projection shortcut is used to match the dimension (done by 1*1 conv) using the following formula

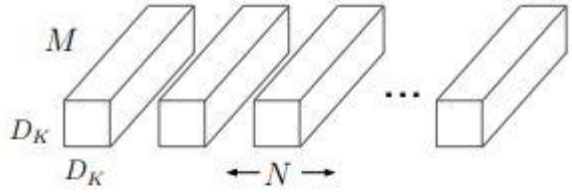
$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s\mathbf{x}. \qquad (2)$$

Residual block function when the input and output dimensions are not same.

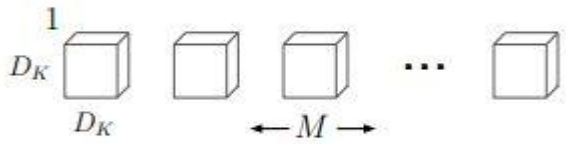
The first case adds no extra parameters, the second one adds in the form of W_{s}

MOBILENET

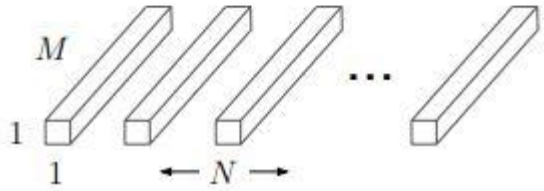
MobileNet is an efficient and portable CNN architecture that is used in real world applications. MobileNets primarily use **depthwise seperable convolutions** in place of the standard convolutions used in earlier architectures to build lighter models.MobileNets introduce two new global hyperparameters(width multiplier and resolution multiplier) that allow model developers to trade off **latency** or **accuracy** for speed and low size depending on their requirements.



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



Architecture

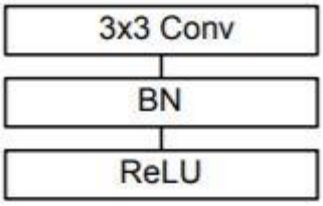
MobileNets are built on depthwise seperable convolution layers.Each depthwise seperable convolution layer consists of a depthwise convolution and a pointwise convolution.Counting depthwise and pointwise convolutions as seperate layers, a MobileNet has 28 layers.A standard MobileNet has 4.2 million parameters which can be further reduced by tuning the width multiplier hyperparameter appropriately.
The size of the input image is 224 × 224 × 3.

The detailed architecture of a MobileNet is given below :

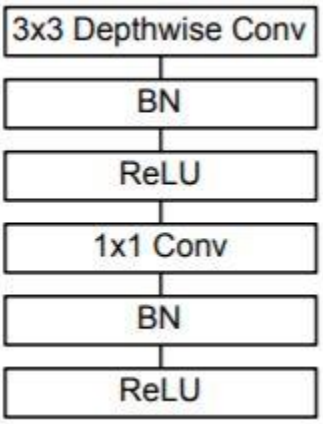
TYPE	STRIDE	KERNEL SHAPE	INPUT SIZE
Conv	2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw	1	3 × 3 × 32	112 × 112 × 32
Conv	1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw	2	3 × 3 × 64	112 × 112 × 64
Conv	1	1 × 1 × 64 × 128	56 × 56 × 64

TYPE	STRIDE	KERNEL SHAPE	INPUT SIZE
Conv dw	1	3 × 3 × 128	56 × 56 × 128
Conv	1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw	2	3 × 3 × 128	56 × 56 × 128
Conv	1	1 × 1 × 128 × 256	56 × 56 × 128
Conv dw	1	3 × 3 × 256	28 × 28 × 256
Conv	1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw	2	3 × 3 × 256	28 × 28 × 256
Conv	1	1 × 1 × 256 × 512	14 × 14 × 256
Conv dw	1	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw	1	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw	1	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw	1	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw	1	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw	2	3 × 3 × 512	14 × 14 × 512
Conv	1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw	2	3 × 3 × 1024	7 × 7 × 1024
Conv	1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool	1	Pool 7 × 7	7 × 7 × 1024
Fully Connected	1	1024 × 1000	1 × 1 × 1024
Softmax	1	Classifier	1 × 1 × 1000

Standard Convoltion layer :
A single standard convolution unit (denoted by **Conv** in the table above) looks like this :

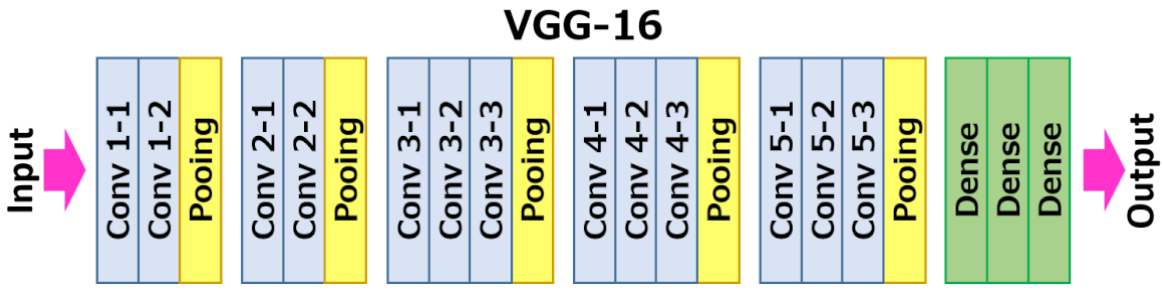


Depthwise seperable Convolution layer
A single Depthwise seperable convolution unit (denoted by **Conv dw** in the table above) looks like this :

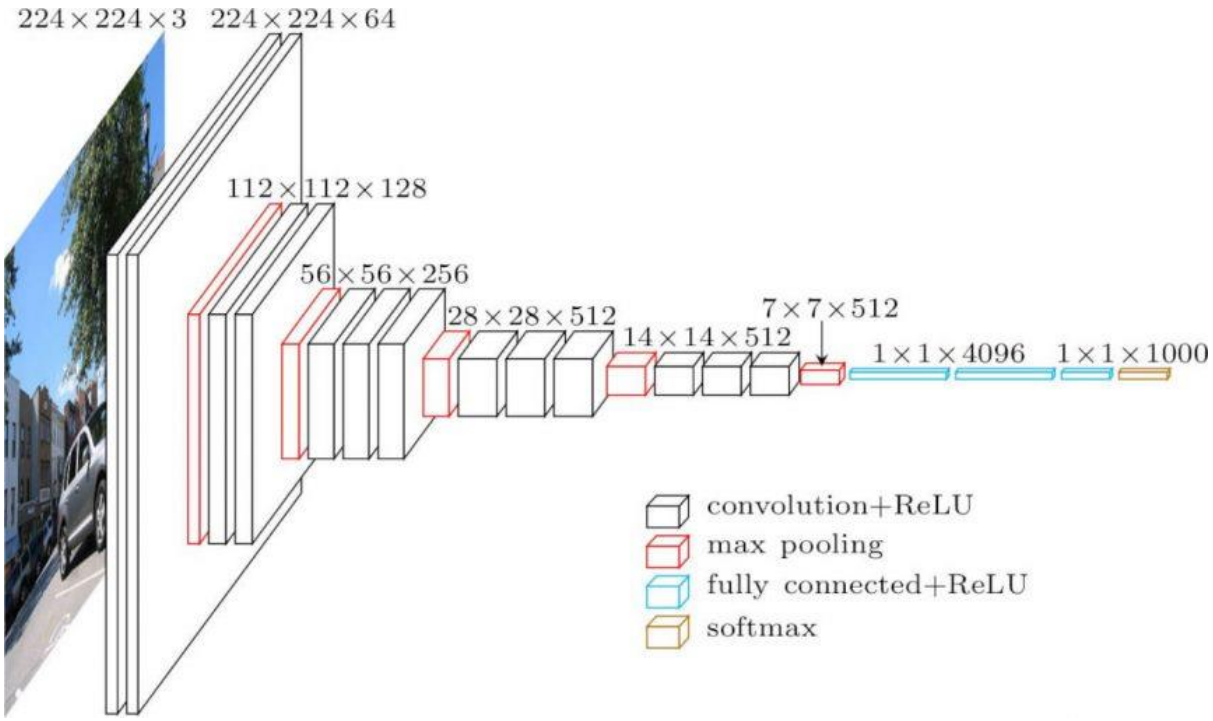


VGG16

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU’s.



The architecture depicted below is VGG16.



VGG16 Architecture

The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

Code:

```
import pandas as pd
import tensorflow as tf
from tensorflow.keras import models,Sequential,layers,preprocessing
import os
from tensorflow.keras.applications.vgg16 import VGG16

file_names=os.listdir("/content/train")
dogorcat=[]
for name in file_names:
```

```
    category=name.split('.')[0]
    if category=='dog':
        dogorcat.append("DOG")
    else:
        dogorcat.append("CAT")
train=pd.DataFrame({
    'filename':file_names,
    'category':dogorcat
})

base=VGG16(include_top=False,input_shape=(224,224,3),weights='imagenet')
base.trainable=False
model=models.Sequential()
model.add(base)
model.add(layers.Flatten())
model.add(layers.Dense(120,activation="relu"))
model.add(layers.Dense(2,activation="softmax"))

model.compile(optimizer="adam",loss='categorical_crossentropy',metrics=['accur
acy'])

from sklearn.model_selection import train_test_split
from keras.preprocessing.image import ImageDataGenerator,load_img
train_df,validate_df = train_test_split(train,test_size=0.1, random_state=42)
train_df = train_df.reset_index(drop=True)
validate_df = validate_df.reset_index(drop=True)
training = preprocessing.image.ImageDataGenerator(rotation_range=15, rescale=1
./255, shear_range=0.1, zoom_range=0.2, horizontal_flip=True, width_shift_rang
e=0.1, height_shift_range=0.1)
total_train=train_df.shape[0]
total_validate=validate_df.shape[0]

trainingdata = training.flow_from_dataframe(train_df,"/content/train",x_col='f
ilename',y_col='category',target_size=(224,224),class_mode='categorical',batch
_size=4)
validation = ImageDataGenerator(rescale=1./255)
validationdata = validation.flow_from_dataframe(validate_df,"/content/train",
x_col='filename',y_col='category',target_size=(224,224),class_mode='categorica
l',batch_size=4)

model.fit(trainingdata,validation_data=validationdata,epochs=10,steps_per_epoc
h=total_train//200,validation_steps=total_validate//200)

model.evaluate(validationdata)
```

Results:

```
model.fit(trainingdata,validation_data=validationdata,epochs=10,steps_per_epoch=total_train//200,validation_steps=total_validate//200)

Epoch 1/10
112/112 [==============================] - 256s 2s/step - loss: 0.7865 - accuracy: 0.7366 - val_loss: 0.8776 - val_accuracy: 0.6667
Epoch 2/10
112/112 [==============================] - 257s 2s/step - loss: 0.5440 - accuracy: 0.7723 - val_loss: 0.1737 - val_accuracy: 0.9167
Epoch 3/10
112/112 [==============================] - 255s 2s/step - loss: 0.4249 - accuracy: 0.8393 - val_loss: 0.5555 - val_accuracy: 0.8125
Epoch 4/10
112/112 [==============================] - 255s 2s/step - loss: 0.3796 - accuracy: 0.8549 - val_loss: 0.3619 - val_accuracy: 0.8125
Epoch 5/10
112/112 [==============================] - 252s 2s/step - loss: 0.3398 - accuracy: 0.8527 - val_loss: 0.2243 - val_accuracy: 0.9375
Epoch 6/10
112/112 [==============================] - 254s 2s/step - loss: 0.3460 - accuracy: 0.8482 - val_loss: 0.1937 - val_accuracy: 0.9167
Epoch 7/10
112/112 [==============================] - 254s 2s/step - loss: 0.3037 - accuracy: 0.8661 - val_loss: 0.2235 - val_accuracy: 0.8750
Epoch 8/10
112/112 [==============================] - 255s 2s/step - loss: 0.2772 - accuracy: 0.8884 - val_loss: 0.2347 - val_accuracy: 0.8958
Epoch 9/10
112/112 [==============================] - 255s 2s/step - loss: 0.3160 - accuracy: 0.8728 - val_loss: 0.2730 - val_accuracy: 0.8958
Epoch 10/10
112/112 [==============================] - 253s 2s/step - loss: 0.3103 - accuracy: 0.8728 - val_loss: 0.4754 - val_accuracy: 0.7917
<tensorflow.python.keras.callbacks.History at 0x7fcd9d05cbe0>

[ ] model.evaluate(validationdata)

625/625 [==============================] - 1250s 2s/step - loss: 0.3219 - accuracy: 0.8564
[0.3218751549720764, 0.8564000129699707]
```

Conclusion:

Thus, we have understood the importance of Transfer learning and also the usage of transfer learning in tensorflow