Caltech-256

By http://places2.csail.mit.edu/PAMI_places (http://places2.csail.mit.edu/PAMI_places) (given by Dr.yang)and reference code ,learning the method to generate the training and testing dataset.

In [ ]:
```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import tarfile
import cv2
import pickle
%matplotlib inline
```

In [ ]:
```
path = '/content/drive/My Drive/Colab Notebooks/VGG16_Practice/256_ObjectCa
```

In [ ]:
```
os.chdir(path)
```

In [ ]:
```
folders = os.listdir()
```

In [ ]:
```
folder_paths = []
all_images = []
all_classes = []
```

In [ ]:
```
img_size = 128
```

Read the pictures in each class

In [ ]:
```
from PIL import Image

def make_square(image, min_size=img_size, fill_color=(0, 0, 0, 0)):
    size = (min_size, min_size)
    image.thumbnail(size, Image.ANTIALIAS)
    background = Image.new('RGB', size, (255, 255, 255, 0))
    background.paste(
        image, (int((size[0] - image.size[0]) / 2), int((size[1] - image.si
    )

    new_img = np.array(background)
    new_img.flatten()
    return new_img
```

In [ ]:
```
for folder in range(len(folders)):
    folder_paths = path+str(folders[folder])+str('/')

    os.chdir(folder_paths)
    image_in_folder = os.listdir()

    for image in range(len(image_in_folder)):
        img = Image.open(image_in_folder[image])
        img = make_square(img)

        all_images.append(img.flatten()/255)
        all_classes.append(folders[folder])
```

Reading raw data into a pickle file, Divide data such that it is less than 4GB, the advantage of this is that after the test, you no longer need to read the data from the original photo, but directly load pickle file, and then get the train Data and test data

Pickel

In [ ]:
```
import sys
sys.getsizeof(all_images_df)
```

In [ ]:
```
all_images_df1 = all_images_df[:10000,:]
all_images_df2 = all_images_df[10000:20000,:]
all_images_df3 = all_images_df[20000:,:]
print('all_images_df1:'+str(sys.getsizeof(all_images_df1)))
print('all_images_df2:'+str(sys.getsizeof(all_images_df2)))
print('all_images_df3:'+str(sys.getsizeof(all_images_df3)))
```

Save pickel data，Save the pickle file, then test it, you can get the train data and test data directly from the pickle file.

In [ ]:
```

```

In [ ]:
```
picklepath = 'G:/Caltech256/'
os.chdir(picklepath)
```

In [ ]:
```
import pickle

pickle_out = open("pickle_all_images_df1.pickle","wb")
pickle.dump(all_images_df1, pickle_out)
pickle_out.close()

pickle_out = open("pickle_all_images_df2.pickle","wb")
pickle.dump(all_images_df2, pickle_out)
pickle_out.close()

pickle_out = open("pickle_all_images_df3.pickle","wb")
pickle.dump(all_images_df3, pickle_out)
pickle_out.close()

pickle_out = open("pickle_all_classes.pickle","wb")
pickle.dump(all_classes, pickle_out)
pickle_out.close()
```

In [ ]:
```

```