TypeB_AntsBees_VGG16_PyTorch_DataSet_02

January 6, 2022

1 Binary classification model : AntsBees

2 Coding Stye: TypeB

3 Section: DataSet Class

[1]:
```
%pwd
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).
```

[2]:
```
import numpy as np
from PIL import Image
from torchvision import transforms
import matplotlib.pyplot as plt
import os
import torch
import torch.utils.data as data
from pathlib import Path
```

[3]:
```
data_dir = '/content/gdrive/My Drive/Colab Notebooks/AntsBees/data'
root_dir= '/content/gdrive/My Drive/Colab Notebooks/AntsBees/data/
 ↪hymenoptera_data'
```

[4]:
```
os.chdir('/content/gdrive/My Drive/Colab Notebooks/AntsBees')
```

4 *DataSet Class*

4.0.1 1ImageTransform Class

[5]:
```
from util.ImageTransform import ImageTransform
```

4.0.2 (2) make_path_list function

4.0.3 train_list, val_list,file_list

[6]:

```
from dsets.dsets import make_path_list
```

[7]:

```
import pprint
#
train_list = make_path_list(phase='train',root_dir=root_dir)
val_list = make_path_list(phase='val',root_dir=root_dir)

file_list={'train':train_list,'val':val_list}

#print(train_list)
print(len(train_list))

# 5
print('train')
print(train_list[:2])
pprint.pprint(train_list[-2:-1])
# 5
print('val')
print(val_list[:2])
pprint.pprint(val_list[-3:-1])
```

```
242
train
[CandidateInfoTuple(label=0, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/train/ants/1030023514_aad5c608f9.jpg'),
CandidateInfoTuple(label=0, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/train/ants/0013035.jpg')]
[CandidateInfoTuple(label=1, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/train/bees/2634617358_f32fd16bea.jpg')]
val
[CandidateInfoTuple(label=0, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/val/ants/10308379_1b6c72e180.jpg'),
CandidateInfoTuple(label=0, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/val/ants/1053149811_f62a3410d3.jpg')]
[CandidateInfoTuple(label=1, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/val/bees/759745145_e8bc776ec8.jpg'),
 CandidateInfoTuple(label=1, path='/content/gdrive/My Drive/Colab
Notebooks/AntsBees/data/hymenoptera_data/val/bees/936182217_c4caa5222d.jpg')]
```

4.0.4 create train_dataset, val_dataset

4.0.5 3MakeBalancedDataset Class

[8]:

```
from dsets.dsets import MakeBalancedDataset
```

[10]:

```
#
# ()
SIZE = 224
# RGB
MEAN = (0.485, 0.456, 0.406) # ImageNet
# RGB
STD = (0.229, 0.224, 0.225)  # ImageNet


#
size, mean, std = SIZE, MEAN, STD

# MakeDataset
train_dataset = MakeBalancedDataset(
    file_list=file_list, #
    ratio_int=True,
    transform=ImageTransform(size, mean, std), #
    phase='train',records=300)
# MakeDataset
val_dataset = MakeBalancedDataset(
    file_list=file_list, #
    ratio_int=True,
    transform=ImageTransform(size, mean, std), #
    phase='val',records=200)
```

[11]:

```
print(len(train_dataset))
print(len(val_dataset))
```

```
300
200
```

4.0.6 4dataloader

[24]:

```
from torch.utils.data import DataLoader

#
batch_size = 32

# :(, 3, 224, 224)
train_dl = DataLoader(train_dataset, batch_size=batch_size, shuffle=True)

# :(, 3, 224, 224)
val_dl = DataLoader(val_dataset, batch_size=batch_size, shuffle=False)
```

[25]:

```
print(len(train_dl))
print(len(train_dl.dataset))
```

```
10
300
```

[26]:

```
print(len(val_dl))
print(len(val_dl.dataset))
```

```
7
200
```

4.0.7 check

[27]:

```
batch_iterator = iter(train_dl)
images, target = next(batch_iterator)

print(images.size())
print(target)
```

```
torch.Size([32, 3, 224, 224])
tensor([0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0,
        0, 1, 0, 1, 1, 1, 1, 1])
```

[29]:

```
batch_iterator = iter(val_dl)
images, target = next(batch_iterator)

print(images.size())
print(images.shape)
print(target)
```

```
torch.Size([32, 3, 224, 224])
torch.Size([32, 3, 224, 224])
tensor([1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
        1, 0, 1, 0, 1, 0, 1, 0])
```

4.0.8 5For Loop with DataLoader

[18]:

```
import sys
for batch_ndx, batch_tup in enumerate(train_dl):
  input_t, label_t = batch_tup
  print(batch_ndx,input_t.shape)
  #print(label_t)
  var_name = 'input_t'
  print(sys.getsizeof(eval(var_name)))
```

```
0 torch.Size([32, 3, 224, 224])
88
1 torch.Size([32, 3, 224, 224])
88
2 torch.Size([32, 3, 224, 224])
88
3 torch.Size([32, 3, 224, 224])
88
4 torch.Size([32, 3, 224, 224])
88
5 torch.Size([32, 3, 224, 224])
88
6 torch.Size([32, 3, 224, 224])
88
7 torch.Size([32, 3, 224, 224])
88
8 torch.Size([32, 3, 224, 224])
88
9 torch.Size([12, 3, 224, 224])
88
```

[22]:

```
import sys
for batch_ndx, batch_tup in enumerate(val_dl):
  input_t, label_t = batch_tup
  print(batch_ndx,input_t.shape)
  print(label_t)
  var_name = 'input_t'
  print(sys.getsizeof(eval(var_name)))
```

```
0 torch.Size([32, 3, 224, 224])
88
1 torch.Size([32, 3, 224, 224])
88
2 torch.Size([32, 3, 224, 224])
88
3 torch.Size([32, 3, 224, 224])
88
4 torch.Size([32, 3, 224, 224])
88
5 torch.Size([32, 3, 224, 224])
88
6 torch.Size([8, 3, 224, 224])
88
```

5 END