

## Convert Pascal VOC formatted datasets to COCO datasets

### 1. Collect Pascal VOC datasets (down tập dữ liệu Pascal VOC)

Ở đây e sử dụng tập hình ảnh và chú thích ở link này:

<https://github.com/howl0893/custom-object-detection-datasets>

### 2. Create class labels.txt file: (Tạo lớp tập label.txt)

Tạo một tập văn bản mới và nhập từng tên lớp trong tập dữ liệu được phân tách bằng dấu cách hoặc dòng mới. Trong trường hợp này e dùng tập card trong file dataset, tạo một tập văn bản mới có tên 'label.txt' để chứa:

*Ace*  
*King*  
*Queen*  
*Jack*  
*Ten*  
*Nine*

### 3. Create test.txt and train.txt annotation files(Tạo tập chú thích test.txt và train.txt):

- Mở test\_labels.csv và train\_labels.csv
- Tạo hai tập văn bản mới có tên là 'test-ext.txt' và 'train-ext.txt'
- Sao chép nội dung của cột đầu tiên trong mỗi tập csv bỏ qua hàng đầu tiên(cope từ excel vào bỏ dòng file name đi) và dán vào tập văn bản tương ứng( ở đây e dán vào notetap)



```

1 # written by: Matthew Howlett
2 # date: September 24, 2019
3 # python script for editing text files_in
4
5 files_in = ['C:/Users/Matt/Desktop/custom-object-detection-datasets/datasets/cards/test-ext.txt']
6 files_out = ['C:/Users/Matt/Desktop/custom-object-detection-datasets/datasets/cards/test.txt']
7
8 # clean up file format
9 for i in range(len(files_in)):
10     with open(files_in[i], 'r') as file_in, \
11         open(files_out[i], 'w') as file_out:
12         data = file_in.read()
13         data = data.replace(".", "")
14         data = data.replace("jpg", "")
15         data = data.replace("JPG", "\t")
16         file_out.write(data)
17
18 # import os
19
20 # directory = 'C:/Users/Matt/OneDrive/Documents/Kettering/courses/summer-2019/ce-senior-design/creating-
21
22 # for filename in os.listdir(directory):
23 #     if filename.endswith(".xml"):
24 #         os.remove(directory + filename)
25 #     else:
26 #         continue
27

```

(Đổi tên ở phần khoanh tròn,

```

1 # written by: Matthew Howlett
2 # date: September 24, 2019
3 # python script for editing text files_in
4
5 files_in = ['C:/Users/Matt/Desktop/custom-object-detection-datasets/datasets/cards/train-ext.txt']
6 files_out = ['C:/Users/Matt/Desktop/custom-object-detection-datasets/datasets/cards/train.txt']
7
8 # clean up file format
9 for i in range(len(files_in)):
10     with open(files_in[i], 'r') as file_in, \
11         open(files_out[i], 'w') as file_out:
12         data = file_in.read()
13         data = data.replace(".", "")
14         data = data.replace("jpg", "")
15         data = data.replace("JPG", "\t")
16         file_out.write(data)
17
18 # import os
19
20 # directory = 'C:/Users/Matt/OneDrive/Documents/Kettering/courses/summer-2019/ce-senior-design/creating-kitti-datasets/cocosynth-master/datasets/car
21
22 # for filename in os.listdir(directory):
23 #     if filename.endswith(".xml"):
24 #         os.remove(directory + filename)
25 #     else:
26 #         continue
27

```

```

Anaconda Prompt
(chase) C:\Users\Matt>cd desktop\custom-object-detection-datasets\python
(chase) C:\Users\Matt\Desktop\custom-object-detection-datasets\python>python clean-up.py
(chase) C:\Users\Matt\Desktop\custom-object-detection-datasets\python>python clean-up.py
(chase) C:\Users\Matt\Desktop\custom-object-detection-datasets\python>_

```

4. Run voc2coco.py script(Chạy tập lệnh voc2coco.py):

Ở đây e dùng đoạn source code của cái này:

<https://github.com/yukkyo/voc2coco>

sau đó e mở Anaconda để prompt vs enter lệnh này

(Đoạn code này e sửa khai báo đường dẫn trong máy e, đoạn code gốc ở link a)

```
cd desktop/custom-object-detection-datasets/python
```

```
python voc2coco.py
```

```
--ann_dir ../datasets/cards/images/test (cái này e chỉ đường dẫn)
```

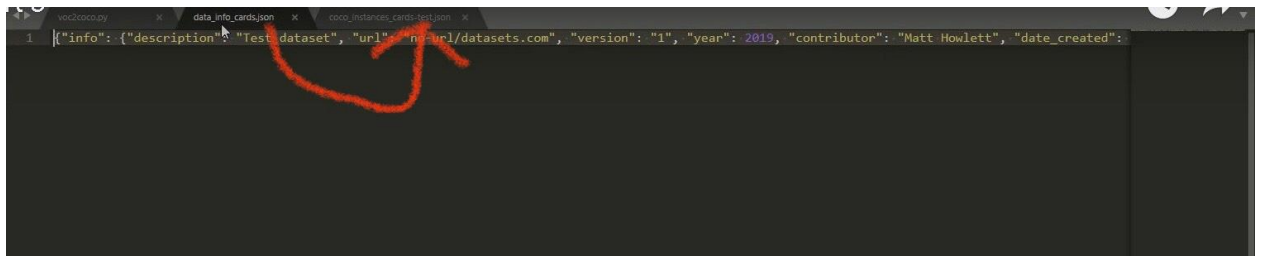
```
--ann_ids ../datasets/cards/test.txt
```

```
--labels ../datasets/cards/labels.txt
```

```
--output ../datasets/cards/annotations/coco_instances_cards-  
test.json (đuôi ra là đuôi.json)
```

```
--ext xml
```

\*Làm điều này cho tập hợp hình ảnh và tệp xml. Sau khi hoàn tất, hãy mở tệp data\_info\_cards.json, sao chép toàn bộ nội dung của nó và dán vào đầu của cả hai tệp coco\_instance.



*\*Copy hết r dán lên đầu file bên kia, nhớ có khoảng cách*

\***LƯU Ý:** File train đổi giống file **test** chỉ cần thay **test** bằng train là xong