# Selected Topics in Visual Recognition using Deep Learning Homework 2 announcement

TA: 楊証琨, Jimmy

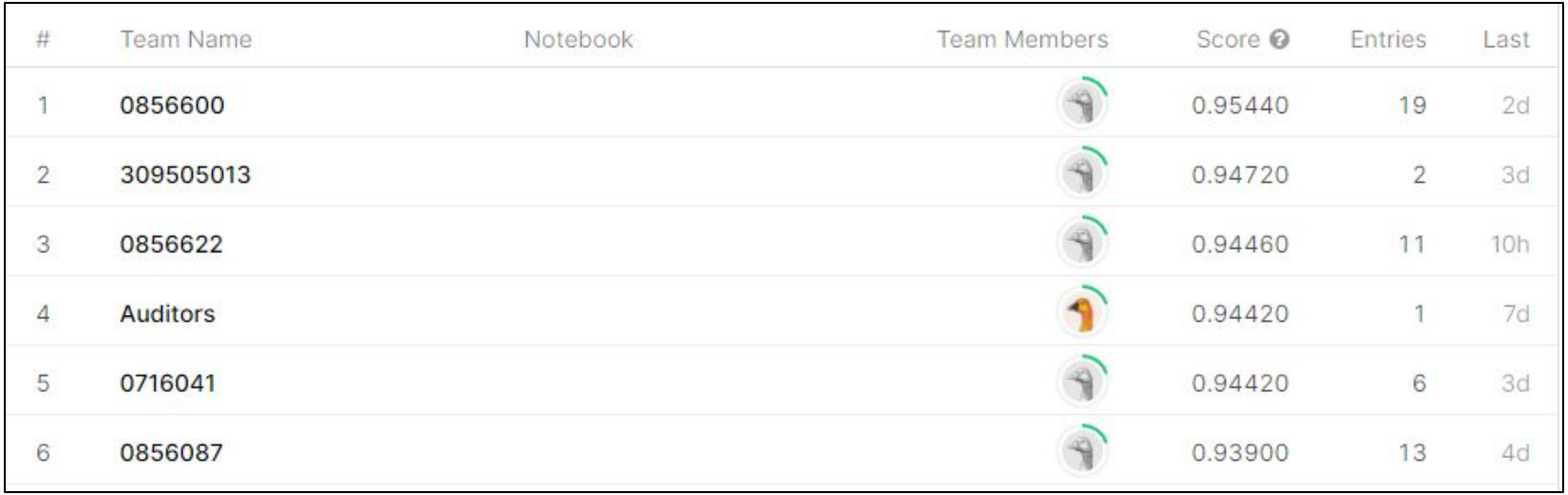
Ph.D. student at National Taiwan Universitiy

d08922002@ntu.edu.tw

# Homework 1 reminder!

- Deadline: 11/12, Thr at 23:59

1. Finish the Kaggle competition (check the leaderboard)

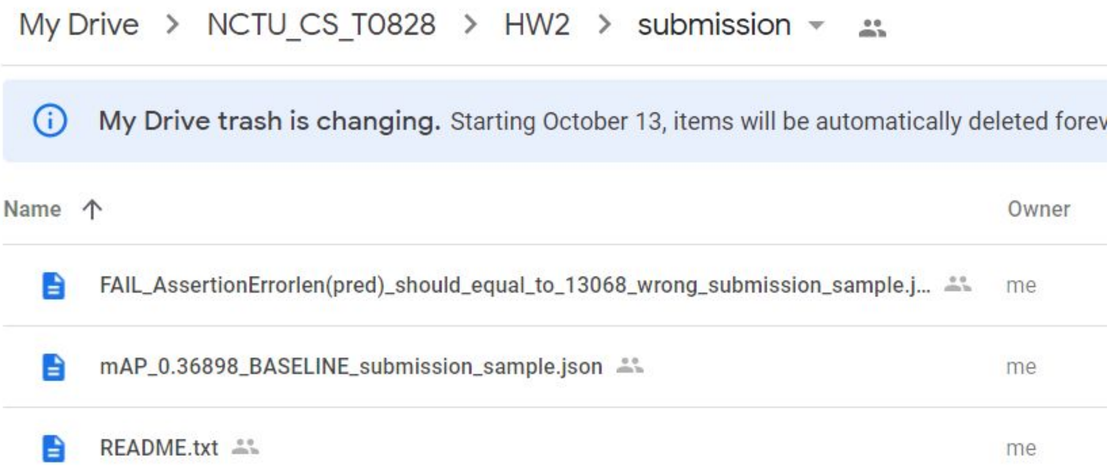


| # | Team Name | Notebook | Team Members | Score ? | Entries | Last |
|---|---|---|---|---|---|---|
| 1 | 0856600 | | | 0.95440 | 19 | 2d |
| 2 | 309505013 | | | 0.94720 | 2 | 3d |
| 3 | 0856622 | | | 0.94460 | 11 | 10h |
| 4 | Auditors | | | 0.94420 | 1 | 7d |
| 5 | 0716041 | | | 0.94420 | 6 | 3d |
| 6 | 0856087 | | | 0.93900 | 13 | 4d |

2. Upload your reports to **E3 systems**

# Homework 2: Digits detection

- Deadline: 11/26, Thr at 23:59
  1. Upload your **report.pdf** to E3
  2. **Submit prediction file** to this Google drive
  3. Test your model inference speed by the Google colab



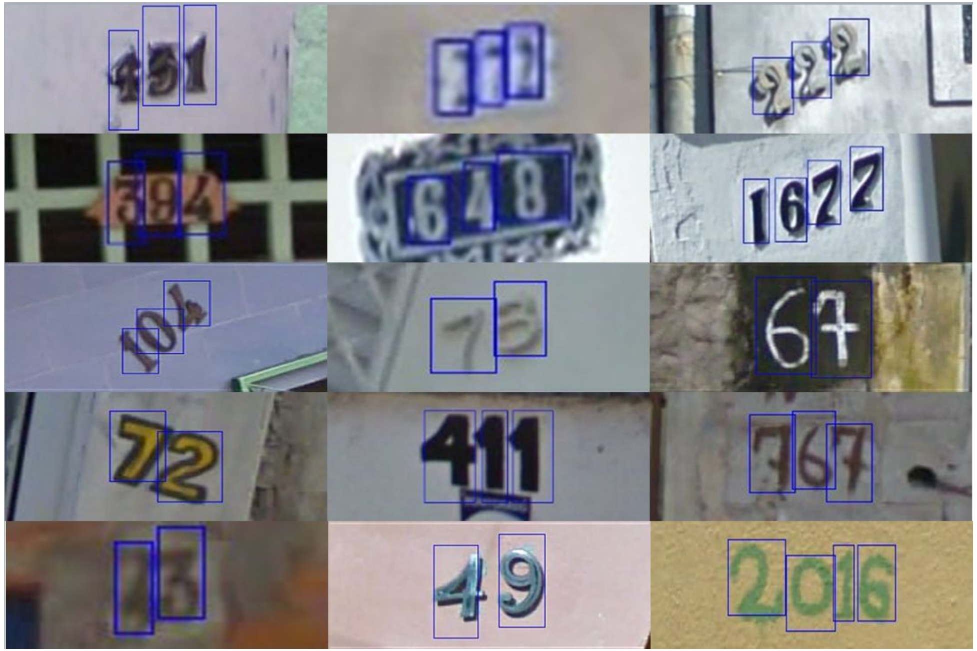
My Drive > NCTU_CS_T0828 > HW2 > submission

**My Drive trash is changing.** Starting October 13, items will be automatically deleted forev

| Name ↑ | Owner |
|---|---|
| FAIL_AssertionErrorlen(pred)_should_equal_to_13068_wrong_submission_sample.j... | me |
| mAP_0.36898_BASELINE_submission_sample.json | me |
| README.txt | me |

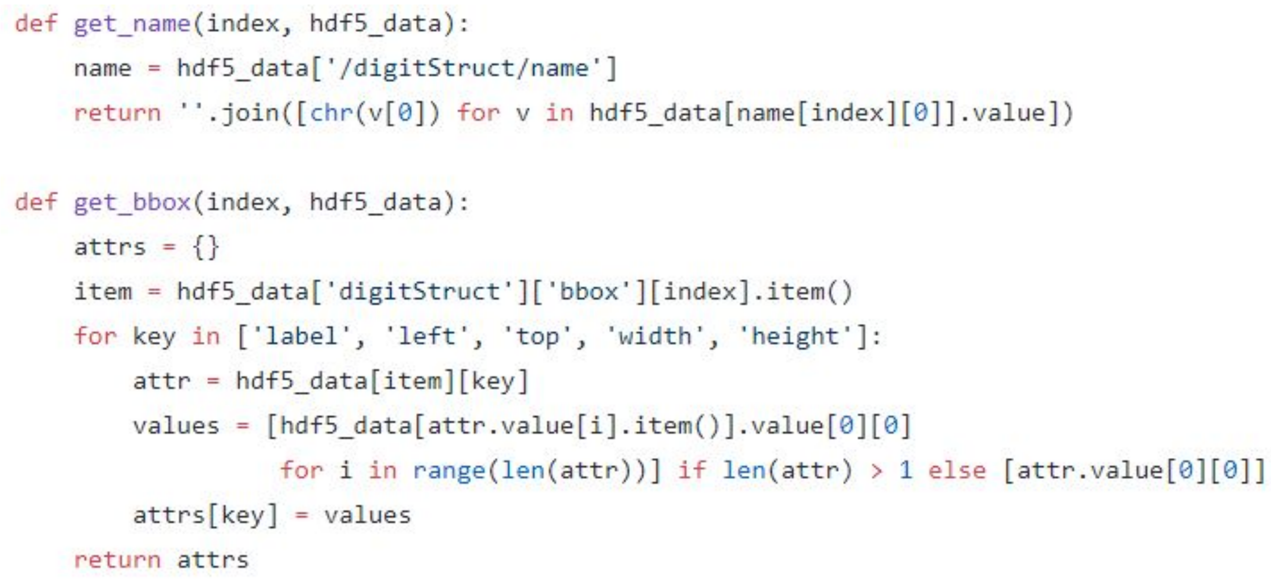
# **HW3 Introduction:** Street View House Numbers

- SVHN dataset contains 33,402 trianing images, 13,068 test images
- Train a not only **accurate** but **fast** digit detector!



# HW3 Get the dataset

- Download the dataset from this Google Drive
- The annotations are save in **.h5** file. Here are Python code to parse the annotation file

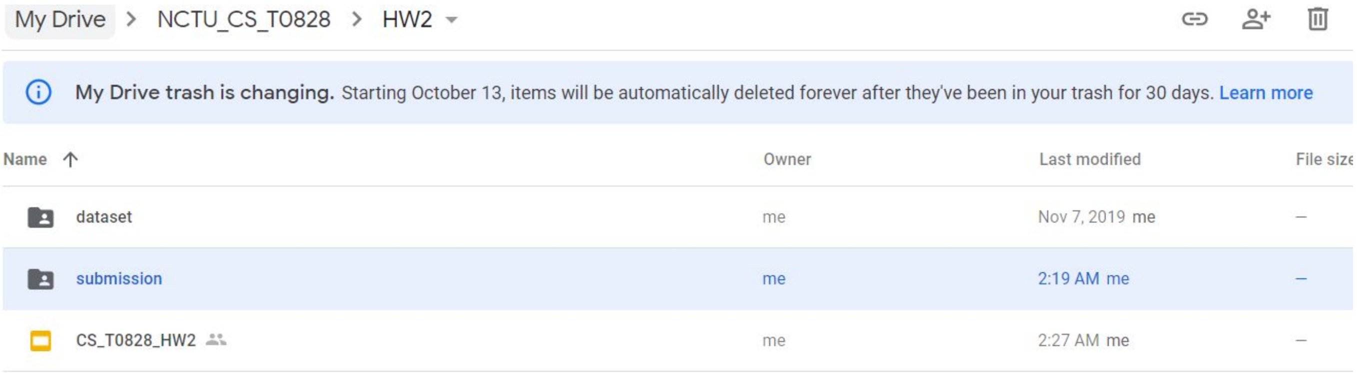


```python
def get_name(index, hdf5_data):
    name = hdf5_data['/digitStruct/name']
    return ''.join([chr(v[0]) for v in hdf5_data[name[index][0]].value])

def get_bbox(index, hdf5_data):
    attrs = {}
    item = hdf5_data['digitStruct']['bbox'][index].item()
    for key in ['label', 'left', 'top', 'width', 'height']:
        attr = hdf5_data[item][key]
        values = [hdf5_data[attr.value[i].item()].value[0][0]
                  for i in range(len(attr))] if len(attr) > 1 else [attr.value[0][0]]
        attrs[key] = values
    return attrs
```

# Upload your submission.json file to HERE

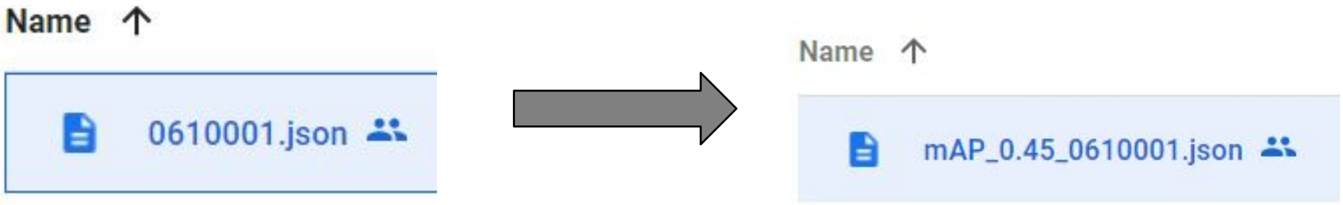
- Free version Kaggle doesn’t provide metrics of detection :(
- Upload you submission file into the Google Drive, we will inference it and return the performance on your filename
- Filename should be <STUDENT_ID>.json



My Drive > NCTU_CS_T0828 > HW2

My Drive trash is changing. Starting October 13, items will be automatically deleted forever after they've been in your trash for 30 days. Learn more

| Name ↑ | Owner | Last modified | File size |
|---|---|---|---|
| dataset | me | Nov 7, 2019 me | – |
| submission | me | 2:19 AM me | – |
| CS_T0828_HW2 | me | 2:27 AM me | – |



Name ↑

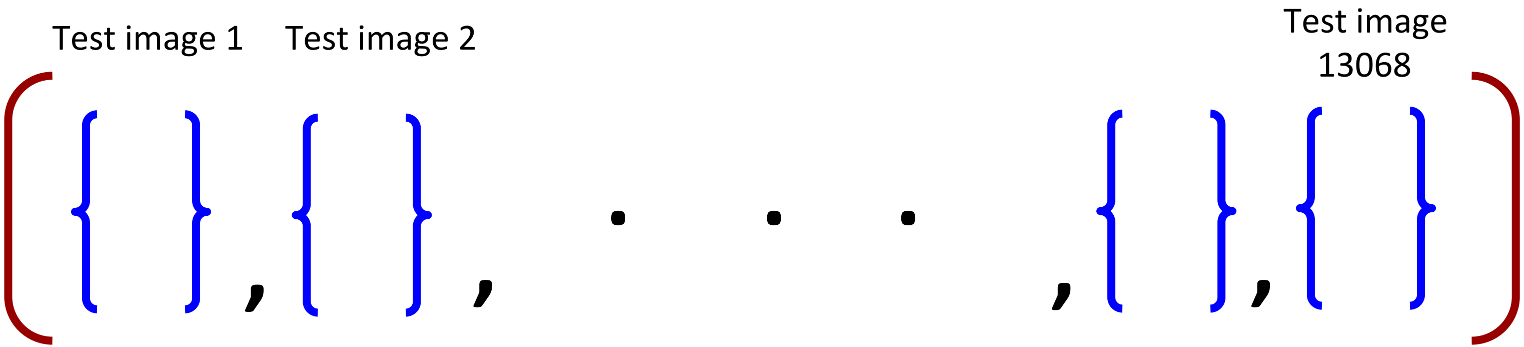
0610001.json

Name ↑

mAP_0.45_0610001.json

# File format of submission.json

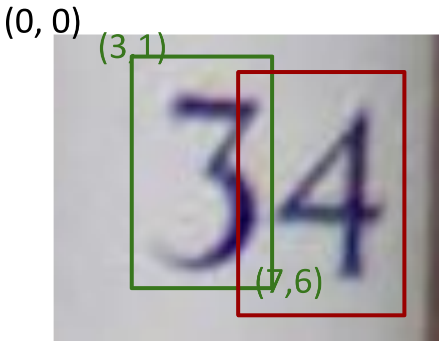
- List of dictionaries, len(list)=number of test images. **Order matters**!
- Each dictionary contains three keys
  - **“bbox”**: list of bounding boxes in (y1, x1, y2, x2). (top,left,right,bottom)
  - **“score”**: list of probability for the class
  - **“label”**: list of label

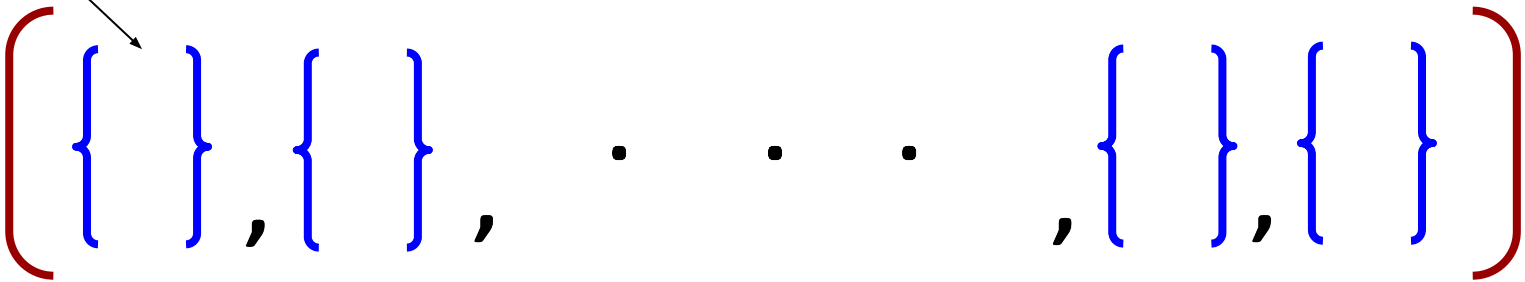


# Submission.json file format

- E.g., your model predict two boxes on an image, the dictionary should be

```
                 y1,  x1,  y2,  x2
dict = {“bbox”: [(1, 3, 6, 7), (4, 5, 8, 12)] ,
          “score”:[0.87, 0.61],
          “label”: [3, 5]}
```

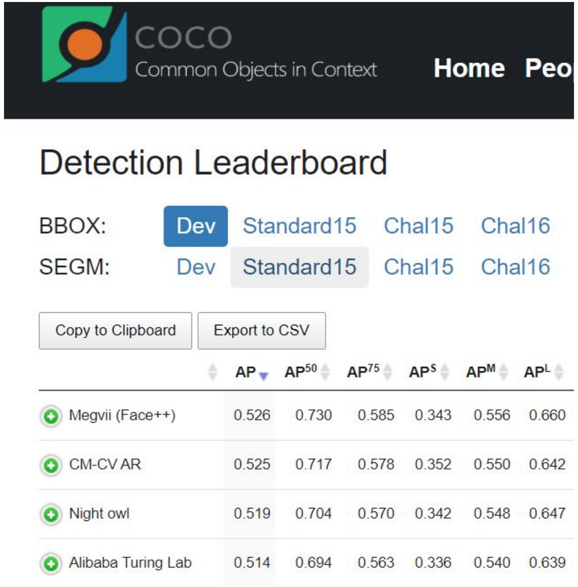




國立交通大學
National Chiao Tung University

# Evaluation metrics: mean Average Precision

- Practical metric for object detection
- Measure the precision on different threshold then average it



COCO
Common Objects in Context

Home Peo

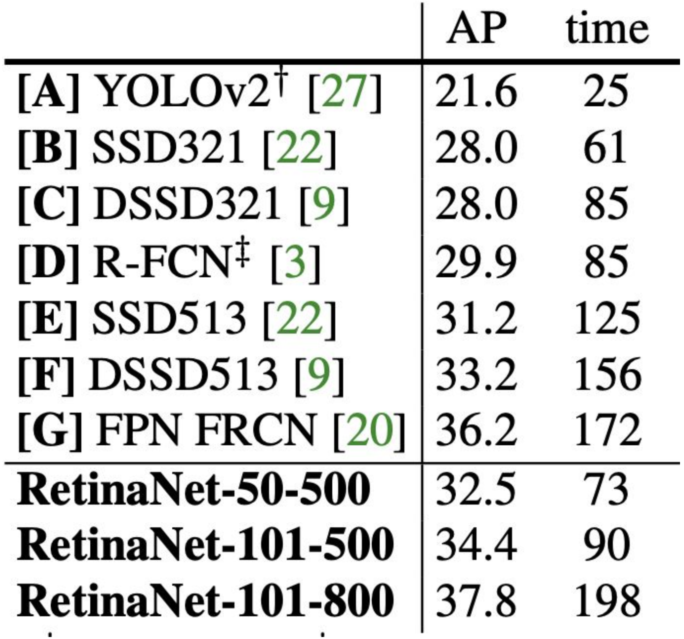
Detection Leaderboard

BBOX: Dev Standard15 Chal15 Chal16

SEGM: Dev Standard15 Chal15 Chal16

Copy to Clipboard Export to CSV

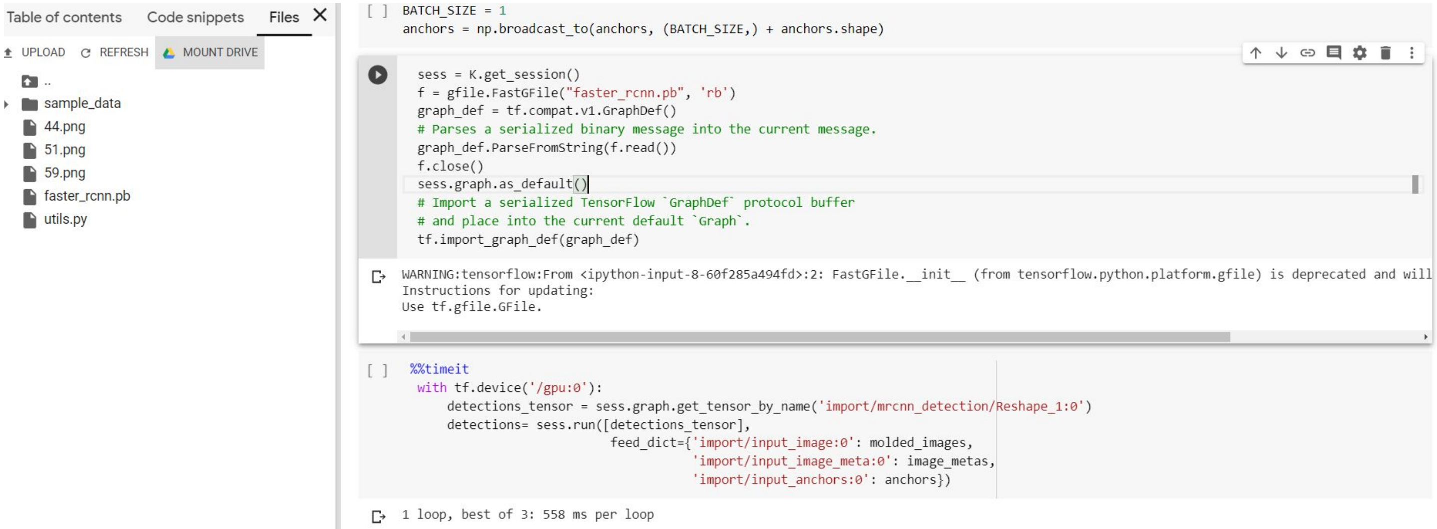
| | AP | $AP^{50}$ | $AP^{75}$ | $AP^{S}$ | $AP^{M}$ | $AP^{L}$ |
|---|---|---|---|---|---|---|
| Megvii (Face++) | 0.526 | 0.730 | 0.585 | 0.343 | 0.556 | 0.660 |
| CM-CV AR | 0.525 | 0.717 | 0.578 | 0.352 | 0.550 | 0.642 |
| Night owl | 0.519 | 0.704 | 0.570 | 0.342 | 0.548 | 0.647 |
| Alibaba Turing Lab | 0.514 | 0.694 | 0.563 | 0.336 | 0.540 | 0.639 |



| | AP | time |
|---|---|---|
| [A] YOLOv2† [27] | 21.6 | 25 |
| [B] SSD321 [22] | 28.0 | 61 |
| [C] DSSD321 [9] | 28.0 | 85 |
| [D] R-FCN‡ [3] | 29.9 | 85 |
| [E] SSD513 [22] | 31.2 | 125 |
| [F] DSSD513 [9] | 33.2 | 156 |
| [G] FPN FRCN [20] | 36.2 | 172 |
| **RetinaNet-50-500** | 32.5 | 73 |
| **RetinaNet-101-500** | 34.4 | 90 |
| **RetinaNet-101-800** | 37.8 | 198 |

# Model speed benchmark by Google Colab

- To evaluate your model by GPU, you need to transfer the code and weights of your model into Colab. Then run inference to test the speed of your model
- **Please include this part in your reports**



```
BATCH_SIZE = 1
anchors = np.broadcast_to(anchors, (BATCH_SIZE,) + anchors.shape)
```

```
sess = K.get_session()
f = gfile.FastGFile("faster_rcnn.pb", 'rb')
graph_def = tf.compat.v1.GraphDef()
# Parses a serialized binary message into the current message.
graph_def.ParseFromString(f.read())
f.close()
sess.graph.as_default()
# Import a serialized TensorFlow `GraphDef` protocol buffer
# and place into the current default `Graph`.
tf.import_graph_def(graph_def)
```

```
WARNING:tensorflow:From <ipython-input-8-60f285a494fd>:2: FastGFile.__init__ (from tensorflow.python.platform.gfile) is deprecated and will
Instructions for updating:
Use tf.gfile.GFile.
```

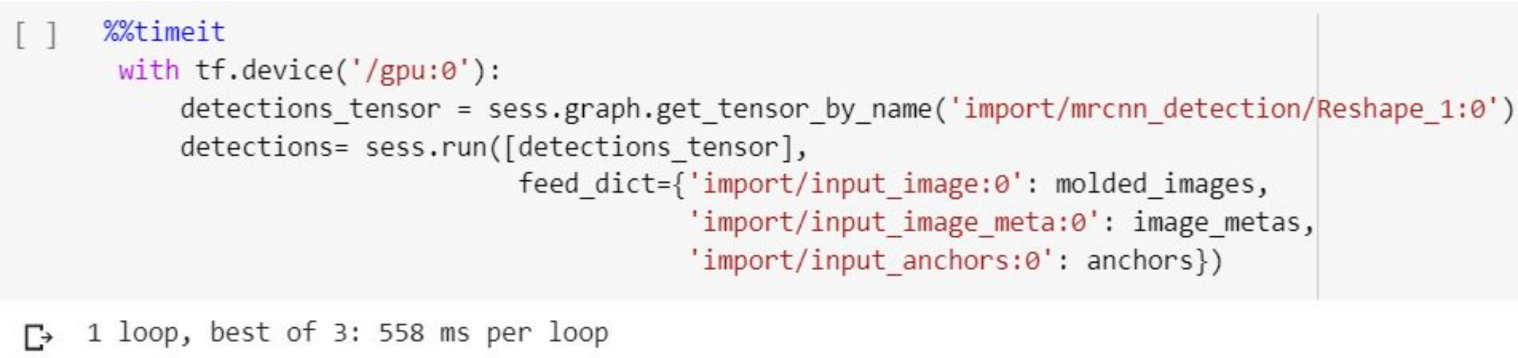
```
%%timeit
with tf.device('/gpu:0'):
    detections_tensor = sess.graph.get_tensor_by_name('import/mrcnn_detection/Reshape_1:0')
    detections= sess.run([detections_tensor],
                          feed_dict={'import/input_image:0': molded_images,
                                     'import/input_image_meta:0': image_metas,
                                     'import/input_anchors:0': anchors})
```

```
1 loop, best of 3: 558 ms per loop
```

# Grading policy: Model performance (70 points)

- 50 points for the accuracy ranking
- 20 points for the speed benchmark ranking
- Pass the each baseline will get 80% of that points

- mAP baseline: 0.36898
- Speed baseline: 558 ms per image



```
[ ]  %%timeit
      with tf.device('/gpu:0'):
          detections_tensor = sess.graph.get_tensor_by_name('import/mrcnn_detection/Reshape_1:0')
          detections= sess.run([detections_tensor],
                               feed_dict={'import/input_image:0': molded_images,
                                          'import/input_image_meta:0': image_metas,
                                          'import/input_anchors:0': anchors})
```

```
1 loop, best of 3: 558 ms per loop
```

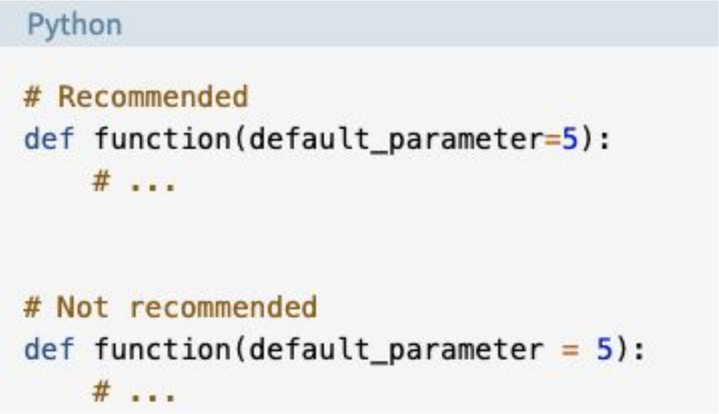
國立交通大學
National Chiao Tung University

# Grading policy: Reports (20 points)

- Document your work (**in PDF**)
    - GitHub/ GitLab link of your code
    - **reference if you used code from GitHub**
    - **Speed benchmark**
    - Brief introduction
    - Methodology (Data pre-process, Model architecture, Hyperparameters,...)
    - Findings or Summary

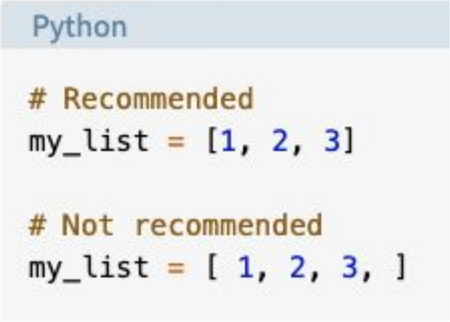
# Grading policy: Code readability (10 points)

- Write beautiful Python code with PEP8 guidelines for readability. Base requirement: use whitespace correctly!



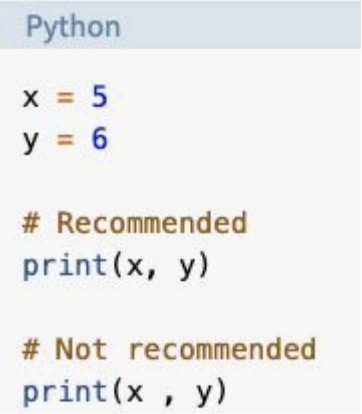
```python
# Recommended
def function(default_parameter=5):
    # ...


# Not recommended
def function(default_parameter = 5):
    # ...
```



```python
# Recommended
my_list = [1, 2, 3]

# Not recommended
my_list = [ 1, 2, 3, ]
```



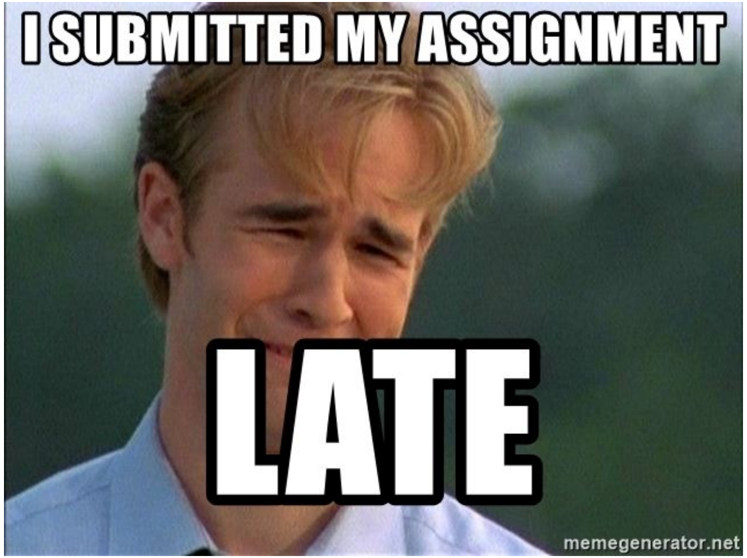
```python
x = 5
y = 6

# Recommended
print(x, y)

# Not recommended
print(x , y)
```

# Late Policy

- We will deduct a late penalty of 20 points per additional late day
- For example, If you get 90% of HW but delay for two days, your will get only 90 points- (20 points x 2) = 50 points!



# Keywords

- Beat the baseline
  - yolo, SSD, Retina-Net, Faster RCNN

- Rank Top 3!
  - Read some paper from CVPR’2020, ECCV’2020 and try to implement it!

# FAQ

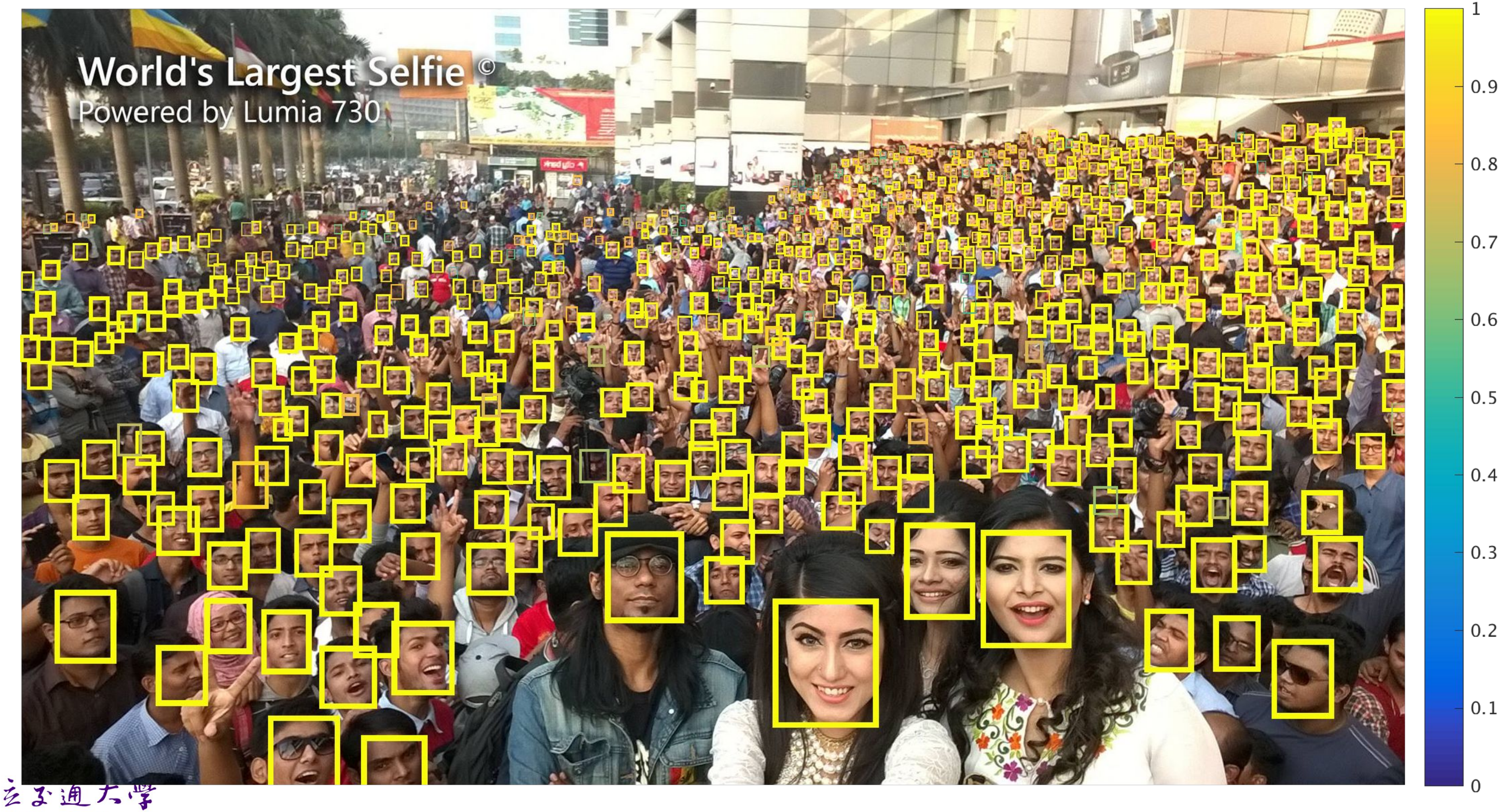
- Can I use any code/tools/Library from GitHub or other resources?
    - Yes! We encourage you to learn how to apply existing tools on your own task, such as Keras-Retinanet, Pytorch-mmdetection, TF-object-detection-API

      **But DO NOT copy code from your classmate!**
- Why my testing results are so bad?
    - If you have done any image translation (resize, padding), you will need to transfer the coordinates into original image dimension
- How do I set the score threshold for box predictions?
    - Figure out how mAP is computed, you will get the answer!

# Notice

- Check your email regularly, we will mail you if there are any updates or problems of the homework
- If you have any questions or comments for the homework, please mail me and cc Prof. Lin
  - Prof. Lin: lin@cs.nctu.edu.tw
  - Jimmy: d08922002@ntu.edu.tw (3pm-4pm, Thur., EC118)
  - 佳諭: mylifeai1116@gmail.com (3pm-4pm, Thur., EC118)
  - 玉霖: oscar861201@gmail.com (3pm-4pm, Thur., EC118)

# Have fun!


World's Largest Selfie ©
Powered by Lumia 730
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0