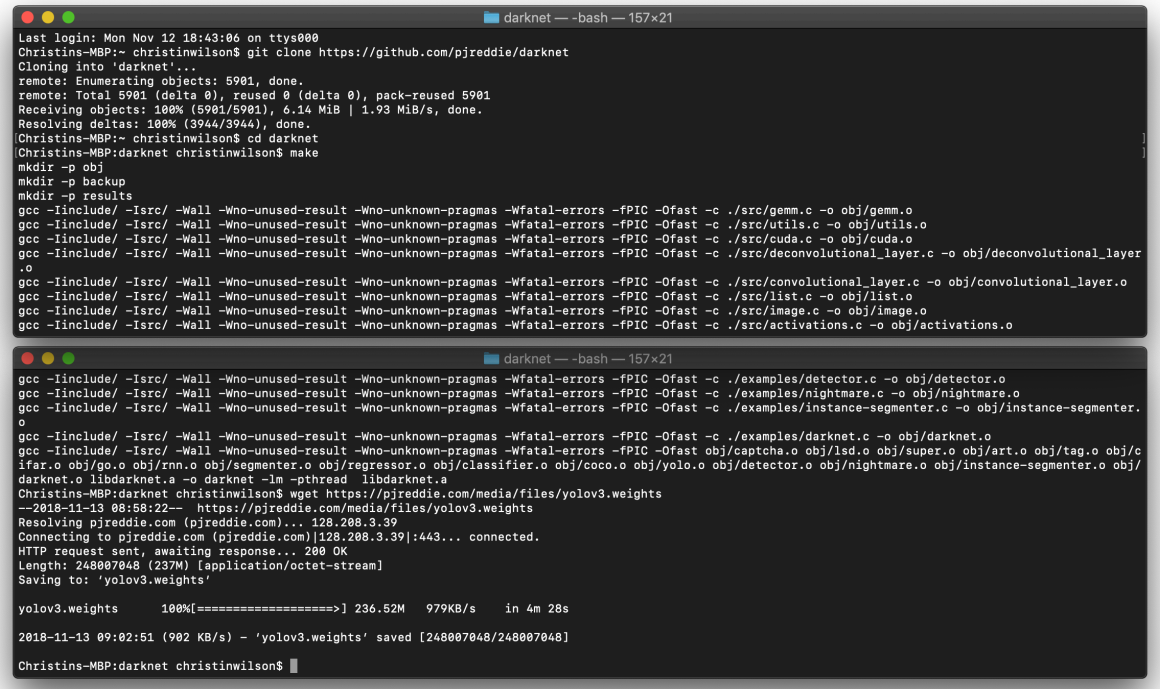
Project 3 - PII Detection Using YOLO

1. Installed darknet, downloaded the pre-trained weight file and finished setting up yolo.



2. Downloaded the dataset. There are 110 images in the private folder and 101 in the public folder.

3. Ran the detector on both the private and public images and stored the objects in two separate txt files.

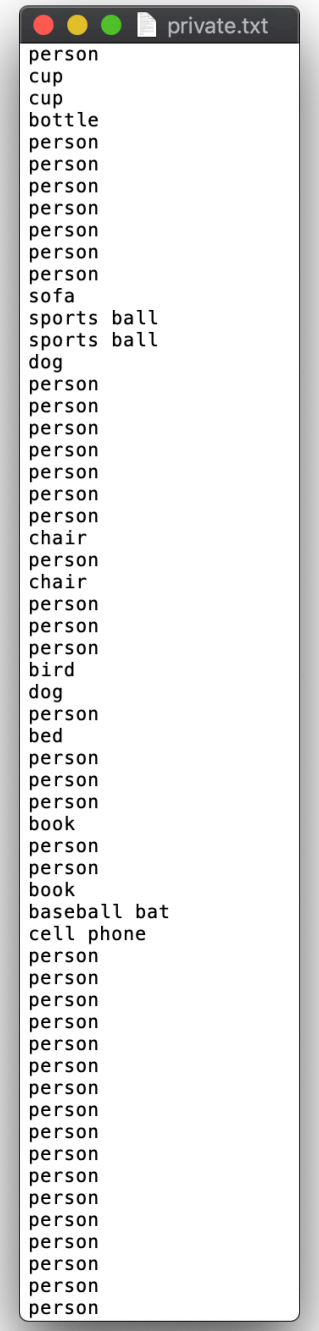
priv.txt

```
person: 100%
cup: 60%
cup: 52%
bottle: 81%
person: 100%
person: 96%
person: 94%
person: 100%
person: 99%
person: 100%
person: 100%
sofa: 52%
sports ball: 62%
sports ball: 100%
dog: 98%
person: 99%
person: 51%
person: 100%
person: 100%
person: 59%
person: 100%
person: 95%
chair: 99%
person: 54%
chair: 99%
person: 100%
person: 100%
person: 100%
bird: 69%
dog: 75%
person: 100%
bed: 50%
person: 100%
person: 100%
person: 100%
book: 98%
person: 100%
person: 100%
book: 63%
baseball bat: 64%
cell phone: 52%
person: 100%
person: 100%
person: 98%
person: 98%
person: 98%
person: 97%
person: 95%
person: 92%
person: 90%
person: 89%
person: 88%
person: 87%
person: 82%
person: 67%
person: 67%
person: 58%
person: 56%
```

pub.txt

```
person: 99%
vase: 99%
train: 88%
truck: 100%
person: 94%
person: 94%
person: 82%
person: 98%
person: 60%
person: 97%
person: 96%
person: 100%
person: 100%
person: 100%
person: 100%
person: 94%
person: 91%
person: 66%
handbag: 55%
person: 100%
person: 100%
handbag: 57%
cake: 61%
person: 95%
person: 99%
person: 100%
train: 100%
person: 100%
person: 97%
person: 83%
giraffe: 100%
giraffe: 100%
giraffe: 99%
giraffe: 97%
hot dog: 57%
diningtable: 60%
car: 99%
car: 99%
person: 100%
person: 50%
person: 77%
person: 98%
bicycle: 72%
bicycle: 100%
person: 92%
car: 90%
person: 69%
person: 99%
boat: 76%
person: 97%
person: 95%
person: 92%
person: 91%
person: 91%
person: 88%
person: 87%
person: 84%
person: 82%
```

4. Processed the txt files and removed the percentages to get only the objects.



5. Ran a unix command to sort and count the objects in each dataset and created a table from this data and got the top 10 private and public objects.

```
sort public.txt | uniq -c | sort -nr
sort private.txt | uniq -c | sort -nr
```

Public

No of occurences	Object
124	person
15	bottle
6	knife
6	chair
4	giraffe
4	car
4	book
3	train
2	handbag
2	diningtable
2	cup
2	cat
2	cake
2	boat
2	bird
2	bicycle
1	vase
1	umbrella
1	truck
1	tie
1	oven
1	keyboard
1	hot dog
1	broccoli
1	bed
1	backpack
1	apple

private

No of occurrences	object
156	person
26	cup
8	wine glass
8	dog
6	chair
5	tvmonitor
5	diningtable
5	cell phone
5	bottle
4	sports ball
4	knife
4	book
3	fork
3	cat
3	car
3	bowl
3	bed
2	suitcase
2	sofa
2	microwave
2	bicycle
1	vase
1	truck
1	tie
1	spoon
1	remote
1	refrigerator
1	handbag
1	carrot
1	cake
1	broccoli

No of occurrences	object
1	boat
1	bird
1	baseball bat

Observations:

Both tables have ‘person’ in the top. This means that people prefer to keep their photos both private and public. Thus it is the environment that they are in that are actually deciding whether the image should be private or public.
When the image has objects like train, giraffe, car, book and handbag, they are usually public images.
When the image has cups, wine glasses, and screens like monitors or cellphone, they are usually private images.

APPENDIX

public.py

CODE:

```
import os
from subprocess import Popen, PIPE
f=open("pub.txt", "a+")
images_dir = “/users/christinwilson/darknet/data/dataset"

#loop to run the detector on every image in the folder
for img in os.listdir(os.getcwd()+"/data/dataset/public"):
    count=0
    setting = "public"
    p = Popen(['./darknet', 'detect', 'cfg/yolov3.cfg',
'yolov3.weights', images_dir + '/' + setting + '/' + img],cwd = '/
users/christinwilson/darknet', stdout = PIPE, stderr = PIPE)
    stdout, stderr = p.communicate()

    #get the number of characters in stdout
    for line in stdout:
        count+=1

    #get the position where the first line of stdout ends
    pos=stdout.find(‘\n')

    #obtain the substring of stdout with just objects and percentage
    stdout=stdout[pos+1:count-1]

    #write it to the file
    f.write(stdout)
f.close()
```

private.py

CODE:

```
import os
from subprocess import Popen, PIPE
f=open("priv.txt", "a+")
images_dir = “/users/christinwilson/darknet/data/dataset"

#loop to run the detector on every image in the folder
for img in os.listdir(os.getcwd()+"/data/dataset/private"):
    count=0
    setting = "private"
    p = Popen(['./darknet', 'detect', 'cfg/yolov3.cfg',
'yolov3.weights', images_dir + '/' + setting + '/' + img],cwd = '/
users/christinwilson/darknet', stdout = PIPE, stderr = PIPE)
    stdout, stderr = p.communicate()


    #get the number of characters in stdout
    for line in stdout:
        count+=1

    #get the position where the first line of stdout ends
    pos=stdout.find(‘\n')

    #obtain the substring of stdout with just objects and percentage
    stdout=stdout[pos+1:count-1]

    #write it to the file
    f.write(stdout)
f.close()
```

Program to Delete percentages:

CODE:

```
lines=[line.rstrip('\n') for line in open('priv.txt')]
f=open("private.txt", "a+")
for x in lines:
     pos=x.find(':')
     x=x[0:pos]
     f.write(x+'\n')
f.close()
lines=[line.rstrip('\n') for line in open('pub.txt')]
f=open("public.txt", "a+")
for x in lines:
     pos=x.find(':')
     x=x[0:pos]
     f.write(x+'\n')
f.close()
```