In my inference engine I have divided the task of predicting if a person is wearing a mask and safety helmet into **two separate tasks:**

Task 1: First, an object detector network is used to predict bounding boxes containing heads in the image

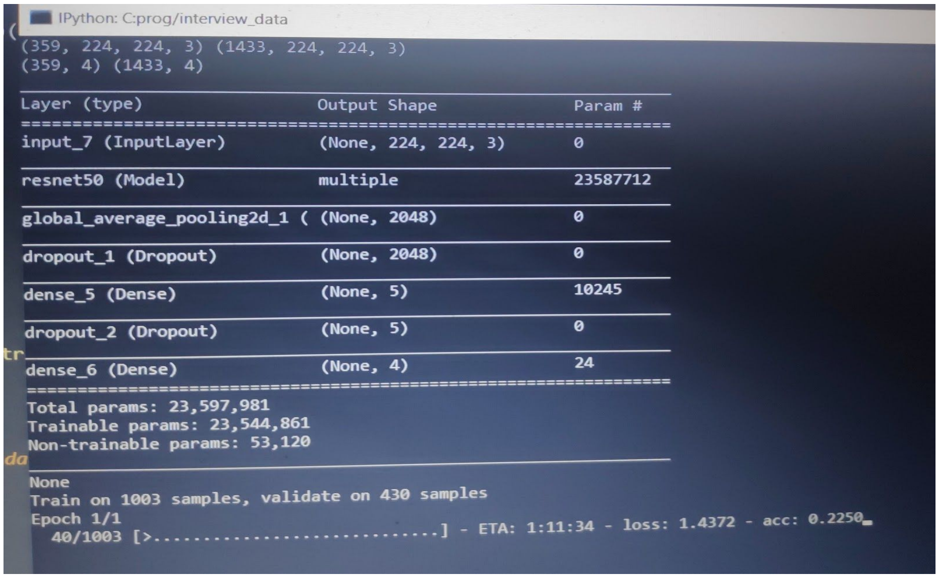
Task 2: Next, two parallel image classifier networks are used to classify the detected head bounding boxes into 4 classes for the presence of mask(yes, no, invisible, wrong) and 2 classes for the presence of safety helmets(yes, no).

So, the **engine** consists of **two stages**:

Stage-1: **Object detector network(Faster RCNN.)** that takes an RGB image as input and produces bounding box locations for heads in the image. The object classifier branch of faster Rcnn classifies each ROI into ***three classes : head, nonhead and background.*** The object bounding box regressor branch produces 4 coordinates xtl, ytl, xbr, ybr specifying the bounding box location.

Stage-2 : 2 parallel Image classifier networks for mask and helmet.

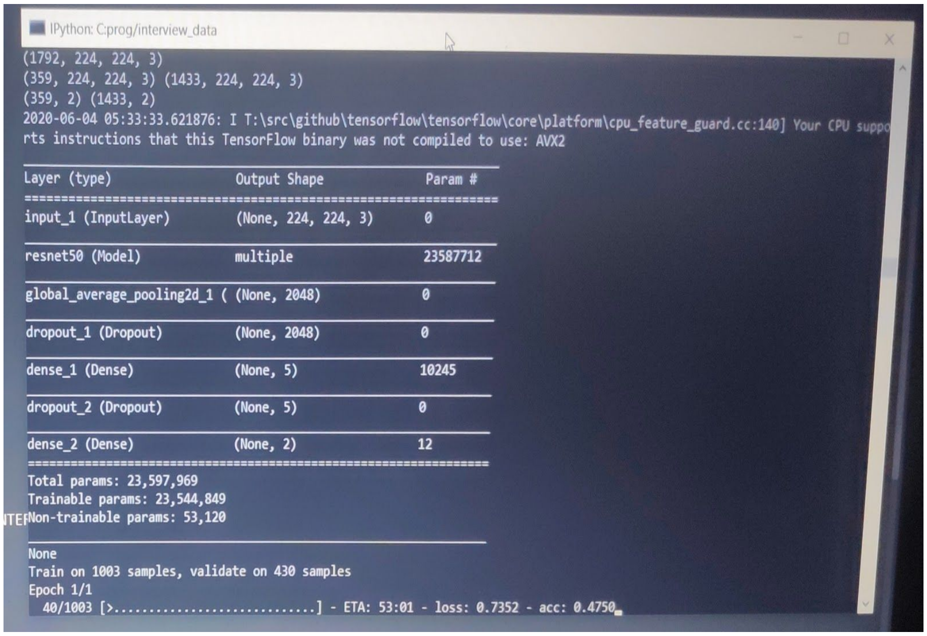
Classifier 1(mask classifier) : Takes the bounding boxes labeled as head from the faster Rcnn detector and crops the bounding boxes from the image and takes them as input. The output has ***4 classes(depending on values present in annotation) viz (invisible, no, wrong, yes)*** denoting the presence of masks in the image. The architecture summary is shown below. Note that the final dense layer has 4 classes.



```
(359, 224, 224, 3) (1433, 224, 224, 3)
(359, 4) (1433, 4)
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_7 (InputLayer)         (None, 224, 224, 3)       0
_________________________________________________________________
resnet50 (Model)             multiple                  23587712
_________________________________________________________________
global_average_pooling2d_1 ( (None, 2048)              0
_________________________________________________________________
dropout_1 (Dropout)          (None, 2048)              0
_________________________________________________________________
dense_5 (Dense)              (None, 5)                 10245
_________________________________________________________________
dropout_2 (Dropout)          (None, 5)                 0
_________________________________________________________________
dense_6 (Dense)              (None, 4)                 24
=================================================================
Total params: 23,597,981
Trainable params: 23,544,861
Non-trainable params: 53,120
_________________________________________________________________
None
Train on 1003 samples, validate on 430 samples
Epoch 1/1
  40/1003 [>.............................] - ETA: 1:11:34 - loss: 1.4372 - acc: 0.2250
```

Mask classifier

Classifier 2(safety helmet classifier) : Takes the bounding boxes labeled as head from the faster Rcnn and crops the bounding boxes from the image and takes them as input. The output has ***2 classes(depending on values present in annotation) viz (yes, no)*** denoting the presence of safety helmets in the image. The architecture summary is shown below. Note that the final dense layer has 2 classes.



```
(1792, 224, 224, 3)
(359, 224, 224, 3) (1433, 224, 224, 3)
(359, 2) (1433, 2)
2020-06-04 05:33:33.621876: I T:\src\github\tensorflow\tensorflow\core\platform\cpu_feature_guard.cc:140] Your CPU suppo
rts instructions that this TensorFlow binary was not compiled to use: AVX2
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 224, 224, 3)       0
_________________________________________________________________
resnet50 (Model)             multiple                  23587712
_________________________________________________________________
global_average_pooling2d_1 ( (None, 2048)              0
_________________________________________________________________
dropout_1 (Dropout)          (None, 2048)              0
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 10245
_________________________________________________________________
dropout_2 (Dropout)          (None, 5)                 0
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 12
=================================================================
Total params: 23,597,969
Trainable params: 23,544,849
Non-trainable params: 53,120
_________________________________________________________________
None
Train on 1003 samples, validate on 430 samples
Epoch 1/1
  40/1003 [>.............................] - ETA: 53:01 - loss: 0.7352 - acc: 0.4750
```

Helmet classifier