Model Optimization for Face Detection

November 16, 2020

1 FaceDetection:

GroundTruth:
Here I used "Matlab Video Labeller" to generate ground truth. This video labeller returns .mat file , which has all the labels and the cordinates of the object.
Then I converted them to csv files for ease of use in python.

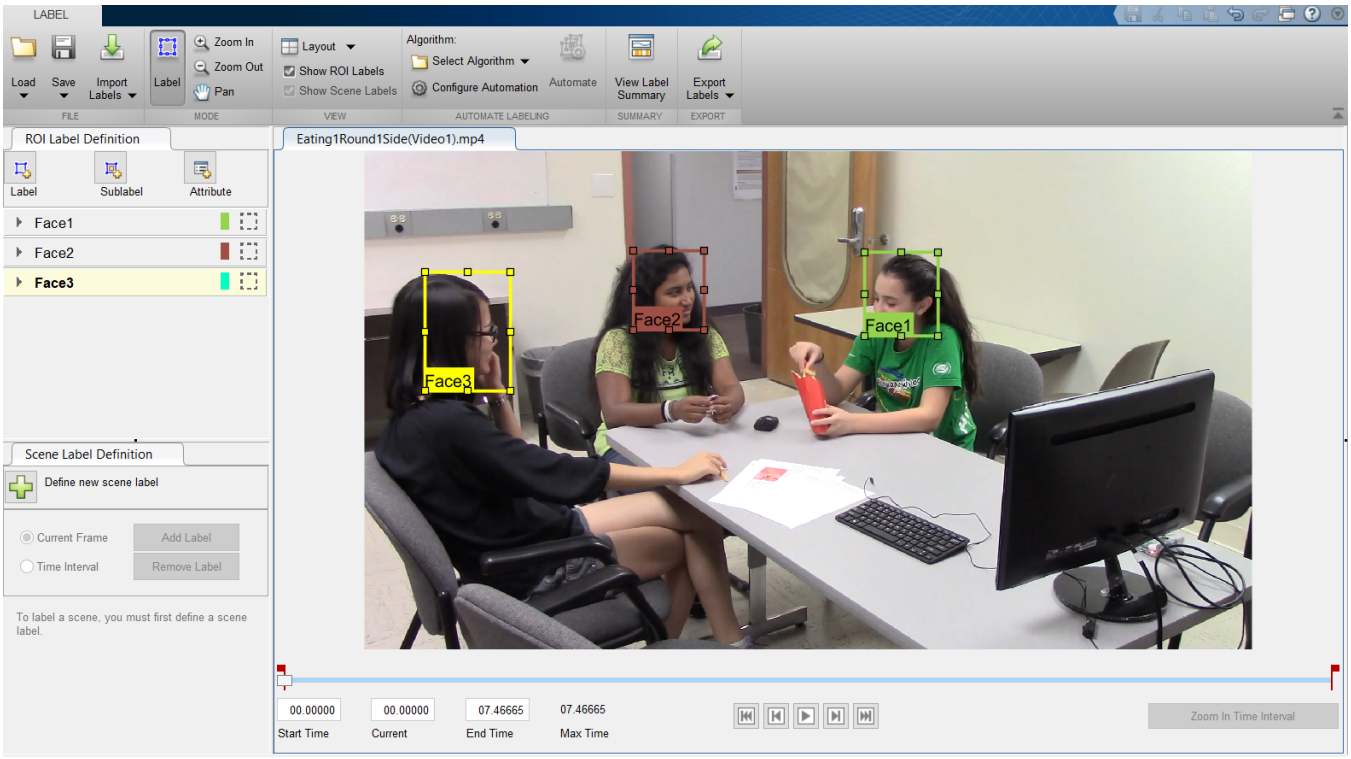

Figure 1: Image with anchor points

This initial ground truth is used as reference for generating actual ground truth for the classifier.

Generating Anchor points:
To generate ground truth , anchor points are placed across the image.Anchor points are the points which are placed at equal intervals on the image as shown in the following figure. Here in this image I used,row space=7 and col space=7 and placed anchor points.

Generating Anchor boxes:
With these anchor points as center and Aspect ratios= $\{1, 1.2\}$ and scales=$\{1, 1.2\}$,
anchorboxes are generated. I used unit scale spacing= 170 So now with the combination of Aspect ratio=1 and scale=1
Number of BoxRows and BoxCols are calculated as below
BoxCols = Scale*UnitScaleSpacing
BoxRows = AspectRatio*NumOfCols

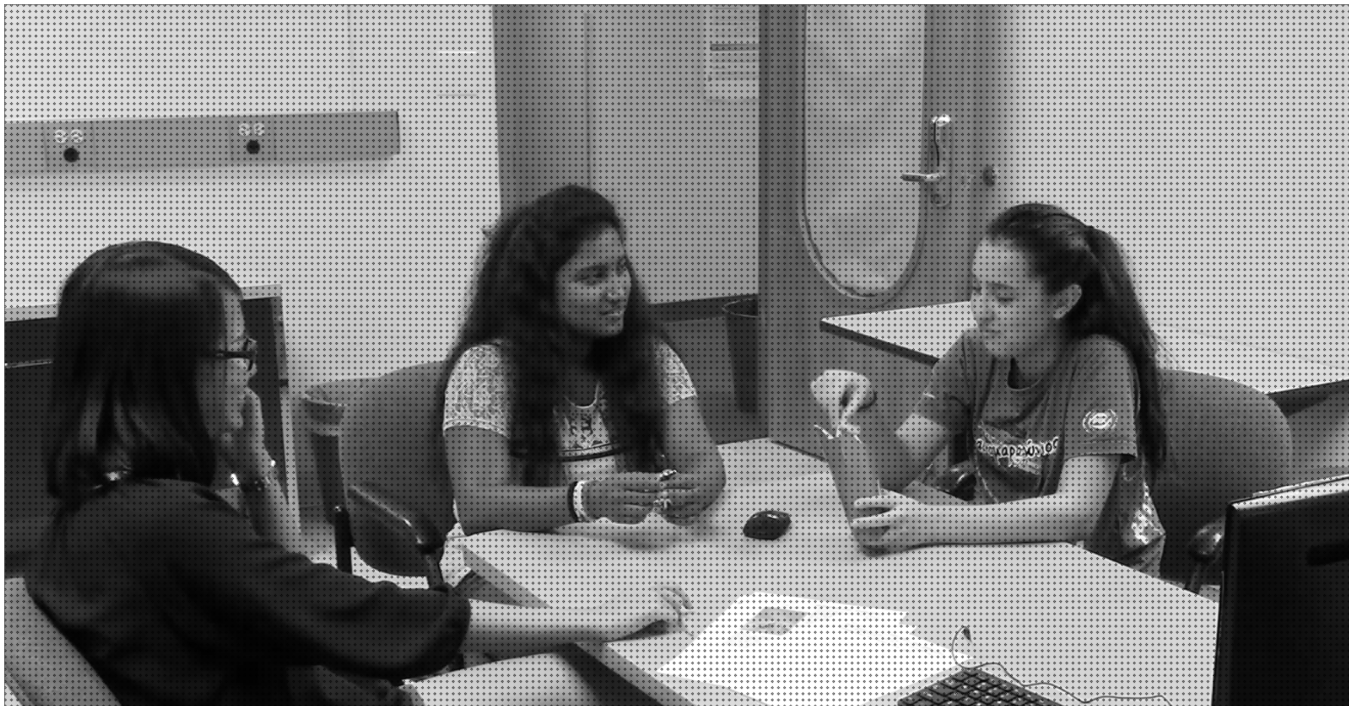

Figure 2: Image with anchor points

$\Longrightarrow$ Boxrows=170 and BoxCols=170 Now boxes are generated using BoxCols and BoxRows with anchor points as center.

Validating Anchor points:

All the boxes generated may not be valid ones. Suppose getting a box at point (10,10) is not valid because it is at extreme corner and doing this will result in bad results.

So to avoid this, I am validating all the anchor points, So the below is the picture with valid anchor points.

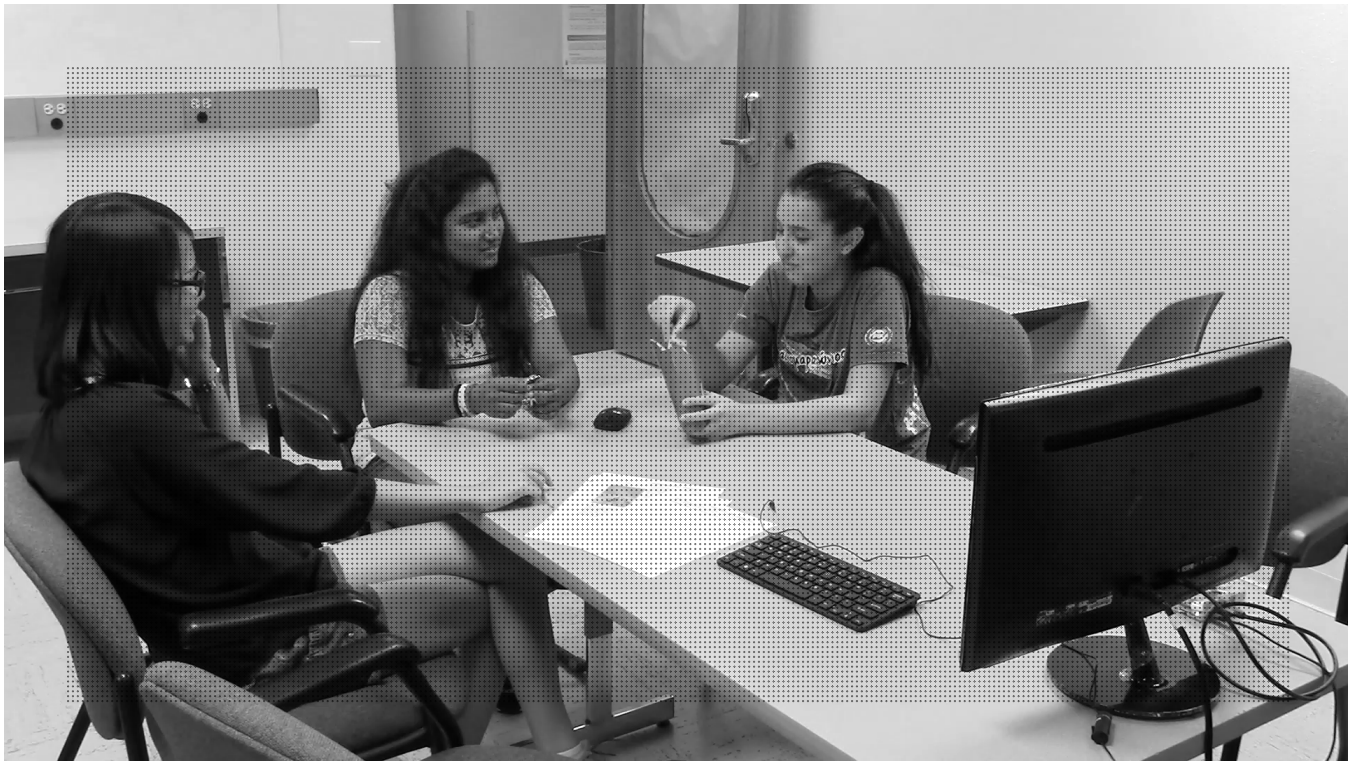

Figure 3: Image with valid anchor points

Deciding ground truth for boxes generated:
Here the overlap ratio=Intersection/union is used to decide if the box is face or not a face.
For each box, the overlap ratio with all the three faces is calculated and if any of these ratios is greater than 0.6, the face is labeled as face. For each frame, three face boxes and three no face boxes are taken for training the classifier. In the following figure, all the three black boxes are ground-truth generated by mat-lab and white box is the one we are trying to classify as background or face.

Some of the images which are classified as face and background are as below

From 9 videos I used 157 frames , which are taken at half the frame rate to generate ground truth. From every frame, 3 face and 3 non face boxes are generated.

Figure 4: Calculating overlap ratio

Figure 5: examples of ground truth for face

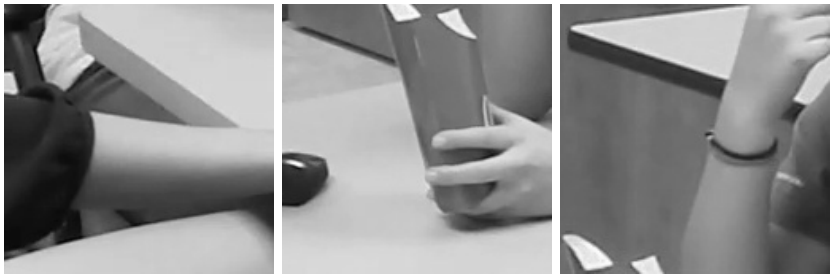

Figure 6: examples of ground truth for background

Finding optimal parameters:
Here nested cross validation is used to find the best parameters which fits the classifier. Pipeline of PCA(Prinicipal Component Analysis) and SVM is used as classifier. Here within the whole dataset, 80 percent is used for training+validation and I tested on remaining 20 percent.
Results for four classifiers are tabulated below
Validation Accuracies and Best parameters:

Comparision of four Classifiers				
Metric	Classifier-1 (A=,S=1)	Classifier-2 (A=1.2,S=1)	Classifier-3(A=1,S=1.2)	Classifier-4 (A=1.2,S=1.2)
Best parameters(C)	1000	1000	1000	1000
gamma	0.0001	0.0001	0.0001	0.0001
NumOfComponents	10	10	10	10
Accuracy(best)	92.56	95.8	91.7	85.9
Accuracy(Mean)	93	95	92	83

For testing(not full frame):

Comparision of four Classifiers(testing)				
Metric	Classifier-1 (A=,S=1)	Classifier-2 (A=1.2,S=1)	Classifier-3(A=1,S=1.2)	Classifier-4 (A=1.2,S=1.2)
Confusion matrix	$\begin{bmatrix} 87 & 7 \\ 6 & 88 \end{bmatrix}$	$\begin{bmatrix} 87 & 7 \\ 2 & 92 \end{bmatrix}$	$\begin{bmatrix} 86 & 8 \\ 5 & 89 \end{bmatrix}$	$\begin{bmatrix} 71 & 23 \\ 10 & 84 \end{bmatrix}$
Accuracy	93.08	95.2	93	83.89

Final Classifier: After finding the best parameters from cross validation, we fit the training data for classifier using these best parameters. Final Classifier is implemented for full frame. Here 80 percent of data is used for training and 20 percent for validation and testing is done on new full frame.

Testing results:

Frame-1				
Metric	Classifier-1 (A=,S=1)	Classifier-2 (A=1.2,S=1)	Classifier-3(A=1,S=1.2)	Classifier-4 (A=1.2,S=1.2)
Confusion matrix	$\begin{bmatrix}715 & 35\\ 122 & 262\end{bmatrix}$	$\begin{bmatrix}681 & 61\\ 25 & 444\end{bmatrix}$	$\begin{bmatrix}705 & 45\\ 22 & 447\end{bmatrix}$	$\begin{bmatrix}595 & 155\\ 4 & 628\end{bmatrix}$
Accuracy(validation)	95.7	93.6	93.6	88.2
Accuracy(testing)	86.15	92.4	94.5	88.4
Frame-2				
Metric	Classifier-1 (A=,S=1)	Classifier-2 (A=1.2,S=1)	Classifier-3(A=1,S=1.2)	Classifier-4 (A=1.2,S=1.2)
Confusion matrix	$\begin{bmatrix}695 & 55\\ 127 & 257\end{bmatrix}$	$\begin{bmatrix}724 & 26\\ 113 & 369\end{bmatrix}$	$\begin{bmatrix}684 & 66\\ 16 & 453\end{bmatrix}$	$\begin{bmatrix}698 & 52\\ 0 & 632\end{bmatrix}$
Accuracy(validation)	90.7	97.6	87	90.3
Accuracy(testing)	83.9	88.7	32.7	92.4
Frame-3				
Metric	Classifier-1 (A=,S=1)	Classifier-2 (A=1.2,S=1)	Classifier-3(A=1,S=1.2)	Classifier-4 (A=1.2,S=1.2)
Confusion matrix	$\begin{bmatrix}714 & 36\\ 133 & 251\end{bmatrix}$	$\begin{bmatrix}726 & 24\\ 105 & 377\end{bmatrix}$	$\begin{bmatrix}705 & 45\\ 13 & 456\end{bmatrix}$	$\begin{bmatrix}642 & 108\\ 0 & 632\end{bmatrix}$
Accuracy(validation)	92.0	89.5	93.6	90.6
Accuracy(testing)	85.0	89.5	95.2	92.4

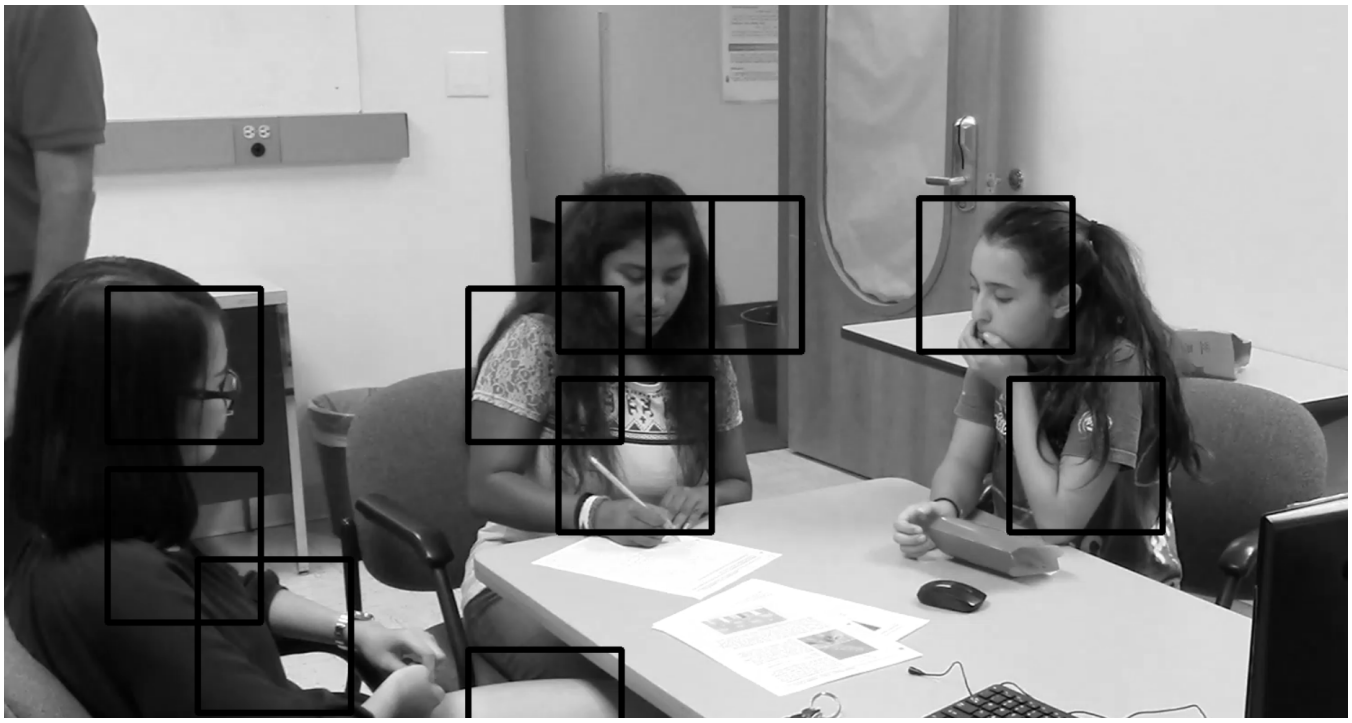

Figure 7: Test on full frame

The above picture shows that there are more false positives and it is evident even from confusion matrix.

I changed aspect ratios and scales from 2 to 1.2 because with '2' and unit space=170, the boxes generated are big and it is hard to get ground truth for higher thresholds.