Project Milestone 1

Mohamed Kamal AbdEl rahman 60

Mohamed Elmaghraby Mohamed 55

Explanation of the problem statement

- Our project focuses on moving object detection which deals with identifying and locating objects of certain classes in a stream of videos.
- The best accuracy achieved so far is on coco dataset which is 37.4 by CenterNet on 52 fps.

We would use MOT video of labeled bounding boxes as a data sets and IOU and precession accuracy.


55.96

Related work

- CenterNet where the network uses Hourglass blocks and detects the center point of the object. 37.4% mAP
- CornerNet with its variation (original, squeeze) 36.5% mAP
- SSD detectors 26.8% mAP
- Yolo V2 28% mAP
- RetinaNet 31% mAP

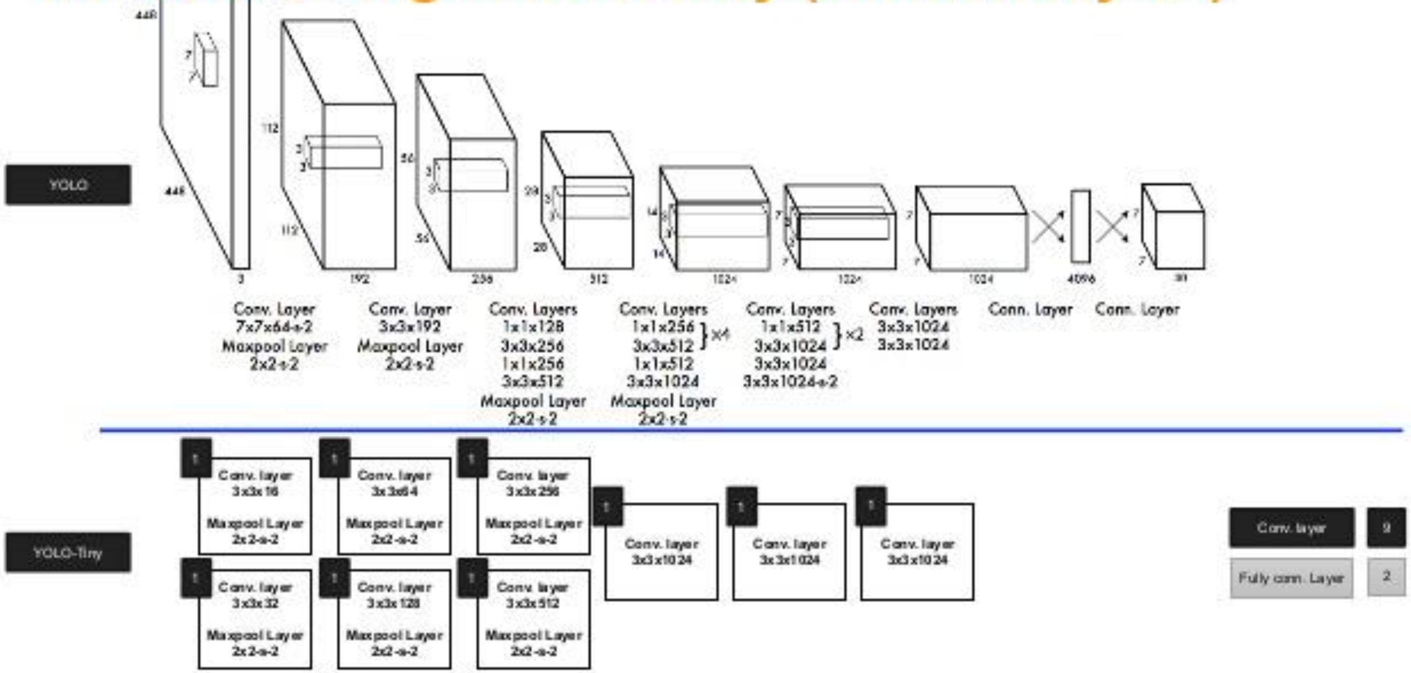
Loss Function

We selected to use Tiny Yolo as

- Because Yolo considered the first model to achieve good good performance near real-time.
- Use the tiny version for speed of training and inference.
- And has a simple architecture

$$
\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \sum_{i=0}^{S^2} 1_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \qquad (3)
\end{aligned}
$$

Network Design: YOLO-tiny (9 Conv. Layers)



Performance evaluation metric AP(average precision)

- IOU to measure overlap between two bounding boxes (detected and ground truth)
- True Positive (TP): A correct detection. Detection with IOU ≥ threshold
- False Positive (FP): A wrong detection. Detection with IOU < threshold
- False Negative (FN): A ground truth not detected
- Typical threshold used is .5 or .75 or .95
- Precision recall curve: An object detector is considered good if its precision stay high as recall increase.
- Average Precision (AP) is a quantity to judge how good is precision recall curve it represent the area under the curve so the higher it is the better the object detector.

proposed solution

Feature fusion?

Combining two or more different feature representation for the data two come up with combined feature that more representative than both of original features.

Example

Is that enough?

No, because the foreground mask is not always that accurate due to:

- Possibilities of noise.
- It is hard to accurately compute it in dynamic background.

Current models achieved respectable accuracy but face the real time constraint on this problem so we proposed

- since we are working on a stream of videos we use a background subtraction technique to use a foreground mask and integrated it as another channel to the image input.
- The main goal is to increase the accuracy without increasing the inference time of the model.

expectations

pros:

- We expect increase the accuracy of detecting moving objects without a significant increase in computational power because the foreground mask is sparse and expected to carry most of the information of the moving object.

- train model faster (intuition).

cons:

- Overfitting may happen (due specific background scene)

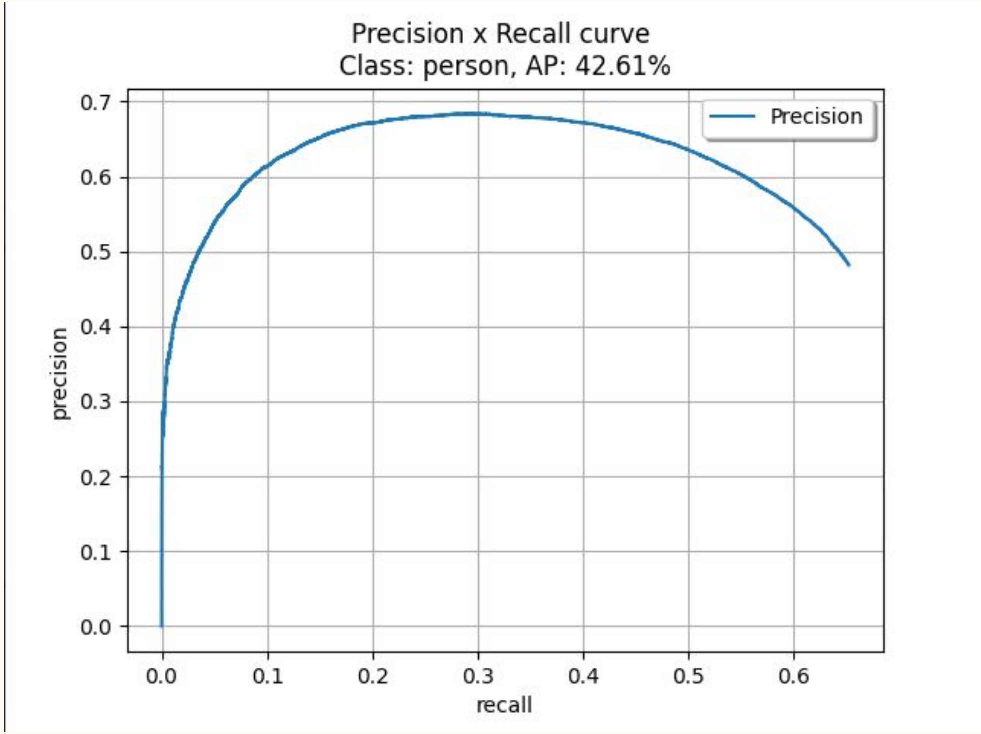
Progress report

Current progress

- Implemented the utility functions
  - Convert image to grayscale
  - Background subtraction
  - Resize image
  - Evaluate accuracy function
- Download Yolo Tiny weights and configurations
- Download MOT datasets
- Build Yolo Tiny model on keras
- Run existing models
- Run proposed model and get initial result

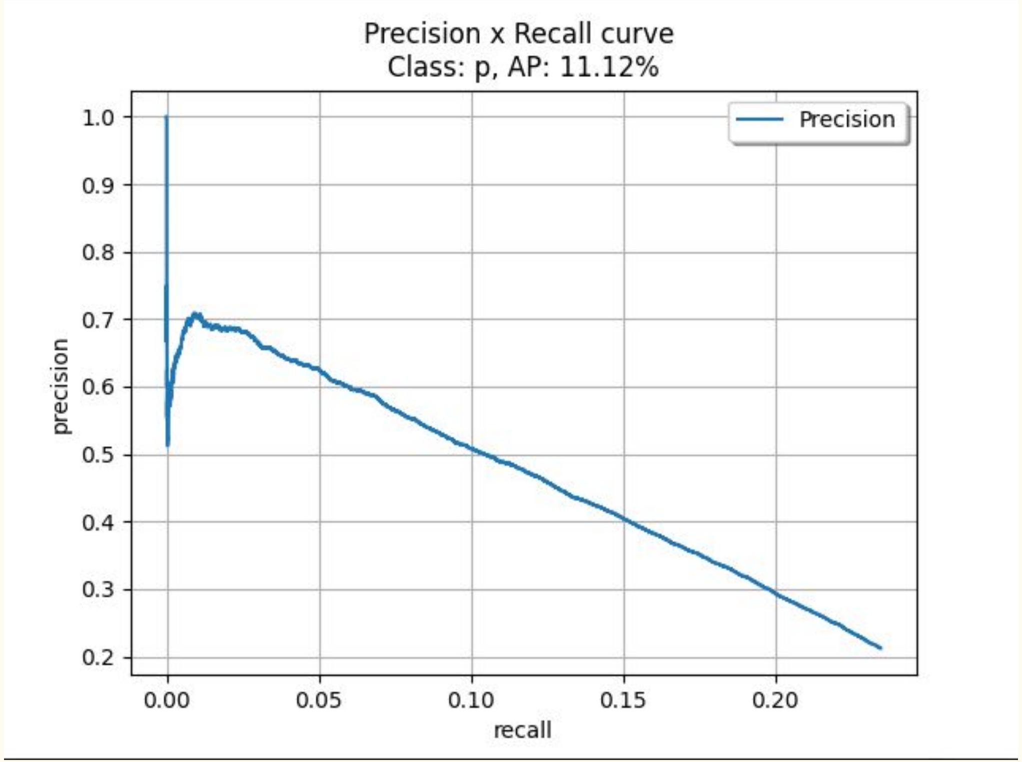
Initial Results of original models with its pretrained weights.

- Yolo v3



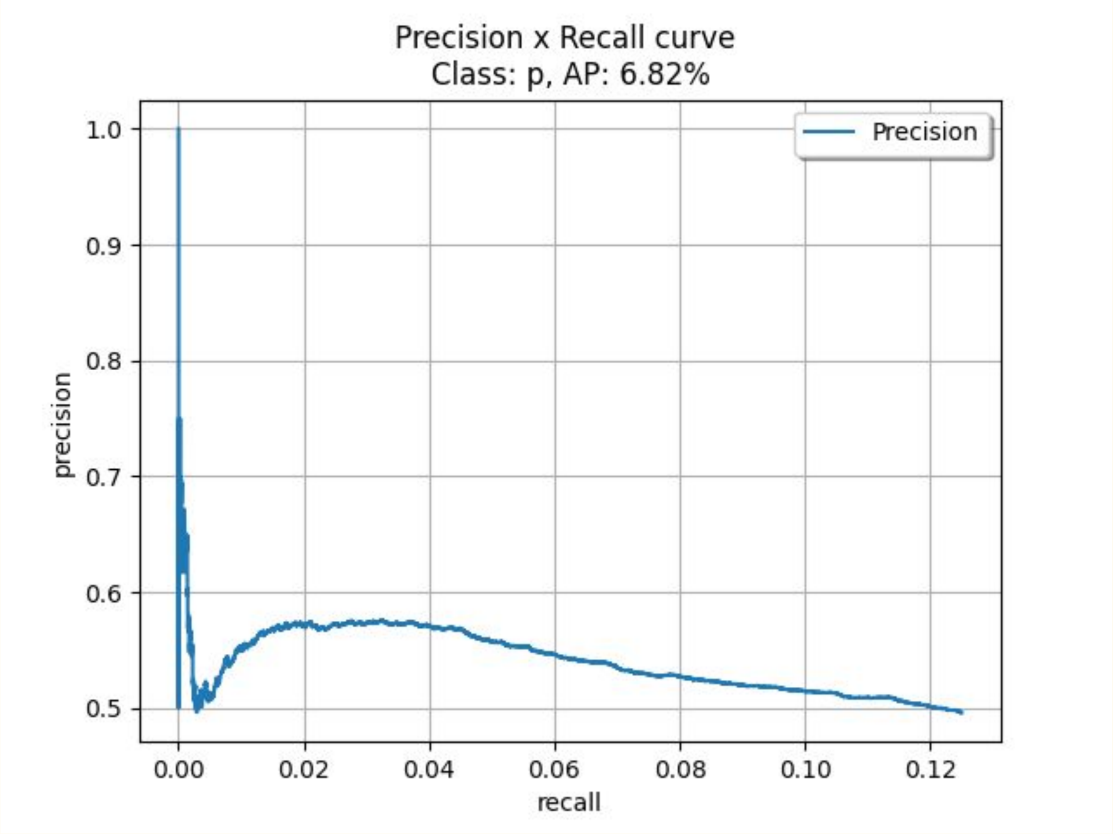
Initial Results of original models with its pretrained weights.

- Tiny-Yolo



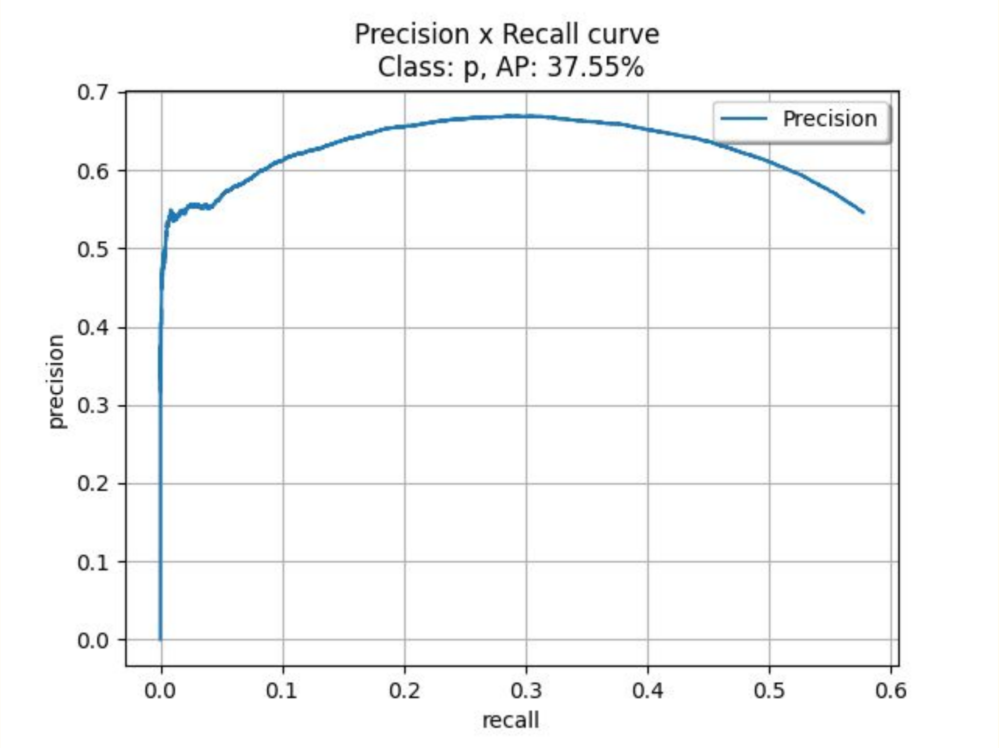
Initial Results of original models with its pretrained weights.

- ssd_mobilenet



Initial Results of original models with its pretrained weights.

- fast_rnn_resnet



Inference time:

Model name	Inference time (ms)	Average precision (%)
Yolov3	83.73	42.61
tiny-Yolo	35.6	11.12
ssd_mobilenet	115.07	6.82
fast_rnn_resnet	94.83	37.55

Next steps

- Train the model, tune hyperparameters to get best accuracy of the proposed model.
- Test our approach with different models (Yolo, RetinaNet, etc)

Resources

1. Tensorflow object detection API
2. mAP (mean Average Precision) for Object Detection
3. Object-Detection-Metrics
4. Yolo implementaion

Thank You