Project report

Grid-Slid-Flow automated tracking

Gia Son Nguyen

giason@ualberta / giasonnguyen995@gmail.com

I. Introduction

The project aims to track an object without the need of registering a region-of-interest, this is something like object detection, but without manual selection of region-of-interest for training data like how YOLO needs for training. This project utilizes light-weighted CNN binary classification models alongside a grid-based approach to automatically detect and track desired object, eliminating the need for manual region-of-interest selection in registered tracking. Unlike state-of-the-art object detection method YOLO (you only look once), this approach does not require to manual selection of region of interest for training data. This method is suitable for tracking object for a long period of time, especially when there are obstructions within the time frame.

II. High level explanations

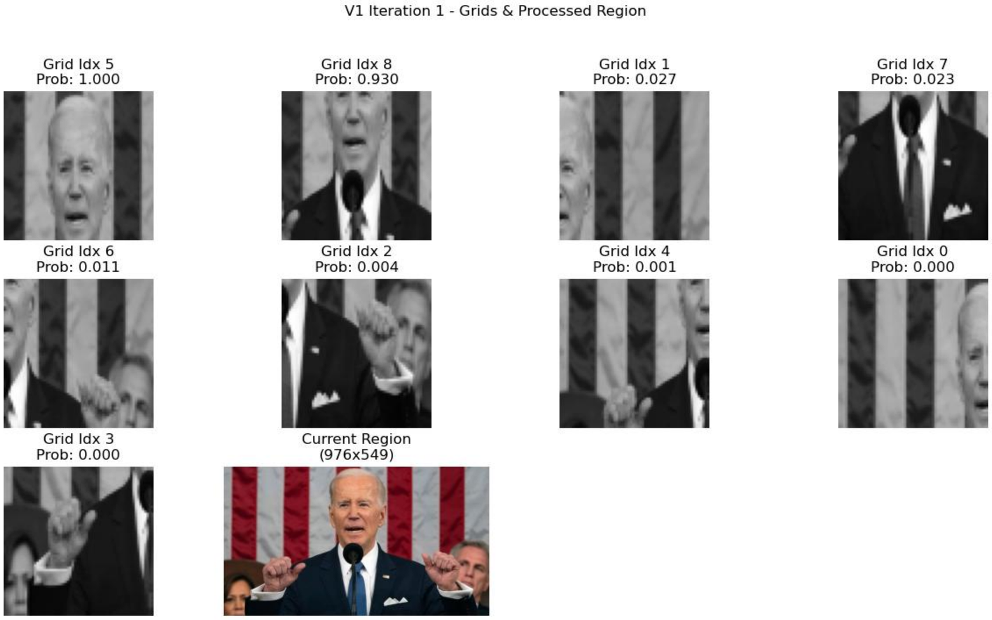
Take a frame from the video or live camera, we divide the frame into 9 cells (4 quarters cells, 4 cells each overlap the 4 lines that separate the quarters cells, and 1 cell overlap the cross of the lines). Then we would stack the cells and use it as input to the trained model, and the model would give us the probability that a cell contains the object. And from the promising cells, we merge them into a promising region and repeat the process again until we can no longer get a promising cell (or until the promising region is smaller than a threshold). After that, a simple sliding window method is used to further refine the region of interest, then that region is used for a 2DoF Lucas Kanade tracker. The template image of the tracker, which is the region of interest we got earlier, will be reevaluate every n frames and the region of interest will be redefined if the template image still contains the object. (n is a number of frame that we see fit, a parameter)

III. Overall process and result:

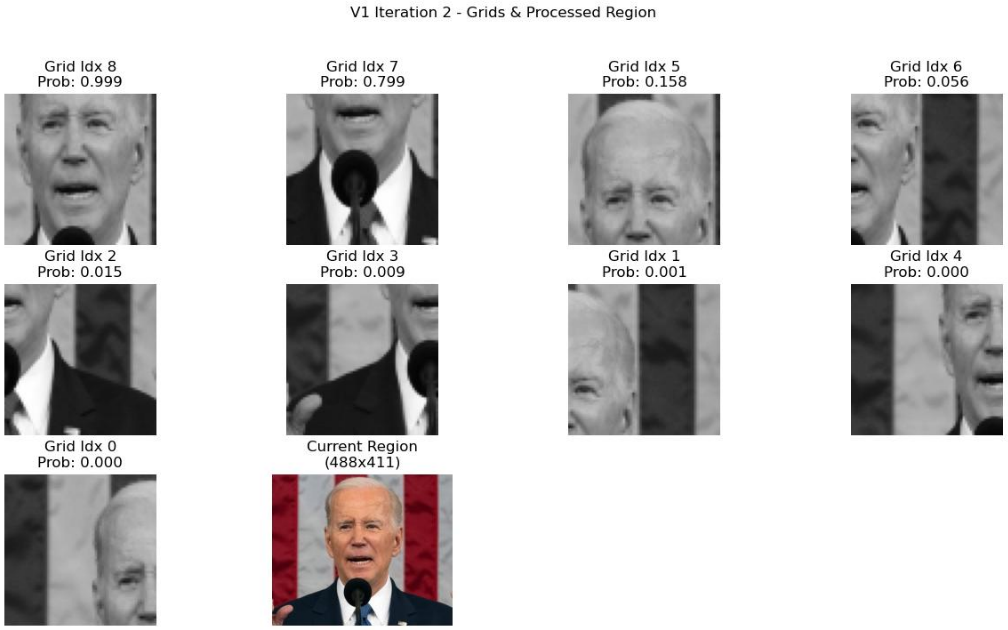
1) Training the model

The model was trained using images of objects from other multiple classes and solid colors as the negative class, and human faces (even though a lot of images are portrait-style) as the positive class. The model takes in a stack of grayscale image and output a vector of probabilities of the human face is in the corresponding image.(data used for training: Natural Images, Human Faces, Gender Recognizer, note, the best model, BB3, did not used all these dataset)

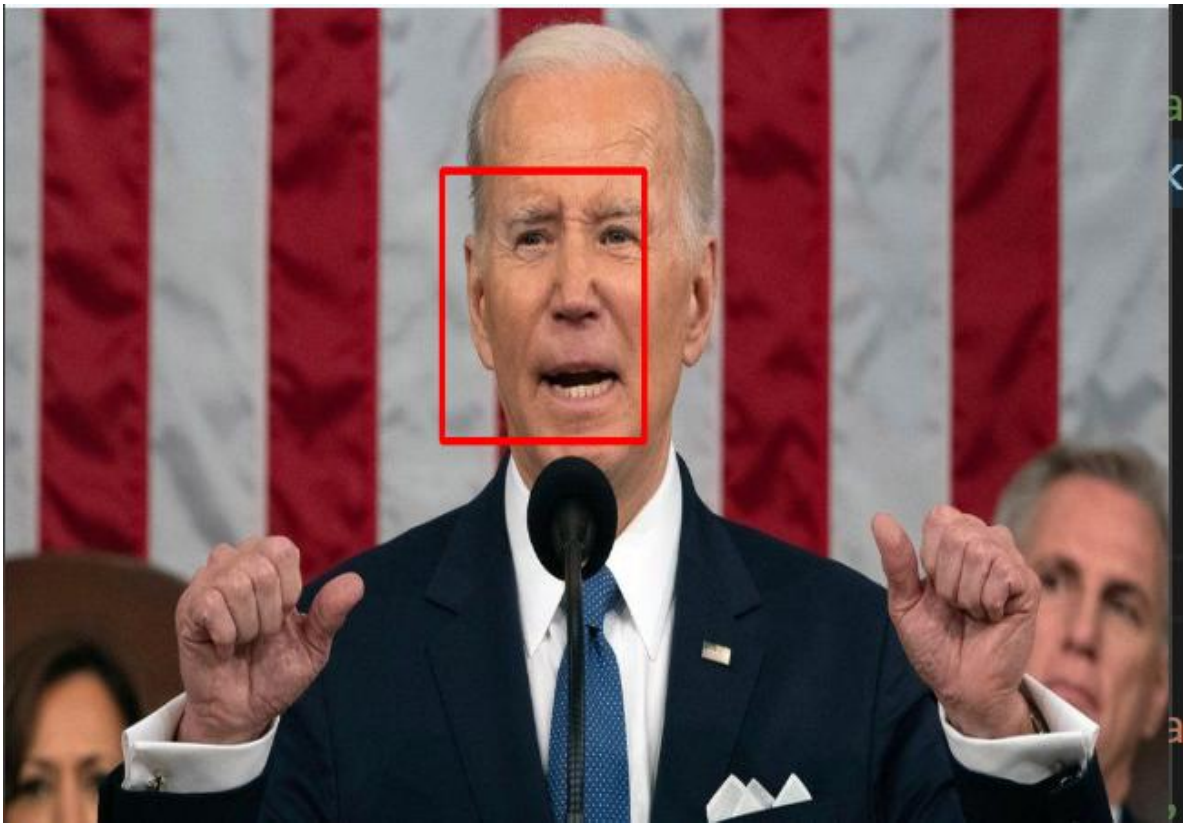
2) Divide the image and get their probability. (image source: https://news.wttw.com/2023/03/08/biden-s-budget-aims-cut-deficits-nearly-3-trillion-over-10-years)



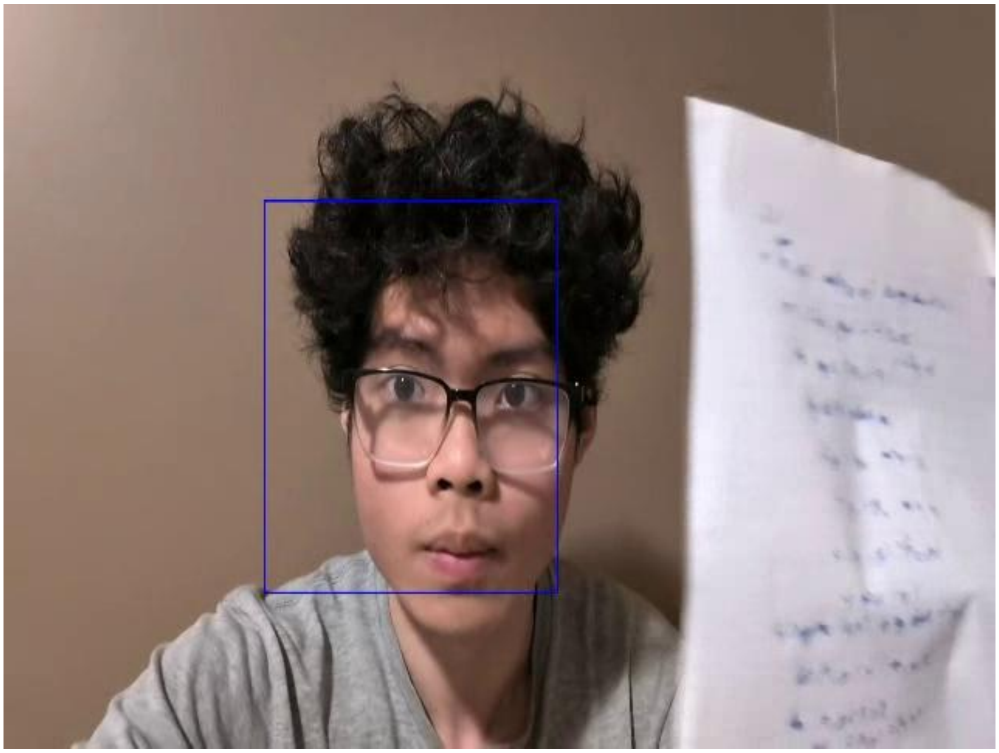
3) Repeat on the promising region until the cells are not good enough anymore or size is under a threshold (40% in this case)



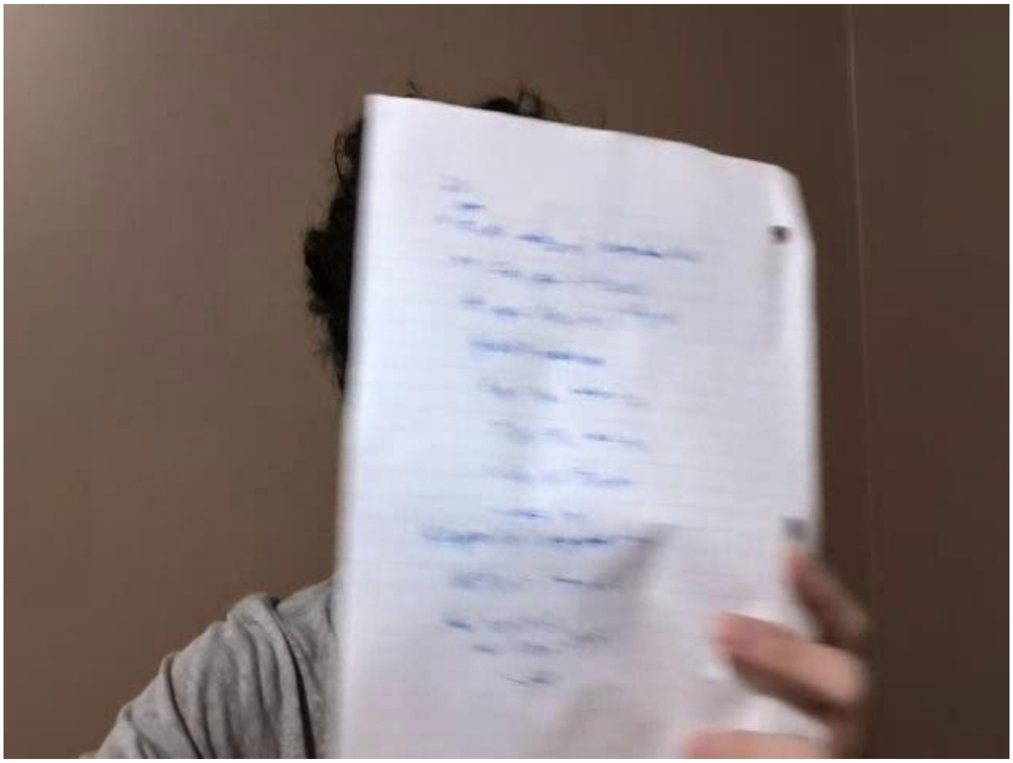
4) And then use a sliding window to refine the region of interest.



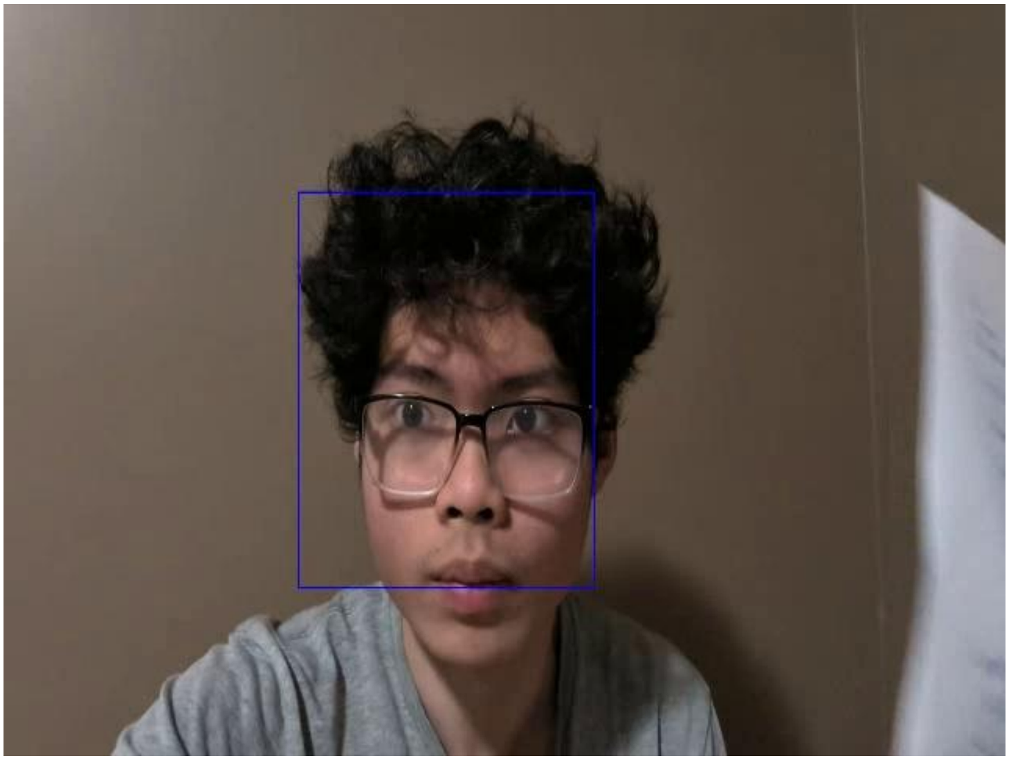
5) Apply Lucas Kanade tracking (new example).



6) Reevaluate the template image (this one does not match because the object is obstructed)



7) The object got redetected when the object is no longer obstructed.



IV. Advantages of this Approach (The Good):

1) There is no need to register the region-of-interest in tracking.
2) Reduce the runtime of window sliding by reducing the area it needs to run on.
3) More generalized than template matching.
4) No need for manually selecting region-of-interest for training data like YOLO object detection since it’s running based on classification.
5) If tracking loses sight of the object or the object is covered, this approach can actually re-detect the region-of-interest.
6) Faster than expected (can work live even on a CPU).

V. Disadvantages of this Approach (The Bad):

1) Does not work quite well with small object.
2) The bounding box can be not as great as registered tracking.
3) There’s the need to effectively divide the grid.
4) Cannot track multiple objects.

VI. The Ugly:

Heavily dependent on the trained model. If the model is good, it would work well, if it’s bad, it won’t. It all depends on the model and trained data.

VII. Future considerations and possible way of improvements.

1) Can we use 2 models? One for object within a lot of background, one is object with little background to find the promising cells and region better.
2) Is there a way to better divide the grid so that we can separate multiple objects?