AMLI Summer 2022 Capstone Project Report

Student Detection, Tracking & Counting

07/29/2022

Ellion Dison, CSCE, elliondison@gmail.com

Yasser Hassan, EE, yasserhassan1738@gmail.com

Gabriel Young, CSCE, gabrielgabe06@gmail.com

**1 Abstract**

The MOT (Multiple Object Tracking) is an important tool in the modern world. It has various uses like object detection, counting objects, security tools ,etc. The goal of this project is to detect and count the number of students walking in and out a classroom. To accomplish this, we used object tracking implemented with YOLO, Deep Sort, Pytorch, and TensorFlow. You Only Look Once (YOLO) is a state-of-the-art algorithm that uses deep convolutional neural networks to perform object detections. We used yolov4 and yolov5. Compared to yolov4, yolov5 missed detections were signed reduced, and the number of target ID switches. We can take the output of YOLO and feed these object detections into Deep SORT (Simple Online and Realtime Tracking with a Deep Association Metric) to create a highly accurate object tracker. Deep Sort detects and tracks objects in videos frame by frame, and then predicts the target position. Also, unlike the general yolo object detection tool which detects all objects at the same time, this MOT system also detects only objects which are needed to be detected by the user and thus helps in improving the performance of the system.

**2 Introduction**

Deep sort is a multi-object tracking algorithm. Tracking is one of the necessary technologies needed for the upcoming world. Tracking can broadly be divided into multiple Object Tracking (MOT) and single object tracking. Multiple Object Tracking (MOT) plays an important role in solving many basic problems in computer vision. Tracking multiple objects in videos requires detection of objects in individual frames and combining those across multiple frames. Many Computer Vision techniques have been used to build MOT systems, and day to day the technology is growing rapidly providing an area of opportunities called image processing is done by providing a labeled dataset which is trained and is used as a model for the system which can detect objects in different frames comparing to the objects that was provided in the model by mapping the same pattern of model in the frame [2]. YOLO(“You Only Look Once”), OPENCV, PYTORCH, COCO dataset, TKINTER and GPU are the methodology used to detect, count, and track the objects in MOT. People tend to get object detection and object tracking confused. In object detection, we detect an object in a frame, put a bounding box or a mask around it and [classify the object](https://nanonets.com/blog/how-to-classify-fashion-images-easily-using-convnets/). Note that, the job of the detector ends there. It processes each frame independently and identifies numerous objects in that frame. Now, an object tracker on the other hand needs to track a particular object across the entire video. This work by taking an initial set of object detections and develops a unique identification for each of the initial detections and then tracks the detected objects as they move around frames in a video. If the detector detects 5 bikes in the frame, the object tracker must identify the 5 separate detections and needs to track it across the subsequent frames (with the help of a unique ID).

**The contributions and role of each member:**

**•** Video Manager: Yasser Hassan

• Program Manager: Ellion Dison

• Resource Manager: Gabriel Young

**1. Phase 1:**

Timeline: June 25, 2021 - July 2, 2021

Goals: Our team is composed of Ellion Dison, Gabriel Young, and Yasser Hassan. We will use a pre-trained model that detects and track humans in real time in a classroom. (I) Form the team; (ii) Define the capstone topic; (iii) Decide on who does what for the coming phases; (iv) Estimate the task and goal for the next Phases: fill out the content for Phase 2, Phase 3, Phase 4.

**2. Phase 2:**

Timeline: July 2, 2021 - July 09, 2021

Goals:

• Collect and modify the video data

– Yasser

• Reproduce the code

– Ellion

• Understand the input data so that you can run the code on your custom data

– Gabriel

• Investigate ethical implications and impacts of our model.

– Yasser

**3. Phase 3:**

Timeline: From July 9, 2021 - July 19, 2021

Goals:

• Load yolo models and deep sort repositories

– Ellion

• Consider Results and determine if a better model is needed

– Yasser

• Test live video on webcam

– Ellion

• Prepare for mid-project presentation

– Gabriel

**4. Phase 4:**

Timeline: July 19, 2021 - July 28, 2021

• Final review of the model and prepare live demo

– Ellion

• Prepare data and ethics analyses for presentation

– Yasser

• Edit and review Capstone Report

– Gabriel

• Prepare results analysis for presentation

– Gabriel

**3 Project Implementation**

Our project was implemented in 5 steps. First, we took a video of the class and reduced the video quality, cut out sound, and reduced its size. Second, we determined which tracking method to use and which backbone to use for it. Third, we wrote and repurposed code for our chosen model and tracking algorithm. Fourth, we used our models and tracking algorithm on our recorded and live video, then adjusted it until the result desired was achieved. Lastly, we present the results of our tracking algorithm. At the end, there was discussion about the challenges that we encountered during the implementation of these pretrained models.

**4 Experimental Results**

**4.1 Testing**

We used the yolov4 and yolov5 pre-trained models. We used different confidence thresholds between 30 and 60 percent. We began testing these models with different videos of people online and with videos of our classroom. Then we used a live webcam these models with deep sort on the class.

**4.2 Evaluation**

Yolov4 performed well on videos fed to the model. There was a case when it detected a computer as a person while their hand was on the computer. When testing yolov4 on the webcam, the video was very delayed due to each frame being processed at 1.18 to 1.30 frames per second. Yolov5 performed better and faster than yolov4. Yolov5 had less ID switching in the video and the time to process a video was reduced by 15-20 minutes. The webcam stream was a lot smoother since each frame was being processed from .1 to .3 seconds.

**4.3 Results**

Our deep sort tracker was able to accurately track and count the number of students in the classroom using yolov4 and yolov5. Yolov5 was a lot faster than yolov4 at processing videos/live feed. A confidence threshold of 50 percent was good for filtering out possible false detections. Overall, yolov5 had less switching and it was a little more accurate than yolov4 at detecting and tracking humans.

**5 Conclusion**

My team was able to use object detection and object tracking to accurately track students going in and out of our classroom. We implemented object tracking using deep sort with yolov4 and yolov5 as the backbone. Both models performed well on recorded videos but yolov4 did not work well the live webcam. Overall, yolov5 was the better model due to less switching, more accuracy, and faster video processing speeds. Deep sort is one of many object tracking algorithms. To improve our results, we could try using other object trackers for human detection like Byte Track or ROLO. Also using a computer with better CPU and GPU would also improve the performance of our tracker.

**6 Acknowledgement**

First, we want to express our gratitude for this chance. This summer's internship has been fantastic in every way. Collaborative projects, group projects, and the capstone process have all taught us a great deal academically. The bonds we created will leave a permanent imprint on our life. It was interesting to experience the environment within the classroom and it made it easier to get to know the instructors and teaching assistants. Each professor who gave a lecture to us throughout the summer made us feel quite at ease, and we liked getting to know them. We would like to express our gratitude to Dr. Le, our capstone lecturer, for helping us through the entire process and offering comments. She was a strong advocate for a better final product. We also would like to thank NACME and Google for funding this program and the University of Arkansas for hosting us.

**References**

Krishnan, K., Baby, B., & Viji, K. S. (n.d.). *Multiple Object Tracking using Deep Learning with YOLO V5*. Retrieved July 29, 2022, from <https://www.ijert.org/research/multiple-object-tracking-using-deep-learning-with-yolo-v5-IJERTCONV9IS13010.pdf>

Guy, T. A. (2022, May 22). *yolov4-deepsort*. GitHub. <https://github.com/theAIGuysCode/yolov4-deepsort>

Jocher, G. (2020, August 21). ultralytics/yolov5. GitHub.

<https://github.com/ultralytics/yolov5>

nwojke. (2019, April). nwojke/deep\_sort. GitHub.

<https://github.com/nwojke/deep_sort>