Sotheanith Sok

CECS 553

10/29/2021

Report 2

1. What have you accomplished this week? (Specify what each person has done)

Ans: For this week, I have completed the following:

- Decide on the possible class.
  - Current classes
    - Tom,
    - Jerry
    - Spike
- Collect images
  - Current class distribution
    - Tom: 562
    - Jerry: 490
    - Spike: 538
- Annotate images

2. Any challenge you faced during this week? If so, how are you planning to resolve it? Any solutions or ideas?

Ans:

Initially, I want to use a single video from the Wile E. Coyote & Road Runner series as the dataset. I plan to have two classes: Wile E. Coyote and Road Runner with a minimum of 125 images per class. The train, validation, and test datasets will be frames randomly extracted from the video and the resulting model will be used to detect characters over the entire video. Unfortunately, after collecting the dataset and examining it, there are multiple apparent issues. To start with, there is a huge classes imbalance problem where there are approximately three times more Wile E. Coyote samples and null samples than Road Runner samples. Then, there are problems with the characters themselves. In multiple instances, characters are either obscure by something or they are too small and thus, they lack details.

To quantify the problems, I decide to train a YOLOv5 model with the above dataset and it performs poorly. During the testing phase, the model achieves around 0.65mAP when detecting Wile E. Coyote and 0.23mAP when detecting Road Runner. When the model is used to detect characters over the entire video, it tends to misidentify characters in dimly lit scenes or when fast-moving objects are presented in the scene.

In order to overcome the above problems, I decided to re-evaluate dataset choice and size. To start with, instead of a single video from Wile E. Coyote & Road Runner, I will use two videos from Tom & Jerry series where one video will be used to generate train, validation, and test datasets and the other video will be used to demonstrate the model detection capability over a video. In addition, I also increase the number of classes by one and thus, the current classes are Tom, Jerry, and Spike.

Last but not least, the size of the dataset has been increased to 1000 and it is composed of approximately 500 samples per class and minimal null samples.