Clara Rovira Noëlle

Q_1

1. What do you think it occurred during this model development (trainning & evaluation)?

One of the most common problems that occur when using deep learning models is data quality. Poor data quality directly affects the model performance. One of the common causes can be the following:

- There are no high-quality labeled datasets
- Images or videos used for training are low-definition
- Not enough training data for the required task

Also, in computer vision, one of the biggest challenges is scene understanding. It is not enough to identify a single object, the model must understand how the objects in the scene are related. In this specific case, the bald head and the ball were visually alike in the video. As a consequence, if the model structure together with the poor data quality is not well structured, the model can be confused. In case the model had been constructed considering not only the main object (the ball) but also the context, and its environment, maybe it wouldn't have occurred.

2. How would you fix this behavior? Please provide at least 2 options explaining their pros and drawbacks

1. Regarding the data problem, in case there is not enough training data, which is a common problem, a possible solution could be to add synthetic data to the dataset. A model using real-world and synthetic data has more probabilities to improve the model performance and accuracy.

Pros: it is easy and cost-effective to create. Also, ensures privacy and confidentiality of the data which is a common limitation in the obtention data process.

Cons: the quality of synthetic data is highly dependent on the quality of the model that created the synthetic dataset. Moreover, a verification process is necessary to compare the outputs generated by using the real and synthetic data.

2. Regarding scene understanding, there exist some datasets for scene understanding (COCO dataset) that can be used as an initial point of reference to training the model.

Pros: It is a dataset with 330K images and dozens of classes already labeled and ready to use. Also, it is constantly improving. Finally, it has been mentioned to be the reference dataset for common perception tasks that can help to improve the model.

Cons: This dataset can't be directly used, and a transfer learning process would be necessary.

3. Extra: Do you know any tracking algos (Deep learning based) that could be used here?

I could mention the CNN networks and their variants such as the R-CNN, Fast R-CNN, and the Faster R-CNN. Also, I know the YOLO algorithm. However, I haven't used them, only know that they are used for object tracking and detection.

Q_2

4. Which algorithms would you use to achieve this? Please provide at least 2 proposals

Some algorithms that combine object detection & classification would be R-CNN, fast R-CNN, faster R-CNN, YOLO or SSD.

5. Which potential bottlenecks may this problem encounter? (i.e. non-static camera). Just list them (if any)

- brightness change
- ability to expand and shrink the object (zoom in and zoom out the scene)
- distorted objects due to noise
- rotation of the object (usual players are standing but sometimes they can be laid or rotating on the floor as a consequence of a fall)

6. Could your proposals be scaled to other team-based sports? (i.e. hockey, rugby, ...) Which would be worst-case scenarios?

The worst-case scenarios could be the ones that are less similar to football. For example, badminton requires an object on the hands of each player and also a net in the middle, it could confuse the model. Furthermore, in rugby, most of the time, players are in physical contact and players wouldn't be fully detected. Also, the model could interpret some player from the opposite team because he or she is over the player