Course Project Part 3: Dataset Plan

Hamza Aziz
University of Saskatchewan
Saskatoon, Saskatchewan
hamza.aziz@usask.ca

Abstract

A project dataset plan for the SoccerNet Tracking competition that is associated with the 9th International Workshop on Computer Vision in Sports(CVsports) at CVPR 2023.

1. Introduction

Sports have for a long time now, have played a major role in both society as a whole, as well as how many individuals enjoy their lives. Whether it is playing or spectating, sports have the ability to really grasp the hearts of numerous people. With people's accessibility and desire to watch sports online increasing rapidly in recent years, so does the effort in using technology to enhance the viewing of sports. With the visual displaying for sports becoming more dynamic, it has made it more difficult to track objects (players, balls, etc.) and accurately identify them, in order to collect useful data. This data is important for the benefit of the players and teams, in order to allow them to review their performances and improve upon themselves to achieve a higher level of gameplay. Not only would this higher level of gameplay enhance the viewing experience for spectators, but the data itself allows for viewers to compare players, teams, and the games, adding another dimension to the viewing experience.

For the previously stated reasons, I wanted to work on a project for the SoccerNet Tracking competition, that is associated with the 9th International Workshop on Computer Vision in Sports(CVsports) at CVPR 2023. The task for this competition is Multiple Object Tracking (MOT), which expects competitors to both detect all objects of interest (players, goalkeepers, referees, staff, ball) and make the association.

The dataset used for this competition comes from a dataset called SoccerNet-v2, which has data collected from 12 complete soccer games, only using the main camera, which is not easy to find. As well as having the full 12 soccer games recorded (during the 2019 Swiss Super League), there is also 200 clips of 30 seconds each, and a complete halftime, both annotated with tracking data. The 30 second clips in particular are clips of key moments, or moments that would be harder for tracking to perform well on, such as corners, fouls, goals, free kicks, penalties, etc. These clips can be difficult to track multiple objects in due to the objects being closer together, sometimes overlapping, or certain fast paced motions of both the objects and the camera itself. "Note that a subset of this data is used in this first challenge. In particular, this accounts for 57 30-seconds clips for the train set, 49 clips for the test set, 58 clips for our first public challenge, and 37 clips for our future challenges, including the entire half-time video in the latter." [1]. The first public challenge refers to the 2022 iteration of this competition, and so the 2023 version of this competition includes the 37 clips for the challenge set.

Figure 1: Sample image of a corner kick clip from the SoccerNet-v2 Tracking dataset.

Each 30 second clip/sample is stored as 750 jpg images, which is 25 frames per second. Each jpg image has a dimension of 1920x1080, and the 750 images together take a size of 127 mb. The images do not contain any components usually displayed on TV, like advertisements, details of the score and time, as the data collected is directly from the main camera used for these games, as shown in Figure 1. Along with all the pictures, each sample also has two INI files called "gameinfo" and "seqinfo". The seqinfo files have information stored that define attributes of the 30 second clip, such as framerate, sequence length, image size measurements, and image data type. Within the gameinfo files, labels for the sample are stored. Some of the labels are the name of the sample, gameID, both the game time start and end, visibility, and so on. One of the labels is called "actionClass", which defines what the action within the 30 second clip is. For example, the actionClass might be a foul, or a corner, or a penalty kick, etc. The gameinfo file also contains a label for the number of unique tracklets throughout the 30 second clip, and a label exists for each tracklet ID going from 1 to x, x being the number of tracklets. The gameinfo file is only present in samples within the train and test set, and not within the challenge set.

Each clip is also stored along with the ground truth and detections for the bounding boxes of the multiple objects, both being stored in comma-separate csv files. These csv files have 10 columns, “These values correspond in order to: frame ID, track ID, top left coordinate of the bounding box, top y coordinate, width, height, confidence score for the detection (always 1. for the ground truth) and the remaining values are set to -1 as they are not used in our dataset, but are needed to comply with the MOT20 requirements.” [1]. The ground truth and detections data is only present with samples in the train and test sets, not in the challenge set. The 2022 dataset kept the detections data in the challenge set, as it is essentially all the bounding boxes ground truth without identifying the tracklets, allowing for competitors to focus on association. Using this data and focusing purely on association is also an acceptable task for the competition this year, as it was the main task of the previous year’s competition for CVsports at CVPR 2022, but this project will also be focusing on detecting the multiple objects on its own as well as associating them, as the new version of the SoccerNet Tracking competition states. The SoccerNet-v2 dataset allows for a large variety of use, but there are also other dataset’s out there with the purpose of soccer video understanding.

One of the similar datasets to SoccerNet-v2, is “SoccerTrack” a Dataset for soccer footage recorded with drones and fish-eye cameras [2]. The purpose of SoccerTrack is a solution to more accurately track objects on a soccer field and collect data. Since they use a camera with a fish-lens attached to a drone above the soccer field, it allows them to always view the entire pitch and all of the multiple objects. Another similar dataset is called SoccerDB which contains 171,191 video segments from 346 different high-quality soccer games [3]. This competition uses the second version of SoccerNet which has added more classes of actions (original SoccerNet only had goal, yellow/red card, and substitution), more annotations and has made a huge focus public tasks with reproduceable benchmarks for each one, to allow for events(like this competition) which would allow the public to help improve both the dataset and the technology of this area. There is also a third version of SoccerNet (SoccerNet-v3 which is not used for these events yet [4]. SoccerNet-v3 try’s to connect the replay sequence which comes at a different camera angle, to the original live action frame

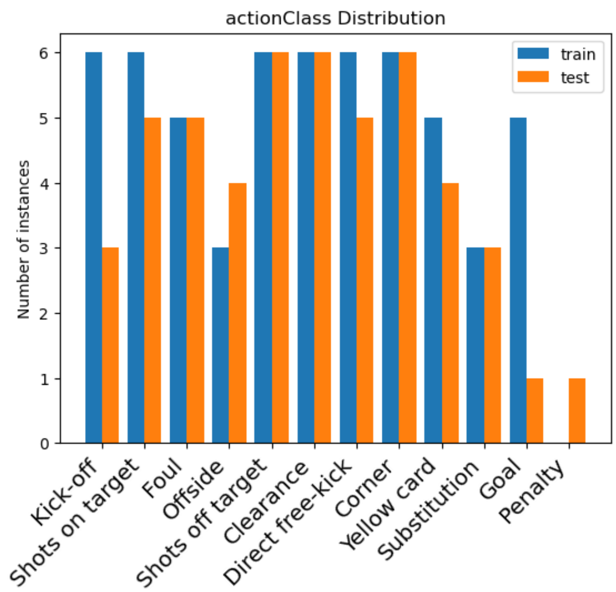

Figure 2: Bar chart representing the distribution of actionClass labels for the train and test sets.

from the main camera, which includes each of the two having the lines and goals annotated and players framed within a bounding box, and then having the all the annotations and bounding boxes in the two different views linked. It is still a work in progress, but it is trying to utilize the improved techniques from people attempting the tasks for SoccerNet-v2.

2. Visualization of The Data Labels

When analyzing the labels for the dataset, one of the first things that was considered was the “actionClass” label. One of SoccerNet-v2’s biggest improvements was the increase of action labels, originally only having classes for goals, yellow/red cards, and substitutions. I wanted to investigate both what the new classes were in SoccerNet-v2, as well as the distribution of these actionClass labels in the testing and training sets given for this tracking competition. This is important to better understand what actions are occurring in the samples given for the competition, as well as to ensure there is a relatively even distribution between the testing and training sets. As displayed in Figure 2, the total number of unique actionClass labels went from 3 in the old SoccerNet dataset, to 12 in the given dataset for this competition from SoccerNet-v2. The distribution of labels within the train and test set’s are fairly evenly distributed, with the goal and penalty actionClass labels being the exception. The test set only has one goal and one penalty sample, and the train set does not have any penalty samples. However since the focus of the competition is tracking and not action spotting, this is fine.

Another label that was considered was the “visibility” label. At first what this label suggested was unclear, and the initial assumption made was that some of the samples might not have as clear visibility of the pitch and objects as other samples, causing these samples to be difficult to deal with when tracking multiple objects. Upon further investigation, out of the 57 samples in the train set and the 49 samples in the test set, almost all the samples had a visibility label equal to “visible”. Two samples in the train set and two samples in the test set had their visibility label equal to “not shown”. Once these samples were investigated, it showed that the visibility refers to the main action of the clip. For example, in a clip where the main action is an offside call, if the player is not visible from the camera when the incident occurs, the action is “not shown” and thus the visibility label is defined. This consideration of when the objects are in the clip and how many lead to the analysis of the “num_tracklets” label. In the “gameinfo.ini” file, the num_trcklets label defines how many unique objects ever entered the cameras view during the clip, and each of these tracklets also have an ID stored in the file. On average throughout the test and train sets, 24.714 unique tracklets entered the camera’s view in the samples, with a median value of 25. The clip with the least amount of tracklets had 20, and the most tracklets in any of the clips was 31. Finally, the last statistical analysis performed on the dataset related to the labels, was the analysis of the distribution of bounding box shapes.

Using the ground truth files given with each sample within the testing and training set, Figure 3 was created and represents the average size for all bounding boxes, for both sets individually, with the bounding boxes for the balls separated from all other bounding boxes. The figure shows that the bounding boxes for the balls are significantly smaller, which suggests that the bounding boxes are being properly wrapped around the detected object. The bounding boxes in the test set seem to be on average larger than the train set’s bounding boxes. This can make sense because the train set has more clips than the test set, and the objects(players) spend most of their time is a standing position, meaning their bounding boxes will mostly be closer to being the largest they will ever be, with incidents like slide tackles or crouching being less common, where the bounding boxes would get smaller. Thus, I conclude that there are no apparent issues with the database related to the labels.

3. Visualization of The Data Inputs

The input data for the Soccernet-v2 dataset is very consistent according to the publishers. In particular, the input data for the SoccerNet Tracking competition’s dataset, has consistent image characteristics. All samples have the same .jpg data type, a frame rate of 25, and an image size of 1920x1080. This data comes from the

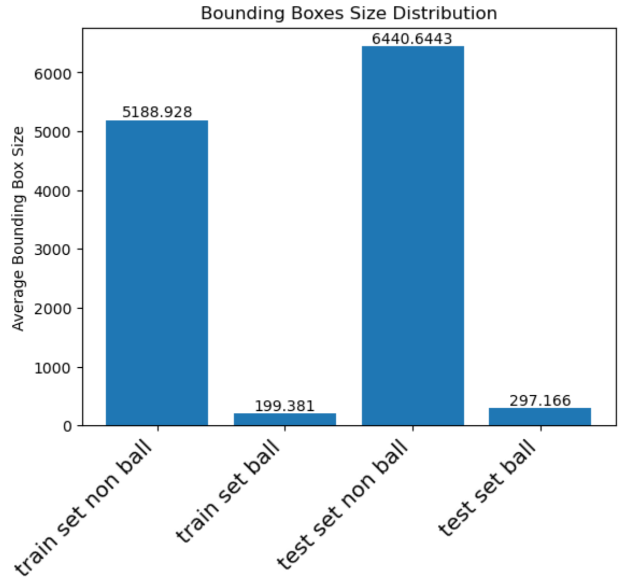

Figure 3: Bar chart representing the distribution of the Bounding Boxes for both the training set and testing set. Bounding Boxes for the ball are separated. The average sizes are calculated via width * height.

“seqinfo.ini” file as mentioned before, so I decided to perform an analysis to confirm that all the input data was of the correct and consistent format as the publishers stated. Figure 4 shows that all the images are size 1920x1080, and also are colour images. It even shows that the frame rate for the clips are correct, since the train, test and challenge set have a total of 143 clips, and 30 seconds each with a 25 frame rate, equals a total of 107250 images, which is the amount I found in the sets. The image size was also confirmed directly from the images, instead of the “seqinfo.ini” files, and the files were required to have “,jpg” type, so that also confirms all the image files were the correct data type.

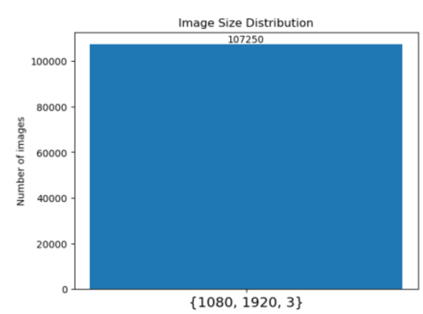

Figure 4: Bar chart representing the distribution of the image sizes, for both the test, train and challenge set combined. (There is no distribution, they are all the same, which is what the publishers of the dataset intended)

Using a clustering analysis approach when analyzing the data inputs is difficult due to there not being different dimensions that can be compared and are the same data type (like multiple different integer values that have a comparable meaning). Since the clips within the dataset are consistent with image characteristics, and the reason stated above, there is not really a feasible cluster analysis approach that would make sense for this dataset. Instead, I decided to analyze the distribution of samples among the 12 complete soccer games in the SoccerNet-v2 dataset. Figures 5 and 6 show that within both the test and training set, the games which the samples originate from(indicated by the gameID) are evenly distributed, however the training set has samples from game's 4,6 and 9, while the test set has samples from games 7,8 and 11. This means none of the samples from the training set come from the same game as any of the samples in the test set. This makes me assume that the case is the same for the challenge set as well. This is likely intended, so that the processes between the training, testing, and challenge stages are all properly separated.

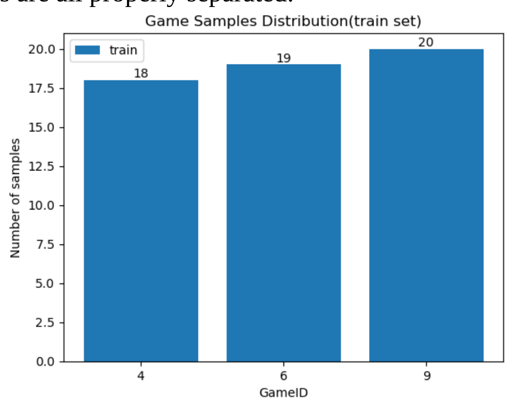

Figure 5: Bar chart representing the distribution of the gameID's for the samples of the train set.

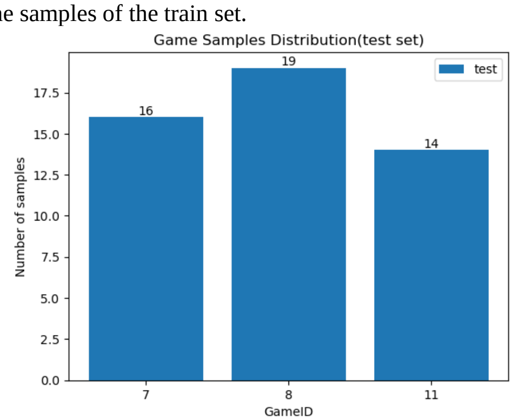

Figure 6: Bar chart representing the distribution of the gameID's for the samples of the test set.

4. Selecting a Smaller Sub-dataset

For the smaller sub-dataset, I decided to take samples from the original set based on the actionClass labels. Specifically, my subset will take 1 sample of a corner action from the test set, and 1 sample of a corner action from the train set. The sample names are "SNMOT-167" from the training set, and "SNMOT-125" from the testing time. The corner action events are debatably some of the more difficult ones, as they are most likely to have more objects in the camera view, as well as many objects being close together and overlapping. The Corner actionClass samples are also well distributed between the testing and training sets, so both sets will be evenly balanced. Since this smaller sub-dataset will have a combination of small number of balanced samples, as well as samples difficult for multiple object tracking, it will be a very useful dataset and efficient dataset for quickly testing models and hyperparameters. I considered using all 12 of the corner samples, but the smaller sub-dataset wants to be within the ~100's of images range, and each sample takes 750 images, so more than two samples would be too many for the smaller sub-dataset. Figure 7 shows the bounding boxes size distribution of the two selected samples, and Figure 8 shows the new game samples distribution of the two samples.

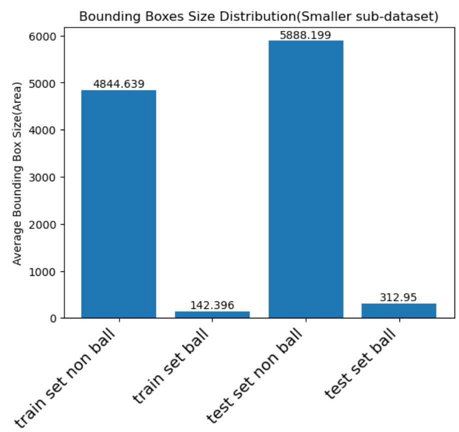

Figure 7: Bar chart representing the distribution of the Bounding Boxes for both the training set and testing set(smaller sub-set version). The average sizes are calculated via width * height. The balls in this figure are the only two since there is only one sample in the train set, and one sample in the test set.

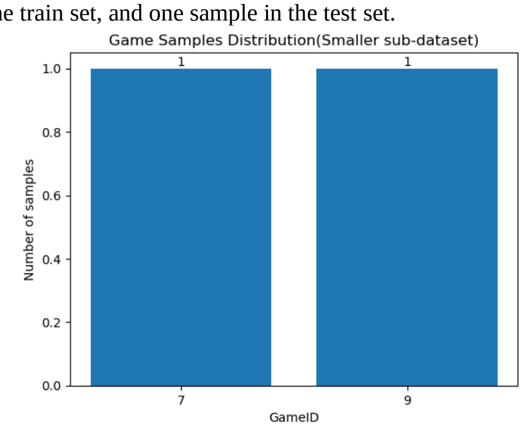

Figure 8: Bar chart representing the distribution of the gameID's for the samples of the smaller sub-dataset. The sample from gameID 7 is in the test set, and the sample from gameID 9 is in the training set.

References

[1] Anthony Cioppa, Silvio Giancola, kangle, mars. (2022). SoccerNet - Tracking. https://github.com/SoccerNet/sn-tracking

[2] Atom Scott, Ikuma Uchida, Masaki Onishi, Yoshinari Kameda, Kazuhiro Fukui, Keisuke Fujii. (2022). SoccerTrack: A Dataset and Tracking Algorithm for Soccer With Fish-Eye and Drone Videos https://github.com/AtomScott/SoccerTrack

[3] Yudong Jiang, Kaixu Cui, Leilei Chen, Canjin Wang, Changliang Xu. (2020). SoccerDB: A Large-Scale Database for Comprehensive Video Understanding. https://arxiv.org/pdf/1912.04465.pdf

[4] Anthony Cioppa, Adrien Deliège, Silvio Giancola. (2022). SoccerNet-v3 https://github.com/SoccerNet/SoccerNet-v3