Course Project Part 2: CVPR Project Proposal

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

Abstract

A project plan proposal for the SoccerNet Tracking competition that is associated with the 8th International Workshop on Computer Vision in Sports(CVsports) at CVPR 2022.

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 had 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 8th International Workshop on Computer Vision in Sports(CVsports) at CVPR 2022. The 9th iteration of this workshop also has this competition, but further details have yet to be released, so for now I am proposing for the competition in the 8th iteration but may adapt to the new competition if details are released soon. The competition has two stages/tasks, with the first task is focused on association using a subset of data. Given ground-truth detections of the players and ball, we must associate what is detected, the second task has a focus on implementing both detection and association.

This competition is difficult there are several factors that have to be dealt with when tracking in soccer. A lot of the players are similar, making I third to accurately identify them. Throughout all the clips, there are many times where a player will leave and re-enter the scene, every time that requires them to be reidentified. The cameras in soccer games show many individuals who are not meant to be tracked, such as fans, referee's, ball boys, players on the bench, as well as coaching and medical staff. Distinguishing who we want to track and who we don't want to track can be difficult at times. The recorded video clips of soccer gameplay include fast motion, as well as scenarios where vision of the objects is obscure, which make both detection and association difficult.

This task being performed is specifically for soccer, which would make it initially hard to be used in other cases, for example in other sports. However, it could be adjusted to work for similar scenarios, like for rugby or hockey, which have a similar set up in the sport, but the solutions would have to be unique to the sport/setting.

The problem itself for the competition is a classification problem, but the classes are not considered for the competition's evaluation metrics. As stated earlier, in the first challenge/task we would be provided with a subset of clips that have all the objects already detected, they just have to be identified at every moment of the clips, and in the second challenge they must be detected as well as identified. For both tasks, the performance would be measured by comparing the output with a predefined ground truth (stored in comma separated csv files) to calculate a HOTA (Higher Order Tracking Accuracy) value. HOTA is a performance measurement that aims to determine the balance between sub metrics, specifically DetA (Detection Performance) and AssA.(Association Performance). The performance can be evaluated by competitors at any time on the challenge site:

`https://eval.ai/web/challenges/challenge-page/1539/overview`

However, there are limitations on how often you are allowed to submit an evaluation.

The dataset used for this competition comes 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 1019 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, penalties, etc. These clips can be difficult due to objects we are trying to detect and identify being closer together, 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].

In terms of the feasibility of this potential project, since the development kit for the competition is provided, which includes three baselines to improve on (DeepSORT, FairMOT, ByteTrack), I believe the project is quite achievable with reasonable performance. From what I have seen, the development kit provides code to start with for detection, but most of the higher achieving competitors use their own detected system to get better performance, so it may be difficult to compete without doing so as well. If I want to adapt to submit for the 2023 competition once it is released (it is confirmed, but without further details), there is a risk of it being either too late to adapt or being potentially not feasible due to changes in the competition.

For potential required resources to complete the project, for reference, the second placed solution for the SoccerNet Tracking Challenge in 2022 mentioned in their paper that “Due to the simplicity of our C-BIoU tracker, it can reach 680 FPS on the testing set, by using an Intel Xeon Silver 4216 CPU. For our offline processing, extracting the appearance feature from images costs about 40 mins on the testing set (on a single Nvidia GeForce 1050 GPU)” [2]. Considering that the testing set is about a quarter of the total 30 second clips, this means a similar solution would be feasible, and if we guessed a total of 8 training runs with the training set, which is only slightly larger than the test set, than this would only take less than 8 hours for computing time when training the model. Another solution in on GitHub for this competition mentioned needing “Google Colab Pro for GPU for background execution. Google Drive Storage for 200 GB.” for compute resources [3]. However, it is hard to say my potential solution would have similar compute times or space requirements, since not many of the previous submissions for this competition made public articles.

2. Existing Competition

During the 2022 version of the SoccerNet Tracking competition, which was the first tracking competition SoccerNet hosted, the competition had evaluation servers hosted on “eval.ai”. The test set was blind, but after a certain date, participants were able to use the evaluation server on the challenge set, before having the challenge set determine the winners later on. The competition had 12 competitors, with the average final HOTA score being 84.7858, the standard deviation being 13.775 and the winning final HOTA score was 93.64 (for reference, the second highest score was 93.25, the lowest final score was 51.03, and the baseline score was 70.79. Only 2 participants got a score lower than the baseline score).

Of the sub metrics for HOTA that evaluate detection and association independently, everyone besides the two lowest ranked competitors got a DetA (Detection) score of more than 99.5 (the baseline was 82.97), but the AssA (association) scores varied a lot more, and generally were not as high, with the highest score being 88.06 (the baseline was 60.68) and being the determining factor for how the competitors placed in the final rankings. This suggests that implementing for association is more difficult than for detection with the current common approaches for this competition’s problem. The competition is archived, but the evaluation server does not have an ending date until 2099, allowing people to evaluate their solutions even when there is not an active competition currently live.

2.1. Analysis of An Existing Solution

The 2022 SoccerNet Tracking competition’s top winner and runner up both have publicly submitted their paper’s for the competition, however both of these two papers did not provide the code to their solution. I did manage to find a public GitHub repository for a project for this tracking competition at the following link:

```
https://github.com/adityaabothra/SoccerNet_V2_MultiPlayerTracking
```

In their solution, they used the provided FairMOT baseline to base their solution on. They cloned the following GitHub repository:

```
https://github.com/PaddlePaddle/PaddleDetection.git
```

which is another development toolkit for Object Detection. It seems they used the baseline FairMOT which is a baseline for one-shot multi-object tracking, and this 3rd party development toolkit called PaddleDetection for detection. They used the SoccerNet Tracking competition’s evaluation script provided in the development kit to test and print the scores, and their results don’t seem to be too far from the competition’s baseline evaluation values. The reason they used the FairMOT baseline was because it is “Anchor free, State of the Art Accuracy, Integrated Reld, Transfers well” [3]. I attempted to run their solution through Google Colab, but was not able to due to issues with Google Drive Storage space. I would need to create an account with no preexisting data (or move the preexisting data on my account) and pay for the space required. I decided not to do

so since their code was saved in a .pynb file so I could see the output already.

3. Past Research Papers

The following sections contains summaries on three different relevant research papers created in past years. The first paper is covered in more depth as it is the most closely related.

3.1. The Second-place Solution for CVPR 2022 SoccerNet Tracking Challenge

"The Second-place Solution for CVPR 2022 SoccerNet Tracking Challenge" manages to accomplish the SoccerNet Tracking Challenge by creating two separate phases of tracking [2]. First they have a short term online tracking system, and then in the second phase they have an offline long-term tracking system. In the online tracking, they solely used geometrical features when created short-term tracklets. They replaced IoU, which is "commonly used to calculate cross-frame geometric affninity between observations", with a Buffer-IoU (BIoU) which "adds buffers that are proportional to the original box size for calculating IoU scores" [2]. Adding this buffer around the box size, allows for better detection when dealing with objects that can be moving very fast and erratically, like soccer players.

With the offline tracking, to deal with objects leaving and re-entering the screen, they needed to recover long-term tracklets. To do this they used appearance features and hierarchical clustering. They used training data from the SoccerNet 2022 re-id challenge to initialize their re-id model, and then at the end applied hierarchical clustering to "cluster short-term tracklets to long-term ones". They were able to get a good HOTA score of near 90 with just the online tracking, but adding the offline tracking allowed them to enhance their performance and get a HOTA score of around 93.2.

An issue the authors of this paper noticed that may have negatively affected their results, is that in the SoccerNet-Tracking dataset, there were some instances where the bounding boxes were not accurate. There were also 2 ideas they explored upon to improve their results, and may be a possibility for future exploration:

-Removing camera motions for tracking

-Performing tracking on a Bird's-Eye-View (BEV) soccer field

3.2. Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space

"Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space" is an article written by the same authors as the last section and is in fact an article specifically for going more in depth on the Cascaded Buffered IoU (C-BIoU) that was used for their SoccerNet Tracking Challenge submission [4]. The solution itself is quite straightforward in that it simply "adds buffers to expand the matching space of detections and tracks, which mitigates the effect of irregular motions in two aspects: one is to directly match identical but non-overlapping detections and tracks in adjacent frames, and the other is to compensate for the motion estimation bias in the matching space" [4]. The C-BIoU was also a key factor in their submission in the ECCV'22 MOTComplex DanceTrack challenges.

There is a new component to the C-BIoU tracker that was not present in their SoccerNet Tracking Challenge submission, and that is the use of cascading matching. Basically this is done by "first matching alive tracks and detections with a small buffer, and then matching unmatched tracks and detections with a large buffer" [4]. This helps reduce of the risk of adding too large a buffer to the matching space.

3.3. SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos

"SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos" won the best paper award at the 8th International Workshop on Computer Vision in Sports (CVsports) at CVPR 2022 [5]. This paper is also written by the people hosting the SoccerNet-Tracking challenge at the workshop. Their paper talks about the importance of tracking multiple objects in soccer, and release the novel dataset called "SoccerNet-Tracking dataset", which is the dataset used for the competition my project is being proposed to be involved in. The paper covers the need and potential impact of a strong multiple object tracking (MOT) system. They also discuss basic MOT approaches like DeepSORT, Tracktor, FairMOT and ByteTrack, including both the benefits of each one, implementation details, and different evaluation metrics. They also talk more in depth about the SoccerNet-Tracking dataset, particularly what data they have, the data format, and a reminder that this is the current best dataset in terms of quality and size for trying to train and test consistent tracking of multiple objects. This paper may be the most critical for me, as it will help me better understand the data itself, the problem, potential starting points for a strong solution, and more.

References

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

[2] Fan Yang, Shigeyuki Odashima, Shoichi Masui, Shan Jiang. (2022). The Second-place Solution for CVPR 2022 SoccerNet Tracking Challenge. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://arxiv.org/pdf/2211.13481v2.pdf

[3] Tyler Bann, Aditya Bothra. (2022). SoccerNet_V2_MultiPlayerTracking. https://github.com/adityaabothra/SoccerNet_V2_MultiPlayerTracking

[4] Fan Yang, Shigeyuki Odashima, Shoichi Masui, Shan Jiang. (2023). Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. https://openaccess.thecvf.com/content/WACV2023/papers/Yang_Hard_To_Track_Objects_With_Irregular_Motions_and_Similar_Appearances_WACV_2023_paper.pdf

[5] Anthony Cioppa, Silvio Giancola, Adrien Deliege, Le Kang, Xin Zhou, Zhiyu Cheng, Bernard Ghanem, Marc Van Droogenbroeck. (2022). SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://openaccess.thecvf.com/content/CVPR2022W/CVSports/papers/Cioppa_SoccerNet-Tracking_Multiple_Object_Tracking_Dataset_and_Benchmark_in_Soccer_Videos_CVPRW_2022_paper.pdf