7. Supplementary Material

7.1. Annotation Guidelines

We provided our annotators with the following annotation guidelines to annotate the actions and the camera shots.

Actions. Following the original SoccerNet [24], we annotate each action with a single timestamp. These actions are illustrated in Figure 9, and their timestamps are defined as:

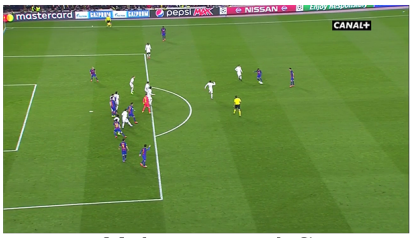
- Ball out of play: Moment when the ball crosses one of the outer field lines.
- Throw-in: Moment when the player throws the ball
- Foul: Moment when the foul is committed
- Indirect free-kick: Moment when the player shoots, to resume the game after a foul, with no intention to score
- Clearance (goal-kick): Moment when the goalkeeper shoots
- Shots on target: Moment when the player shoots, with the intention to score, and the ball goes in the direction of the goal frame
- Shots off target: Moment when the player shoots, with the intention to score, but the ball does not go in the direction of the goal frame
- Corner: Moment when the player shoots the corner
- Substitution: Moment when the replaced player crosses one of the outer field lines
- Kick-off: Moment when, at the beginning of a half-time or after a goal, the two players in the central circle make the first pass
- Yellow card: Moment when the referee shows the player the yellow card
- Offside: Moment when the side referee raises his flag
- Direct free-kick: Moment when the player shoots, to resume the game after a foul, with the intention to score or if the other team forms a wall
- Goal: Moment when the ball crosses the line
- Penalty: Moment when the player shoots the penalty
- Yellow then red card: Moment when the referee shows the player the red card
- Red card: Moment when the referee shows the player the red card

Camera shots. We define the following 13 types of cameras, illustrated in Figure 10:

- Main camera center: Camera shown most of the time. It is placed high in the stadium and is centered on the middle field line. It films the players with a wide field of view, allowing an easy understanding of the game.
- Main camera left: Camera placed high in the stadium on the left side of the field. It is mostly used to allow for an easy overview of what is happening close to the left goal. It can also be used sometimes to show the right side of the field from a further perspective, mostly for an artistic effect. It is also sometimes called the 16-meter left camera.
- Main camera right: Counterpart of the main camera left but on the right side of the field.
- Main behind the goal: Camera placed behind the goal, either on a moving crane or in the stadium. It allows for a perpendicular field of view compared to the other main cameras.
- Goal line technology camera: Camera often placed next to the main camera left or right, but aligned with the goal line. It is used to check if the ball entirely crosses the line in contentious goal cases.
- Spider camera: Camera placed above the field and able to move freely in 3 dimensions thanks to long cables. It is often used in replays for a dynamic immersion in the action.
- Close-up player or field referee: Camera placed on ground-level, either fixed or at the shoulder of an operator, filming the players or the referees on the field with a narrower field of view.
- Close-up side staff: Located similarly to close-up player cameras, films the reaction of the coaches and the staff outside the field. This also includes players on the bench or warming-up.
- Close-up corner: Camera often on the shoulder of an operator filming the player that shoots the corner.
- Close-up behind the goal: Camera either on the shoulder of an operator or fixed on the ground and filming the goalkeeper or the players from behind the goal.
- Inside the goal: Camera placed inside the goal that is sometimes shown during replays for an artistic effect.
- Public: Camera possibly located at different places in the stadium with the objective of filming the reaction of the public.
- Other: all other types of cameras that may not fit in the above definitions and that are most often used for artistic effects (*e.g.* the helicopter camera or a split screen to show simultaneously two different games).

Figure 9. **Actions.** An example of each action identified in SoccerNet-v2.

Main camera center

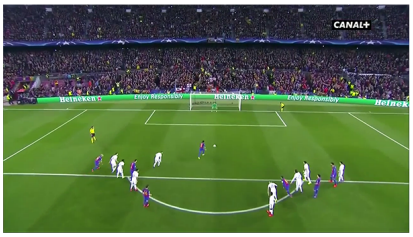


Main camera left

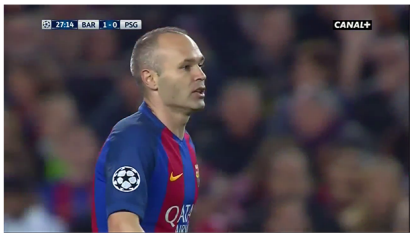
Main camera right

Main behind the goal

Goal line technology



Spider camera



Close-up player or field referee

Close-up side staff

Close-up corner

Close-up behind the goal

Inside the goal

Public

Other

Figure 10. **Cameras.** An example of each camera shot identified in SoccerNet-v2.

7.2. Annotation Process

We developed two tools for the annotations: the first for annotating the actions, shown in Figure 11, the second for the camera changes and replay grounding, shown in Figure 12. For each video, a .json annotation file is created, which constitutes our annotations. The structure of the .json file is illustrated hereafter.

Listing 1. Example of an action annotation in json.

```
"UrlLocal": "path/to/game",
"annotations": [
    {
        "gameTime": "1 - 06:35",
        "label": "Offside",
        "position": "395728",
        "team": "away",
        "visibility": "visible"
    },
```

Listing 2. Example of a camera change annotation in json.

```
"UrlLocal": "path/to/game",
"annotations": [
    {
        "change_type": "logo",
        "gameTime": "1 - 06:57",
        "label": "Main behind the goal",
        "link": {
            "half": "1",
            "label": "Offside",
            "position": "395728",
            "team": "away",
            "time": "06:35",
            "visibility": "visible"
        },
        "position": "417414",
        "replay": "replay"
    },
```

These tools were given to our 33 annotators, who are engineering students and soccer fans. Each annotator is attributed a given annotation task with detailed instructions and a set of matches to annotate. In case of doubt, they always have the possibility to contact us so that we control their work in ambiguous situations.

The total annotation time amounts to ~1600 hours. Annotating all the actions of a single game takes ~105 minutes; annotating all the camera changes requires ~140 minutes per game, while only associating each replay shot of a game with its action takes ~70 minutes.

7.3. Human Level Performances

Manually labeling events with timestamps raises the question of the sharpness of the annotations. In Charades [70], the average tIoU of human annotators on temporal boundaries is only of 72.5%[1], and 58.7% on MultiTHUMOS [84]. Alwassel *et al.*[2] also observe some variability on ActivityNet [30], but note that a reasonable level of label noise still allows performance improvements and keeps the challenge relevant.

Although all the annotations of our SoccerNet-v2 dataset are based on a set of well-defined rules, some uncertainty still resides in the timestamps. To quantify it, we determine an average human level performance on a common match shared across all the annotators as follows. We assess the performance of an annotator against another by considering one as the predictor, the other as the ground truth. Then, we average the performances of an annotator against all the others to obtain his individual performance. Finally, we average the individual performances to obtain the human level performance. This yields an Average-mAP of 80.5% for action spotting, a mIoU of 69.0% for camera segmentation, and a mAP of 90.2% for camera shot boundary detection. These metrics indicate that label noise is present but that current algorithms are still far from solving our tasks with a human-level cognition of soccer, as seen in Tables 2, 3, 4 of the main paper.

[1] Gunnar A. Sigurdsson, Olga Russakovsky, and Abhinav Gupta. What actions are needed for understanding human actions in videos? In *IEEE International Conference on Computer Vision (ICCV)*, pages 2156-2165, October 2017.

[2] Humam Alwassel, Fabian Caba Heilbron, Victor Escorcia, and Bernard Ghanem. Diagnosing error in temporal action detectors. In *European Conference on Computer Vision (ECCV)*, pages 264-280, September 2018.

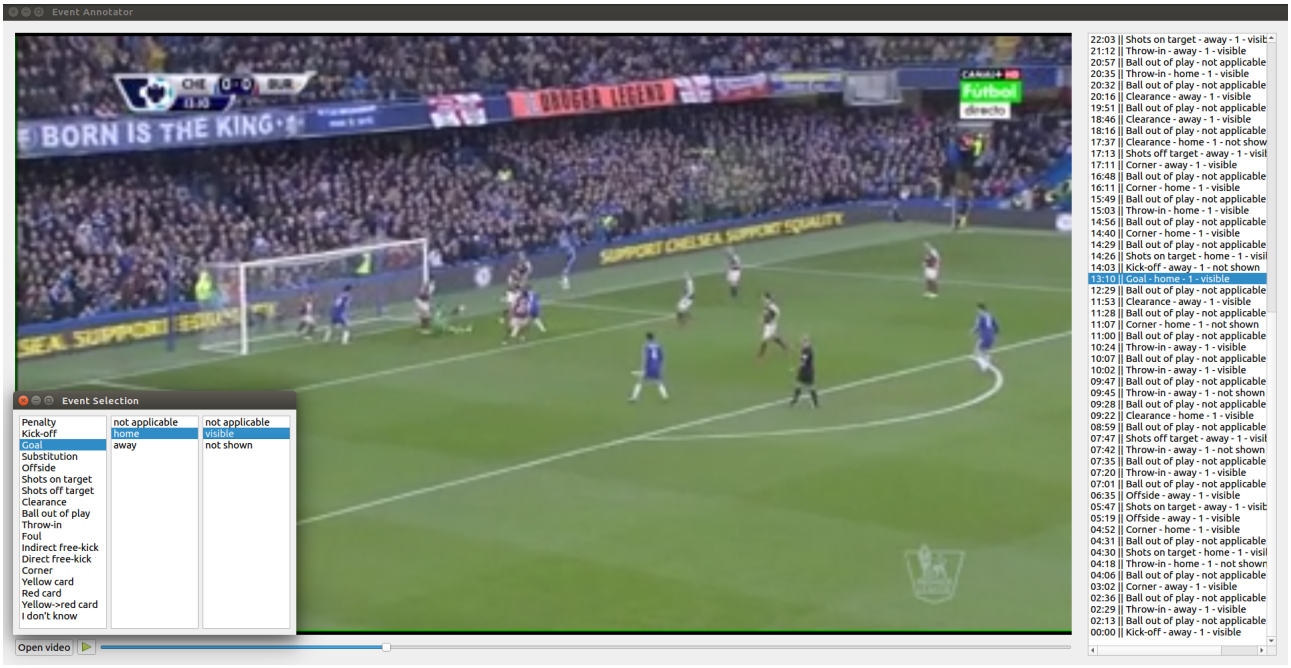


Figure 11. **Actions annotation tool.** When an action occurs, the annotator pauses the video to open the annotation menu (bottom left) and selects the action, the team that performs it, and whether it is shown or unshown in the video. The right column provides all the actions already annotated for that game, sorted chronologically.

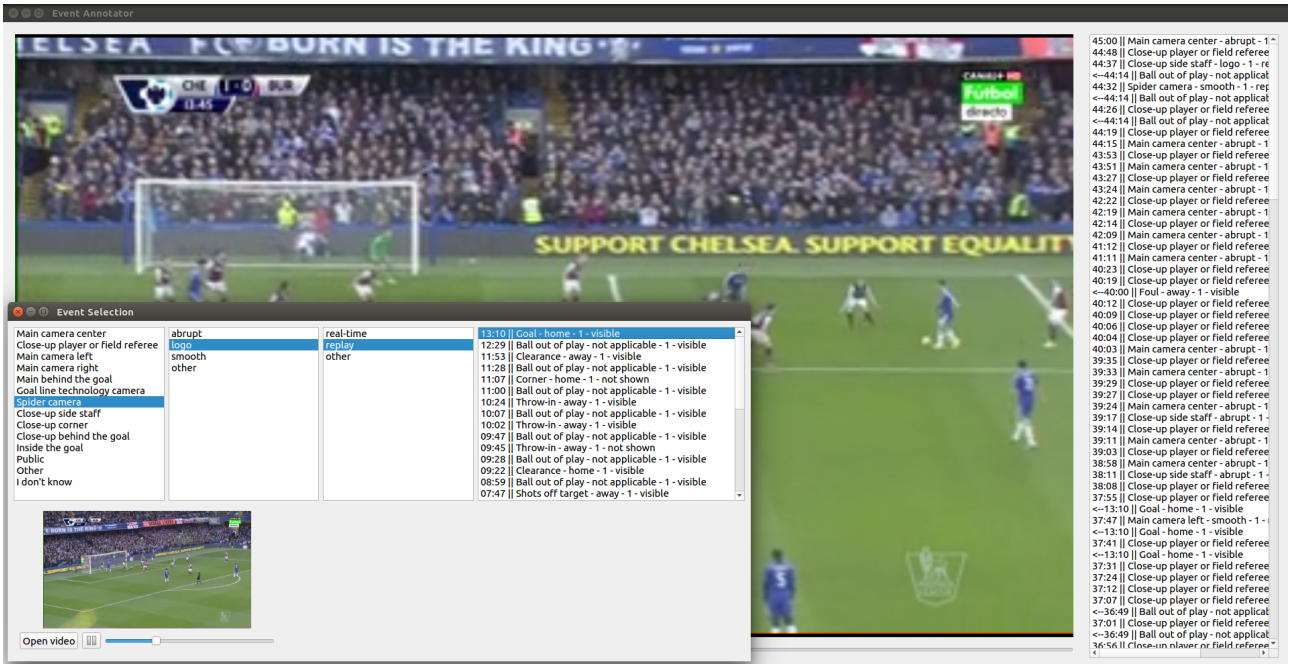


Figure 12. **Cameras annotation tool.** When a camera transition occurs (in this case, just before for a better visualization), the annotator pauses the video to open the annotation menu (bottom left) and selects the type of camera, the upcoming transition, and the real-time or replay characteristic of the current shot. In the case of a replay, as shown here, the annotator selects the action replayed in the last column, with the possibility to visualize a short clip around the action selected to ensure the correctness of the annotation. The large column on the right provides all the camera shots already annotated for that game, sorted chronologically.