iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations

Zhutian Chen
Harvard University
Cambridge, MA, USA
ztchen@g.harvard.edu

Qisen Yang*
Zhejiang University
Hangzhou, Zhejiang, China
qs_yang@zju.edu.cn

Jerry Shan*
UC Berkeley
Berkeley, CA, USA
jiarui.shan@berkeley.edu

Tica Lin
Harvard University
Cambridge, MA, USA
tlin@g.harvard.edu

Johanna Beyer
Harvard University
Cambridge, MA, USA
jbeyer@g.harvard.edu

Haijun Xia
UC San Diego
La Jolla, CA, USA
haijunxia@ucsd.edu

Hanspeter Pfister
Harvard University
Cambridge, MA, USA
pfister@g.harvard.edu

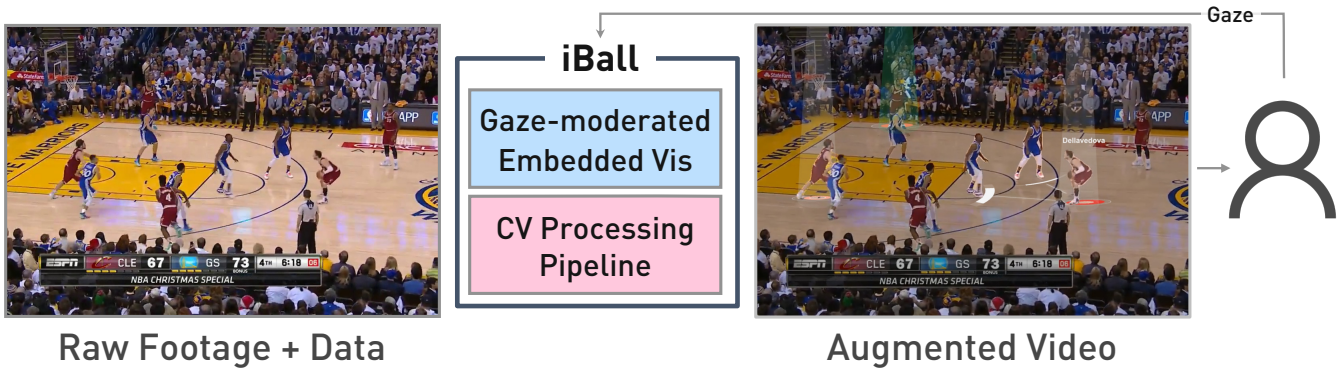

Figure 1: iBall augments basketball videos with gaze-moderated embedded visualizations to facilitate game understanding and engagement of casual fans. It embeds data visualizations into basketball raw footage using a well-designed computer vision pipeline, and automatically adapts the visualizations based on the game context and users' gaze.

ABSTRACT

We present iBall, a basketball video-watching system that leverages gaze-moderated embedded visualizations to facilitate game understanding and engagement of casual fans. Video broadcasting and online video platforms make watching basketball games increasingly accessible. Yet, for new or casual fans, watching basketball videos is often confusing due to their limited basketball knowledge and the lack of accessible, on-demand information to resolve their confusion. To assist casual fans in watching basketball videos, we compared the game-watching behaviors of casual and die-hard fans in a formative study and developed iBall based on the findings. iBall embeds visualizations into basketball videos using a computer vision pipeline, and automatically adapts the visualizations based on the game context and users' gaze, helping casual fans appreciate basketball games without being overwhelmed. We confirmed the usefulness, usability, and engagement of iBall in a study with 16 casual fans, and further collected feedback from 8 die-hard fans.

KEYWORDS

Augmented Sports Videos, Embedded Visualization, Gaze Interaction, Sports Visualization, Video-based Visualization

ACM Reference Format:
Zhutian Chen, Qisen Yang, Jerry Shan, Tica Lin, Johanna Beyer, Haijun Xia, and Hanspeter Pfister. 2023. iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3544548.3581266

1 INTRODUCTION

Basketball, one of the most popular team sports, has continued to attract new fans over the past decades due to the proliferation of video broadcasting and online video platforms. However, as our formative study will show, unlike experienced fans, new or casual fans often get confused when watching basketball videos. This is because they lack sufficient basketball knowledge to understand the players' complex teamwork and in-game decisions. Existing methods of providing extra information, such as scoreboards in broadcasting videos and online webpages (e.g., ESPN [23]), often

*This work was done when Qisen Yang and Jerry Shan were interns at Harvard University. They contributed equally to this research.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581266

fail to adequately address their confusion. These methods either cannot provide on-demand data or distract fans from the game by showing data in separate windows. Thus, finding a way to enable seamless access to extra data while watching basketball games would particularly benefit casual fans in their understanding and engagement of games.

Embedded visualizations provide a promising opportunity to allow audiences to access extra data without being distracted from the game video by directly displaying data in the actual scenes. Given such benefits, various commercial products [19, 76] and research systems [15, 16] leverage embedded visualizations to augment sports videos. However, these systems either focus on post-game analysis rather than game-watching scenarios, or only use simple, non-interactive text labels and progress bars to show data. Recently, researchers have started to explore the design space of embedded visualizations for game-watching scenarios but used low-fidelity simulated environments (e.g., 3D simulated sports games [41], moving charts on white backgrounds [80]). Little is known about how to design and implement interactive embedded visualizations in real sports videos and how they can facilitate game understanding and increase engagement of fans when watching games.

In this work, we aim to fill this gap by developing interactive embedded visualizations to assist casual fans in watching basketball game videos. To understand the particular practices, pain points, and solutions of casual fans in watching basketball game videos, we compared the game-watching behaviors of 8 casual and 8 die-hard fans in a formative study. Findings revealed that the casual fans were confused about key players and their in-game decisions from time to time, and had trouble seeking customized data during the game. Informed by the study, we developed iBall (Fig. 1), a basketball game viewing system that automatically highlights key players and visualizes their performance through gaze-moderated embedded visualizations.

We developed iBall by tackling two main challenges. First, embedding visualizations into actual scenes is recognized as a grand challenge [22], especially for basketball, where players overlap heavily and the camera moves rapidly. To tackle this challenge, we contribute a CV pipeline that pre-processes team sports videos for embedding visualizations. We also conducted experiments to evaluate our pipeline quantitatively and discuss potential methods to extend the pipeline to process live videos. Second, it remains unclear how to design embedded visualizations that are informative but not overwhelming for individual audiences, who may have various levels of game literacy, data needs, and personal interests. We designed a set of gaze-moderated embedded visualizations that leverage the user's gaze to seamlessly present the data the user is interested in and suppress others. To evaluate iBall, we conducted a user study with 16 casual fans to compare the game-watching experiences between raw video (RAW), video + embedded visualizations (AUG), and video + gaze-moderated embedded visualizations (FULL). Participants spoke highly of our system, ranked FULL as the best, and confirmed that our embedded visualizations and gaze interactions were useful, usable, and engaging. We further collected and discuss the feedback on iBall from another 8 die-hard fans. We discuss our observations and design implications learned from the study for future research inspiration.

In summary, through developing iBall, we make the following four main contributions: 1) a formative study that identifies the pain points of casual fans in watching basketball videos and solicits plausible solutions from die-hard fans, 2) an open-source CV pipeline to process team sports videos for embedding visualizations, 3) a set of gaze-moderated embedded visualizations for basketball game videos, and 4) a user study that assesses our system and provides insightful feedback on using gaze-moderated embedded visualizations in team sports videos. Finally, we will open source our system at https://github.com/ASportsV/iBall.

2 RELATED WORK

We review prior work on personalized game viewing systems, embedded visualizations in sports videos, computer vision for embedded visualizations, and applications of gaze interactions.

2.1 Personalized Game Viewing Systems

Visualization has long been used in sports to present data [56], including box scores [25], tracking data [18, 44, 55, 81], and metadata [77]. Sports visualizations are mainly used for post-game analysis or in-game informing purposes. This work mainly focuses on the latter.

Sports games usually involve complex in-game decision-making. To better understand, analyze, and appreciate players' in-game decisions, spectators often look for additional information when watching a sports game [41]. To fulfill individual spectators' information needs, prior research has explored the design of interactive game-watching systems. ARSpectator [87], for example, presents a concept design of using mobile AR to enhance the experience of live sports events. Gamebot [86] uses a conversational interface to help users request data visualizations in watching NBA games. GameViews [85] uses simple visualizations (*e.g.*, line charts) to show in-game box scores of basketball games. Omnioculars [41] uses interactive embedded visualizations to support in-game analysis of basketball games. CourtVision [19] is a commercial product that allows inspectors to review basketball in-game data through simple, non-interactive embedded visualizations (e.g., text labels, progress bars). Compared to traditional sports, most E-Sports already provide a personalized game viewing experience by default. Multiplayer Online Battle Arena (MOBA) games, such as Defense of the Ancients2 [1] (Dota2) and League of Legends [3] (LoL), allow spectators to interact with the systems to inspect in-game data (*e.g.*, points over time) of players or teams. Nevertheless, these systems either display the data in separated panels or require viewers to explicitly interact with the system to request the data, inevitably distracting viewers from the game. In contrast, we propose to use embedded visualizations and gaze interactions to present extra data in game videos, providing an intuitive, seamless, and engaging watching experience.

2.2 Embedded Visualizations in Sports Videos

Embedded visualizations have been widely used for sports data due to their ability to show the data into its physical context (*e.g.*, a basketball court). Early works mainly embedded the data into static court diagrams. Examples such as CourtVision [28] (basketball), StatCast Dashboard [40] (baseball), and SnapShot [57] (ice

hockey) display density maps on top of court diagrams to show sports events, such as successful shots. Recent progress in CV now allows embedding visualizations directly into sports videos instead of just court diagrams. For example, Stein et al. [68, 69] developed a method to automatically extract and visualize data from and in soccer videos. Chen et al. [15, 16, 42] explored the design of augmented sports videos and introduce fast prototyping tools to help users create augmented videos for racket-based sports by using direct manipulation and textual comments. However, these works mainly target experts for analytic and authoring purposes. More recently, researchers have started to explore embedded visualizations in live game-watching scenarios. Yao et al. [80] proposed the notion of *visualization in motion* to depict visualizations that are moving relative to the viewer and summarized a design space for it. Lin et al. [41] presented a design framework for embedded visualizations to facilitate in-game analysis when watching basketball games. Yet, all the above works only evaluated visualizations in simulated scenarios (e.g., moving charts on white backgrounds, 3D virtual sports games). We design our embedded visualizations based on these prior works but particularly target real basketball videos, with the aim to understand how embedded visualizations can improve casual fans' game-watching experience.

2.3 Computer Vision for Embedded Visualizations

Recent years have shown remarkable advances in CV techniques based on deep learning. Researchers have achieved unprecedented success in a broad range of tasks including object detection [26, 43], object tracking [7, 83], pose estimation [79], and segmentation [14]. Thanks to this progress, more and more data can be extracted from videos (e.g., [20, 30, 36, 59]), opening new opportunities for sports analytics. For example, the positions of the players and the ball [67], as well as other tracking data [4, 5], of each NBA game are extracted and shared online. We refer the reader to Shih [63] for a comprehensive survey on content-aware video analysis for sports. Furthermore, these new CV techniques ease the embedding of visualizations into the video scenes, which is recognized as a grand challenge in situated visualization [22]. Embedding visualizations into sports videos requires a CV pipeline to complete tasks such as detecting, tracking, and segmenting the players from the video, estimating their pose, calibrating the camera [84], and sometimes reconstructing the 3D scene [47]. Prior works [15, 16, 69] applied a simplified CV pipeline to process racket-based sports videos, in which the players are separated, and the camera is mostly static. However, it is much more difficult to embed visualizations into team sports videos (especially basketball) since players overlap heavily and the camera typically moves rapidly. While commercial systems [19] can achieve good embedding results, they require videos collected from multiple cameras [78] to register the visualizations. To the best of our knowledge, there is no existing CV solution that can embed visualizations into basketball videos based solely on broadcasting videos. The lack of such a solution inevitably hinders the research of embedded visualizations in complex, dynamic scenarios, such as team sports. In this work, we contribute a CV pipeline that consists of open-sourced modular components to process team sports videos for embedding visualizations.

2.4 Applications of Gaze Interactions

There is a long history of interest in leveraging gaze for interactions due to its efficiency, expressiveness, and applicability in hands-free scenarios [11, 45, 46]. Gaze interactions either *explicitly* or *implicitly* leverage the gaze to interact with digital content. We focus on implicit methods and refer readers to a more comprehensive review [45] for further reading.

Implicit gaze-based systems use gaze as an implicit input source, usually in combination with other input modalities, to facilitate interactions [24, 45]. Given that reliable eye trackers are now affordable enough to be integrated into desktop and laptop computers, researchers have leveraged implicit gaze interactions to support a variety of applications, such as content annotation [17, 70], video editing [35, 48, 58], and remote collaborations [32, 38]. The most relevant to our endeavor are attempts at adapting viewing content based on users' gaze. The gaze-contingent display [21], for example, shows a higher resolution on the area the user is focusing on. Other examples include adjusting the playback speed of lecture videos based on the user's gaze [51], or a tourist guide that directs a user's gaze to highlighted features in a panorama and adapts the audio introductions accordingly [39]. Kurzhals et al. [37] have proposed a gaze-adaptive system that dynamically adjusts video captions' placement to optimize the viewing experience. We also aim to use gaze to adjust video content but focus on augmented sports videos.

In the visualization field, research related to gaze mainly focuses on visualizing gaze data [8] and analyzing users' gaze in viewing visualizations [9, 12]. Only a few works [53, 62, 65, 66] have explored leveraging the gaze to interact with visualization systems. Silva et al. [64] give a systematic review on eye tracking for visual analytics systems and current challenges. We draw on this line of research and, to the best of our knowledge, are the first to explore gaze-aware embedded visualizations to improve the sports-watching experience.

3 FORMATIVE STUDY WITH BASKETBALL FANS

To understand the practices, pain points, and solutions of casual fans in watching basketball videos, we conducted a formative study.

3.1 Study Setup

3.1.1 *Participants*

We recruited participants using university mailing lists and forums and pre-screened participants based on their fandom level, game-watching frequency, and basketball knowledge. In total, we recruited 8 casual fans (P1-P8; M=3, F=5; Age: 18 - 35), who only knew *"basic rules of basketball"* and watched *"1 - 10 games per year"*. To better identify the pain points specific to casual fans, we further recruited 8 die-hard fans (P9 - P16; M=8; Age: 18 - 55), who knew *"basketball tactics and pros and cons of specific players"* and watched *"at least 1 game per week"*. No female die-hard fan responded to us. All participants had normal vision or wore contact lenses or glasses to correct to normal vision.

3.1.2 *Procedure*

We started each session by introducing our research motivation and study protocol. The experimenter then conducted a semi-structured

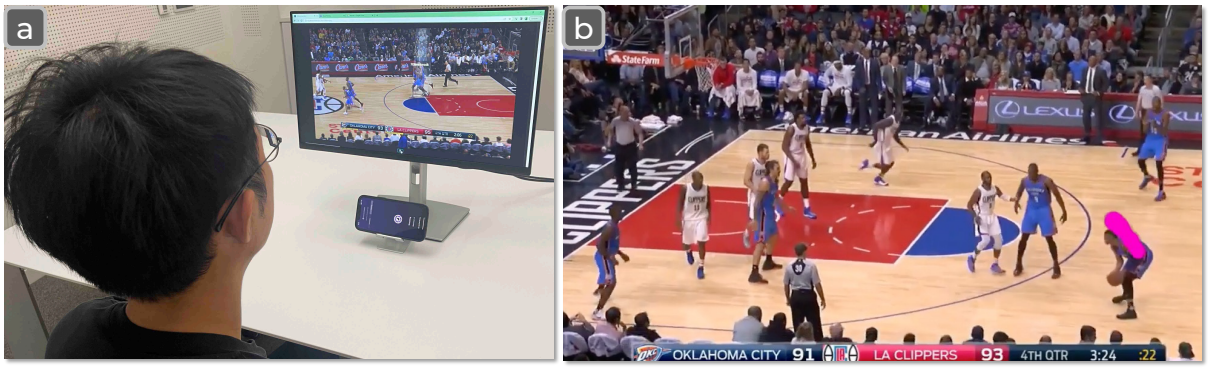

Figure 2: a) Formative study setup. b) In the review phase, the participant's gaze is visualized and overlaid on the video.

Table 1: The two game videos in the formative study.

	Date	Teams	Quarter	Duration (m:ss)
G1	2015.12.25	GSW vs. CLE*	4	9:02
G2	2015.12.22	OKC vs. LAC*	4	9:30

* GSW (Golden State Warriors), CLE (Cleveland Cavaliers), OKC (Oklahoma City Thunder), and LAC (Los Angeles Clippers)

interview with each participant, focusing on their current practices, pain points, and solutions when watching live basketball games. Next, we followed the format of *contextual query* [31] to ask participants to watch two videos (Table 1) on a 24-inch monitor. These two games were rated as top-30 games of the season and have been watched millions of times. We collected think-aloud and gaze data during the game-watching process. To collect the gaze data, we used Eyeware Beam [2], a commercial software that leverages Apple's TrueDepth camera [6] to track the participant's head and gaze. The participants sat approximately 60cm from the screen and were asked to adjust the chair before watching the videos (Fig. 2a). The system was then calibrated and the participants were allowed to move the head freely after the calibration. We used a TrueDepth camera-based tracker as it provided sufficient accuracy [29] for inspecting what video objects participants were looking at while watching the game, at a much lower budget. For more fine-grained gaze data (e.g., saccades, fixation), more proficient eye-tracker would be required.

After watching each game video, we asked participants to re-watch the game with their gaze data overlaid (Fig. 2b) and to elaborate on any confusion, data needs, insights, and excitement they had felt when watching the game for the first time. Participants could pause the video in the review phase. Each participant was compensated with a $20 gift card for their time (1 hour).

3.1.3 *Analysis*

Interviews and think-alouds were audio-recorded, transcribed, and analyzed using a reflexive thematic analysis [10]. Three authors coded independently on the transcriptions to form sets of plausible codes and iteratively refined the codes to converge on a single coding schema. Besides, three authors analyzed the gaze data by manually annotating the video objects each participant was looking at while watching the games. The categories of objects (Fig. 3 x-axis) were generated based on the data and prior knowledge. We classified participants as looking at an object only when their gaze rested on the object for at least 0.25 seconds (fixation duration [54]). The duration when the gaze was moving to the object was also annotated as looking at the object.

3.2 Findings and Discussions

All the casual fans only watched *"important games, such as semi-finals or finals."* (P1) They were neither familiar with basketball nor the NBA. In comparison, the die-hard fans watched basketball games much more frequently. They had a rich knowledge of basketball (e.g., tactics), knowing almost all NBA players and even their strengths and shortcomings. TV was the main way for all the participants to watch live basketball games. Overall, for the casual fans, watching live basketball games was a leisure activity, such as hiking, but it was a more serious hobby for the die-hard fans.

3.2.1 *Casual Fans' Confusion in Watching Basketball Games*

In terms of the watching experience, all 16 participants confirmed that they were confused from time to time when watching basketball games and that they would like to seek extra information, other than the data provided by the scoreboard and commentaries. Some confusion is common among both casual and die-hard fans, such as questions like *"who got a foul?"* and *"which team called the timeout?"* These questions can usually be resolved by *"watching the replay"* (P15) or simply by searching Google. However, we did identify some confusing aspects specific to casual fans that cannot be easily resolved by the current methods and thus lead to a poor watching experience:

C1: Casual fans are unsure about which players they should focus on. When watching basketball games, the casual fans often could not identify the important players and felt that the players were just *"moving objects."* (P4) The casual fans' inability to identify key players was also reflected in their gaze patterns. In our study, we found that casual fans spent more time on the player with the ball than the die-hard fans (Fig. 3), since they *"didn't notice other players' [off-ball] movement"* (P1) when watching the game. As a result, the casual fans often missed important off-ball movements and felt that the ball *"magically fly to an open player."* (P1) Moreover, in some casual fans' gaze, we noticed some rapid zigzag movement between the player with the ball and the other players, revealing their attempts to scan through the players. P4, for example, explained that

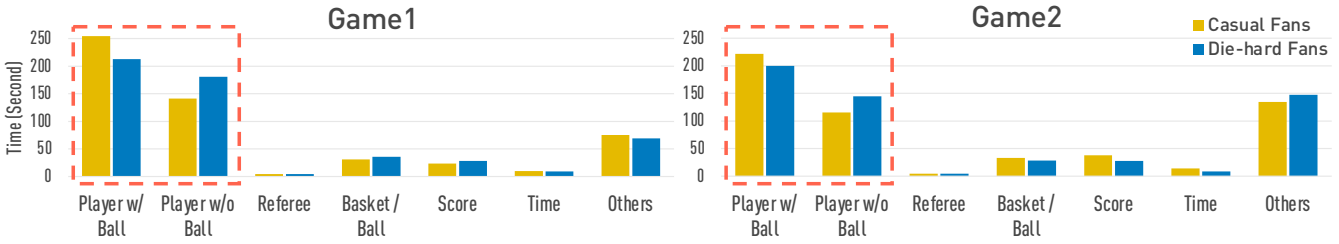

Figure 3: The gaze distribution in seconds shows that, compared to die-hard fans, casual fans spent more time watching the player with the ball in both games. Due to the small sample size, we discuss these results with a descriptive approach and focus more on other behavioral observations.

she was *"scan[ning] other players"* to predict the ball receiver at the next pass while keeping an eye on the player with the ball, leading to a heavy cognitive load. In contrast, the die-hard fans scanned through the players much more predictively and often could directly identify the next ball receiver.

C2: Casual fans are confused about the in-game decisions of players. The casual fans could hardly understand the in-game decisions of the players, since the situated factors (e.g., players' abilities) behind these decisions were hard to interpret from the videos. Consequently, the casual fans could not appreciate the game at the same deep level as die-hard fans and had difficulties keeping pace with their experienced friends. This was also revealed in the think-aloud data of the participants. When watching the two videos, the most frequent verbal comments from the casual fans were interjections, e.g., *"Oops"*, *"Wooooow!"* Even in the follow-up review session, casual fans could hardly describe their thoughts while watching the games. P3 acknowledged that she sometimes actually *"didn't totally understand"* what was going on but just felt excited. By contrast, the die-hard fans could clearly elaborate, comment on, and even suggest players' tactics when watching the games. Generally speaking, our study suggested that the experience of watching games for the casual fans was closer to "feeling" while the experience for the die-hard fans was closer to "reading".

C3: Casual fans have trouble seeking customized data while watching game videos. All the casual fans never searched the internet to seek data to resolve their confusion when watching the games. This was because the games were so fast and overwhelming that they could miss key events when looking up websites. Additionally, the casual fans sometimes could not search for a player's data because they did not know the player's name. In contrast, the die-hard fans would search websites (e.g., ESPN) when watching the games, though they also complained about the context switching between the games and the webpages. According to the casual fans, perhaps the best way to seek information about game understanding was to *"ask my [experienced] friends."* (P2) Otherwise, they would just *"let it [the confusions] go."*

3.2.2 *Die-hard Fans' Suggestions for Understanding Basketball Games*

Since casual fans preferred to *"ask experienced friends"* to seek information, we were interested in what information die-hard fans suggest for understanding a live basketball game. Several critical insights were suggested by die-hard fans:

Distinguishing between offense and defense. Basketball, from a certain perspective, is a turn-based game. A basketball game consists of multiple possessions (i.e., turns), in which the team that has possession of the ball is on offense, and the other team is on defense. A player can have completely different roles, tactics, and behaviors between offense and defense. Being aware of players' offense and defense status can help casual fans better understand and follow the game.

Identifying Key Players. While basketball is a team sport, the importance of each player, especially when she/he is on offense, is different. Generally speaking, on the offensive side, *the player with the ball* and *the ball receiver* at the next pass are the most important ones. *Players with open spaces* are also critical to the offensive team as they have a higher chance of making the goal. On the defensive side, all the defenders guarding the player with the ball are important. By identifying these key players, the die-hard fans could watch the game more effectively and predictably. In addition to the aforementioned key players, we also discussed other players with the die-hard fans, such as offensive helpers who play screens. Overall, they suggested not helping casual fans identify these players, as their contributions to the possession outcome (e.g., a goal) are not explicit and thus can confuse casual fans.

Understanding In-game Decisions. Knowing players' offensive and defensive abilities is essential to understanding their in-game decisions. The die-hard fans suggested two metrics to help casual fans understand the players' abilities. For offensive players, we can present their location-based *expected point value*, which measures how many points a player is expected to make if they shoot at a specific location. For defensive players, we can present their location-based *percentage points difference*, which measures how much the field goal percentage of a player changes when being defended by the defensive player. Both metrics can be calculated or directly obtained by using the data from the Official NBA Stats website [5]. The die-hard fans also suggested visualizing the one-on-one relationships between offensive and defensive players, which can reveal interactions between the players and their tactics (e.g., defensive switching).

3.3 Summary

In summary, the casual fans were often confused about the key players and their in-game decisions, but rarely sought data to resolve their confusion because the searching process is slow and

Table 2: Three design requirements for assisting casual fans in game watching derived from the formative study.

Findings	Design Requirements
C1: Casual fans are unsure about which players they should focus on.	R1: Guide the user's attention to the important offensive and defensive players. (Sec. 5.1)
C2: Casual fans are confused about the in-game decisions of players.	R2: Visualize players' offensive and defensive abilities. (Sec. 5.2)
C3: Casual fans have trouble seeking customized data while watching game videos.	R3: Provide a fast and seamless method to retrieve data of interest. (Sec. 5.3)

distracting. To help casual fans better understand the game, the die-hard fans suggested a few critical insights, including distinguishing between offense and defense, identifying key players, and understanding players' in-game decisions. By interpreting these findings and suggestions, we derived three design requirements (Table 2) and designed iBall. Next, we first introduce a CV pipeline (Sec. 4) to enable iBall, followed by a set of gaze-moderated embedded visualizations (Sec. 5).

4 A CV PIPELINE FOR EMBEDDING VISUALIZATIONS

To embed visualizations into a basketball video, we need to recognize the players (e.g., bounding box, identity, and key points) and segment them from the background. To this end, we designed a CV pipeline (Fig. 4) to pre-process team sports videos.

4.1 Recognizing the Players

To embed visualizations for a player, one must first recognize the player in the video. For example, to display a label with the name of a player, the system needs to detect the video object that corresponds to the player (bounding box and identity) and the player's key body joints (key points) for placing the label. Given a raw video frame, we obtain this information for each player via three steps:

Step 1. Player Detection *(Fig. 4a).* To obtain the players' bounding boxes and identities in a video frame, we use an object detection model to locate and classify each player into different categories. Different from common object detection tasks, we use the players' identities as their categories. In our implementation, we fine-tuned a COCO [13] pretrained YoLoX [27] model on an NBA player dataset (details in Appendix A). The output of the model is a set of bounding boxes associated with their identities and confidence scores (i.e., $score_c$). By convention, only those bounding boxes with $score_c$ greater than a threshold T_{high} are considered successful detections.

Step 2. Post-Processing *(Fig. 4b).* One limitation of the object detector is that it only utilizes the players' visual appearance information to determine their confidence score. Consequently, the detector can assign low confidence scores to players whose visual qualities are low (e.g., when they are occluded by others) and filter them out. We thus use object trackers to exploit the players' motion information to complement the detector. An object tracker stores the history of an object's bounding boxes in the previous frames and can predict the object's bounding box in the next frame by using a Kalman filter. We use object trackers as follows:

(1) For a frame F_t, we divide all the detected bounding boxes into three clusters based on their $score_c$: high-quality boxes ($score_c > T_{high}$), low-quality boxes ($T_{low} < score_c < T_{high}$), and rejected boxes ($score_c < T_{low}$).
(2) For each high-quality box, we match it with the trackers in the previous frame F_{t-1} by calculating the Intersection over Union (IoU) between the box and the predicted boxes of the trackers. A tracker is considered as *matched* with the high-quality box if it maximizes the IoU. If matching is successful, we assign the matched tracker to the box; otherwise, we initialize a new object tracker for the box.
(3) For each low-quality box, we match it with the remaining trackers (i.e., those that have not been matched with any high-quality boxes). If matching is successful, we assign the matched tracker to the box.
(4) Finally, we output all the boxes with matched trackers.

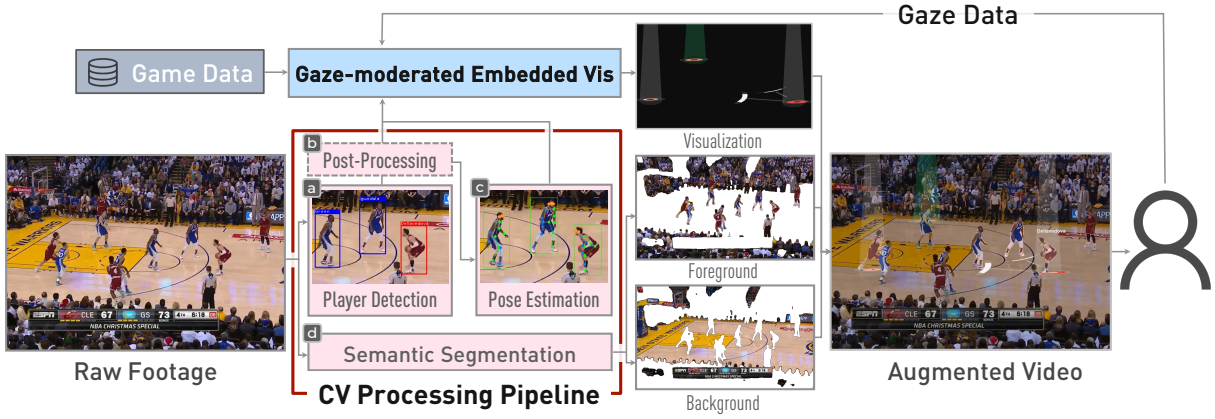

Figure 4: Our CV pipeline takes a raw video as the input, outputs the bounding box, identity, and key points of each player, and separates the image frame into the foreground (humans) and background (all others). The bounding boxes, identities, and key points are used to create visualizations, which are then composited with the foreground and background to form the augmented video.

Intuitively, this method uses the motion information of the players to select some low-quality bounding boxes to complement the output of the detector. We refer the readers to Bot-SORT [7] for more details about the tracker and the matching process.

Step 3. Pose Estimation *(Fig. 4c).* We use a pose estimation model to obtain the players' key points, such as head, hands, hip, and feet. In our implementation, we first used the bounding boxes produced in Step. 2 to extract the players from the video frame and then fed those boxes to ViTPose [79] to estimate the key points.

4.2 Separating Foreground and Background

According to previous works [16, 41], embedded visualizations for sports, such as empty areas, are often placed on the ground, beneath the players' feet. To achieve this, we need to separate the video frame into the foreground (the objects on the ground) and background (the ground), draw the visualizations onto the background, and finally overlay the foreground on the background to form an augmented video frame (Fig. 4d). Ideally, all the objects should be segmented from the ground. To simplify the segmentation process, we decided to only segment humans from the video as the foreground and leave the remaining pixels as the background, as humans are the major objects on the ground in a basketball video. In our implementation, we used a ViT-Adapter [14] trained on COCO 164K [13] to perform binary semantic segmentation to segment the humans.

4.3 Computational Evaluation

To evaluate the performance of our pipeline, we conducted several experiments focused on three main questions: 1) Can the object detector detect the players?; 2) Can the post-processing step improve the detections?; 3) How much time does each step take? To answer these three questions, we manually annotated the bounding box and identity of each player in each frame of the two game videos used in Sec. 3 (i.e., G1 and G2). We then split each video into clips and allocated 70% for training and 30% for testing. To accelerate the training process, we sampled every tenth frame from the training clips and used only these frames for training. This is because the consecutive frames often contain redundant information. Despite this, the testing was conducted on all frames in the testing clips. The details of the dataset can be found in Appendix A. We trained and evaluated the detector on G1 and G2 separately, using their default hyperparameters whenever possible.

We did not evaluate the accuracy of the Pose Estimation and Semantic Segmentation steps because we used off-the-shelf models for their standard tasks without any fine-tuning in these two steps. Yet, their performance for basketball videos can be qualitatively evaluated by inspecting the augmented videos provided in the supplemental material.

Table. 3 shows the performance of our fine-tuned object detector on the testing clips of G1 and G2. To access the detector, we followed the convention to calculate the Average Precision (AP) metrics over different IoUs. The higher the AP, the better it is. $AP_{50:95}$ is the average AP over different IoU, from 0.5 to 0.95 with step 0.05. AP_{50} and AP_{75} are the APs calculated at IoU 0.5 and 0.75, respectively. The larger the IoU, the stricter the metric will be. Overall, the results reveal that our fine-tuned object detector can perform well in detecting players. Furthermore, all the APs increase after applying the post-processing step, which shows that the post-processing step is useful and can complement the detector to improve its results.

Table 3: Average Precision of the Player Detection and Post-processing steps.

Dataset	Step	$AP_{50:95}$	AP_{50}	AP_{75}
G1	Player Detection	65.4	83.6	76.2
	Post-Processing	69.2 (+3.8)	87.9 (+4.3)	79.4 (+3.2)
G2	Player Detection	70.7	86.1	82.3
	Post-Processing	75.0 (+4.3)	90.0 (+3.9)	85.3 (+3.0)
COCO*	YoLoX	51.2	69.6	55.7

*Due to the lack of benchmarks, we provide YoloX's performances on COCO as a reference. However, it does not serve as a comparative baseline.

Table 4: Time cost of each step.

Step	Time (ms)
Player Detection	31.98
Post-Processing	2.40
Pose Estimation	121.00
Semantic Seg.*	3674.96

*Semantic Segmentation can run in parallel with other steps.

In terms of time performance, Table. 4 shows the average time in milliseconds (ms) each step takes to process a video frame. We tested the pipeline on a machine with a Nvidia Tesla V100 graphic card and only counted the inference time of the models by excluding the model and dataset loading time. Overall, the Player Detection and Post-processing steps use 34ms for one frame, almost achieving 30FPS. Other steps, especially the Semantic Segmentation step, need longer to process one frame. These results show that the semantic segmentation model we used is the bottleneck for extending the pipeline to support real-time scenarios.

4.4 Extendibility, Generalizability, and Limitations

The contribution of our pipeline does not lie in the individual components but a workable solution that shows which CV models are required and how they can be composited together to process basketball videos for the purpose of embedding visualizations into videos. To inspire future research, we further discuss the extendibility, generalizability and limitations of the pipeline:

Extendibility. Our pipeline can be extended for better performance. To improve the accuracy, we can try using better models or adding extra components to the post-processing step to improve the detections. For example, in our implementation, we further interpolated and smoothed the bounding boxes for the user study. To improve the efficiency, we can use faster models, more powerful graphic cards, or remove the Semantic Segmentation step if visualizations on the ground are not needed. Overall, our pipeline can serve as a reference for other researchers to develop their own systems for their specific scenarios, videos, and tasks.

Generalizability. The CV pipeline can be applied to other basketball videos and even other team sports videos. For example, there are about 450 players in the NBA [49]. To generalize the pipeline to other NBA game videos, we need to develop a player dataset of

these 450 players to fine-tune the detector. Note that it is not necessary to develop a player dataset for each video. Our experiments showed that the detector could detect players on unseen testing clips even if it was trained only on the training clips. If the player dataset is large enough, the detector fine-tuned on it can be applied to any NBA game video. This is not impossible as modern deep learning-based image classifiers can achieve superhuman performance on tasks with more than 1000 classes [82] and many priors can be used to optimize the model results, e.g., there are no more than 24 players in a game.

Limitations. The pipeline and the evaluation have a few limitations. First, as shown in Table. 4, the processing time of our pipeline for one frame is about 4 seconds. While the Semantic Segmentation step can run in parallel with others, our implementation can only pre-process the game videos instead of running in real-time. Second, our pipeline only extracts 2D information from the video, limiting the design space of available embedded visualizations. For example, without the camera parameters, we cannot display visualizations that are static relative to the ground, such as trajectories. In reality, the camera parameters can be provided by the producer of the video or estimated using camera calibration techniques (e.g., [34, 61]). In our implementation, similar to prior research [15, 16], we treated the camera parameters as partially known meta information to display visualizations that are static relative to the players. Third, we only evaluated the pipeline on G1 and G2. A larger video dataset with more ground truth labels is required to fully test the pipeline. We consider developing such a sports video dataset beyond the scope of this work and leave it for the future.

5 GAZE-MODERATED EMBEDDED VISUALIZATIONS

Based on the identified design requirements from the formative study, we designed a set of gaze interactions that can naturally guide and respond to the user's attention through gaze tracking without explicit user input. Our gaze-moderated embedded visualizations 1) guide audiences' attention and 2) reveal players' offensive and defensive abilities and 3) update the embedded visualizations based on gaze. The system flow is shown in Fig. 5a-c.

5.1 Guiding Audiences' Attention

To help casual fans identify the important players (R1), we first ranked the players' importance levels according to die-hard fans' suggestions and then highlighted the players accordingly.

5.1.1 *Ranking Players' Importance Levels*

Based on the formative study, we adopted an offensive-first method to rank the players' importance into three levels:

Lv3 - Key offensive players: The player with the ball, the next ball receiver, and the players with open spaces are considered as the most important offensive players. When pre-processing the game videos, we used positional tracking data [67] to identify which players had the ball or were with open spaces in each frame. Meanwhile, we looked ahead 1.8 seconds (selected empirically) to find the next ball receiver. To extend our system to livestream scenarios in the future, potential approaches could be to use machine learning models [60] or the buffer time in video streaming to detect the next ball receiver.

Lv2 - Key defensive players: The players who are defending the player with the ball are considered as the important defenders. In our implementation, inspired by previous work [72], we detect important defenders by checking which defenders were closest to the player with the ball within a time interval.

Lv1 - Other players: All other players who do not belong to **Lv3** and **Lv2** fall into this level.

To detect if a player is in offense or defense, we also used the positions of the players and the ball. If a player or one of her/his teammates is the closest player to the ball within a predefined time interval (0.5s in our implementation), she/he is in offense; otherwise, in defense. Note that we deliberately ignored some important players, such as those who play screens or specific tactics, since casual fans usually cannot understand why these players are important.

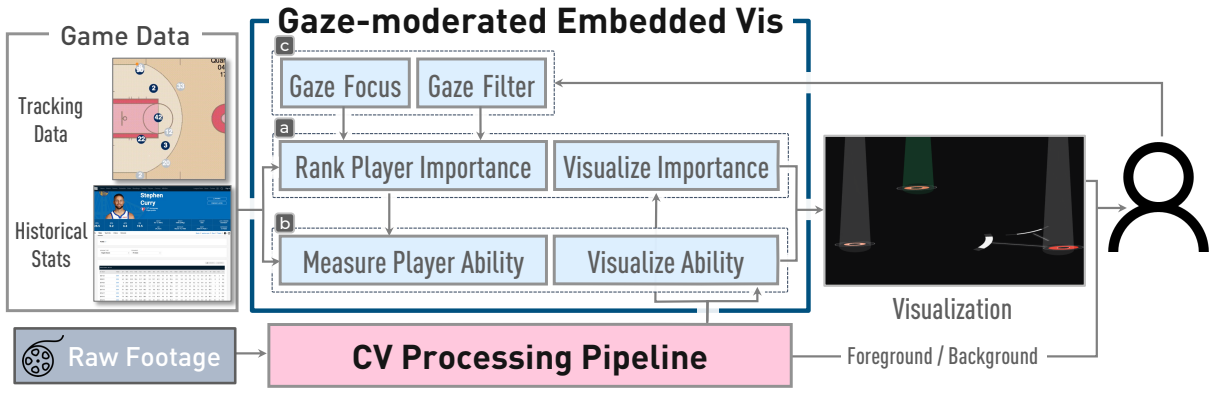

Figure 5: The system takes positional tracking data and historical stats as input to calculate the players' importance and offensive and defensive abilities. Only the important players and their offensive and defensive abilities will be highlighted and visualized in the video. The user can use gaze points to adjust the players' importance levels, as well as controlling whose abilities to show.

Figure 6: Visualization of various importance levels: Lv3, key offensive players, highlighted by a sportlight; Lv2.5, players of interest to the user, triggered by Gaze Focus, highlighted by a glowing effect (will be introduced in Sec. 5.3); Lv2, key defensive players, highlighted by extra brightness; Lv1, other players, no highlighting.

5.1.2 *Visualizing Importance Levels*

We designed multiple highlight effects to guide user attention to players at different importance levels (Fig. 6). We considered the effectiveness and aesthetics of the visualizations and iterate our designs to make them intuitive and distinguishable, i.e., the one for a higher importance level is more attractive. We displayed the name of players with importance levels greater than **Lv2**. We also colored the name of "star" players in gold with an icon showing their roles (i.e., for good shooter and for good defender). Furthermore, Lv3 spotlight encodes the different offense roles with color, *green* for players with open space and *white* for other key offensive players.

5.2 Revealing Players' Abilities

To help casual fans understand the players' abilities (R2), we computed and visualized two location-based metrics of the players whose attention level is higher than **Lv1** (Fig. 5b).

5.2.1 *Measuring Players' Offensive and Defensive Abilities*

We used two well-established metrics to indicate the players' offensive and defensive abilities:

- **Offense - Expected Point Value (EPV)** measures how many points a player is expected to make if he/she shoots from the current position. In basketball, it is a value between 0 and 3. Fundamentally, the goal of offensive tactics in basketball games is to maximize the EPV of the shooter. Thus, visualizing the EPV can help casual fans better understand and evaluate the in-game decisions of offensive players (e.g., pass or shoot). We obtained the EPV for each player based on their historical shot records. Specifically, we created a hexbin shot chart [71] for each player based on their historical shot records, in which the bins are grouped based on the shooting regions (defined by Official NBA Stats [5]). We then calculated the EPV per region by multiplying the player's field goal percentage and points they can make in the region. The results were cached as an EPV map for efficient access in each frame. Figure 7 shows an example EPV map.
- **Defense - Percentage Points Difference (DIFF%)** is a measure of a defender's ability to affect a shooter's shot percentage. Good defenders will have a negative DIFF% since they hold their opponent to a lower percentage than normal. For example, Stephen Curry's DIFF% is −3.6%, which means on average, a shooter's shot percentage will decease by 3.6% when being guarded by Curry. We acquired DIFF% by regions for each player directly from NBA Stats [5]. Besides DIFF%, the distance between a defender and the offensive player with the ball (DIST) is critical to the defensive performance. We calculated DIST based on the positional tracking data.

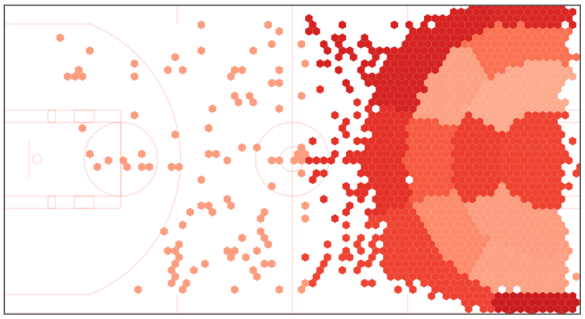

Figure 7: An EPV map of Stephen Curry based on his shooting records in the 2015-16 season. A darker color indicates a higher EPV.

5.2.2 *Visualizing Players' Offensive and Defensive Abilities*

Grounded in the design space proposed by prior work [41], we designed three embedded visualizations to present the offensive (EPV) metric, defensive (DIFF% and DIST) metric, and the one-on-one relationship between the defenders and the offensive player with the ball:

- **Offense Ring** (Fig. 8a) presents the player's location-based EPV, where a larger ring with darker color indicates better offensive ability. The inner and outer rings represent the minimum and maximum of possible EPV (i.e., 0 and 3). We used both the size and color of the middle ring to encode the player's EPV at the current position for easier interpretation.
- **Defense Shield** (Fig. 8b) represents the defender's location-based DIFF% and DIST in an arc shape, where a thicker and longer arc indicates better defensive ability. The thickness of the "shield" encodes the inverse DIFF% (a negative value) to make the visualization intuitive. The arc length of the "shield" encodes the subtraction of DIST from *maximum guarding distance*, since a larger DIST indicates lower pressure from the defender to the player with the ball. We displayed an outer border of the "shield" to show the maximal guarding distance (empirically selected as 12 feet) for comparison.

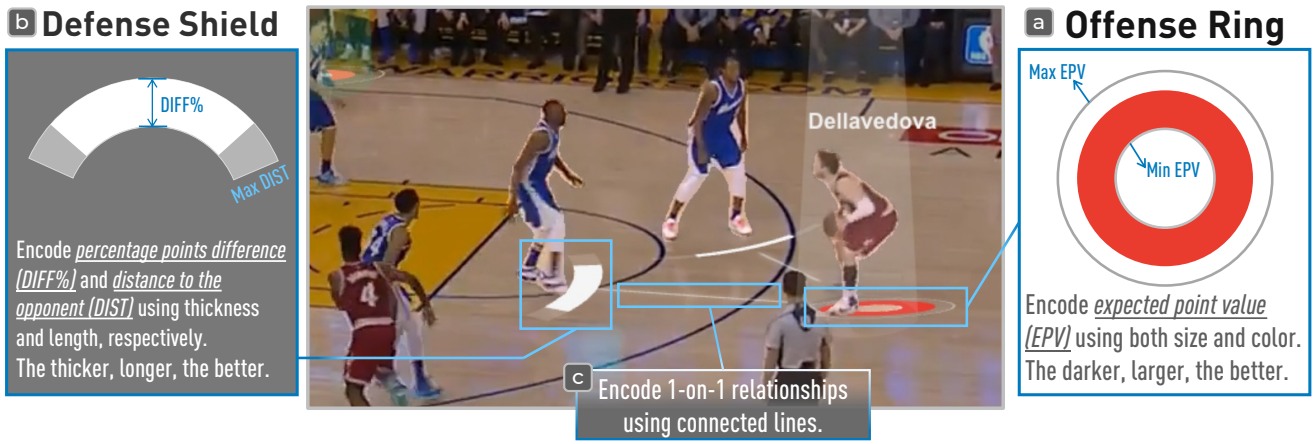

Figure 8: Three embedded visualizations for in-game data: a) Offense Ring shows the offensive performance of an offensive player. The darker, larger, the better. b) Defense Shield shows the defensive performance of the defender. The thicker, longer, the better. c) One-on-one Line shows the one-on-one relationship between the offensive player with the ball and the defenders.

- **One-on-one Line** (Fig. 8c) visualize the one-on-one relationship between the key defenders and the offensive player with the ball. The player with the ball can be defended by multiple defenders.

All these three visualizations are updated dynamically in the game based on the players' positions. We also darkened the background image to provide enough contrast for reading the visualizations.

5.2.3 *Design Process and Alternatives*

We finalized our designs through multiple rounds of iterations, especially for Offense Ring. Two considerations mainly drove our decision to use a ring placed on the ground — the visualization should 1) tightly connect with the target player and 2) avoid occluding other objects. Similar designs were used in previous research [41] and basketball video games [50]. Figure. 9a-c show some alternative designs we explored but none of them were satisfactory.

When designing the visual encoding of Offense Ring, we first used the size of the ring to encode a player's shooting frequency and a divergent color scale to encode the player's EPV, with the league average EPV as the midpoint. Figure. 9d shows an EPV map we created based on this encoding schema. However, in a pilot study, we found that this encoding schema was too complex to interpret for casual fans. For example, when the size of the ring is small (low shooting frequency) but the color is dark blue (high EPV), casual fans cannot judge if this is a good chance for the player to shoot or not. Thus, we decided to remove the encoding of shooting frequency and use both size and color to encode EPV. However, this could still be confusing as the size scale is sequential but the color scale is divergent. Consequently, we decided to use a sequential color scale instead of a conventional divergent one. This design was found to be easy to understand with clear messages (i.e., the bigger and darker, the better) to improve game understanding for casual fans. Different design decisions could be made for other purposes or fans, e.g., for analytic purposes or die-hard fans [41].

5.3 Gaze-based Interactions

To help casual fans seamlessly and efficiently access data of players they are interested in while watching the game (R3), we explored

Figure 9: Left: Three design alternatives for Offense Ring. a) Displaying the data on top of the player can occlude other players. b) Moving the visualization higher (e.g., the design in CourtVision [19]) can make it hard to connect to the target player. c) Displaying the data aside of the players (e.g., the shot meter in NBA 2K [50]) can also occlude other players. Right: An experimental EPV map of Steven Curry encodes his shooting frequency and EPV by using the size and divergent color scale. Different from Fig. 7, the bins in this EPV map are not grouped by regions.

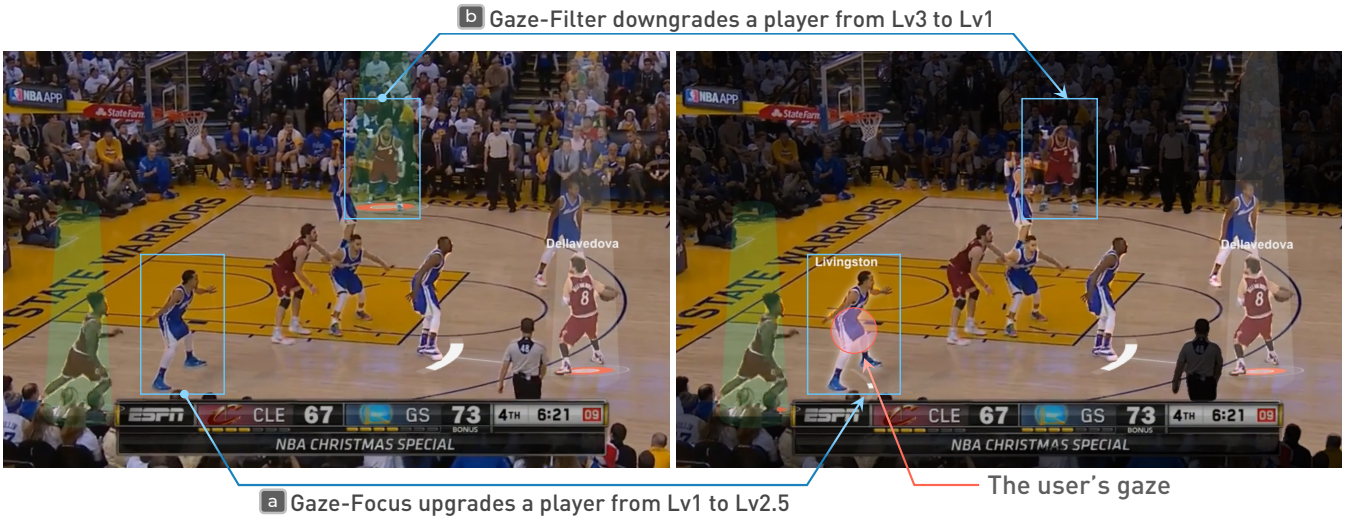

Figure 10: Two gaze interactions to adjust the embedded visualizations: a) Gaze Focus lifts the importance level and shows in-game data of players that are of interest to the user. b) Gaze Filter drops the importance level of video objects who are not the user's focus (e.g., open players and audiences out of focus).

using gaze as an input signal and designed two gaze interactions (Fig. 5c). The fans can still be guided by the visualizations to identify important offensive players.

5.3.1 *Gaze Focus – Fetching Data of Players of Interest to the User*

Gaze Focus allows the user to express their interests in players through gaze dwelling, which lifts the importance level of the players. Gaze Focus comprises the following three considerations:

- **Trigger**: According to our formative study, the users' gaze can move rapidly between, and across, players. To avoid showing data of players glanced over by the user accidentally, we defined a "dwell time" [54] for the interaction. The user needs to dwell her/his gaze on a player for 0.25 seconds to trigger the interaction.
- **Visual feedback**: To help the user realize that she/he is gazing at a player and triggering the interaction, we designed a highlight effect (Fig. 6c) in which the glow of the player will gradually increase when the user is gazing at the player, until the interaction is triggered. This design provides continuous visual feedback for the user while conforming with the visual design of importance levels.
- **Outcome**: Once the user triggers the interaction, iBall lifts the targeted player to **Lv2.5** if she/he is currently at a lower level (Fig. 10a). As a result, the system will also visualize the name and offensive or defensive data of the player. Lastly, when the user moves her/his gaze away from the player, the player will stay in **Lv2.5** for 1.8s (selected empirically) before reverting to their original player importance level. We designed such a lasting duration to cope with the users' rapid saccade in game watching.

5.3.2 *Gaze Filter – De-emphasizing Video Objects Out of the Sight*

To prevent users from being overwhelmed by too many **Lv3** players, we designed *Gaze Filter* to turn off the green spotlights of open players beyond a pre-defined filter radius. Gaze Filter incorporates three considerations:

- **Trigger**: Generally speaking, the system should always avoid overwhelming the user. Thus, Gaze Filter is consistently triggered and updated when the user moves his/her gaze. It centers at the user's gaze point with a filter radius of 650px (empirically selected).
- **Visual feedback**: To indicate the user that the interaction is being triggered, we designed a radial blurring effect that darkens the audience outside the filter radius and updates dynamically. We smooth the movement of the radial blurring effect to prevent it from abruptly changing location due to the user's saccade. Note that the blurring effect will not be applied to players and the court to ensure their readability. This visual feedback can notify the user about the existence of the interaction while also creating a theater mode that helps the user to focus on and engage with the game.
- **Outcome**: The green spotlights, which are used to highlight offensive players with open spaces, outside the filter radius will be turned off (Fig. 10b). An ease-in-out effect is applied to the change.

6 USER STUDY

To assess the usefulness, usability, and engagement of using our gaze-moderated embedded visualizations in watching basketball videos, we conducted a comparative study between three modes – watch with raw footage (RAW), with solely embedded visualizations (AUG), and with gaze-moderated embedded visualizations (FULL).

6.1 Study Setup

6.1.1 *Participants and Apparatus*

We recruited 16 casual fans (F1-F16; M = 10, F = 6; Age: 18 - 35) through university mailing lists and forums after screening for their fandom levels and basketball knowledge. All participants watched *"less than 10 games per year"* and only *"know the basic rules"* of basketball. All participants had not participated in our formative study. The study was conducted in the lab with a 24-inch monitor. We followed the same ergonomic settings in the formative study but changed to use a Tobii eye tracker 5 (133Hz) [74] for more accurate gaze interactions. The study took about one hour and each participant was compensated with a $20 gift card.

6.1.2 *Design and Procedure*

The study consisted of two tasks, namely, **Task1 - RAW vs. AUG**, and **Task2 - AUG vs. FULL**. Each task compared two modes. For each task, we used a game video from the formative study and evenly split it into two video clips (each lasting around 4.5 minutes) for each mode. The videos and the order of modes in each task were counterbalanced across participants. Each session included the following phases:

Phase 1. Introduction (10mins). The study started with an introduction of the research motivation, the purpose of the study, and the protocol. After the participant signed a consent form, we conducted a warm-up interview about basketball game-watching experiences.

Phase 2. Comparative Tasks (40mins). We asked participants to finish two comparative tasks (each lasted 20 mins): In **Task1 - RAW vs. AUG**, the participants watched two video clips in RAW and AUG modes, respectively. Before AUG mode, we conducted a training session to walk the participant through the four embedded visualizations in a separate video (about 20 seconds). We only proceeded to the task when participants were clear and confident enough to use the embedded visualizations. In **Task2 - AUG vs. FULL**, the participants watched another two video clips in AUG and FULL mode, respectively. Again, a training session was conducted before starting FULL mode to ensure the participant were confident to use the gaze interactions. Participants were encouraged to think aloud about their game observations when watching the videos. At the end of each video clip, we performed a post-video interview to collect the participants' feedback on the mode they had just experienced. At the end of each task, participants filled out a post-task questionnaire to rate their experiences.

Phase 3. Post-study Questionnaire (10mins). We asked participants to complete a post-study questionnaire of their subjective ratings on the overall system, rank the three modes, and provide feedback on the entire system.

6.1.3 *Measures*

We collected quantitative measurements of user subjective ratings in the post-task and post-study questionnaires. At the end of each task, participants were asked to rate the usefulness, engagement, and usability of the features they had just experienced, including the four visualizations (i.e., Player Highlight, Offense Ring, Defense Shield, and One-on-one Line) in **Task1** and the two gaze interactions (i.e., Gaze Focus and Gaze Filter) in **Task2**, on a 7-point Likert scale. In the post-study questionnaires, we asked participants to rate the overall system on five questions about system usefulness and engagement [52] and ranked the three modes.

6.2 Study Results

We first report the ratings of the overall system and the rankings of the three modes, and then discuss the feedback on the usefulness, engagement, and usability of individual visualizations and interactions.

6.2.1 *The overall user experience of iBall was predominantly positive, with FULL being most preferred*

Figure 11 left shows that the majority rated iBall as *"helpful"* and *"fun"*, felt *"in control"* and *"encouraged"* when using the system, and were *"likely to use"* it for watching basketball games. Figure 11 right presents the rankings of the three modes, showing that FULL was the most preferred mode by 12 participants, followed by AUG by 4 and RAW by none. The four participants who didn't rank FULL as the best were mainly concerned about the blurring effect of the audience in Gaze Filter, stating that a game video without audiences seemed abnormal. However, they agreed that the filtering of open players (highlighted in green) was useful. Thus, the system should allow users to turn off the blurring effect.

6.2.2 *Embedded visualizations are more useful if they are more attractive and informative*

Participants rated positively on the usefulness of each visualization (Fig. 12a). Among the four embedded visualizations, Player Highlight was considered to be the most useful in *"predicting the ball*

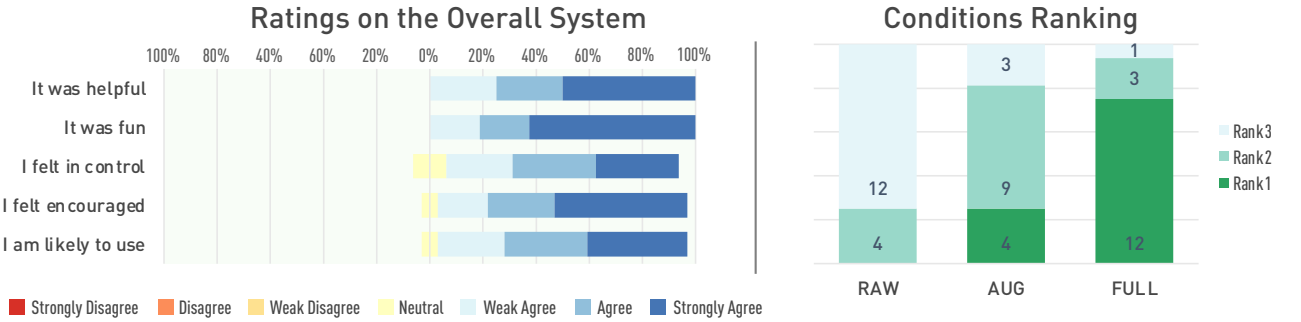

Figure 11: Left: Ratings on the overall system. Right: Rankings of different modes, i.e., RAW – watch with raw footage, AUG – watch with embedded visualizations solely, FULL – watch with gaze interaction and embedded visualizations.

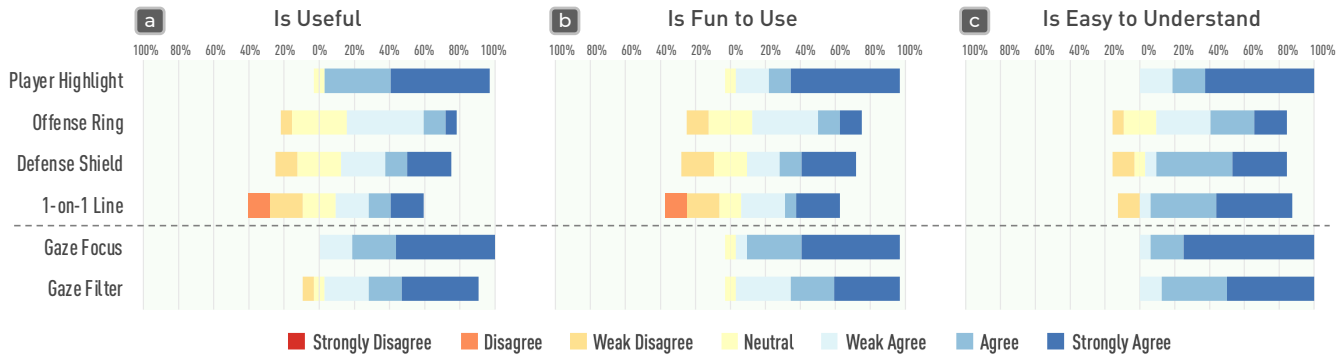

Figure 12: Ratings on the usefulness, engagement, and usability of the embedded visualizations and gaze interactions in iBall.

receiver" at the next pass, *"highlighting the open players"*, and thus *"making the game much clearer"*, as mentioned by the participants.

Participants also rated Offense Ring and Defense Shield as useful to understand players' in-game decisions. For example, F8 thought that Offense Ring could *"help me expect when he's gonna shoot. Otherwise, it's just like this person can shoot anytime."*; F3 commented that Defense Shield clearly showed *"who's able to defend"* and explained why *"James can easily score on Iguodala"*. However, some participants found them less useful since the visualizations were placed on the ground, different from the ball at hand-height, and thus often hard to notice in a tense game. Nevertheless, most of the participants agreed that *"it was nice to have"* Offense Ring and Defense Shield in the system.

One-on-one Line was controversial in terms of usefulness. Some participants considered it helpful *"to follow the game and to see who was involved [in defending]"* (F6) while others felt it suffers from the same limitations as Offense Ring and Defense Shield (i.e., cannot be noticed) with less useful information that was already *"clear from the video."* (F1)

6.2.3 *Gaze interactions are useful as they satisfy audiences' personal information and cognitive needs*

Both Gaze Focus and Gaze Filter received very positive ratings regarding usefulness (Fig. 12a). All participants enjoyed using Gaze Focus as it could help them immediately *"know the name of the player [who I am looking at]"* with simple and effective interaction. Gaze Filter also improved the game understanding of participants as it could *"make the scene tidier"* (F2) and *"focus the information on what I'm looking [at]."* (F4)

6.2.4 *Embedded visualizations are engaging since they provide a deeper epistemic pleasure*

The participants felt that the embedded visualizations were engaging (Fig. 12b) since they could help them *"understand and follow the game"*. This is also reflected in the positive relationship between the usefulness and engagement of the embedded visualizations – the more useful, the more engaging. For example, F1 commented that *"it is nice to know who the star players are because otherwise, they all look like regular players to me."* F5 remarked that he *"always tried to predict the receiver of a pass but always failed."* With the visualizations, he felt much more confident in predicting the receiver and felt quite a sense of accomplishment whenever he was correct.

6.2.5 *Gaze interactions are engaging since they promote proactive game viewing experiences*

Gaze interactions are also engaging (Fig. 12b) since they make the passive watching experience interactive and proactive. F2 said when the video scene responded to her gaze, she felt that she was *"a part of the game."* F4 particularly enjoyed using the gaze interactions, which made him *"almost feel like I'm there."* Moreover, participants provided that using the gaze interactions was *"natural"* and would not add further cognitive cost to the game watching since *"you can actively control it ... you don't have to [use it]"* (F5). Overall, the participants' feedback provides a strong hint that the engagement of watching sports videos can be improved if video content responds to the audience's gaze.

6.2.6 *The gaze-moderated embedded visualizations are easy to understand and use*

All participants confirmed the usability of the embedded visualizations and gaze interactions in iBall (Fig. 12c). They thought both the visualizations and gaze interactions were *"easy to understand"* and *"use"*. We noticed that while some participants thought that One-on-one Line was not that useful, they still agreed that it was easy to understand.

6.3 Implications for Designing Gaze-moderated Embedded Visualizations

We now discuss the design implications we learned from the feedback and observations in the user study.

6.3.1 *Visual attention matters when designing embedded visualizations*

Instead of being overwhelmed by the embedded visualizations, some participants sometimes even could not notice several of them. This may be due to a phenomenon known as Inattentional Blindness [33], in which viewers can fail to perceive visually salient objects or activities. This finding implies that designers should consider **how to properly direct the user's attention when designing embedded visualizations** to effectively convey information and avoid overwhelming the user. For instance, F5 suggested that we should highlight Offense Ring, instead of players, to help audiences efficiently identify the player with the highest EPV; Highlighting visualizations that are linked to immediate actions can increase the

information salience, e.g., highlighting the Offense Ring when a player is about to shoot.

6.3.2 *Synchronization of seeing and hearing matters when designing embedded visualizations*

F4 mentioned that he paid less attention to the commentaries in AUG mode but listened to them more carefully in RAW mode, which implied that the embedded visualizations could "overlay" the commentaries. In FULL mode, several participants reported using gaze interactions to search for players mentioned in the commentaries. These observations indicate that there is an interaction between the participants' perception of the embedded visualizations and the commentaries. Such an interaction between vision and hearing has long been identified (e.g., McGurk effect [73]). This highlights the importance of **synchronizing the embedded visualizations and the commentaries to create a consistent watching experience**. Some participants suggested leveraging the commentaries to create embedded visualizations, as explored in a recent paper [15].

6.3.3 *Gaze interactions shift the game from explanatory to exploratory*

In FULL mode, the participants spent more time using their gaze to highlight players, while in AUG mode, participants tended to follow the players highlighted by the system. This difference suggests that AUG mode is more explanatory while the addition of gaze interactions can shift it towards exploratory. This is not surprising, as the gaze interactions allows the audiences to actively explore the game more. When designing gaze interactions for game viewing systems or, broadly speaking, any situated visualization systems that involve visual guidance, designers must **consider the ultimate goal of the systems and strike a balance between explanatory and exploratory**.

6.3.4 *Gaze interactions enable active learning in game watching*

The gaze interactions can also help audiences learn basketball knowledge progressively. For example, Gaze Filter only highlights open players with green spotlights when the players are near the user's gaze. F3 used this feature to verify his hypotheses of team tactics by moving his gaze to some areas and seeing if the system *"showed green [highlighting]"* there. F15 elaborated that Gaze Focus helped him better recognize players by showing the name of a player to confirm that he was looking at the right person. Such a hypothesis-testing process made the participants feel more confident in interpreting the game. We see this as an interesting opportunity to **leverage gaze-moderated embedded visualization to develop long-term impact for the users** beyond improving their watching experience within individual games.

6.3.5 *Suggestions*

While the participants generally spoke highly of iBall, they did mention a few limitations related to the system implementations. For example, F8 disliked the visual artifacts introduced by the imperfect segmentation model; some participants said that Gaze Focus was not accurate when the players crowded together. These limitations can potentially be resolved with more advanced models or eye trackers. The participants also suggested several improvements for iBall, such as providing more customization options (e.g., for the visualizations and gaze interactions) through a *"Preference"* panel and allowing the visualizations to adapt to the pace of the game (e.g., e.g., showing more details in slow-paced and less in fast-paced situations). Another point worth mentioning is that a few participants wished the system could generate play-by-play replays with embedded visualizations to explain the game in detail. This could be an interesting direction for future research.

6.4 Feedback from Broader Users

While iBall is designed for casual fans, we also conducted a follow-up study with die-hard fans to explore its potential use beyond our original target users. We recruited 8 die-hard fans (D1-D8; M = 8; Age: 18 - 35), who knew *"basketball tactics and pros and cons of specific players"* and watched *"at least 1 game per week"*. No female die-hard fan responded to us. We followed the same process as in Sec. 6.1 to help the die-hard fans experience our system with a focus on collecting feedback on the real-world use of iBall. We discuss their major opinions that differed from those of the casual fans, as well as how iBall can be further improved.

6.4.1 *iBall can improve game understanding and engagement for die-hard fans*

All die-hard fans confirmed the usefulness of iBall in watching basketball videos. Unlike the casual fans (F1-F16), the die-hard fans could gain a deeper understanding of the game with the embedded visualizations. For example, they could further recognize the offensive tactics of the team from the highlighted open players. For some die-hard fans, the usefulness of iBall extends beyond understanding the games. D4 - D7, for example, were basketball players themselves and believed that iBall could help them improve their in-game decisions and tactics. On the other hand, the die-hard fans did request extra in-depth data that were currently not supported by iBall, such as the trajectories of the players' off-ball movements. Besides game understanding, the die-hard fans also agreed that iBall could enhance their engagement in game watching, especially the gaze interactions, which gave them a feeling of *"participating in the game."* (D1)

6.4.2 *Customization is indispensable for the die-hard fans*

The feedback from the die-hard fans indicates that there is no one-size-fits-all design that will satisfy everyone. Compared to the casual fans, the die-hard fans had more diverse opinions on the features. For instance, D5 preferred highlighting fewer open players, while D6 preferred highlighting more; D1 thought that showing the names could help him learn about unfamiliar players, while some participants only cared about the "star" players; D7 and D2 wanted more gaze interactions, but D4 found them distracting. While their preferences vary, they all agreed that there are valid reasons for the different design choices and that the best solution is to give users the option to customize the system, which is aligned with findings by Lin et al. [41]. One interesting question is how to help users efficiently express their preferences for customization, as the range of possible configurations can be very large.

6.4.3 *Embedded visualizations do not need to be displayed throughout the entire game*

Seven out of 8 die-hard fans thought that they did not need the embedded visualizations to be displayed throughout the entire game. They explained that iBall is most useful when players are executing the coach's strategy, such as in the first two quarters. However, when the game is decided by the "star" players' in-game states and improvisations (during crunch time), the embedded visualizations may be less useful. In addition, some participants (e.g., D4, D5) felt that watching a full game with gaze interactions can be exhausting, as they would *"keep using the interactions"*. This echoes our finding in Sec. 6.3.3 that gaze interactions promote proactive analysis. The participants suggested that the system should allow users to decide when to display the embedded visualizations.

6.4.4 *Gaze interactions should provide more adaptive data for the die-hard fans*

When using the gaze interactions, the die-hard fans wanted more adaptive data for different teams, players, and game events. For example, the retrieved data for the Golden State Warriors could focus on teamwork, while the data for the Cleveland Cavaliers could emphasize the performance of their "star" players. When gazing two "star" players facing off against each other, such as LeBron James vs. Steph Curry, the system could display their historical one-on-one records. Besides, the data could adapt to specific game events, such as dunking, or the intensity of the game. These suggestions, which would require a more intelligent and sophisticated system, are left for future research.

7 DISCUSSIONS

In this section, we will discuss potential future research directions and limitations of our current study.

Reproducible Environments For Embedded Visualizations Research. Compared to traditional web-based visualizations, embedded or situated visualizations are particularly challenging to research since the physical context where they are registered in is inherently difficult to reproduce, distribute, and benchmark. It can be more difficult, or even impossible, to reproduce the physical context if it is dynamic (e.g., sports scenarios). This perhaps is the major reason why most existing research [41, 80] uses reproducible simulated environments (e.g., virtual reality) to study embedded visualizations. In this work, we use videos to explore the design of interactive embedded visualizations in dynamic, complex scenarios. To advance research in embedded visualizations, we will open source our video-based environments (i.e., code and data) so that others can reproduce our system and develop their own.

Gaze Interactions for Embedded Visualizations. In recent years, eye-tracking technology has become increasingly affordable. Compared to other input modalities such as keyboard, mouse, and voice, gaze input can enable fast, intuitive, and implicit interactions. In fact, gaze interactions are widely supported in head-mounted displays [75] for augmented or virtual reality (AR/VR), which are the main scenarios for using embedded visualizations. Our research shows that even only using simple gaze data (i.e., the 2D position of the gaze point on the screen) can significantly increase the usefulness and engagement of embedded visualizations. However, gaze interactions have their own limitations, such as that they cannot be used in multi-viewers scenarios (e.g., TV in living rooms). Besides, we have not yet explored using advanced gaze events (e.g., fixation, saccade, pursuit) or combining gaze with other input modalities (e.g., speech) to achieve more adaptive or customized embedded visualizations, which we consider as promising future directions.

Towards Augmenting Live Game Viewing. In Sec. 4.4, we discussed the technical challenges and potential solutions for extending the CV pipeline to livestreams. Additionally, to support the embedded visualizations in livestreams, the system also requires real-time tracking data of the players. If this data is unavailable, a potential solution is to use camera calibration techniques [34, 61] to estimate the camera parameters, which can be used to estimate the players' positions and calculate the offensive and defensive metrics. On the other hand, real-world scenarios also provide additional resources for improving the system, including videos with a higher resolution and framerate, buffer time in streaming, camera parameters, and steering from human experts. Thus, we believe our system can be extended to live game videos by the broadcasters or researchers once these additional resources are available.

Augmenting Real-world Games Beyond Videos. With the development of sensing techniques and AR devices such as head-mounted displays, it becomes increasingly possible to augment real-world games with digital information. While our research provides a step-stone towards augmenting real-world environments, several issues must be taken into account when adopting it to AR, including the limited field of view, the effect of stereoscopy vision and depth perception, and the ability to freely change viewing perspective. We hope that the lessons learned from the present research can inspire and provide a solid foundation for future research on augmenting in-person game watching scenarios, ultimately generalizing embedded visualizations to general real-world environments.

Study Limitations. Our user study only evaluated the system on G1 and G2 rather than videos of entire basketball games. The study results, including the ratings and gaze distributions, only provide qualitative evidences. The designs of the embedded visualizations and gaze interactions in iBall are derived based on the 16 participants in our formative study. Further explorations are thus suggested for different scenarios and user groups.

8 CONCLUSION

This work explores using gaze-moderated embedded visualizations to facilitate game understanding and engagement of casual fans. We compared the game-watching behaviors of casual and die-hard fans in a formative study to identify the particular pain points of casual fans in watching basketball videos. Based on the findings, we developed a CV pipeline to support iBall, a basketball video-watching system equipped with gaze-moderated embedded visualizations. With iBall, casual fans can effectively identify key players, understand their in-game decisions, and personalize the game-viewing experiences through natural gaze interactions. We evaluated the CV pipeline with computational experiments. A user study with 16 casual fans confirmed the usefulness, engagement, and usability of iBall. We further collected feedback on iBall from 8 die-hard fans. The feedback of these 24 participants provides useful suggestions to improve iBall and insightful implications for future research in interactive embedded visualizations for sports game viewing.

ACKNOWLEDGMENTS

The authors wish to thank Salma Abdel Magid for her beautiful voice and help on the video narration. This research is supported in part by the NSF award III-2107328, NSF award IIS-1901030, NIH award R01HD104969, and the Harvard Physical Sciences and Engineering Accelerator Award.

REFERENCES

[1] 2022. DOTA 2. https://www.dota2.com/.
[2] 2022. Eyeware Beam. https://beam.eyeware.tech.
[3] 2022. League of Legends. https://www.leagueoflegends.com/.
[4] 2022. NBA Shot Charts. http://nbashotcharts.com/.
[5] 2022. NBA Stats. https://www.nba.com/stats/.
[6] 2022. Streaming Depth Data from the TrueDepth Camera. https://developer.apple.com/documentation/avfoundation/additional_data_capture/streaming_depth_data_from_the_truedepth_camera.
[7] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. 2022. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv preprint arXiv:2206.14651* (2022).
[8] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data. In *Proc. of EuroVis*. Eurographics Association. https://doi.org/10.2312/eurovisstar.20141173
[9] Michelle A. Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2016. Beyond Memorability: Visualization Recognition and Recall. *IEEE Trans. Vis. Comput. Graph.* 22, 1 (2016), 519–528. https://doi.org/10.1109/TVCG.2015.2467732
[10] Virginia Braun and Victoria Clarke. 2019. Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806
[11] Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Comput.* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86
[12] Zoya Bylinskii, Nam Wook Kim, Peter O'Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Frédo Durand, Bryan C. Russell, and Aaron Hertzmann. 2017. Learning Visual Importance for Graphic Designs and Data Visualizations. In *Proc. of UIST*. ACM, 57–69. https://doi.org/10.1145/3126594.3126653
[13] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. 2018. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1209–1218.
[14] Zhe Chen, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, and Yu Qiao. 2022. Vision Transformer Adapter for Dense Predictions. *arXiv preprint arXiv:2205.08534* (2022).
[15] Zhutian Chen, Qisen Yang, Xiao Xie, Johanna Beyer, Haijun Xia, Yingcai Wu, and Hanspeter Pfister. 2022. Sporthesia: Augmenting Sports Videos Using Natural Language. *To appear in IEEE Transactions on Visualization and Computer Graphics* (2022).
[16] Zhutian Chen, Shuainan Ye, Xiangtong Chu, Haijun Xia, Hui Zhang, Huamin Qu, and Yingcai Wu. 2022. Augmenting Sports Videos with VisCommentator. *IEEE Trans. Vis. Comput. Graph.* 28, 1 (2022), 824–834. https://doi.org/10.1109/TVCG.2021.3114806
[17] Shiwei Cheng, Zhiqiang Sun, Lingyun Sun, Kirsten Yee, and Anind K. Dey. 2015. Gaze-Based Annotations for Reading Comprehension. In *Proc. of CHI*. ACM, 1569–1572. https://doi.org/10.1145/2702123.2702271
[18] Xiangtong Chu, Xiao Xie, Shuainan Ye, Haolin Lu, Hongguang Xiao, Zeqing Yuan, Zhutian Chen, Hui Zhang, and Yingcai Wu. 2022. TIVEE: Visual Exploration and Explanation of Badminton Tactics in Immersive Visualizations. *IEEE Trans. Vis. Comput. Graph.* 28, 1 (2022), 118–128. https://doi.org/10.1109/TVCG.2021.3114861
[19] Clippers. 2020. Court Vision. https://www.clipperscourtvision.com/.
[20] Adrien Deliège, Anthony Cioppa, Silvio Giancola, Meisam Jamshidi Seikavandi, Jacob V. Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B. Moeslund, and Marc Van Droogenbroeck. 2021. SoccerNet-v2: A Dataset and Benchmarks for Holistic Understanding of Broadcast Soccer Videos. In *Proc. of CVPR*. Computer Vision Foundation / IEEE, 4508–4519. https://doi.org/10.1109/CVPRW53098.2021.00508
[21] Andrew T. Duchowski, Nathan Cournia, and Hunter A. Murphy. 2004. Gaze-Contingent Displays: A Review. *Cyberpsychology Behav. Soc. Netw.* 7, 6 (2004), 621–634. https://doi.org/10.1089/cpb.2004.7.621
[22] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim Dwyer, Jens Grubert, Jason H. Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matt Whitlock, and Yalong Yang. 2021. Grand Challenges in Immersive Analytics. In *Proc. of CHI*. ACM, 459:1–459:17. https://doi.org/10.1145/3411764.3446866
[23] ESPN. 2022. ESPN. https://www.espn.com.
[24] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun K. Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proc. of CHI*. ACM, 1118–1130. https://doi.org/10.1145/3025453.3025599
[25] Yu Fu and John T. Stasko. 2022. Supporting Data-Driven Basketball Journalism through Interactive Visualization. In *Proc. of CHI*. ACM, 598:1–598:17. https://doi.org/10.1145/3491102.3502078
[26] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. 2021. YOLOX: Exceeding YOLO Series in 2021. *arXiv preprint arXiv:2107.08430* (2021).
[27] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. 2021. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430* (2021).
[28] K Goldsberry. 2012. Courtvision: New visual and spatial analytics for the NBA, in '2012 MIT Sloan Sports Analytics Conference'. In *MIT Sloan Sports Analytics Conference*.
[29] Robert Greinacher and Jan-Niklas Voigt-Antons. 2020. Accuracy Assessment of ARKit 2 Based Gaze Estimation. In *Proc. of HCII (Lecture Notes in Computer Science, Vol. 12181)*. Springer, 439–449. https://doi.org/10.1007/978-3-030-49059-1_32
[30] Chunhui Gu, Chen Sun, David A. Ross, Carl Vondrick, Caroline Pantofaru, Yeqing Li, Sudheendra Vijayanarasimhan, George Toderici, Susanna Ricco, Rahul Sukthankar, Cordelia Schmid, and Jitendra Malik. 2018. AVA: A Video Dataset of Spatio-Temporally Localized Atomic Visual Actions. In *Proc. of CVPR*. Computer Vision Foundation / IEEE Computer Society, 6047–6056. https://doi.org/10.1109/CVPR.2018.00633
[31] Rex Hartson and Pardha S Pyla. 2012. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.
[32] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. GazeChat: Enhancing Virtual Conferences with Gaze-aware 3D Photos. In *Proc. of UIST*. ACM, 769–782. https://doi.org/10.1145/3472749.3474785
[33] Christopher G. Healey and James T. Enns. 2012. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 7 (2012), 1170–1188. https://doi.org/10.1109/TVCG.2011.127
[34] Min-Chun Hu, Ming-Hsiu Chang, Ja-Ling Wu, and Lin Chi. 2011. Robust Camera Calibration and Player Tracking in Broadcast Basketball Video. *IEEE Trans. Multim.* 13, 2 (2011), 266–279. https://doi.org/10.1109/TMM.2010.2100373
[35] Eakta Jain, Yaser Sheikh, Ariel Shamir, and Jessica K. Hodgins. 2015. Gaze-Driven Video Re-Editing. *ACM Trans. Graph.* 34, 2 (2015), 21:1–21:12. https://doi.org/10.1145/2699644
[36] Will Kay, João Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The Kinetics Human Action Video Dataset. *CoRR* abs/1705.06950 (2017). arXiv:1705.06950 http://arxiv.org/abs/1705.06950
[37] Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proc. of CHI*. ACM, 1–12. https://doi.org/10.1145/3313831.3376266
[38] Grete Helena Kütt, Kevin Lee, Ethan Hardacre, and Alexandra Papoutsaki. 2019. Eye-Write: Gaze Sharing for Collaborative Writing. In *Proc. of CHI*. ACM, 497. https://doi.org/10.1145/3290605.3300727
[39] Tiffany C. K. Kwok, Peter Kiefer, Victor R. Schinazi, Benjamin Adams, and Martin Raubal. 2019. Gaze-Guided Narratives: Adapting Audio Guide Content to Gaze in Virtual and Real Environments. In *Proc. of CHI*. ACM, 491. https://doi.org/10.1145/3290605.3300721
[40] Marcos Lage, Jorge Piazentin Ono, Daniel Cervone, Justin Chiang, Carlos A. Dietrich, and Cláudio T. Silva. 2016. StatCast Dashboard: Exploration of Spatiotemporal Baseball Data. *IEEE Computer Graphics and Applications* 36, 5 (2016), 28–37. https://doi.org/10.1109/MCG.2016.101
[41] Tica Lin, Zhutian Chen, Yalong Yang, Daniele Chiappalupi, Johanna Beyer, and Hanspeter Pfister. 2022. The Quest for Omnioculars: Embedded Visualization for Augmenting Basketball Game Viewing Experiences. *To appear in IEEE Transactions on Visualization and Computer Graphics* (2022).
[42] Jingyuan Liu, Nazmus Saquib, Zhutian Chen, Rubaiat Habib Kazi, Li-Yi Wei, Hongbo Fu, and Chiew-Lan Tai. 2022. VCoach: A Customizable Visualization and Analysis System for Video-based Running Coaching. *CoRR* abs/2204.08805 (2022). https://doi.org/10.48550/arXiv.2204.08805 arXiv:2204.08805
[43] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
[44] Antonio G. Losada, Roberto Therón, and Alejandro Benito. 2016. BKViz: A Basketball Visual Analysis Tool. *IEEE Computer Graphics and Applications* 36, 6 (2016), 58–68. https://doi.org/10.1109/MCG.2016.124
[45] Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 39–65.
[46] Raphael Menges, Chandan Kumar, and Steffen Staab. 2019. Improving User Experience of Eye Tracking-Based Interaction: Introspecting and Adapting Interfaces. *ACM Trans. Comput. Hum. Interact.* 26, 6 (2019), 37:1–37:46. https://doi.org/10.1145/3338844

[47] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2022), 99–106. https://doi.org/10.1145/3503250

[48] K. L. Bhanu Moorthy, Moneish Kumar, Ramanathan Subramanian, and Vineet Gandhi. 2020. GAZED- Gaze-guided Cinematic Editing of Wide-Angle Monocular Video Recordings. In *Proc. of CHI*. ACM, 1–11. https://doi.org/10.1145/3313831.3376544

[49] NBA. 2022. NBA Players & Team Rosters. https://www.nba.com/players.

[50] NBA2K23. 2022. NBA2K23. https://nba.2k.com/2k23/.

[51] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. In *Proc. of CHI*. ACM, 2093–2097. https://doi.org/10.1145/2858036.2858137

[52] Heather L. O'Brien and Elaine G. Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (Jan. 2010), 50–69. https://doi.org/10.1002/asi.21229

[53] Mershack Okoe, Sayeed Safayet Alam, and Radu Jianu. 2014. A Gaze-enabled Graph Visualization to Improve Graph Reading Tasks. *Comput. Graph. Forum* 33, 3 (2014), 251–260. https://doi.org/10.1111/cgf.12381

[54] Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber. 2012. Designing for the eye: design parameters for dwell in gaze interaction. In *Proc. of OzCHI*. ACM, 479–488. https://doi.org/10.1145/2414536.2414609

[55] Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete. 2013. SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Trans. Vis. Comput. Graph.* 19, 12 (2013), 2506–2515. https://doi.org/10.1109/TVCG.2013.192

[56] Charles Perin, Romain Vuillemot, Charles D. Stolper, John T. Stasko, Jo Wood, and Sheelagh Carpendale. 2018. State of the Art of Sports Data Visualization. *Comput. Graph. Forum* 37, 3 (2018), 663–686. https://doi.org/10.1111/cgf.13447

[57] Hannah Pileggi, Charles D. Stolper, J. Michael Boyle, and John T. Stasko. 2012. SnapShot: Visualization to Propel Ice Hockey Analytics. *IEEE Trans. Vis. Comput. Graph.* 18, 12 (2012), 2819–2828. https://doi.org/10.1109/TVCG.2012.263

[58] Kranthi Kumar Rachavarapu, Moneish Kumar, Vineet Gandhi, and Ramanathan Subramanian. 2018. Watch to Edit: Video Retargeting using Gaze. *Comput. Graph. Forum* 37, 2 (2018), 205–215. https://doi.org/10.1111/cgf.13354

[59] Vignesh Ramanathan, Jonathan Huang, Sami Abu-El-Haija, Alexander N. Gorban, Kevin Murphy, and Li Fei-Fei. 2016. Detecting Events and Key Actors in Multi-person Videos. In *Proc. of CVPR*. IEEE Computer Society, 3043–3053. https://doi.org/10.1109/CVPR.2016.332

[60] Yusuke Sano and Yohei Nakada. 2019. Improving Prediction of Pass Receivable Players in Basketball: Simulation-Based Approach with Kinetic Models. In *Proceedings of the Tenth International Symposium on Information and Communication Technology*. ACM, 328–335. https://doi.org/10.1145/3368926.3369671

[61] Long Sha, Jennifer Hobbs, Panna Felsen, Xinyu Wei, Patrick Lucey, and Sujoy Ganguly. 2020. End-to-End Camera Calibration for Broadcast Videos. In *Proc. of CVPR*.

[62] Lin Shao, Nelson Silva, Eva Eggeling, and Tobias Schreck. 2017. Visual Exploration of Large Scatter Plot Matrices by Pattern Recommendation based on Eye Tracking. In *Proc. of ESIDA@IUI*. ACM, 9–16. https://doi.org/10.1145/3038462.3038463

[63] Huang-Chia Shih. 2017. A Survey of Content-Aware Video Analysis for Sports. *IEEE TCSVT* 28, 5 (2017), 1212–1231.

[64] Nelson Silva, Tanja Blascheck, Radu Jianu, Nils Rodrigues, Daniel Weiskopf, Martin Raubal, and Tobias Schreck. 2019. Eye tracking support for visual analytics systems: foundations, current applications, and research challenges. In *Proc. of ETRA*. ACM, 11:1–11:10. https://doi.org/10.1145/3314111.3319919

[65] Nelson Silva, Tobias Schreck, Eduardo E. Veas, Vedran Sabol, Eva Eggeling, and Dieter W. Fellner. 2018. Leveraging eye-gaze and time-series features to predict user interests and build a recommendation model for visual analysis. In *Proc. of ETRA*. ACM, 13:1–13:9. https://doi.org/10.1145/3204493.3204546

[66] Nelson Silva, Lin Shao, Tobias Schreck, Eva Eggeling, and Dieter W. Fellner. 2016. Visual Exploration of Hierarchical Data Using Degree-of-Interest Controlled by Eye-Tracking. In *Proc. of Forum Media Technology and All Around Audio Symposium (CEUR Workshop Proceedings, Vol. 1734)*. CEUR-WS.org, 82–89. http://ceur-ws.org/Vol-1734/fmt-proceedings-2016-paper10.pdf

[67] SportVU. 2022. NBA Sportvu Dataset. https://paperswithcode.com/dataset/nba-sportvu.

[68] Manuel Stein, Thorsten Breitkreutz, Johannes Häussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel A. Keim, and Halldor Janetzko. 2018. Revealing the Invisible: Visual Analytics and Explanatory Storytelling for Advanced Team Sport Analysis. In *Proc. of BDVA*. IEEE, 1–9. https://doi.org/10.1109/BDVA.2018.8534022

[69] Manuel Stein, Halldor Janetzko, Andreas Lamprecht, Thorsten Breitkreutz, Philipp Zimmermann, Bastian Goldlücke, Tobias Schreck, Gennady L. Andrienko, Michael Grossniklaus, and Daniel A. Keim. 2018. Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2018), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

[70] Thomas Templier, Kenan Bektas, and Richard H. R. Hahnloser. 2016. Eye-Trace: Segmentation of Volumetric Microscopy Images with Eyegaze. In *Proc. of CHI*. ACM, 5812–5823. https://doi.org/10.1145/2858036.2858578

[71] Daniel Teo. 2022. NBA Shot Charts Part 2: Building the viz in Tableau. https://datavizardry.com/2020/02/03/nba-shot-charts-part-2/.

[72] Changjia Tian, Varuna De Silva, Michael Caine, and Steve Swanson. 2019. Use of machine learning to automate the identification of basketball strategies using whole team player tracking data. *Applied Sciences* 10, 1 (2019), 24.

[73] Kaisa Tiippana. 2014. What is the McGurk effect? , 725 pages.

[74] Tobii. 2022. Tobii Eye Tracker 5. https://gaming.tobii.com/product/eye-tracker-5.

[75] Vive. 2022. Vive Pro2. https://www.vive.com/us/product/vive-pro2/overview.

[76] Vizrt. 2022. Viz Libero. https://www.vizrt.com/products/viz-libero.

[77] Romain Vuillemot and Charles Perin. 2016. Sports Tournament Predictions Using Direct Manipulation. *IEEE Computer Graphics and Applications* 36, 5 (2016), 62–71. https://doi.org/10.1109/MCG.2016.90

[78] Wikipedia. 2022. Player tracking (National Basketball Association). https://en.wikipedia.org/wiki/Player_tracking_(National_Basketball_Association).

[79] Yufei Xu, Jing Zhang, Qiming Zhang, and Dacheng Tao. 2022. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. arXiv:2204.12484 [cs.CV]

[80] Lijie Yao, Anastasia Bezerianos, Romain Vuillemot, and Petra Isenberg. 2022. Visualization in Motion: A Research Agenda and Two Evaluations. *IEEE Trans. Vis. Comput. Graph.* (2022).

[81] Shuainan Ye, Zhutian Chen, Xiangtong Chu, Yifan Wang, Siwei Fu, Lejun Shen, Kun Zhou, and Yingcai Wu. 2021. ShuttleSpace: Exploring and Analyzing Movement Trajectory in Immersive Visualization. *IEEE Trans. Vis. Comput. Graph.* 27, 2 (2021), 860–869. https://doi.org/10.1109/TVCG.2020.3030392

[82] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. 2022. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917* (2022).

[83] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. 2022. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. (2022).

[84] Zhengyou Zhang. 2000. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 11 (2000), 1330–1334. https://doi.org/10.1109/34.888718

[85] Qiyu Zhi, Suwen Lin, Poorna Talkad Sukumar, and Ronald A. Metoyer. 2019. GameViews: Understanding and Supporting Data-driven Sports Storytelling. In *Proc. of CHI*. ACM, 269. https://doi.org/10.1145/3290605.3300499

[86] Qiyu Zhi and Ronald A. Metoyer. 2020. GameBot: A Visualization-augmented Chatbot for Sports Game. In *Proc. of Extended Abstracts of CHI*. ACM, 1–7. https://doi.org/10.1145/3334480.3382794

[87] Stefanie Zollmann, Tobias Langlotz, Moritz Loos, Wei Hong Lo, and Lewis Baker. 2019. ARSpectator: Exploring Augmented Reality for Sport Events. In *SIGGRAPH Asia 2019 Technical Briefs*. ACM, 75–78. https://doi.org/10.1145/3355088.3365162

A PLAYER DATASET

Table 5: Statistics for our player dataset.

	Num of Frames	Num of Labels
G1 - train	9552	8837
G1 - test	4029	3596
G2 - train	8312	6530
G2 - test	3485	3485

Table 6: Number of labeled instances for each player class in G1.

Class Label	# of instances
G30	12574
G9	12953
G23	12404
C0	12418
C8	12461
C23	12354
G11	11760
G34	11834
C4	10629
C2	6663
C5	5268
C13	1508
G31	643

Table 7: Number of labeled instances for each player class in G2.

Class Label	# of instances
O0	10375
L3	11115
O9	9761
O3	9407
O35	9194
L6	9774
L32	9401
O12	9235
L1	5576
L11	5632
L12	3864
L4	3632
O21	653
L33	642
O2	506

We created a dataset for the two videos (G1 and G2) used in our formative study. For both videos, we first removed all transition scenes (e.g., replays) since transition scenes typically show close-up views of the players and can be noise for the detector. We gathered 13581 and 11797 frames for G1 and G2, respectively. For each frame, we identified the players with at least half of the body in the scene and labeled their bounding boxes with the players' unique IDs. Occluded players were also labeled if at least half of their bodies were in the scene. In total, there are 13 and 15 unique IDs (i.e., classes) in G1 and G2, respectively. Table. 5 shows an overview statistics of our dataset. Table. 6 and Table. 7 provide a detailed breakdown of the number of labeled instances the datasets have for each class. The first letter of the class names (except Negative) indicates the player's team (where G = Golden State, C = Cleveland, O = Oklahoma City, and L = Los Angeles), and the number that follows is the number of the player's labels.