Machine Learning-Based Play Evaluations from Racket Sports Videos

Ning Ding

Abstract

Artificial Intelligence (AI) technology has fundamentally transformed our daily lives in areas such as autonomous vehicles, healthcare, and gaming. For instance, AlphaGo, a well-known AI, has proven its superiority by defeating top-tier Go players. However, applying AI to sports analysis poses unique challenges, primarily due to the unpredictability and complexity of sports, contrasting with the more contained and predictable environments of board games. In sports, assessing play performance is vital for effective coaching decisions. While traditional manual evaluation methods are time-consuming and struggle with complexity, machine learning-based assessment offers an efficient and objective alternative. Thus, in this thesis, a machine learning-based approach is employed to comprehensively evaluate games, considering their context and dynamics. In particular, play assessment is explored in racket sports, which, despite their widespread popularity, present unique challenges in performance assessment due to their dynamic nature and rapid gameplay.

Utilizing deep learning techniques, two distinct frameworks are proposed to assess play performance based on video in both singles and doubles racket sports. Firstly, for play evaluations in singles matches, a novel evaluation method of play performance for badminton matches from broadcast videos utilizing deep reinforcement learning is proposed. Unlike traditional methods that primarily focus on outcomes, this approach uses historical data, including information about the tactical and technical performance of players, to learn the next-score probability as a Q-function. This function is then applied to assess the value

of each stroke, offering a detailed understanding of performance on a stroke-by-stroke basis. Secondly, for play evaluations in doubles matches, a pioneering framework of deep neural networks is proposed to estimate the control area probability map for badminton doubles from drone video. This research introduces the first annotated drone dataset for badminton doubles, captured from the top and back views. Additionally, a practical application is proposed for assessing optimal positioning, providing insights that can be instrumental for coaching.

These approaches are validated in this thesis by comparing them with various baselines and examining the correlations between evaluation results and multiple standard success measures. The findings indicate that the proposed frameworks could effectively provide quantitative evaluations for singles and doubles matches. Furthermore, these assessments are demonstrated visually, offering a nuanced and insightful approach to coaching interventions, ultimately leading to improved outcomes for players and teams.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Artificial Intelligence (AI) has led to significant advancements and has profoundly transformed our lifestyles and work methods. For instance, autonomous vehicles can navigate complex traffic scenarios without human intervention. In healthcare, AI-powered diagnostic tools can detect and predict illnesses with unprecedented accuracy. The capabilities of AI were vividly displayed when AlphaGo [3] astonishingly defeated top-tier Go players. Recently, Japan's Shogi AI, which provides real-time winning probabilities for players, has enriched the viewing experience for audiences. However, sports analysis poses unique challenges for AI. The unpredictable nature of player movements, swift game dynamics, and varying strategies in sports make them harder for AI to analyze compared to the finite moves in board games.

In sports, evaluating player performance has always been a focal point. Athletes, regardless of their chosen discipline, share a common goal; the pursuit of excellence. The ability to assess and analyze play performance has significant implications for enhancing skills, making informed player selections, and scouting new talents.

Traditionally, performance evaluation in racket sports heavily relied on manual assessments conducted by experienced experts, introducing subjectivity and potential biases into the evaluation process [4, 5]. Additionally, manual evaluation struggles to unravel intricate patterns owing to its inherent limitations in handling complexity. This approach often demands the involvement of multiple analysts, consuming valuable time and resources. In contrast, machine learning-based evaluation methods offer a more objective alternative, mitigating human biases and excelling in managing the complexity inherent in analyzing racket sports plays [6, 7]. Once trained and deployed, these models can rapidly evaluate extensive volumes of data, making them highly scalable and efficient.

Deep learning, a subset of machine learning, is a powerful tool for pattern recognition and decision-making, closely mirroring the neural networks of the human brain. Its applications span various domains, including image recognition, computer vision, and natural language processing, and have shown remarkable promise in sports analytics [8, 9, 10]. Therefore, it stands out as an ideal candidate for learning-based play evaluations in racket sports owing to its intrinsic capacity to handle and interpret vast and intricate datasets. Within sports, data instances are manifold and often intertwined comprising positional data, motion trajectories, and event sequences. As a result, deep learning excels in uncovering complex patterns and relationships [11, 12] that are usually invisible to manual evaluators.

Racket sports, despite their widespread popularity, present unique challenges in play performance assessments due to their dynamic nature and rapid gameplay. This gap between the existing achievements of deep learning methods and the complexities of racket sports underscores the need for a better understanding of game tactics and strategies. A rally in racket sports like badminton involves a series of strokes, each contributing to the final points and reflecting strategic thinking, just as in games like chess, Shogi, and Go [13]. In both racket sports and these games, players must consider every move carefully from a long-term perspective. Thus, assessing the value of each stroke can provide insights into which steps are relatively critical in winning or losing a point. Additionally, participating in doubles games offers a different playing experience. Even if two players possess high-level technical skills in badminton, it does not guarantee that they can effectively cooperate to achieve a performance exceeding the sum of their individual skills, often described as the "$1 + 1 > 2$" effect. Top-class players also need to become familiar with each other and collaborate extensively. To achieve an "$1 + 1 > 2$" outcome, the initial step for players is to understand where to move based on the game situation and their partner's position. Comprehending such dynamics is crucial, requiring a study of the relationship not only between the players but also between the players and the shuttlecock/ball in a spatial context. These personal insights into the subtleties of racket sports emphasize the importance of detailed and accurate play analysis. However, three key challenges arise when attempting to translate these personal insights using a deep learning method.

Firstly, data acquisition in sports such as badminton or tennis poses significant challenges owing to the high-speed nature of these games. They require high-frame-rate recording systems to capture players' rapid movements and ball trajectories adequately. Existing datasets for racket sports are limited, with some providing partial game information, such as player IDs, strokes, and match points [14, 15], while others remain unreleased [16, 17]. Additionally, conventional broadcast videos, while being a rich source of data, present several disadvantages, notably the frequent perspective changes, occlusions, and partial view issues [18, 19].

Secondly, in contrast to many team sports where collective movement and strategy often overshadow individual technical actions, racket sports, with fewer participants and a heavy reliance on individual technical prowess, demand detailed pose analysis. Every stance, swing, and step in these sports can significantly influence the match outcome [20, 21]. Consequently, an integrated multi-modal data processing system is required to effectively manage visual, spatial, and temporal data.

Lastly, the challenge in visual analysis for racket sports lies in advancing beyond traditional discrete event analysis. While existing studies have made valuable contributions to

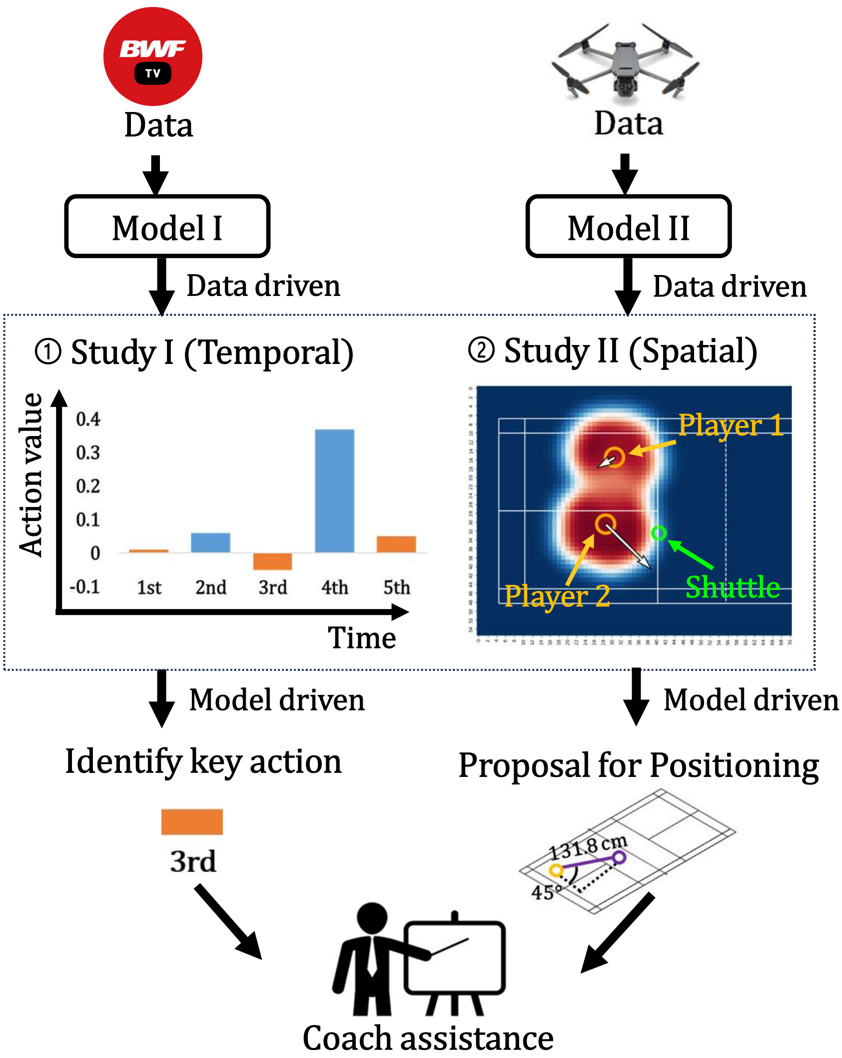


Figure 1.1: Overview of the thesis. Play performance is evaluated in both studies. Study I focuses on the temporal dynamics of action value in a singles match, while Study II focuses on the spatial dynamics in a doubles match to provide coaching insights.

understanding player performance and basic patterns, there is a need to explore and develop advanced visual analysis techniques capable of capturing the intricate spatial relationships among players, shuttlecock/ball, and court.

Fig. 1.1 provides a comprehensive overview of this thesis. Study I focuses on the temporal dynamics of individual player movements in a singles match, while Study II examines spatial team dynamics in a doubles match. Temporal visual analysis helps in understanding which actions are critical, whereas spatial visual analysis provides insights into how to achieve effective positioning. Together, they offer a comprehensive view of performance in racket sports. Both studies contribute to a more nuanced understanding of sports performance.

The primary objective of this thesis is to offer a quantitative assessment of play performance in singles and doubles badminton matches through learning-based evaluation from video. Before diving into the chosen methodologies, it is crucial to specify the challenges primarily addressed in this thesis:

- This thesis tackles the challenge of integrating detailed players' poses into the deep neural networks framework. Study I incorporates these poses as components of the 'State' in deep reinforcement learning, while Study II utilizes graph convolutional embeddings to integrate players' poses.
- This thesis addresses the challenge of inadequate quantitative assessments and visualizations for spatial and temporal analysis. Study I quantitatively evaluates individual strokes and offers temporal visualizations, while Study II assesses team plays in densely distributed areas and focuses on providing spatial visualizations.
- Study II addresses the challenge of providing only partial information about the game in broadcast videos by utilizing drone technology.

1.2 Scope of the Thesis

This thesis focuses on the quantitative analysis of plays in singles and doubles matches in racket sports, particularly emphasizing badminton, a domain where such quantitative evaluation methodologies have rarely been explored. Two innovative frameworks are proposed for incorporating pose information into play performance analysis and quantification from video. These frameworks utilize advanced deep learning techniques and drone technology. Additionally, this research aims to visualize evaluation results and offer actionable coaching insights.

Firstly, two distinct studies that advance the methodologies for play performance evaluation are discussed in this thesis. The challenge of quantitatively assessing play-by-play evaluations in a singles badminton match is addressed (Study I). Secondly, the challenge presented by the absence of quantitative assessments in densely distributed spatial domains for racket sports is tackled (Study II). The scope of these two studies is detailed as below.

1.2.1 Study I: Action Value Assessment

Study I offers a more nuanced analysis than traditional game scores by evaluating the action value of each play and introducing an innovative evaluation method using Deep Reinforcement Learning (DRL). A DRL model provides a unified evaluation criterion; Each time a player wins a rally, they are rewarded with one point. Utilizing time series data that captures both tactical and technical performances, the proposed method predicts the probabilities of the subsequent score. As a result, a granular assessment of players' actions is proposed, thereby advancing performance evaluation in racket sports.

1.2.2 Study II: Control Area Estimation

Study II investigates the dynamics of teamwork in doubles matches, with a particular emphasis on synchronized positioning, which is crucial for winning a match. Recognizing the significance of court control for doubles teams, the first annotated drone dataset capturing top and back views in doubles badminton matches is created. This dataset facilitates the estimation of control area probability maps. The proposed framework is based on deep neural networks and employs a dual-stream approach to capture both tactical and technical performances for control area estimation. Since data have been annotated for direct modeling, the proposed method is more straightforward and accurate than reinforcement learning. Furthermore, practical applications that enable optimal positioning strategies are proposed for coaching guidance.

1.3 Thesis Overview

The remainder of this thesis is organized as follows: Chapter 2 discusses related work on sports evaluations. Chapter 3 introduces Study I, which focuses on the development of a DRL method for racket sports to estimate the action value of each player from an input video. Chapter 4 introduces Study II, which includes the introduction of an annotated drone dataset for badminton doubles matches and proposes an efficient deep neural network framework for evaluating doubles teamwork. Finally, Chapter 5 summarizes the contributions of this thesis and discusses future work.

2 Literature Review

The content of this thesis can be categorized into three distinct research categories. The first category relates to play evaluation methods in sports, covering quantitative analysis across various sports, specifically racket sports. The second category relates to research on visual analytics, including its applications in team sports and racket sports. The third category relates to sports datasets in team sports and racket sports. In this chapter, literature on each category is introduced and summarized at the end.

2.1 Play Evaluation

2.1.1 Quantitative Play Evaluation in Team Sports

Traditional methods for evaluating sports plays primarily involve manual evaluation, typically obtained from video or statistical data. The manual evaluation methods through video include the systematic review of game footage by coaches and analysts. They observe and watch game videos multiple times to evaluate both individual and team performances. For instance, Franks et al. [22] developed a system enabling observers to systematically record and view the behavior of athletes during team sports competitions. Other traditional methods relied on basic statistical data such as goals, assists, and other easily quantifiable metrics [23]. For example, the distribution of the number of batters faced and the number of runs scored in an inning is used to model pitcher performance in Major League Baseball (MLB) [24]. An econometric model is developed to connect individual player statistics in the National Basketball Association (NBA) games with team wins, offering a method to measure a player's productivity and their marginal contribution to the team's success [25].

Machine learning-based evaluations have been applied to various sports contexts in many team sports. Some useful metrics have been proposed to provide a comprehensive overview of a player's performance in a match. For example, Cervone et al. [26] proposed the Expected Possession Value (EPV) to evaluate players' decisions and actions in basketball. This method uses spatio-temporal tracking data to quantify the expected value of ball possession at any moment, considering the positions and movements of all players and the ball.

Rather than assessing a player's overall play performance in a match, other methods employ models to evaluate individual actions taken during the match. In soccer, the Valuing Actions by Estimating Probabilities (VAEP) method and its variants, as referenced in [27, 28], assess the impact of individual player actions on the likelihood of scoring by calculating the probabilities of various outcomes following a player's action. A comprehensive survey on these methods in team sports was conducted by Fujii et al. [29]. Additionally, to evaluate the shooting action of players, researchers have extended the concept of Expected Possession Value (EPV) [26, 30, 31] and the value of space [32, 33] by utilizing advanced techniques like Voronoi diagrams [34], which partition the playing field into regions based on the proximity of players to assess spatial control and influence.

In addition to these approaches, reinforcement learning frameworks have been applied in research papers for team sports. These studies have utilized inverse planning methods in reinforcement learning, which involves estimating the action model or reward function from observed data using statistical learning techniques. For instance, the state-action value function (Q-function) of a player was estimated using a recurrent neural network [35, 36], providing insights into the effectiveness of player decisions in specific game states. This approach was further analyzed using a linear model tree [37] to enhance its comprehensibility for human interpretation, addressing issues of interpretability that often accompany deep learning models. However, the broader applicability and adaptation to sports with different dynamics, like racket sports, remain a challenge, primarily due to the unavailability of detailed pose and movement data specific to these sports.

2.1.2 Quantitative Play Evaluation in Racket Sports

In racket sports, methods for action evaluation can be categorized into manual evaluation and machine learning-based approaches. Manual evaluation methods are straightforward and widely adopted by analysts. Analysts organize game videos in various ways and repeatedly assess stroke performances based on their expertise and knowledge [4, 5, 38]. However, manual evaluation relies on the expertise and experience of evaluators, which can introduce subjectivity and bias into the assessment process and are also time-consuming. Other methods rely on basic statistical data including serve percentages, unforced error percentages, aces, winners, and scores. For example, a novel test, known as the Leuven Tennis Performance Test (LTPT), has been proposed to assess stroke performance in elite tennis players under match-like conditions [4]. Additionally, a tennis skill test for forehand and backhand drive strokes was developed, utilizing a pneumatic ball machine and time-based scoring, and has proven to be both reliable and valid in assessing the skills of novice tennis players [39]. Other existing methods emphasize notational and temporal variables, using statistical data to analyze player behavior, such as forced or unforced errors, to evaluate the likelihood of a player's victory [40, 41].

Despite the popularity of racket sports, only a limited number of works have offered computer vision-based solutions for play evaluation tasks. Machine learning-based evaluation, also known as data-driven evaluation, involves the extraction of stroke features from videos [42, 43, 44] to characterize and model the competition process in advance. Subsequently, it employs deep learning or other approaches to evaluate stroke performance. In DeCoach [45], a distance-based methodology was proposed to compare a player's pose with that of a professional player, which is then utilized for stroke evaluation.

Pfeiffer et al. [42] pioneered in adopting the Markov model approach for play performance diagnosis in table tennis. McGarry and Franks [46] used the Markov model to explain championship performance in squash. However, these studies discretized the coordinates of location and time, resulting in the loss of information and the inability to generalize the unobservable parts of the state space. Wang et al. [47] integrated the knowledge of analysts and trained a classifier that learned to evaluate strokes based on their assessments, yielding quantified evaluation results. However, their approach utilized domain knowledge specific to table tennis, which differs from the general approach used for modeling various racket sports. More recently, Wang et al. [12] introduced a specialized badminton language to describe the process and predict the winning probability in a rally. However, their method could not directly estimate the value of each stroke.

2.2 Visual Analytics

2.2.1 Visual Analytics in Team Sports

In general, visual analytics can be described as "the science of analytical reasoning facilitated by interactive visual interfaces" [48]. It has developed rapidly and is widely used in analyzing various sports [49, 50]. In soccer, the Time-to-intercept method is used to calculate a pitch control function that quantifies and visualizes the regions of the pitch controlled by each team [51]. Another approach uses a generative model for multi-agent trajectory data and visualizes the predicted trajectory of players in both soccer and basketball [52].

Other game analysis methods use non-sports-specific visualizations. One of these visualizations is the heatmap, which visualizes the most frequent locations of game events by density. Recently, a deep neural network called SoccerMap has been proposed to estimate the full probability surfaces of potential passes in soccer [53]. SnapShot [54] introduced a specific type of heatmap called radial heatmap to display shot data in ice hockey, while CourtVision [55] quantified and visualized the shooting range of players in basketball. Another commonly used visualization method is the flow graph, where the nodes' size shows each player's role, and the links show the connections between them [56]. Besides, the glyph-based visualization method has also been applied to sports. For example, MatchPad [57] adopted a glyph-based visual design to analyze players' performances during rugby games. All these works demonstrate the need for visual analytics, its impact, and its potential in sports.

2.2.2 Visual Analytics in Racket Sports

Visual analytics in racket sports involves the use of data visualization and interactive tools to analyze and interpret large datasets related to aspects like player performance, match statistics, and tactical patterns.

Most previous studies focused on analyzing and assessing players in singles matches and visualizing discrete representations such as strokes. In tennis, CourtTime [58] introduced a novel visual metaphor to facilitate pattern detection, while TenniVis [59] visualized statistical data, such as score and service information. In table tennis, Tac-Miner [60] developed a visual analytics system to facilitate simulative analysis based on the Markov chain, while iTTVis [61] used a matrix to reveal the relationship among multiple attributes within strokes. Tac-Simur [43] provided an interactive exploration of diverse tactical simulation tasks and visually explained the simulation results. RallyComparator [62] introduced an interactive visualization system to analyze and compare stroke sequences in table tennis, enabling experts to identify new, complex playing patterns efficiently. In badminton, TIVEE [63] studied tactic analysis in a 3D environment and proposed immersive visual analytics. Recently, Haq et al. [64] tracked players and visualized their position statistics on a heatmap. While previous studies have made significant strides in visualizing certain aspects of the game, such as stroke analysis, scoring, and service information, they generally have not delved deeply into the spatial and temporal dynamics of player movements and game events. One reason the exploration of spatial distributions of game events in racket sports remains limited is due to the differing importance of understanding spatial distribution in singles and doubles matches. In singles matches, the analysis of spatial distribution, such as court coverage, is not as critical as in doubles. This oversight leads to the neglect of dense spatial distribution analysis, especially in doubles matches.

2.3 Sports Datasets

2.3.1 Team Sports Datasets

The development of sports datasets has significantly evolved over time, with a marked shift towards providing more detailed and comprehensive information. Initially, datasets such as the University of Central Florida (UCF) sports action dataset [65], Olympics Sports [66], and Sports-1M [67] focused on categorizing different types of sports through the action recognition of individual players. These datasets played a crucial role in the early stages of sports analysis by enabling the recognition and categorization of various sports actions through video analysis.

With the further advancement of sports analytics, there was a growing need for datasets that not only recognize actions but also localize them precisely within the video frames.

This need led to the creation of more advanced datasets. Recent efforts in this direction include the development of larger broadcast sports video datasets. In soccer, SoccerNet [68] and SoccerNet-v2 [69] have been proposed for event spotting. Another large-scale dataset SoccerDB [70] offers a comprehensive benchmark for video understanding, suitable for object detection, action recognition, temporal action detection, and video highlight detection. In basketball, the National Collegiate Athletic Association (NCAA) basketball dataset [71] includes broadcast videos of matches and provides dense, temporal event annotations in long videos. Northwestern Polytechnical University (NPU) RGBD Dataset [72] provides not only RGB frames but also depth maps and the skeleton of players. In ice hockey, broadcast videos of National Hockey League (NHL) games have been extensively utilized to create datasets [73, 74]. However, broadcast videos typically do not show the entire pitch, providing only partial information about the game. To overcome this limitation, a fisheye camera has recently been used in basketball [75]. Similarly, SoccerTrack [76] utilized fish-eye and drone cameras to capture the whole field in a single frame.

2.3.2 Racket Sports Datasets

Similarly, the evolution and characteristics of datasets in racket sports have changed over time. Early efforts in data collection for racket sports predominantly involved manual methods, focusing on basic statistics such as scores and player movements. However, these methods often provide only partial game information, such as player IDs, strokes, and match points.

With technological advancements, there is a shift toward automated data collection methods, notably through the use of high-definition cameras and motion-tracking systems. In tennis, Zhao et al. [77] gathered data from a tennis court equipped with the PlaySight system, consisting of six High-Definition (HD) cameras. Pulgarin-Giraldo et al. [78] collected tennis data using a Motion Capture (MoCap) system. These works collected data using multiple cameras and business data analysis systems, which are noted for their high cost and complexity. THree dimEnsional TennIs Shots (THETIS) action dataset [79] contains 12 types of tennis shots. However, this dataset was collected indoors using Microsoft Kinect, focuses on swinging motions in an exercise context, and cannot be used to effectively capture player motions typically observed in real tennis matches. For table tennis, the TTStroke-21 dataset [80] is composed of 129 videos at 120 fps, recorded from an egocentric perspective in real match conditions. However, this dataset does not capture the entire table tennis scene in a single frame. The SPIN dataset [81] is a new high-resolution, high-frame-rate stereo video dataset. It was captured using two high-speed cameras and can be utilized for multiple tasks, including ball tracking, pose estimation, and spin prediction. Some of these self-collected datasets are available while others remain unreleased [16, 17]. Additionally, conventional broadcast videos, particularly those from the Badminton World Federation (BWF) on YouTube[1] have been widely used in academic data analysis studies of badminton [18, 19]. However, broadcast videos often suffer from challenges such as frequent perspective changes and occlusion issues [18, 82]. Although some image processing methods are focused on addressing these issues, recognizing and tracking the ball or multiple players on the court remains a challenging task.

Moreover, some researchers have collected and built their datasets based on specific task requirements, thereby providing an optimal perspective for analysis. For instance, in table tennis, TTNet [83] introduced the OpenTTGames dataset for game event detection and semantic segmentation tasks. Blank et al. [84] utilized inertial sensors attached to rackets to collect stroke data, while Kulkarni et al. [85] strategically positioned their cameras and vibration sensors to capture the most optimal view for detailed stroke analysis.

[1] `https://www.youtube.com/c/bwftv` (Accessed Dec. 29, 2023)

Meanwhile, different from ordinary cameras, drone technology offers a unique perspective for filming. Drones provide an optimal solution with superior adaptability and flexibility, especially in capturing dynamic sporting events. Drone cameras can provide more accurate 2D coordinates without occlusion if they can capture the court from a bird-view. However, unfortunately, there is currently no dataset available for racket sports. Thus, in this thesis, a dataset composed of drone videos that capture the entire pitch is created, effectively mitigating the challenges of severe occlusion during analysis.

2.4 Summary

The development and current state of play evaluation methods, visual analytics, and sports datasets across various sports, with a particular focus on racket sports have been overviewed in the literature review above. This thesis contributes by proposing innovative solutions to existing challenges. It addresses challenges in previous studies by proposing a method to estimate the value of each stroke and addressing the neglect of dense spatial distribution analysis in doubles matches by estimating the dense control area probability map. Additionally, this thesis introduces a novel dataset created from drone videos to overcome the challenges of frequent view changes and occlusions in traditional video footage.

Compared to previous studies that focus on overall play performance in a match, this thesis includes two studies aimed at providing detailed insights into a rally, which consists of a series of strokes. These studies aim to answer questions such as when and where improvements can be made, and how they can be achieved. Study I is implemented to estimate the action value of each play, identifying key moments of advantage and fluctuation, and addressing the need for deeper insight into the extent of changes in action value during a rally. Study II fills the gap in understanding the game's dynamics, particularly focusing on how player positioning and movement influence game outcomes, which is a critical aspect in doubles matches.

These two studies offer a comprehensive understanding of play performance in both singles and doubles, enriching the analysis of racket sports. Moreover, the practical implications of these insights are expected to benefit players, coaches, and analysts in real match situations.

3 Player Evaluation via Deep Reinforcement Learning

3.1 Introduction

Player evaluation has received increasing attention as it can assess and appraise the actions observed in a game to players, coaches, and other staff in order to facilitate decision-making (i.e., tactics) and improve technical skills, thus providing a competitive advantage to an individual or a team.

There are two main approaches for player evaluation. The first is to use various statistics to sum up "the total contributions of a player to his/her team" into a numerical value. The second approach is to assign values to the actions performed during a match. In the second approach, traditional methods (e.g., [86]) demonstrate significant limitations as they can only evaluate the actions that directly lead to a score (e.g., shooting), but are unable to evaluate those actions that indirectly lead to a score. Recently, the Markov model has been used to address this issue, which has the advantage of unified evaluation criteria (actions are evaluated on the same scale by anticipating expected outcomes). These approaches are based on the analysis of event stream data (including optical data) that describe the actions performed in a game. However, in racket sports, the task of action value evaluation of players is almost unexplored, because, in such a sport, technical whole-body movements have to be evaluated in addition to the tactics.

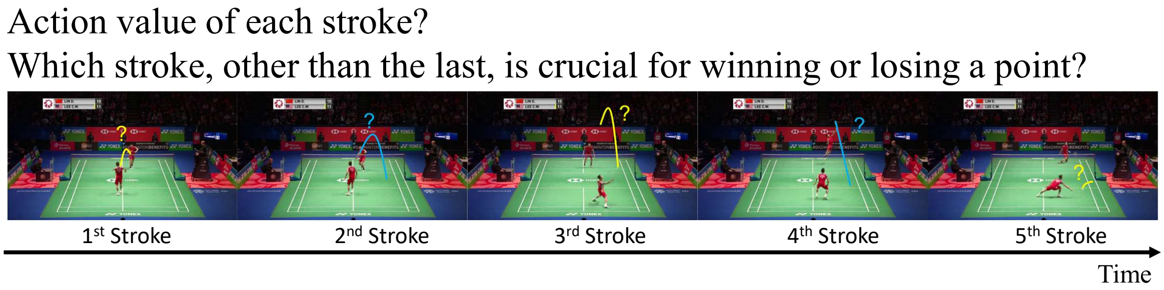


Figure 3.1: Example of a rally consisting of 5 strokes, focusing on the action value of each stroke. This analysis helps determine which stroke, aside from the last, is crucial in deciding the outcome of winning or losing a point.

In racket sports, most previous studies on player evaluation were limited to traditional methods. Shen et al. evaluated a single stroke type (lob) in badminton and defined a good/bad lob on whether the subsequent action of the opponent is a smash and leads to a score [87]. This study has constraints and cannot be used to evaluate the value of each stroke in a match. In racket sports like badminton, a rally is made up of many strokes as shown in Fig. 3.1. Each stroke contributing to the score and involving strategies similar to chess, Shogi, and Go [13]. In these sports and games, players need to think about every move in terms of long-term strategy. Understanding the value of each stroke is crucial to identify key moments that lead to winning or losing a point. While the final stroke is critical in determining the point's outcome, the preceding strokes also play an important role in shaping this outcome. Therefore, it is essential to estimate the value of each stroke for a straightforward understanding. The Q-function in the Markov model has the advantage of evaluating all actions (all stroke types) on the same scale by looking ahead to the outcomes. Thus, inspired by this idea, the Markov model approach (i.e., the second approach in the above paragraph) is adopted in this chapter using technical whole-body movement and tactical information for player evaluation in actual games.

With the recent advances in deep learning, Deep Reinforcement Learning (DRL) [88] has been applied in various fields [89, 90, 91, 92], and has shown significant promise in player evaluation in complex and dynamic environments. A previous method used the Markov model in team sports (e.g., ice hockey) to evaluate the tactical performance, but ignored the effect of technical performance on the value of the action. As racket sports have the characteristic of fewer participants (single- or double-player games), the result of a game depends largely on individual technical skills. Therefore, when compared to team sports, the specific sports-based technical performance of players should be analyzed through videos.

Most previous studies have used active Reinforcement Learning (RL) to calculate optimal strategies for complex continuous-flow games [93, 94]. Similar to the approach proposed by Liu et al. [35], here, a prediction (not control) problem is solved in the passive learning (fixed policy) setting. Note that RL is used as a behavioral analytics tool for real human agents, not to control artificial agents.

In this study, a player evaluation approach with play contexts in badminton is proposed, which leverages historical match data containing tactical and technical information to assign a rating to an action (e.g., smash) performed by a player in a match. For a given badminton game, deep learning methods are used to extract the features from videos that contain information related to both the tactical and technical performance of the players. Through experiments, the effect of technical and tactical contexts on player evaluation is examined by applying a DRL method to a badminton dataset (videos from the Badminton World Federation (BWF)). This research aims to provide coaches with some insights into the influence of the movements of players on the advantages and disadvantages in specific competition situations. Therefore, concrete movements and their influence are analyzed.

In summary, the primary contributions of this study are as follows:

- A player evaluation method in racket sports based on DRL is proposed that can analyze

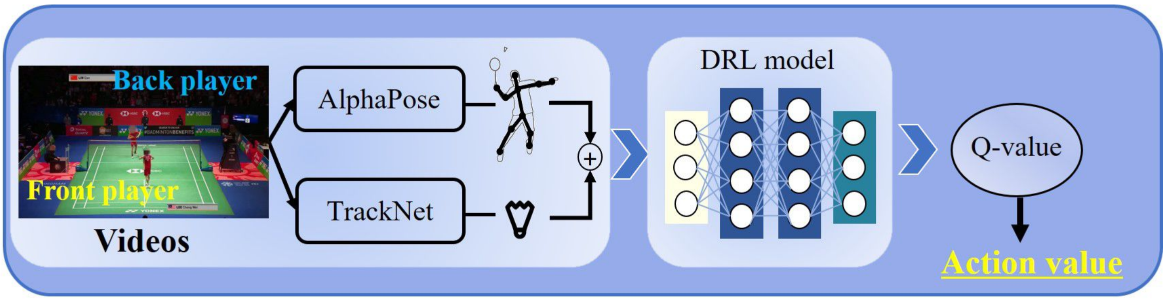


Figure 3.2: Overview of the proposed method designed to estimate action values. It utilizes AlphaPose [1] for precise player pose estimation and TrackNet [2] for accurate shuttlecock position tracking, followed by a DRL method to compute the Q-function, thereby quantifying action values of each player based on changes in Q-values due to player actions.

the motion of a player in more detail, instead of analyzing the outcome (i.e., scores).

- Methodologically, the proposed method leverages historical data including the tactical and technical performance information of players to learn the next-score probability as a Q-function, which is used to value the actions of the players.
- In the experiment, the proposed approach is verified by comparing various baselines. Its effectiveness is confirmed through use cases that analyze the performance of the elite badminton players in World-class events.

3.2 Player Evaluation Method

In this section, the proposed method for estimating the action value of a player is described. An overview of the proposed method is illustrated in Fig. 3.2.

Badminton is a competitive sport, where the winner of a match is determined based on the best performance out of three games, and each game is played for 21 points. A rally starts with a serve and ends when a point is won. Here, to describe a badminton match, a rally as

Table 3.1: Complete feature list used for stroke value evaluation.

Name	Type	Representation
X, Y Coordinate of shuttle	Continuous	(x_s, y_s)
X, Y Coordinate of player	Continuous	(x_p, y_p)
Pose	Continuous	17 keypoints
Action (Stroke type)	Discrete	One-hot representation

the analysis unit is considered. The course of the rally can be described as the transition from one state to another. A rally in badminton comprises a sequence of strokes with outcomes.

By using video data of badminton games, first each game in a match is segmented into several rallies. For each rally, the XY-coordinate values of 17 body parts, as defined in the Microsoft Common Objects in Context (COCO) dataset [95], are estimated to identify joint positions. These include the nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. This is achieved using AlphaPose [1], a popular, high-precision, multi-person body-pose estimation system. Additionally, the position of the shuttlecock is detected using TrackNet [2], an object tracking network noted for its decent tracking capability in fast-paced games like badminton.

As a prepossessing procedure, joint positions that were not properly estimated owing to overlap were annotated through the COCO annotator. Moreover, the midpoint of the two ankles is assumed to indicate the position of the player, and the pose represents the coordinate values relative to the position of the player. For the pose of a left-handed player, the corresponding relative coordinate values are reversed. The list of the complete features used in this method is shown in Table 3.1

The outputs are then combined as an input feature vector of a DRL model. The DRL method is applied to estimate the Q-values, and finally, the action value from the Q-values is estimated for evaluating the performance of a player. Players are labeled as front or back

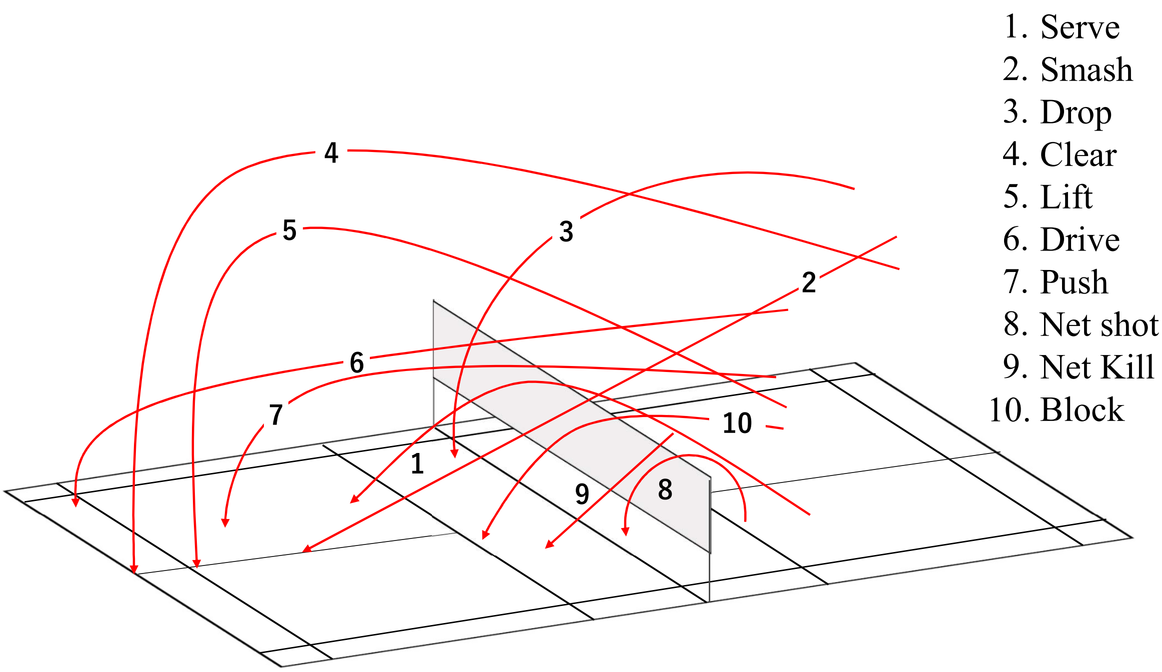


Figure 3.3: Shuttle's typical trajectories of different stroke types.

players based on their relative position to the camera. The player closer to the camera is referred to as the front player, and the one farther away as the back player. In the following, they are sometimes denoted as "front" and "back", respectively.

3.2.1 Formulation

A reinforcement learning framework is adopted in the proposed method, specifically based on the recent sports-related work [35]. The **reward** R specifies the player who wins a point at the end of a rally. Here, the strokes are generalized into nine types, namely Serve, Drop, Smash, Clear, Lift, Drive, Block, Net kill, and Net shot. Fig. 3.3 illustrates the shuttle's typical trajectories of different stroke types. **Action** a_t is one of the stroke types.

To describe the **state**, the feature vector at each hit time (the moment when the racket contacts the shuttlecock) is considered as state representation s_t at time step t.

The Q-function $Q(s,a)$, represents the conditional probability that the front or back player wins a point at the end of the rally such that

$$Q_{\text{front/back}}(s,a) = P(\text{point} = 1 | s_t = s, a_t = a). \tag{3.1}$$

The Q-function computes the expected rewards for an action taken in a given state. Different Q-functions can be used to study different outcomes of interest, such as goals and penalties [96]. A previous study [42] used “point” and “fault” as the expected outcomes for the model, whereas in this study, the probability of the next score is used, ranging from 0 to 1. The advantages of using the next score probability are as follows. (1) Compared to the outcome of a rally (“point” or “fault”), the next-score probability function is highly interpretable, as it models the probability of an event. (2) It can provide coaches with a more detailed overview of player performance during a rally.

3.2.2 Learning the Q-function

Fig. 3.4 shows the architecture of the proposed DRL model. Dynamic two-layer Long Short-Term Memory (LSTM) networks are used to learn the Q-function and estimate the Q-values. The networks take a sequence of states s_t and actions a_t at the moment the player hits the shuttlecock (hit frame) as the input. This model is used to simultaneously evaluate both the front and back players in a given rally.

The LSTM networks have an input layer composed of 256 nodes, a hidden layer with 256 nodes, and a dense output layer with 3 output nodes. The three output nodes, $Q_{\text{Back}}(s,a)$, $Q_{\text{Front}}(s,a)$, and $Q_{\text{Rally\_end}}(s,a)$, represent the probability that the front/back player wins the next point according to the present state and action and the probability that a rally ends according to the present state and action. A rally ends when the shuttle drops on the ground.

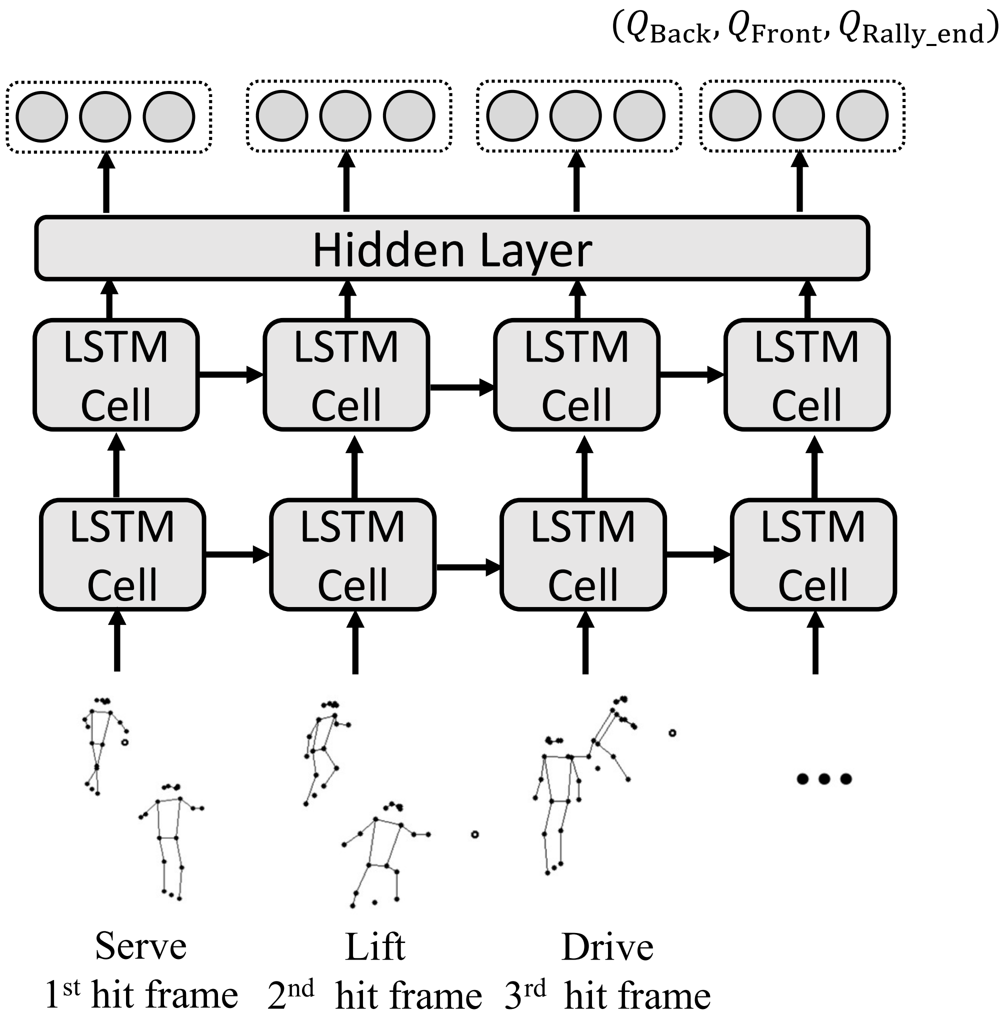


Figure 3.4: Architecture of the DRL model composed of two layers of LSTM networks. The input is a combination of feature vectors of XY-coordinates of the positions of the front and back players, their poses, the position of the shuttlecock, and the action at each hit time.

The Q-functions for each team share weights and the output values are normalized to probabilities. The LSTM networks require four types of input data for model training, namely the XY-coordinates of the positions of the front and back players, both their poses, the position of the shuttlecock, and the action.

In this study, the Q-function is learned via a neural network, which is called a function approximation approach. It is approximated by LSTM networks as:

$$Q(s,a) \approx q(s,a;w), \tag{3.2}$$

where LSTM networks are parameterized by w. The positions of the front and back players, both their poses, and the position of the shuttlecock are used to describe the state s, and the action a in $q(s,a;w)$.

The state–action–reward–state–action algorithm is applied, which is an on-policy reinforcement learning algorithm for estimating Q-values. The Q-value for a state–action is updated using the following equation:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha \left[R + \gamma Q\left(s_{t+1},a_{t+1}\right) - Q(s_t,a_t)\right], \tag{3.3}$$

where s_{t+1} and a_{t+1} denote the state and action at time step $t+1$. α is the learning rate, and γ is the discount factor. Instead of tabular learning, the neural networks are considered as the function approximators.

The network in this study is trained with loss $\mathscr{L}$ as:

$$\mathscr{L} = \frac{1}{n}\sum_{t=1}^{n}\left(Q(s_t,a_t) - R - \gamma Q\left(s_{t+1},a_{t+1}\right)\right)^2, \tag{3.4}$$

using the Adam optimizer [97] with a learning rate of 10^{-5} for 90 epochs. The hyperparameters were set to $\gamma = 0.3$.

3.2.3 Stroke Evaluation

For each action of the front/back player in a rally, the action value A can be computed as:

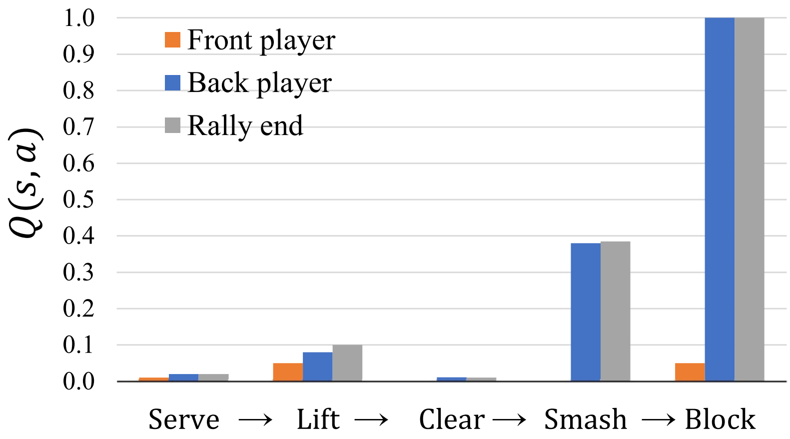
$$A_{t+1} = Q_{\text{player}}\left(s_{t+1},a_{t+1}\right) - Q_{\text{player}}(s_t,a_t). \tag{3.5}$$

Here, the probability that the front/back player will win the next point (Q-value) before service is set as 0. First, the action value of each stroke type and the corresponding number of each stroke type that a player takes in each game are calculated, and then the average action value of each stroke type performed by a player in a game is calculated.

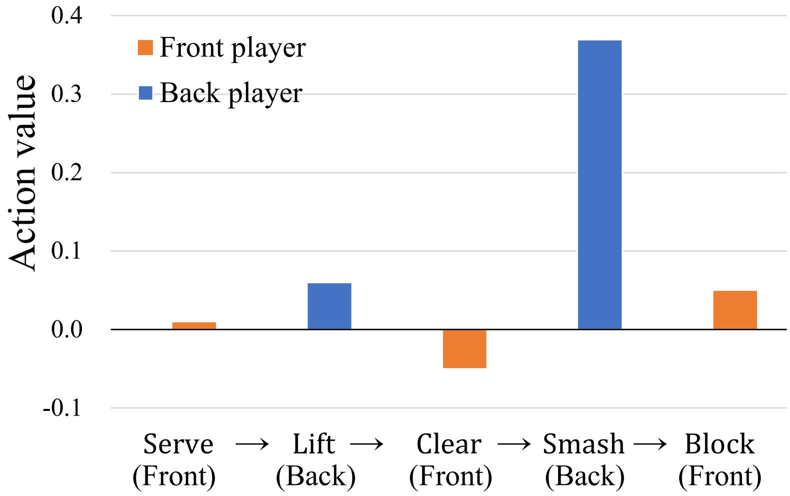
A typical example is presented to demonstrate the effectiveness of this approach in Fig. 3.5. The graph in Fig. 3.5 (a)shows the dynamic changes in the Q-values of a rally in a match between the back player and the front player. The back player won this rally in the end. The figure plots the values of the three output nodes.

Meanwhile, Fig. 3.5 (b)shows the action value of each stroke performed by the front and back players. A greater action value in a rally leads to major changes in the scoring chance and it also shows that the proposed method can capture small changes associated with every action under different technical and tactical contexts. According to the bar chart, the fourth stroke, which was a "smash" performed by the back player, shows the greatest action value, and the fifth stroke, which was a "block" performed by the front player, has a minor action value. The result of the action value is consistent with observations from the video that the "smash" performed by the back player was significantly powerful, and the front player failed to intercept the shuttle through the "block" stroke.

Moreover, this figure provides more insights into the rally. The observations gained from the video may make the coach erroneously attribute the loss of a point to the poor defensive "block" performed by the front player. However, the result shows the action value of the "block" is positive, albeit a small one. Conversely, the third stroke, which was a "clear" stroke performed by the front player, has a negative action value; after this stroke, the back player seized the chance to attack using the offensive "smash" action. Observations of the video identified that the front player returned the shuttle to a position closer to the back player with the "clear" action; consequently, the back player had sufficient time to jump and



(a) Front/back player next score and rally end probablity



(b) Each stroke value of front/back player

Figure 3.5: Typical example of the front/back player scoring probability and action value analysis in a rally.

attack with a "smash." The figure reflects that the previous actions in a rally have an indirect impact on the final result. Therefore, it can be concluded that the result of this approach can provide useful clues for further analysis over the course of multiple strokes in a rally.

3.3 Experiments

This section presents a series of experiments designed to evaluate the proposed method. For validation, videos from the Badminton World Federation (BWF) TV channel[1] are utilized, focusing on assessing the impact of various input features in the LSTM networks, player characteristics, and the relationship between stroke value, match scores, and player rankings. Moreover, the proposed method is compared with a baseline method to highlight its advancements in sports analytics. The goal of these experiments is to thoroughly evaluate the proposed method in the context of high-level badminton matches.

3.3.1 Dataset

The dataset is composed of videos from the BWF TV channel[1]. To ignore the cross-view problem, only videos captured from a single view were selected as shown in Fig. 3.2. Here, the "cross-view problem" in the context of broadcast videos, particularly in sports, refers to the challenge posed by frequent changes in camera angles and perspectives. This issue is especially relevant in dynamic sports broadcasting, where multiple cameras are used to capture the action from various angles to enhance the viewing experience. The constructed dataset provides information regarding game contexts and player actions for the 2018–2020 BWF season, which contains 21 matches, covering 22 players and 320 rallies. The total length of the video is 1,432 minutes, with the average length of each rally being approximately 10 seconds after manual clipping. 80% of the data are used for training and the remaining 20% are reserved for testing. Points scored in each rally are employed as the ground-truth data.

[1] `https://www.youtube.com/c/bwftv` (Accessed Dec. 29, 2023)

Table 3.2: Comparison of each input feature by eliminating it from the total input. The value of loss function $\mathscr{L}$ according to Eq. (3.4) is shown here, which was evaluated on the badminton dataset.

Error	Full model	−Shuttle position	−Player position	−Player pose
$\mathscr{L}$	**0.0261**	0.0289	0.0436	0.0506

Furthermore, this dataset has been published[2] to benefit the broader research community and is expected to be useful for various research purposes.

3.3.2 Verification of the Proposed Method

The design choices of the proposed method are evaluated in terms of input components by comparing the performance of the LSTM networks with different inputs.

As presented in Table 3.2, a model trained with all the input components (Full model) achieved the best performance, indicating that all the input components contribute to estimating an accurate Q-function. While each input component contributes to the model's accuracy, the player pose appears to be especially crucial, as indicated by the significant increase in the loss function value when it is omitted. The reason could be that the player pose offers detailed insights into a player's current actions and potential next moves, which are vital for a precise estimation of the Q-function in the context of badminton.

Next, the effect of the stroke type is examined as shown in Table 3.3. As the action is a primary element in reinforcement learning, a fully supervised LSTM model (non-reinforcement learning) is used to examine the effect of action (stroke types) by eliminating it from the total input. The results show that the stroke type feature can help in improving the accuracy of the

[2]The dataset and related codes are available at `https://github.com/Ning-D/Evaluate_Badminton_Stroke`

Table 3.3: Comparison of LSTM model performance with and without stroke type feature.

Error	With stroke type	Without stroke type
$\mathscr{L}$	0.0995	0.1274

fully supervised LSTM model when compared to a model that does not consider this feature.

3.3.3 Characteristics of Players

Analyzing the average actions of different stroke types is instrumental in helping coaches and players identify areas of strength and weakness. This understanding facilitates targeted training aimed at enhancing specific skills, which contributes to overall performance improvement. To illustrate the practical application of this study, four matches from the 2018 BWF Tour are examined. These matches include one final, two semi-finals, and one quarter-final men's singles matches, which are the matches in the All England Open Tournament between Lin Dan and Shi Yuqi, Huang Yuxiang and Lin Dan, Son Wan-ho and Shi Yuqi, and Lin Dan and Lee Chong Wei. The performance of the players is analyzed from the perspective of the average action value of each stroke type performed by a player in a match (Fig. 3.6).

For example, when Lin was playing against Huang and Lee (Figs. 3.6 (b)and (d)), the strokes performed by Lin were superior to those by his opponents. Lin had fewer stroke types whose action values were substantially below zero, especially when he was playing against Lee. His offensive strokes such as "smash" and "net shot" were better than those of Lee. However, when playing against Shi (Fig. 3.6 (a)), Lin did not display significant advantages, which would explain why he lost the final match. During the match between Shi and Son (Fig. 3.6 (c)), Shi showed superior performance in offensive strokes such as "smash"

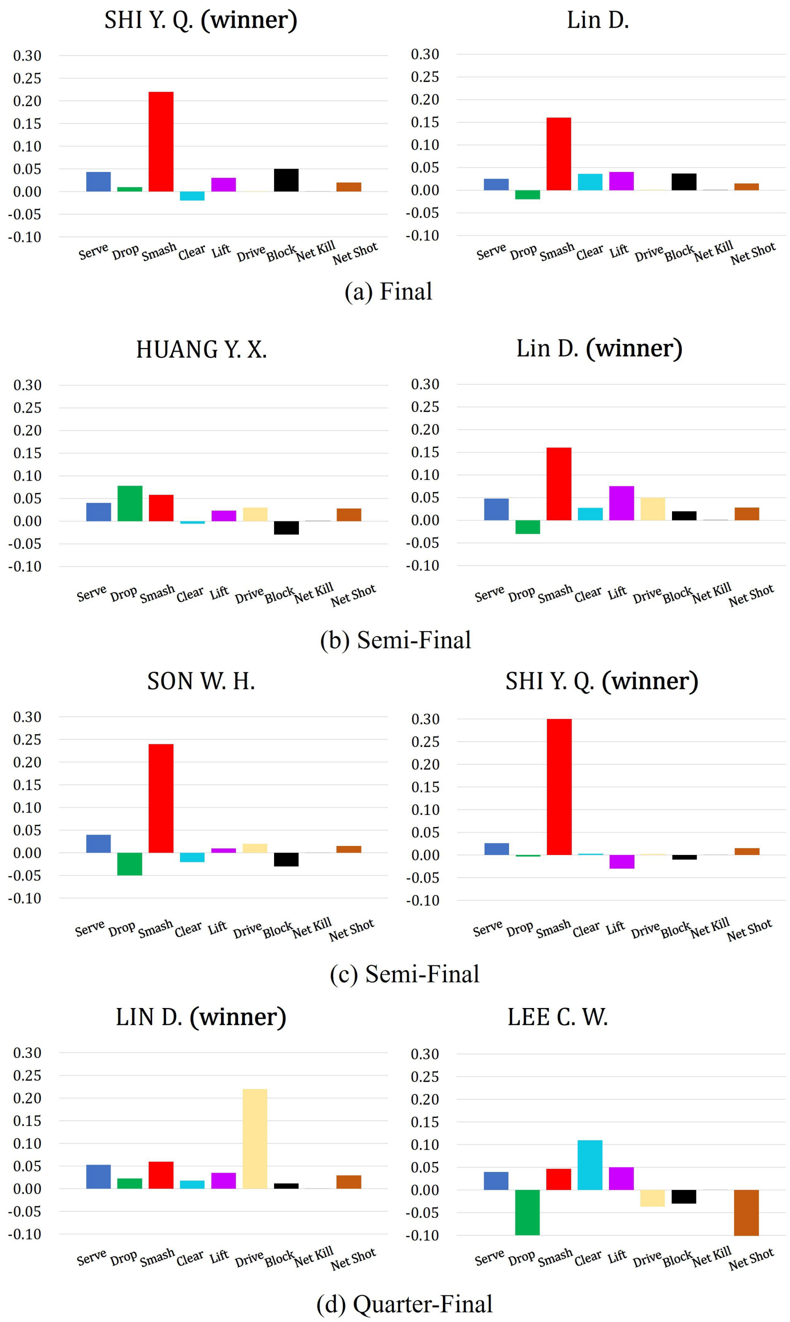


Figure 3.6: Average action value of each stroke type in four badminton finals.

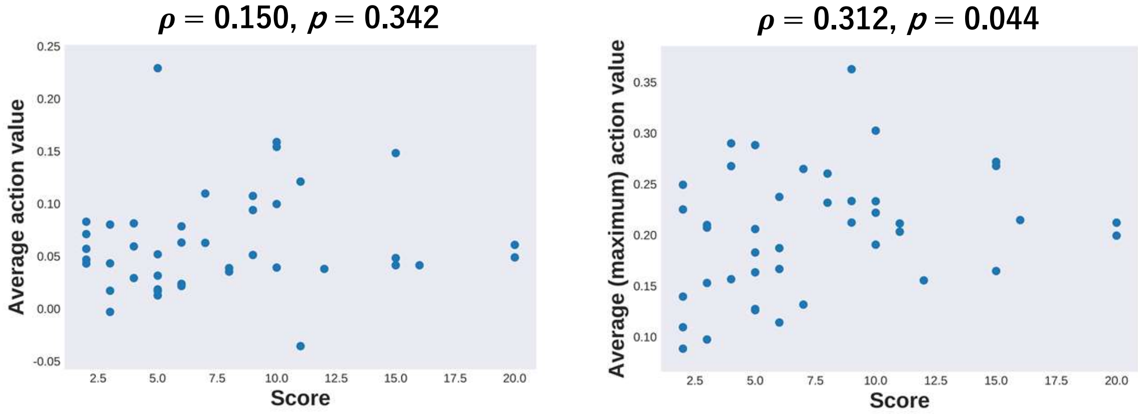
Table 3.4: Player action values in the 2018–2020 BWF Tour

Game	Players	Average action value	Average (maximum) action value	Point	Rank
1	LIN D. vs. LEE C. W.	0.0392 vs. 0.0387	0.2218 vs. 0.2317	10 vs. 8	BWF2018 11 vs. 49
2	NG K. L. vs. NISHIMOTO	0.0169 vs. 0.1074	0.1634 vs. 0.3628	5 vs. 9	BWF2018 16 vs. 9
3	LIN D. vs. SHI Y. Q.	0.0519 vs. 0.0630	0.1829 vs.0.1871	5 vs. 6	BWF2018 11 vs. 4
4	CHOU T. C. vs.TSUNEYAMA	0.0814 vs. -0.0030	0.2900 vs. 0.2099	4 vs. 3	BWF2020 6 vs. 25
5	LEE Z. J. vs. CHRISTIE	0.0414 vs. 0.0432	0.2147 vs. 0.0974	16 vs. 3	BWF2020 10 vs. 19
6	CHOU T. C. vs. AXELSEN	0.0802 vs. 0.0171	0.1530 vs. 0.2073	3 vs. 3	BWF2018 1 vs. 20
7	HUANG Y. X. vs. LIN D.	0.0512 vs. 0.0482	0.2122 vs. 0.2678	9 vs. 15	BWF2018 17 vs. 11
8	LEE D. K. vs. LEE C. W.	0.0785 vs. 0.0354	0.2374 vs. 0.2603	6 vs. 8	BWF2018 30 vs. 49
9	VITTINGHUS vs. LIN D.	0.0315 vs. 0.0997	0.2059 vs. 0.1905	5 vs. 10	BWF2018 23 vs. 11
10	LEVERDEZ vs. KIDAMBI	0.0379 vs. 0.0940	0.1555 vs. 0.2333	12 vs. 9	BWF2018 24 vs. 14
11	LEE Z.J. vs. AXELSEN	0.0594 vs. 0.2291	0.2676 vs. 0.2883	4 vs. 5	BWF2020 10 vs. 1
12	CHOU T.C. vs. ANTONSEN	0.0468 vs. 0.0571	0.2251 vs. 0.1094	2 vs. 2	BWF2020 6 vs. 2
13	LEE Z.J. vs. CHEN L.	0.0829 vs. 0.0711	0.2493 vs. 0.0884	2 vs. 2	BWF2020 10 vs. 28
14	LU G.Z. vs. AXELSEN	0.1588 vs. 0.1097	0.3024 vs. 0.2649	10 vs. 7	BWF2019 19 vs. 5
15	SON W.H. vs. SHI Y.Q.	0.0431 vs. 0.0237	0.1395 vs. 0.1667	2 vs. 6	BWF2018 46 vs. 23
16	ANTONSEN vs. SHI Y.Q.	0.0126 vs. 0.1541	0.1274 vs. 0.2332	5 vs. 10	BWF2019 27 vs. 4
17	SHI Y.Q. vs. AXELSEN	0.0184 vs. -0.0356	0.1262 vs. 0.2114	5 vs. 11	BWF2020 7 vs. 1
18	CHEN L. vs. GINTING	0.0627 vs. 0.1483	0.1317 vs. 0.2719	7 vs. 15	BWF2019 4 vs. 6
19	MOMOTA vs. GINTING	0.0414 vs. 0.1211	0.1646 vs. 0.2032	15 vs. 11	BWF2019 1 vs. 6
20	GINTING vs. ANTONSEN	0.0216 vs. 0.0292	0.1141 vs. 0.1566	6 vs. 4	BWF2019 6 vs. 7
21	MOMOTA vs. ANTONSEN	0.0608 vs. 0.0489	0.2121 vs. 0.1994	20 vs. 20	BWF2019 1 vs. 7

and “net shot.” The summary of all player action values from the 2018–2020 BWF Tour is listed in Table 3.4, and the relationships between action value and point/rank are examined in 3.3.4 and 3.3.5.

3.3.4 Relationship with the Score of the Match

Owing to the cross-view scenes in badminton videos, only a portion of the rallies could be used in each match. Therefore, to further demonstrate the effectiveness of the proposed



(a) Correlation between score and average action value

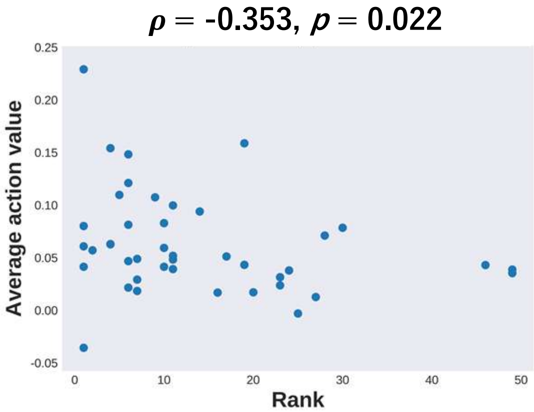
(b) Correlation between score and average (maximum) action value

Figure 3.7: Correlation between the score and average action value/average (maximum) action value.

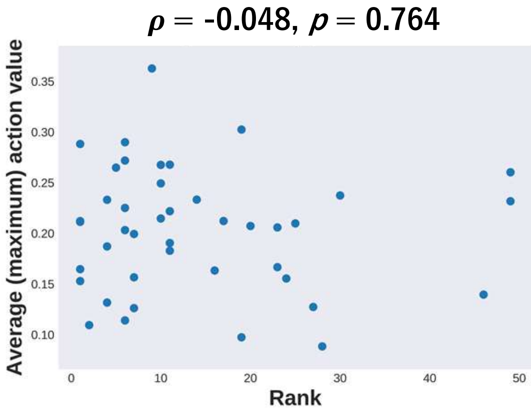
method, the relationship between average (maximum) action values and the score is examined, as shown in Figs. 3.7 (a) and (b). Here, the average (maximum) action values refer to the average value of the maximum value of the actions performed by a player in each rally over several rallies, and the score indicates the number of points scored by the front/back player in a match.

Spearman's rank correlation coefficient ρ was applied to quantify the correlation because the relationship with the rank of the players is also examined, as described in 3.3.5. Fig. 3.7 (a) shows that $\rho = 0.150$ ($p > 0.05$), indicating that there is no correlation between the score and average action value. Meanwhile, Fig. 3.7 (b) shows that $\rho = 0.312$ ($p < 0.05$), which reflects a weak positive monotonic correlation between the score and average (maximum) action value.

The results suggest that the average (maximum) action value could be associated with the score, and if the maximum action value of a player in a rally is greater than that of his/her opponent, the action can lead to obtaining the score of the rally.



(a) Correlation between rank and average action value



(b) Correlation between rank and average (maximum) action value

Figure 3.8: Correlation between the rank of the player and average action value/average (maximum) action value.

3.3.5 Relationship with the Rank of a Player

The relationship between the rank of the player and average (maximum) action values is also examined, as shown in Figs. 3.8 (a)and (b). Here, the average action values refer to the average value of all actions performed by a player over several rallies. For the player's ranking, the official BWF ranking data were referred to. Fig. 3.8 (a)shows that $\rho = -0.353$ ($p < 0.05$), which reflects a weak negative monotonic correlation between the rank of the player and average action value. Fig. 3.8 (b)shows that $\rho = -0.048$ ($p > 0.05$), which indicates that there is no correlation between the rank of the player and average (maximum) action value.

The results suggest that the average action value can be associated with the rank of the player, and higher-ranked players tend to perform actions with higher action values. These results also imply that higher-ranked players, who are typically more skilled, do not necessarily perform actions with consistently higher average (maximum) values. This could be attributed to various factors, including the diverse range of strategies employed by different players, their individual styles, and the situational context of each game.

3.3.6 Correlation Comparison with a Baseline Method

No previous studies have utilized the DRL model to evaluate each stroke in racket sports like badminton and it is important to note that the dynamics of badminton and ice hockey are inherently different, which makes a direct comparison between the proposed method and the method applied in ice hockey (Liu et al. [35]) challenging.

Here, Spearman's rank correlation coefficient ρ is computed to assess the relationship between the action value of the stroke estimated by the baseline model and standard success measures (score/rank). This statistical approach is crucial in understanding how well the model's estimations align with actual game outcomes. The results are summarized in Table 3.5, which shows that there is no correlation between player evaluation metrics and standard success for the baseline model ($p > 0.05$). However, the proposed method (Full model), which includes pose information, shows a weak monotonic correlation with the two success measures. This suggests that incorporating pose information into the model provides some important insights, potentially enhancing the model's ability to evaluate player performance more accurately.

3.4 Summary

This study proposed a new evaluation method for racket sports based on DRL, which can analyze the motion of a player in more detail, rather than only considering the results (i.e., scores). The proposed method used historical data including information related to the tacti-

Table 3.5: Spearman's correlation coefficient values computed between action value and standard success measures (score/rank) using both the proposed method and the baseline method. Values in bold indicate statistically significant correlations ($p < 0.05$).

Model	Full model		− Player pose (baseline)	
	Score	Rank	Score	Rank
Average action value	–0.150	**–0.353**	−0.154	−0.122
Average (maximum) action value	**0.312**	–0.048	−0.061	0.163

cal and technical performance of players to learn the next-score probability as a Q-function, which is used to value the actions of the players. Two layers of LSTM networks were leveraged for the learning of the Q-function with the poses of the players and the position of the shuttlecock as the input, which are identified by the AlphaPose [1] and TrackNet [2] algorithms, respectively. The proposed method was verified by comparing various baselines and its effectiveness was demonstrated through use cases that analyze the performance of the top badminton players in World-class events. In addition, valuable insights were discovered regarding the correlation between the action value and the score/rank. By collecting videos from the BWF channels, a badminton dataset with ground-truth annotations of both stroke types and game scores was constructed and made public. However, it should be noticed that the proposed method used a fixed policy, which poses a challenge for individual player modeling and may impact the accuracy of action evaluation and correlation results. Furthermore, the use of 2D feature extraction currently presents an additional challenge in accurately modeling the match.

In future work, improvements will be made to the framework of the proposed method to address existing challenges. This will include modeling individual players using player-specific policies. Additionally, efforts will focus on utilizing 3D features to further improve match modeling.

4 Control Area Estimation with Pose Information from Drone Videos

4.1 Introduction

Visual tracking has become an increasingly popular research area due to its diverse applications in domains such as human-computer interaction, robotics, autonomous driving, and medical imaging [98, 99, 100, 101]. Object tracking and pose estimation are two essential components of visual tracking, allowing for accurate and efficient tracking of objects in videos. However, despite significant progress in object tracking [102] and pose estimation [103], applying these technologies to the sports field remains a challenging problem due to partial occlusion, changes in viewpoint, and complex movements of athletes in a dynamic environment.

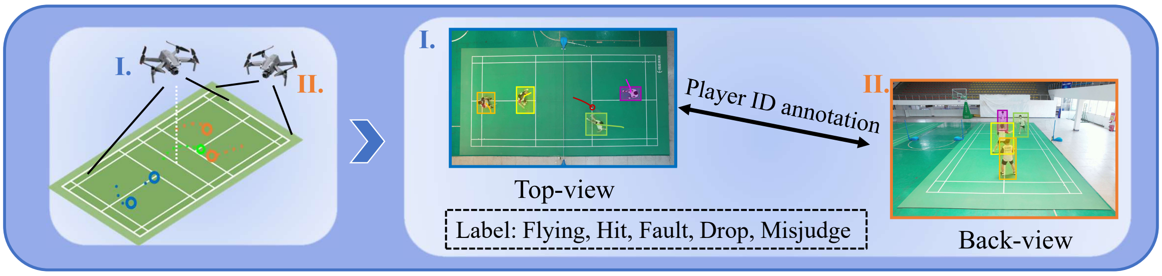
As technology has advanced in the past few years, data collection has become more in-depth and can be conducted with relative ease. Significant effort has been focused on building larger broadcast sports video datasets [69, 104]. Specifically, publicly available badminton datasets primarily consist of broadcast videos sourced from the Badminton World Federation (BWF) channel on YouTube[1]. However, large-scale broadcast videos are primarily accessible to wealthy professional sports teams, making it challenging for teams with fewer resources to benefit from them. Additionally, broadcast videos do not capture the entire

[1] https://www.youtube.com/c/bwftv (Accessed Dec. 29, 2023)

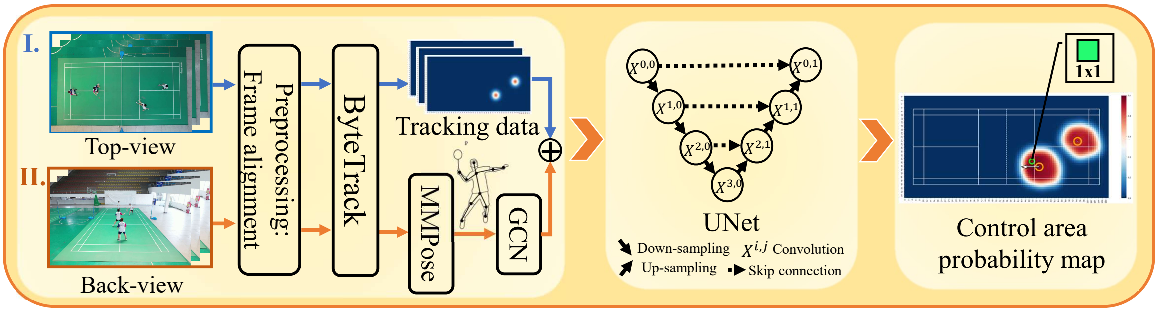
pitch, offering only partial information about the game. To be specific, these videos frequently feature perspective changes, and they also suffer from significant occlusions, especially in team sports. In soccer, drone cameras are used to capture the entire pitch in a single frame [76], providing greater adaptability and flexibility. Unfortunately, there is currently no such dataset for any racket sport.

Visual analytics, including heatmaps, have become prevalent in various sports, especially in team sports like soccer [33, 105]. Advanced methods are being developed to estimate probabilities of actions, such as passes, and other performance metrics based on spatiotemporal data. For example, SoccerMap [53] estimates probability surfaces of potential passes from high-frequency spatiotemporal data in team sports like soccer. However, there has been little exploration of fine-grained spatial analysis in racket sports. In racket sports, most previous studies have focused on analyzing and assessing singles players in broadcast videos and discrete representations such as stroke analysis (e.g., [19, 82, 106]). This is because singles matches are more straightforward due to the involvement of only two players. This simplicity facilitates the application of methods like stroke analysis, offering more straightforward and clear insights. In contrast, doubles matches involve more complex interactions and spatial dynamics due to the presence of four players on the court. Consequently, the increased complexity and analytical challenges associated with doubles matches have led to a predominant focus on singles matches in previous studies, often resulting in a neglect of the meaningful spatial distributions that are critical in doubles scenarios.

Therefore, the objective of this study is to address these limitations by focusing on quantifying spatial value occupation and providing a quantitative position evaluation metric in a doubles badminton match. To achieve this, a self-collected drone dataset is created and leveraged to estimate the control area probability map, as depicted in Figure 4.1. The proposed method in this study employs a two-stream network architecture (Figure 4.1 (b)) that



(a) Proposed open-source dataset



(b) Control area estimation

Figure 4.1: Overview of the two key contributions in this study. Firstly, an open-source dataset that includes labeled data on shuttlecock status, player positions and poses, and player IDs for both top-view and back-view videos is introduced. Secondly, a novel visual analysis method is proposed for estimating control area in badminton doubles.

combines positional information from a top view with pose information from a back view. By utilizing a 3-layer U-Net [107] model, the proposed method accurately predicts the probability map, allowing for an effective evaluation of the team's control area.

The contributions of this study are as follows:

- A men's doubles badminton drone dataset is constructed and made public, which includes annotations of bounding boxes, shuttlecock locations (Hit/Drop), and poses.

- An efficient framework of deep neural networks is proposed that enables the calculation of full probability surfaces, which utilizes the embedding of a Gaussian mixture map of players' positions and graph convolution of their poses. The effectiveness of this fine-grained analysis is validated in badminton game situations.

- A practical application is proposed for assessing optimal positioning in badminton that can increase the probability of a successful shot.

4.2 Dataset

By providing both top and back-view videos, this study offers a comprehensive perspective essential for an accurate and detailed understanding of player movements and game strategies. This dual viewpoint is particularly advantageous for precisely capturing player and shuttlecock positions, as well as player poses, thereby overcoming the limitations of existing badminton datasets. In terms of its characteristics, this dataset is distinguished by its comprehensive coverage and high-quality video data. Utilizing two drones, it captures every aspect of the badminton match.

4.2.1 Video Collection

Video data from 2 vs. 2 men's doubles badminton matches played among members of a college badminton club were recorded. Prior to data collection, approval was obtained from Anhui Normal University's ethics committee, and the study was conducted in compliance with the principles of the Declaration of Helsinki [108]. All participants provided signed informed consent. To capture the entire badminton court, two DJI Air 2S drones (Da-Jiang Innovations Science and Technology Co., Ltd., China)[3] were used to provide top and back views. The first drone was positioned directly above the center of the court at a height of

[3] `https://www.dji.com/jp/air-2s` (Accessed Dec. 29, 2023)

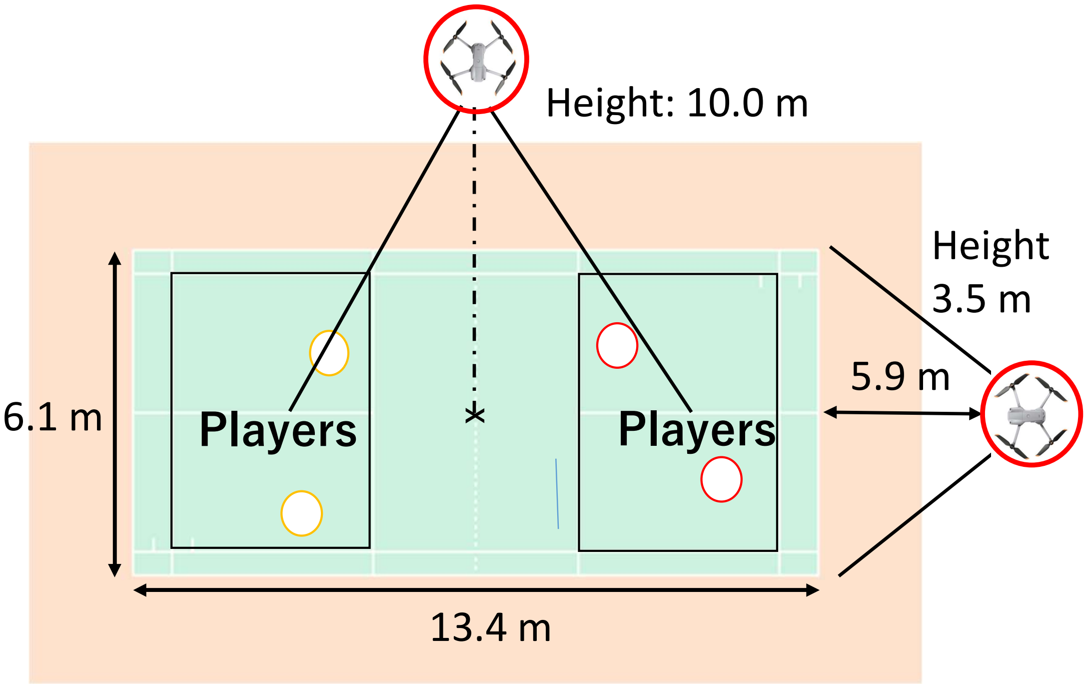


Figure 4.2: Configuration of the drones in relation to the court.

10.0 m, filming from above, while the second drone, placed 5.9 m behind the court at a height of 3.5 m, captured the back view, as illustrated in Fig. 4.2. The video resolution is 4K (3,840 $\times$ 2,160 pixels), with a frame rate of 30 fps. The raw video data includes 39 matches involving 14 pairs, 11 players, and a total of 1,347 rallies. The total video length is 702 minutes, and the average length of each rally is approximately 6 seconds

It is important to highlight that publicly available badminton datasets primarily consist of broadcast videos sourced from the Badminton World Federation (BWF) channel on YouTube[1]. However, these videos typically only offer back-view footage and lack the crucial top-view videos required for the proposed method.

[1] `https://www.youtube.com/c/bwftv` (Accessed Dec. 29, 2023)

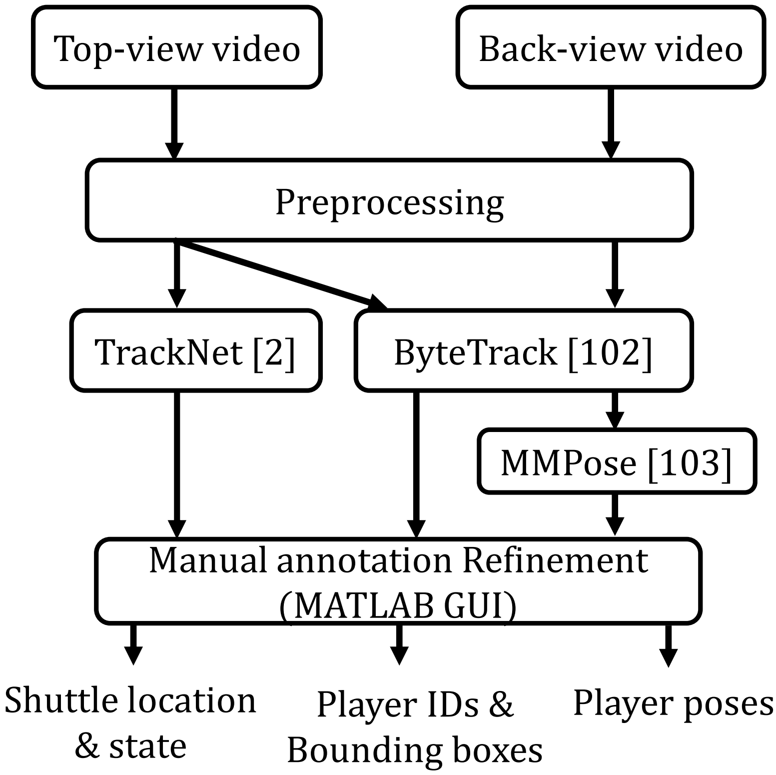


Figure 4.3: Dataset annotation workflow for top-view and back-view drone videos.

4.2.2 Data Annotation and Structure

This section outlines the efficient annotation of bounding boxes and shuttlecock locations in drone videos (Fig. 4.1 (a)). Manual annotation is a time-consuming process that can take several hours to annotate a single rally. Therefore, well-established computer vision techniques are introduced to shorten the process as shown in Fig. 4.3.

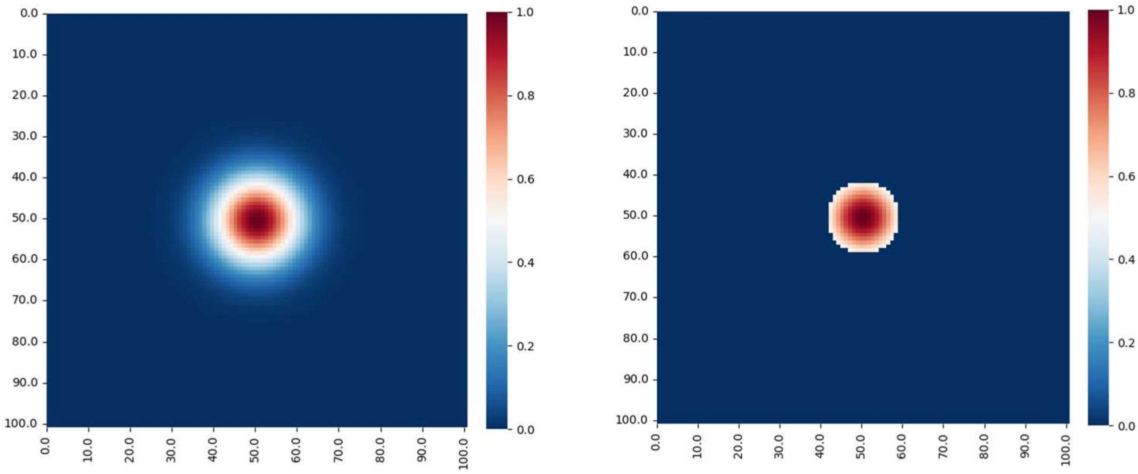
The dataset is structured such that each rally is accompanied by two annotation files in a simple Comma-Separated Value (CSV) format. For shuttlecock detection, each line of the CSV file contains five values: frame number, visibility, X-coordinate, Y-coordinate, and status. The status of the shuttlecock in each frame was annotated with one of five types: Flying,

Table 4.1: Overview of badminton doubles drone dataset.

	Top & Back view camera
Device	DJI Air 2S
Resolution	4 K
Frame rate	30 FPS
Player location	✓
Bounding box of player	✓
Shuttlecock location	✓
Shuttlecock status	✓
Player ID	✓

Hit, Fault, Drop, and Misjudge. For players, corresponding bounding box coordinates are also provided in CSV format.

To track players and the shuttlecock in the raw video data, the largest court lines in each frame are first converted to a uniform size of $3,500 \times 1,600$ pixels for calibration. This is achieved by applying homography transformation [109], which helps to eliminate the offset problem caused by the drone's perspective. Each game is segmented into several rallies, and for each rally, the XY-coordinate values were estimated for the locations of the four players using tracking, specifically ByteTrack [102], a popular high-precision multi-object detection and tracking system. Meanwhile, the shuttlecock location was detected using TrackNet [2], an object tracking network that has been proven to exhibit decent tracking capability in games that involve small, high-speed balls such as shuttlecocks. Additionally, player poses are also estimated using MMPose [103], based on the bounding boxes captured by ByteTrack [102]. Any outlier is adjusted using a simple labeling tool with a MATLAB Graphical User Interface (GUI), modifying the code originally created and utilized in the study of TrackeNet [2] to fit the requirements of this study. Moreover, Direct Linear Transforma-



(a) *Typical heatmap range is* $(0 \leq p \leq 1)$. (b) *Heatmap in this study* $p = 0$ *or* $p \geq 0.5$

Figure 4.4: Heatmap comparison in probability analysis. In a typical heatmap, the probability ranges from 0 to 1. The heatmap in this study only shows areas where $p = 0$ *and* $p \geq 0.5$*, using deep red to highlight higher probabilities and deep blue for probabilities at 0.*

tion (DLT) is applied to re-identify players' IDs and any discrepancy is manually corrected. Combining detection results with manual adjustments significantly expedits the annotation process. An overview of the badminton doubles drone dataset is presented in Table 4.1. The dataset is made publicly available[4].

4.3 Estimation Method

A framework is proposed for estimating the probability map of the control area, where the probability values, ranging from 0 to 1, are used to express the likelihood of a team successfully receiving the shuttlecock. It should be noted that the proposed method, unlike a typical heatmap displaying probability values ranging from 0 to 1 as shown in Figure 4.4 (a), exclu-

[4] `https://github.com/Ning-D/Drone_BD_ControlArea`

sively shows values of either $p = 0$ or $p \geq 0.5$, as depicted in Figure 4.4 (b). This is because, in this study, only the area where $p \geq 0.5$ is considered the control area of a team. Areas with a deep red color indicate a higher probability of receiving the shuttlecock, whereas white areas represent regions where the probability value is approximately 0.5. Regions with $p \leq 0.5$ are represented as $p = 0$, depicted in deep blue.

A neural network that learns the relationship between tracking/event data and a probability map in a data-driven manner is constructed. Details of the neural network architecture and its training are described below.

4.3.1 Model Architecture

The model architecture consists of a two-stream network. The first stream is based on the top view and captures information about the location and velocity of players, while the second stream is based on the back view and captures information about the pose of players. The outputs of the two streams are then combined to serve as the input of a 3-layer U-Net model, which predicts the control probability map.

A 3-layer U-Net [107] network is constructed to estimate the full probability map of the control area, considering that the input image size is 112 $\times$ 56 pixels. Here, the term '3-layer' refers to the fact that both the downsampling layer (Max pooling) and the upsampling layer (Up-convolution) in the network consist of three layers each, as shown in Figure 4.5. U-Net is a convolutional network architecture specifically designed for fast and accurate image segmentation. This network is expected to be effectively utilized here to segment and identify areas that are controllable or not. The network takes as input: (1) a Gaussian mixture probability map centered on the location of two players (same team) who receive the shuttlecock, (2) the X-velocity and Y-velocity of two players (same team) who receive the shuttlecock, and (3) the poses of two players (same team) who receive the shuttlecock.

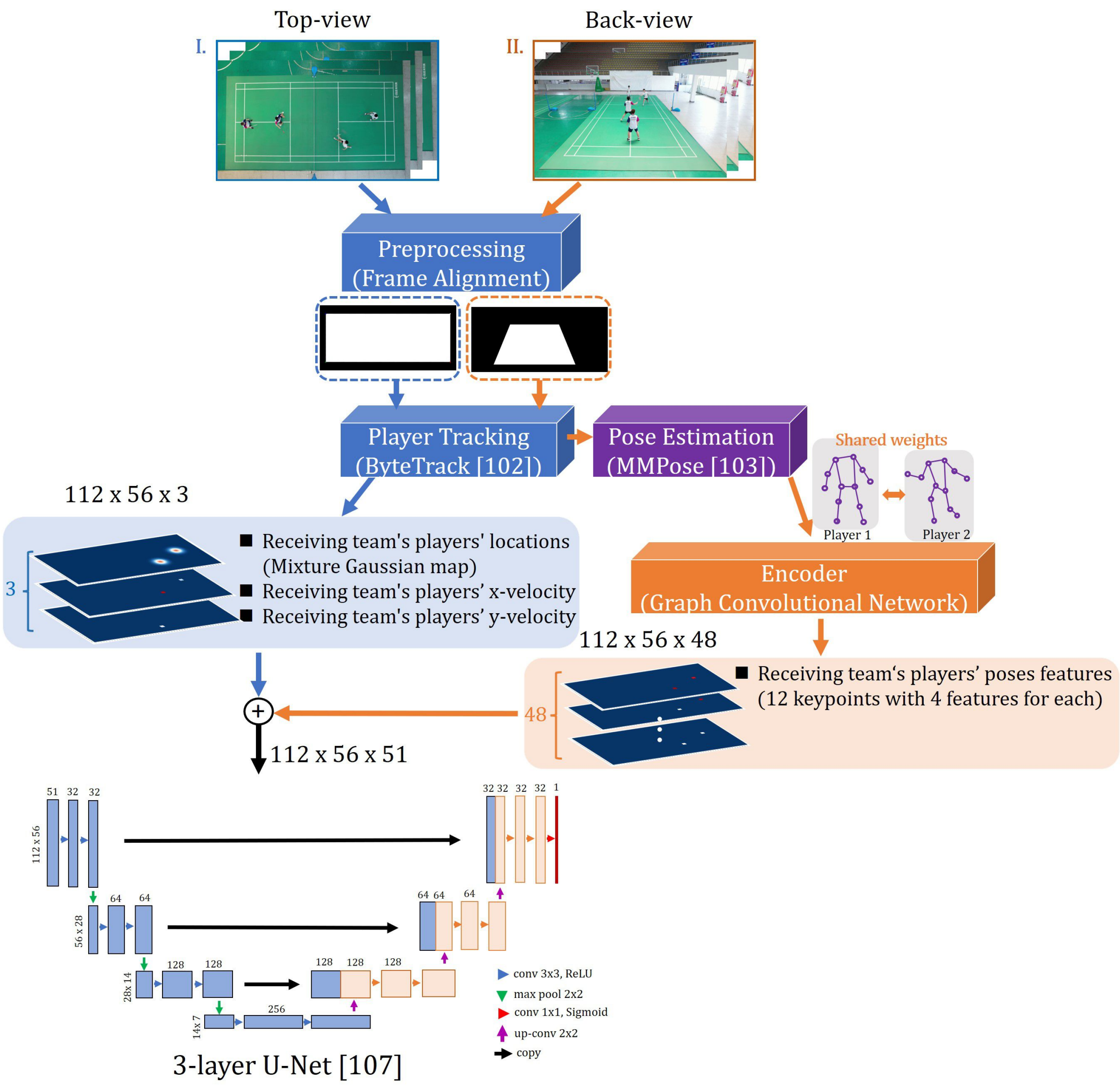


Figure 4.5: Detailed architecture diagram of the proposed model, showing the 2-stream network and the 3-layer U-Net model used to generate the control area probability map.

The location where the player hits the shuttlecock or where the shuttlecock lands (drop) is used as the target location to obtain the control area probability map.

4.3.2 Learning

The proposed loss function L in this study consists of a Focal Loss (FL) L_f [110] and a constraint on spatial continuity L_c, denoted as:

$$L = L_f + \mu L_c. \tag{4.1}$$

Here, μ represents the weight for balancing the two constraints, which is set as $\mu = 0.03$. As the objective function of the model, FL is used to address the class imbalance problem in the dataset, defined as:

$$\mathrm{FL}(p_\mathrm{t}) = -\alpha (1 - p_\mathrm{t})^\gamma \log(p_\mathrm{t}), \tag{4.2}$$

$$p_\mathrm{t} = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \tag{4.3}$$

where p_t is the estimated probability of the model. Here, α is set as 0.8 and γ is set as 3. The Focal loss L_f used here can then be written as:

$$L_f = \mathrm{FL}\left(y_{\mathrm{loc}_k}, f(x_k;\theta)_{\mathrm{loc}_k}\right), \tag{4.4}$$

where x_k is the game state at time k, loc_k represents the location where the player hit the shuttlecock or where the shuttlecock dropped at time k, and y_{loc_k} represents the ground-truth control probability at time k.

An additional constraint L_c is introduced here following Kim et al.'s practice [111]. This loss term is used to encourage spatial smoothness and continuity in the estimated control probabilities. The spatial continuity loss is defined as:

$$L_c = \sum_{i=1}^{W-1} \sum_{j=1}^{H-1} \left\| v_{i+1,j} - v_{i,j} \right\|_1 + \left\| v_{i,j+1} - v_{i,j} \right\|_1, \tag{4.5}$$

where W [pixels] and H [pixels] represent the width and height of the image, respectively, and $v_{i,j}$ denotes the control probability of the pixel at coordinates (i, j). The loss is calculated by summing the L_1-norm differences between adjacent pixels in both horizontal and vertical directions. Minimizing this loss term encourages the network to generate spatially continuous control probabilities with smooth transitions and discourages the presence of complex patterns or abrupt changes in the probability map. At the same time, it can also mitigate the impact of data amount on the results to some extent.

4.3.3 Optimal Positioning

The model in this study can estimate the control area and evaluate the players in a doubles badminton match. However, determining the optimal positioning and movement strategy for players, which is a critical aspect of successful performance, is currently unknown. In this study, a method to identify optimal positioning strategies is proposed for doubles badminton players based on data-driven analysis.

This method consists of four steps as below:

- **Moving the Receiver's Position**: Here, a "receiver" in a team is defined as the player who receives or attempts to receive the shuttlecock. First, the position of one player is fixed and the receiver is moved to all possible grid positions within a 56×56 matrix. Subsequently, the team's control probability map is calculated by using the model, and the control probability at each position is denoted as $P_c(x, y)$.
- **Determining Candidate Positions**: Choose the candidates for the optimal position based on two conditions. Firstly, the shuttlecock's control probability for that position should be equal to p or higher than p ($P_c(x,y) \geq p$). Secondly, that position should be close to the receiver's actual position. To reduce the impact of randomness or

variability, the set of n-nearest grid locations to the actual play position of the receiver that is within $P_c(x,y) \geq p$ are selected first.

- **Clustering Candidates**: As the n-nearest grid locations may be separated and not closely located together as shown in Fig. 4.6 (a). To address this issue, the step of unsupervised clustering is added here, specifically using hierarchical clustering [112], to group these candidate positions and identify the largest cluster.

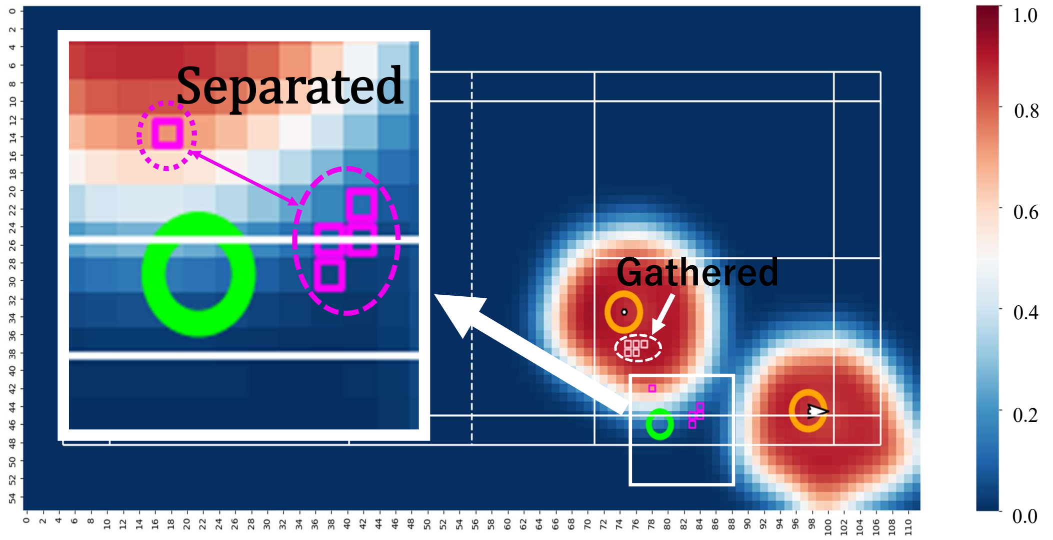
- **Calculating** the Average Position: Calculate the average position value of candidates within the largest cluster $C_{\max}$ to obtain the recommended position for the receiving player that increases the probability of controlling the shuttlecock to p as shown in Fig. 4.6 (b). Eq. (4.6) shows that the recommended position for the receiver is obtained by calculating the average value of all grid locations in the largest cluster $C_{\max}$, which is denoted by $(x_{\text{rec}}, y_{\text{rec}})$.

$$(x_{\text{rec}}, y_{\text{rec}}) = \frac{1}{n(C_{\max})} \sum_{(x,y)\in C_{\max}} (x,y) \tag{4.6}$$

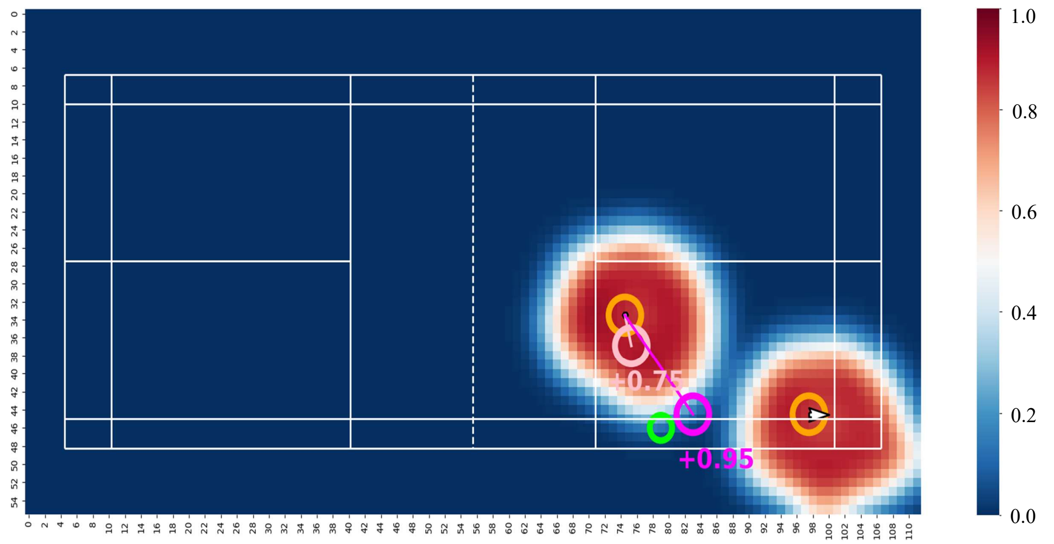
  Here, $n(C_{\max})$ denotes the number of grid locations in the largest cluster. For example, as shown in Fig. 4.6 (b), the average position values of only four candidates are calculated (four squares in magenta color) and are considered as the optimal position. Overall, this study's approach provides a data-driven and actionable recommendation for players and coaches to apply in practice and competition.

4.4 Experiments

In this section, several experiments are conducted to demonstrate the performance of the proposed method. These experiments include not only the training of the control area estimation model but also verification experiments to confirm the reliability of the estimated



(a) Example of separated optimal position candidates



(b) Recommended optimal positions

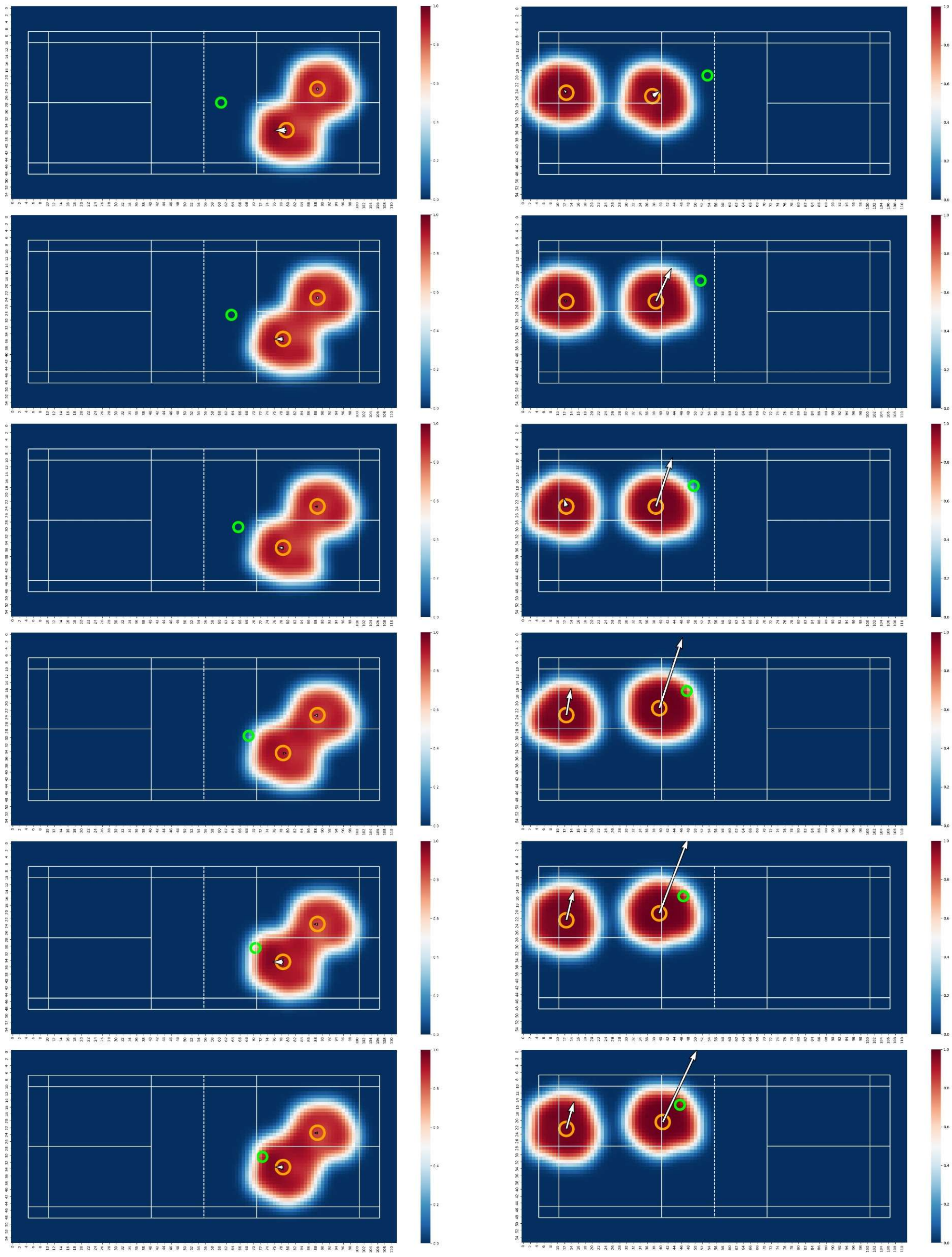
Figure 4.6: Optimal position computation. (a) Five optimal position candidates (pink and magenta squares) with $P_c(x,y) \geq 0.75$ and 0.95, respectively, with the pink squares gathered together while magenta squares separated into two clusters. The green circle indicates the shuttlecock's position and the yellow circles represent the actual player positions. (b) Optimal positions for $P_c(x,y) \geq 0.75$ and $P_c(x,y) \geq 0.95$ are determined by averaging the positions of candidates within the largest cluster that corresponds to these probabilities.

results. Specifically, the outcomes are verified by visualizing the estimated control areas and assessing their accuracy. Subsequently, practical insights of the proposed method are explored by investigating the relationship between the control area and the score. Additionally, the optimal positions for drop samples based on the proposed method are presented in 4.4.4, and its application on a real-world scale badminton court is demonstrated.

4.4.1 Control Area Estimation

In the training process, the control area estimation model is trained using the Adam optimizer [97] with a learning rate of 10^{-6} across 30 epochs and with batch sizes of 16. The dataset for the training is augmented with a horizontal flip, comprising 12,658 hit samples and 796 drop samples. 80% of the hit samples and 50% of the drop samples are used for training, with the remaining data reserved for testing.

After the training phase, the changes in the control area during the period when a team responds to an incoming shuttlecock after it crosses the net can be visualized. In sports like badminton, the concept of the control area becomes significantly meaningful for players only once the shuttlecock has crossed the net. Observing changes in the control area can aid in understanding the reasons for losing points. It is important to note here that the control area is defined as the region where the control probability is greater than or equal to 0.5. Fig. 4.6 presents the changes in the control area probability map of the receivers (on both sides) during a catch in a rally. Receivers can hit (receive) the shuttlecock in cases shown in Fig. 4.6 (a), (b), and (c). In the last case shown in Fig. 4.6 (d), the left-side receivers failed to catch the shuttlecock. The results of the control area are consistent with human expectations, and here, no assumptions is made about the distribution's shape within the control area (i.e., learned from data). Additionally, it can be observed that the model may learn the player's speed for estimating the occupied spaces of the control area.



(a) (b)

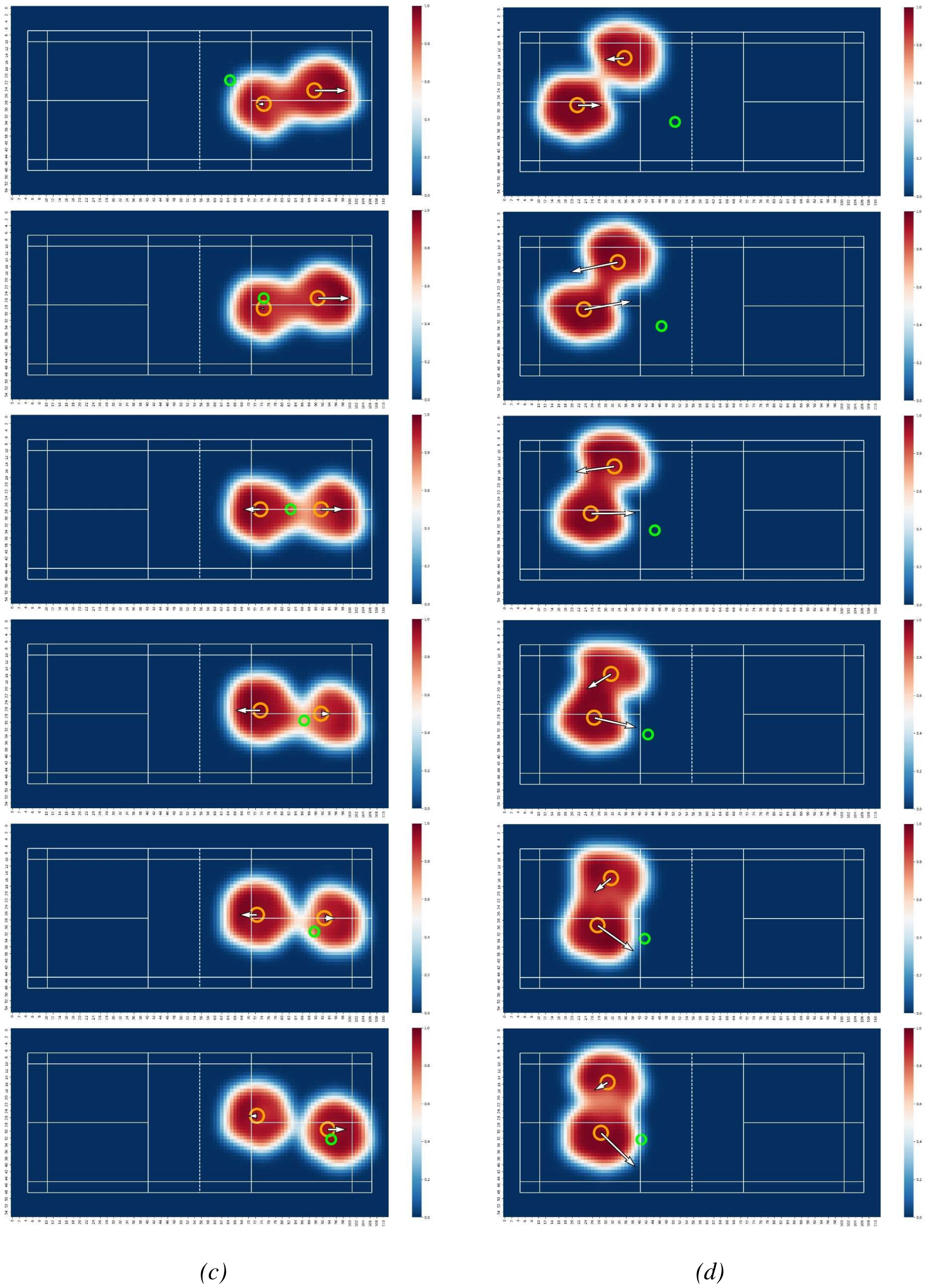


Figure 4.6: *Visualization of control area changes of the receivers during catches in a rally (time passes from top to bottom). Orange and lime circles represent players' locations and the shuttlecock location, respectively, while the arrows indicate the velocity vector for each player. A darker shade of red denotes a higher probability of control.*

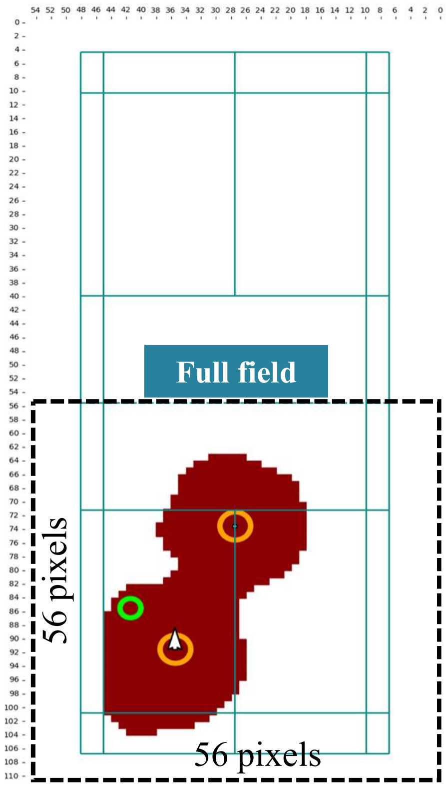
Table 4.2: Comparison of L_1 classification loss for each input feature by eliminating the different input features from the total input.

Error	Full	$-$Pose + Bbox (top)	$-$Pose	$-$Players' velocity
Control (hit)	**0.085**	0.092	0.100	0.118
Non-control (drop)	**0.238**	0.252	0.294	0.244
All	**0.094**	0.101	0.110	0.125

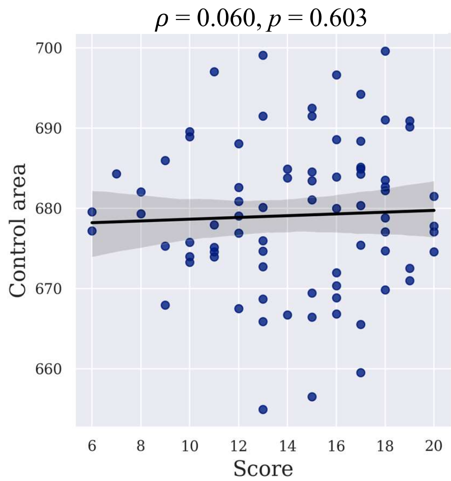
4.4.2 Verification of the Proposed Method

Here, the performance of the model is evaluated, and the impact of players' velocity ($-$ Players' velocity), players' pose ($-$ Pose), and computed Bounding box (Bbox) height and width from top-view ($-$ Pose + Bbox (top)) on the estimation performance is examined.

The last one replaced the players' poses in the full model with height and width measurements of bounding boxes (bboxes) as input to examine the effectiveness of the pose information. To evaluate the model, the L_1 classification loss between the ground-truth and estimated positions is computed for hit/drop samples. As presented in Table 4.2, a model trained with the full components (players' velocity and pose) achieved the best performance for both control (hit) and non-control (drop) samples, indicating that both the players' velocity and pose information contributed to accurately estimating the control area probability map. Notably, Bbox (top) was not included in the full components. Using the back view poses seems to have produced more accurate results compared to using the bboxes from the top view. In the model verification stage, an overall L_1 classification loss of 0.094 was achieved for the test samples with hit and drop shuttlecocks. Specifically, the L_1 classification loss for hit and drop samples were 0.085 and 0.238, respectively.



(a) Definition of the full field



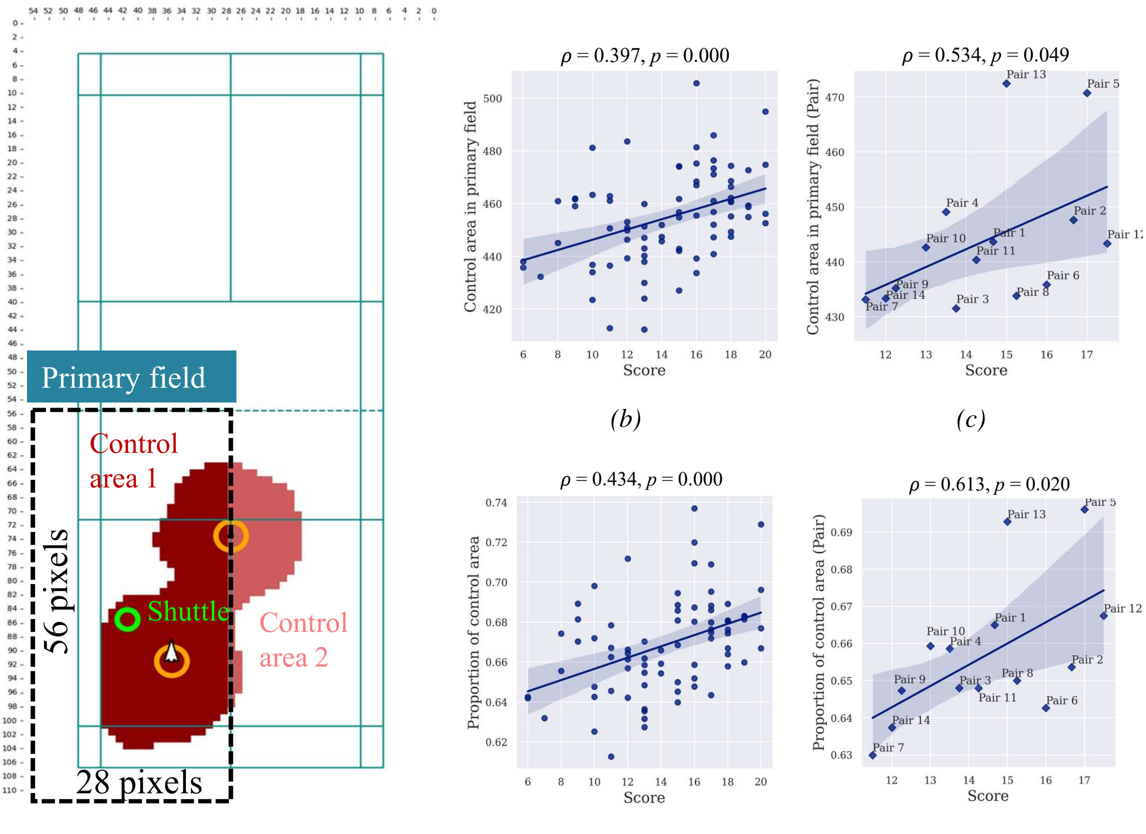
(b) Correlation between the score and the size of the control area

Figure 4.7: Correlation between control area and score.

4.4.3 Analysis of Relationship of Factors with the Score of the Match

4.4.3.1 Control Area in the Full Field

The full field is defined as half of the input map size (56×56 pixels) on the side of the receiving team, as shown in Fig. 4.7 (a). First, the relationship between the score and the size of the control area on the full field is examined. Fig. 4.7 (b) indicates that there is no correlation between the score and the size of the control area ($p > 0.05$). This can be speculated that the size of the control area may be related to the velocity of two players



(a) Definition of the primary field in relation to the shuttlecock's location.

(b) *(c)* *(d)* *(e)*

Figure 4.8: Analysis of score correlation with primary field's control area (b,c) and control area proportion (d,e) across games (b,d) and identified player pairs (c,e).

rather than the team's performance (defense capabilities). Therefore, in the next section, the size of the control area in the primary area is analyzed instead of the full field.

4.4.3.2 Control Area in the Primary Field

Next, the primary field is defined as the field where the shuttlecock is located, which has a size of one-quarter of the control area probability map (56×28 pixels). In Fig. 4.8, the control area in the primary field is analyzed, which depends on the shuttlecock location and the proportion of the control area to evaluate the team's performance in coverage, i.e., defense capabilities. The score indicates the number of points scored by the team in a game, and the control area in the primary field refers to $C_{\text{primary}} = \text{Control Area 1}$ if the shuttlecock is located in the left of the field as shown in Fig. 4.8 (a). The proportion of the control area refers to $P_{\text{control area}} = \frac{\text{Control Area 1}}{\text{Control Area 1} + \text{Control Area 2}}$ if the shuttlecock is located in the left of the field.

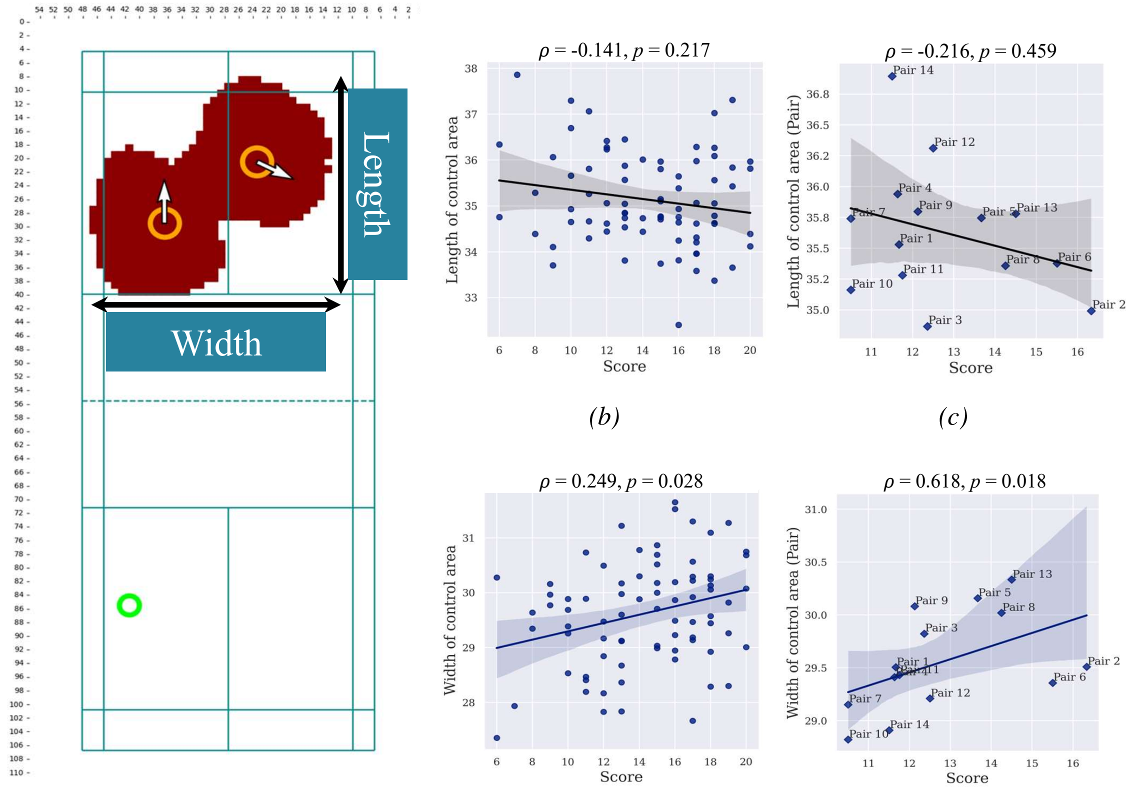
A moderate positive monotonic correlation between the score and the control area was found in the primary field (Fig. 4.8 (b), $\rho = 0.397$ ($p < 0.05$)) as well as between the proportion of the control area (Fig. 4.8 (d), $\rho = 0.434$ ($p < 0.05$)). In addition, for each pair, a strong positive monotonic correlation was observed between the score and the control area in the primary field (Fig. 4.8 (c), $\rho = 0.534$ ($p < 0.05$)), as well as between the proportion of the control area (Fig. 4.8 (e), $\rho = 0.613$ ($p < 0.05$)). These results suggest that the pair with better team performance tends to have a larger control area in the primary field where the shuttlecock is located.

4.4.3.3 Length/Width of the Control Area

In badminton, mastering the "doubles rotation" skill is crucial to maintaining court coverage and preventing any gap during the match. The ability to effectively cover the field is a valuable indicator of player performance.

To measure field coverage, the control area of each team is analyzed at the moment their opponents hit the shuttlecock. This is the time when both players in the team should prepare and position themselves for the next stroke (Fig. 4.9 (a)).

For each game and each pair, no correlation between the score and the length of the control area ($p > 0.05$) was found, as shown in Fig. 4.9 (b) and (c). However, a weak positive



(a) Definition of the length and width of the control area

(b)

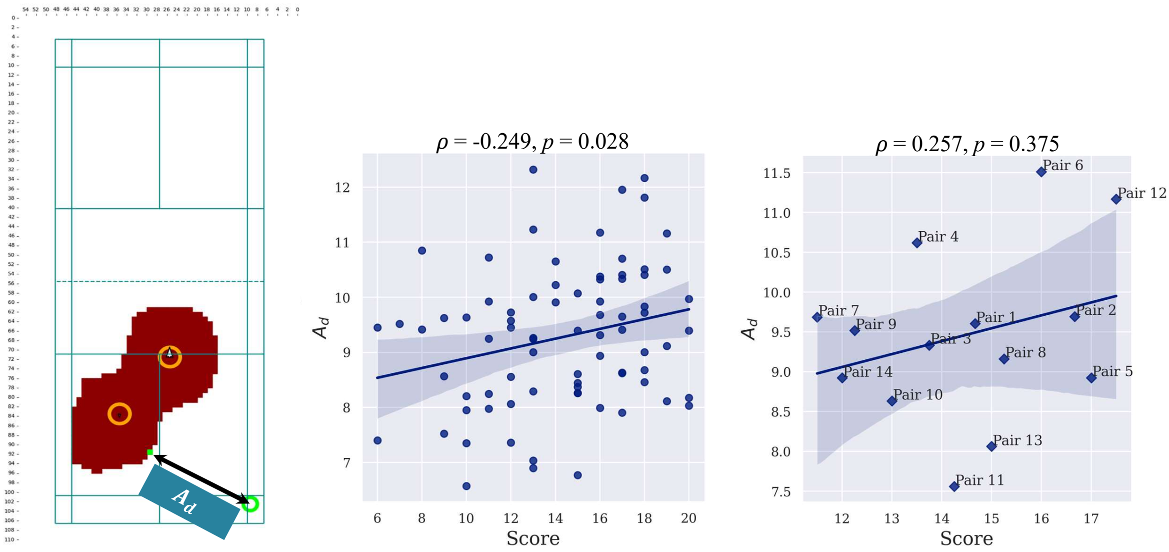
(c)

(d)

(e)

Figure 4.9: Analysis of score correlations with the (b, c) length and (d, e) width of the control area across (b, d) games and (c, e) identified player pairs.

monotonic correlation was observed between the score and the width of the control area for each game (Fig. 4.9 (d), $\rho = 0.249$ and $p < 0.05$), and a strong positive monotonic correlation between the score and the width of the control area for each pair (Fig. 4.9 (e), $\rho = 0.618$ and $p < 0.05$). These results suggest that teams with better performance tend to cover more width of the field when preparing for the next stroke.



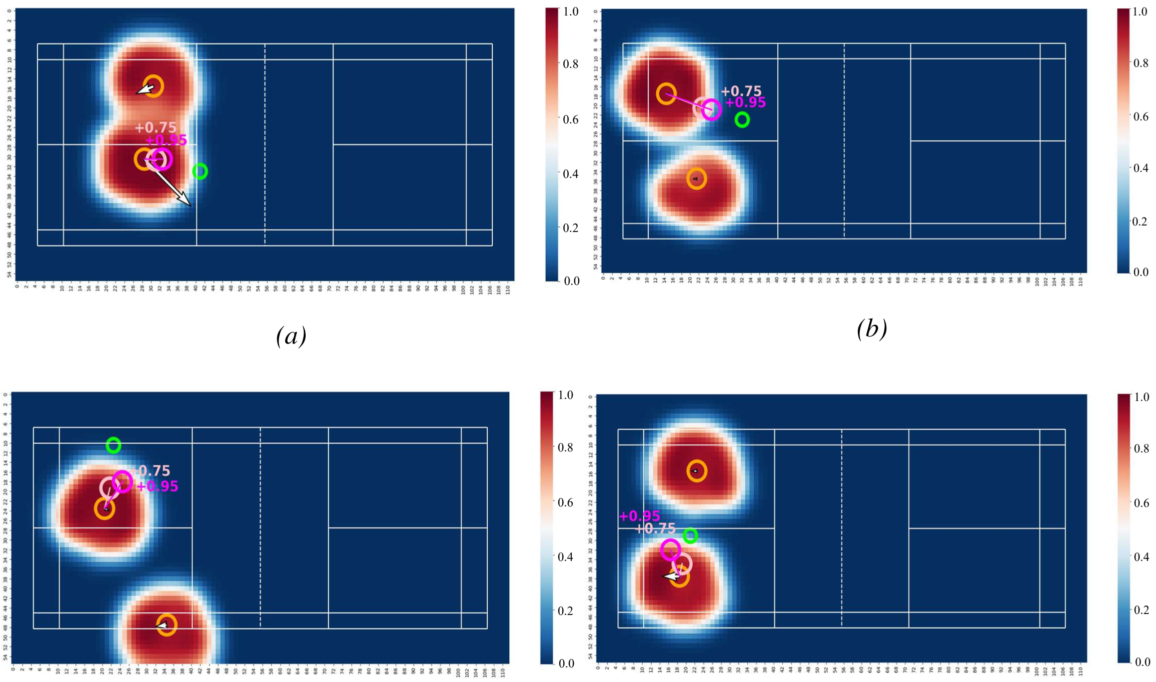
(a) Definition of aiming technique distance *(b) Correlation between the score and A_d for each game* *(c) Correlation between the score and A_d for each identified pair*

Figure 4.10: Correlation between aiming technique distance and score.

4.4.3.4 Aiming Technique

In a doubles badminton match, players look for opportunities to hit the shuttlecock towards areas where the opponent's formation is relatively distant during the moment of hitting. This strategy aims to increase the likelihood of their opponents making an error. Therefore, the maximum distance between the position where the player aims to land the shuttlecock and the control area of opponents across all rallies is used as a measure of this aiming technique, denoted as A_d, as shown in Fig. 4.10 (a). In cases where the player's shot is returned by the opponent, the player's aiming position with the opponent's actual hitting position is approximated.

A weak positive monotonic correlation was observed between the score and the A_d for each game (Fig. 4.10 (b), $\rho = 0.249$ and $p < 0.05$) For each pair, no correlation was found



(a) *(b)*

(c) *(d)*

Figure 4.11: Optimal positions for drop samples. Orange circles indicate the current positions of players, with arrows representing players' velocity. Pink and magenta circles indicate recommended positions for the receiver. Positioning in these spots can increase the probability of successfully receiving the shuttlecock to either 0.75 or 0.95.

between the score and the A_d ($p > 0.05$), as shown in Fig. 4.10 (c). These results suggest that the aiming technique measured by A_d has some impact on the overall score in a doubles badminton match. However, for individual pairs, other factors such as teamwork, communication, and individual skills, may also play important roles in determining the success of a doubles badminton pair.

4.4.4 Assessment of Optimal Positioning

As proposed in 4.3.3, optimal locations are defined as those that are near the receiver and provide a higher probability of a successful shot than the current location.

To improve the chances of controlling the shuttlecock during drop shots in doubles, it is recommended to move along the shortest path while maintaining the hitting pose. Specifically, it would be recommended to fix one player's location and consider all possible grid locations (56×56 pixels) for the receiving player.

In this case, the control probability model is used to identify the five nearest grid locations ($n = 5$) to the actual play position of the receiver where the probability of controlling the shuttlecock is greater than or equal to 0.75 and 0.95 ($P_c(x,y) \geq 0.75$ and $P_c(x,y) \geq 0.95$). Then, a hierarchical clustering algorithm is used to group these five locations into clusters and the largest cluster among them is identified. Finally, the average value of all locations is calculated in the largest cluster, which gives us a recommended position for the receiving player that increases the probability of controlling the shuttlecock to 0.75 (pink circle) and 0.95 (magenta circle), as shown in Fig. 4.11.

Furthermore, the research results are applied to a real badminton court and the distance a player needs to move in order to increase the control probability of the shuttlecock to either 0.75 or 0.95 is calculated. The distances of movement, as well as the distances in the x and y directions in centimeters, along with the angles, are marked in Fig. 4.12. This detailed analysis is crucial as it provides players with concrete data on the distance and direction they need to move to transition from their current position to the optimal one. By understanding these distances and angles, players can move more intelligently on the court, thereby reducing wasted effort and increasing efficiency. This experiment demonstrates that the proposed methods can be applied to real-world scenarios, effectively bridging the gap between theoretical research and practical application. It offers players and coaches a valuable tool for

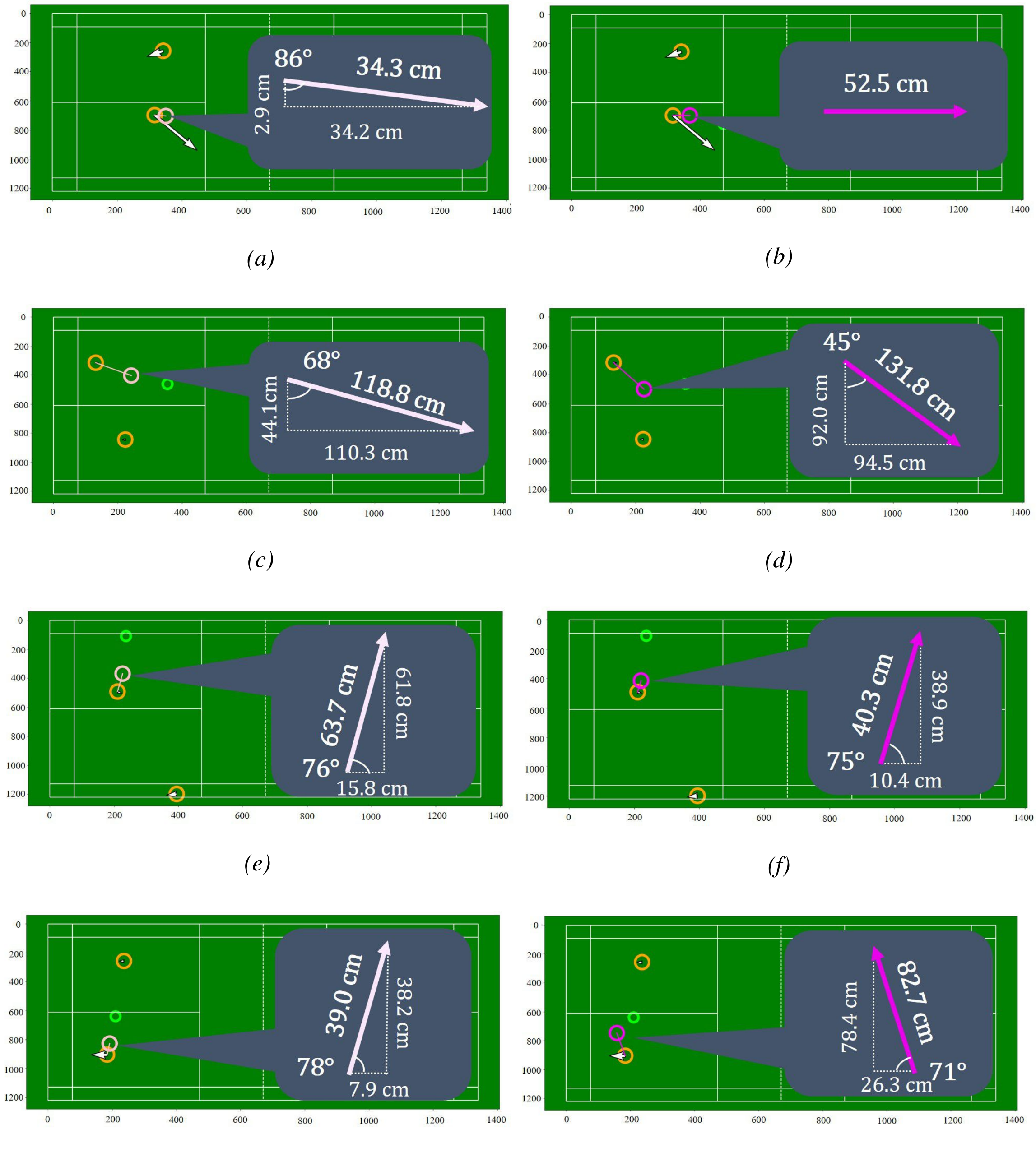


Figure 4.12: Optimal positions on an actual badminton court scale. (a), (c), (e), and (g) depict the positions for a shuttlecock control probability of 0.75 (indicated by pink circles), derived from drop shot samples. (b), (d), (f), and (h) show the positions for a higher control probability of 0.95 (magenta circles), also based on drop shot samples. Movement distances, measurements in x and y directions (cm), and angles are included.

improving play performance through smarter and more strategic movement on the court.

4.5 Summary

In this study, two primary contributions were made to the evaluation of doubles badminton matches using drone videos. Firstly, the first annotated drone video dataset for men's badminton doubles matches was introduced. This dataset serves as a valuable resource for researchers seeking to gain deeper insights into player movements and tactics in badminton doubles. Secondly, a deep neural network framework was developed to estimate the control area probability map from the drone videos. The proposed method utilized a Gaussian mixture map of player positions and incorporated graph convolution techniques based on their poses. To validate the proposed method, it was compared with various baselines and significant insights into the relationship between scoring and control areas were identified. Furthermore, a practical application was proposed to assess optimal positioning, which can be used to provide in-game instructions. By applying the proposed method to an actual badminton court, we could calculate the movement distance, allowing players to understand the effort needed to transition from their current position to the optimal one in real-world scenarios. A limitation of this study is that it estimated the average control area of players, rather than the control area specific to each pair. Recently, a method that can learn individual player styles [113] has emerged, providing a promising direction for future work. Additionally, based on the current study in badminton, there is potential for the proposed approach to be applied to other racket sports, such as table tennis and tennis. Overall, the proposed method in this study provides both a visual and quantitative evaluation of players' movements, offering invaluable insights into teamwork in doubles matches.

5 Conclusion

5.1 Summary of the Thesis

This thesis addressed the complex challenge of evaluating player performance in racket sports, specifically focusing on singles and doubles badminton matches. To overcome the limitations of traditional assessment methods, advanced deep learning techniques were employed to develop two innovative frameworks. These frameworks not only facilitate a thorough and precise evaluation of individual players and teams but also present the findings in a visually intuitive manner, greatly enhancing the comprehensibility and accessibility of the results.

Chapter 1 introduced the background and scope of this thesis, highlighting the unique challenges faced by deep learning-based methods in sports analysis. It outlined two main studies: Study I and Study II.

Chapter 2 presented a literature review related to this thesis, organized into three categories. The first category was related to play evaluation methods in sports, covering quantitative analysis across various sports, specifically racket sports. The second category was related to research on visual analytics, including its applications in team sports and racket sports. The third category included datasets related to racket sports.

In Chapter 3, for play evaluations in singles matches, a novel evaluation mechanism based on deep reinforcement learning was proposed (Study I). Contrary to traditional methods that predominantly focused on outcomes, this approach utilized historical data, including information about the tactical and technical performance of players, to learn the next-score probability as a Q-function. This function was then applied to assess the value of each stroke, offering a detailed understanding of performance on a stroke-by-stroke basis.

In Chapter 4, for play evaluations in doubles matches, a pioneering framework of deep neural networks to estimate the control area probability map for badminton doubles using drone video was proposed (Study II). This study introduced the first annotated drone dataset for badminton doubles, captured from both top and back perspectives. Additionally, a practical application to assess optimal positioning was proposed, offering insights that could be instrumental for coaching.

Overall, the two studies in this thesis shared the objective of advancing the evaluation of play performance in racket sports, though they each focus on different aspects. The first study (Study I) provided a quantitative evaluation of action value and visually presented the results to illustrate the evolution of player performance over time. This analysis highlighted that a match consists of a series of actions, with key actions offering insights into the reasons for winning or losing a point. The second study (Study II) provided a quantitative evaluation of the control area. It visually presents the results to enhance understanding of court utilization, positioning, and movement strategies in doubles matches.

Combined, these studies offer a more comprehensive understanding and actionable insights for coaching. Temporal analysis emphasizes the dynamic aspects of performance, while spatial analysis reveals strategic use of space and movement in racket sports. Together, they complement each other to provide a more comprehensive understanding and insights for actionable coaching.

5.2 Future Work

Although the methods proposed in this thesis can accurately quantify play performance and present these quantitative results visually, leading to beneficial visualizations for coaching strategies and enhancing the effectiveness of coaching methods, there remain several challenges that need to be addressed to fully realize the potential of these methods. Firstly, there is the issue of limited interpretability of pose data, which suggests a need for more advanced algorithms or techniques that can provide clearer and more actionable insights from player pose. Secondly, the generalization of these methods to other sports is a significant challenge. Since each sport has its own unique set of rules and play styles, tailored approaches are required for effective analysis. Lastly, there is the challenge of player-specific performance analysis.

Limited Interpretability of Pose Data

The research presented in this thesis currently falls short in interpreting the pose representations used in the model. The two studies included do not sufficiently demonstrate how adjusting the positions of the upper and lower limbs can optimize stroke quality or effectively increase the range of control areas. Therefore, the correlation between limb positions and play performance should be investigated more deeply, employing advanced analytics to understand the nuances of optimal limb arrangement. Understanding these correlations is key to developing training programs that can significantly enhance an athlete's performance. As a future plan, applying advanced data-driven deep learning models to identify various limb motion patterns would be beneficial. Additionally, developing simulation models to predict and analyze the outcomes of different limb positions is considered helpful in understanding theoretically best practices before applying them in real-world scenarios.

Generalization to Other Sports

Even within the domain of racket sports, different games may have unique characteristics that require specific modeling. For instance, table tennis involves players engaging in swift exchanges over short distances. The nature of this game demands rapid reflexes and precision, as players must hit the ball alternately after it bounces once on the table. This fast-paced environment with limited spatial movement presents distinct challenges for data collection and analysis. In contrast, tennis is played over larger areas, offering players a diverse range of shot options. The rules of tennis allow a player to strike the ball either before or after it bounces, which adds another layer of complexity to the game. In this thesis, badminton was examined, in which, players must return the shuttlecock before it touches the ground. This is a different dynamics compared to the return of a tennis or table tennis ball. Therefore, models developed for badminton might not be immediately applicable to other sports. Thus, in the future, expert knowledge of racket sports should be integrated with deep learning techniques to refine the models for each sport. This integration will enable the development of refined models tailored to the individual characteristics of each sport. Such specialized models would take into account the unique physical dynamics, play styles, and strategic elements inherent in each game. By doing so, it becomes possible to provide more accurate and useful insights for coaches and players in each specific sport.

Player-specific Performance

While generalized or average evaluations are invaluable for understanding broader trends and guiding large-scale strategic decisions, they may not capture the nuanced abilities and contributions of individual players. The proposed methods in this thesis are not able to model the unique strengths and potential of each athlete. However, player-specific performance assessments are crucial for tapping into the full potential of individual athletes. Such targeted evaluations enable coaches, managers, and analysts to pinpoint and cultivate the distinct skills of each player, leading to more personalized training and development strategies. Additionally, an individualized performance evaluation can be pivotal for career planning for professional players. Therefore, as future work, it is necessary to create an individualized performance assessment framework that can recognize and quantify each player's unique abilities. Methods such as updating models based on player IDs to estimate each player's performance, or adding player information like body size, are considered to be useful for player-specific performance evaluation.

Real-time Strategy Adaptation and Tactical Awareness

A player's strategy and gameplay often evolve in real-time, responding to the progression of a match, the strategies employed by the opponent, and internal team communication. This ability to dynamically adjust and react represents a core competency of an athlete. However, the methods currently detailed in this thesis do not adequately capture these dynamic aspects. As future work, it is necessary to explore effective methods for capturing and evaluating these real-time strategic shifts. Gaining insights into a player's decision-making, tactical awareness, and synergy with teammates can provide a more comprehensive evaluation of play performance, thereby offering targeted training suggestions for coaches. By integrating technologies such as motion capture systems, wearable sensors, and real-time data tracking systems, including eye tracking, it would be possible to gather comprehensive data on a player's decision-making process, their tactical awareness, and their non-verbal communication with teammates during the match.

Overall, these future directions aim to bridge the gap between theoretical models and practical applications in sports analytics. The future of sports performance analysis is not just about enhancing the physical aspects of play, but also about revolutionizing the entire ecosystem of sports. Fans will have access to more detailed and insightful analytics, deepening their understanding of the game and their connection to the players and teams they support. The integration of advanced technologies such as Artificial intelligence (AI), Virtual Reality (VR), and biometric data analysis will refine athlete performance and revolutionize coaching strategies and scouting processes. As a result, sports performance analysis will elevate not only the level of play but also the cultural and social aspects of sports, thereby making them more engaging and enjoyable for all involved.

Acknowledgement

First and foremost, I wish to extend my deepest appreciation to Prof. Kazuya Takeda of Nagoya University for his unwavering supervision and encouragement throughout my doctoral course. His guidance in research and his timely assistance during critical moments have been invaluable. Without Prof. Takeda's mentorship, completing this work would have been impossible.

I must also express my profound gratitude to Associate Prof. Keisuke Fujii of Nagoya University. His insightful suggestions and relentless support have been fundamental throughout my academic journey. He taught me the intricacies of conducting experiments and the art of paper writing. Prof. Fujii not only guided us in our research but also set an exemplary standard through his own actions. He illuminated the skills and qualities essential for a researcher, serving as our compass on the journey of academic growth. His genuine care for his students goes beyond academic progress; he has been a pillar of emotional support. The path to a Ph.D. is rarely straightforward or wholly pleasant. In times of stress and uncertainty, he stood by my side, patiently guided me, and helped me face challenges and find solutions.

Being a part of this lab and having the privilege to work under the tutelage of Prof. Takeda and Prof. Fujii has truly been a blessing. The success of this research is a testament to their unwavering guidance and support. In addition to my advisors, I want to express my heartfelt gratitude to Prof. Tomoki Toda and Associate Prof. Ming Ding. Their invaluable insights and comments have been instrumental throughout my doctoral journey.

Furthermore, my gratitude extends to Assistant Prof. Kazushi Tsutsui of Nagoya University for his kind support. I also owe a tremendous debt of gratitude to two esteemed researchers from the sports education field in China: Associate Prof. Yingjiu Bei of Anhui Normal University and Prof. Wenhui Jin of Wuhu Institute of Technology. Along with Mr. Yundong Yu of Wuhu Institute of Technology, their invaluable assistance with data collection was instrumental to the success of this work.

Special appreciation goes to Mr. Tatsuya Yoshikawa, Mr. Scott Atom, as well as other members of the sports behavior group in Takeda Lab. I had the pleasure of working closely with them, and their contributions and teamwork were invaluable to our collaborative efforts.

Last but not least, I would like to express my love and gratitude to my family for their unconditional love, understanding, and support over these years.

References

[1] H. Fang, S. Xie, Y. Tai, and C. Lu, “RMPE: Regional multi-person pose estimation,” in *Proceedings of the 16th IEEE International Conference on Computer Vision*, 2017, pp. 2334–2343.

[2] Y. Huang, I. Liao, C. Chen, T. İk, and W. Peng, “TrackNet: A deep learning network for tracking high-speed and tiny objects in sports applications,” in *Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2019, 8 pages.

[3] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, G. Thore, and H. Demis, “Mastering the game of Go without human knowledge,” *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[4] L. Vergauwen, A. J. Spaepen, J. Lefevre, and P. Hespel, “Evaluation of stroke performance in tennis,” *Occupational Health and Industrial Medicine*, vol. 5, no. 39, pp. 238–239, 1998.

[5] I. Malagoli Lanzont, R. Di Michele, and F. Merni, “Reliability of selected table tennis performance indicators,” *International Journal of Table Tennis Sciences*, vol. 7, pp. 62–65, 2012.

[6] X. Qiu, H. Zhang, J. Wei, and J. Liu, "Machine learning based movement analysis and correction for table tennis," in *Proceedings of the IEEE 8th International Conference on Cloud Computing and Intelligent Systems*, 2022, pp. 150–154.

[7] I. Ghosh, S. R. Ramamurthy, and N. Roy, "StanceScorer: A data driven approach to score badminton player," in *Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops*, 2020, 6 pages.

[8] S. Buch, V. Escorcia, C. Shen, B. Ghanem, and J. Carlos Niebles, "SST: Single-stream temporal action proposals," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2911–2920.

[9] M. Ullah and F. Alaya Cheikh, "A directed sparse graphical model for multi-target tracking," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1816–1823.

[10] L. Bridgeman, M. Volino, J.-Y. Guillemaut, and A. Hilton, "Multi-person 3D pose estimation and tracking in sports," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 2487–2496.

[11] N. Miyahara, T. Tezuka, and Y. Nakauchi, "Pattern recognition for tennis tactics using hidden Markov model from rally series," in *Proceedings of the 2019 IEEE/SICE International Symposium on System Integration*, 2019, pp. 751–755.

[12] W. Wang, T. Chan, H. Yang, C. Wang, Y. Fan, and W. Peng, "Exploring the long short-term dependencies to infer shot influence in badminton matches," in *Proceedings of the 21st IEEE International Conference on Data Mining*, 2021, pp. 1397–1402.

[13] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, "A

general reinforcement learning algorithm that masters chess, Shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.

[14] M. Sharma, N. Kumar, and P. Kumar, "Badminton match outcome prediction model using naïve Bayes and feature weighting technique," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8441–8455, 2021.

[15] K.-W. Ban, J. See, J. Abdullah, and Y. P. Loh, "BadmintonDB: A badminton dataset for player-specific match analysis and prediction," in *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, 2022, pp. 47–54.

[16] J. Wang, D. Deng, X. Xie, X. Shu, Y. Huang, L. Cai, H. Zhang, M. Zhang, Z. Zhou, and Y. Wu, "Tac-Valuer: Knowledge-based stroke evaluation in table tennis," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2021, pp. 3688–3696.

[17] K.-S. Chang, W.-Y. Wang, and W.-C. Peng, "Where will players move next? Dynamic graphs and hierarchical fusion for movement forecasting in badminton," in *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023, pp. 6998–7005.

[18] M. Archana and M. Kalaiselvi Geetha, "An efficient ball and player detection in broadcast tennis video," in *Proceedings of the 2016 International Symposium on Intelligent Systems Technologies and Applications*, 2016, pp. 427–436.

[19] N. Ding, K. Takeda, and K. Fujii, "Deep reinforcement learning in a racket sport for player evaluation with technical and tactical contexts," *IEEE Access*, vol. 10, pp. 54 764–54 772, 2022.

[20] G. Munivrana, L. Z. Petrinović, and M. Kondrič, "Structural analysis of technical-tactical elements in table tennis and their role in different playing zones," *Journal of Human Kinetics*, vol. 47, pp. 197–214, 2015.

[21] N. S. Kolman, B. C. Huijgen, C. Visscher, and M. T. Elferink-Gemser, "The value of technical characteristics for future performance in youth tennis players: A prospective study," *PloS One*, vol. 16, no. 1, pp. e0 245 435_1–13, 2021.

[22] I. M. Franks and P. Nagelkerke, "The use of computer interactive video in sport analysis," *Ergonomics*, vol. 31, pp. 1593–1603, 1988.

[23] V. Sarlis and C. Tjortjis, "Sports analytics —Evaluation of basketball players and team performance," *Information Systems*, vol. 93, no. 101562, pp. 1–19, 2020.

[24] B. Rosner, F. Mosteller, and C. Youtz, "Modeling pitcher performance and the distribution of runs per inning in major league baseball," *The American Statistician*, vol. 50, no. 4, pp. 352–360, 1996.

[25] D. J. Berri, "Who is "most valuable"? Measuring the player's production of wins in the National Basketball Association," *Managerial and Decision Economics*, vol. 20, no. 8, pp. 411–427, 1999.

[26] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, "POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data," in *Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference*, 2014, 9 pages.

[27] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis, "Actions speak louder than goals: Valuing player actions in soccer," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1851–1861.

[28] K. Toda, M. Teranishi, K. Kushiro, and K. Fujii, "Evaluation of soccer team defense based on prediction models of ball recovery and being attacked," *PLoS One*, vol. 17, no. 1, pp. e0 263 051_1–14, 2022.

[29] K. Fujii, "Data-driven analysis for understanding team sports behaviors," *Journal of Robotics and Mechatronics*, vol. 33, no. 3, pp. 505–514, 2021.

[30] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, "A multiresolution stochastic process model for predicting basketball possession outcomes," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 585–599, 2016.

[31] J. Fernández, L. Bornn, and D. Cervone, "Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer," in *Proceedings of the 13th Annual MIT Sloan Sports Analytics Conference*, 2019, 20 pages.

[32] D. Cervone, L. Bornn, and K. Goldsberry, "NBA court realty," in *Proceedings of the 10th Annual MIT Sloan Sports Analytics Conference*, 2016, 8 pages.

[33] J. Fernandez and L. Bornn, "Wide open spaces: A statistical technique for measuring space creation in professional soccer," in *Proceedings of the 12th Annual MIT Sloan Sports Analytics Conference*, 2018, 19 pages.

[34] T. Taki and J.-I. Hasegawa, "Visualization of dominant region in team games and its application to teamwork analysis," in *Proceedings of the Computer Graphics International 2000 Conference*, 2000, pp. 227–235.

[35] G. Liu and O. Schulte, "Deep reinforcement learning in ice hockey for context-aware player evaluation," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3442–3448.

[36] G. Liu, Y. Luo, O. Schulte, and T. Kharrat, “Deep soccer analytics: Learning an action-value function for evaluating soccer players,” *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1531–1559, 2020.

[37] X. Sun, J. Davis, O. Schulte, and G. Liu, “Cracking the black box: Distilling deep sports analytics,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3154–3162.

[38] M. Fuchs, R. Liu, I. Malagoli Lanzoni, G. Munivrana, G. Straub, S. Tamaki, K. Yoshida, H. Zhang, and M. Lames, “Table tennis match analysis: A review,” *Journal of Sports Sciences*, vol. 36, no. 23, pp. 2653–2662, 2018.

[39] K. Purcell, “A tennis forehand-backhand drive skill test which measures ball control and stroke firmness,” *Research Quarterly for Exercise and Sport*, vol. 52, no. 2, pp. 238–245, 1981.

[40] M. A. Gomez, A. S. Leicht, F. Rivas, and P. Furley, “Long rallies and next rally performances in elite men’s and women’s badminton,” *PloS one*, vol. 15, no. 3, pp. e0 229 604_1–16, 2020.

[41] M.-Á. Gómez-Ruano, A. Cid, F. Rivas, and L.-M. Ruiz, “Serving patterns of women’s badminton medalists in the Rio 2016 Olympic Games,” *Frontiers in Psychology*, vol. 11, no. 136, pp. 1–9.

[42] M. Pfeiffer, H. Zhang, and A. Hohmann, “A Markov chain model of elite table tennis competition,” *International Journal of Sports Science & Coaching*, vol. 5, no. 2, pp. 205–222, 2010.

[43] J. Wang, K. Zhao, D. Deng, A. Cao, X. Xie, Z. Zhou, H. Zhang, and Y. Wu, "Tac-Simur: Tactic-based simulative visual analytics of table tennis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 407–417, 2019.

[44] S. Wenninger and M. Lames, "Markov simulation by numerical derivation in table tennis," in *Proceedings of the 10th International Symposium on Computer Science in Sports*, 2016, pp. 161–169.

[45] I. Ghosh, S. R. Ramamurthy, A. Chakma, and N. Roy, "DeCoach: Deep learning-based coaching for badminton player assessment," *Pervasive and Mobile Computing*, vol. 83, no. 101608, 20 pages.

[46] T. McGarry and I. M. Franks, "Development, application, and limitation of a stochastic Markov model in explaining championship squash performance," *Research Quarterly for Exercise and Sport*, vol. 67, no. 4, pp. 406–415, 1996.

[47] J. Wang, D. Deng, X. Xie, X. Shu, Y.-X. Huang, L.-W. Cai, H. Zhang, M.-L. Zhang, Z.-H. Zhou, and Y. Wu, "Tac-Valuer: Knowledge-based stroke evaluation in table tennis," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3688–3696.

[48] K. A. Cook and J. J. Thomas, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Richland, WA, USA: National Visualization and Analytics Center, 2005.

[49] M. Du and X. Yuan, "A survey of competitive sports data visualization and visual analysis," *Journal of Visualization*, vol. 24, no. 1, pp. 47–67, 2021.

[50] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale, "State of the art of sports data visualization," *Computer Graphics Forum*, vol. 37, no. 3, pp. 663–686, 2018.

[51] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop, "Physics-based modeling of pass probabilities in soccer," in *Proceeding of the 11th Annual MIT Sloan Sports Analytics Conference*, 2017, 14 pages.

[52] R. A. Yeh, A. G. Schwing, J. Huang, and K. Murphy, "Diverse generation for multi-agent sports games," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4610–4619.

[53] J. Fernández and L. Bornn, "SoccerMap: A deep learning architecture for visually-interpretable analysis in soccer," in *Proceeding of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2020*, vol. 5, 2021, pp. 491–506.

[54] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko, "Snapshot: Visualization to propel ice hockey analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2819–2828, 2012.

[55] K. Goldsberry, "CourtVision: New visual and spatial analytics for the NBA," in *Proceeding of the 6th Annual MIT Sloan Sports Analytics Conference*, 2012, pp. 12–15.

[56] C. Perin, R. Vuillemot, and J.-D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2506–2515, 2013.

[57] P. A. Legg, D. H. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen, "MatchPad: Interactive glyph-based visualization for real-time sports performance analysis," vol. 31, pp. 1255–1264, 2012.

[58] T. Polk, D. Jäckle, J. Häußler, and J. Yang, "CourtTime: Generating actionable insights into tennis matches using visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 397–406, 2019.

[59] T. Polk, J. Yang, Y. Hu, and Y. Zhao, "TenniVis: Visualization for tennis match analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2339–2348, 2014.

[60] J. Wang, J. Wu, A. Cao, Z. Zhou, H. Zhang, and Y. Wu, "Tac-Miner: Visual tactic mining for multiple table tennis matches," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 6, pp. 2770–2782, 2021.

[61] Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, and H. Zhang, "iTTVis: Interactive visualization of table tennis data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 709–718, 2017.

[62] J. Lan, J. Wang, X. Shu, Z. Zhou, H. Zhang, and Y. Wu, "RallyComparator: Visual comparison of the multivariate and spatial stroke sequence in table tennis rally," *Journal of Visualization*, vol. 25, no. 1, pp. 143–158, 2022.

[63] X. Chu, X. Xie, S. Ye, H. Lu, H. Xiao, Z. Yuan, Z. Chen, H. Zhang, and Y. Wu, "TIVEE: Visual exploration and explanation of badminton tactics in immersive visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 118–128, 2021.

[64] M. A. Haq, S. Tarashima, and N. Tagawa, "Heatmap visualization and badminton player detection using convolutional neural network," in *Proceedings of the 2022 International Electronics Symposium*, 2022, pp. 627–631.

[65] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH a spatio-temporal maximum average correlation height filter for action recognition," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, 8 pages.

[66] J. C. Niebles, C.-W. Chen, and L. Fei-Fei, "Modeling temporal structure of decomposable motion segments for activity classification," in *Proceedings of the 11th European Conference on Computer Vision*, vol. Part II, 2010, pp. 392–405.

[67] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[68] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "SoccerNet: A scalable dataset for action spotting in soccer videos," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1711–1721.

[69] A. Deliege, A. Cioppa, S. Giancola, M. J. Seikavandi, J. V. Dueholm, K. Nasrollahi, B. Ghanem, T. B. Moeslund, and M. Van Droogenbroeck, "SoccerNet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos," in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4508–4519.

[70] Y. Jiang, K. Cui, L. Chen, C. Wang, and C. Xu, “SoccerDB: A large-scale database for comprehensive video understanding,” in *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports*, 2020, 8 pages.

[71] V. Ramanathan, J. Huang, S. Abu-El-Haija, A. Gorban, K. Murphy, and L. Fei-Fei, “Detecting events and key actors in multi-person videos,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3043–3053.

[72] C. Ma, J. Fan, J. Yao, and T. Zhang, “NPU RGBD dataset and a feature-enhanced LSTM-DGCN method for action recognition of basketball players,” *Applied Sciences*, vol. 11, no. 4426, 27 pages.

[73] E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthankar, “Violence detection in video using computer vision techniques,” in *Proceedings of the 14th International Conference on Computer Analysis of Images and Patterns*, 2011, pp. 332–339.

[74] K. Vats, P. Walters, M. Fani, D. A. Clausi, and J. S. Zelek, “Player tracking and identification in ice hockey,” *Expert Systems with Applications*, vol. 213, no. 119250, 2023, 14 pages.

[75] Y. Yang, M. Xu, W. Wu, R. Zhang, and Y. Peng, “3D multiview basketball players detection and localization based on probabilistic occupancy,” in *Proceedings of the 2018 International Conference on Digital Image Computing: Techniques and Applications*, 2018, pp. 267–274.

[76] A. Scott, I. Uchida, M. Onishi, Y. Kameda, K. Fukui, and K. Fujii, “SoccerTrack: A dataset and tracking algorithm for soccer with fish-eye and drone videos,” in *Proceed-*

ings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3569–3579.

[77] H. Zhao, S. Wang, G. Zhou, and W. Jung, "TennisEye: Tennis ball speed estimation using a racket-mounted motion sensor," in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, 2019, pp. 241–252.

[78] J. D. Pulgarin-Giraldo, A. M. Álvarez-Meza, L. G. Melo-Betancourt, S. Ramos-Bermudez, and G. Castellanos-Dominguez, "A similarity indicator for differentiating kinematic performance between qualified tennis players," in *Proceedings of the 22nd Iberoamerican Congress on Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 2017, pp. 309–317.

[79] S. Gourgari, G. Goudelis, K. Karpouzis, and S. Kollias, "THETIS: Three dimensional tennis shots a human action dataset," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 676–681.

[80] P.-E. Martin, J. Benois-Pineau, R. Péteri, and J. Morlier, "Sport action recognition with Siamese spatio-temporal CNNs: Application to table tennis," in *Proceedings of the 2018 International Conference on Content-Based Multimedia Indexing*, 2018, 6 pages.

[81] S. Schwarcz, P. Xu, D. B. D'Ambrosio, J. Kangaspunta, A. Angelova, H. Phan, and N. Jaitly, "SPIN: A high speed, high resolution vision dataset for tracking and action recognition in ping pong," *Computing Research Repository arXiv Preprint*, 2019, arXiv:1912.06640.

[82] A. Ghosh, S. Singh, and C. Jawahar, “Towards structured analysis of broadcast badminton videos,” in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision*, 2018, pp. 296–304.

[83] R. Voeikov, N. Falaleev, and R. Baikulov, “TTNet: Real-time temporal and spatial video analysis of table tennis,” in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 884–885.

[84] P. Blank, J. Hoßbach, D. Schuldhaus, and B. M. Eskofier, “Sensor-based stroke detection and stroke type classification in table tennis,” in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 2015, pp. 93–100.

[85] K. M. Kulkarni and S. Shenoy, “Table tennis stroke recognition using two-dimensional human pose estimation,” in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4576–4584.

[86] J. Ensum, R. Pollard, and S. Taylor, “Applications of logistic regression to shots at goal in association football,” in *Proceedings of the 5th World Congress on Science and Football*, 2005, pp. 214–221.

[87] L. Shen, H. Zhang, M. Zhu, J. Zheng, and Y. Ren, “Measurement and performance evaluation of lob technique using aerodynamic model in badminton matches,” in *Proceedings of the 12th International Symposium on Computer Science in Sport*, 2019, pp. 53–58.

[88] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT press, 2018.

[89] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Diele-

man, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[90] A. E. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electronic Imaging*, vol. 2017, no. 19, pp. 70–76, 2017.

[91] Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang, and Y. Wang, "Deep reinforcement learning for dynamic treatment regimes on medical registry data," in *Proceedings of the 2017 IEEE International Conference on Healthcare Informatics*, 2017, pp. 380–385.

[92] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation*, 2017, pp. 3389–3396.

[93] M. Hausknecht and P. Stone, "Deep recurrent q-learning for partially observable MDPS," in *Proceedings of the 2015 AAAI Fall Symposium Series*, pp. 31–39.

[94] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, S. Ostrovski, Georg Petersen, A. Beattie Charles, Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[95] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proceedings of the 13th European Conference on Computer Vision*, vol. Part V, 2014, pp. 740–755.

[96] K. Routley and O. Schulte, “A Markov game model for valuing player actions in ice hockey,” in *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, 2015, pp. 782–791.

[97] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *Computing Research Repository arXiv Preprint*, 2014, arXiv:1412.6980.

[98] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, “GANerated hands for real-time 3D hand tracking from monocular RGB,” in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 49–59.

[99] R. Boutteau, R. Rossi, L. Qin, P. Merriaux, and X. Savatier, “A vision-based system for robot localization in large industrial environments,” *Journal of Intelligent & Robotic Systems*, vol. 99, pp. 359–370, 2020.

[100] K. Dasgupta, A. Das, S. Das, U. Bhattacharya, and S. Yogamani, “Spatio-contextual deep network-based multimodal pedestrian detection for autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 940–15 950, 2022.

[101] N. Wawrzyniak, T. Hyla, and A. Popik, “Vessel detection and tracking method based on video surveillance,” *Sensors*, vol. 19, no. 5230, 2019, 14 pages.

[102] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, “ByteTrack: Multi-object tracking by associating every detection box,” in *Proceedings of the 17th European Conference on Computer Vision*, vol. Part XXII, 2022, pp. 1–21.

[103] "OpenMMLab: Pose estimation toolbox and benchmark," https://github.com/open-mmlab/mmpose, 2020 (accessed: Dec. 29, 2023).

[104] S. Giancola and B. Ghanem, "Temporally-aware feature pooling for action spotting in soccer broadcasts," in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4490–4499.

[105] H. Cho, H. Ryu, and M. Song, "Pass2vec: Analyzing soccer players' passing style using deep learning," *International Journal of Sports Science & Coaching*, vol. 17, no. 2, pp. 355–365, 2022.

[106] T. Hsu, C. Chen, N. P. Jut, T. İk, W. Peng, Y. Wang, Y. Tseng, J. Huang, Y. Ching, C. Wang, and Y. Lin, "CoachAI: A project for microscopic badminton match data collection and tactical analysis," in *Proceedings of the 20th Asia-Pacific Network Operations and Management Symposium*, 2019, 8 pages.

[107] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 3, 2015, pp. 234–241.

[108] W. M. Association, "World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects," *The Journal of the American Medical Association*, vol. 310, no. 20, pp. 2191–2194, 2013.

[109] E. Dubrofsky, "Homography estimation," Master's thesis, Vancouver, British Columbia, Canada: Department of Computer Science, University of British Columbia, 2009.

[110] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the 16th IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[111] W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," *IEEE Transactions on Image Processing*, vol. 29, pp. 8055–8068, 2020.

[112] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[113] P. Srinivasan, R. Subramanian, and W. Knottenbelt, "Thinking the GOAT: Imitating tennis styles," in *Proceedings of the 17th Annual MIT Sloan Sports Analytics Conference*, 2023, 17 pages.

List of Publications

Journal Papers

[1] N. Ding, K. Takeda, W. Jin, Y. Bei, and K. Fujii, “Estimation of control area in badminton doubles with pose information from top and back view drone videos,” *Multimedia Tools and Applications*, August 2023. DOI: 10.1007/s11042-023-16362-1

[2] N. Ding, K. Takeda, and K. Fujii, “Deep reinforcement learning in a racket sport for player evaluation with technical and tactical contexts,” *IEEE Access*, vol. 10, pp. 54764–54772, May 2022. DOI: 10.1109/ACCESS.2022.3175314

Conferences

[3] N. Ding, K. Takeda, Y. Bei, and K. Fujii, “Visual analysis of control area in badminton doubles using drone video dataset,” in *Proceedings of the 10th MathSport International Conference*, Budapest, Hungary, June 2023, 6 pages.

[4] N. Ding, K. Takeda, and K. Fujii, “Player evaluation in a racket sport via deep reinforcement learning with technical and tactical contexts,” in *Proceedings of the 36th Annual Conference of the Japanese Society for Artificial Intelligence*, Kyoto, Japan, June 2022, no. 1S1IS304, 8 pages.