Submitted in fulfillment of the requirements for the degree of
Master of Science in Computer Science:
Computer Networks and Distributed Systems

Graph Neural Networks for Event Spotting in Football

Karol Zioło

Adviser: Prof. Tim Verdonck

Supervisor: Leonid Kholkine

Thomas Servotte

Faculty of Science
2024

Contents

1 Introduction **1**

1.1 Sports analytics: A market overview 2

1.2 Evolution of sports analytics: A research perspective 4

1.3 Framework for event detection 7

1.4 Research objectives and challenges 9

2 Background and related work **12**

2.1 Related work for event detection 13

2.2 Spatial analysis: Graph Neural Networks 17

2.3 Spatial analysis: NetVLAD 21

2.4 Temporal analysis: Context Aware Loss Function 22

2.5 Final thoughts 24

3 Methodology **27**

3.1 Data collection and preparation 28

3.2 Segmentation module 31

3.3 Spotting module 36

3.4 Training 38

4 Results **40**

4.1 Data analysis 41

4.2 Model performance 44

4.3 Comparison of the model outputs 48

5 Discussion **53**

5.1 Interpretation of results 54

5.2 Research Findings 60

5.3 Implications 61

5.4 Limitations 62

5.5 Future work 63

6 Conclusion **65**

Bibliography **68**

Acknowledgements

I would like to send my sincere gratitude to my supervisors, Leonid Kholkine and Thomas Servotte, as well as my advisor, Professor Tim Verdonck, for their invaluable feedback, insights, and guidance throughout the course of this research. Their expertise and thoughtful direction helped shape both the vision and execution of this thesis.

I am also thankful to all the professors who have shared their knowledge over the past two years, laying the educational foundation necessary for me to undertake and complete this work. Their teachings were crucial in equipping me with the skills required to navigate complex research challenges.

A special thank you to the research group IDLab, which provided the essential hardware resources needed for my research. Their support played a key role in facilitating the computational aspects of my work, allowing me to explore and implement advanced technological solutions.

On a personal note, I am extremely grateful to my family and friends, whose strong support and encouragement have been my constant source of strength and motivation. Their belief in my abilities and their emotional support during challenging times have been irreplaceable.

This journey would not have been possible without the collective support and encouragement from each individual mentioned above, and I am deeply appreciative of everyone who contributed to my academic and personal growth during this significant phase of my life.

Summary

In this thesis we set the goal of improving event detection tools in football through an advanced framework using positional data. A thorough exploration of the research landscape highlighted current trends, leading to the innovative integration of Graph Neural Networks (GNNs) with a Context-Aware Loss function, and exploring the potential of NetVLAD in a slightly unique way. These techniques aimed to extract detailed spatial and temporal information essential for analyzing the dynamic nature of football. Moreover, state-of-the-art applications in sports analytics have focused on detecting less frequent events such as offsides or goals within a relatively broad time frame. This thesis also focused on these limitations by enhancing frame-level event prediction and addressing both rare and common occurrences, such as passes, within game dynamics.

The findings of this research reveal several critical insights. While the models developed are not yet precise enough for exact event detection in real-time applications, they have a potential to be used as a tools for identifying key phases of gameplay, which can be invaluable for strategic analysis. Moreover, the reliance only on positional data proved insufficient for complete event understanding, suggesting a need for integration with visual or audio data to enrich the context and accuracy of event detection. This thesis not only advances the field of sports analytics by addressing current shortcomings but also sets the stage for future improvements that could fully realize the potential of advanced event detection in football.

List of Tables

3.1 Challenges in detecting specific events. 29

4.1 Segmentation evaluation results. 45

4.2 Spotting evaluation results. 47

5.1 Conclusions derived from the qualitative analysis. 59

List of Figures

1.1 Spatiotemporal event detection workflow. 8

2.1 Behaviour of the Context Aware Loss Function (CALF). . . . 25

3.1 The pipeline of the event detection model. 31

3.2 Optimised Context Aware Loss Function. 34

3.3 Context Aware Loss Function impact on the segmentation outputs. 35

3.4 Artificially generated actual probabilities. 35

4.1 Player heat maps Brasil vs Belgium WC2018. 42

4.2 Snapshot of the animation with edges. 43

4.3 Snapshot of the animation with directions. 43

4.4 Number of events from the event dataset 43

4.5 Belgium vs Brasil game predictions. 50

4.6 Animation layout. 51

4.7 Predictions with context segments. 52

Chapter 1

Introduction

The topic of this thesis is event detection in football, a complex and dynamic challenge that sits at the intersection of sports analytics and data science. This introductory chapter provides a broad overview of the sports analytics market, presenting its current trends, significant growth, and the technological advancements that have revolutionized this field. Additionally, the chapter covers the evolution of sports analytics from a research perspective, highlighting how advancements in methodologies and computational techniques have reshaped our understanding and analytical capabilities. Furthermore, this chapter will lay the groundwork for the main exploration of this thesis by outlining the specific research objectives and challenges associated with event detection in football. It will discuss the details of detecting and analyzing events through spatiotemporal data, the necessary steps to overcome existing limitations, and the potential methodologies that could effectively address these challenges. This serves as the foundation for the detailed investigations and innovative developments that will be explored throughout the subsequent chapters of this thesis.

1.1 Sports analytics: A market overview

Football's status as a leading global entertainment phenomenon is explicitly illustrated by its economic impact, as detailed in Deloitte's Football Money League report [1]. According to this report, the combined revenue of the world's top 20 football clubs reached a record high of €10.5 billion for the 2022/23 season, marking a significant 14% increase from the previous year. This substantial growth highlights the sport's huge fascination and its ability to generate notable economic activity even in challenging times. Deloitte's analysis points to optimized matchday strategies, innovative broadcasting rights agreements, and diverse commercial ventures as key drivers of this revenue spike. The popularity of football, together with its effective commercial strategies, not only highlights its role as a dominant player in the global entertainment market but also presents the growing role of advanced analytics in shaping business decisions and enhancing fan engagement. As clubs increasingly use data analytics for performance optimization and market expansion, football's interconnection with technological advancements such as AI and big data is set to develop even further.

The popularity of football has surpassed its traditional role as only a sport, influencing various sectors beyond the business domain and demonstrating significant impacts on social, technological, and educational fields. As highlighted by [2], football is claimed as "a perfect catalyst for social, technological, and educational development" reflecting its broad utility beyond just entertainment. In terms of technology, the sport has seen the integration of advanced systems like the Video Assistant Referee (VAR) and goal-line technology, which have transformed officiating, enhancing fairness and accuracy in gameplay. These innovations, discussed in [3], illustrate how technology is reshaping the rules and experiences of football. Moreover, the introduction of wearable technology and the development of smart stadiums, as mentioned by [4], further highlight the technological evolution within the sport. These advancements not only optimize player performance through detailed health and activity tracking but also enhance the fan experience by offering greater connectivity and enriched interactive engagement within stadium environments.

The impact of these technological innovations extends to players and coaches, providing them with novel tools for performance analysis and team management, as noted by [5]. This digital revolution in football also significantly enhances the fan experience through interactive apps and real-time engagement platforms that bring fans closer to the game they love. Research from [6] provides empirical support for these observations. Their findings indicate a strong belief among industry stakeholders that technology will play an increasingly critical role in football, viewed as development opportunity. While there is some caution about technology potentially replacing specialist skills, the overall sentiment is that it supports and enhances game-related processes. However, challenges remain, particularly in making these technologies accessible to players globally, suggesting a need for broader distribution and integration to truly use its potential.

These rapid trends, particularly in the use of technology and data within football, have significantly accelerated the growth of the sports analytics sector. According to a report by [7], the sports analytics market is experiencing strong expansion, projected to grow from USD 4.81 billion in 2024 to USD 32.31 billion by 2032. This rapid development indicates sector's increasing relevance and the crucial role analytics now plays in sports management and strategy.

Particularly within football, the analytics segment is positioned for remarkable growth, expected to register the highest compound annual growth rate (CAGR) of over 24% through 2030 [7]. Such growth is largely driven by football's immense popularity globally, with notable fanbases in European countries such as Germany, Spain, and the U.K. This widespread enthusiasm for the sport creates an ideal environment for advanced analytical tools and practices that enhance performance analysis and audience engagement.

One of the primary catalyst behind these developments is the advancements in machine learning (ML), artificial intelligence (AI), and big data. These technologies have not only transformed the sports analytics landscape but have also led to the emergence of various sub-branches within the field. From video analytics that offers detailed analyses of match footage to bioanalytics that provides insights into players' physiological data, and smart wearable

technology that tracks athletes' performance metrics in real-time, the scope of sports analytics has expanded dramatically. These advancements enable more precise assessments of player performance and game dynamics, driving the evolution of football into a more data-informed sport [8].

Liverpool Football Club has emerged as a leading proponent of sports analytics, utilizing advanced mathematical and data science techniques to refine their approach to football. According to a detailed analysis in [9], Liverpool's strategic integration of these technologies has been a crucial factor in their rise on the global stage. The club has adopted sophisticated predictive modeling and performance analytics, which have enhanced player recruitment, tactical planning, and in-game decisions. This data-driven strategy has not only optimized physical performance and injury prevention among players but also revolutionized scouting and player development. By embedding data science deeply into their operational framework, Liverpool has set a benchmark in football for how analytics can drive success both on and off the pitch, making them a standout example in the sports analytics field.

1.2 Evolution of sports analytics: A research perspective

In the paper [10] the impact of sports analytics across various sectors of the sports industry is carefully examined. It provides a foundational understanding of its broad applications. The review explains how sports analytics has significantly transformed not just player performance metrics but has also deeply influenced other important areas such as marketing, team management, betting, and match strategy. Through detailed analysis, the paper highlights how data-driven insights are used in marketing to better understand fan preferences and engagement, leading to more effectively targeted marketing strategies. Additionally, it explores the use of analytics in enhancing team management decisions, optimizing betting odds, and refining match strategies through predictive modeling and real-time data analysis. However, the primary goal of that section is to investigate data analytics in team sports to identify trend patterns in methodologies. This analysis is

expected to highlight some general characteristics of team sports analytics compared to those in endurance or individual disciplines.

Building upon the foundational insights into sports analytics, the application of predictive analytics has significantly expanded the scope of research into sports-related events. Initially rooted in basic econometric and statistical techniques, predictive analytics included methods such as multivariate regression and other foundational statistical approaches. These techniques, early in the development of sports analytics, were important for processing data to forecast outcomes and reveal patterns that would guide strategic decisions. This early use of predictive analytics set the stage for the deeper, more complex analyses that would follow in the field. For instance, in [11], authors employed statistical techniques to predict fans' perceptions of competitive balance, providing insights that are crucial for league organizers to maintain engagement and fairness. Similarly, [12] used Ordinary Least Squares regression to analyze data from professional basketball games to predict the likelihood of missed shots during free throws, offering valuable information for coaching and player development. In [13], authors employed Logistic Regression to assess player performance for team selection, optimizing team compositions based on predictive performance outcomes to maximize the effectiveness of team strategies during matches.

Afterwards, sports analytic's advancements have seen a significant shift towards the integration of machine learning models, which have enabled more nuanced and predictive insights into various aspects of sports. For example, [14] applied machine learning techniques to perform classifications, that helps in identifying player types or game situations from complex datasets. Additionally, [15] employed principal components analysis, a technique that reduces data dimensionality while preserving variance, to distill key factors that influence outcomes in sports contexts.

The integration of machine learning extends further into predictive models, as shown by the work of [16] in their study on football league outcomes and player performance predictions. Applying historical data and advanced statistical methods, their research forecasts team performances prior to the season's start. This approach demonstrates the practical application of basic

machine learning techniques like KNN or SVM to effectively predict long-term sports outcomes, offering valuable insights into team dynamics and player contributions.

Currently, the recent increase in the popularity of deep learning within sports analytics represents a significant movement towards more advanced and predictive analytical techniques. This trend is well documented in [17] which highlights a variety of innovative applications of deep learning that enhance both the analysis and prediction capabilities in sports contexts.

In [18], authors introduced a recently popular approach using Generative Adversarial Networks (GANs) to simulate defensive movements in response to offensive plays in team sports. By inputting the movement of the ball and offensive players, their model generates realistic defensive trajectories, enabling teams to forecast and strategize against potential reactions from opponents. In [19], authors advanced the field by applying Convolutional Neural Networks (CNNs) to transform monocular videos of soccer games into dynamic three-dimensional reconstructions. This development allows for a more detailed spatial analysis of player movements and game dynamics, offering new insights into the tactical aspects of soccer.

Further enhancing the utility of deep learning, [20] developed "SoccerMap," a deep learning architecture based on CNNs that estimates full probability surfaces for potential passes in soccer. Utilizing high-frequency spatial-temporal data, this innovative tool enhances decision-making during the game by providing detailed visualizations of passing opportunities. On the basketball court, [21] utilized Recurrent Neural Networks (RNNs) for sequence modeling to predict the success of three-point shots. This application of RNNs demonstrates the capability of deep learning to handle temporal dependencies and sequences in sports data, offering predictive insights into specific basketball events. Similarly, [22] explored the impact of off-ball events on possession success using a combination of Long Short-Term Memory Networks (LSTMs) and neural embeddings, analyzing subtle yet significant actions away from the ball, such as screens and cuts, to determine their contribution to strategic play outcomes. In [23], the authors employed Mixture Density Networks (MDNs) combined with team lineup encoding to predict the distribution of

final score differences in NBA games. Their model addresses contextual uncertainties within win probability models, offering a nuanced approach to forecasting game outcomes based on the current state of play.

Expanding on complex player interactions, in [24] the authors applied hierarchical Variational Recurrent Neural Networks (VRNNs) with a shared macro intents architecture to model complex interactions among offensive basketball players, generating realistic multi-agent trajectories over extended periods. This work provides deep insights into team dynamics and player interactions during games. Innovation in player movement simulation and team strategy analysis is further exemplified by [25] and [26], who have pushed the boundaries with advanced imitation learning techniques and ghosting models. In [25], authors introduced a coordinated multi-agent imitation learning algorithm that progressively learns from sequenced player actions, enhancing the simulation of real-game scenarios. Meanwhile, in [26], the authors developed a model that learns to mimic the movements of multiple players, enhancing the understanding of player roles and team strategies by simulating various in-game situations.

The main conclusion from that section is that the current shift towards deep learning methods has enabled addressing more complex problems in team sports. It was also observed that valuable insights in team sports rely on the phases of the game. Consequently, a tool for game event detection was recognized as very useful for further analytical approaches. Ultimately, the decision was made to focus this thesis on developing such a tool specifically for detecting football events.

1.3 Framework for event detection

As it was mentioned in the previous section the aim of that thesis is to develop tool for football events detection. To construct it, it's essential to establish a solid framework. Drawing from insights of [27] this work identifies critical elements necessary for a advanced event detection system as illustrated by a Figure 1.1.

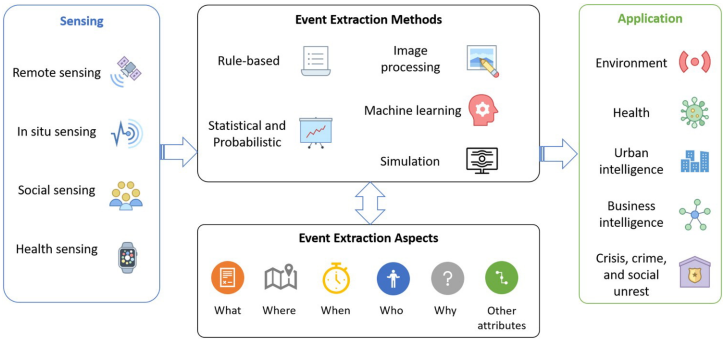

Figure 1.1: *Spatiotemporal event detection workflow. It presents the components of event detection systems. The sensing component aims to provide the necessary data. The feature extraction methods aim to identify and interpret significant events. However, the application component enhances various practical domains. Figure taken from the work of [27].*

The first component is sensing which includes various data collection methods such as remote sensing, in situ sensing, and social sensing. These methods gather diverse and rich data sets, providing the foundational layer for event analysis. From a sports analytics perspective, this is a particularly important aspect due to the multivariate nature of the data. Typically, four different data types can be distinguished. The first type is events data, which consists of sequences of events occurring during games [28]. This data is extremely important for event detection tasks as it allows for obtaining labels. The second type is video data, usually captured from broadcast videos [19]. The third type is positional data, which can be obtained from wearable technology [8]. Lastly, audio data from commentators is also very important, as it can be used as an additional feature for event detection [29].

Moving to event extraction methods, the framework includes a range of techniques from rule-based systems [30] and image processing to advanced machine learning models [20] and simulations [25]. These methodologies are applied to the sensed data to identify and interpret significant events within a football match, such as scoring actions, fouls, or strategic moves.

The final component, application, uses the events extracted to enhance various practical domains including health, urban intelligence, and in our case, sports performance analysis. Here, the insights derived from event detection are applied to optimize team strategies [13], enhance player performance,

and boost fan engagement [11]. Each element of this framework is interconnected, ensuring a fluid transition from data collection to practical application, thereby enabling a deeper and more effective exploration of event detection in football.

In this thesis we are mostly focused on extraction methods. Particularly, on machine learning models that can analyze and interpret the vast amounts of data generated during matches. Such models must be capable of recognizing subtle patterns and interactions that are not immediately obvious, enabling them to predict events before they happen. The most challenging task for these models is the analysis of the spatial and temporal dynamics. Football is a highly contextual sport where the relevance and impact of events can differ significantly based on their occurrence within the spatial layout and timing of the game. Understanding how spatial configurations and temporal sequences influence game outcomes requires advanced analytical models that can parse and interpret complex spatiotemporal data. These models must account for the fluid nature of the game, where the interactions between players and the evolving strategies significantly impact the dynamics observed on the field.

1.4 Research objectives and challenges

The implementation of football analytics, particularly in the areas of event detection, faces several challenges that impact the effectiveness of data-driven strategies. Addressing these challenges is essential for the advancement of analytical methods and their application in real-world scenarios. One of them arises from the collection and preprocessing of positional data. Ensuring accuracy, completeness, and consistency of the data is necessary. Challenges include data inaccuracies due to limitations of tracking technologies and environmental interferences that introduce noise and gaps in data capture. Moreover, the completeness of data can be compromised by technical issues during matches. Effective preprocessing must address these issues through precise filtering and interpolation techniques to maintain the integrity of data. Additionally, synchronization of positional data with specific game events is crucial for contextual analysis, necessitating precise alignment of timestamps

and player positions. Ensuring standardized data formats across various sources and matches is also critical for enabling consistent and comparative analysis across datasets.

Moreover, distinguishing between event detection and event spotting presents unique challenges within sports analytics, particularly due to the uncertainty in defining the start and end of events [31]. Event detection typically involves identifying the boundaries of actions within a game. This traditional approach suits scenarios where events have a clear duration and distinct endpoints, which is feasible for many sports actions. However, the task of defining these boundaries can become problematic for certain types of events. For instance, some events in football, like a goal may occur in a brief, indistinct moment, making it difficult to specify their exact starting and ending points. Similarly, continuous or overlapping events pose additional challenges, as distinguishing between concurrent actions becomes complex. Given these difficulties, some researchers advocate for a different approach known as event spotting. This method focuses on pinpointing the occurrence of an event at a specific moment, especially when the boundaries are uncertain or the event duration is extremely short. Event spotting simplifies the localization of such events by identifying a key moment within an acceptable error tolerance window around the event's occurrence, avoiding the need to define precise start and end points. This approach is particularly advantageous in football, where many significant events happen instantaneously and require rapid recognition. Adopting event spotting facilitates the use of standard evaluation metrics like mean average precision, enhancing the analysis and understanding of complex event dynamics in sports.

Additionally, developing strong features that effectively capture the essential aspects of football dynamics and designing suitable model architectures that can process and analyze these features present significant technical challenges. Feature generation must focus on identifying elements that accurately reflect the strategic and tactical aspects of the game, requiring deep domain knowledge and complicated data engineering techniques. Similarly, model architectures must be capable of handling the high-dimensional and often non-linear relationships natural in football data. This involves selecting or designing advanced models.

In the context of football analytics, the spatiotemporal event detection approach is critical due to its dynamic nature. Based on our conclusions derived from the event detection workflow presented in [27] and all mentioned challenges, we were able to establish two objectives for our research.

Objectives:

1. **Develop Advanced Detection Algorithms:** Apply and adapt spatiotemporal event detection algorithms to identify and classify key events in football, such as passes or tackles.
2. **Validate Detection Models:** Test and validate the effectiveness of the detection models in real-world scenarios, ensuring they are both accurate and reliable for practical use.

Ultimately, these objectives aim to guide the research towards developing a detection tool that addresses challenges present in the dynamic domain of football. Although the thesis will focus on testing the theoretical aspects, it aims to establish a strong foundation for future practical applications.

Chapter 2

Background and related work

In this section, we explore the essential concepts and previous studies that form the foundation of the methodologies used in this thesis, setting the context within the evolving field of sports analytics. This examination is important for understanding the current state of event detection technologies, the advancements in spatial and temporal analyses, and how these elements integrate to enhance our understanding and capabilities within sports contexts.

We begin by reviewing the landscape of event detection in sports. This subsection investigates the diverse methodologies that have been developed, the challenges that have arisen, and the innovative solutions that researchers have designed over the years. This narrative not only traces the evolution of event detection but also prepares for discussing the newer, more advanced approaches that are the focus of this thesis.

Next, we turn our attention to the spatial analysis of sports data, specifically through the perspective of Graph Neural Networks (GNNs) and the NetVLAD layer. The discussion on GNNs will illustrate how these networks can use the complex relational data within sports, effectively capturing the dynamic interactions between players and their strategic formations. Following this, we explore the NetVLAD layer, originally adapted from computer vision, which has proven essential in enhancing the neural networks' ability to manage and interpret large sets of spatial data, thereby strengthening the

models' accuracy and robustness.

Finally, we address the temporal dimension of sports analytics through the original use of Context Aware Loss. This segment highlights how integrating contextual information into model training processes significantly improves the handling of time-dependent data, crucial for predicting and analyzing events that evolve over the course of a game.

Together, these discussions form a coherent narrative that not only presents a thorough review of the field but also clearly defines the technological and methodological advancements that this thesis contributes to sports analytics. This background sets a solid foundation for appreciating the novel approaches developed in this research, highlighting their significance in pushing the boundaries of what can be achieved in sports event detection and analysis.

2.1 Related work for event detection

In exploring the landscape of related work for event detection in football, it is important to acknowledge the foundational efforts that have shaped current methodologies. One of the earliest approaches to this challenge was introduced in the paper [32], which proposed a framework that allows for real-time event detection using cinematic features. They refer to elements such as shot boundaries or camera motions derived from common video composition and production rules. This framework also incorporates filtering of slow-motion replay shots by object-based features for semantic labeling. All these together paved the way for automated analysis in sports broadcasting.

Progressing in complexity, the authors of [33] further developed the field by implementing an action recognition system based on the Bag of Words (BoW) approach. Video frames were encoded using a BoW model, where each frame was represented as a histogram of visual words, and these histograms were used to create a sequence of "phrases" that were then classified using SVM with a string kernel for prediction. The evolution of these techniques continued with the integration of neural networks, by incorporating

Long Short-Term Memory (LSTM) cells [34].

More recently, the development of deep learning models has introduced even more sophisticated analytical capabilities. Transformer architectures and RNNs have been used to predict events based only on past occurrences, focusing on the temporal behavior of the game without the need for traditional video analysis. [35]

Building on the foundation laid by earlier research, the adoption of CNNs marked a significant advancement in the field of event detection in football. CNNs have been especially influential in improving how features are extracted from broadcast videos. For example, in a 2017 study, authors demonstrated this progression by applying pre-trained CNNs to extract both spatial and temoporal features effectively.[33] The study further investigated the use of Autoencoders to fuse these different information streams, specifically focusing on the detection of goals. This approach demonstrates the ability of CNNs to enhance the precision of detecting specific types of events by analyzing a combination of movement and positioning data within game footage. Another significant development was presented in [36], where broadcasting videos were processed using a combination of pre-trained CNNs and RNNs. This hybrid model segments and classifies video content, capturing an integrated view of spatial and temporal dynamics, thus providing a more comprehensive framework for event detection.

Further innovations in CNN application were explored in [37], which introduced a two-stream deep convolutional descriptor (TDD). This novel architecture separates the convolutional networks into spatial and temporal nets, each designed to process respective types of data, enhancing the accuracy and efficiency of event recognition in complex game scenarios. The approach was improved further in [38], which employs a Two-Stream Convolutional Neural Network alongside a Dilated Recurrent Neural Network. This configuration allows for detailed processing of both short-range temporal information and mid to long-range temporal correlations between frames, offering a nuanced understanding of event dynamics over time.

Lastly, the studies [39] and [40] pushed the boundaries by integrating 3D CNNs to analyze volumetric data from videos. This method enhances the

detection capabilities by considering the depth aspect, crucial for interpreting more complex events accurately and in real-time.

As positional data became more accessible and detailed, there has been a notable shift in football analytics towards using this data to enhance event detection and strategic insights. This transition reflects the evolving capabilities of data capture technologies and the growing sophistication of analytical models aimed to exploit this rich data source.

This trend is also visible in [41] where authors aggregated temporal information within features derived from positional data and used XgBoost to detect counterpressing tactics effectively. This approach not only enhances the understanding of team dynamics during pressing sequences but also highlights the value of integrating temporal dynamics with spatial positioning. Further advancements were presented in [42], which introduces a possession algorithm that analyzes each frame to determine which player has ball possession and whether the ball is in play. This continuous assessment of possession and play status, based on positional data, allows for precise determination of in-game events following the laws of football, offering a detailed view of game flow and player interactions. Another significant contribution is found in [30] where authors used positional data to reason about complex events, such as determining the outcomes of specific passes or crosses that lead to goals. By formalizing these events within a structured framework based on principled Interval Temporal Logic (ITL), the study demonstrates how positional data can be deeply analyzed to understand causative relationships and event sequences in football.

The next step in the analysis of positional data in sports analytics was the integration of GNNs, which brought about a significant advancement in the depth and precision of understanding football gameplay. The significant advantage provided by GNNs is their ability to capture complex interactions and relationships between players. A great example of usage GNNs is detailed in [43] where authors used them to extract features directly from positional data, enabling the analysis of defensive performance in football. By employing Graph Convolutional Networks (GCNs), the study not only identifies the positions and movements of players but also how these elements

associate during defensive plays. The approach allows for the creation of three distinct statistical measures that provide insights into the effectiveness of defensive strategies, showcasing the power of GNNs in making complex tactical aspects of the game more predictable and analyzable. Further development of GNNs usage was by introducing a Context Aware Loss Function that enhances the GCN's ability to incorporate temporal information into the feature extraction process. [44] That approach was quite appealing therefore it was decided to use it in our framework. The details about it are provided in the next section. But the general idea is that by integrating this specialized loss function, the model gains sensitivity to the timing and sequence of player actions, which is important for understanding and predicting game events accurately. This method uses the natural strengths of GCNs in handling spatial relationships while enhancing their capacity to process temporal dynamics, providing a more comprehensive analytical framework.

In the evolving field of sports analytics, particularly in football, researchers have increasingly tried to enhance model performance by combining different techniques and feature representations. This multifaceted approach is evident in several recent studies that integrate various methodologies to increase the accuracy and depth of event detection and analysis.

Different pooling methods are explored in [45], including the introduction of NetVLAD, a technique that enhances the classification capabilities of video clip analysis by efficiently aggregating features over entire sets of data. Further integration of multimodal features is seen in [46], which combines visual and audio features to improve the detection and classification of events. The combination of audio and video streams is also emphasized in [29], demonstrating how dual-stream data can enhance the spotting of nuanced soccer actions. Additionally, techniques like Reinforcement Learning are applied in [47] to dynamically adjust models based on ongoing gameplay, further refining the accuracy and responsiveness of analytical models.

The review of the existing literature on event detection in football reveals notable gaps that may delay the broader application of these technologies. Primarily, many studies focus on predicting events from the perspective of isolated clips. While this approach is suitable for analyzing events that occur

infrequently, it falls short when addressing mixed or continuous events that require a more holistic view of the game dynamics. Furthermore, there is a limitation in the application of GNNs within the field. Current implementations of GNNs in these studies are often restricted by the number of features they can process effectively, and there is a noticeable lack of adoption of more advanced GNN architectures that might capture a richer set of relationships and interactions within the data. This limitation highlights the need for further development in GNN methodologies, which could expand the capacity of analytical models to handle complex, multi-faceted sports events more effectively. Addressing these gaps could significantly enhance the predictive accuracy and applicability of event detection systems in football, leading to more nuanced and comprehensive game analysis.

2.2 Spatial analysis: Graph Neural Networks

GNNs, with their unique ability to process data represented as graphs, are employed to model and analyze the complex and dynamic spatial interactions that occur between players and play elements. By mapping player positions and movements to graph structures, GNNs enable a deeper understanding of game patterns, providing insights that are not readily apparent through traditional analysis methods.

2.2.1 Message Passing Neural Networks

Message Passing Neural Networks (MPNNs) [48] represent a base framework for dealing with graph-based data, which effectively uses node features and the graph structure to perform various types of predictions and classifications. In MPNNs, the core operation is message passing, where each node sends and receives messages to and from its neighbors, allowing information to be aggregated across the graph.

The general process in an MPNN consists of two phases: the message passing

phase and the readout phase. In the message passing phase, nodes update their states by aggregating messages from their neighbors. These messages are computed based on the features of the nodes themselves and their neighboring nodes, summarised in the formula:

$$m_v^{(t+1)} = \sum_{w \in N(v)} M_t(h_v^{(t)}, h_w^{(t)}, e_{vw})$$

Here, $m_v^{(t+1)}$ represents the aggregated message received by node v at time step $t+1$, $N(v)$ denotes the neighbors of node v, M_t is the message function at time step t, $h_v^{(t)}$ and $h_w^{(t)}$ are the hidden states of nodes v and w respectively at time t, and e_{vw} is the edge feature between nodes v and w. [48].

Following the aggregation of messages, the hidden state of each node is updated through a vertex update function U_t:

$$h_v^{(t+1)} = U_t(h_v^{(t)}, m_v^{(t+1)})$$

In this formula, $h_v^{(t+1)}$ is the updated state of node v at time $t+1$, incorporating the aggregated information from its neighbors. [48]

The readout phase occurs after several iterations of the message passing phase, during which the network computes a global graph descriptor or properties specific to nodes or edges. This is typically done by some form of pooling or aggregation across node states, ensuring the output is invariant to node permutations, preserving the graph's isomorphic properties, as discussed in [48].

Overall, MPNNs are adept at capturing both local connectivity and global structural information, making them suitable for tasks where the interaction between components (such as atoms in molecules or users in social networks) plays a critical role in the system's behavior. They also serve as foundational elements for other more advanced applications.

2.2.2 Graph Convolutional Network

The Graph Convolutional Network (GCN) [49] represents an advanced development of MPNNs, fitted for effective handling of graph-structured data. The primary advancement brought about by GCN is its ability to perform semi-supervised learning on nodes with only a subset of labels available, which is typical in real-world applications.

The core operation of GCN, the layer-wise propagation rule, can be understood as a first-order approximation of spectral graph convolutions. This is significant because it simplifies the computational demands of graph convolutions, traditionally a costly operation due to the need to calculate eigendecompositions of large matrices. By using this approximation, the computational complexity is reduced to a linear function of the number of edges, which greatly enhances scalability and efficiency. Here's how the basic convolution operation is defined in a GCN:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

where $\tilde{A} = A + I_N$ is the adjacency matrix of the graph G with added self-connections, $\tilde{D}$ is the degree matrix corresponding to $\tilde{A}$, σ denotes an activation function, and $W^{(l)}$ is a layer-specific trainable weight matrix. [49]

By integrating such a model, one can effectively encode both the graph structure and the node features into the learning process, enabling more nuanced insights and predictions. This method has shown superior performance compared to other semi-supervised techniques in numerous studies, not only enhancing the accuracy but also the interpretability of the results within a football analysis context.[49]

2.2.3 Graph Attention Network

Graph Attention Networks (GATs) [50] mark a significant evolution in handling graph-structured data, crucial for domains like football analytics where relational data is key. GATs uniquely use an attention mechanism that allows

nodes to focus on their most relevant neighbors, effectively learning to weigh the influence of each connection dynamically. This capacity to assign variable importance to different nodes enhances the model's adaptability and depth, making it particularly suited for football, where the significance of interactions can shift drastically across different plays.

The core mechanism of GATs involves computing attention coefficients that determine the influence of each node's features on another, which is important when analyzing player interactions in a football match. This approach is defined mathematically as:

$$h_i^{(t+1)} = \sigma\left(\sum_{j \in N(i) \cup \{i\}} \alpha_{i,j} W^{(t+1)} h_j^{(t)}\right),$$

$$\alpha_{ij} = \text{softmax}_j\left(\text{LeakyReLU}\left(a^T [Wh_i \parallel Wh_j]\right)\right),$$

where h_i is the feature vector of node i, W is a weight matrix, a is a weight vector of the attention mechanism, and $\parallel$ denotes concatenation. The softmax function is applied over all nodes j that are neighbors of i, ensuring the coefficients are normalized, making them easily comparable across different nodes. This formulation allows the GAT to handle varying sized neighborhoods, which is common in football where player influence can vary throughout the game. [50]

2.2.4 Graph Isomorphism Network

Graph Isomorphism Networks (GINs) [51] represent an advancement in graph neural network architectures, specifically addressing the limitations of previous models such as GCNs and GATs. GINs are designed to be as powerful as the Weisfeiler-Lehman graph isomorphism test, which is a classic algorithm for testing graph isomorphism capable of distinguishing a broad class of graphs.

The unique feature of GINs is their use of a more expressive function for aggregating node information. Where GCNs may use mean aggregators and GATs may use attention mechanisms, GINs employ a sum aggregator which enables the network to capture more of the graph's structure. The sum

aggregator is crucial because it preserves information about the count of different types of nodes in a node's neighborhood, which is information that mean or max-pooling would lose. [51]

The architecture of a GIN can be described with the following update rule:

$$h_v^{t+1} = MLP^{t+1}\left((1+\epsilon^{t+1}) \cdot h_v^t + \sum_{u \in N(v)} h_u^t\right)$$

In this formula, h_v^t denotes the new feature vector of node v, MLP stands for a multi-layer perceptron which is used to approximate complex functions, ϵ is a learnable parameter that adjusts the importance of a node's own features relative to its neighbors' features, and $N(v)$ signifies the set of neighbors of v. [51]

This structure allows GINs to distinguish not only between different graph structures but also to capture the similarity of graph structures, providing a powerful tool for both classification and regression tasks on graph-structured data. This is particularly relevant in the domain of football analytics, where discerning and exploiting complex patterns can provide a competitive edge in player and game analysis.

2.3 Spatial analysis: NetVLAD

This subsection focuses on the realm of feature pooling within neural network architectures, an important aspect when addressing tasks that require the consolidation of spatial information from extensive datasets. This approach has seen significant application in the field of sports analytics, particularly in the automated understanding and recognition of complex events in football videos.

The NetVLAD (Vector of Locally Aggregated Descriptors) pooling layer has emerged as a transformative approach in the realm of computer vision, particularly for tasks that benefit from robust and discriminative global descriptors derived from local features. Originating from the conventional VLAD

method, NetVLAD extends this concept into a differentiable layer that is compatible with end-to-end training within convolutional neural networks. This allows it to learn more effective image and video representations by aggregating local deep features into a compact global descriptor.

NetVLAD was originally designed to enhance place recognition tasks by producing a spatially invariant descriptor capable of handling variations in viewpoints, scales, and lighting conditions, which are common challenges in image-based localization. The architecture of NetVLAD includes a clustering component analogous to a learnable "codebook", where each local feature extracted from the CNN is assigned to a cluster, and the differences (or residuals) between features and their corresponding cluster centers are accumulated to form the final global descriptor. This methodology is detailed comprehensively in [52] which introduces the concept and its applications in detail.

In the context of sports analytics and event detection, the ability of NetVLAD to capture essential spatial features makes it a valuable tool for analyzing sequences of sports actions where context and temporal dynamics are important. For example, in [45], NetVLAD has been employed to enhance the performance of action spotting models by providing a strong way to pool spatial features across entire video frames, which aids in the accurate detection and classification of specific sports events like goals, penalties, or substitutions. This application showcases how advanced pooling techniques like NetVLAD can significantly impact the development of analytical tools in dynamic and complex environments such as football matches.

2.4 Temporal analysis: Context Aware Loss Function

This subsection introduces the Context-Aware Loss Function (CALF), applied in conjunction with 1D Convolutional Neural Networks (1D CNNs) for temporal analysis in football event detection. Given the complexity and richness of data in sequences of player actions and game dynamics, 1D CNNs are

uniquely suited for analyzing the timing and order of events within a match due to their efficiency in handling temporal data.

Building upon the capabilities of 1D CNNs, the CALF [44] is specifically designed to meet the demands of temporal event detection. This advanced approach not only aims for the accurate localization of event timings relative to video frames but also strategically adjusts the network's training focus. By employing event slicing, the model prioritizes actions based on their closeness to key events: actions that occur far from an event have their influence reduced to avoid weakening the model's predictive accuracy. Meanwhile, actions immediately preceding an event are treated with neutrality, acknowledging that they may not invariably result in significant outcomes. Most crucially, actions following an event are given enhanced attention due to their clear contextual relevance, enabling the model to focus on moments of high predictive value. This methodical integration of 1D CNNs with the context-aware loss framework facilitates a more nuanced understanding and improved precision in detecting and analyzing temporal events in football.

The $L(p, s)$ function, integral to this loss framework, dynamically modifies the loss computation based on the temporal position s of a predicted event. This is characterized by:

$$L(p,s) = \begin{cases} -\ln(1-p) & \text{if } s \leq K_1^c \\ -\ln\left(1 - \frac{K_2^c - s}{K_2^c - K_1^c} p\right) & \text{if } K_1^c < s \leq K_2^c \\ 0 & \text{if } K_2^c < s < 0 \\ -\ln\left(\frac{s}{K_3^c} + (1 - \frac{K_3^c - s}{K_3^c})p\right) & \text{if } 0 \leq s < K_3^c \\ -\ln\left(1 - \frac{s - K_3^c}{K_4^c - K_3^c} p\right) & \text{if } K_3^c \leq s < K_4^c \\ -\ln(1-p) & \text{if } s \geq K_4^c \end{cases}$$

- **Before and After Event:** For predictions that occur too early or too late relative to an event ($s \leq K_1^c$ or $s \geq K_4^c$), the loss is increased to discourage the model from making predictions that are far from important event times, thus promoting more accurate timing in predictions.

- **Critical Zone Near Event:** Within a narrowly defined time window just after the event ($0 \leq s < K_3^c$), the loss function rewards highly

accurate predictions that closely match the event's timing, promoting precision.

- **Neutral Zone Just Before Event:** Immediately before the event, where $K_2^c < s < 0$, the loss is set to zero. This neutral stance reflects the uncertain nature of potential build-ups to events, acknowledging that while these moments are crucial, they do not guarantee a significant outcome.

The modified loss $\tilde{L}(p, s)$ incorporates a logarithmic term based on predefined maximum and minimum thresholds (τ_{max} and τ_{min}), which serves to normalize the impact of varying confidence levels in predictions across different frames:

$$\tilde{L}(p,s) = \begin{cases} \max(0, L(p,s) + \ln(\tau_{max})) & \text{if } 0 \leq s < K_3^c \\ \max(0, L(p,s) + \ln(1 - \tau_{min})) & \text{otherwise} \end{cases}$$

Finally, the final loss L aggregates these individual losses across all frames x_i and across all classes c, providing a comprehensive measure of the model's performance across the entire video clip:

$$L = \frac{1}{CN_F} \sum_{i=1}^{N_F} \sum_{c=1}^{C} \tilde{L}(p^c(x_i), s^c(x_i))$$

The behaviour of the loss function is presented in the Figure 2.1. This nuanced approach to calculating loss ensures that models trained for event detection in football videos are not only accurate in detecting events but also precise in pinpointing their temporal occurrence, consequently enhancing the analytical capabilities for strategic planning and performance analysis in sports.

2.5 Final thoughts

The analysis of the sports analytics market and research domain has confirmed both the demand and the increasing popularity of sports analytics.

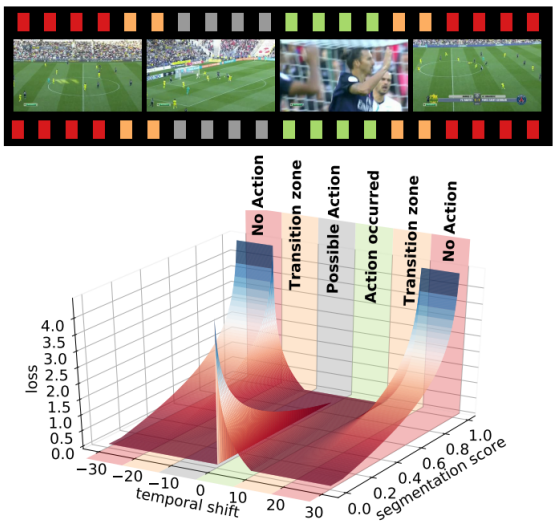

Figure 2.1: *The Figure represents the behaviour of the Context Aware Loss Function (CALF). The function heavily penalizes the frames far distant from the action and decrease the penalty for those gradually closer. It does not penalize the frames just before the action to avoid providing misleading information as its occurrence is uncertain, but it heavily penalizes those just after, as the action has occurred. Figure taken from the work of [44].*

This growing interest highlights the importance of developing a football event detection tool. Our research has identified the creation of advanced models that accurately capture spatiotemporal information as a key and challenging task in achieving this goal. Additionally, it is necessary to provide reliable validation tools that can contextualize the development of these models. Both objectives are essential and drive our research in the right direction.

The literature review has demonstrated the variety of approaches tested in the field and highlighted the rising popularity of deep learning methods. We have identified a slight gap in the analysis of event detection tools based on positional data. With the development of GNNs, there is notable potential in using positional data to detect football events. Different GNN architectures have shown a good ability to capture spatial information. However, to effectively capture temporal information, events are segmented and captured by the Context Aware Loss Function.

All these observations have led to the creation of research questions that establish clear quantitative goals:

RQ1: How do temporal dynamics influence the accuracy of event predictions in football, and what strategies can be implemented with advanced GNN architectures to effectively capture and utilize these dynamics for improved predictive performance?

RQ2: How do different GNN architectures, such as Graph Convolutional Networks, Graph Attention Networks, and Graph Isomorphism Networks, perform in predicting football events?

RQ3: How does the combination of the NetVLAD pooling method and Context Aware Loss function enhance the performance of Graph Neural Networks in accurately predicting football events?

Chapter 3

Methodology

In this chapter, the thesis will cover the thorough methodology used to develop predictive models for football event detection. The process outlined involves several key phases: data collection and preparation, model development, and validation. Each phase is critical to ensure the accuracy and effectiveness of the models in real-world applications.

Initially, the methodology section will describe the dual-source data collection approach, integrating positional data from the Royal Belgian Football Association (RBFA) with event annotations from SoccerNet [53]. This integration requires precise synchronization to ensure the integrity and usability of the dataset for subsequent analyses.

Following data preparation, the thesis will cover the details about two modules, segmentation and spotting. Each module is tasked for different purposes deeply explained in their architecture subsections. Additionally, for each of them we used different evaluation methodologies which are also deeply covered in that part.

Finally, the thesis will discuss the training settings used to obtain the models. This includes different approaches to handling data imbalance, enhancing model performance, and optimizing hyperparameters to achieve the best possible results. The necessary hardware used for training the models is also presented.

3.1 Data collection and preparation

The thesis uses a multi-source dataset including positional data collected from the Royal Belgian Football Association (RBFA) and event annotations from SoccerNet [53]. This datasets gather data from 12 Belgian games during the World Cup 2018, Euro Cup 2020, and World Cup 2022. The integration of datasets was important for constructing a detailed representation of match events. The positional data was captured at a rate of 5 frames per second (fps), providing a detailed view of player movements and ball positioning throughout the game. However, the event data included only timestamps of the events. Synchronizing the positional data with the event annotations presented significant challenges due to differing data structures and time codes. The timestamps from SoccerNet marks only the initiation of events, which required an expanded approach to data marking. To reduce timing errors and enhance dataset alignment, the study expanded the temporal window for each event to include two frames before and after the identified start time. Ultimately, data setting forced the model to predict to detect beginning of the events.

The SoccerNet event dataset includes a diverse array of annotated events, each presenting unique analytical challenges which are summarised in the Table 3.1.

These event types exhibit natural imbalances, reflecting the varied nature of a football game's flow and the different frequencies of event occurrences. It adds another layer of complexity to the predictive modeling. In addition to the event annotations from SoccerNet, the ball is out of play event was generated from positional data. Unlike the other event types that are marked only at the start of an occurrence, the 'dead ball' event is annotated for its entire duration which also might confuse the model.

Besides these annotations, a rich set of features was extracted from the positional data to assist in the analysis. These features include x and y positions of players, their distance to the ball, speed, directionality (x-direction, y-direction, and movement direction in radians), order of average position, team affiliation, red card flag, and both average velocity and acceleration

Event	Challenges
Pressure	Difficult to quantify effectively because not all pressure results in tangible outcomes like possession loss. Analyzing pressure involves assessing player positioning and the immediate impact on gameplay, which requires advanced tracking and context interpretation.
Foul Committed	It is challenging because fouls result from complex player interactions, often influenced by aggressive defensive tactics or errors in judgment.
Ball Recovery	Predicting ball recovery demands insights into post-possession dynamics, including the anticipation of where the ball will end up after tackles, deflections, or incomplete passes.
Duel	Outcomes are highly dependent on player attributes and situation, making prediction complex. Understanding the nuances of each duel requires detailed player data and situational context to forecast outcomes accurately.
Shot	Shots are influenced by opportunities created through complex team movements and individual player decisions, requiring models to evaluate potential shooting lanes and player tendencies under varying degrees of pressure.
Dribble	There might be troubles to capture the nuanced movements and techniques, from positional data, that characterize different players' dribbling styles, leading to poor representation and prediction accuracy.
Clearance	It involves understanding defensive strategies and the immediate pressure faced by players, as clearances typically occur under threat close to or within the goal area.
Goalkeeper Actions	It is challenging due to the high stakes and varied nature of these actions, which include saves, distributions, and positioning decisions.
Pass	Predicting passes in football is challenging due to their high variability and the complexity of interactions between players, which are influenced by tactical contexts and the dynamic state of play.

Table 3.1: *The events available in the SoccerNet [53] dataset and there challenges in automatic identification.*

over time. This detailed feature engineering supports deeper insights into player behaviors and game dynamics.

Advantages of Extracted Features:

- **Spatial and temporal positioning (x and y positions)**: Enables the analysis of player formations and movements across the field.
- **Distance to the ball**: Provides insights into player engagement and their closeness to key play actions, useful for assessing defensive and offensive strategies.
- **Speed and acceleration**: Offers a quantitative measure of player dynamics, which is critical for providing useful context.
- **Directionality**: Helps in understanding the strategic directions players are taking, which is crucial for predicting future movements and potential play developments.
- **Order of average position and team affiliation**: Useful for analyzing team structures and player roles within different tactical setups.
- **Red card flag**: Indicates significant shifts in team dynamics.
- **Average velocity and acceleration**: Aggregated metrics that give a broader view of overall team and player performance trends over time.

For the application of GNNs in later modeling phases, it was crucial to define the edges in the player interaction graphs with a specific cut-off value. This strategic limitation prevented the GNNs from processing excessive information, ensuring computational efficiency and focusing analysis on the most critical player interactions. This balance is a key to extract actionable insights from the complex web of data points captured during football matches.

The last step of the data preparation was the data augmentation. It played a very important role in enriching the training dataset, particularly due to the limited size of available data. We decided to mirror the position of the players on the field horizontally, vertically, and in both directions. Effectively

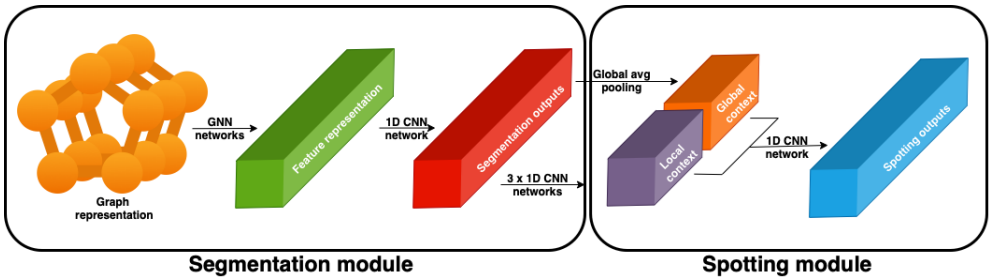

Figure 3.1: *The pipeline of the models used in the process. Initially, it receives a graph representation of frames from the clip. These frames are processed through GNN layers to extract relevant features. Subsequently, a 1D CNN layer is applied, and the outputs are segmented based on the frame's distance from the event occurrence. The CALF is then used for training. The spotting module processes the segmented outputs separately through 1D CNNs and global average pooling to capture both local and global contexts, which are then concatenated. Finally, a 1D CNN layer is applied to obtain the final results.*

we increased the number of our dataset four times, providing the models with varied perspectives of the same play, which helps in enhancing the robustness and accuracy of the models.

3.2 Segmentation module

This thesis introduces a dual-module framework, consisting of segmentation and spotting components, detailed in the diagram presented in the Figure 3.1. The primary motivation for segregating the analysis into two distinct stages is to effectively use the segmentation module's capabilities to generate rich and contextualized data. This data then serves as a foundation for the spotting module, which analyzes this information to detect and classify specific events within the dynamic environment of football games. This structured approach allows for a more targeted and detailed analysis, increasing the accuracy and relevance of event detection.

3.2.1 Architecture

The segmentation module is designed to capture the essential context surrounding events, as most events are directly associated with other behaviors on the field. For instance, when a goal is scored, players often gather to celebrate, and such contextual details are important for a detailed understanding of the events. To achieve this, the segmentation module generates features from the clip, chronologically and then analyzes them along the time dimension.

We decided to use GNNs and one-dimensional 1D CNNs for this task. The module receives a graph representation for each frame from the analyzed clip. These graph representations are then processed through a 4-layer GNN module. To explore the capabilities of different GNN architectures, we created separate models utilizing GCNs, GATs, and GINs as feature extractors.

Each of these GNN architectures offers unique advantages. GCNs are currently considered state-of-the-art due to their efficiency and strong performance in various applications. GATs, on the other hand, are expected to focus on significant neighbors, adapting to the dynamic contexts of a game by emphasizing the roles and interactions that matter most in specific situations. GINs are anticipated to distinguish more subtle graph structures, offering a different perspective on the data.

After the graph representations are processed through the GNN layers, a global descriptor is applied to provide an invariant representation of the graph. We chose a global mean pooling technique over global max pooling to capture more information from the graph structures. The resultant stack of feature representations from the clip is then processed through a 1D CNN with a kernel size of 5. This configuration is designed to capture the context around each frame, allowing the model to recognize patterns across time, which ultimately aids in understanding how events evolve sequentially.

Additionally, we explored an alternative approach to enhance the performance of the segmentation module by integrating a NetVLAD layer, a popular tool in other applications. Typically, NetVLAD aggregates descriptors across an entire video clip. However, in our case, we retained the descriptors

for each frame and then applied the 1D CNN. By incorporating the NetVLAD layer after the GNN layers, followed by the 1D CNN, we aimed to use its capability to capture detailed descriptors while preserving the temporal context provided by the frame-wise analysis.

3.2.2 Evaluation

A very important part of this research is the proper validation of the models. Based on proper evaluation and analytical tools, we are able to understand how the model operates and performs. Therefore, before using segmentation outputs by the spotting module, it was necessary to check if the segmentation model effectively captures the context around events. To do so, we established two metrics: the Predictive Overlap Ratio (POR) and the Actual Coverage Ratio (ACR).

To understand the intuition behind these metrics, we need to perform an analysis on the impact of the Context Aware Loss Function (CALF). Mathematically, if we want to optimize the loss value for one event across a 60-second clip, we would obtain outputs as presented in Figure 3.2. Based on it, we can distinguish that the CALF aims to decrease the segmentation outputs to zero for the frames located far away from the event. The outputs of the frames immediately before the event have a random value for reasons justified in the previous chapter. On the other hand, the segmentation outputs for frames immediately after events aim to reach 1.

However, in practice, due to the backpropagation method during training and different marginal loss gains for different frames, the probabilities converge to a shape similar to a normal distribution around the event and K parameters, as shown in Figure 3.3. Ultimately, it was decided to use this observation to derive our metrics. For that case, it was decided to draw the normal distributions around the events individually using K parameters to determine the widths, as shown in Figure 3.4. These values were then aggregated using the maximum function to receive the values called actual probabilities. It should be highlighted that these values were generated only for evaluation purposes. We just wanted to generate the targets that de-

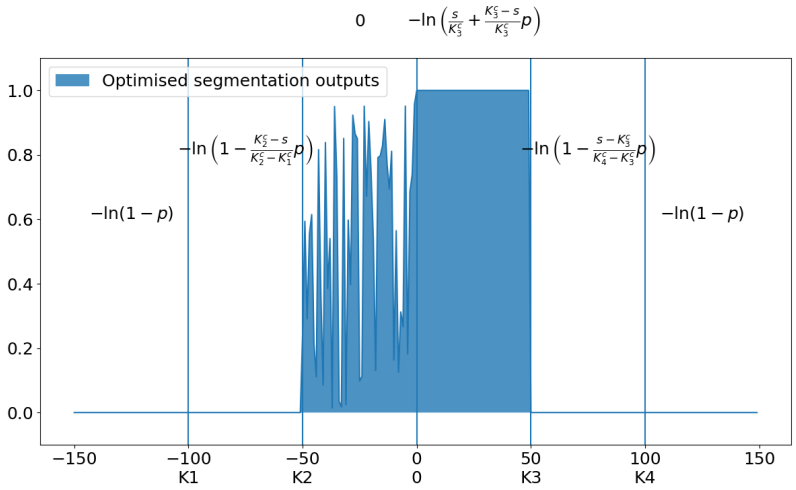

Figure 3.2: *Mathematically optimised CALF. It depicts the segmentation outputs that minimize the function depending on the frame distance from the event. The x-axis represents the distances from the event and y-axis the segmentation outputs.*

scribe the event context so the segmentation outputs could be compared to them. The segmentation outputs, however, are only determined by the event locations and K parameters in CALF.

Thus, for evaluation purposes, we assume that the context of the events can be described with a normal distribution and then the segmentation outputs are compared to it. For this purpose, we use the POR and ACR metrics, which are based on the logic of common evaluation metrics, precision, and recall.

Predictive Overlap Ratio (POR):

$$POR = \frac{\sum_{i=1}^{n} \min(p_i, a_i)}{\sum_{i=1}^{n} p_i}$$

Actual Coverage Ratio (ACR):

$$ACR = \frac{\sum_{i=1}^{n} \min(p_i, a_i)}{\sum_{i=1}^{n} a_i}$$

The difference lies in the fact that, commonly in machine learning tasks, we compare entities in a binary way. In this case, we want to compare the distributions of the segmentation outputs with artificially generated ones. Thus,

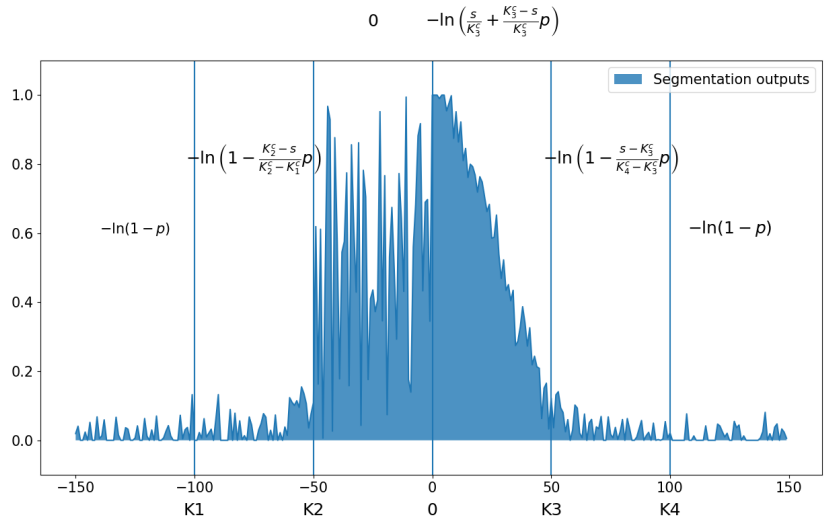

Figure 3.3: *CALF impact on the segmentation outputs. It depicts the probabilities obtained during training with CALF. The x-axis represents the distances from the event and y-axis the segmentation outputs.*

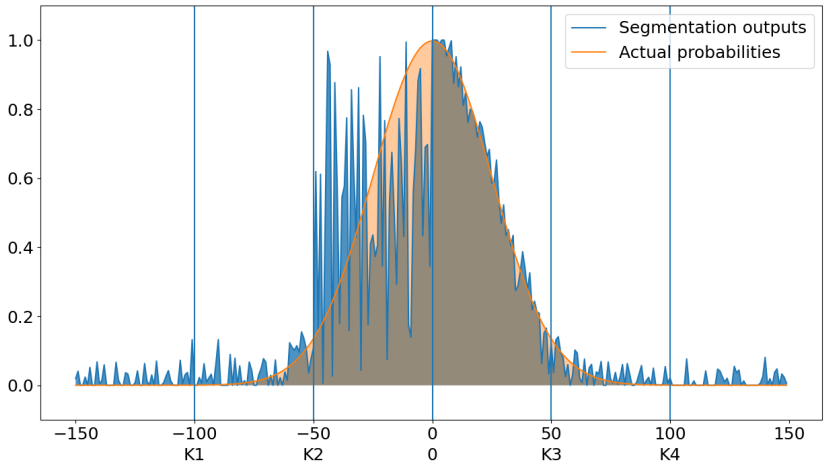

Figure 3.4: *Actual probabilities. It depicts the probabilities obtained from normal distribution that aim to present the context distribution. Used only for evaluation purposes. The x-axis represents the distances from the event and y-axis the segmentation outputs.*

the Predictive Overlap Ratio (POR) measures the proportion of the segmentation outputs (p_i) that overlap with the actual probabilities (a_i), which are artificially generated from the normal distributions around the events. The intuition behind POR is to assess the model's precision in predicting the context of the events.

Similarly, the Actual Coverage Ratio (ACR) measures the proportion of actual probabilities (a_i) covered by the segmentation outputs (p_i). The intuition here is to evaluate the model's recall in predicting the context of the events. Together, POR and ACR provide a comprehensive analysis of the model's ability to provide contextual information, offering insights into both precision and recall aspects of the model's performance.

3.3 Spotting module

As the segmentation module had a task to generate the contextual information for different events, the spotting one has a goal to analyse it and detect the event occurrence. As most applications predict events within relatively broad range of error, our tries to predict the event occurrence still on the frame level. This is a difficult task as it requires deep interpretation of the context provided from segmentation module.

3.3.1 Architecture

The input for the spotting module is a probability distribution that contains the contextual information of events. Therefore, it is crucial for the model to detect various patterns from this data. To achieve this, we built an architecture capable of detecting different levels of patterns.

Firstly, we determined that a 5 fps setting might be too detailed for event spotting and could be somewhat redundant. Thus, we decided to make detections at a 1 fps setting by using max pooling to reduce the size of the segmentation output. Next, we applied two parallel operations. The first operation involved calculating the average pooling and extending it to the

size of the clip to provide the global context. The second operation consisted of three sequential 1D CNN layers to capture patterns at different local levels. At the end, all these results were concatenated, followed by the application of a final classification 1D CNN layer.

To refine the reliability of its predictions, we used calibrators based on Platt Scaling [54]. This statistical technique adjusts the output of the convolutional network by fitting a logistic regression model to the initial predictions. Platt Scaling transforms the raw output scores into calibrated probabilities, providing a more accurate and interpretable measure of certainty regarding each predicted event. This calibrated output is crucial for practical applications, as it ensures that the predictions not only accurately reflect the likelihood of events but are also useful for decision-making processes in real-time tasks.

3.3.2 Evaluation

The primary metric for evaluating the spotting module is the Mean Average Precision (MAP) score. While MAP is widely used in various object detection and information retrieval applications for its efficient evaluation of precision at multiple recall levels, it is particularly sensitive to the exactness of event timing. Delays in detection can disproportionately lower the MAP score, potentially obscuring the practical usefulness of the model in operational sports analytics settings.

To provide a more useful evaluation, an event-centric method was adopted, where small delays in detection haven't been overly penalized. This approach is crucial because real-world applications often require a degree of temporal flexibility, acknowledging that precise frame-level synchronization may not always be feasible or necessary. This approach involves identifying specific events within the game footage, specified by their start and end frames, and assessing whether predictions were made in the immediate surrounding of these events. If a prediction aligns closely with the actual event, it is considered a positive detection. Conversely, the algorithm also generates negative events by identifying periods where no event occurred and dividing these into fixed-size subevents. If a prediction occurs within these non-event

periods, it is marked as a false positive.

This methodology allows for the calculation of Precision, Recall, and F1-scores for specific probability thresholds, providing a detailed measure of the model's accuracy and its ability to distinguish between actual events and non-events. Additionally, this approach enables the computation of the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) score, which are important metrics for evaluating the performance of binary classification systems. The ROC curve plots the true positive rate against the false positive rate at various threshold settings, while the AUC score provides an aggregate measure of performance across all possible classification thresholds. This robust analysis helps evaluate the effectiveness of the event detection model in a football analytics context, offering insights into its operational efficiency and precision in real-world scenarios.

In addition to quantitative metrics, animations were generated to visually depict the alignment between model predictions and actual game events. These visual tools are important for a hands-on evaluation, allowing for the identification of recurring patterns or discrepancies in model behavior. This detailed evaluation approach not only quantifies model success but also enriches understanding of its practical deployment in real-world settings.

3.4 Training

During the training of our model, we tackled several challenges, including data imbalance, overfitting, and lack of data, which needed to be addressed to create robust models. The issue of data imbalance arose because some events occurred significantly more often than others and often co-occurred. For instance, when selecting a clip with a shot event, there is a high probability of also having pressure and pass events within that clip. To address this, we decided that the data loader should select rarer events with higher probability. Ultimately, it selected events according to the inverse of their occurrence probability. This approach ensured a balanced representation of various event types in the training process, helping to mitigate the effects of data imbalance and improve the overall performance and reliability of the

model. Additionally, the position of the event within each clip was randomized rather than centered, enhancing the model's ability to generalize across different temporal contexts and reducing the model's tendency to overfit to a specific event placement.

To prevent overfitting, an early stopping criterion was employed. This technique monitors the validation loss during training and halts the training process if the validation loss discontinues to improve, thus preserving the model's generalization capabilities.

In terms of training tools and methodologies, the Adam optimizer was selected for both the segmentation and spotting modules due to its effectiveness in handling sparse gradients and adapting the learning rate during training. For the segmentation module, the CALF [55] was used to enhance context capture, while the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) was employed in the spotting module to provide a stable training process by directly working with logits and incorporating a sigmoid layer.

Hyperparameters were carefully chosen through a methodical approach. A grid search was conducted to determine the optimal clip size, balancing the need for sufficient context without excessive computational load. It was conducted by comparing POR and ACR of initial segmentation framework. The K values for CALF were set based on qualitative assessments of necessary context lengths for different events, recognizing that shorter events like passes require less contextual data compared to more complex events like shots.

Regarding software and computational resources, the training used PyTorch along with the PyTorch Geometric library for handling graph-based data structures efficiently. Due to the demanding nature of training deep learning models, especially those involving large networks and extensive data, additional computational resources were necessary. For this purpose, the GPU lab was used. For the models that have GATs as a feature extractor we used the NVIDIA A100 80GB PCIe with 80 GB RAM and for the rest of them Tesla V100-SXM3-32GB with 32 GB RAM.

Chapter 4

Results

This chapter of the thesis is dedicated to an analysis of the findings derived from our experimentation and modelling processes. This analysis is divided into three main sections, each designed to provide distinct insights into different aspects of our research. The first section, investigates the observations and patterns identified within our dataset. This part of the analysis is important as it sets the foundation for understanding the dynamics and characteristics of the data we are working with. It includes a variety of visualizations and animations that illustrate the data's properties and verify the efficacy of our preprocessing steps.

The next section is focused on a detailed quantitative assessment of the various models developed throughout the thesis. By comparing evaluation metrics across different phases of the research, this assessment highlights the relative strengths and weaknesses of each model configuration. The metrics include all measures presented previously that collectively offer a solid evaluation framework.

The final section shifts from quantitative to qualitative analysis, presenting a deeper dive into the operational effectiveness of the best-performing model. It includes a discussion on the practical implications of the model outputs, supported by qualitative data from animations that demonstrate the models' behavior in real-world scenarios. But more importantly, this part is important as it allows to draw conclusions that might be insightful for further

developments.

4.1 Data analysis

In the initial phase of data analysis, a thorough quality check of the positional data was conducted to ensure its integrity and completeness for subsequent modelling efforts. This step was crucial in establishing a reliable foundation for the research, as any inconsistencies or gaps in the data could significantly impact the accuracy and reliability of the event detection models.

The examination of the positional data began with verifying the continuity of data frames throughout each match. This process revealed no instances of missing frames, indicating a consistent capture of data during the games. Additionally, the presence of all players on the field was confirmed for each frame, which is vital for accurate spatial analysis and player interaction modelling. However, there was one exception, where one player in the game between Belgium and Russia has received a red card. As a result of this finding, a preprocessing adjustment was made to the dataset to include a 'red card' flag for the affected player.

To further ensure the accuracy of the positional data, we employed heat maps to visualize the distribution of player positions throughout the games. This technique provided a clear graphical representation of where players spent most of their time on the field, offering an intuitive method to assess the correctness of the recorded positions. An example of such a visualization is provided in the Figure 4.1, where the heat maps from a specific game illustrate the typical positional patterns expected from player movements during a match.

The heat maps were helpful in confirming the accuracy of the data, as they showed expected patterns such as midfielders covering large central areas, wingers primarily remained near the sidelines or forwards concentrating near the opposing goal. The consistency of these patterns with known football strategies and player roles substantiated the reliability of the data collected.

Additionally, the creation of game animations, which depicted the flow of

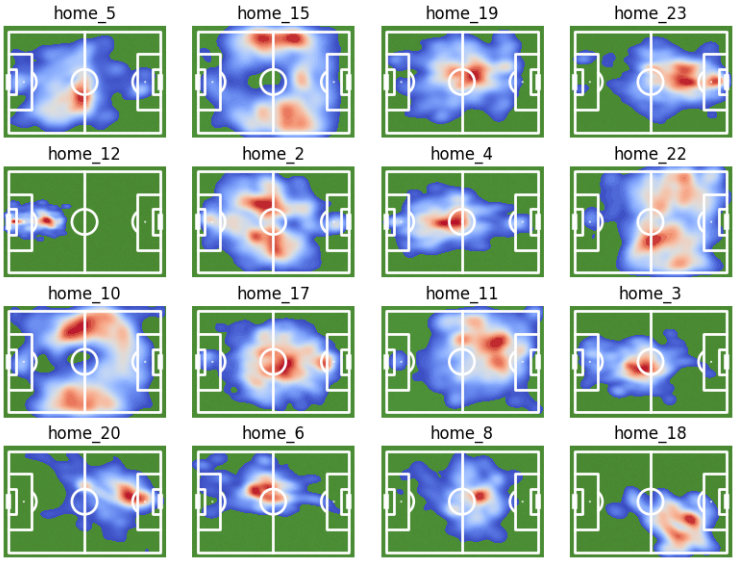

Figure 4.1: *Player heat maps Brasil vs Belgium WC2018. It provides the most common areas covered by specific players.*

player movements over time, further supported the accuracy of the positional data. These animations allowed for a dynamic review of how players interacted and moved across the pitch, providing visual confirmation that the positional data accurately reflected actual game flow. Beyond only confirming player locations, these animations were also useful for verifying other key features necessary for our graph-based analysis. Specifically, they enabled us to examine the edges between players to ensure they were correctly set, as these connections form the backbone of our graph neural network models; the snapshot of the animation is visually demonstrated in Figure 4.2. Moreover, the animations facilitated the verification of the accuracy in the generation of movement directions, an important aspect for analyzing player dynamics and strategies, as illustrated in Figure 4.3. This complete validation approach, including both positional accuracy and relational dynamics, ensured a solid foundation for the subsequent modeling phases of the thesis.

Moving on to the event data in our study, each football event is annotated only with the timestamp of its occurrence, without additional contextual information. The bar plot in Figure 4.4 presents the distribution of different event types across the dataset, highlighting a significant imbalance in their frequency. The 'Pass' event dominates, occurring more than 12,000 times,

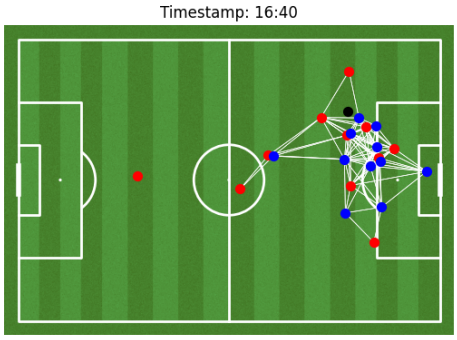

Figure 4.2: *Snapshot of the animation with edges. Presents player movements and connections between them based on the distances between them.*

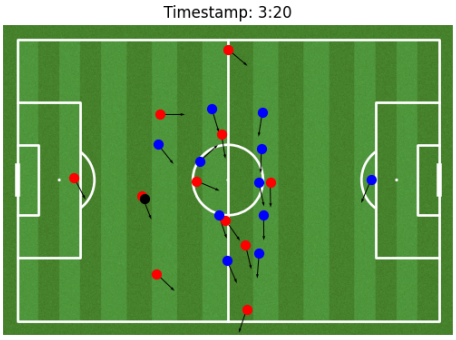

Figure 4.3: *Snapshot of the animation with directions. Presents players movements with their directions.*

whereas other events like 'Pressure', 'Foul Committed', and 'Ball Recovery' are considerably less frequent. This disparity in event counts highlights the challenges in training models for event detection in football, as the overwhelming occurrence of certain events could bias the learning process towards more frequently occurring types, potentially undermining the ability to accurately recognize and predict less common but equally important in-game actions.

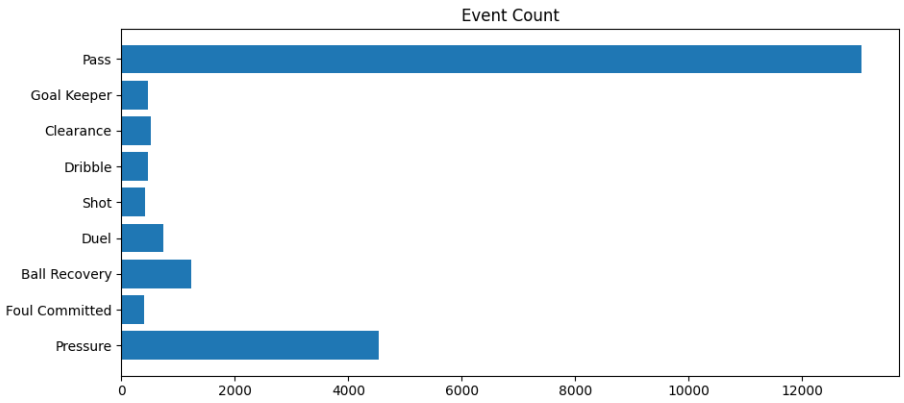

Figure 4.4: *Number of events from the event dataset*

4.2 Model performance

To explore the effectiveness of different GNN architectures in capturing contextual information, we created five distinct models. The first three utilized GCNs (base backbone), GATs, and GINs as their respective backbones or feature extractors and we decided to call them BackboneGCN, BackboneGIN and BackboneGAT. Furthermore, we introduced a fourth model which employed a GCN backbone complemented by a NetVLAD layer. This addition was intended to capture more complex spatial features. To remind, unlike typical applications where NetVLAD aggregates features across the entire clip, our approach treated each frame individually with convolutions to maintain temporal resolution. We called that model NetVLAD. The final model used a pre-trained model with a GAT backbone, subsequently fine-tuning it independently for each event. Therefore, each model could learn event-specific features therefore we called that model FineTuned.

4.2.1 Segmentation results

Initially, our focus was on understanding the impact of the Context-Aware Loss function, which was designed to enhance the model's ability to recognize and interpret the contextual dynamics of football events. This structured analysis was crucial for improving our approach and determining the most effective strategies for event detection in football, setting the stage for a detailed examination of the different models implemented during the research.

For evaluating the segmentation models, we employed two metrics: POR and ACR. These metrics were important in providing quantitative evaluations of the models' performance, focusing on how effectively the predicted probabilities aligned with artificially generated context data surrounding each event. These evaluations are important for understanding the precision with which temporal dynamics are handled by the model, ensuring that the predictions not only identify the occurrence of events but also accurately reflect their timing and context within the game's flow. This assessment helps improve the models further, pushing for upgrades in both the precision and accuracy

Results											
	Pressure	Foul	Ball recovery	Duel	Shot	Dribble	Clearance	Goal keeper	Pass	Dead	Average
BackboneGCN											
POR	0.46	0.13	0.24	0.12	0.20	0.13	0.17	0.22	0.63	0.77	0.31
ACR	0.59	0.31	0.48	0.26	0.40	0.31	0.36	0.42	0.88	0.77	0.48
BackboneGIN											
POR	0.46	0.14	0.27	0.12	0.20	0.13	0.18	0.22	0.62	0.77	0.31
ACR	0.62	0.30	0.40	0.30	0.42	0.33	0.38	0.43	0.89	0.79	0.49
BackboneGAT											
POR	**0.48**	**0.14**	**0.27**	**0.13**	**0.21**	**0.13**	0.18	**0.24**	**0.63**	**0.79**	**0.32**
ACR	0.65	0.28	0.40	0.26	0.41	0.31	0.39	0.43	0.90	0.80	0.48
NetVLAD											
POR	0.43	0.14	0.23	0.12	0.17	0.11	0.14	0.19	0.62	0.78	0.29
ACR	0.70	0.26	0.43	0.25	0.38	0.30	0.36	0.40	0.90	0.75	0.47
FineTuned											
POR	0.36	0.06	0.15	0.08	0.11	0.06	0.08	0.12	0.56	0.69	0.23
ACR	**0.90**	**0.36**	**0.59**	**0.44**	**0.51**	**0.41**	**0.56**	**0.62**	**0.95**	**0.88**	**0.62**

Table 4.1: *Segmentation evaluation results. It provides the POR and ACR results for all tested models. The BackboneGAT model slightly outperformed other models in terms of the POR scores. On the other hand, the FineTuned model proved to be the best at the ACR scores but at the cost of the POR scores.*

of event detection. Table 4.1 summarizes the performance of each model, providing a clear comparison of how each configuration progressed in terms of POR and ACR metrics across different event types. The general conclusion from that table indicates that BackboneGAT model slightly outperformed other models in terms of the POR scores. However, the FineTuned model proved to be the best at ACR scores but at the cost of POR scores.

4.2.2 Spotting results

The context derived from the segmentation models served as the foundational input for the spotting models, which were tasked with detecting specific events based on the analyzed contexts. To ensure an extensive evaluation of these spotting models, several evaluation techniques were employed. These techniques were important as they allowed us to validate the models beyond just the theoretical performance indicated by previous metrics, which were influenced by artificially generated targets.

Initially, all models were compared using the Mean Average Precision (MAP) statistic, which evaluates the accuracy of event predictions at the frame level. This metric is particularly useful for understanding how well each model identifies the precise frames where events occur, thus providing a direct measure of the spotting models' effectiveness in real-time event detection. The results of this comparison are summarized in the Table 4.2, which presents the MAP scores across all tested models, illustrating their relative performance in capturing and predicting event occurrences within video frames.

To evaluate the models from an event-centric perspective, Precision, Recall, and F1-Score metrics were calculated with 0.6 threshold. These metrics provide a more holistic view of the models' effectiveness by measuring not only the accuracy of the event detection but also the completeness and relevance of the detected events. Precision assesses the proportion of correctly predicted events out of all predicted events, indicating the accuracy of the detection. Recall evaluates the proportion of actual events that were correctly identified, reflecting the model's ability to capture all relevant instances. The F1-Score, a harmonic mean of precision and recall, provides a single metric that bal-

RESULTS											
	Pressure	Foul	Ball recovery	Duel	Shot	Dribble	Clearance	Goal keeper	Pass	Dead	Average
BackboneGCN											
MAP	0.24	**0.17**	0.09	0.04	0.24	**0.04**	0.17	0.22	**0.41**	0.94	0.26
P	0.42	0.18	0.19	0.09	0.24	**0.07**	0.18	0.23	0.66	0.58	0.28
R	0.58	0.66	0.45	0.52	0.73	0.59	0.65	0.72	0.77	0.87	0.65
F1	0.48	0.28	0.26	0.15	0.36	**0.13**	0.28	0.35	0.71	0.69	0.37
BackboneGIN											
MAP	0.25	0.15	**0.10**	**0.05**	0.29	**0.04**	0.17	0.23	0.40	0.96	0.26
P	0.42	0.16	0.20	**0.10**	**0.27**	**0.07**	**0.19**	**0.28**	**0.66**	0.55	0.29
R	0.55	**0.69**	0.41	0.50	**0.81**	0.53	0.66	0.78	0.80	0.90	0.66
F1	0.48	0.26	0.27	0.17	**0.40**	0.12	0.29	0.39	**0.72**	0.68	0.38
BackboneGAT											
MAP	**0.26**	**0.17**	**0.10**	**0.05**	**0.31**	**0.04**	0.18	**0.26**	**0.41**	**0.97**	**0.27**
P	**0.44**	**0.22**	**0.29**	0.09	**0.27**	**0.07**	**0.19**	**0.28**	**0.66**	**0.60**	**0.30**
R	0.65	0.64	0.43	0.63	0.79	**0.62**	0.67	0.80	0.80	0.89	0.69
F1	**0.53**	**0.33**	**0.28**	0.16	**0.40**	0.12	**0.30**	**0.41**	**0.72**	**0.72**	**0.40**
NetVLAD											
MAP	0.24	0.14	0.08	0.04	0.18	0.03	0.11	0.16	0.39	0.95	0.23
P	0.39	0.17	0.17	0.09	0.19	**0.07**	0.12	0.20	0.64	0.56	0.26
R	0.64	0.59	0.38	0.46	0.72	0.56	0.65	0.72	**0.82**	0.83	0.64
F1	0.49	0.26	0.23	0.15	0.30	0.12	0.20	0.31	**0.72**	0.67	0.34
FineTuned											
MAP	0.24	0.12	0.09	**0.05**	0.25	**0.04**	**0.20**	0.20	0.37	0.92	0.25
P	0.42	0.13	0.17	**0.10**	**0.27**	**0.07**	0.18	0.22	0.65	0.55	0.28
R	**0.71**	0.66	**0.69**	0.52	0.64	**0.62**	**0.75**	**0.83**	0.66	**0.91**	**0.70**
F1	**0.53**	0.21	**0.28**	0.17	0.38	0.12	**0.30**	0.35	0.66	0.68	0.37

Table 4.2: *Spotting evaluation results. It provides the MAP, precision (P), recall (R), and F1 scores for all tested models. The BackboneGAT model slightly outperformed other models in terms of the MAP, precision and F1 scores. On the other hand, the FineTuned model proved to be the best at the recall scores.*

ances both the precision and the recall, offering a comprehensive measure of a model's overall performance in detecting events accurately and completely. These metrics collectively offer a rounded assessment of how well each model performs in real-world scenarios where timing exactitude is balanced with the need for accurate event recognition. All these results are presented in the following Table 4.2.

4.3 Comparison of the model outputs

For this analysis, BackboneGAT model was selected as the primary focus due to its superior performance demonstrated in the quantitative evaluation. This model, which effectively utilized a GAT architecture, showed a consistent ability to capture complex patterns in the data, making it a strong candidate for deeper qualitative review.

The Figure 4.5 illustrates a detailed comparison between the game predictions made by the model and the actual ground truths for each event. This visual representation is important for a general understanding of the model's performance across different types of events within football matches. It presents how closely the model's predictions align with the reality of the events as they occurred on the field, highlighting both the accuracy and the areas where discrepancies exist. By examining these visual comparisons, we can evaluate the model's precision in capturing the timing and classification of events, providing general insights into its reliability and potential areas for improvement in real-world applications.

The figure depicting the entire game's predictions and ground truths provides a general overview but lacks detailed, actionable insights. To address this and investigate deeper model's performance, a script was developed specifically to generate animations of the most significantly missed events which synchronizes the prediction plots with the video footage. This visual layout is detailed in the Figure 4.6. For the qualitative analysis, ten different animations were generated to illustrate missed annotations for each event type. Since these animations cannot be displayed here, Figure 4.7 was created to show selected predictions and annotations for these events. Additionally,

segments providing context derived from the animations have been included. These contextual insights allow us to identify specific situations where the model struggles to detect events accurately, thus highlighting areas for potential improvement in model performance.

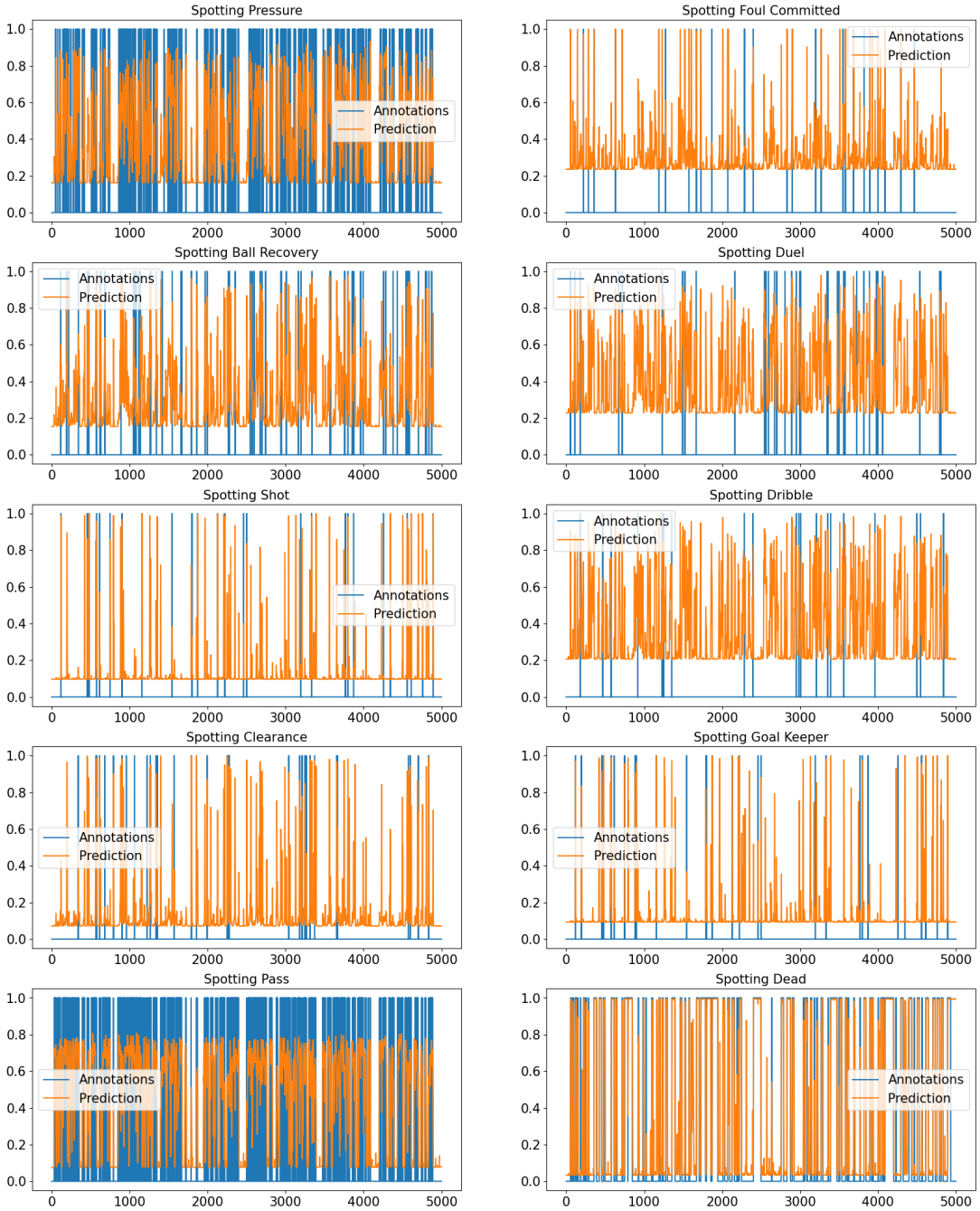

Figure 4.5: *Belgium vs Brasil game predictions. The x-axis of the plots presents the frame number and the y-axis presents the probability of event occurrence. Provides a general understanding of the model's performance across different types of events.*

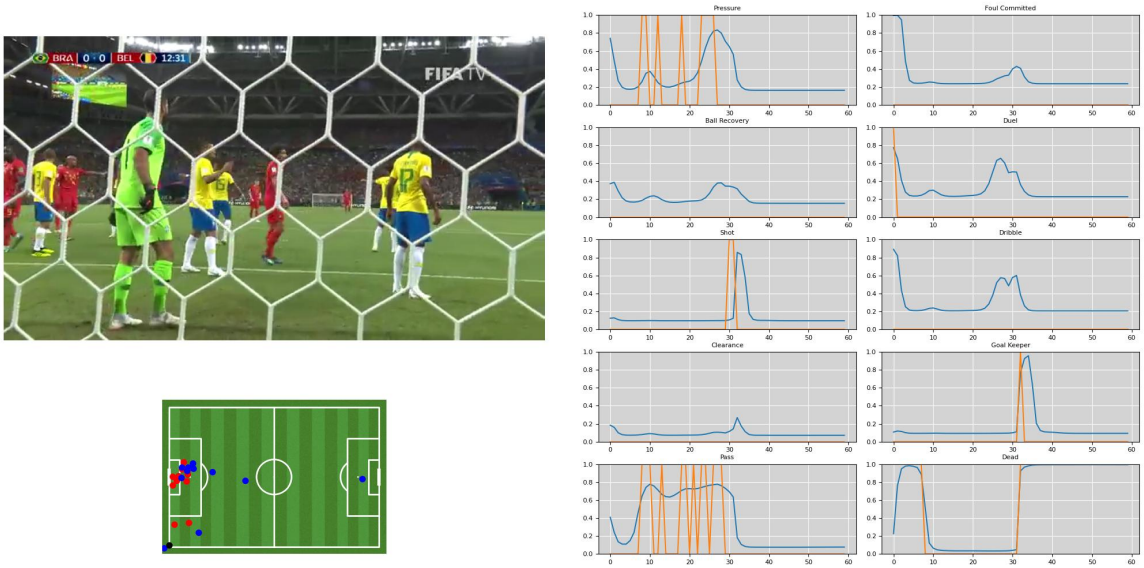

Figure 4.6: *Animation layout. The animation provides insight into how the model behaves as time progresses. It enables a comparison of its results with the broadcast video of the game.*

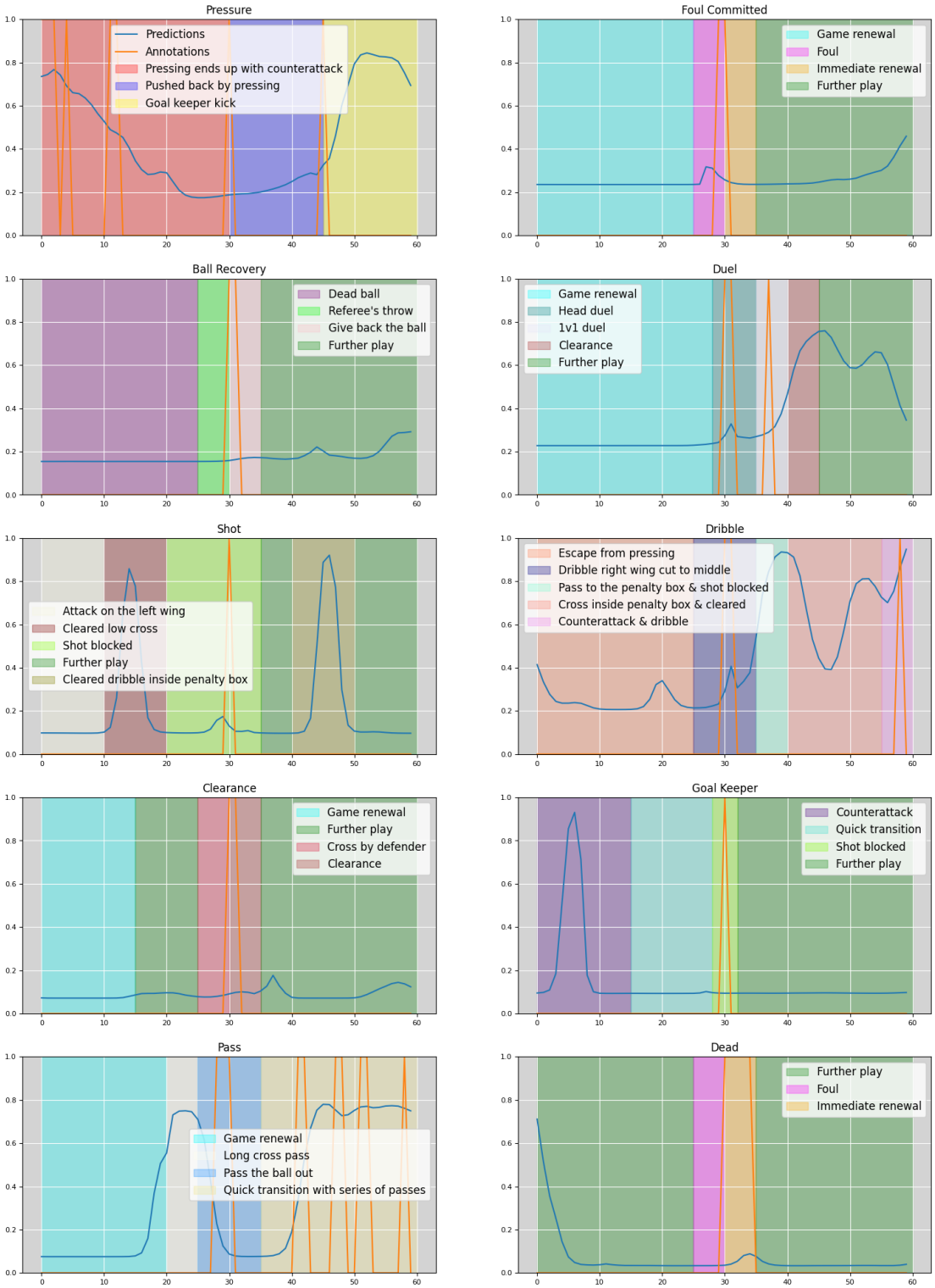

Figure 4.7: *Predictions with context segments. The figure consists of ten different clips where each event was misclassified. The colored spans offer contextual information derived from the generated animations.*

Chapter 5

Discussion

The "Discussion" chapter of this thesis explores thoroughly the broader implications, detailed interpretations, and potential enhancements related to the methodologies and findings outlined in previous chapters. Structured to provide a detailed understanding, this chapter is divided into four subsections. In the first one where the data derived from the segmentation and spotting models is thoroughly analyzed. This analysis aims to clarify the significance of the observed outcomes and evaluate how they align with the theoretical frameworks established in the methodology.

The following sections explore the practical and theoretical consequences of the study's findings. It discusses the impact of advanced event detection techniques, such as the use of graph neural networks, on future research directions in sports analytics and their potential to transform industry practices.

The discussion then moves to the limitations of the study, acknowledging the constraints posed by the data, analytical tools, and methodologies employed. This evaluation not only outlines the boundaries of the study's findings but also sets the stage for future questions by highlighting potential fields for further investigation.

The last subsection proposes adjustments and future research paths. It suggests improvements to modelling approaches, data collection protocols, and analysis techniques aimed at boosting the precision and utility of event detec-

tion systems in sports analytics. This forward-looking perspective is designed to catalyze advancements in the field, pushing the boundaries of what can be achieved in sports data analysis.

5.1 Interpretation of results

This section is an integral part of the discussion where we provide a detailed analysis of the findings from the segmentation and spotting models used in this thesis. This analysis aims to describe the performance of the different models within the framework of football event detection. The discussion will offer insights into how various factors such as model architecture techniques influenced the overall system performance, providing a deeper understanding of the strengths and limitations observed during the study.

5.1.1 Interpretation of the segmentation results

In the evaluation of the segmentation models, BackboneGAT, which utilized a GAT as its feature extractor, demonstrated the best performance out of all the models we experimented on. It performed especially well in detecting pressures and shots. This suggests that the attention mechanism in GATs may enhance the model's capability to capture the relevant contextual nuances necessary for accurate event detection in dynamic football environments.

BackboneGIN, featuring a GIN as a feature extractor, also outperformed the baseline BackboneGCN model in most categories. This enhancement can be attributed to the GIN's ability to more effectively model complex patterns in the data, which are important for interpreting the complex interactions during football matches. These results highlight the potential advantages of advanced GNN architectures in improving context capture within the event detection framework.

Contrarily, model that included the NetVLAD pooling layer alongside a GCN as feature extractor, did not yield an improvement over the simpler GCN

model. Despite the intention for NetVLAD to enhance the extraction of complex spatial features by aggregating more expressive descriptors, it resulted in lower Predictive Overlap Ratio (POR) scores across most events compared to the baseline. This could indicate that while NetVLAD enhances spatial feature representation, it may not synergize well with the contextual dynamics required for event detection in football, as modeled by the context-aware loss function.

Finally, FineTuned model, which combines separate pre-trained GAT models for each event, performed significantly worse than all other models in terms of the POR scores. This decline in performance highlights the challenges associated with fine-tuning in scenarios where the model may overfit to less representative aspects of the data, losing generalizability across various football gameplay situations.

Analyzing the Actual Coverage Ratio (ACR) results further confirms the enhanced capabilities of advanced GNN architectures over the baseline GCN model. The data reveals that both GIN (BackboneGIN) and GAT (BackboneGAT) architectures tend to provide more complete coverage of actual events context compared to the GCN (BackboneGCN). Their advantage was visible in several key event categories like Pass and Goalkeeper actions. This suggests that these architectures are more adept at capturing the full context of events, which is crucial for effective event detection in football.

Interestingly, NetVLAD model, shows ACR results similar to those of the baseline GCN model, indicating no substantial improvement in the coverage of actual events. The similarity in ACR alongside lower POR suggests a potential overfitting issue with NetVLAD model, where it might be capturing event contexts accurately but not aligning as closely with the actual event timings. However, its performance in detecting Pressure events is notably higher than other models, which could point to specific contexts where the NetVLAD layer adds significant value.

Moreover, FineTuned model shows a markedly higher ACR across all events, significantly outperforming other models. This suggests that while FineTuned model might be over-predicting, it also captures a broader range of event-related data, making it the most complete in terms of context coverage.

This trait, although potentially leading to higher false positives, underscores the model's ability to ensure that no relevant event goes undetected, which can be particularly advantageous in scenarios where capturing every potential context is necessary.

5.1.2 Interpretation of the spotting results

Analysis of the Mean Average Precision (MAP) results from the spotting models provides a detailed perspective on how each model performs at the frame level across various football events. This table highlights the effectiveness of the models in identifying the correct timing and occurrence of events, which is crucial for real-time sports analytics and automated event detection.

BackboneGAT model, shows the highest average MAP score at 0.2732. This better performance is consistent across multiple event types, especially excelling in 'Shot' and 'Goal keeper' events, and achieving the highest score in the 'Dead' category, which involves detecting moments when the ball is out of play. The GAT's ability to focus on important nodes within the graph dynamically helps in capturing the nuances of such complex event sequences more effectively than the other models.

BackboneGIN model, also shows strong performance with an average MAP of 0.2617. It leads in the 'Shot' category, suggesting that GIN's capability to capture more subtle graph patterns is beneficial, especially in events requiring detailed spatial analysis like shot attempts.

BackboneGCN and NetVLAD models show comparatively lower performance, with average MAP scores of 0.2560 and 0.2305, respectively. The integration of NetVLAD layer does not seem to provide an advantage in this specific application, potentially indicating that while NetVLAD can help in aggregating spatial features, it may not be as effective in capturing the precise timing required for accurate event spotting in football.

The performance of FineTuned model, is slightly below BackboneGAT, which suggests that additional fine-tuning did not yield the expected improvements and might have led to overfitting or insufficient adaptation to the specific characteristics of event detection in football.

From a feature-specific perspective, the MAP scores reveal some interesting insights into the model's capabilities and potential areas for improvement. The categories 'Duel' and 'Dribble' consistently show the lowest performance across all models, which was not entirely anticipated given their relatively moderate results in the segmentation phase. This discrepancy suggests that while the models may be capturing some aspects of these events, there could be a significant number of false positives affecting the precision of event spotting. This issue justifies closer examination in the qualitative analysis to better understand the underlying causes and to improve the models accordingly.

On the other hand, the 'Dead' event category exhibits the highest performance, which aligns with expectations due to the distinct and often clearer nature of these events within the game data. Events categorized as 'Dead' typically have well-defined start and end points, making them easier to detect compared to more dynamically developed events.

Additionally, the events 'Pass', 'Shot', and 'Pressure' also perform relatively well, showcasing the models' effectiveness in detecting significant actions that have substantial impacts on the flow of the game. These results are encouraging, as they indicate that the models are not only capable of recognizing clear-cut events but are also proficient in identifying important moments that could influence game outcomes.

To ensure a complete evaluation of the event detection models that is not excessively influenced by minor timing delays, an event-centric approach was adopted for analysing performance. This method allows us to evaluate the models based on their ability to identify the occurrence of an event within its expected error, rather than just evaluating the accuracy at the exact frame level.

The evaluation based on precision, recall, and F1 scores (Table 4.2) highlights the complex capabilities of the models when analyzed from an event perspective. Among the models, BackboneGAT model consistently presents better performance, confirming its robustness in event detection across various football scenarios. This model strikes an optimal balance between detecting events accurately (high precision) and ensuring most relevant events

are detected (high recall), as reflected in its leading F1 scores.

FineTuned model, shows variability in its performance but excels in ensuring the detection of events (high recall), which contributes to its higher F1 scores in certain events like 'Pressure' and 'Clearance'. This suggests that combining separate pre-trained GAT models for each event can enhance the model's sensitivity to events, although sometimes at the cost of precision.

BackboneGCN and NetVLAD models while generally lower in performance compared to GAT-based models, highlight the challenges and potential limitations natural in the GCN architecture or the integration approach with NetVLAD when applied to complex, dynamic sports events.

Certain features such as 'Dead' and 'Pass' show consistently high performance across all metrics, indicating that these event types are well-captured by the models due to perhaps clearer defining characteristics or less uncertainty in their occurrences. On the contrary, events like 'Dribble' and 'Duel' present lower scores, highlighting ongoing challenges in detecting more fluid, rapidly changing situations on the field. These events require models to dynamically adapt to swift changes in player positions and interactions, which may not always be effectively captured by the current model architectures. The high recall scores seen in 'Goalkeeper' events across most models suggest that these events are generally easier to detect, likely due to the distinct and localized nature of goalkeeper actions. However, the precision varies, indicating potential over-detection or misclassification issues.

In the qualitative evaluation of model outputs, the finer details revealed through visual analysis allow for a deeper understanding of model behavior beyond what is quantitatively captured. Figure 4.5, which presents the predictions covered with actual annotations for an entire game, serves as a compelling illustration of this point. From these visualizations, we observe certain patterns confirming previous suspicions regarding model performance. Particularly, events that occur less frequently, such as duels and dribbles, tend to generate a high number of false positives, suggesting a sensitivity issue or a lack of sufficient training examples. On the other hand, well-represented events like passes, pressure situations, and dead ball scenarios are accurately detected, showcasing the model's effectiveness in identifying common game

Event	Observation
Pressure	The model tends to slightly delay in recognizing pressing actions, particularly in midfield scenarios. This might indicate a need for adjustments in temporal sensitivity.
Foul	The model struggles to recognize fouls followed by immediate game renewals, suggesting that the contextual transition from a foul to game continuation is not being captured effectively.
Ball Recovery	An unusual scenario occurred during a referee throw where the Belgium team was required to return the ball to the opposing team, which the model failed to classify correctly.
Duel	The model showed delays in recognizing both head duels and one-on-one duels. This might be due to overlapping actions that confuse the model.
Shot	Blocked shots were frequently missed, and low crosses from the wings were often mistaken for shots, indicating a confusion in the model between different types of ball movements.
Dribble	There was a noticeable delay in capturing the initiation of dribble movements, suggesting that the model may benefit from more reactive temporal features.
Clearance	The model failed to recognize some obvious clearances, pointing towards a potential gap in understanding defensive actions.
Goalkeeper	Actions blocked by the defender were labeled as goalkeeper interventions, indicating misclassifications between defensive actions and goalkeeper-specific movements.
Pass	The model accurately detected long crosses but failed in situations where a pass was immediately followed by the ball going out of play. Additionally, it successfully detected quick series of passes, highlighting its effectiveness in capturing rapid exchanges, though these were recognized as joint events rather than separate ones.
Dead Ball	Dead ball with immediate renewals were not captured correctly, suggesting the model may not effectively handle quick transitions back into active play.

Table 5.1: *Conclusions derived from the qualitative analysis. The table illustrates the patterns in which models encountered difficulties in classification, serving as a valuable resource for further improvements..*

occurrences. Shots and ball recovery events also show decent performance, although with room for development.

To perform a deeper qualitative analysis and uncover more nuanced behavioral patterns missed by the model, an animation tool was developed, generating clips to visually represent and analyze the events. Through this approach, specific patterns of model inaccuracies were identified across different event types as shown in Figure 4.7. In the Table 5.1 some observations and conclusions are presented.

5.2 Research Findings

This study aimed to advance the field of sports analytics by developing advanced detection algorithms and validating them. In pursuit of this goal, we formulated research questions to guide our investigation. In this section, we aim to provide answers that will help clarify key aspects of our research and contribute to the ongoing evolution of sports analytics. As a reminder, the research questions are presented below:

RQ1: How do temporal dynamics influence the accuracy of event predictions in football, and what strategies can be implemented with advanced GNN architectures to effectively capture and utilize these dynamics for improved predictive performance?

RQ2: How do different GNN architectures, such as Graph Convolutional Networks, Graph Attention Networks, and Graph Isomorphism Networks, perform in predicting football events?

RQ3: How does the combination of the NetVLAD pooling method and Context Aware Loss function enhance the performance of Graph Neural Networks in accurately predicting football events?

Regarding the first research question, the thesis particularly took into account temporal dynamics by applying 1D CNNs and CALF during training. The quantitative and qualitative assessments confirmed the relative success of this approach in capturing contextual information. This method effec-

tively shaped and detected the context around events, which is essential in the football domain.

For the second research question, the thesis employed different GNN architectures to develop algorithms capable of capturing nuanced spatial information. Based on the evaluation results, we concluded that GAT networks performed the best in analyzing football dynamics. Additionally, combining these solutions for capturing temporal dynamics from the previous research question allowed us to build algorithms and methods that effectively mapped the complex dynamics of football gameplay, reflecting both spatial interactions and the temporal sequence of events.

Lastly, the novel application of NetVLAD combined with CALF did not entirely enhance the performance of the models. However, further investigation into parameter tuning might increase performance.

5.3 Implications

Even though the model performance is not yet robust enough to be deployed for live event detection during games, it holds significant potential for post-game analysis. The predictive outputs, especially in scenarios like high shot probability, might signal critical phases of the game such as increased danger to the goal. Consequently, these models could be useful for teams and coaches looking to identify specific game phases for detailed strategic review. By analyzing these moments, coaching staff can enhance tactical planning, adjust training focus, and ultimately improve overall team performance. This application of the models allows for a deeper understanding of game dynamics and player behaviors, offering a valuable tool for sports analysts and team strategists.

Additionally, the model could be used to enhance fan engagement during live matches or replays. By identifying key moments and phases within a game, such as potential scoring opportunities or critical defensive maneuvers, broadcasters and content creators can highlight these instances, enriching the viewing experience for fans. This application not only can help in maintain-

ing viewer interest through dynamic content but also educates the audience about complex game strategies and player movements. Such enriched content can lead to a more informed and engaged fanbase, which is beneficial for both broadcasters and sports teams in terms of increased viewership and fan loyalty.

5.4 Limitations

In the development and application of the models presented in this thesis, several limitations were identified that could affect the generalizability and effectiveness of the findings. These limitations are essential to consider for future research and practical applications of the models.

One primary limitation is the data synchronization bias. The research relied on two main data sources: one for tracking data and the other for event annotations. Misalignments or inconsistencies between these sources could lead to inaccuracies in the models' training and predictive performance, impacting the reliability of event detection.

Additionally, the distinction between event detection and event spotting presents a fundamental challenge. The models are designed to detect events within a broader temporal context, which may not align perfectly with the precise and instantaneous requirements of event spotting. This difference can result in delays or inaccuracies, particularly when real-time detection is crucial.

The tracking data utilized in this study may also be insufficient for capturing all relevant event characteristics. While it provides detailed information on player positions and ball locations, it lacks three-dimensional depth—such as player elevation or jump heights—that could enhance event characterization. Furthermore, integrating visual and audio features from game footage could significantly improve the detection of events, offering information that positional data alone cannot provide.

A significant limitation is the competition-specific bias introduced by exclusively analyzing games involving the Belgium national team. This focus

may limit the models' applicability to different teams or leagues with varying styles of play and tactical approaches. The models' performance might not translate as effectively to other contexts, reducing their utility in broader football analytics scenarios.

5.5 Future work

In light of the identified limitations, several potential improvements can be considered to enhance the accuracy and applicability of the models developed in this thesis. These improvements aim to address the specific challenges observed during the research and provide a roadmap for future work in sports analytics.

One possible improvement is the adoption of Recurrent Neural Networks (RNNs) in place of Convolutional Neural Networks (CNNs) for handling temporal information. RNNs are particularly well-suited for sequence prediction tasks, as they can capture temporal dependencies and dynamics more effectively than CNNs. This change could lead to more accurate predictions of events over time, particularly in complex game scenarios where the sequence of actions plays a critical role.

Including additional data types such as video and audio features could also significantly enhance model performance. Visual and auditory information can provide context that is not captured through positional data alone, such as the crowd's reaction, referee's whistle, or visual information that are connected with key events like fouls or goals. Integrating these features could lead to a richer, more valuable dataset and improve the model's ability to detect and classify events accurately.

Expanding the temporal context considered by the models—both in terms of how far back in time the models look when making predictions and how they handle ongoing events—could provide additional benefits. A longer context window might capture more of the buildup to events, offering deeper insights into the conditions leading to specific outcomes.

Revising the labelling strategy used in training data could also yield im-

provements. Current models might benefit from labels that better reflect the complexities of real-world sports scenarios, including more detailed or differently categorized event types.

Lastly, conducting further analysis with different sets of hyperparameters and model configurations could uncover more optimal approaches to model training and architecture. Exploring a wider range of parameters, especially those related to the structure and depth of neural networks, could increase the models' abilities to generalize across diverse datasets and scenarios.

By implementing these improvements, future research can build on the foundational work of this thesis to develop more robust, accurate, and flexible models for event detection in football and potentially other sports.

Chapter 6

Conclusion

In this thesis, we engaged in an exploration of advanced machine learning techniques for event detection in football. We focused on the integration of GNNs and the Context-Aware Loss Function. Our research aimed to enhance the quality and synchronization of positional data, develop sophisticated algorithms for event detection, and create validation tools to assess these algorithms effectively.

Ultimately, our research successfully combined advanced GNN architectures, specifically GCNs, GATs and GINs, with the Context-Aware Loss Function, achieving significant improvements in the temporal accuracy of event predictions. This combination allowed for a nuanced understanding of the dynamics within football games. Furthermore, we explored the use of the NetVLAD pooling layer within our models, which, although variable in its effectiveness, showed potential in enhancing spatial feature representation in specific contexts.

The methodologies developed offer valuable tools for sports analytics, particularly in enhancing game analysis by enabling the detection of different phases within a match. These insights can be useful for coaches and analysts in strategizing and improving player performance based on detailed, data-driven event analysis. The models and tools from this thesis provide a foundation for deeper tactical analysis.

Despite these advancements, the study faced several limitations. The exclusive focus on Belgian football games may affect the generalizability of the models to other contexts with varying playing styles. Synchronization discrepancies between different data sources presented challenges that could impact the precision of event detection. Moreover, relying only on tracking data limited our ability to capture all aspects of game events, as additional insights could potentially be captured from audio-visual data.

Future research could build on these findings by incorporating multi-dimensional data sources, including visual and auditory data, to enrich the feature sets used for event detection. Further exploration of diverse neural network architectures and expanding the dataset to include a wider range of teams and leagues would help to overcome the limitations noted in this study.

From a personal perspective, this thesis has significantly broadened my knowledge across various domains. Firstly, it enabled me to explore the field of graph learning, motivating me to truly grasp its potential. It also provided the opportunity to work with spatio-temporal data, which is often challenging due to its complex patterns. Furthermore, it highlighted the importance of utilizing a diverse array of validation tools, including both quantitative and qualitative methods.

Through this process, I gained valuable insights into developing the detection model for football events. My general observation is that different approaches should be customized to different event types based on their characteristics. For instance, the positional approach performed well for pressure events and could be effective for detecting off-sides. Particularly in scenarios involving multiple player interactions. Conversely, for events such as fouls or duels, incorporating broadcast video features would be beneficial, as these events involve player contacts that are not visible in positional data. For simpler and more frequent actions like passes, a rule-based approach might be more suitable. Overall, this experience has deepened my understanding of machine learning applications in sports analytics and has inspired new ideas for future research. It also confirmed my understanding of the complexity involved in applying sports analytics to the football domain.

Ultimately, this thesis marks a step toward transforming sports analytics

through advanced machine learning, as it has provided valuable insights into the general challenges encountered in sports analytics. By addressing these challenges and offering innovative solutions, this research lays the foundation for future advancements in the field, ultimately enhancing our understanding and application of data-driven insights in sports.

Bibliography

[1] Deloitte. Deloitte football money league. https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html, 2023.

[2] World Football Summit. Impact of sports beyond the pitch. https://worldfootballsummit.com/impact-of-sports-beyond-the-pitch/, 2023.

[3] Omar Mohamed. The impact of technology on football: Challenges and benefits. *Medium*, 2023.

[4] The Daily Guardian. The impact of technology on football. https://thedailyguardian.com/the-impact-of-technology-on-football/, 2023.

[5] FC Business. Navigating the digital revolution: The impact of technology on modern football. https://fcbusiness.co.uk/news/navigating-the-digital-revolution-the-impact-of-technology-on-modern-footbal

[6] FIFA. Navigating the digital revolution: Whu research report on technology in football. https://digitalhub.fifa.com/m/31cbb0f9e12a57b1/original/211012_WHU_Research_Report_Fifa_EN_Digital_RZ.pdf, 2021.

[7] Fortune Business Insights. Sports analytics market size, share & industry analysis. https://www.fortunebusinessinsights.com/sports-analytics-market-102217, 2024.

[8] Jordy Post. Moneyball and soccer data: A game-changing approach. 2023.

[9] Viroshan Naicker. How math and data science made liverpool the best team on the planet. *Medium*, 2020.

[10] Nitin Singh. Sport analytics: A review. *The International Technology Management Review 9*, 2020.

[11] Georgios Nalbantis, Tim Pawlowski, and Dennis Coates. The fans' perception of competitive balance and its impact on willingness-to-pay for a single game. *Journal of Sports Economics*, 18(5):479–505, 2017.

[12] Mattie Toma. Missed shots at the free-throw line: Analyzing the determinants of choking under pressure. *Journal of Sports Economics*, 18(6):539–559, 2017.

[13] Bharathan, RP Sundarraj, and Abhijeet. A self-adapting intelligent optimized analytical model for team selection using player performance utility in cricket. 2015.

[14] Victor Cordes and Lorne Olfman. Sports analytics: Predicting athletic performance with a genetic algorithm. 2016.

[15] Verónica Baena. Analyzing online and mobile marketing strategies as brand love drivers in sports teams. findings from real madrid. *International Journal of Sports Marketing and Sponsorship*, 17:1–18, 07 2016.

[16] Victor Chazan Pantzalis and Christos Tjortjis. Sports analytics for football league table and player performance prediction. pages 1–8, 2020.

[17] Gervas Mgaya, H. Liu, and Bo Zhang. *A Survey on Applications of Modern Deep Learning Techniques in Team Sports Analytics*, pages 434–443. 04 2021.

[18] Chieh-Yu Chen, Wenze Lai, Hsin-Ying Hsieh, Wen-Hao Zheng, Yu-Shuen Wang, and Jung-Hong Chuang. Generating defensive plays in basketball games. 2018.

[19] Konstantinos Rematas, Ira Kemelmacher-Shlizerman, Brian Curless, and Steve Seitz. Soccer on your tabletop. 2018.

[20] Javier Fernández and Luke Bornn. *SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer*, page 491–506. Springer International Publishing, 2021.

[21] Rajiv Shah and Rob Romijnders. Applying deep learning to basketball trajectories, 2016.

[22] Anthony Sicilia, Konstantinos Pelechrinis, and Kirk Goldsberry. Deephoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data, 2019.

[23] Sujoy Ganguly and Nathan Frank. The problem with win probability.

[24] Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, and Patrick Lucey. Generating multi-agent trajectories using programmatic weak supervision, 2019.

[25] Hoang M. Le, Yisong Yue, Peter Carr, and Patrick Lucey. Coordinated multi-agent imitation learning, 2018.

[26] Hoang Minh Le, Peter Carr, Yisong Yue, and Patrick Lucey. Data-driven ghosting using deep imitation learning. 2017.

[27] Manzhu Yu, Myra Bambacus, Guido Cervone, Keith C. Clarke, Daniel Q. Duffy, Qunying Huang, Jing Li, Wenwen Li, Zhenlong Li, Qian Liu, Bernd Resch, Jingchao Yang, and Chaowei Phil Yang. Spatiotemporal event detection: a review. *International Journal of Digital Earth*, 13:1339 – 1365, 2020.

[28] Ahmed Walid. How are football stats collected? visiting statsbomb's data centre. 2023.

[29] Mohammed Yassine, Lamia Kazi Tani, Lamia Fatiha, Abdelghani Ghomari, Ahmed Bella, and Algeria Oran. An offside soccer detection system using ontology and deep learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 13:377–387, 06 2022.

[30] Lia Morra, Francesco Manigrasso, Giuseppe Canto, Claudio Gianfrate, Enrico Guarino, and Fabrizio Lamberti. *Slicing and Dicing Soccer: Au-*

tomatic Detection of Complex Events from Spatio-Temporal Data, page 107–121. Springer International Publishing, 2020.

[31] Behzad Mahaseni, Erma Rahayu Mohd Faizal, and Ram Gopal Raj. Spotting football events using two-stream convolutional neural network and dilated recurrent neural network. *IEEE Access*, 9:61929–61942, 2021.

[32] A. Ekin, A.M. Tekalp, and R. Mehrotra. Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, 12(7):796–807, 2003.

[33] Lamberto Ballan, Marco Bertini, Alberto Del Bimbo, and Giuseppe Serra. Action categorization in soccer videos using string kernels. In *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, pages 13–18, 2009.

[34] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Action classification in soccer videos with long short-term memory recurrent neural networks. In Konstantinos Diamantaras, Wlodek Duch, and Lazaros S. Iliadis, editors, *Artificial Neural Networks – ICANN 2010*, pages 154–159, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[35] Ian Simpson, Ryan J. Beal, Duncan Locke, and Timothy J. Norman. Seq2event: Learning the language of soccer using transformer-based match event prediction. New York, NY, USA, 2022. Association for Computing Machinery.

[36] Haohao Jiang, Yao Lu, and Jing Xue. Automatic soccer video event detection based on a deep neural network combined cnn and rnn. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 490–494, 2016.

[37] Himangi Saraogi, Rahul Sharma, and Vijay Kumar. Event recognition in broadcast soccer videos. pages 1–7, 12 2016.

[38] Behzad Mahaseni, Erma Rahayu Mohd Faizal, and Ram Gopal Raj. Spotting football events using two-stream convolutional neural network

and dilated recurrent neural network. *IEEE Access*, 9:61929–61942, 2021.

[39] Olav A. Nergård Rongved, Steven A. Hicks, Vajira Thambawita, Håkon K. Stensland, Evi Zouganeli, Dag Johansen, Cise Midoglu, Michael A. Riegler, and Pål Halvorsen. Using 3d convolutional neural networks for real-time detection of soccer events. *International Journal of Semantic Computing*, 15(02):161–187, 2021.

[40] Olav A. Norgård Rongved, Steven A. Hicks, Vajira Thambawita, Håkon K. Stensland, Evi Zouganeli, Dag Johansen, Michael A. Riegler, and Pål Halvorsen. Real-time detection of events in soccer videos using 3d convolutional neural networks. In *2020 IEEE International Symposium on Multimedia (ISM)*, pages 135–144, 2020.

[41] Pascal Bauer and Gabriel Anzer. Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction. *Data Mining and Knowledge Discovery*, 35, 09 2021.

[42] Ferran Vidal-Codina, Nicolas Evans, Bahaeddine El Fakir, and Johsan Billingham. Automatic event detection in football using tracking data, 2022.

[43] Michael Stöckl, Thomas Seidl, Daniel Marley, and Paul Power. Making offensive play predictable -using a graph convolutional network to understand defensive performance in soccer. 04 2021.

[44] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Bernard Ghanem, Marc Van Droogenbroeck, Rikke Gade, and Thomas B. Moeslund. A context-aware loss function for action spotting in soccer videos, 2020.

[45] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. Soccernet: A scalable dataset for action spotting in soccer videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2018.

[46] Olav Rongved, Markus Stige, Steven Hicks, Vajira Thambawita, Cise Midoglu, Evi Zouganeli, Dag Johansen, Michael Riegler, and Pål

Halvorsen. Automated event detection and classification in soccer: The potential of using multiple modalities. *Machine Learning and Knowledge Extraction*, 3:1030–1054, 12 2021.

[47] Saikat Sarkar, Dipti Prasad Mukherjee, and Amlan Chakrabarti. Reinforcement learning for pass detection and generation of possession statistics in soccer. *IEEE Transactions on Cognitive and Developmental Systems*, 15(2):914–924, 2023.

[48] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry, 2017.

[49] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[50] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

[51] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019.

[52] Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition, 2016.

[53] Silvio Giancola et al. Soccernet: A scalable dataset for action spotting in soccer videos. `https://www.soccer-net.org`, 2018.

[54] John Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. 1999.

[55] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[56] Grigorios Tsagkatakis, Mustafa Jaber, and Panagiotis Tsakalides. Goal!! event detection in sports video. *Electronic Imaging*, 2017:15–20, 01 2017.

[57] Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Relational pooling for graph representations, 2019.