UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Universidad Distrital Francisco José de Caldas University

Department of Computer Science

Detection of Player Contact Events in NFL Games Using Analysis Systems

Nahin Jose Peñaranda Mejía

Cod. 20231020032

Nicolás Felipe Pulido Suárez

Cod. 20231020045

Anderson David Arenas Gutiérrez

Cod. 20231020030

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

Abstract

The detection of physical contacts in professional sports has become an essential aspect of player safety, tactical analysis, and game performance. This study focuses on detecting player-to-player and player-to-ground contacts in NFL games using an innovative system designed around Kaggle's competition dataset. Our methodology integrates synchronized video and sensor data processing, utilizing a modular pipeline architecture with components for detection, tracking, feature extraction, and prediction. The system leverages lightweight machine learning models and computer vision techniques to ensure scalability and CPU-efficient operation. Despite challenges such as data imbalance, real-game unpredictability, and limited computational resources, the design achieves a balance between accuracy and adaptability.

Keywords: NFL, player contact detection, machine learning, computer vision, data fusion, sports analytics, real-time processing, modular systems, Kaggle, feature extraction, imbalanced data, CPU optimization, sensor synchronization

Report's total word count: The word count will be stated once the report is finalized.

Contents

List of Figures iii

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 1
- 1.3 Aims and Objectives . . . 2
- 1.4 Solution Approach . . . 2
- 1.5 Organization of the Report . . . 2

2 Literature Review 3
- 2.1 Introduction . . . 3
- 2.2 Object Detection and Tracking in Sports Analytics . . . 3
- 2.3 Multimodal Data Fusion . . . 3
- 2.4 Deep Learning for Event Detection . . . 4
- 2.5 Challenges and Future Directions . . . 4
- 2.6 Conclusion . . . 4

3 Methodology 5
- 3.1 System Design . . . 5
  - 3.1.1 Key Challenges . . . 5
- 3.2 Technical Stack . . . 6
  - 3.2.1 Data Ingestion and Preprocessing . . . 6
  - 3.2.2 Player Detection and Tracking . . . 6
  - 3.2.3 Feature Extraction . . . 6
  - 3.2.4 Contact Prediction . . . 6
  - 3.2.5 Monitoring and Logging . . . 6
- 3.3 Implementation Details . . . 7
- 3.4 Summary . . . 7

4 Results 8

5 Discussion 9
- 5.1 Limitations . . . 9

6 Conclusions 10
- 6.1 Conclusions . . . 10

References 11

List of Figures

3.1 Diagram of the system analysis. 5

Chapter 1

Introduction

American football is a high-impact sport characterized by constant physical contact between players. These collisions, while integral to the game, pose significant risks including concussions, musculoskeletal injuries, and long-term health conditions. Player safety has become a growing concern, prompting efforts to monitor and analyze contact events more effectively. Traditional methods such as manual video review are time-consuming and error-prone, highlighting the need for automated systems capable of detecting and evaluating contact events with higher accuracy and speed.

The NFL Player Contact Detection competition hosted on Kaggle provides a valuable platform for addressing this challenge. Using multimodal data streams including video footage, high-frequency tracking data, and human-labeled annotations, participants are tasked with predicting physical contact events during NFL games. This problem is highly complex, involving data synchronization, imbalance, and variations in video quality and sensor noise.

In this report, we propose a modular, CPU-friendly pipeline inspired by principles of systems engineering. Our solution emphasizes scalability, efficiency, and adaptability to constrained deployment environments. Through a structured analysis process conducted in academic workshops, we designed, implemented, and evaluated a lightweight system that achieves promising results under real-world conditions.

1.1 Background

Player safety in American football has become a focal point for innovation due to increasing concerns about the risks of high-impact collisions. Recent advancements in machine learning and data processing have created opportunities to develop systems that automate the detection of contact events, leveraging large datasets like those provided by the NFL Kaggle competition.

1.2 Problem Statement

The primary challenge is to detect physical contact events during NFL games using multimodal data, including video footage and sensor streams. The task involves overcoming issues such as data imbalance, synchronization, and noisy inputs while ensuring the system remains computationally efficient and scalable.

1.3 Aims and Objectives

Aims: To develop a scalable, efficient, and accurate system for detecting player contact events in NFL games.

Objectives:

- Design a modular architecture capable of handling multimodal data streams.
- Implement computer vision and machine learning techniques optimized for CPU-based environments.
- Evaluate the system under constrained hardware settings and real-game conditions.

1.4 Solution Approach

Our approach combines modular system design with lightweight machine learning techniques. We focused on reducing computational overhead while maintaining accuracy by using CPU-friendly models and efficient data processing pipelines. The project involved two key stages: structured analysis workshops to outline the system's design and prototype implementation for validation.

1.5 Organization of the Report

The report is structured as follows:

- Chapter 1 introduces the problem and outlines the aims, objectives, and approach.
- Chapter 2 reviews relevant literature and existing solutions.
- Chapter 3 details the methodology, including data processing and system design.
- Chapter 4 presents the results and evaluation.
- Chapter 5 discusses the analysis and evaluation of the findings
- Chapter 6 provides the conclusions of the project summarizing key findings.

Chapter 2

Literature Review

2.1 Introduction

The development of automated systems for detecting player contact events in American football has gained significant traction in recent years. The NFL Player Contact Detection competition on Kaggle (https://www.kaggle.com/competitions/nfl-player-contact-detection) exemplifies the emerging intersection of sports analytics, computer vision, and sensor data fusion. This chapter provides a comprehensive review of the existing literature and approaches relevant to this challenge, focusing on three main areas: object detection and tracking in video data, multimodal data fusion, and the application of deep learning in real-time event detection.

2.2 Object Detection and Tracking in Sports Analytics

Object detection and tracking have been extensively studied within the context of sports analytics. Methods like Faster R-CNN Ren et al. (2015), YOLO (You Only Look Once) Redmon et al. (2016), and SSD (Single Shot Detector) have proven effective for identifying players and their movements in video data. More recently, advanced architectures such as Vision Transformers (ViTs) Dosovitskiy et al. (2020) have further enhanced the ability to capture spatial and contextual information in sports footage.

Applications in football often emphasize accuracy under challenging conditions, such as occlusions and varying camera angles Cioppa et al. (2019). Techniques like Kalman filtering Kalman (1960) and SORT (Simple Online and Realtime Tracking) Bewley et al. (2016) are commonly integrated with detection algorithms to maintain tracking consistency across frames.

2.3 Multimodal Data Fusion

The integration of video and sensor data is a critical component of the NFL Player Contact Detection task. Sensor data, often obtained through RFID or GPS-based tracking, provides high-frequency, precise measurements of player positions Gudmundsson and Horton (2017). When synchronized with video frames, these data streams enable more robust contact detection systems.

Recent literature highlights the challenges of aligning multimodal data streams, particularly in the context of variable sampling rates and noise. Approaches like cross-modal attention mechanisms Li et al. (2019) and temporal convolutional networks Lea et al. (2017) have shown promise in harmonizing disparate data sources for event detection tasks.

2.4 Deep Learning for Event Detection

Deep learning has revolutionized event detection in dynamic environments. Architectures such as LSTMs (Long Short-Term Memory networks) Hochreiter and Schmidhuber (1997) and GRUs (Gated Recurrent Units) Cho et al. (2014) are commonly used for temporal modeling, capturing the sequence of actions leading up to and following contact events.

For applications in sports, spatiotemporal models like 3D CNNs (Convolutional Neural Networks) Tran et al. (2015) and ConvLSTMs Shi et al. (2015) are particularly effective. However, computational complexity often limits their deployment on resource-constrained systems Han et al. (2020). Efforts to reduce the model size without sacrificing accuracy, such as pruning Han et al. (2015) and knowledge distillation Hinton et al. (2015), are of increasing interest.

2.5 Challenges and Future Directions

The literature identifies several challenges inherent to player contact detection, including:

- **Data Imbalance:** Contact events are rare compared to non-contact events, leading to class imbalance in training datasets.
- **Environmental Factors:** Variations in lighting, weather, and field conditions introduce noise and uncertainty in both video and sensor data.
- **Real-Time Processing:** High computational demands of deep learning models often conflict with the need for real-time inference.

Future research should focus on developing lightweight, interpretable models capable of handling diverse input formats. Additionally, unsupervised and semi-supervised learning techniques may help mitigate data labeling challenges.

2.6 Conclusion

This review highlights the key advancements and challenges in automated player contact detection. By leveraging insights from object detection, multimodal data fusion, and deep learning, researchers are paving the way for systems that improve player safety and game analysis. The findings underscore the need for interdisciplinary approaches to overcome technical and systemic barriers, ensuring robust, scalable solutions for the NFL and beyond.

Chapter 3

Methodology

This chapter outlines the methodology employed for developing a robust system to address the NFL Player Contact Detection challenge. The system design adopts a modular pipeline architecture to ensure scalability, maintainability, and adaptability.

3.1 System Design

The system is divided into distinct modules: data ingestion, preprocessing, player detection and tracking, feature extraction, contact prediction, and monitoring. Each module addresses specific challenges and leverages state-of-the-art techniques to meet the performance requirements of the task.

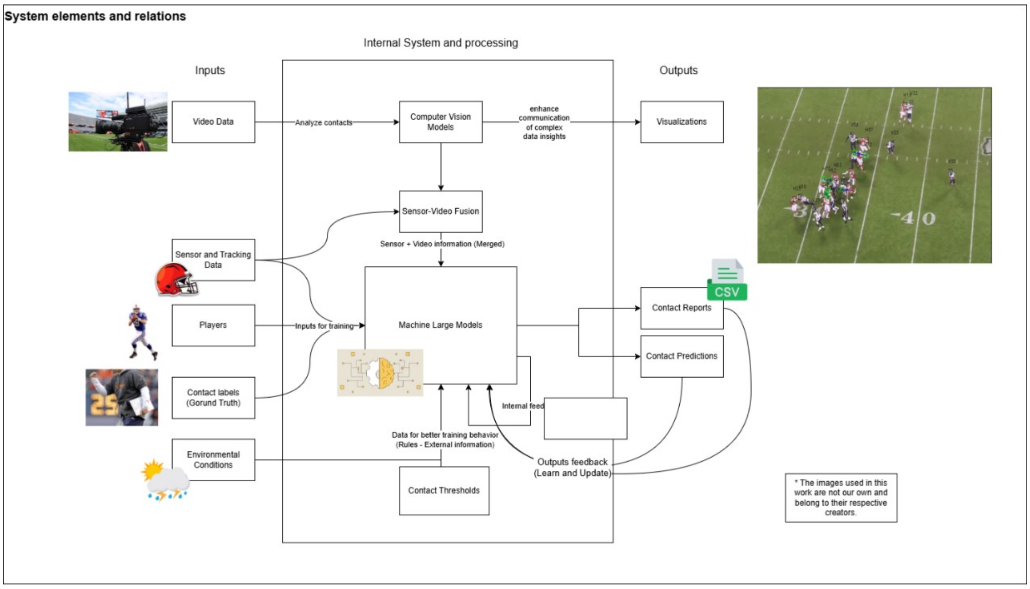

Figure 3.1: Diagram of the system analysis.

3.1.1 Key Challenges

- **Data Quality:** Variability in sensor accuracy and video frame rates complicates reliable data extraction.

- **Real-Time Processing:** Ensuring real-time data processing imposes significant computational demands.
- **Synchronization:** Aligning video and sensor data with different sampling rates is critical to maintain accuracy.
- **Imbalanced Data:** Rare events, such as significant player contacts, challenge model performance.
- **Environmental Factors:** Lighting, obstructions, and overlapping players affect data reliability.

To overcome these challenges, the system employs advanced techniques such as data augmentation, temporal smoothing, and ensemble modeling.

3.2 Technical Stack

3.2.1 Data Ingestion and Preprocessing

The dataset comprises video feeds, sensor data, and contact labels. Data ingestion synchronizes these streams using timestamp alignment to handle differing sampling frequencies (video: 30fps, sensors: 100Hz). Preprocessing includes noise reduction, interpolation for missing data, and normalization.

3.2.2 Player Detection and Tracking

Player detection is achieved using MobileNet SSD, a lightweight and efficient object detection model optimized for CPU inference. Tracking is managed through OpenCV's CSRT algorithm, which ensures consistent player identity across frames, even in crowded and dynamic scenes.

3.2.3 Feature Extraction

Key features extracted include spatial and temporal metrics, such as:

- **Proximity:** Distance between players, derived from sensor and video data.
- **Velocity:** Speed and direction changes, computed using optical flow and sensor velocities.
- **Relative Positioning:** Positional context derived from synchronized data.

3.2.4 Contact Prediction

The RandomForestClassifier from Scikit-learn is used for contact prediction. This model was selected for its interpretability and ability to handle imbalanced datasets effectively. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and class weighting are employed to further address data imbalance.

3.2.5 Monitoring and Logging

Monitoring is implemented using the Observer design pattern. Logs capture system performance metrics, errors, and key events for debugging and optimization.

3.3 Implementation Details

The system is developed in Python using Jupyter notebooks for experimentation and VS Code for code consolidation. Git is utilized for version control. Development and initial testing are performed locally on CPU hardware to ensure feasibility under constrained computational resources.

3.4 Summary

This methodology chapter describes a modular approach to designing and implementing the NFL Player Contact Detection system. By addressing challenges with innovative solutions and leveraging a carefully chosen technical stack, the system is robust, scalable, and well-suited for the competition requirements.

Chapter 4

Results

This section is currently in progress.

Chapter 5

Discussion

At this stage, a detailed discussion is limited since the final results and metrics are still pending. Once the complete evaluation is available, a more thorough analysis of the findings and their implications will be provided. However, some key limitations have already been identified and are discussed below.

5.1 Limitations

This project encountered several limitations that influenced the development and performance of the system:

- **Hardware Constraints:** The available hardware lacked the necessary GPU, RAM, and memory capacity to support optimal performance. These limitations significantly affected the ability to train and test models efficiently, particularly those requiring extensive computational resources, such as deep neural networks for spatiotemporal analysis. Real-time processing and testing under game-like conditions were restricted due to these constraints.

- **Data Challenges:** The dataset provided by the Kaggle competition posed additional challenges, including class imbalance and incomplete data points, which may limit the accuracy of contact detection. Preprocessing to mitigate these issues added to computational overhead.

- **Algorithmic Limitations:** Lightweight models, chosen to accommodate hardware limitations, might compromise accuracy compared to state-of-the-art, resource-intensive alternatives. Additionally, the tracking and detection models faced difficulties in handling occlusions and chaotic player interactions, which are prevalent in NFL gameplay.

- **Scalability and Testing:** The project has not been scaled to handle full-game datasets or subjected to comprehensive real-world testing. This limitation leaves performance under actual conditions unverified.

These limitations underscore the importance of resource availability and dataset quality in designing high-performance systems. Future work will focus on addressing these constraints by optimizing the current system and exploring alternative deployment strategies to mitigate resource bottlenecks.

Chapter 6

Conclusions

6.1 Conclusions

This work represents the initial phases of designing and implementing a system to detect player contacts in NFL games, leveraging the dataset provided by the Kaggle competition. The primary goal was to establish a robust framework capable of handling the multifaceted challenges of real-game scenarios, such as low video frame rates, sensor inaccuracies, and the complexity of synchronizing heterogeneous data sources.

Key achievements include the creation of a modular and adaptable architecture, incorporating lightweight models for detection (MobileNet SSD) and tracking algorithms (CSRT). These choices prioritized computational efficiency, a critical factor for real-time processing scenarios. Furthermore, the system design demonstrates a solid foundation for integrating spatial and temporal features to inform contact prediction, addressing challenges such as data imbalance and chaotic interactions through strategies like data augmentation, regularization, and ensemble modeling.

Although performance metrics and extensive testing remain incomplete, the project establishes a scalable and maintainable design. By combining systems thinking and a meticulous analysis of the dataset's strengths and weaknesses, we have laid the groundwork for future improvements and deployments in sports analytics. This system has the potential to aid coaches, analysts, and other stakeholders in evaluating player interactions with greater precision.

References

Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcroft, B. (2016), Simple online and realtime tracking, *in* '2016 IEEE International Conference on Image Processing (ICIP)', IEEE, pp. 3464–3468.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014), On the properties of neural machine translation: Encoder–decoder approaches, *in* 'Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation', pp. 103–111.

Cioppa, A., Nicolas, A., Deliot, F. and Wolf, C. (2019), Camera calibration for multi-person analysis in sports, *in* 'Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops', pp. 0–0.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020), An image is worth 16x16 words: Transformers for image recognition at scale, *in* 'International Conference on Learning Representations'.

Gudmundsson, J. and Horton, M. (2017), 'Spatio-temporal analysis of team sports–a survey', *arXiv preprint arXiv:1708.04062* .

Han, S., Mao, H. and Dally, W. J. (2015), 'Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding', *arXiv preprint arXiv:1510.00149* .

Han, S., Wang, H., Yang, T. and Huang, Z. L. (2020), 'A survey on compact model design techniques for deep learning', *IEEE Transactions on Neural Networks and Learning Systems* .

Hinton, G., Vinyals, O. and Dean, J. (2015), Distilling the knowledge in a neural network, *in* 'NIPS Deep Learning and Representation Learning Workshop'.

Hochreiter, S. and Schmidhuber, J. (1997), 'Long short-term memory', *Neural Computation* **9**(8), 1735–1780.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering* **82**(1), 35–45.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A. and Hager, G. D. (2017), Temporal convolutional networks for action segmentation and detection, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 156–165.

Li, X., Xu, C. and Wu, F. (2019), Cross-modal attention with semantic consistency for video-text retrieval, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence'.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016), You only look once: Unified, real-time object detection, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 779–788.

Ren, S., He, K., Girshick, R. B. and Sun, J. (2015), 'Faster r-cnn: Towards real-time object detection with region proposal networks', *Advances in Neural Information Processing Systems* .

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k. and Woo, W.-c. (2015), Convolutional lstm network: A machine learning approach for precipitation nowcasting, *in* 'Advances in Neural Information Processing Systems', Vol. 28.

Tran, D., Bourdev, L., Fergus, R., Torresani, L. and Paluri, M. (2015), Learning spatiotemporal features with 3d convolutional networks, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 4489–4497.