jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 3, PP. 1–8
Manuscript received February 5, 2025; revised June 2025. doi: 10.24425/ijet.2025.153617

Deep Learning in motion analysis for false start detection in speedway racing

Jacek Krakowian, and Łukasz Jeleń

***Abstract*—Accurately identifying false starts in speedway racing is a very challenging task due to the subtle nature of pre-start movements. Manual detection methods, often dependent on the judgment of race officials, are prone to errors and subjectivity, leading to inconsistencies in decision-making. This paper introduces an automated approach that leverages computer vision methods to enhance detection precision. Here, we have expanded its use to detect false starts in speedway racing. The proposed approach introduces image processing techniques with 3D Convolutional Neural Networks (CNNs) and Long-Short-Term Memory (LSTM) networks to analyze rider movements during the starting procedure. Unlike manual detection, which often misses fine movements at the start line, our method uses 3D CNNs to monitor racer movements and applies LSTM networks to assess time-based motion patterns that signal false starts. The presented results show that the 3D CNN achieved an accuracy of 86.36% with a higher precision when compared to traditional methods. This automated process not only enhances fairness in competitive racing, but also illustrates the broader capability of emerging technologies to refine decision-making in sports.**

***Keywords*—False start detection; 3D Convolutional Neural Networks (3D CNNs); Long Short-Term Memory (LSTM) networks; motion analysis; speedway racing**

I. Introduction

SPEEDWAY racing is a type of motorcycle racing where riders compete on oval dirt tracks at very high speeds. Races usually consist of short laps and riders must maneuver their bikes through tight turns without brakes, relying on throttle management and skills. This fast-paced sport requires precision especially at the starting line, where even the slightest premature movement can not only have a major impact on race outcomes, but can give racers an unfair advantage. Detection of these movements remains a significant challenge in speedway racing, making accurate false start detection essential for maintaining the integrity of the sport.

Historically, false start detection in speedway has been managed by officials using television video footage. Such false start recognition heavily depend on human judgment, introducing variability and the risk of mistakes. Race officials often depend on their ability to visually detect subtle movements or analyze video footage, but this manual process is inherently limited by the events' rapid pace and complexity. As a result, false start decisions can be inconsistent as human observations frequently lacks the precision needed to capture small, yet significant, movements.

Over the years, several approaches have been developed to address the limitations of manual detection [1]–[3]. One common method involves the use of high-speed cameras and is widely used in many racing environments [2]. This method allows officials to slow down footage and review frame-by-frame details of the start. Such an approach enhances the ability to detect subtle false start movements but it is still a manual process that depends on the experience and interpretation of the official reviewing the footage and can result in reaching different conclusions based on the same video evidence by two different referees.

Another approach involves sensor-based systems that detect physical movements or changes in position before official start. Here, we can distinguish pressure-sensitive start gates that detect racer-induced force and accelerometers mounted on vehicles that measure sudden movements [4]. Although these technologies can provide better precision, the implementation of such sensor systems requires significant infrastructure changes and can be expensive to implement on a large scale. Recent developments in the field of computer vision and machine learning have emerged as leading solutions to automating complex detection tasks, including false start detection [5], [6]. Furthermore, literature review shows that the application of these techniques in various sports and safety-related domains is an active research field. These technologies have been used to analyze the performance of the skeleton push start in winter sports [7] and to detect service height faults in badminton [8]. In motorsports, fast video image detection has been implemented for racing cars using TensorFlow Mobile Networks [9]. Other applications include advanced video analysis for a variety of applications, such as deception detection and facial micro-movement analysis [10]. These technologies have also shown promising results in tasks such as detecting players and balls, tracking movements, and recognizing strategies in a range of different sport fields.

In the field of speedway, one of the crucial moments of the race is fair start, and therefore accurate false start detection is crucial for maintaining the integrity of the sport. Using advanced vision-based algorithms, such as those seen in accident detection systems, there is potential to significantly improve

Jacek Krakowian is with the Foundation for the Support of Innovation, Research and Development Wrocław, Poland (e-mail: jacek.krakowian@wibir.org);

Łukasz Jeleń is with the Department of Computer Engineering, Wrocław University of Science and Technology, Wrocław, Poland (e-mail: lukasz.jelen@pwr.edu.pl).

the reliability of false start identification in racing environments [11], [12]. These methods greatly improve our ability to detect fast, subtle motions that may not be visible to the naked eye. Moreover, advanced machine learning models, such as 3D Convolutional Neural Networks (CNNs), can now process spatial data from image sequences and video data, allowing real-time tracking of racer movements with high accuracy [13]. These networks extend 2D Convolutional Neural Networks by adding a temporal dimension, allowing them to capture both spatial and temporal information [13], [14]. 3D CNNs have shown superior performance in various applications, including human action recognition [15], [16], medical imaging [17], silent speech interfaces [18], video super-resolution [19], and gesture recognition using high-density surface electromyography [20]. These networks can process raw input data without the need for complex handcrafted features, making them particularly effective for tasks involving motion analysis and spatio-temporal pattern recognition. Because of their ability to capture the spatial dynamics of movements, these models seem to be very beneficial in identifying slight changes that indicate a potential false start. Furthermore, long-short-term memory (LSTM) networks can capture spatial dependencies and contextual information in images [21], making them suitable for analyzing temporal sequences and therefore ideal for monitoring motion patterns over time. LSTMs can effectively capture spatial dependencies and contextual information in images [22]. They have been successfully applied to scene labeling [22], action recognition in videos [23], and human re-identification [24]. These networks can process images as sequences of pixels or high-level features extracted using convolutional neural networks [25] [26]. In speedway, LSTMs can be used to predict whether a series of movements is consistent with a legitimate start or indicative of a false start based on historical data [6].

This paper presents a novel approach that leverages the strengths of both 3D CNNs and LSTM networks to overcome the difficulties of false start detection in speedway racing. By combining spatial and temporal analysis, the proposed system aims to deliver an automated solution that reduces human error and ensures more consistent and objective results.
The results presented in this paper can be beneficial beyond speedway racing, as in many sports, false start detection, and other motion-based judgments remain reliant on human observers or basic timing devices. Recently, one can observe a growth of demands for accuracy and fairness in sports. Therefore, the need for advanced automated systems is becoming more apparent. Implementing machine learning-driven detection tools like the one proposed in this paper has the potential to establish a new standard for motion analysis in competitive sports. Such systems could provide referees and officials with tools they could use to make faster and more precise decisions.

In subsequent sections, a cutting-edge approach is described that integrates the latest advances in machine learning and computer vision to tackle the problem of false start detection in speedway racing. By automating false-start detection, we aim to bring a new level of fairness, consistency, and technological advancement to the world of competitive racing.

II. Materials and Methods

In this section, materials used and methods for false start detection are described. This study focuses on a combination of frame difference algorithms, 3D Convolutional Neural Networks (3D CNNs), and long-short-term memory (LSTM) networks. Video footage was collected and preprocessed to enhance image quality, followed by the application of a frame difference algorithm to identify motion. 3D CNNs were then used to extract spatio-temporal features from the video sequences, enabling the detection of subtle racer movements. LSTM networks were used to analyze the temporal progression of these movements, distinguishing false starts from valid ones [27] [28] [29] [30].

In speedway racing, a false start occurs when a rider moves from their starting position before the official start signal is given, typically indicated by the rise of the starting tapes. Riders are required to remain stationary at the starting line until the signal is received, but due to the high tension and quick reaction times involved, even the slightest movement before the tapes rise can lead to a false start. This premature movement gives the rider an unfair advantage by allowing them to begin accelerating ahead of their competitors. When a false start is detected, the race is usually stopped, and the offending rider receives a waning or is disqualified if his bike touches the tape. As mentioned previously, detecting false starts manually can be challenging due to the subtlety and speed of the movements, which is why automated detection systems might be important in ensuring race integrity.

A. Speedway False Start Detection Dataset Description

To perform the detection of rider movements that are the indication of false starts in speedway racing, we have prepared a database comprising high-resolution video sequences. The database is thoroughly designed to capture and highlight subtle movements that precede a false start, providing a robust foundation for developing and testing detection methodologies focused on subtle movement analysis. The data set collected comes from two different sources. The primary source comes from official speedway world championships that were made publicly available on the Official Speedway GP YouTube Channel. In addition, we recorded videos during practice sessions at the Opole Speedway track. Official videos were recorded at a frame rate of 30 frames per second (FPS) and a resolution of 960 × 540 pixels, utilizing Open Broadcaster Software (OBS) to ensure consistent quality and format. We conducted a thorough investigation of 20 real-life video streams to identify rider movements during the starting procedures. From these streams, we extracted 75 distinct starts with an average duration of 2 seconds, of which in 64 a false start occurred, and in 11 of them the procedure was misclassified. The misclassification is the situation in which the rider movement was visible but the procedure was not stopped. Supplemental data was recorded in a controlled environment during practice. Here, a total of 40 test videos were recorded at a frame rate of 30 FPS and a resolution of 1920 x 1080 pixels. These videos were used to verify the correctness and robustness of the proposed false start detection method.

The collected database is a valuable collection of annotated video data for false start detection algorithms in speedway racing. By including both competition and practice videos, the dataset provides a diverse range of scenarios to improve detection accuracy in real-world applications.

B. Research Methodology

In the described research methodology, we adopt a comprehensive approach to evaluate techniques for false start detection based on subtle rider movements that precede a false start. The methodology is based on insights gathered from the literature and the specific requirements of speedway racing. Our approach integrates a series of computer vision algorithms such as Long Short-Term Memory Networks and 3D Convolutional Neural Networks. The proposed networks were compared with traditional movement detection methods based on frame differences and the Gaussian mixture model.

1) **Frame Difference Methods:** Frame difference methods belong to one of the simplest and most computationally efficient techniques for detecting motion in video sequences. These methods work by comparing the pixel values of consecutive video frames to detect changes, which typically correspond to moving objects in the scene [31]. In this study, we have compared three variations of this method: a two-frame difference (TFD), a three-frame difference (ThFD), and a custom two-frame difference (CTFD). These variations offer different balances between motion sensitivity and noise robustness.

In the two-frame difference method, the difference between two consecutive frames is computed. If the absolute difference between the corresponding pixels exceeds a predefined threshold, it is assumed that motion has occurred in that region. Mathematically, the difference for a frame at time t can be represented as:

$$D_t(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \tag{1}$$

where $I_t(x, y)$ and $I_{t-1}(x, y)$ represent the pixel intensities at position (x, y) in the frames at time t and $t-1$ respectively. The advantage of this method is its ability to quickly highlight regions with sudden changes in the scene. However, it may struggle with noisy video sequences, as small fluctuations in lighting or camera noise can trigger false positives. Additionally, fast-moving objects may leave gaps or be undersegmented since only the difference between two frames is considered.

The three-frame difference method extends the two-frame difference by incorporating an additional frame into the comparison [32]. It calculates the difference between the current frame and both the previous and next frames, as shown by:

$$D_t(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \cup |I_t(x, y) - I_{t+1}(x, y)| \tag{2}$$

By comparing the current frame with both adjacent frames, this method improves the detection of continuous motion and reduces sensitivity to noise. It helps eliminate some of the flickering issues present in the two-frame method, as it considers changes over a wider temporal window. This makes it more reliable in detecting gradual movements or objects that are moving slowly, at the cost of a slight increase in computational complexity.

The custom two-frame difference is our modification of a traditional two-frame difference algorithm that takes two video frames (F_k) and (F_{k+n}) as input, where k is the frame number, $(k + n)$ is the number of the next frame, and n represents the frame step. Unlike the standard approach, the proposed algorithm does not restrict the input to two consecutive frames, but also calculates a difference between the foreground and background values. In the proposed algorithm, the difference between the foreground and background is calculated using a threshold applied to a blurred version of the input frames at gray level. Motion is detected if the subtraction of binary foreground values from the background yields a result less than zero. This approach improves detection flexibility by allowing the comparison of non-consecutive frames and enhances accuracy in motion detection by incorporating background subtraction with thresholded, binary images.

2) Gaussian Mixture Model: A Gaussian Mixture Model (GMM) is a probabilistic model that represents the presence of multiple Gaussian distributions in a dataset. This method is often used in motion detection to model the background of a scene, where each pixel's intensity is modeled as a mixture of several Gaussian distributions [33]. This technique allows for handling complex and dynamic backgrounds, such as moving trees, rippling water, or varying lighting conditions. In motion detection, the primary task is to separate moving foreground objects from a static or semi-static background. GMM tackles this by assuming that the pixel intensities of the background can vary and are best represented by a mixture of Gaussian distributions [34]. Each pixel in the video frame is modeled as a combination of K Gaussian distributions. The probability of observing a pixel intensity X_t at a location (x, y) at time t is given by the mixture model described by equation 3.

$$P(X_t(x, y)) = \sum_{i=1}^{K} w_i \cdot \mathcal{N}(X_t(x, y)|\mu_i, \sigma_i^2) \tag{3}$$

where w_1 is the weight (or probability) associated with the $i-th$ Gaussian distribution and $\mathcal{N}(X_t(x, y)|\mu_i, \sigma_i^2)$ represents the $i-th$ Gaussian distribution with mean μ and variance σ_i^2. Initially, each pixel in the scene is modeled by a set of K Gaussian distributions with random mean values and variances. As more video frames are processed, the model parameters are updated to reflect the pixel's evolving background behavior. For each new pixel intensity the algorithm checks whether it fits any of the existing Gaussian distributions within a threshold determined with a Mahalanobis distance. To detect motion, the GMM method examines each pixel and determines whether its intensity value fits any of the existing background models. If the pixel value does not match any Gaussian model representing the background, it is classified as foreground, which is considered as motion.

3) Long Short-Term Memory Networks: Long-Short-Term Memory networks are a type of recurrent neural network (RNN) designed to handle sequential data to overcome the limitations of traditional RNNs, such as the problem of vanishing and exploding gradients [35]. These networks are particularly effective in learning long-term dependencies in the data, making them suitable for time series tasks, including video motion detection.

The architecture of LSTMs is consists of memory cells that control the flow of information, enabling the model to retain, forget, or update information as needed. Each LSTM unit consists of the input gate that controls how much of the new information from the current input should be written to the cell state, the forget gate that decides which parts of the cell state to forget, and the output gate that regulates the amount of information from the cell state that should be output to the next time step.

In video-based motion detection, these networks can be used to capture temporal dependencies in the sequence of video frames. Unlike traditional frame-difference methods, which only analyze two or three consecutive frames, LSTMs have the ability to learn from extended sequences of frames and extract motion patterns over time. This makes them particularly effective for detecting complex movements, tracking objects, or identifying subtle changes in the scene that evolve over time. By processing such time sequences, LSTMs can predict whether motion is occurring in the current frame or provide continuous tracking of moving objects [35].

4) 3D Convolutional Neural Networks: 3D Convolutional Neural Networks (3D CNNs) extend the concept of traditional convolutional neural networks by applying convolutional operations in three dimensions: height, width, and time. This makes them particularly powerful for tasks involving spatio-temporal data, such as video analysis and motion detection. Unlike 2D CNNs, which are limited to processing spatial features of individual frames, 3D CNNs capture both spatial and temporal features by convolving multiple consecutive video frames [14].

Typically in 3D CNNs, the input consists of a sequence of video frames represented as a 3D tensor of dimensions (T, H, W, C), where T is the number of frames (temporal depth), H and W are the height and width of the frame, and C is the number of channels. A 3D convolutional layer applies filters that have a shape of (d_t, d_h, d_w, C), where d_t represents the temporal depth and d_h and d_w represent the spatial dimensions of the filter. The output of a 3D convolutional operation is another 3D feature map, which allows the network to capture spatio-temporal patterns across consecutive frames [36].

By processing a stack of consecutive video frames, 3D CNNs can learn to identify complex motion patterns that span across multiple frames. For example, they can detect activities such as walking, running, or jumping by understanding how objects move over time. This makes 3D CNNs particularly effective for action recognition, human behavior analysis, and object tracking in videos. Compared to traditional frame-based motion detection methods, 3D CNNs can model both short-term and long-term dependencies in video data, resulting in more accurate and robust motion detection.

In conclusion, the materials and methods described in this section provide a solid framework for detecting false starts in speedway racing. The proposed combination of methods and analysis, such as frame difference methods, Gaussian mixture models, long-short-term memory networks, and 3D convolutional neural networks, can capture and classify critical moments at the start of each race with great reliability. For comparison purposes, we have implemented both traditional and advanced machine learning approaches to capture and analyze crucial spatio-temporal patterns associated with rider movement. These methods allow for precise detection of motion irregularities that signal a false start that can be used in a reliable system to monitor and detect false starts in speedway competitions.

III. Results

Detecting false starts in speedway racing presents several challenges, including unpredictable environmental conditions, the minimal distance that separates riders, and the need for precise timing to distinguish between legitimate and false movements at the start line. Traditional human-based assessments are often error-prone, given the high stakes and pressure associated with these starting sequences. In response to this problem, our proposed methodology aims to provide a more accurate and automated approach to analyzing image sequences for false start identification. In this section, we present the results obtained from implementing various methods for false start detection in speedway racing. By applying frame difference methods, Gaussian mixture models (GMMs), Long- and short-term memory (LSTM) networks, and 3D Convolutional Neural Networks (3D CNNs), we evaluated the system's ability to detect them. Each method was assessed based on its accuracy and robustness to environmental variations. The results demonstrate the effectiveness of these approaches in identifying early motion patterns and distinguishing legitimate starts from false ones, providing a comprehensive evaluation of their performance under real-world racing conditions. The performance of the algorithms was evaluated on the videos from the collected database and compared with traditional human assessments to emphasize their comparative strengths and potential areas for enhancement.

The frame difference and Gaussian mixture methods were first applied to the practice sequences of the Opole track and the laboratory setup. Here, all the methods showed very good results reaching 99% accuracy for subtle movement detection. This level of precision arises from focusing solely on one object against a fairly consistent background. In the next step, the algorithms were applied to the 70 database image sequences from official video footage. An example of the results obtained is shown in Fig. 1 and the detected motion is marked in green. The comparison of the results demonstrate that the TFD method was able to detect the movement of only one rider (see Fig. 1a) while the proposed CFTD method identified the front and back wheels of a red helmet rider as well as an advertisement rotation (see Fig. 1b). The clear motion is also visible in Fig. 1c that shows an example of a ThFD method. Here, a motion of additional bike parts was detected, such as the clutch, spokes, and a part of the engine. The last method, namely GMM algorithm was bale to detect the movement of a yellow rider with his front wheel (Fig. 1d).

In the remaining experiments, we tested the possibility of detecting false starts based on subtle rider movements with long-short-term memory and 3D convolutional neural networks. The LSTM was implemented due to its proficiency in handling sequential data and capturing temporal dependencies. The network was trained with the annotated data set comprising the official competition focusing on sequences where minimal pre-start motions were critical for detection accuracy. Upon evaluation, the LSTM network achieved an overall accuracy of 69.21% in correctly identifying false starts. This performance reflects the model's ability to recognize and learn from the temporal patterns associated with subtle rider movements preceding a false start. The accuracy indicates a moderate level of effectiveness, suggesting that while the LSTM network can capture essential temporal features in the data, there is room for improvement. 3D CNN, on the other hand, is particularly well-suited for video analysis as it processes consecutive frames simultaneously, allowing it to learn motion patterns over time. The network was trained using the same dataset and the evaluation showed an overall precision of 76. 19%. This performance marks a significant improvement not only in the accuracy of the LSTM network but also over the GMM and frame-difference methods. The increased accuracy suggests that the 3D-CNN is more effective in differentiating between normal prerace activities and movements that constitute a false start. The complete results are summarized in Table I and the 3D CNN traing perfomance is shown in Fig. 2.

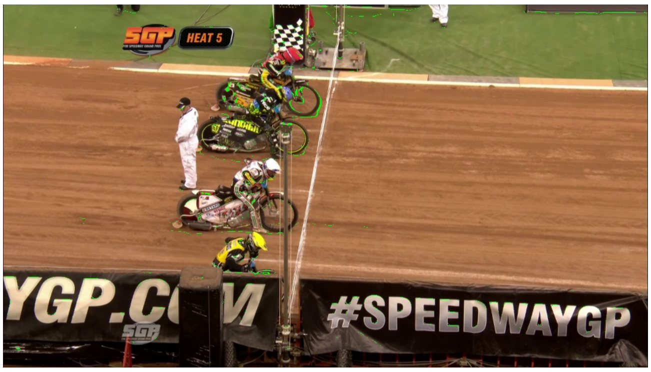
TABLE I. False start detection results.

Method	TFD	CTFD	ThFD	GMM	LSTM	3D CNN
Detected	70 %	65 %	50 %	75 %	69.21%	76.19%
Not detected	30 %	35 %	50 %	25 %	30.79%	23.81%

IV. Result Discussion

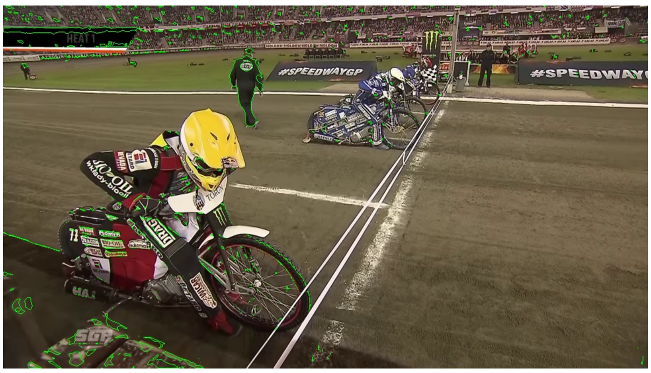
The primary objective of this study was to develop and evaluate machine learning models capable of detecting false starts in speedway racing by analyzing subtle movements in video sequences. The detection of such minimal movements is critical for ensuring fair competition, but presents significant challenges due to the nature of the motion and the potential noise in the data. We implemented two neural network architectures to address this challenge and compared their performance. Additionally, we compared the performance with conventional frame difference and Gaussian mixture methods.

(a) Two Frame Difference.



(b) Custom Two Frame Difference.

(c) Three Frame Difference.



(d) Gaussian Mixture Model.

Fig. 1. Motion Detection Results.

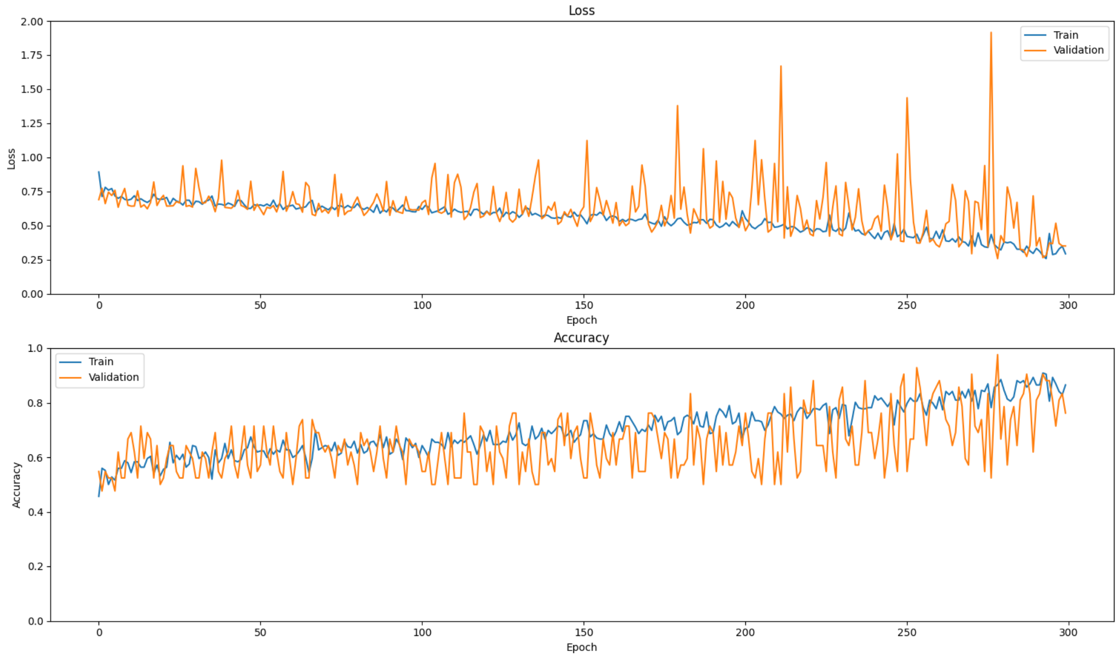


Fig. 2. Results of 3D Convolutional Neural Network for 300 epoch.

This study showed that employing advanced image processing and machine learning techniques we were able to accurately detect false starts despite the major challenges that lie in the close proximity of the racers and the critical timing precision required at the start.
The results presented in this study demonstrate that the described techniques substantially surpass manual detection to accurately identify false starts. On average, 20% of the referee misclassified false starts were correctly classified with the proposed framework. Furthermore, the results show that 3D Convolutional Neural Network networks outperform all other methods and proved to be advantageous in classifying movements at the start line, capturing even subtle and progressive shifts that are challenging for human referees. Although some of the simpler methods, such as the traditional two-frame difference (TFD) and the three-frame difference (ThFD), offered lower accuracy in the case of competitive event footage, their performance in controlled environments was nearly flawless. This suggests that these methods could be valuable in specific contexts, such as practice settings or situations with fewer visual complexities.

Overall, the application of neural network approaches provides a foundational approach for modeling temporal dynamics in rider movements and the combination of traditional motion detection algorithms with more sophisticated machine learning models presents a strong case for an automated false start detection system. The results clearly show that the proposed methods contribute valuable insights into the development of more robust false start detection that can support human referees in making critical decisions, thus improving fairness in speedway racing.

V. Conclusions

This study aimed to address challenges faced during the detection of small movements in video sequences for the detection of false starts by developing machine learning models capable of detecting these minimal pre-start motions using high-resolution video data. For this purpose, we curated a specialized dataset comprising annotated videos from official competitions and practice sessions, focusing on capturing the nuanced movements indicative of false starts. The database was then used to train and test several detection methods described in sec. II-B, including traditional image processing techniques and advanced machine learning models. The obtained results demonstrate the potential of advanced machine learning techniques, particularly 3D convolutional neural networks, in improving the detection of false starts in speedway racing through the analysis of subtle rider movements. Although traditional methods like GMM and frame differences provide a reasonable baseline, the enhanced performance of the 3D CNN indicates significant advantages in using deep learning models for this application. By achieving the highest detection rate of 76.19%, the 3D CNN showcases its effectiveness in capturing complex spatio-temporal patterns that are critical for accurate false start detection. The two-frame difference methods (TFD and CTFD) and the three-frame difference method (ThFD) yielded moderate detection rates, highlighting the limitations of relying solely on simple frame difference techniques. The presented outcomes have significant implications beyond speedway racing. Many other motorsports encounter similar challenges where precise movement recognition is important. Therefore, continued efforts are needed to refine these models and address current limitations to further enhance detection accuracy.

The findings from this study have not only demonstrated the potential of advanced detection methods but have also endorsed a clear course for future research. By comparing traditional image processing techniques with deep learning models, we have gained valuable insights into the strengths and limitations of each approach in detecting subtle rider movements. A key area for future work is the extension of the database. Expanding the dataset to include a larger number of videos from various competitions, tracks, and environmental conditions will enhance the robustness and generalizability of the detection models. Incorporating additional data, such as higher-resolution footage, different camera angles, and more instances of both false starts and correctly started races, will allow for more comprehensive training and validation of the models. Another promising direction for future work would be integrating these systems with existing sports technologies to create a more comprehensive framework for monitoring and analyzing competitive events.

In conclusion, the presented research showcases pathways to integrating advanced image analysis and machine learning to speedway racing by ensuring that false starts are identified with greater precision, ultimately promoting fair play and competitiveness in the sport.

References

[1] J. B. Lee, R. Mellifont, B. Burkett, and D. James, "Detection of illegal race walking: A tool to assist coaching and judging," *Sensors (Basel, Switzerland)*, vol. 13, pp. 16 065–16 074, 2013. [Online]. Available: https://doi.org/10.3390/s131216065

[2] J. Wu, K. Zhang, Y. Zhang, X. Xie, and G. Shi, "High-speed object tracking with dynamic vision sensor," *Proceedings of the 5th China High Resolution Earth Observation Conference (CHREOC 2018)*, 2018. [Online]. Available: https://doi.org/10.1007/978-981-13-6553-9_18

[3] K. L. Strobel, S. Zhu, R. Chang, and S. Koppula, "Accurate, low-latency visual perception for autonomous racing: Challenges, mechanisms, and practical solutions," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1969–1975, 2020. [Online]. Available: https://doi.org/10.1109/IROS45743.2020.9341683

[4] M. Brodie, A. Walmsley, and W. Page, "Fusion motion capture: A prototype system using inertial measurement units and gps for the biomechanical analysis of ski racing," *Sports Technology*, vol. 1, pp. 17 – 28, 2008. [Online]. Available: https://doi.org/10.1080/19346182.2008.9648447

[5] V. Wiley and T. Lucas, "Computer vision and image processing: A paper review," *International Journal of Artificial Intelligence Research*, 2018. [Online]. Available: https://doi.org/10.29099/ijair.v2i1.42

[6] J. Krakowian and Ł. Jeleń, "Preliminary study on the detection of subtle variations in image sequences for identifying false starts in speedway racing," in *System Dependability - Theory and Applications*, W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, and J. Kacprzyk, Eds. Cham: Springer Nature Switzerland, 2024, pp. 111–120. [Online]. Available: https://doi.org/10.1007/978-3-031-61857-4_11

[7] L. Needham, M. Evans, D. P. Cosker, and S. L. Colyer, "Using computer vision and deep learning methods to capture skeleton push start performance characteristics," *ISBS Proceedings Archive*, vol. 38, no. 1, 2020. [Online]. Available: https://commons.nmu.edu/isbs/vol38/iss1/191

[8] G. L. Goh, G. D. Goh, J. W. Pan, P. S. P. Teng, and P. W. Kong, "Automated Service Height Fault Detection Using Computer Vision and Machine Learning for Badminton Matches," *Sensors*, vol. 23, no. 24, p. 9759, dec 11 2023. [Online]. Available: https://doi.org/10.3390/s23249759

[9] S. Akkas, S. S. Maini, and J. Qiu, "A Fast Video Image Detection using TensorFlow Mobile Networks for Racing Cars," in *2019 IEEE International Conference on Big Data (Big Data)*, vol. abs 1804 2767. IEEE, 12 2019, pp. 5667–5672. [Online]. Available: https://doi.org/10.1109/BigData47090.2019.9005689

[10] W. Khan, K. Crockett, J. O'Shea, A. Hussain, and B. M. Khan, "Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection," *Expert Systems with Applications*, vol. 169, p. 114341, 5 2021. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.114341

[11] M. S. Sharif, A. Zorto, V. K. Brown, and W. Elmedany, "Scalable machine learning model for highway cctv feed real-time car accident and damage detection," in *Proceedings of the International Conference on Innovative Computing and Technologies (3ICT)*, 2023. [Online]. Available: https://doi.org/10.1109/3ict60104.2023.10391543

[12] U. Jartarghar, D. Sanghvi, M. Nidgundi, K. Kumar, and S. Varur, "Vision transformer-based multi-phase accident detection and false positive mitigation," *Preprint*, 2024. [Online]. Available: https://doi.org/10.21203/rs.3.rs-3903862/v1

[13] C. Wang, "A review on 3d convolutional neural network," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, 2023, pp. 1204–1208. [Online]. Available: https://doi.org/10.1109/ICPECA56706.2023.10075760

[14] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013. [Online]. Available: https://doi.org/10.1109/TPAMI.2012.59

[15] R. Al-Akam, D. Paulus, and D. Gharabaghi, "Human Action Recognition based on 3d Convolution Neural Networks from RGBD Videos," in *CSRN*. Západočeská univerzita, 2018. [Online]. Available: https://doi.org/10.24132/CSRN.2018.2803.3

[16] S. Gurkirt and C. Fabio, "Recurrent Convolutions for Causal 3d CNNs," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1811.07157

[17] S. Tiwari, G. Jain, D. K. Shetty, M. Sudhi, J. M. Balakrishnan, and S. R. Bhatta, "A Comprehensive Review on the Application of 3d Convolutional Neural Networks in Medical Imaging," in *RAiSE-2023*, vol. 25. MDPI, dec 11 2023, p. 3. [Online]. Available: https://doi.org/10.3390/engproc2023059003

[18] T. L. and H. S. Amin, "3d Convolutional Neural Networks for Ultrasound-Based Silent Speech Interfaces," *International Conference on Artificial Intelligence and Soft Computing*, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-61401-0_16

[19] K. S., L. Jeongyeon, N. Taeyoung, and K. Munchurl, "3dsrnet: Video Super-resolution using 3d Convolutional Neural Networks," *arXiv.org*, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1812.09079

[20] J. Chen, S. Bi, G. Zhang, and G. Cao, "High-Density Surface EMG-Based Gesture Recognition Using a 3d Convolutional Neural Network," *Sensors*, vol. 20, no. 4, p. 1201, feb 21 2020. [Online]. Available: https://doi.org/10.3390/s20041201

[21] S. Sun, L. Mu, L. Wang, and P. Liu, "L-unet: An lstm network for remote sensing image change detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020. [Online]. Available: https://doi.org/10.1109/LGRS.2020.3041530

[22] B. Wonmin, B. T., R. Federico, and L. M., "Scene labeling with LSTM recurrent neural networks," *Computer Vision and Pattern Recognition*, 2015. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298977

[23] U. Amin, A. Jamil, M. Khan, S. M., and B. S., "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features," *IEEE Access*, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2017.2778011

[24] R. V. Rahul, S. Bing, L. Jiwen, X. Dong, and W. G., "A Siamese Long Short-Term Memory Architecture for Human Re-identification," *European Conference on Computer Vision*, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-46478-7_9

[25] S. Nitish, M. Elman, and S. R., "Unsupervised Learning of Video Representations using LSTMs," *International Conference on Machine Learning*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1502.04681

[26] G. Kuldeep, "Image Classification using Sequence of Pixels," *arXiv.org*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2209.11495

[27] J. Li, S. Zhang, and T. Huang, "Multi-scale 3d convolution network for video based person re-identification," 2018. [Online]. Available: https://arxiv.org/abs/1811.07468

[28] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1412.0767

[29] K. Liu, W. Liu, C. Gan, M. Tan, and H. Ma, "T-c3d: Temporal convolutional 3d network for real-time action recognition," in *AAAI Conference on Artificial Intelligence*, 2018. [Online]. Available: https://doi.org/10.1609/aaai.v32i1.12333

[30] J. Mänttäri, S. Broomé, J. Folkesson, and H. Kjellström, "Interpreting video features: a comparison of 3d convolutional networks and convolutional lstm networks," 2020. [Online]. Available: https://arxiv.org/abs/2002.00367

[31] C. Zhan, X. Duan, S. Xu, Z. Song, and M. Luo, "An improved moving object detection algorithm based on frame difference and edge detection," in *Fourth International Conference on Image and Graphics (ICIG 2007)*, 2007, pp. 519–523. [Online]. Available: https://doi.org/10.1109/ICIG.2007.153

[32] E. Şimşek and B. Ozyer, "Selected three frame difference method for moving object detection," *International Journal of Intelligent Systems and Applications in Engineering*, 2021. [Online]. Available: https://doi.org/10.18201/ijisae.2021.233

[33] C. Zhang, B. Cheng, and X. Gao, "A motion detection algorithm integrating foreground matching into gaussian mixture modeling," *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 01, pp. 366–369, 2018. [Online]. Available: https://doi.org/10.1109/ISCID.2018.00089

[34] G. Chen, Z. Yu, Q. Wen, and Y. Yu, "Improved gaussian mixture model for moving object detection," in *Artificial Intelligence and Computational Intelligence*, 2011. [Online]. Available: https://doi.org/10.1007/978-3-642-23881-9_23

[35] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," in *IEEE Transactions on Neural Networks and Learning Systems*, 2016, pp. 2222–2232. [Online]. Available: https://doi.org/10.1109/TNNLS.2016.2582924

[36] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4489–4497. [Online]. Available: https://doi.org/10.1109/ICCV.2015.510