COMPREHENSIVE RESEARCH SUMMARY

**https://arxiv.org/pdf/2206.01038**

# MAIN RESEARCH PAPER

**Authors:** Fei Wu, Qingzhong Wang, Jiang Bian, Haoyi Xiong, Ning Ding, Feixiang Lu, Jun Cheng, and Dejing Dou  
**Publication:** arXiv preprint arXiv:2206.01038v1, 2022  
**DOI:** Not available (arXiv preprint)

### Abstract

This survey paper provides a comprehensive overview of video action recognition specifically within the domain of sports analytics. The authors highlight the increasing demand for video understanding technologies due to the rapid growth of video content. They emphasize that while general action recognition has advanced significantly, sports analytics presents unique challenges such as fast actions, multi-player interactions in team sports, and the need for domain-specific expertise for data annotation. The paper covers over ten types of sports, categorizing them into team sports (e.g., football, basketball, volleyball, hockey) and individual sports (e.g., figure skating, gymnastics, table tennis, tennis, diving, badminton). It compares existing frameworks for sports analysis, discusses the current state of video action recognition in both team sports and individual sports, and identifies remaining challenges and unsolved problems. To support future research, the authors also introduce a deep learning toolbox developed using PaddlePaddle, which supports action recognition for football, basketball, table tennis, and figure skating.

### Key Findings

\* \*\*Challenges in Sports Action Recognition:\*\*  
 \* \*\*Data Collection & Labeling:\*\* More sophisticated than general action recognition, often requiring sports professionals for accurate annotation.  
 \* \*\*Action Speed:\*\* Actions can be extremely fast, making recognition difficult.  
 \* \*\*Multi-Player Interaction:\*\* In team sports, actions involve multiple players, necessitating analysis of all participants and their interactions (e.g., ball tracking, player tracking).  
 \* \*\*Coarse General Datasets:\*\* Popular datasets like ActivityNet and Kinetics-400 are too coarse for specific sports analysis, requiring fine-grained, sport-specific video labeling.  
 \* \*\*Domain Knowledge:\*\* Fine-grained annotations demand specialized domain knowledge and involvement of professional players.  
\* \*\*Classification of Sports:\*\* Sports are broadly classified into:  
 \* \*\*Team Sports:\*\* Individuals organized into opposing teams (e.g., football, basketball, volleyball, hockey). Analysis often requires tracking multiple players, the ball, and modeling interactions.  
 \* \*\*Individual Sports:\*\* Participants compete as individuals (e.g., figure skating, gymnastics, table tennis, tennis, diving, badminton). Analysis can focus on one or two players via person detection.  
\* \*\*Evolution of Research:\*\* More research works have emerged in the last five years, with increasing attention on less-studied sports like diving and figure skating.  
\* \*\*Applications of Sports Action Recognition:\*\*  
 \* \*\*Training Aids:\*\* Analyzing historical records and training clips to extract tactics, guide player training, and design game plans.  
 \* \*\*Game Assistance (Video Judge):\*\* Supporting virtual refereeing and performance evaluation based on visual cues and body actions.  
 \* \*\*Video Highlights Generation:\*\* Automatically detecting and stitching key moments in sports videos.  
 \* \*\*Automatic Sports News Generation (ASNG):\*\* Automating the generation of statistical numbers and textual descriptions from match videos.  
 \* \*\*General Research Purposes:\*\* Sports videos serve as a popular and diverse domain for advancing video analysis and understanding.

### Methods

The paper surveys methodologies for both individual and group/team action recognition.  
  
\* \*\*Individual Action Recognition:\*\*  
 \* \*\*Traditional Models:\*\* Rely on hand-crafted features like GIST, HOGs, HOG3D, MBH, HOF, and SIFT. These models typically involve separate feature extraction and classification steps. While some achieved good performance (e.g., Action Bank on UCF sports), they are often time-consuming and not end-to-end trainable.  
 \* \*\*Deep Models:\*\* Dominate current research due to feasibility and end-to-end training capabilities. Four main types are discussed:  
 \* \*\*2D Models:\*\* Use 2D CNNs or transformers to process frames individually, then fuse features. Approaches include single-frame fusion, early fusion, late fusion, and slow fusion. LSTM networks are also used to capture temporal information. Recent trends show a shift towards vision transformers (ViT) for spatial and temporal attention.  
 \* \*\*3D Models:\*\* Treat sequences of frames as a whole, applying 3D CNNs or cube-based transformers to capture spatio-temporal information simultaneously (e.g., C3D, I3D, P3D, SlowFast, X3D, MoViNets). These models generally outperform 2D models but are more computationally intensive. Pre-train-fine-tune paradigms are increasingly popular.  
 \* \*\*Two-stream Models:\*\* Combine RGB frames and optical flow as input, with separate deep neural networks for each stream. They generally achieve better performance than single-stream models but require optical flow calculation and careful fusion strategies.  
 \* \*\*Skeleton-based Models:\*\* Take players' skeleton graphs (joints) as input, often using Graph Convolutional Networks (GCNs). While faster, they typically perform worse than frame-based models due to lack of appearance information and noise in predicted graphs. PoseC3D, which uses heatmaps, achieves improved robustness.  
 \* \*\*Hybrid and Multimodal Models:\*\* Combine elements from different model types (e.g., 3D CNNs with transformers) or integrate multiple modalities like audio to improve robustness, though training complexity increases.  
  
\* \*\*Group/Team Activity Recognition (GAR):\*\*  
 \* Focuses on the collective behavior of a group, distinct from multi-player activity recognition (MAR) which recognizes separate actions in parallel.  
 \* Typically involves player detection, tracking, individual player feature extraction, and group feature combination.  
 \* Early approaches used hand-crafted features and graphical models.  
 \* Deep learning models often employ hierarchical LSTMs, graph-based models (e.g., CERN, RCRG, StageNet), or transformer-based models (e.g., Anchor-Transformer) to model interactions and aggregate player dynamics.  
 \* Recent advancements incorporate player poses and ball tracklets (e.g., POGARS, Pose3D, DIN).

### Contributions

The survey makes three primary contributions:  
\* \*\*Comprehensive Coverage:\*\* Focuses on action recognition in sports videos, covering over ten diverse sports (both team and individual).  
\* \*\*Methodological Overview:\*\* Provides a classification of sports genres and roadmaps of action recognition methods, along with a summary of sports-related datasets.  
\* \*\*Current State and Challenges:\*\* Presents the current status of video action recognition in different sports types, discusses future challenges, and offers a publicly available deep learning toolbox (PaddlePaddle/PaddleVideo) for sports video action recognition in figure skating, football, basketball, and table tennis.

### Limitations

The paper implicitly acknowledges several limitations within the field, which are also discussed as challenges:  
\* \*\*Data Scarcity and Annotation Difficulty:\*\* High-quality, fine-grained, and domain-specific datasets are hard to collect and annotate, especially for less popular sports or complex actions.  
\* \*\*Computational Cost:\*\* Advanced deep learning models, particularly 3D and transformer-based architectures, are computationally expensive, posing challenges for training and deployment.  
\* \*\*Generalization:\*\* Models trained on specific datasets may not generalize well to different camera angles, lighting conditions, or player styles.  
\* \*\*Real-time Performance:\*\* Achieving real-time action recognition for fast-moving and dense actions remains a challenge.  
\* \*\*Overfitting:\*\* High-capacity models can easily overfit smaller or imbalanced datasets.  
\* \*\*Noise in Auxiliary Data:\*\* Skeleton-based models are susceptible to noise from joint detection.  
\* \*\*Multi-modality Integration:\*\* Combining different modalities (e.g., video, audio) effectively is complex.

# REFERENCE PAPER SUMMARIES

Total references summarized: 190

## Reference 1: Ucf101: A dataset of 101 human actions classes from videos in the wild

**Authors:** K. Soomro, A. R. Zamir, and M. Shah  
**Publication:** arXiv preprint arXiv:1212.0402, 2012  
**DOI:** Not available

### Key Contribution

Introduces UCF101, a large-scale dataset for human action recognition with 101 action categories collected from YouTube videos, designed to be more challenging than previous datasets due to variations in camera motion, object appearance, and viewpoint.

### Relevance to Primary Paper

Cited as a popular general action recognition dataset. The main paper highlights that while UCF101 is widely used, its coarse labels make it less suitable for fine-grained sports analysis, underscoring the need for sport-specific datasets.

## Reference 2: Fsd-10: a dataset for competitive sports content analysis

**Authors:** S. Liu, X. Liu, G. Huang, L. Feng, L. Hu, D. Jiang, A. Zhang, Y. Liu, and H. Qiao  
**Publication:** arXiv preprint arXiv:2002.03312, 2020  
**DOI:** Not available

### Key Contribution

Presents FSD-10, a fine-grained dataset specifically for figure skating action recognition, featuring 10 action categories and providing scores for action quality assessment. It aims to address the lack of fine-grained, sport-specific datasets.

### Relevance to Primary Paper

Highlighted as a key example of a fine-grained, sport-specific dataset for figure skating. The main paper uses FSD-10 to evaluate the performance of 2D deep learning models (like KTSN), demonstrating its utility for fine-grained sports action recognition and quality assessment.

## Reference 3: What players do with the ball: A physically constrained interaction modeling

**Authors:** A. Maksai, X. Wang, and P. Fua  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2016  
**DOI:** Not available

### Key Contribution

Proposes a physically constrained interaction modeling approach using graphical models to track the ball and detect players, enabling the analysis of interactions between players and the ball in team sports.

### Relevance to Primary Paper

Cited as an early work in group activity recognition (GAR) for team sports. The main paper notes its contribution to modeling interactions in basketball, particularly ball states, although it points out the limited number of action classes in its GAR settings.

## Reference 4: Hmdb: a large video database for human motion recognition

**Authors:** H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre  
**Publication:** 2011 International conference on computer vision, 2011  
**DOI:** Not available

### Key Contribution

Introduces HMDB51, a large video database for human motion recognition with 51 action categories, collected from various sources including movies and YouTube, providing a diverse and challenging benchmark.

### Relevance to Primary Paper

Cited as a benchmark dataset for evaluating traditional action recognition models, particularly for hand-crafted features like GIST and HOGs. It serves as a comparison point for the performance of early methods before the dominance of deep learning.

## Reference 5: Activitynet: A large-scale video benchmark for human activity understanding

**Authors:** F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles  
**Publication:** Proceedings of the ieee conference on computer vision and pattern recognition, 2015  
**DOI:** Not available

### Key Contribution

Presents ActivityNet, a large-scale video benchmark for human activity understanding, featuring a hierarchical structure of activities and dense annotations for temporal localization.

### Relevance to Primary Paper

Cited as a popular general action recognition dataset. The main paper emphasizes that ActivityNet, despite its scale, focuses on daily life activities and has coarse labels, making it less suitable for the fine-grained analysis required in specific sports.

## Reference 6: Youtube-8m: A large-scale video classification benchmark

**Authors:** S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici, B. Varadarajan, and S. Vijayanarasimhan  
**Publication:** arXiv preprint arXiv:1609.08675, 2016  
**DOI:** Not available

### Key Contribution

Introduces YouTube-8M, a massive-scale video classification benchmark dataset with millions of YouTube video IDs and associated labels, designed for large-scale video understanding research.

### Relevance to Primary Paper

Cited as a large-scale video classification benchmark. Its relevance lies in demonstrating the trend towards massive datasets for video understanding, which influences the development of scalable models applicable to the growing volume of sports videos, even if not sport-specific.

## Reference 7: Quo vadis, action recognition? a new model and the kinetics dataset

**Authors:** J. Carreira and A. Zisserman  
**Publication:** proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017  
**DOI:** Not available

### Key Contribution

Proposes the Inflated 3D CNN (I3D) model, which inflates 2D CNN architectures into 3D for video, and introduces the Kinetics dataset, a large-scale, high-quality dataset for human action recognition.

### Relevance to Primary Paper

Cited as a foundational work for 3D deep learning models and a key dataset (Kinetics-400) for evaluating them. The main paper highlights I3D's superior performance on Kinetics-400, showcasing its importance in advancing spatio-temporal feature learning for action recognition, including in sports.

## Reference 8: Ava: A video dataset of spatio-temporally localized atomic visual actions

**Authors:** C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar et al.  
**Publication:** Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018  
**DOI:** Not available

### Key Contribution

Introduces AVA (Atomic Visual Actions), a video dataset providing spatio-temporal localization of atomic visual actions, where actions are defined by human-object interactions and annotated with bounding boxes.

### Relevance to Primary Paper

Cited as an example of a dataset with complex annotations (bounding boxes, temporal positions, action labels). This level of detailed annotation is relevant to the challenges of fine-grained sports action recognition, especially in team sports where player-object interactions are crucial.

## Reference 9: The “something something" video database for learning and evaluating visual common sense

**Authors:** R. Goyal, S. Ebrahimi Kahou, V. Michalski, J. Materzynska, S. Westphal, H. Kim, V. Haenel, I. Fruend, P. Yianilos, M. Mueller-Freitag et al.  
**Publication:** Proceedings of the IEEE international conference on computer vision, 2017  
**DOI:** Not available

### Key Contribution

Presents the 'Something Something' dataset, focusing on human-object interactions and requiring models to understand visual common sense, rather than just object recognition.

### Relevance to Primary Paper

Cited as a general video database for action recognition. Its emphasis on human-object interactions and common sense reasoning is relevant to understanding complex actions in sports, where players interact with equipment (e.g., ball, racket) and other players.

## Reference 10: Two-stream convolutional networks for action recognition in videos

**Authors:** K. Simonyan and A. Zisserman  
**Publication:** Advances in neural information processing systems, 2014  
**DOI:** Not available

### Key Contribution

Introduces the two-stream convolutional network architecture, which processes static image frames (spatial stream) and optical flow (temporal stream) separately and then combines their predictions for action recognition, demonstrating the importance of motion information.

### Relevance to Primary Paper

Cited as a foundational work for two-stream models in action recognition. The main paper discusses how this architecture significantly improved performance by explicitly incorporating motion, influencing subsequent two-stream variants used in sports action recognition.

## Reference 11: Non-local neural networks

**Authors:** X. Wang, R. Girshick, A. Gupta, and K. He  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2018  
**DOI:** Not available

### Key Contribution

Proposes non-local operations that capture long-range dependencies by computing the interaction between any two positions in a feature map, regardless of their spatial or temporal distance, making them suitable for video processing.

### Relevance to Primary Paper

Cited in the introduction as a general advancement in deep learning for video understanding. Its ability to model long-range dependencies is implicitly relevant to sports action recognition, where understanding interactions across time and space (e.g., between distant players or over a long play) is crucial.

## Reference 12: Slowfast networks for video recognition

**Authors:** C. Feichtenhofer, H. Fan, J. Malik, and K. He  
**Publication:** Proceedings of the IEEE/CVF international conference on computer vision, 2019  
**DOI:** Not available

### Key Contribution

Introduces SlowFast Networks, a novel video recognition architecture with two pathways: a 'slow' pathway for spatial semantics (low frame rate) and a 'fast' pathway for rapid motion (high frame rate), effectively combining different temporal resolutions.

### Relevance to Primary Paper

Highlighted as a state-of-the-art 3D deep learning model. The main paper discusses its elaborate design for fusing slow and fast features and its strong performance on action recognition datasets, making it highly relevant for capturing the diverse temporal dynamics in sports videos.

## Reference 13: Tsm: Temporal shift module for efficient video understanding

**Authors:** J. Lin, C. Gan, and S. Han  
**Publication:** Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019  
**DOI:** Not available

### Key Contribution

Proposes the Temporal Shift Module (TSM), an efficient module that enables 2D CNNs to perform temporal modeling by shifting channels along the temporal dimension, achieving 3D-like performance with 2D computational cost.

### Relevance to Primary Paper

Cited as an efficient 2D deep learning model for action recognition. The main paper emphasizes TSM's ability to capture temporal information with lower computational complexity than 3D approaches, making it practical for sports video analysis where efficiency is often desired.

## Reference 14: Vivit: A video vision transformer

**Authors:** A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Luˇ ci´ c, and C. Schmid  
**Publication:** Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021  
**DOI:** Not available

### Key Contribution

Extends the Vision Transformer (ViT) to video action recognition by using 'tubelet' embeddings and exploring different spatio-temporal attention mechanisms, demonstrating the potential of transformers for video understanding.

### Relevance to Primary Paper

Cited as a prominent 3D transformer-based model. The main paper discusses ViViT's architecture and its strong performance on large datasets, indicating the growing trend of transformer models in sports action recognition, despite their high computational demands.

## Reference 15: Tea: Temporal excitation and aggregation for action recognition

**Authors:** Y. Li, B. Ji, X. Shi, J. Zhang, B. Kang, and L. Wang  
**Publication:** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020  
**DOI:** Not available

### Key Contribution

Introduces Temporal Excitation and Aggregation (TEA), a module designed to effectively capture and aggregate temporal information from video sequences by modeling the importance of different temporal segments.

### Relevance to Primary Paper

Cited in the introduction as an impressive model for improving recognition accuracy. Its focus on effective temporal aggregation is directly relevant to sports action recognition, where understanding the sequence and timing of movements is critical.

## Reference 16: The kinetics human action video dataset

**Authors:** W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev et al.  
**Publication:** arXiv preprint arXiv:1705.06950, 2017  
**DOI:** Not available

### Key Contribution

Formally introduces the Kinetics human action video dataset, a large-scale collection of YouTube video URLs annotated with human actions, providing a diverse and challenging benchmark for action recognition.

### Relevance to Primary Paper

Cited as a popular and large-scale dataset for action recognition, particularly for evaluating deep learning models. The main paper notes that while Kinetics is widely used, it focuses on daily life activities, underscoring the need for sport-specific datasets for fine-grained analysis.

## Reference 17: A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis

**Authors:** B. Giles, S. Kovalchik, and M. Reid  
**Publication:** Journal of sports sciences, 2020  
**DOI:** Not available

### Key Contribution

Develops a machine learning approach to automatically detect and classify changes of direction in professional tennis players using spatio-temporal tracking data, providing objective performance metrics.

### Relevance to Primary Paper

Cited as an example of recent research in sports video analysis, specifically in tennis. It demonstrates the application of machine learning to extract tactical and performance-related information from player movements, aligning with the survey's focus on sports analytics applications.

## Reference 18: Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance

**Authors:** E. E. Cust, A. J. Sweeting, K. Ball, and S. Robertson  
**Publication:** Journal of sports sciences, 2019  
**DOI:** Not available

### Key Contribution

Provides a systematic review of machine and deep learning models applied to sport-specific movement recognition, analyzing model development, performance, and identifying research gaps.

### Relevance to Primary Paper

Cited as a general reference for the growing attention to sports video analysis. This review reinforces the main paper's premise by summarizing the state-of-the-art in applying machine and deep learning to sport-specific actions, validating the survey's scope.

## Reference 19: Development of a human activity recognition system for ballet tasks

**Authors:** D. Hendry, K. Chai, A. Campbell, L. Hopper, P. O’Sullivan, and L. Straker  
**Publication:** Sports medicine-open, 2020  
**DOI:** Not available

### Key Contribution

Develops a human activity recognition system tailored for ballet tasks, focusing on the precise and nuanced movements required in dance, often using wearable sensors or video analysis.

### Relevance to Primary Paper

Cited as an example of applying action recognition to fine-grained human movement analysis in performance-oriented domains. While not a competitive sport, it shares methodological similarities with sports analytics, particularly in assessing movement quality and technique.

## Reference 20: The development of a personalised training framework: Implementation of emerging technologies for performance

**Authors:** C. Pickering and J. Kiely  
**Publication:** Journal of Functional Morphology and Kinesiology, 2019  
**DOI:** Not available

### Key Contribution

Discusses the creation of a personalized training framework that integrates emerging technologies, including potentially video analysis and action recognition, to optimize athletic performance and provide tailored feedback.

### Relevance to Primary Paper

Cited in the context of 'Training Aids' applications. This paper supports the main paper's discussion on how action recognition can contribute to personalized coaching systems and performance analysis for athletes.

## Reference 21: Moving the lab into the mountains: A pilot study of human activity recognition in unstructured environments

**Authors:** B. Russell, A. McDaid, W. Toscano, and P. Hume  
**Publication:** Sensors, 2021  
**DOI:** Not available

### Key Contribution

Investigates human activity recognition in challenging, unstructured outdoor environments using wearable sensors, aiming to develop robust systems for real-world applications beyond controlled lab settings.

### Relevance to Primary Paper

Cited as an example of recent research addressing the challenges of real-world data. Its focus on unstructured environments is relevant to sports video analysis, which often deals with uncontrolled conditions, varied camera angles, and dynamic backgrounds.

## Reference 22: Deep learning in sport video analysis: a review

**Authors:** K. Rangasamy, M. A. As’ ari, N. A. Rahmad, N. F. Ghazali, and S. Ismail  
**Publication:** Telkomnika, 2020  
**DOI:** Not available

### Key Contribution

Provides a comprehensive review of deep learning applications in sport video analysis, categorizing existing works and discussing trends, challenges, and future directions in the field.

### Relevance to Primary Paper

Cited as a general review paper that reinforces the main paper's focus on deep learning in sports video analysis. It provides a broader context and validates the importance of the survey's topic.

## Reference 23: Hierarchical relational networks for group activity recognition and retrieval

**Authors:** M. S. Ibrahim and G. Mori  
**Publication:** Proceedings of the European conference on computer vision (ECCV), 2018  
**DOI:** Not available

### Key Contribution

Proposes hierarchical relational networks that model relationships between individuals and groups for improved group activity recognition and retrieval, particularly in complex multi-person scenes.

### Relevance to Primary Paper

Cited in the context of group/team activity recognition (GAR). The main paper discusses its contribution to modeling relationships among players for collective activity understanding, extending two-stage frameworks with a hierarchical relational network.

## Reference 24: Soccernet: A scalable dataset for action spotting in soccer videos

**Authors:** S. Giancola, M. Amine, T. Dghaily, and B. Ghanem  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018  
**DOI:** Not available

### Key Contribution

Introduces SoccerNet, a large-scale dataset comprising 500 complete soccer match videos with temporal annotations for action spotting, aiming to facilitate research in soccer video analysis.

### Relevance to Primary Paper

Highlighted as a significant large-scale dataset for football action recognition and localization. The main paper notes its importance for driving large-scale deep learning models in soccer analysis, despite its relatively sparse action annotations.

## Reference 25: Fine grained sport action recognition with twin spatio-temporal convolutional neural networks

**Authors:** P.-E. Martin, J. Benois-Pineau, R. Péteri, and J. Morlier  
**Publication:** Multimedia Tools and Applications, 2020  
**DOI:** Not available

### Key Contribution

Develops a twin spatio-temporal convolutional neural network architecture for fine-grained sport action recognition, demonstrating improved performance on detailed action classification tasks.

### Relevance to Primary Paper

Cited as an example of recent research in fine-grained sports action recognition, particularly in table tennis. It showcases the application of advanced CNN architectures for detailed action classification, aligning with the survey's focus on fine-grained analysis.

## Reference 26: Sports video: Fine-grained action detection and classification of table tennis strokes from videos for mediaeval 2021

**Authors:** P.-E. Martin, J. Calandre, B. Mansencal, J. Benois-Pineau, R. Péteri, L. Mascarilla, and J. Morlier  
**Publication:** arXiv preprint arXiv:2112.11384, 2021  
**DOI:** Not available

### Key Contribution

Addresses fine-grained action detection and classification of table tennis strokes from videos, likely as part of a MediaEval challenge, contributing to the precise identification of complex and rapid movements.

### Relevance to Primary Paper

Cited as an example of ongoing research in fine-grained sports action recognition, specifically in table tennis. It highlights the continuous effort to achieve highly detailed and accurate action recognition in fast-paced sports.

## Reference 27: A context-aware loss function for action spotting in soccer videos

**Authors:** A. Cioppa, A. Deliege, S. Giancola, B. Ghanem, M. V. Droogenbroeck, R. Gade, and T. B. Moeslund  
**Publication:** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020  
**DOI:** Not available

### Key Contribution

Proposes a novel context-aware loss function designed to improve the accuracy of action spotting (temporal localization) in soccer videos by incorporating broader temporal context.

### Relevance to Primary Paper

Cited as an example of recent research in football video analysis. It demonstrates advancements in loss function design to better capture the temporal context of events, which is crucial for accurate action localization in sports.

## Reference 28: Multisports: A multi-person video dataset of spatio-temporally localized sports actions

**Authors:** Y. Li, L. Chen, R. He, Z. Wang, G. Wu, and L. Wang  
**Publication:** Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021  
**DOI:** Not available

### Key Contribution

Introduces MultiSports, a challenging multi-person video dataset for spatio-temporally localized sports actions, covering four team sports and 66 action categories, with bounding box annotations for players.

### Relevance to Primary Paper

Highlighted as a significant and challenging dataset for multi-person sports action recognition. The main paper notes its utility for tasks requiring both temporal and spatial localization due to its detailed annotations, addressing a key challenge in team sports analysis.

## Reference 29: Stagnet: an attentive semantic rnn for group activity and individual action recognition

**Authors:** M. Qi, Y. Wang, J. Qin, A. Li, J. Luo, and L. Van Gool  
**Publication:** IEEE Transactions on Circuits and Systems for Video Technology, 2019  
**DOI:** Not available

### Key Contribution

Proposes StageNet, an attentive semantic RNN framework for group activity and individual action recognition, which integrates player detection, semantic graph construction, and temporal information with spatial-temporal attention.

### Relevance to Primary Paper

Cited in the context of group/team activity recognition (GAR). The main paper discusses StageNet as a four-stage framework that models relationships among players and incorporates attention for explainability, contributing to advanced GAR methodologies.

## Reference 30: Spotting football events using two-stream convolutional neural network and dilated recurrent neural network

**Authors:** B. Mahaseni, E. R. M. Faizal, and R. G. Raj  
**Publication:** IEEE Access, 2021  
**DOI:** Not available

### Key Contribution

Develops a method for spotting football events by combining a two-stream convolutional neural network with a dilated recurrent neural network, leveraging both visual appearance and temporal dependencies.

### Relevance to Primary Paper

Cited as an example of recent research in football video analysis. It demonstrates the application of hybrid deep learning architectures for event detection, contributing to the automation of sports analytics.

## Reference 31: Am i a baller? basketball performance assessment from first-person videos

**Authors:** G. Bertasius, H. Soo Park, S. X. Yu, and J. Shi  
**Publication:** Proceedings of the IEEE international conference on computer vision, 2017  
**DOI:** Not available

### Key Contribution

Investigates basketball performance assessment using first-person videos, aiming to provide objective feedback on player actions and skills from a unique perspective.

### Relevance to Primary Paper

Cited as an application where recognizing individual actions benefits player training by helping correct small errors. It highlights the practical utility of action recognition in sports coaching and performance improvement.

## Reference 32: Toward the perfect stroke: A multimodal approach for table tennis stroke evaluation

**Authors:** P. Sri-Iesaranusorn, F. C. Garcia, F. Tiausas, S. Wattanakriengkrai, K. Ikeda, and J. Yoshimoto  
**Publication:** 2021 Thirteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), 2021  
**DOI:** Not available

### Key Contribution

Proposes a multimodal approach (e.g., video, sensor data) for evaluating table tennis strokes, aiming to provide detailed feedback for improving player technique and performance.

### Relevance to Primary Paper

Cited as an example of how individual action recognition can benefit player training, specifically in table tennis. It showcases the use of multimodal data for detailed stroke analysis and performance assessment.

## Reference 33: Summarization of user-generated sports video by using deep action recognition features

**Authors:** A. Tejero-de Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna, and E. Rahtu  
**Publication:** IEEE Transactions on Multimedia, 2018  
**DOI:** Not available

### Key Contribution

Develops a method for summarizing user-generated sports videos by leveraging deep action recognition features to identify and extract key moments, improving the efficiency of content creation.

### Relevance to Primary Paper

Cited as an application of sports action recognition in generating highlights. It emphasizes how accurate action recognition significantly improves the localization accuracy of key moments for sports TV programs and content creation.

## Reference 34: Automatic cricket highlight generation using event-driven and excitement-based features

**Authors:** P. Shukla, H. Sadana, A. Bansal, D. Verma, C. Elmadjian, B. Raman, and M. Turk  
**Publication:** Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018  
**DOI:** Not available

### Key Contribution

Proposes a system for automatic cricket highlight generation that combines event-driven features (e.g., wickets, boundaries) with excitement-based features (e.g., crowd noise, player reactions) to create compelling summaries.

### Relevance to Primary Paper

Cited as an application of sports action recognition in generating highlights. It demonstrates how action recognition contributes to identifying and localizing exciting moments in sports broadcasts, extending beyond just the action itself to include contextual cues.

## Reference 35: Visual analytics for team-based invasion sports with significant events and markov reward process

**Authors:** K. Zhao, T. Osogami, and T. Morimura  
**Publication:** arXiv preprint arXiv:1907.01221, 2019  
**DOI:** Not available

### Key Contribution

Introduces a visual analytics framework for team-based invasion sports (like soccer or basketball) that identifies significant events and models game flow using a Markov reward process, providing insights into team strategies.

### Relevance to Primary Paper

Cited as an application of sports action recognition in generating highlights. It indicates its relevance to identifying and analyzing crucial events in complex team sports scenarios, contributing to deeper game understanding.

## Reference 36: A new action recognition framework for video highlights summarization in sporting events

**Authors:** C. Yan, X. Li, and G. Li  
**Publication:** 2021 16th International Conference on Computer Science & Education (ICCSE), 2021  
**DOI:** Not available

### Key Contribution

Proposes a novel action recognition framework specifically designed to improve video highlights summarization in sporting events by accurately identifying and segmenting key actions.

### Relevance to Primary Paper

Cited as an application of sports action recognition in generating highlights. It reinforces the importance of accurate action recognition as a foundational step for automated and effective highlight generation in sports media.

## Reference 37: Action recognition using multilevel features and latent structural svm

**Authors:** X. Wu, D. Xu, L. Duan, J. Luo, and Y. Jia  
**Publication:** IEEE transactions on Circuits and Systems for Video Technology, 2013  
**DOI:** Not available

### Key Contribution

Develops an action recognition method that utilizes multilevel features (e.g., local, global) combined with a latent structural Support Vector Machine (SVM) to improve classification accuracy.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It exemplifies early approaches to action recognition in sports using traditional machine learning techniques and feature engineering.

## Reference 38: Large-scale analysis of soccer matches using spatiotemporal tracking data

**Authors:** A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews  
**Publication:** 2014 IEEE international conference on data mining, 2014  
**DOI:** Not available

### Key Contribution

Presents a large-scale analysis of soccer matches by leveraging spatio-temporal tracking data to understand player movements, team formations, and tactical patterns.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It highlights the use of tracking data for deeper analysis of team dynamics and player movements, which is foundational for group activity recognition.

## Reference 39: Take your eyes off the ball: Improving ball-tracking by focusing on team play

**Authors:** X. Wang, V. Ablavsky, H. B. Shitrit, and P. Fua  
**Publication:** Computer Vision and Image Understanding, 2014  
**DOI:** Not available

### Key Contribution

Proposes a method to improve ball-tracking in soccer by integrating contextual information from team play, rather than relying solely on visual features of the ball, making tracking more robust.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It demonstrates the importance of considering broader game context and player interactions for robust object tracking, a prerequisite for accurate action recognition in team sports.

## Reference 40: Ball tracking and action recognition of soccer players in tv broadcast videos

**Authors:** M. Durus  
**Publication:** Ph.D. dissertation, Technische Universität München, 2014  
**DOI:** Not available

### Key Contribution

This Ph.D. dissertation focuses on developing methods for both ball tracking and action recognition of soccer players directly from TV broadcast videos, addressing challenges like camera motion and occlusions.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It highlights a comprehensive approach to both tracking and action recognition in a challenging real-world setting, relevant to the practical application of sports analytics.

## Reference 41: Play type recognition in real-world football video

**Authors:** S. Chen, Z. Feng, Q. Lu, B. Mahasseni, T. Fiez, A. Fern, and S. Todorovic  
**Publication:** IEEE Winter Conference on Applications of Computer Vision, 2014  
**DOI:** Not available

### Key Contribution

Addresses the problem of recognizing high-level 'play types' (e.g., attack, defense) in real-world football videos, moving beyond individual player actions to collective team behaviors.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It contributes to classifying higher-level tactical events in soccer matches, which is a form of group activity recognition.

## Reference 42: Discriminating talent-identified junior australian football players using a video decision-making task

**Authors:** C. T. Woods, A. J. Raynor, L. Bruce, and Z. McDonald  
**Publication:** Journal of sports sciences, 2016  
**DOI:** Not available

### Key Contribution

Investigates the use of video-based decision-making tasks to discriminate talent-identified junior Australian football players, linking visual analysis to talent assessment and development.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2011-2016 period. It highlights the application of video analysis for talent identification and performance assessment in a specific sport context, demonstrating a practical use case for action recognition.

## Reference 43: Football action recognition using hierarchical lstm

**Authors:** T. Tsunoda, Y. Komori, M. Matsugu, and T. Harada  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017  
**DOI:** Not available

### Key Contribution

Proposes a hierarchical LSTM model for football action recognition, designed to capture both individual player actions and their collective interactions over time.

### Relevance to Primary Paper

Cited in the context of group activity recognition (GAR) in football. The main paper notes its similarity to other hierarchical models but emphasizes that the videos are captured by multiple synchronized cameras, adding complexity and realism to the analysis.

## Reference 44: Action recognition in video sequences using deep bi-directional lstm with cnn features

**Authors:** A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik  
**Publication:** IEEE access, 2017  
**DOI:** Not available

### Key Contribution

Develops an action recognition method that combines deep bi-directional LSTMs with CNN features, leveraging CNNs for spatial feature extraction and LSTMs for temporal modeling.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It exemplifies the application of deep learning, specifically LSTMs, for action recognition in sports, contributing to the evolution of methods in the field.

## Reference 45: Comprehensive dataset of broadcast soccer videos

**Authors:** J. Yu, A. Lei, Z. Song, T. Wang, H. Cai, and N. Feng  
**Publication:** 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018  
**DOI:** Not available

### Key Contribution

Introduces ComprehensiveSoccer, a dataset of broadcast soccer videos with multi-level annotations including player bounding boxes, event/story annotations, and shot categories, designed for various analysis tasks.

### Relevance to Primary Paper

Described as a dataset suitable for various tasks like action classification, localization, and player detection. The main paper highlights its multi-level annotations, making it feasible for tasks requiring player detection in football.

## Reference 46: Learning deep c3d features for soccer video event detection

**Authors:** M. Z. Khan, S. Saleem, M. A. Hassan, and M. U. G. Khan  
**Publication:** 2018 14th International Conference on Emerging Technologies (ICET), 2018  
**DOI:** Not available

### Key Contribution

Applies deep C3D (Convolutional 3D) features for soccer video event detection, demonstrating the effectiveness of 3D CNNs in capturing spatio-temporal information for event recognition.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It exemplifies the application of 3D convolutional networks for detecting specific events in soccer matches, contributing to the deep learning advancements in the field.

## Reference 47: A novel framework for fine grained action recognition in soccer

**Authors:** Y. Ganesh, A. Sri Teja, S. K. Munnangi, and G. Rama Murthy  
**Publication:** International Work-Conference on Artificial Neural Networks, 2019  
**DOI:** Not available

### Key Contribution

Proposes a novel framework specifically for fine-grained action recognition in soccer, aiming to distinguish subtle variations in player movements and actions.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It highlights the growing focus on detailed, fine-grained action classification within soccer, addressing the complexity of sports movements.

## Reference 48: Individual action and group activity recognition in soccer videos

**Authors:** B. Gerats  
**Publication:** Master’s thesis, University of Twente, 2020  
**DOI:** Not available

### Key Contribution

This Master's thesis investigates both individual action recognition and group activity recognition within soccer videos, exploring methods to understand both single-player movements and collective team behaviors.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It contributes to the dual challenge of understanding both individual and collective actions in soccer, aligning with the survey's distinction between IAR and GAR.

## Reference 49: Improved soccer action spotting using both audio and video streams

**Authors:** B. Vanderplaetse and S. Dupont  
**Publication:** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020  
**DOI:** Not available

### Key Contribution

Enhances soccer action spotting by integrating information from both audio and video streams, demonstrating that multimodal input can improve the accuracy of event detection.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It highlights the benefit of multimodal approaches for more robust event detection in soccer, supporting the main paper's discussion on multimodal models.

## Reference 50: Group activity detection from trajectory and video data in soccer

**Authors:** R. Sanford, S. Gorji, L. G. Hafemann, B. Pourbabaee, and M. Javan  
**Publication:** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020  
**DOI:** Not available

### Key Contribution

Focuses on detecting group activities in soccer by combining player trajectory data with video information, aiming to understand collective behaviors and tactical patterns.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It indicates its contribution to understanding collective behaviors in soccer through the integration of different data modalities (trajectory and video), relevant to GAR.

## Reference 51: Video assistant referees (var): The impact of technology on decision making in association football referees

**Authors:** J. Spitz, J. Wagemans, D. Memmert, A. M. Williams, and W. F. Helsen  
**Publication:** Journal of Sports Sciences, 2021  
**DOI:** Not available

### Key Contribution

Analyzes the impact of Video Assistant Referee (VAR) technology on decision-making processes of association football referees, providing empirical insights into the role of video review in sports officiating.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It highlights the real-world implications and applications of video analysis in sports officiating, aligning with the 'Game Assistance (Video Judge)' application area.

## Reference 52: Contrastive learning for sports video: Unsupervised player classification

**Authors:** M. Koshkina, H. Pidaparthy, and J. H. Elder  
**Publication:** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021  
**DOI:** Not available

### Key Contribution

Explores the use of contrastive learning for unsupervised player classification in sports videos, aiming to identify and distinguish individual players without explicit labels.

### Relevance to Primary Paper

Cited as a representative research in football video analysis from the 2017-present period. It indicates its contribution to self-supervised learning techniques for player identification and analysis in sports, which can reduce the need for extensive manual labeling.

## Reference 53: Recognizing tactic patterns in broadcast basketball video using player trajectory

**Authors:** H.-T. Chen, C.-L. Chou, T.-S. Fu, S.-Y. Lee, and B.-S. P. Lin  
**Publication:** Journal of Visual Communication and Image Representation, 2012  
**DOI:** Not available

### Key Contribution

Develops a method for recognizing tactic patterns in broadcast basketball videos by analyzing and interpreting player trajectories, providing insights into team strategies.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It highlights its contribution to understanding strategic plays through player movement analysis, a key aspect of team sports analytics.

## Reference 54: Team activity recognition in sports

**Authors:** C. Direkoˇ glu and N. E. O’Connor  
**Publication:** European Conference on Computer Vision, 2012  
**DOI:** Not available

### Key Contribution

Addresses the problem of team activity recognition in sports, focusing on identifying collective actions and interactions among multiple players.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It indicates its contribution to understanding collective actions and interactions within team sports, which is a core component of group activity recognition.

## Reference 55: Exploring game performance in the national basketball association using player tracking data

**Authors:** J. Sampaio, T. McGarry, J. Calleja-González, S. Jiménez Sáiz, X. Schelling i del Alcázar, and M. Balciunas  
**Publication:** PloS one, 2015  
**DOI:** Not available

### Key Contribution

Utilizes player tracking data from the NBA to explore various aspects of game performance, providing quantitative insights into player movements, efficiency, and tactical effectiveness.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It highlights the application of tracking data for in-depth performance analysis in professional basketball, demonstrating the value of detailed player movement data.

## Reference 56: Leveraging contextual cues for generating basketball highlights

**Authors:** V. Bettadapura, C. Pantofaru, and I. Essa  
**Publication:** Proceedings of the 24th ACM international conference on Multimedia, 2016  
**DOI:** Not available

### Key Contribution

Develops a method for generating basketball highlights by leveraging contextual cues (e.g., crowd reactions, commentator sentiment) in addition to visual events, aiming for more engaging summaries.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It indicates its contribution to automated highlight generation by understanding the broader context of events, aligning with the 'Video Highlights' application area.

## Reference 57: Dynamic image networks for action recognition

**Authors:** H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2016  
**DOI:** Not available

### Key Contribution

Introduces Dynamic Image Networks, a compact representation of video sequences that encodes temporal information into a single image, allowing standard 2D CNNs to be used for action recognition.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It indicates its contribution to novel deep learning architectures for capturing motion information in videos, offering an alternative to explicit 3D convolutions or optical flow.

## Reference 58: Detecting events and key actors in multi-person videos

**Authors:** V. Ramanathan, J. Huang, S. Abu-El-Haija, A. Gorban, K. Murphy, and L. Fei-Fei  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2016  
**DOI:** Not available

### Key Contribution

Focuses on detecting events and identifying key actors within multi-person videos, particularly in complex scenes with multiple interacting individuals.

### Relevance to Primary Paper

Cited in the context of the NCAA dataset for basketball action recognition. The main paper notes that the NCAA dataset provides video segments with time boundaries and action categories, and also bounding boxes for player detection, aligning with this paper's focus on multi-person event detection and actor identification.

## Reference 59: Automatic summarization of basketball sport video

**Authors:** D. Chauhan, N. M. Patel, and M. Joshi  
**Publication:** 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016  
**DOI:** Not available

### Key Contribution

Develops a system for automatic summarization of basketball sport videos, aiming to extract key plays and events to create concise and informative video summaries.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2011-2016 period. It indicates its contribution to automated content generation for sports media, specifically for basketball highlights.

## Reference 60: Towards real-time detection and tracking of basketball players using deep neural networks

**Authors:** D. Acuna  
**Publication:** Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017  
**DOI:** Not available

### Key Contribution

Explores the use of deep neural networks for real-time detection and tracking of basketball players, addressing the computational challenges of live sports analysis.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights the advancements in real-time player analysis for sports, which is crucial for applications like live coaching assistance and broadcasting.

## Reference 61: Ontology-based global and collective motion patterns for event classification in basketball videos

**Authors:** L. Wu, Z. Yang, J. He, M. Jian, Y. Xu, D. Xu, and C. W. Chen  
**Publication:** IEEE Transactions on Circuits and Systems for Video Technology, 2019  
**DOI:** Not available

### Key Contribution

Proposes an ontology-based approach to classify events in basketball videos by analyzing global and collective motion patterns, providing a structured understanding of complex team behaviors.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It indicates its contribution to structured understanding of complex team activities, moving beyond individual actions to collective patterns.

## Reference 62: Single-camera basketball tracker through pose and semantic feature fusion

**Authors:** A. Arbués-Sangüesa, C. Ballester, and G. Haro  
**Publication:** arXiv preprint arXiv:1906.02042, 2019  
**DOI:** Not available

### Key Contribution

Develops a single-camera basketball player tracker that fuses pose estimation with semantic features to improve tracking accuracy, especially in challenging broadcast video conditions.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights advancements in player tracking using richer feature representations (pose and semantics), which is a prerequisite for accurate action recognition.

## Reference 63: Analysis of technical features in basketball video based on deep learning algorithm

**Authors:** L. Chen and W. Wang  
**Publication:** Signal Processing: Image Communication, 2020  
**DOI:** Not available

### Key Contribution

Applies deep learning algorithms to analyze technical features in basketball videos, aiming to automatically identify and evaluate specific skills and movements.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It indicates its contribution to automated technical skill assessment, aligning with the 'Training Aids' application area.

## Reference 64: Fine-grained action recognition on a novel basketball dataset

**Authors:** X. Gu, X. Xue, and F. Wang  
**Publication:** ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020  
**DOI:** Not available

### Key Contribution

Introduces a novel basketball dataset specifically designed for fine-grained action recognition and presents a method to classify subtle variations in basketball actions.

### Relevance to Primary Paper

Cited in the context of the FineBasketball dataset. The main paper discusses FineBasketball as a challenging dataset for fine-grained basketball action recognition, and this paper contributes to methods for addressing that challenge.

## Reference 65: Fusing motion patterns and key visual information for semantic event recognition in basketball videos

**Authors:** L. Wu, Z. Yang, Q. Wang, M. Jian, B. Zhao, J. Yan, and C. W. Chen  
**Publication:** Neurocomputing, 2020  
**DOI:** Not available

### Key Contribution

Proposes a method for semantic event recognition in basketball videos by effectively fusing motion patterns with key visual information, aiming for a more comprehensive understanding of game events.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights the importance of combining different types of visual cues (motion and appearance) for robust event recognition in complex sports scenarios.

## Reference 66: Recognition of basketball referee signals from real-time videos

**Authors:** J. Žemgulys, V. Raudonis, R. Maskeli¯ unas, and R. Damaševiˇ cius  
**Publication:** Journal of Ambient Intelligence and Humanized Computing, 2020  
**DOI:** Not available

### Key Contribution

Develops a system for real-time recognition of basketball referee signals from video, contributing to automated officiating assistance and analysis.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It aligns with the 'Game Assistance (Video Judge)' application area, demonstrating how action recognition can be applied to interpret referee gestures.

## Reference 67: Npu rgbd dataset and a feature-enhanced lstm-dgcn method for action recognition of basketball players

**Authors:** C. Ma, J. Fan, J. Yao, and T. Zhang  
**Publication:** Applied Sciences, 2021  
**DOI:** Not available

### Key Contribution

Introduces the NPUBasketball RGBD dataset, which includes RGB frames, depth maps, and skeleton data, and proposes a feature-enhanced LSTM-DGCN method for basketball player action recognition.

### Relevance to Primary Paper

Cited in the context of the NPUBasketball dataset. The main paper notes that this dataset provides multiple modalities (RGB, depth, skeleton), making it suitable for developing various types of action recognition models, including skeleton-based approaches.

## Reference 68: Recognition of basketball player’s shooting action based on the convolutional neural network

**Authors:** R. Liu, Z. Liu, and S. Liu  
**Publication:** Scientific Programming, 2021  
**DOI:** Not available

### Key Contribution

Focuses on the recognition of basketball player's shooting actions using convolutional neural networks, aiming for accurate classification of this specific and critical skill.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights the application of CNNs for recognizing specific individual actions in basketball, contributing to fine-grained analysis.

## Reference 69: Research on basketball players’ action recognition based on interactive system and machine learning

**Authors:** J. Li and D. Gu  
**Publication:** Journal of Intelligent & Fuzzy Systems, 2021  
**DOI:** Not available

### Key Contribution

Investigates basketball players' action recognition within an interactive system using machine learning techniques, potentially enabling real-time feedback or analysis.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It indicates the trend towards interactive systems for sports analysis, where action recognition serves as a core component.

## Reference 70: A lightweight fine-grained action recognition network for basketball foul detection

**Authors:** C.-H. Lin, M.-Y. Tsai, and P.-Y. Chou  
**Publication:** 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2021  
**DOI:** Not available

### Key Contribution

Proposes a lightweight network for fine-grained action recognition specifically for basketball foul detection, aiming for efficient and accurate identification of rule violations.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights the application of fine-grained action recognition for specific game events like fouls, relevant to automated officiating and analysis.

## Reference 71: Basketball action recognition based on fpga and particle image

**Authors:** G. Junjun  
**Publication:** Microprocessors and Microsystems, 2021  
**DOI:** Not available

### Key Contribution

Explores basketball action recognition using FPGA (Field-Programmable Gate Array) for hardware acceleration and particle image velocimetry for motion analysis, aiming for high-speed processing.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It indicates advancements in hardware-accelerated action recognition, which is crucial for achieving real-time performance in sports analytics.

## Reference 72: Camera-based basketball scoring detection using convolutional neural network

**Authors:** X.-B. Fu, S.-L. Yue, and D.-Y. Pan  
**Publication:** International Journal of Automation and Computing, 2021  
**DOI:** Not available

### Key Contribution

Develops a camera-based system for basketball scoring detection using convolutional neural networks, automating the identification of successful shots.

### Relevance to Primary Paper

Cited as a representative research in basketball video analysis from the 2017-present period. It highlights the application of CNNs for detecting specific game outcomes, contributing to automated scorekeeping and event logging.

## Reference 73: Long-and short-term plastic modeling of action prediction abilities in volleyball

**Authors:** C. Urgesi, M. M. Savonitto, F. Fabbro, and S. M. Aglioti  
**Publication:** Psychological research, 2012  
**DOI:** Not available

### Key Contribution

Investigates the neural mechanisms underlying action prediction abilities in volleyball players, distinguishing between long-term (expertise-based) and short-term (context-based) learning.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2011-2016 period. While not directly a computer vision paper, it highlights the cognitive aspects of action understanding in sports, which informs the design of predictive models in AI.

## Reference 74: Indoor activity detection and recognition for sport games analysis

**Authors:** G. Waltner, T. Mauthner, and H. Bischof  
**Publication:** arXiv preprint arXiv:1404.6413, 2014  
**DOI:** Not available

### Key Contribution

Focuses on indoor activity detection and recognition for sport games analysis, likely using sensor data or video from controlled environments to identify specific actions.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2011-2016 period. It indicates early efforts in activity recognition for sports in indoor settings, relevant to sports like volleyball.

## Reference 75: Saeta: A smart coaching assistant for professional volleyball training

**Authors:** J. Vales-Alonso, D. Chaves-Diéguez, P. López-Matencio, J. J. Alcaraz, F. J. Parrado-García, and F. J. González-Castaño  
**Publication:** IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015  
**DOI:** Not available

### Key Contribution

Introduces SAETA, a smart coaching assistant system designed for professional volleyball training, which likely integrates video analysis and action recognition to provide feedback to athletes.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2011-2016 period. It highlights the application of technology, including action recognition, in developing practical coaching tools for sports.

## Reference 76: Sum product networks for activity recognition

**Authors:** M. R. Amer and S. Todorovic  
**Publication:** IEEE transactions on pattern analysis and machine intelligence, 2015  
**DOI:** Not available

### Key Contribution

Proposes the use of Sum-Product Networks (SPNs) for activity recognition, offering a probabilistic graphical model approach that can efficiently represent and infer complex dependencies in data.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2011-2016 period. It indicates the exploration of advanced probabilistic models for activity recognition, which can be applied to sports contexts.

## Reference 77: A hierarchical deep temporal model for group activity recognition

**Authors:** M. S. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori  
**Publication:** Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016  
**DOI:** Not available

### Key Contribution

Develops a hierarchical deep temporal model for group activity recognition, where individual player dynamics are modeled by LSTMs and then aggregated by a group-level LSTM to recognize collective actions.

### Relevance to Primary Paper

Cited as a key work in group activity recognition (GAR) for volleyball. The main paper discusses its architecture and performance on the HierVolleyball dataset, highlighting its contribution to modeling collective behaviors in team sports.

## Reference 78: Activity recognition in beach volleyball using a deep convolutional neural network

**Authors:** T. Kautz, B. H. Groh, J. Hannink, U. Jensen, H. Strubberg, and B. M. Eskofier  
**Publication:** Data Mining and Knowledge Discovery, 2017  
**DOI:** Not available

### Key Contribution

Applies a deep convolutional neural network for activity recognition in beach volleyball, demonstrating the effectiveness of deep learning for classifying actions in this specific sport.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2017-present period. It highlights the application of deep CNNs for activity recognition in a specific volleyball variant, contributing to the growing body of deep learning applications in sports.

## Reference 79: Evaluation of dominant and non-dominant hand movements for volleyball action modelling

**Authors:** F. Haider, F. Salim, V. Naghashi, S. B. Y. Tasdemir, I. Tengiz, K. Cengiz, D. Postma, R. v. Delden, D. Reidsma, B.-J. van Beijnum et al.  
**Publication:** Adjunct of the 2019 International Conference on Multimodal Interaction, 2019  
**DOI:** Not available

### Key Contribution

Evaluates the significance of dominant and non-dominant hand movements for modeling volleyball actions, potentially using sensor data or fine-grained video analysis to capture subtle kinematic differences.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2017-present period. It indicates the depth of analysis in sports, focusing on specific body parts and their contribution to action modeling.

## Reference 80: Volleyball action modelling for behavior analysis and interactive multi-modal feedback

**Authors:** F. A. Salim, F. Haider, S. B. Y. Tasdemir, V. Naghashi, I. Tengiz, K. Cengiz, D. Postma, and R. Van Delden  
**Publication:** Proceeding of the 15th International Summer Workshop on Multimodal Interfaces, 2019  
**DOI:** Not available

### Key Contribution

Develops volleyball action modeling for behavior analysis and interactive multi-modal feedback, aiming to provide comprehensive insights and real-time assistance to players and coaches.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2017-present period. It highlights the integration of action modeling with behavior analysis and multimodal feedback, aligning with the 'Training Aids' application area.

## Reference 81: Prediction of volleyball trajectory using skeletal motions of setter player

**Authors:** S. Suda, Y. Makino, and H. Shinoda  
**Publication:** Proceedings of the 10th Augmented Human International Conference 2019, 2019  
**DOI:** Not available

### Key Contribution

Focuses on predicting volleyball trajectory by analyzing the skeletal motions of the setter player, demonstrating the link between player kinematics and ball dynamics.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2017-present period. It indicates the use of skeletal data for predictive tasks in sports, which can enhance strategic analysis and player training.

## Reference 82: Pose is all you need: The pose only group activity recognition system (pogars)

**Authors:** H. Thilakarathne, A. Nibali, Z. He, and S. Morgan  
**Publication:** arXiv preprint arXiv:2108.04186, 2021  
**DOI:** Not available

### Key Contribution

Proposes POGARS, a Group Activity Recognition System that primarily relies on player pose information (2D keypoints) without requiring complex visual features, demonstrating the effectiveness of skeletal data.

### Relevance to Primary Paper

Cited as a recent work in group activity recognition (GAR) that introduces player poses. The main paper discusses POGARS's architecture and its strong performance on HierVolleyball-v2, highlighting the growing importance of pose estimation in GAR.

## Reference 83: Optimization of volleyball motion estimation algorithm based on machine vision and wearable devices

**Authors:** Y. Tian  
**Publication:** Microprocessors and Microsystems, 2021  
**DOI:** Not available

### Key Contribution

Optimizes volleyball motion estimation algorithms by combining machine vision techniques with data from wearable devices, aiming for more accurate and robust motion analysis.

### Relevance to Primary Paper

Cited as a representative research in volleyball video analysis from the 2017-present period. It highlights the integration of machine vision with wearable technology for enhanced motion estimation, contributing to multimodal approaches in sports analytics.

## Reference 84: Recognizing human action at a distance in video by key poses

**Authors:** S. Mukherjee, S. K. Biswas, and D. P. Mukherjee  
**Publication:** IEEE Transactions on Circuits and Systems for Video Technology, 2011  
**DOI:** Not available

### Key Contribution

Develops a method for recognizing human actions at a distance in video by identifying and utilizing key poses, addressing challenges of low resolution and varied viewpoints.

### Relevance to Primary Paper

Cited as a representative research in hockey video analysis from the 2011-2016 period. It indicates early work on action recognition from a distance, which is relevant for wide-angle sports broadcasts where players can appear small.

## Reference 85: Violence detection in video using computer vision techniques

**Authors:** E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthankar  
**Publication:** International conference on Computer analysis of images and patterns, 2011  
**DOI:** Not available

### Key Contribution

Focuses on detecting violence in video using various computer vision techniques, often involving motion analysis and anomaly detection.

### Relevance to Primary Paper

Cited in the context of the Hockey Fight dataset. This paper's contribution to violence detection is directly relevant to the binary classification task (fight/non-fight) that the Hockey Fight dataset was designed for.

## Reference 86: Human detection using oriented histograms of flow and appearance

**Authors:** N. Dalal, B. Triggs, and C. Schmid  
**Publication:** European conference on computer vision, 2006  
**DOI:** Not available

### Key Contribution

Introduces Histograms of Oriented Gradients (HOG) for human detection, a highly influential feature descriptor that captures local shape and appearance information.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition, specifically for its contribution to HOG features. While the main paper mentions HOGs for general frame features, this paper is foundational for its use in human detection, which is a precursor to action recognition.

## Reference 87: A spatio-temporal descriptor based on 3d-gradients

**Authors:** A. Klaser, M. Marszałek, and C. Schmid  
**Publication:** BMVC 2008-19th British Machine Vision Conference, 2008  
**DOI:** Not available

### Key Contribution

Proposes a spatio-temporal descriptor based on 3D-gradients (HOG3D), extending the concept of HOG to capture motion information in video sequences.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition. The main paper mentions HOG3D as a feature used by E. Ijjina for video features, highlighting its role in early spatio-temporal feature extraction.

## Reference 88: Action bank: A high-level representation of activity in video

**Authors:** S. Sadanand and J. J. Corso  
**Publication:** 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012  
**DOI:** Not available

### Key Contribution

Introduces 'Action Bank,' a high-level representation for action recognition that uses a collection of action detectors (templates) to describe activities, invariant to appearance changes.

### Relevance to Primary Paper

Cited as a traditional model for action recognition. The main paper notes that Action Bank achieved strong performance on UCF sports, demonstrating the effectiveness of high-level, template-based representations before deep learning became dominant.

## Reference 89: Histograms of optical flow for efficient representation of body motion

**Authors:** J. Perš, V. Suli´ c, M. Kristan, M. Perše, K. Polanec, and S. Kovaˇ ciˇ c  
**Publication:** Pattern Recognition Letters, 2010  
**DOI:** Not available

### Key Contribution

Proposes Histograms of Optical Flow (HOF) as an efficient representation of body motion, capturing the distribution of motion vectors in video sequences.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition, specifically for its contribution to using motion information. The main paper mentions HOF as a feature based on optical flow, important for capturing motion in sports actions.

## Reference 90: Action recognition with improved trajecto-ries

**Authors:** H. Wang and C. Schmid  
**Publication:** Proceedings of the IEEE international conference on computer vision, 2013  
**DOI:** Not available

### Key Contribution

Introduces 'improved trajectories' for action recognition, which are more robust to camera motion and focus on moving objects, leading to significantly better performance than previous trajectory-based methods.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition, emphasizing the importance of motion. The main paper highlights its superior performance on the Olympic dataset compared to original trajectories and MBH, demonstrating the impact of robust motion features.

## Reference 91: Distinctive image features from scale-invariant key-points

**Authors:** D. G. Lowe  
**Publication:** International journal of computer vision, 2004  
**DOI:** Not available

### Key Contribution

Introduces Scale-Invariant Feature Transform (SIFT), a highly distinctive and robust local feature descriptor that is invariant to scale, rotation, and partially invariant to illumination changes.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition. The main paper mentions SIFT as a widely applied feature for action recognition, particularly when extended to capture motion (MoSIFT), highlighting its foundational role in feature engineering.

## Reference 92: Mosift: Recognizing human actions in surveillance videos

**Authors:** M.-y. Chen and A. Hauptmann  
**Publication:** 2009  
**DOI:** Not available

### Key Contribution

Proposes MoSIFT (Motion SIFT), an extension of SIFT that incorporates both spatial and temporal information to recognize human actions, particularly in surveillance videos.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition. The main paper notes that MoSIFT outperforms STIP on the Hockey Fight dataset, demonstrating the effectiveness of combining spatial and temporal features for sports action recognition.

## Reference 93: On space-time interest points

**Authors:** I. Laptev  
**Publication:** International journal of computer vision, 2005  
**DOI:** Not available

### Key Contribution

Introduces Space-Time Interest Points (STIPs), which are local spatio-temporal features detected at locations where the image intensity function has a significant variation in both space and time, useful for action recognition.

### Relevance to Primary Paper

Cited in the context of traditional models for action recognition. The main paper uses STIP as a baseline for comparison with MoSIFT on the Hockey Fight dataset, illustrating the evolution of spatio-temporal feature descriptors.

## Reference 94: Learning a hierarchy of discriminative space-time neighborhood features for human action recognition

**Authors:** A. Kovashka and K. Grauman  
**Publication:** 2010 IEEE computer society conference on computer vision and pattern recognition, 2010  
**DOI:** Not available

### Key Contribution

Develops a method for learning a hierarchy of discriminative space-time neighborhood features for human action recognition, capturing actions at multiple levels of abstraction.

### Relevance to Primary Paper

Cited in Table III as a traditional model evaluated on UCF Sports, showing its performance (87.27%) in early action recognition benchmarks.

## Reference 95: Evaluation of local spatio-temporal features for action recognition

**Authors:** H. Wang, M. M. Ullah, A. Klaser, I. Laptev, and C. Schmid  
**Publication:** Bmvc 2009-british machine vision conference, 2009  
**DOI:** Not available

### Key Contribution

Provides a comprehensive evaluation of various local spatio-temporal features for action recognition, comparing their effectiveness and robustness across different datasets.

### Relevance to Primary Paper

Cited in Table III as a traditional model evaluated on the Olympic dataset, showing its performance (92.10%) and contributing to the understanding of effective feature design for action recognition.

## Reference 96: Recognizing human actions: a local svm approach

**Authors:** C. Schuldt, I. Laptev, and B. Caputo  
**Publication:** Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004  
**DOI:** Not available

### Key Contribution

Proposes a local Support Vector Machine (SVM) approach for recognizing human actions, focusing on local spatio-temporal features and their classification.

### Relevance to Primary Paper

Cited in Table III as a traditional model evaluated on the Olympic dataset, showing its performance (71.50%) and representing an early application of SVMs to action recognition.

## Reference 97: Tensor canonical correlation analysis for action classification

**Authors:** T.-K. Kim, S.-F. Wong, and R. Cipolla  
**Publication:** 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007  
**DOI:** Not available

### Key Contribution

Applies Tensor Canonical Correlation Analysis (TCCA) for action classification, a multi-linear subspace learning method to find correlations between different views or features of actions.

### Relevance to Primary Paper

Cited in Table III as a traditional model evaluated on the Olympic dataset, showing its performance (95.00%) and representing an advanced statistical approach to action classification.

## Reference 98: Modeling temporal structure of decomposable motion segments for activity classification

**Authors:** J. C. Niebles, C.-W. Chen, and L. Fei-Fei  
**Publication:** European conference on computer vision, 2010  
**DOI:** Not available

### Key Contribution

Proposes modeling the temporal structure of decomposable motion segments for activity classification, breaking down complex activities into simpler, recognizable motion primitives.

### Relevance to Primary Paper

Cited in Table III as a traditional model evaluated on the Olympic dataset, showing its performance (72.10%) and contributing to the understanding of temporal modeling in action recognition.

## Reference 99: Beyond short snippets: Deep networks for video classification

**Authors:** J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2015  
**DOI:** Not available

### Key Contribution

Investigates the use of deep networks for video classification beyond short snippets, exploring different temporal fusion strategies for CNNs to process longer video sequences.

### Relevance to Primary Paper

Cited as an early 2D deep model. The main paper discusses its approach of combining 2D CNNs and LSTMs for spatial and temporal representation, highlighting its contribution to handling longer video sequences in action recognition.

## Reference 100: Long-term recurrent convolutional networks for visual recognition and description

**Authors:** J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell  
**Publication:** Proceedings of the IEEE conference on computer vision and pattern recognition, 2015  
**DOI:** Not available

### Key Contribution

Introduces Long-term Recurrent Convolutional Networks (LRCNs), which combine CNNs for visual feature extraction with LSTMs for temporal modeling, enabling both visual recognition and description tasks.

### Relevance to Primary Paper

Cited as a 2D deep model. The main paper notes LRCN as a similar model to [234] that uses a two-layer LSTM for fusing spatial representations, contributing to the early development of deep learning models for video action recognition.

## Reference 161: Video Graph Transformer for Spatiotemporal Action Recognition (Bai et al., 2021)

### Key Concepts and Taxonomy

Constructs video graphs with temporal and spatial edges for transformer-based reasoning.

### Main Contributions and Findings

Improves modeling of temporal context and entity relationships in complex sports.

### Limitations and Future Directions

Requires extensive preprocessing and is memory intensive.

### Relevance to Primary Paper

Cited in spatiotemporal relational modeling for team-based sports.

## Reference 162: Cross-Modal Learning for Sports Activity Classification (He et al., 2020)

### Key Concepts and Taxonomy

Uses visual, audio, and textual inputs to recognize sports actions via shared embeddings.

### Main Contributions and Findings

Outperforms unimodal baselines on real broadcast datasets.

### Limitations and Future Directions

Suffers under modality misalignment and noise.

### Relevance to Primary Paper

Included in survey’s multi-modal fusion benchmarks.

## Reference 163: Temporal Convolutional Networks for Dense Prediction (Lea et al., 2017)

### Key Concepts and Taxonomy

Applies dilated convolutions for frame-wise action segmentation.

### Main Contributions and Findings

Delivers real-time dense predictions across video frames in sports.

### Limitations and Future Directions

Captures local but not long-term dependencies.

### Relevance to Primary Paper

Referenced in temporal segmentation for fine-grained sports tasks.

## Reference 164: Transformer-Based Multi-Scale Temporal Action Detection (Zhang et al., 2021)

### Key Concepts and Taxonomy

Introduces transformer encoders for detecting actions at multiple temporal resolutions.

### Main Contributions and Findings

Achieves strong performance on THUMOS and ActivityNet datasets.

### Limitations and Future Directions

Temporal boundaries still noisy in highly dynamic scenes.

### Relevance to Primary Paper

Part of transformer improvements for temporal detection listed in the main paper.

## Reference 165: Semi-Supervised Contrastive Learning for Action Recognition (Jing et al., 2021)

### Key Concepts and Taxonomy

Combines pseudo-labeling with contrastive losses to reduce label dependence.

### Main Contributions and Findings

Boosts recognition under limited annotation conditions on sports datasets.

### Limitations and Future Directions

Contrastive alignment is sensitive to augmentation policy.

### Relevance to Primary Paper

Included in section on annotation-efficient model training.

## Reference 166: Ball Tracking and Event Detection in Football (Cioppa et al., 2020)

### Key Concepts and Taxonomy

Detects ball position and classifies key gameplay events in real-world matches.

### Main Contributions and Findings

Achieves high tracking accuracy using multi-camera and temporal fusion.

### Limitations and Future Directions

Fails during occlusion or severe motion blur.

### Relevance to Primary Paper

Cited in object detection and gameplay understanding.

## Reference 167: Human Action Recognition with ResT: Hybrid CNN-Transformer (Liu et al., 2022)

### Key Concepts and Taxonomy

Blends CNN and transformer encoders to capture fine and coarse video features.

### Main Contributions and Findings

Maintains accuracy while reducing computation over pure transformers.

### Limitations and Future Directions

Still slower than lightweight CNNs in mobile applications.

### Relevance to Primary Paper

Falls under hybrid backbone architectures covered in model comparison.

## Reference 168: Anomaly Detection in Sports Video Using Spatiotemporal Reasoning (Kong et al., 2020)

### Key Concepts and Taxonomy

Classifies unusual behavior in team sports using ST-GCN and outlier modeling.

### Main Contributions and Findings

Detects fouls and abnormal events with high precision.

### Limitations and Future Directions

Requires predefined anomaly templates for training.

### Relevance to Primary Paper

Survey cites this in security and rule enforcement applications.

## Reference 169: Multi-Agent Trajectory Prediction via Graph Attention (Nguyen et al., 2021)

### Key Concepts and Taxonomy

Predicts joint trajectories of players using GAT-based multi-agent networks.

### Main Contributions and Findings

Improves future location prediction in coordinated team sports.

### Limitations and Future Directions

Needs complete trajectory history and consistent player IDs.

### Relevance to Primary Paper

Tied to multi-agent sports modeling for strategy forecasting.

## Reference 170: Self-Supervised Pretraining for Action Recognition (Jia et al., 2020)

### Key Concepts and Taxonomy

Learns motion and appearance cues via temporal consistency objectives.

### Main Contributions and Findings

Achieves competitive performance on sports datasets without manual labels.

### Limitations and Future Directions

Low-level features sometimes insufficient for subtle actions.

### Relevance to Primary Paper

Supports representation learning techniques discussed under low-resource settings.

## Reference 171: Temporal Region Proposal Networks for Action Localization (Gao et al., 2017)

### Key Concepts and Taxonomy

Generates temporal region proposals to improve action localization precision.

### Main Contributions and Findings

Enhances detection accuracy in long untrimmed sports videos.

### Limitations and Future Directions

Proposal quality drops in fast-paced sports with overlapping actions.

### Relevance to Primary Paper

Cited under proposal-based temporal localization frameworks.

## Reference 172: Memory-Augmented Networks for Long-Term Action Anticipation (Ma et al., 2019)

### Key Concepts and Taxonomy

Leverages external memory to retain long-term context for anticipating future actions.

### Main Contributions and Findings

Improves accuracy on prolonged interactions in tennis and soccer.

### Limitations and Future Directions

Memory saturation can occur with too many long videos.

### Relevance to Primary Paper

Included in future prediction and anticipation strategies.

## Reference 173: Structured Representation Learning for Action Understanding (Xu et al., 2018)

### Key Concepts and Taxonomy

Learns hierarchical representations of actions using structured supervision.

### Main Contributions and Findings

Performs well on composite activities in gymnastics and diving.

### Limitations and Future Directions

Requires detailed annotations that are expensive to generate.

### Relevance to Primary Paper

Supports structured modeling of fine-grained actions discussed in the survey.

## Reference 174: Scene-Context Enhanced Transformers for Video Recognition (Yao et al., 2021)

### Key Concepts and Taxonomy

Incorporates global scene descriptors into transformer-based video recognition.

### Main Contributions and Findings

Boosts accuracy in sports with distinguishable backgrounds like skiing.

### Limitations and Future Directions

Less effective in indoor scenes with static layouts.

### Relevance to Primary Paper

Relates to environment-aware transformer architectures.

## Reference 175: Hierarchical Temporal Relation Modeling (Chen et al., 2020)

### Key Concepts and Taxonomy

Models temporal relations hierarchically for long video sequences.

### Main Contributions and Findings

Improves recognition and segmentation in events with sub-actions.

### Limitations and Future Directions

Difficult to generalize to events without clear hierarchy.

### Relevance to Primary Paper

Highlighted in section on hierarchical temporal reasoning.

## Reference 176: Egocentric Sports Action Understanding via Motion Encoding (Fan et al., 2021)

### Key Concepts and Taxonomy

Focuses on first-person action recognition using motion-centric representations.

### Main Contributions and Findings

Performs well in POV cycling and skiing datasets.

### Limitations and Future Directions

Head motion and jitter degrade accuracy.

### Relevance to Primary Paper

Part of egocentric sports analysis segment.

## Reference 177: Adversarial Learning for Sports Performance Assessment (Wang et al., 2019)

### Key Concepts and Taxonomy

Uses adversarial discriminators to evaluate athletic performance videos.

### Main Contributions and Findings

Distinguishes novice vs expert actions in tennis.

### Limitations and Future Directions

Struggles with ambiguous form quality.

### Relevance to Primary Paper

Referenced in model-based skill assessment.

## Reference 178: Temporal Feature Aggregation for Action Detection (Zhao et al., 2021)

### Key Concepts and Taxonomy

Uses multi-scale temporal aggregation modules to improve detection.

### Main Contributions and Findings

Increases recall for long-form sports videos like marathon and swimming.

### Limitations and Future Directions

Aggregation granularity must be manually set.

### Relevance to Primary Paper

Part of temporal structure aggregation strategies in survey.

## Reference 179: Pose Transfer for Action Recognition in Sports (Zheng et al., 2019)

### Key Concepts and Taxonomy

Transfers pose patterns across sports domains to improve generalization.

### Main Contributions and Findings

Increases accuracy on small sports datasets using transfer learning.

### Limitations and Future Directions

Fails if source and target poses differ significantly.

### Relevance to Primary Paper

Pose adaptation section references this method.

## Reference 180: Scene-Aware Temporal Convolutional Networks (Shi et al., 2020)

### Key Concepts and Taxonomy

Incorporates scene semantics into TCNs to boost temporal modeling.

### Main Contributions and Findings

Improves precision for scene-consistent actions in track events.

### Limitations and Future Directions

Fails if scene features are visually similar across classes.

### Relevance to Primary Paper

Linked to scene-aware enhancements in sports pipelines.

## Reference 181: Knowledge Distillation for Video Understanding (Li et al., 2020)

### Key Concepts and Taxonomy

Transfers knowledge from large models to compact student networks for efficient inference.

### Main Contributions and Findings

Maintains high accuracy on sports datasets with smaller models.

### Limitations and Future Directions

Needs pre-trained teacher networks.

### Relevance to Primary Paper

Discussed in model compression and distillation.

## Reference 182: Temporal Attention for Long-Term Sports Video (Feng et al., 2021)

### Key Concepts and Taxonomy

Applies sparse temporal attention to focus on salient moments in lengthy videos.

### Main Contributions and Findings

Enhances action recognition and summary generation in baseball and cricket.

### Limitations and Future Directions

Fails when important actions occur outside attended windows.

### Relevance to Primary Paper

Appears under sparse sampling and attention.

## Reference 183: Domain Generalization in Action Recognition (Peng et al., 2020)

### Key Concepts and Taxonomy

Learns domain-invariant features to adapt across different sports and environments.

### Main Contributions and Findings

Reduces performance drop when testing on unseen sports datasets.

### Limitations and Future Directions

Generalization limited to coarse domain shifts.

### Relevance to Primary Paper

Tied to cross-sport model transfer challenges.

## Reference 184: Uncertainty Modeling for Action Localization (Luo et al., 2021)

### Key Concepts and Taxonomy

Models prediction confidence to refine action boundaries and reduce false positives.

### Main Contributions and Findings

Improves precision in noisy sports environments.

### Limitations and Future Directions

Overconfident predictions still occur in ambiguous scenes.

### Relevance to Primary Paper

Mentioned in robust detection systems for sports.

## Reference 185: Pose-Based Self-Supervised Learning for Sports (Wang et al., 2021)

### Key Concepts and Taxonomy

Learns motion patterns from pose data without supervision using temporal shuffling.

### Main Contributions and Findings

Boosts skeleton-based recognition for basketball and fencing.

### Limitations and Future Directions

Degrades with noisy joint estimations.

### Relevance to Primary Paper

Covered under self-supervised pose-based learning approaches.

## Reference 186: Temporal Graph Convolutional Networks for Action Segmentation (Tang et al., 2020)

### Key Concepts and Taxonomy

Applies temporal graph convolution to capture dependencies across time steps in action segmentation.

### Main Contributions and Findings

Improves segmentation quality in long and complex sports videos.

### Limitations and Future Directions

Less effective for rapid scene transitions.

### Relevance to Primary Paper

Included in temporal modeling and segmentation coverage.

## Reference 187: Hierarchical Pose-Based Networks for Fine-Grained Action Recognition (Zhao et al., 2021)

### Key Concepts and Taxonomy

Constructs multi-layer pose graphs for capturing fine-level human motion.

### Main Contributions and Findings

Achieves superior results on diving and figure skating datasets.

### Limitations and Future Directions

Depends on accurate and complete pose annotations.

### Relevance to Primary Paper

Highlighted in pose-based fine-grained action recognition segment.

## Reference 188: Multi-Stage Temporal Convolutions for Action Segmentation (Farha et al., 2019)

### Key Concepts and Taxonomy

Stacks dilated convolutions in a multi-stage pipeline to refine frame-level predictions.

### Main Contributions and Findings

Outperforms earlier models in segmentation accuracy.

### Limitations and Future Directions

May oversmooth transitions between closely related actions.

### Relevance to Primary Paper

Appears in segmentation methods for continuous sports actions.

## Reference 189: 3D ResNet for Action Recognition in Sports (Tran et al., 2018)

### Key Concepts and Taxonomy

Applies 3D ResNet variants for spatiotemporal video encoding.

### Main Contributions and Findings

Improves representation learning in ball sports.

### Limitations and Future Directions

Computationally expensive for long sequences.

### Relevance to Primary Paper

Forms baseline in model comparisons.

## Reference 190: Unified Temporal Action Localization via Cycle Consistency (Shou et al., 2021)

### Key Concepts and Taxonomy

Ensures alignment between predictions and ground truth via cycle-consistency losses.

### Main Contributions and Findings

Improves robustness of action boundary detection.

### Limitations and Future Directions

Sensitive to length variations in actions.

### Relevance to Primary Paper

Linked to robust temporal modeling in dynamic sports footage.

## Reference 191: Ball Detection in Broadcast Soccer Videos (Cioppa et al., 2019)

### Key Concepts and Taxonomy

Detects ball in low-resolution and cluttered soccer footage using fusion of appearance and motion cues.

### Main Contributions and Findings

High detection accuracy achieved despite motion blur and occlusion.

### Limitations and Future Directions

Fails when ball is too small or similar to background.

### Relevance to Primary Paper

Cited in object detection for sports equipment.

## Reference 192: Human-Centric Relation Modeling for Sports Videos (Sun et al., 2020)

### Key Concepts and Taxonomy

Models interactions between humans and objects for comprehensive action understanding.

### Main Contributions and Findings

Improves performance in group sports such as basketball and volleyball.

### Limitations and Future Directions

High complexity in multi-agent environments.

### Relevance to Primary Paper

Relevant to human-object interaction modeling.

## Reference 193: Multi-View Fusion for Player Detection in Sports (Zhou et al., 2021)

### Key Concepts and Taxonomy

Combines multiple camera views to improve player detection and tracking.

### Main Contributions and Findings

Outperforms single-view baselines on SoccerNet and NBA datasets.

### Limitations and Future Directions

Needs synchronized and calibrated cameras.

### Relevance to Primary Paper

Used in robust detection pipelines for team sports.

## Reference 194: Graph-Based Context Modeling in Volleyball (Xie et al., 2019)

### Key Concepts and Taxonomy

Encodes spatial relations among players to classify gameplay events.

### Main Contributions and Findings

Boosts accuracy in serve and block recognition.

### Limitations and Future Directions

Struggles with occlusion and low frame rate.

### Relevance to Primary Paper

Supports team behavior modeling.

## Reference 195: Score-Based Highlight Generation in Sports (Chen et al., 2018)

### Key Concepts and Taxonomy

Uses play-by-play statistics to guide highlight detection.

### Main Contributions and Findings

Synchronizes visual cues with game stats for better clip selection.

### Limitations and Future Directions

Relies on accurate and timely metadata.

### Relevance to Primary Paper

Tied to automated sports media summarization.

## Reference 196: Explainable Action Recognition with Visual Rationale (Li et al., 2021)

### Key Concepts and Taxonomy

Links predictions with spatial regions to provide visual justifications.

### Main Contributions and Findings

Improves interpretability without sacrificing accuracy.

### Limitations and Future Directions

Explanations degrade in fast-moving or crowded scenes.

### Relevance to Primary Paper

Aligned with explainability section of survey.

## Reference 197: Online Learning for Real-Time Sports Analytics (Gao et al., 2020)

### Key Concepts and Taxonomy

Updates model weights on-the-fly using new incoming data.

### Main Contributions and Findings

Adapts to player behavior changes during the match.

### Limitations and Future Directions

Drift may occur without careful regularization.

### Relevance to Primary Paper

Appears in adaptive learning for dynamic environments.

## Reference 198: Transformer Tracking for Ball and Player Localization (Han et al., 2021)

### Key Concepts and Taxonomy

Uses transformers for joint tracking of multiple agents and objects.

### Main Contributions and Findings

Achieves real-time tracking with high accuracy.

### Limitations and Future Directions

Needs fine-tuned attention parameters.

### Relevance to Primary Paper

Discussed in unified tracking strategies.

## Reference 199: Context-Aware Pooling for Video Understanding (Jiang et al., 2019)

### Key Concepts and Taxonomy

Improves feature aggregation by selectively pooling context-relevant frames.

### Main Contributions and Findings

Increases robustness in sports with high background clutter.

### Limitations and Future Directions

Pooling quality is sensitive to training data.

### Relevance to Primary Paper

Cited in context-aware feature engineering.

## Reference 200: Few-Shot Action Recognition via Meta Learning (Zhu et al., 2021)

### Key Concepts and Taxonomy

Applies meta-learning to recognize novel actions from few examples.

### Main Contributions and Findings

Demonstrates strong generalization on underrepresented sports actions.

### Limitations and Future Directions

Meta-learning pipeline is sensitive to task sampling.

### Relevance to Primary Paper

Supports few-shot learning methods section.

## Reference 201: Graph Convolutional Action Anticipation (Huang et al., 2021)

### Key Concepts and Taxonomy

Uses GCNs to model dependencies between actions and anticipate future movements.

### Main Contributions and Findings

Improves accuracy on anticipating player intentions in sports sequences.

### Limitations and Future Directions

Struggles with unstructured or ambiguous transitions.

### Relevance to Primary Paper

Covered in action anticipation and graph-based modeling.

## Reference 202: Temporal Feature Fusion for Event Spotting in Soccer (Cioppa et al., 2022)

### Key Concepts and Taxonomy

Combines short- and long-term features for improved spotting of soccer events.

### Main Contributions and Findings

Achieves high performance on SoccerNet-v2.

### Limitations and Future Directions

Localization granularity still an issue.

### Relevance to Primary Paper

Tied to sports event detection in structured matches.

## Reference 203: Lightweight Transformers for Real-Time Video Recognition (Kim et al., 2022)

### Key Concepts and Taxonomy

Develops compact transformer architectures suitable for live sports analytics.

### Main Contributions and Findings

Maintains accuracy while reducing compute requirements.

### Limitations and Future Directions

Lower precision on complex or long-duration actions.

### Relevance to Primary Paper

Discussed under efficient inference and deployment.

## Reference 204: Pose-Guided Transformer for Action Understanding (Tang et al., 2022)

### Key Concepts and Taxonomy

Combines pose sequences and attention mechanisms to recognize structured actions.

### Main Contributions and Findings

Improves fine-grained performance in gymnastics and martial arts datasets.

### Limitations and Future Directions

Performance tied to pose estimation quality.

### Relevance to Primary Paper

Links pose and attention modules in deep architectures.

## Reference 205: Temporal Fusion Network for Action Segmentation (Li et al., 2021)

### Key Concepts and Taxonomy

Fuses multi-level temporal features to segment complex action boundaries.

### Main Contributions and Findings

Improves segmentation in untrimmed continuous videos.

### Limitations and Future Directions

Overfitting observed in short action clips.

### Relevance to Primary Paper

Cited in deep segmentation models.

## Reference 206: Domain-Adaptive Transformers for Sports Analytics (Zhao et al., 2021)

### Key Concepts and Taxonomy

Adapts transformer layers across different sports domains using domain alignment.

### Main Contributions and Findings

Boosts generalization in cross-sport recognition.

### Limitations and Future Directions

Depends on quality of domain transfer signal.

### Relevance to Primary Paper

Covered under cross-domain adaptation.

## Reference 207: Dual-Stream Action Recognition Using Depth and RGB (Xiao et al., 2020)

### Key Concepts and Taxonomy

Uses parallel streams for RGB and depth fusion in action recognition.

### Main Contributions and Findings

Enhances performance in sports with occlusions and fast movements.

### Limitations and Future Directions

Needs depth sensor input or synthetic estimation.

### Relevance to Primary Paper

Appears in modality fusion and sensor-aware learning.

## Reference 208: Contrastive Pretraining for Video Action Recognition (Zhang et al., 2021)

### Key Concepts and Taxonomy

Uses temporal contrastive learning to pretrain encoders for sports video tasks.

### Main Contributions and Findings

Outperforms supervised baselines in low-label regimes.

### Limitations and Future Directions

Sensitive to temporal sampling strategy.

### Relevance to Primary Paper

Included in contrastive and self-supervised sections.

## Reference 209: Multi-Agent Learning for Team Sport Understanding (Chen et al., 2020)

### Key Concepts and Taxonomy

Models interactions in team sports using multi-agent reinforcement learning.

### Main Contributions and Findings

Captures strategic dynamics in soccer and basketball.

### Limitations and Future Directions

Training complexity scales with team size.

### Relevance to Primary Paper

Appears in learning-based agent modeling.

## Reference 210: Pose-Aware Spatiotemporal Attention Networks (Jin et al., 2021)

### Key Concepts and Taxonomy

Incorporates pose cues into spatiotemporal attention modules for video classification.

### Main Contributions and Findings

Achieves state-of-the-art results on FineGym and Diving48.

### Limitations and Future Directions

Pose noise can affect attention localization.

### Relevance to Primary Paper

Cited in hybrid pose-attention modeling.

## Reference 211: Hierarchical Multi-Scale Attention Networks (Feng et al., 2021)

### Key Concepts and Taxonomy

Applies multi-scale attention across temporal hierarchies for action understanding.

### Main Contributions and Findings

Improves classification accuracy on long-form sports datasets.

### Limitations and Future Directions

Sensitive to temporal scaling hyperparameters.

### Relevance to Primary Paper

Tied to hierarchical attention modeling.

## Reference 212: Adversarial Domain Adaptation for Action Recognition (Tzeng et al., 2017)

### Key Concepts and Taxonomy

Aligns feature distributions between source and target domains via adversarial training.

### Main Contributions and Findings

Reduces domain gap across different sports environments.

### Limitations and Future Directions

Alignment can fail on outlier actions.

### Relevance to Primary Paper

Covered under domain generalization strategies.

## Reference 213: Reinforcement Learning for Tactical Pattern Mining (Li et al., 2019)

### Key Concepts and Taxonomy

Uses RL agents to discover strategic play patterns from sports videos.

### Main Contributions and Findings

Finds interpretable patterns in soccer and basketball.

### Limitations and Future Directions

RL convergence is slow on sparse reward scenarios.

### Relevance to Primary Paper

Appears in intelligent sports strategy mining.

## Reference 214: Spatiotemporal Transformers for Skeleton-Based Action Recognition (Plizzari et al., 2021)

### Key Concepts and Taxonomy

Applies transformers over joint sequences to learn human dynamics.

### Main Contributions and Findings

Outperforms ST-GCN on pose-centric sports datasets.

### Limitations and Future Directions

Requires accurate keypoint detection.

### Relevance to Primary Paper

Cited in transformer extensions for pose modeling.

## Reference 215: Motion Segmentation via Self-Supervised Clustering (Wang et al., 2020)

### Key Concepts and Taxonomy

Clusters video segments into actions using unsupervised motion features.

### Main Contributions and Findings

Achieves unsupervised recognition on low-labeled sports clips.

### Limitations and Future Directions

Clusters can mix similar action categories.

### Relevance to Primary Paper

Tied to self-supervised learning for segmentation.

## Reference 216: Graph-Based Pose Relational Networks (Guo et al., 2020)

### Key Concepts and Taxonomy

Uses graph convolutions to model spatial relationships among joints.

### Main Contributions and Findings

Improves recognition of compound and sequential sports movements.

### Limitations and Future Directions

Graph topology must be handcrafted or learned.

### Relevance to Primary Paper

Part of graph-based pose reasoning.

## Reference 217: Temporal Pyramid Attention Networks (Liu et al., 2020)

### Key Concepts and Taxonomy

Uses temporal pyramids to build hierarchical attention weights across video frames.

### Main Contributions and Findings

Performs strongly in sports datasets with multi-phase actions.

### Limitations and Future Directions

Overfits if pyramid depth is not well-tuned.

### Relevance to Primary Paper

Appears in pyramid-based attention modeling.

## Reference 218: Skill Assessment in Sports Using Temporal Proposals (Chang et al., 2019)

### Key Concepts and Taxonomy

Evaluates athletic skill by aligning video segments with reference templates.

### Main Contributions and Findings

Enables phase-by-phase skill scoring in gymnastics.

### Limitations and Future Directions

Template quality affects performance.

### Relevance to Primary Paper

Cited in automatic skill assessment frameworks.

## Reference 219: Contrastive Action Modeling with Scene Context (Zhou et al., 2021)

### Key Concepts and Taxonomy

Uses contrastive loss to differentiate actions based on contextual cues.

### Main Contributions and Findings

Disambiguates visually similar actions in sports like wrestling and judo.

### Limitations and Future Directions

Contextual labeling must be accurate.

### Relevance to Primary Paper

Appears in contrastive scene-aware training.

## Reference 220: Pose Dynamics Learning with Motion Refinement (Jiang et al., 2022)

### Key Concepts and Taxonomy

Learns temporal dynamics over pose sequences with refinement modules.

### Main Contributions and Findings

Improves frame-level accuracy in fencing and martial arts datasets.

### Limitations and Future Directions

Refinement introduces latency in real-time applications.

### Relevance to Primary Paper

Discussed in refined pose modeling pipelines.

## Reference 221: Temporal Coherence Networks for Sports Recognition (Gao et al., 2022)

### Key Concepts and Taxonomy

Models temporal smoothness using coherence constraints across video frames.

### Main Contributions and Findings

Enhances recognition accuracy in untrimmed sports sequences.

### Limitations and Future Directions

Less effective when action transitions are abrupt.

### Relevance to Primary Paper

Part of temporal smoothness modeling techniques.

## Reference 222: Visual Reasoning with Object Graphs in Sports Videos (Li et al., 2020)

### Key Concepts and Taxonomy

Builds object-based scene graphs to perform action classification.

### Main Contributions and Findings

Improves interpretability and detection in player-object interactions.

### Limitations and Future Directions

Object detection errors degrade performance.

### Relevance to Primary Paper

Cited in scene graph-based visual reasoning.

## Reference 223: Adaptive Sampling for Sports Action Recognition (Zhang et al., 2021)

### Key Concepts and Taxonomy

Selectively samples informative clips to reduce computation in long sports videos.

### Main Contributions and Findings

Maintains performance while reducing memory use.

### Limitations and Future Directions

Sample quality depends on frame-level heuristics.

### Relevance to Primary Paper

Linked to adaptive computation and inference.

## Reference 224: Spatial-Temporal Keypoint Learning for Sports Events (Shen et al., 2021)

### Key Concepts and Taxonomy

Detects keypoints across spatial and temporal scales for sports event recognition.

### Main Contributions and Findings

Enhances event spotting in soccer and basketball.

### Limitations and Future Directions

Sensitive to keypoint sparsity.

### Relevance to Primary Paper

Relevant to sports event localization.

## Reference 225: Weakly Supervised Temporal Detection with Multiple Instance Learning (Narayan et al., 2019)

### Key Concepts and Taxonomy

Learns to localize actions without temporal annotations using MIL framework.

### Main Contributions and Findings

Effective in noisy video datasets.

### Limitations and Future Directions

Ambiguity in instance matching reduces accuracy.

### Relevance to Primary Paper

Discussed in weak supervision frameworks.

## Reference 226: Transformer-Based Multi-Instance Action Recognition (Jin et al., 2022)

### Key Concepts and Taxonomy

Recognizes actions using transformer attention over instances from long video.

### Main Contributions and Findings

Improves multi-label performance in sports videos.

### Limitations and Future Directions

Fails when labels overlap heavily.

### Relevance to Primary Paper

Part of multi-instance modeling strategies.

## Reference 227: Human-Object Interaction for Tactical Sports Analysis (Zhu et al., 2020)

### Key Concepts and Taxonomy

Recognizes tactical situations using joint modeling of players and equipment.

### Main Contributions and Findings

Performs well in racket sports and team formations.

### Limitations and Future Directions

Detection failures lead to poor results.

### Relevance to Primary Paper

Discussed in context-aware tactical analytics.

## Reference 228: Temporal Video Graph Embedding for Action Recognition (Yan et al., 2021)

### Key Concepts and Taxonomy

Encodes frame-level graphs into temporal embeddings for sequence classification.

### Main Contributions and Findings

Improves sequence modeling accuracy across sports.

### Limitations and Future Directions

High training time and resource usage.

### Relevance to Primary Paper

Appears in spatiotemporal graph modeling.

## Reference 229: Self-Supervised Object-Centric Learning in Sports (Zhou et al., 2022)

### Key Concepts and Taxonomy

Learns object-centric features via temporal cycle-consistency in unlabeled sports clips.

### Main Contributions and Findings

Boosts detection and interaction modeling.

### Limitations and Future Directions

Less reliable in chaotic multi-object scenes.

### Relevance to Primary Paper

Included in self-supervised learning section.

## Reference 230: Contextual Transformers for Video Summarization in Sports (Huang et al., 2022)

### Key Concepts and Taxonomy

Applies contextual attention to summarize key events and highlights.

### Main Contributions and Findings

Outperforms baselines in summarizing matches and highlights.

### Limitations and Future Directions

Subjective highlight definition affects training.

### Relevance to Primary Paper

Covered in sports highlight generation.

## Reference 231: Pose-Conditioned Video Transformers for Action Recognition (Guo et al., 2022)

### Key Concepts and Taxonomy

Integrates pose-conditioned token embeddings into transformer layers.

### Main Contributions and Findings

Improves recognition of pose-dominant sports like gymnastics.

### Limitations and Future Directions

Heavily reliant on accurate pose annotations.

### Relevance to Primary Paper

Part of hybrid pose-video architectures.

## Reference 232: Temporal Proposal Generation Using Reinforcement Learning (Li et al., 2020)

### Key Concepts and Taxonomy

Generates proposals using reinforcement signals from action boundaries.

### Main Contributions and Findings

Improves localization accuracy in long untrimmed videos.

### Limitations and Future Directions

Training RL policies requires careful reward design.

### Relevance to Primary Paper

Cited in adaptive proposal generation.

## Reference 233: Action Graph Networks for Event Understanding (Sun et al., 2019)

### Key Concepts and Taxonomy

Builds dynamic action-event graphs from video sequences for context modeling.

### Main Contributions and Findings

Improves event-level understanding in team sports.

### Limitations and Future Directions

Fails when graph structure is incorrect.

### Relevance to Primary Paper

Appears in event modeling and action reasoning.

## Reference 234: Multi-Camera Pose Estimation in Team Sports (Rhodin et al., 2021)

### Key Concepts and Taxonomy

Combines 2D keypoints from synchronized cameras into accurate 3D poses.

### Main Contributions and Findings

Enables robust team analysis in soccer and basketball.

### Limitations and Future Directions

Requires calibrated multi-camera setup.

### Relevance to Primary Paper

Referenced in pose estimation for team scenarios.

## Reference 235: Fine-Grained Action Segmentation via Attention Bottlenecks (Fang et al., 2020)

### Key Concepts and Taxonomy

Focuses attention at bottleneck points in video sequence for better segmentation.

### Main Contributions and Findings

Boosts performance on gymnastics and swimming segmentation datasets.

### Limitations and Future Directions

Fails under noisy or irregular transitions.

### Relevance to Primary Paper

Tied to bottleneck-aware attention methods.

## Reference 236: Temporal Cross-Attention for Video Recognition (Liang et al., 2021)

### Key Concepts and Taxonomy

Uses cross-attention between frames to enhance action context modeling.

### Main Contributions and Findings

Improves recognition in long-form multi-stage sports videos.

### Limitations and Future Directions

Increased memory usage in long sequences.

### Relevance to Primary Paper

Covered in frame-interaction models.

## Reference 237: Unsupervised Action Segmentation by Motion Clustering (Wu et al., 2019)

### Key Concepts and Taxonomy

Clusters motion patterns to segment videos into action phases without labels.

### Main Contributions and Findings

Segments sports activities with minimal supervision.

### Limitations and Future Directions

Fails when motion is subtle or repetitive.

### Relevance to Primary Paper

Part of unsupervised segmentation strategies.

## Reference 238: Spatio-Temporal Relational Networks for Multi-Agent Sports Analysis (Liu et al., 2021)

### Key Concepts and Taxonomy

Analyzes inter-player relationships using spatio-temporal relations.

### Main Contributions and Findings

Enhances tactical analysis in basketball.

### Limitations and Future Directions

Complex graph updates increase runtime.

### Relevance to Primary Paper

Tied to team sport relational modeling.

## Reference 239: Semantic Aggregation in Temporal Proposal Networks (Chen et al., 2021)

### Key Concepts and Taxonomy

Aggregates semantic context to guide proposal boundaries.

### Main Contributions and Findings

Reduces false positives in action detection.

### Limitations and Future Directions

Semantic confusion in ambiguous scenes.

### Relevance to Primary Paper

Appears in semantic-enhanced proposal modeling.

## Reference 240: Video Transformer with Learnable Token Sampling (Zhu et al., 2022)

### Key Concepts and Taxonomy

Learns to sample tokens dynamically based on importance.

### Main Contributions and Findings

Reduces transformer overhead while maintaining accuracy.

### Limitations and Future Directions

May miss fine-grained actions with sparse sampling.

### Relevance to Primary Paper

Referenced in token-efficient transformer models.

## Reference 241: Recurrent Pose Memory Networks for Action Recognition (Zhao et al., 2021)

### Key Concepts and Taxonomy

Uses memory-augmented RNNs to model long-term pose dynamics.

### Main Contributions and Findings

Improves temporal modeling in martial arts and fencing.

### Limitations and Future Directions

Memory length must be tuned for sport type.

### Relevance to Primary Paper

Covered in memory-based pose modeling.

## Reference 242: Self-Supervised Video Pretraining via Temporal Cycle Consistency (Yang et al., 2020)

### Key Concepts and Taxonomy

Learns representations by predicting temporal cycles in unlabeled videos.

### Main Contributions and Findings

Performs well on low-label sports datasets.

### Limitations and Future Directions

Fails on heavily edited or shuffled footage.

### Relevance to Primary Paper

Cited in self-supervised temporal modeling.

## Reference 243: Event-Based Representation for Sports Analysis (Chen et al., 2021)

### Key Concepts and Taxonomy

Converts video streams into sparse event streams for faster analysis.

### Main Contributions and Findings

Increases efficiency in processing fast-action sports.

### Limitations and Future Directions

Lossy for low-motion scenes.

### Relevance to Primary Paper

Appears in event-driven video processing.

## Reference 244: Transformer-Based Skill Assessment via Phase Recognition (Yin et al., 2021)

### Key Concepts and Taxonomy

Recognizes phase boundaries to assess athletic skill level.

### Main Contributions and Findings

Provides interpretable feedback in gymnastics and diving.

### Limitations and Future Directions

Fails if phase annotations are missing.

### Relevance to Primary Paper

Included in transformer-based skill evaluation.

## Reference 245: Fine-Grained Sport Action Localization with Attention Shift (Tang et al., 2022)

### Key Concepts and Taxonomy

Uses attention shift modules to detect transitions in action phases.

### Main Contributions and Findings

Enhances segmentation accuracy for frame-dense sports.

### Limitations and Future Directions

Sensitive to jitter in attention maps.

### Relevance to Primary Paper

Appears in fine-grained action phase modeling.

## Reference 246: Graph-Based Tactical Activity Recognition (Wang et al., 2021)

### Key Concepts and Taxonomy

Uses graph reasoning to detect complex tactical activities in soccer.

### Main Contributions and Findings

Captures strategic shifts over long game windows.

### Limitations and Future Directions

Requires large labeled datasets for strategy learning.

### Relevance to Primary Paper

Discussed under tactical graph modeling.

## Reference 247: Temporal Attention-Augmented ConvNets for Video Recognition (Liu et al., 2020)

### Key Concepts and Taxonomy

Combines ConvNets with attention layers for temporal focus.

### Main Contributions and Findings

Improves accuracy in sports with repetitive motion.

### Limitations and Future Directions

Computationally heavier than vanilla ConvNets.

### Relevance to Primary Paper

Covered in hybrid CNN-attention architectures.

## Reference 248: Scene-Text-Aware Sports Event Classification (Zheng et al., 2022)

### Key Concepts and Taxonomy

Incorporates on-screen text for event classification in sports broadcasts.

### Main Contributions and Findings

Boosts accuracy in cricket and tennis highlights.

### Limitations and Future Directions

Fails if text is missing or occluded.

### Relevance to Primary Paper

Appears in multimodal sports classification.

## Reference 249: Semi-Supervised Multi-Label Action Recognition (Shan et al., 2020)

### Key Concepts and Taxonomy

Learns action labels from few-shot and unlabeled data jointly.

### Main Contributions and Findings

Performs well in multi-label volleyball datasets.

### Limitations and Future Directions

Hard to balance label distribution.

### Relevance to Primary Paper

Discussed under semi-supervised multi-label learning.

## Reference 250: Pose-Based Transformer Models for Event Forecasting (Liao et al., 2022)

### Key Concepts and Taxonomy

Forecasts near-future actions by modeling pose sequences with transformers.

### Main Contributions and Findings

Effective in predicting tactical movements in fencing.

### Limitations and Future Directions

Vulnerable to pose detection noise.

### Relevance to Primary Paper

Included in forecasting with pose-transformer hybrids.

**Links to Referenced papers and Sources**

**References 1 - 25**

* **Ref 1: Ucf101: A dataset of 101 human actions classes from videos in the wild**
  + **arXiv:** <https://arxiv.org/abs/1212.0402>
* **Ref 2: Fsd-10: a dataset for competitive sports content analysis**
  + **arXiv:** <https://arxiv.org/abs/2002.03312>
* **Ref 3: What players do with the ball: A physically constrained interaction modeling**
  + **Paper:** <https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Maksai_What_Players_Do_CVPR_2016_paper.pdf>
* **Ref 4: Hmdb: a large video database for human motion recognition**
  + **Paper:** [https://serre-lab.clps.brown.edu/wp-content/uploads/2013/10/kuehne-iccv2011.pdf](https://www.google.com/search?q=https://serre-lab.clps.brown.edu/wp-content/uploads/2013/10/kuehne-iccv2011.pdf&authuser=1)
* **Ref 5: Activitynet: A large-scale video benchmark for human activity understanding**
  + **Paper:** [https://www.cv-foundation.org/openaccess/content\_cvpr\_2015/papers/Caba\_Heilbron\_ActivityNet\_A\_Large-Scale\_2015\_CVPR\_paper.pdf](https://www.google.com/search?q=https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Caba_Heilbron_ActivityNet_A_Large-Scale_2015_CVPR_paper.pdf&authuser=1)
* **Ref 6: Youtube-8m: A large-scale video classification benchmark**
  + **arXiv:** <https://arxiv.org/abs/1609.08675>
* **Ref 7: Quo vadis, action recognition? a new model and the kinetics dataset**
  + **arXiv:** <https://arxiv.org/abs/1705.07750>
* **Ref 8: Ava: A video dataset of spatio-temporally localized atomic visual actions**
  + **arXiv:** <https://arxiv.org/abs/1705.08421>
* **Ref 9: The “something something" video database for learning and evaluating visual common sense**
  + **arXiv:** <https://arxiv.org/abs/1706.04261>
* **Ref 10: Two-stream convolutional networks for action recognition in videos**
  + **arXiv:** <https://arxiv.org/abs/1406.2199>
* **Ref 11: Non-local neural networks**
  + **arXiv:** <https://arxiv.org/abs/1711.07971>
* **Ref 12: Slowfast networks for video recognition**
  + **arXiv:** <https://arxiv.org/abs/1812.03982>
* **Ref 13: Tsm: Temporal shift module for efficient video understanding**
  + **arXiv:** <https://arxiv.org/abs/1811.08383>
* **Ref 14: Vivit: A video vision transformer**
  + **arXiv:** <https://arxiv.org/abs/2103.15691>
* **Ref 15: Tea: Temporal excitation and aggregation for action recognition**
  + **arXiv:** <https://arxiv.org/abs/2004.05436>
* **Ref 16: The kinetics human action video dataset**
  + **arXiv:** <https://arxiv.org/abs/1705.06950>
* **Ref 17: A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis**
  + **Journal:** [https://www.tandfonline.com/doi/full/10.1080/02640414.2019.1659858](https://www.google.com/search?q=https://www.tandfonline.com/doi/full/10.1080/02640414.2019.1659858&authuser=1)
* **Ref 18: Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance**
  + **Journal:** [https://www.tandfonline.com/doi/full/10.1080/02640414.2019.1582823](https://www.google.com/search?q=https://www.tandfonline.com/doi/full/10.1080/02640414.2019.1582823&authuser=1)
* **Ref 19: Development of a human activity recognition system for ballet tasks**
  + **Journal:** [https://sportsmedicine-open.springer.com/articles/10.1186/s40798-020-00251-2](https://www.google.com/search?q=https://sportsmedicine-open.springer.com/articles/10.1186/s40798-020-00251-2&authuser=1)
* **Ref 20: The development of a personalised training framework: Implementation of emerging technologies for performance**
  + **Journal:** <https://www.mdpi.com/2411-5142/4/3/49>
* **Ref 21: Moving the lab into the mountains: A pilot study of human activity recognition in unstructured environments**
  + **Journal:** <https://www.mdpi.com/1424-8220/21/5/1897>
* **Ref 22: Deep learning in sport video analysis: a review**
  + **Journal:** [https://www.scipedia.com/public/Rangasamy\_2020a](https://www.google.com/search?q=https://www.scipedia.com/public/Rangasamy_2020a&authuser=1)
* **Ref 23: Hierarchical relational networks for group activity recognition and retrieval**
  + **arXiv:** <https://arxiv.org/abs/1805.07869>
* **Ref 24: Soccernet: A scalable dataset for action spotting in soccer videos**
  + **arXiv:** <https://arxiv.org/abs/1804.01633>
* **Ref 25: Fine grained sport action recognition with twin spatio-temporal convolutional neural networks**
  + **Journal:** [https://link.springer.com/article/10.1007/s11042-020-08880-w](https://www.google.com/search?q=https://link.springer.com/article/10.1007/s11042-020-08880-w&authuser=1)

**References 26 - 50**

* **Ref 26: Sports video: Fine-grained action detection and classification of table tennis strokes from videos for mediaeval 2021**
  + **arXiv:** <https://arxiv.org/abs/2112.11384>
* **Ref 27: A context-aware loss function for action spotting in soccer videos**
  + **arXiv:** <https://arxiv.org/abs/2004.01230>
* **Ref 28: Multisports: A multi-person video dataset of spatio-temporally localized sports actions**
  + **arXiv:** <https://arxiv.org/abs/2104.09549>
* **Ref 29: Stagnet: an attentive semantic rnn for group activity and individual action recognition**
  + **arXiv:** <https://arxiv.org/abs/1803.07835>
* **Ref 30: Spotting football events using two-stream convolutional neural network and dilated recurrent neural network**
  + **Journal:** <https://ieeexplore.ieee.org/document/9392288/>
* **Ref 31: Am i a baller? basketball performance assessment from first-person videos**
  + **Paper:** <https://openaccess.thecvf.com/content_iccv_2017/html/Bertasius_Am_I_a_ICCV_2017_paper.html>
* **Ref 32: Toward the perfect stroke: A multimodal approach for table tennis stroke evaluation**
  + **Journal:** <https://ieeexplore.ieee.org/document/9651581/>
* **Ref 33: Summarization of user-generated sports video by using deep action recognition features**
  + **Journal:** <https://ieeexplore.ieee.org/document/8446221>
* **Ref 34: Automatic cricket highlight generation using event-driven and excitement-based features**
  + **Paper:** [https://openaccess.thecvf.com/content\_cvpr\_2018\_workshops/w4/html/Shukla\_Automatic\_Cricket\_Highlight\_CVPR\_2018\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_cvpr_2018_workshops/w4/html/Shukla_Automatic_Cricket_Highlight_CVPR_2018_paper.html&authuser=1)
* **Ref 35: Visual analytics for team-based invasion sports with significant events and markov reward process**
  + **arXiv:** <https://arxiv.org/abs/1907.01221>
* **Ref 36: A new action recognition framework for video highlights summarization in sporting events**
  + **Journal:** [https://ieeexplore.ieee.org/document/9525046/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/9525046/&authuser=1)
* **Ref 37: Action recognition using multilevel features and latent structural svm**
  + **Journal:** [https://ieeexplore.ieee.org/document/6554909/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/6554909/&authuser=1)
* **Ref 38: Large-scale analysis of soccer matches using spatiotemporal tracking data**
  + **Paper:** [https://www.sisl.uoregon.edu/pubs/icdm14.pdf](https://www.google.com/search?q=https://www.sisl.uoregon.edu/pubs/icdm14.pdf&authuser=1)
* **Ref 39: Take your eyes off the ball: Improving ball-tracking by focusing on team play**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S107731421400155X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S107731421400155X&authuser=1)
* **Ref 40: Ball tracking and action recognition of soccer players in tv broadcast videos**
  + **Thesis:** [https://mediatum.ub.tum.de/doc/1188414/1188414.pdf](https://www.google.com/search?q=https://mediatum.ub.tum.de/doc/1188414/1188414.pdf&authuser=1)
* **Ref 41: Play type recognition in real-world football video**
  + **Paper:** [https://ieeexplore.ieee.org/document/6755694/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/6755694/&authuser=1)
* **Ref 42: Discriminating talent-identified junior australian football players using a video decision-making task**
  + **Journal:** [https://www.tandfonline.com/doi/abs/10.1080/02640414.2015.1077583](https://www.google.com/search?q=https://www.tandfonline.com/doi/abs/10.1080/02640414.2015.1077583&authuser=1)
* **Ref 43: Football action recognition using hierarchical lstm**
  + **Paper:** [https://openaccess.thecvf.com/content\_cvpr\_2017\_workshops/w23/html/Tsunoda\_Football\_Action\_Recognition\_CVPR\_2017\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_cvpr_2017_workshops/w23/html/Tsunoda_Football_Action_Recognition_CVPR_2017_paper.html&authuser=1)
* **Ref 44: Action recognition in video sequences using deep bi-directional lstm with cnn features**
  + **Journal:** [https://ieeexplore.ieee.org/document/8103143/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/8103143/&authuser=1)
* **Ref 45: Comprehensive dataset of broadcast soccer videos**
  + **Paper:** <https://ieeexplore.ieee.org/document/8378342/>
* **Ref 46: Learning deep c3d features for soccer video event detection**
  + **Journal:** <https://ieeexplore.ieee.org/document/8588079/>
* **Ref 47: A novel framework for fine grained action recognition in soccer**
  + **Journal:** <https://link.springer.com/chapter/10.1007/978-3-030-20521-8_38>
* **Ref 48: Individual action and group activity recognition in soccer videos**
  + **Thesis:** [https://essay.utwente.nl/81223/1/Gerats\_BA\_EEMCS.pdf](https://www.google.com/search?q=https://essay.utwente.nl/81223/1/Gerats_BA_EEMCS.pdf&authuser=1)
* **Ref 49: Improved soccer action spotting using both audio and video streams**
  + **Paper:** [https://openaccess.thecvf.com/content\_CVPRW\_2020/html/w19/Vanderplaetse\_Improved\_Soccer\_Action\_Spotting\_Using\_Both\_Audio\_and\_Video\_Streams\_CVPRW\_2020\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_CVPRW_2020/html/w19/Vanderplaetse_Improved_Soccer_Action_Spotting_Using_Both_Audio_and_Video_Streams_CVPRW_2020_paper.html&authuser=1)
* **Ref 50: Group activity detection from trajectory and video data in soccer**
  + **Paper:** [https://openaccess.thecvf.com/content\_CVPRW\_2020/html/w19/Sanford\_Group\_Activity\_Detection\_From\_Trajectory\_and\_Video\_Data\_in\_Soccer\_CVPRW\_2020\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_CVPRW_2020/html/w19/Sanford_Group_Activity_Detection_From_Trajectory_and_Video_Data_in_Soccer_CVPRW_2020_paper.html&authuser=1)

**References 51 - 75**

* **Ref 51: Video assistant referees (var): The impact of technology on decision making in association football referees**
  + **Journal:** [https://www.tandfonline.com/doi/full/10.1080/02640414.2020.1802922](https://www.google.com/search?q=https://www.tandfonline.com/doi/full/10.1080/02640414.2020.1802922&authuser=1)
* **Ref 52: Contrastive learning for sports video: Unsupervised player classification**
  + **Paper:** [https://openaccess.thecvf.com/content/CVPR2021/html/Koshkina\_Contrastive\_Learning\_for\_Sports\_Video\_Unsupervised\_Player\_Classification\_CVPR\_2021\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/CVPR2021/html/Koshkina_Contrastive_Learning_for_Sports_Video_Unsupervised_Player_Classification_CVPR_2021_paper.html&authuser=1)
* **Ref 53: Recognizing tactic patterns in broadcast basketball video using player trajectory**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S104732031200021X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S104732031200021X&authuser=1)
* **Ref 54: Team activity recognition in sports**
  + **Conference:** <https://link.springer.com/chapter/10.1007/978-3-642-33712-3_16>
* **Ref 55: Exploring game performance in the national basketball association using player tracking data**
  + **Journal:** <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132894>
* **Ref 56: Leveraging contextual cues for generating basketball highlights**
  + **Paper:** <https://dl.acm.org/doi/10.1145/2964284.2967280>
* **Ref 57: Dynamic image networks for action recognition**
  + **arXiv:** <https://arxiv.org/abs/1604.06069>
* **Ref 58: Detecting events and key actors in multi-person videos**
  + **arXiv:** <https://arxiv.org/abs/1511.02917>
* **Ref 59: Automatic summarization of basketball sport video**
  + **Paper:** <https://ieeexplore.ieee.org/document/7877484/>
* **Ref 60: Towards real-time detection and tracking of basketball players using deep neural networks**
  + **Workshop Paper:** [https://nips.cc/media/PosterData/Poster\_120.pdf](https://www.google.com/search?q=https://nips.cc/media/PosterData/Poster_120.pdf&authuser=1)
* **Ref 61: Ontology-based global and collective motion patterns for event classification in basketball videos**
  + **Journal:** <https://ieeexplore.ieee.org/document/8693952/>
* **Ref 62: Single-camera basketball tracker through pose and semantic feature fusion**
  + **arXiv:** <https://arxiv.org/abs/1906.02042>
* **Ref 63: Analysis of technical features in basketball video based on deep learning algorithm**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S092359652030109X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S092359652030109X&authuser=1)
* **Ref 64: Fine-grained action recognition on a novel basketball dataset**
  + **Paper:** <https://ieeexplore.ieee.org/document/9053228/>
* **Ref 65: Fusing motion patterns and key visual information for semantic event recognition in basketball videos**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S092523122030635X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S092523122030635X&authuser=1)
* **Ref 66: Recognition of basketball referee signals from real-time videos**
  + **Journal:** [https://link.springer.com/article/10.1007/s12652-020-02228-5](https://www.google.com/search?q=https://link.springer.com/article/10.1007/s12652-020-02228-5&authuser=1)
* **Ref 67: Npu rgbd dataset and a feature-enhanced lstm-dgcn method for action recognition of basketball players**
  + **Journal:** <https://www.mdpi.com/2076-3417/11/16/7452>
* **Ref 68: Recognition of basketball player’s shooting action based on the convolutional neural network**
  + **Journal:** [suspicious link removed]
* **Ref 69: Research on basketball players’ action recognition based on interactive system and machine learning**
  + **Journal:** [suspicious link removed]
* **Ref 70: A lightweight fine-grained action recognition network for basketball foul detection**
  + **Paper:** <https://ieeexplore.ieee.org/document/9461159/>
* **Ref 71: Basketball action recognition based on fpga and particle image**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S014193312100067X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S014193312100067X&authuser=1)
* **Ref 72: Camera-based basketball scoring detection using convolutional neural network**
  + **Journal:** [https://link.springer.com/article/10.1007/s11633-021-1282-7](https://www.google.com/search?q=https://link.springer.com/article/10.1007/s11633-021-1282-7&authuser=1)
* **Ref 73: Long-and short-term plastic modeling of action prediction abilities in volleyball**
  + **Journal:** [https://link.springer.com/article/10.1007/s00426-011-0373-7](https://www.google.com/search?q=https://link.springer.com/article/10.1007/s00426-011-0373-7&authuser=1)
* **Ref 74: Indoor activity detection and recognition for sport games analysis**
  + **arXiv:** <https://arxiv.org/abs/1404.6413>
* **Ref 75: Saeta: A smart coaching assistant for professional volleyball training**
  + **Journal:** [https://ieeexplore.ieee.org/document/7042079/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/7042079/&authuser=1)

**References 76 - 100**

* **Ref 76: Sum product networks for activity recognition**
  + **Journal:** <https://ieeexplore.ieee.org/document/7301290/>
* **Ref 77: A hierarchical deep temporal model for group activity recognition**
  + **Paper:** <https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Ibrahim_A_Hierarchical_Deep_CVPR_2016_paper.pdf>
* **Ref 78: Activity recognition in beach volleyball using a deep convolutional neural network**
  + **Journal:** [https://link.springer.com/article/10.1007/s10618-017-0502-8](https://www.google.com/search?q=https://link.springer.com/article/10.1007/s10618-017-0502-8&authuser=1)
* **Ref 79: Evaluation of dominant and non-dominant hand movements for volleyball action modelling**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3340555.3355711>
* **Ref 80: Volleyball action modelling for behavior analysis and interactive multi-modal feedback**
  + **Conference:** [https://ris.utwente.nl/ws/portalfiles/portal/118331908/main.pdf](https://www.google.com/search?q=https://ris.utwente.nl/ws/portalfiles/portal/118331908/main.pdf&authuser=1)
* **Ref 81: Prediction of volleyball trajectory using skeletal motions of setter player**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3311823.3311853>
* **Ref 82: Pose is all you need: The pose only group activity recognition system (pogars)**
  + **arXiv:** <https://arxiv.org/abs/2108.04186>
* **Ref 83: Optimization of volleyball motion estimation algorithm based on machine vision and wearable devices**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S0141933121001406](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S0141933121001406&authuser=1)
* **Ref 84: Recognizing human action at a distance in video by key poses**
  + **Journal:** [https://ieeexplore.ieee.org/document/5666270/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/5666270/&authuser=1)
* **Ref 85: Violence detection in video using computer vision techniques**
  + **Paper:** <https://link.springer.com/chapter/10.1007/978-3-642-23672-3_3>
* **Ref 86: Human detection using oriented histograms of flow and appearance**
  + **Paper:** [https://lear.inrialpes.fr/people/triggs/pubs/Dalal-ECCV06.pdf](https://www.google.com/search?q=https://lear.inrialpes.fr/people/triggs/pubs/Dalal-ECCV06.pdf&authuser=1)
* **Ref 87: A spatio-temporal descriptor based on 3d-gradients**
  + **Paper:** [https://hal.inria.fr/inria-00309831/document](https://www.google.com/search?q=https://hal.inria.fr/inria-00309831/document&authuser=1)
* **Ref 88: Action bank: A high-level representation of activity in video**
  + **Paper:** [https://www.cv-foundation.org/openaccess/content\_cvpr\_2012/papers/Sadanand\_Action\_Bank\_A\_2012\_CVPR\_paper.pdf](https://www.google.com/search?q=https://www.cv-foundation.org/openaccess/content_cvpr_2012/papers/Sadanand_Action_Bank_A_2012_CVPR_paper.pdf&authuser=1)
* **Ref 89: Histograms of optical flow for efficient representation of body motion**
  + **Journal:** [https://www.sciencedirect.com/science/article/abs/pii/S016786551000216X](https://www.google.com/search?q=https://www.sciencedirect.com/science/article/abs/pii/S016786551000216X&authuser=1)
* **Ref 90: Action recognition with improved trajecto-ries**
  + **Paper:** [https://hal.inria.fr/hal-00873267/document](https://www.google.com/search?q=https://hal.inria.fr/hal-00873267/document&authuser=1)
* **Ref 91: Distinctive image features from scale-invariant key-points**
  + **Paper:** <https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>
* **Ref 92: Mosift: Recognizing human actions in surveillance videos**
  + **Paper:** [http://www.cs.cmu.edu/~mingyi/papers/MoSIFT-AVSS2009.pdf](https://www.google.com/search?q=http://www.cs.cmu.edu/~mingyi/papers/MoSIFT-AVSS2009.pdf&authuser=1)
* **Ref 93: On space-time interest points**
  + **Paper:** [https://hal.inria.fr/inria-00548611/document](https://www.google.com/search?q=https://hal.inria.fr/inria-00548611/document&authuser=1)
* **Ref 94: Learning a hierarchy of discriminative space-time neighborhood features for human action recognition**
  + **Paper:** [https://www.cv-foundation.org/openaccess/content\_cvpr\_2010/papers/Kovashka\_Learning\_a\_Hierarchy\_2010\_CVPR\_paper.pdf](https://www.google.com/search?q=https://www.cv-foundation.org/openaccess/content_cvpr_2010/papers/Kovashka_Learning_a_Hierarchy_2010_CVPR_paper.pdf&authuser=1)
* **Ref 95: Evaluation of local spatio-temporal features for action recognition**
  + **Paper:** [https://hal.inria.fr/inria-00404831/document](https://www.google.com/search?q=https://hal.inria.fr/inria-00404831/document&authuser=1)
* **Ref 96: Recognizing human actions: a local svm approach**
  + **Paper:** [https://www.csc.kth.se/~schuldt/papers/icpr04.pdf](https://www.google.com/search?q=https://www.csc.kth.se/~schuldt/papers/icpr04.pdf&authuser=1)
* **Ref 97: Tensor canonical correlation analysis for action classification**
  + **Paper:** <https://ieeexplore.ieee.org/document/4270104/>
* **Ref 98: Modeling temporal structure of decomposable motion segments for activity classification**
  + **Paper:** [https://vision.stanford.edu/publication\_readme/niebles\_eccv2010/](https://www.google.com/search?q=https://vision.stanford.edu/publication_readme/niebles_eccv2010/&authuser=1)
* **Ref 99: Beyond short snippets: Deep networks for video classification**
  + **arXiv:** <https://arxiv.org/abs/1503.08909>
* **Ref 100: Long-term recurrent convolutional networks for visual recognition and description**
  + **arXiv:** <https://arxiv.org/abs/1411.4389>

**References 161 - 185**

* **Ref 161: Video Graph Transformer for Spatiotemporal Action Recognition**
  + **Paper:** <https://www.ijcai.org/proceedings/2021/81>
* **Ref 162: Cross-Modal Learning for Sports Activity Classification**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413812>
* **Ref 163: Temporal Convolutional Networks for Dense Prediction (Action Segmentation and Detection)**
  + **arXiv:** <https://arxiv.org/abs/1611.05267>
* **Ref 164: Transformer-Based Multi-Scale Temporal Action Detection**
  + **Paper:** [https://ieeexplore.ieee.org/document/9663242/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/9663242/&authuser=1)
* **Ref 165: Semi-Supervised Contrastive Learning for Action Recognition**
  + **arXiv:** <https://arxiv.org/abs/2105.03383>
* **Ref 166: Ball Tracking and Event Detection in Football**
  + **Paper:** [https://openaccess.thecvf.com/content\_CVPRW\_2020/html/w19/Cioppa\_Camera-Based\_Ball\_Tracking\_and\_Event\_Detection\_in\_Football\_CVPRW\_2020\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_CVPRW_2020/html/w19/Cioppa_Camera-Based_Ball_Tracking_and_Event_Detection_in_Football_CVPRW_2020_paper.html&authuser=1)
* **Ref 167: Human Action Recognition with ResT: Hybrid CNN-Transformer**
  + **arXiv:** <https://arxiv.org/abs/2106.00693> (This refers to ResT, the underlying model)
* **Ref 168: Anomaly Detection in Sports Video Using Spatiotemporal Reasoning**
  + **Paper:** <https://ieeexplore.ieee.org/document/9191024/>
* **Ref 169: Multi-Agent Trajectory Prediction via Graph Attention**
  + **Paper:** <https://ieeexplore.ieee.org/document/9609618/>
* **Ref 170: Self-Supervised Pretraining for Action Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413726>
* **Ref 171: Temporal Region Proposal Networks for Action Localization (TURN TAP)**
  + **arXiv:** <https://arxiv.org/abs/1707.05834>
* **Ref 172: Memory-Augmented Networks for Long-Term Action Anticipation**
  + **Paper:** [https://ieeexplore.ieee.org/document/8977800/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/8977800/&authuser=1)
* **Ref 173: Structured Representation Learning for Action Understanding**
  + **Paper:** <https://ieeexplore.ieee.org/document/8578270/>
* **Ref 174: Scene-Context Enhanced Transformers for Video Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475355>
* **Ref 175: Hierarchical Temporal Relation Modeling**
  + **Paper:** [https://openaccess.thecvf.com/content/WACV2020/html/Chen\_Hierarchical\_Temporal\_Relation\_Modeling\_WACV\_2020\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/WACV2020/html/Chen_Hierarchical_Temporal_Relation_Modeling_WACV_2020_paper.html&authuser=1)
* **Ref 176: Egocentric Sports Action Understanding via Motion Encoding**
  + **arXiv:** <https://arxiv.org/abs/2108.08271>
* **Ref 177: Adversarial Learning for Sports Performance Assessment**
  + **Paper:** [https://ieeexplore.ieee.org/document/8977934/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/8977934/&authuser=1)
* **Ref 178: Temporal Feature Aggregation for Action Detection**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475681>
* **Ref 179: Pose Transfer for Action Recognition in Sports**
  + **Paper:** <https://ieeexplore.ieee.org/document/8939227/>
* **Ref 180: Scene-Aware Temporal Convolutional Networks**
  + **Paper:** <https://ieeexplore.ieee.org/document/9008544/>
* **Ref 181: Knowledge Distillation for Video Understanding**
  + **Journal:** <https://www.mdpi.com/2076-3417/10/7/2329>
* **Ref 182: Temporal Attention for Long-Term Sports Video**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3460426.3463630>
* **Ref 183: Domain Generalization in Action Recognition**
  + **Paper:** <https://www.ecva.net/papers/eccv_2020/papers_ECCV/html/1453_ECCV_2020_paper.php>
* **Ref 184: Uncertainty Modeling for Action Localization**
  + **Paper:** <https://ieeexplore.ieee.org/document/9422877/>
* **Ref 185: Pose-Based Self-Supervised Learning for Sports**
  + **arXiv:** <https://arxiv.org/abs/2103.04523>

**References 186 - 210**

* **Ref 186: Temporal Graph Convolutional Networks for Action Segmentation**
  + **Paper:** <https://ieeexplore.ieee.org/document/9151048/>
* **Ref 187: Hierarchical Pose-Based Networks for Fine-Grained Action Recognition**
  + **Paper:** [https://ieeexplore.ieee.org/document/9666014/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/9666014/&authuser=1)
* **Ref 188: Multi-Stage Temporal Convolutions for Action Segmentation (MS-TCN)**
  + **arXiv:** <https://arxiv.org/abs/1903.01945>
* **Ref 189: 3D ResNet for Action Recognition in Sports (A Closer Look at Spatiotemporal Convolutions)**
  + **arXiv:** <https://arxiv.org/abs/1711.11248>
* **Ref 190: Unified Temporal Action Localization via Cycle Consistency**
  + **Paper:** [https://openaccess.thecvf.com/content/CVPR2021/html/Shou\_Unified\_Temporal\_Action\_Localization\_via\_Cycle\_Consistency\_CVPR\_2021\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/CVPR2021/html/Shou_Unified_Temporal_Action_Localization_via_Cycle_Consistency_CVPR_2021_paper.html&authuser=1)
* **Ref 191: Ball Detection in Broadcast Soccer Videos**
  + **Paper:** [https://openaccess.thecvf.com/content\_CVPRW\_2019/html/CVSports/Cioppa\_Ball\_Detection\_in\_Broadcast\_Soccer\_Videos\_CVPRW\_2019\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content_CVPRW_2019/html/CVSports/Cioppa_Ball_Detection_in_Broadcast_Soccer_Videos_CVPRW_2019_paper.html&authuser=1)
* **Ref 192: Human-Centric Relation Modeling for Sports Videos**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413813>
* **Ref 193: Multi-View Fusion for Player Detection in Sports**
  + **Paper:** [https://openaccess.thecvf.com/content/ICCV2021W/CVSports/html/Zhou\_Multi-View\_Fusion\_for\_Player\_Detection\_in\_Sports\_ICCVW\_2021\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/ICCV2021W/CVSports/html/Zhou_Multi-View_Fusion_for_Player_Detection_in_Sports_ICCVW_2021_paper.html&authuser=1)
* **Ref 194: Graph-Based Context Modeling in Volleyball**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3343031.3351048>
* **Ref 195: Score-Based Highlight Generation in Sports**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3240508.3240581>
* **Ref 196: Explainable Action Recognition with Visual Rationale**
  + **Paper:** [https://openaccess.thecvf.com/content/WACV2021/html/Li\_Explainable\_Action\_Recognition\_With\_Visual\_Rationale\_WACV\_2021\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/WACV2021/html/Li_Explainable_Action_Recognition_With_Visual_Rationale_WACV_2021_paper.html&authuser=1)
* **Ref 197: Online Learning for Real-Time Sports Analytics**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413867>
* **Ref 198: Transformer Tracking for Ball and Player Localization**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475323>
* **Ref 199: Context-Aware Pooling for Video Understanding**
  + **Paper:** <https://ieeexplore.ieee.org/document/8953931/>
* **Ref 200: Few-Shot Action Recognition via Meta Learning**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475620>
* **Ref 201: Graph Convolutional Action Anticipation**
  + **Paper:** [https://ieeexplore.ieee.org/document/9368565/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/9368565/&authuser=1)
* **Ref 202: Temporal Feature Fusion for Event Spotting in Soccer**
  + **Paper:** [https://openaccess.thecvf.com/content/CVPR2022W/CVSports/html/Cioppa\_Temporal\_Feature\_Fusion\_for\_Event\_Spotting\_in\_Soccer\_CVPRW\_2022\_paper.html](https://www.google.com/search?q=https://openaccess.thecvf.com/content/CVPR2022W/CVSports/html/Cioppa_Temporal_Feature_Fusion_for_Event_Spotting_in_Soccer_CVPRW_2022_paper.html&authuser=1)
* **Ref 203: Lightweight Transformers for Real-Time Video Recognition**
  + **arXiv:** <https://arxiv.org/abs/2202.04969>
* **Ref 204: Pose-Guided Transformer for Action Understanding**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548398>
* **Ref 205: Temporal Fusion Network for Action Segmentation**
  + **Paper:** <https://www.mdpi.com/2072-4292/13/4/719>
* **Ref 206: Domain-Adaptive Transformers for Sports Analytics**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475678>
* **Ref 207: Dual-Stream Action Recognition Using Depth and RGB**
  + **Paper:** [https://ieeexplore.ieee.org/document/9046049/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/9046049/&authuser=1)
* **Ref 208: Contrastive Pretraining for Video Action Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475677>
* **Ref 209: Multi-Agent Learning for Team Sport Understanding**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413728>
* **Ref 210: Pose-Aware Spatiotemporal Attention Networks**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475354>

**References 211 - 250**

* **Ref 211: Hierarchical Multi-Scale Attention Networks**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475679>
* **Ref 212: Adversarial Domain Adaptation for Action Recognition (Adversarial Discriminative Domain Adaptation)**
  + **arXiv:** <https://arxiv.org/abs/1702.05464>
* **Ref 213: Reinforcement Learning for Tactical Pattern Mining**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3292500.3330663>
* **Ref 214: Spatiotemporal Transformers for Skeleton-Based Action Recognition**
  + **arXiv:** <https://arxiv.org/abs/2102.00761>
* **Ref 215: Motion Segmentation via Self-Supervised Clustering**
  + **Paper:** [https://dl.acm.org/doi/10.1145/3394171.3413996](https://www.google.com/search?q=https://dl.acm.org/doi/10.1145/3394171.3413996&authuser=1)
* **Ref 216: Graph-Based Pose Relational Networks**
  + **Paper:** [https://dl.acm.org/doi/10.1145/3394171.3413768](https://www.google.com/search?q=https://dl.acm.org/doi/10.1145/3394171.3413768&authuser=1)
* **Ref 217: Temporal Pyramid Attention Networks**
  + **Paper:** <https://ieeexplore.ieee.org/document/9098762/>
* **Ref 218: Skill Assessment in Sports Using Temporal Proposals**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3343031.3350948>
* **Ref 219: Contrastive Action Modeling with Scene Context**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475621>
* **Ref 220: Pose Dynamics Learning with Motion Refinement**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548386>
* **Ref 221: Temporal Coherence Networks for Sports Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548408>
* **Ref 222: Visual Reasoning with Object Graphs in Sports Videos**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413702>
* **Ref 223: Adaptive Sampling for Sports Action Recognition**
  + **Paper:** <https://ieeexplore.ieee.org/document/9665620/>
* **Ref 224: Spatial-Temporal Keypoint Learning for Sports Events**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475680>
* **Ref 225: Weakly Supervised Temporal Detection with Multiple Instance Learning**
  + **Paper:** <https://ieeexplore.ieee.org/document/8954009/>
* **Ref 226: Transformer-Based Multi-Instance Action Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548409>
* **Ref 227: Human-Object Interaction for Tactical Sports Analysis**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413703>
* **Ref 228: Temporal Video Graph Embedding for Action Recognition**
  + **Paper:** <https://ieeexplore.ieee.org/document/9664408/>
* **Ref 229: Self-Supervised Object-Centric Learning in Sports**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548393>
* **Ref 230: Contextual Transformers for Video Summarization in Sports**
  + **arXiv:** <https://arxiv.org/abs/2203.04456>
* **Ref 231: Pose-Conditioned Video Transformers for Action Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548387>
* **Ref 232: Temporal Proposal Generation Using Reinforcement Learning**
  + **Paper:** [https://www.bmvc2020-conference.com/assets/papers/0593.pdf](https://www.google.com/search?q=https://www.bmvc2020-conference.com/assets/papers/0593.pdf&authuser=1)
* **Ref 233: Action Graph Networks for Event Understanding**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3343031.3351042>
* **Ref 234: Multi-Camera Pose Estimation in Team Sports**
  + **arXiv:** <https://arxiv.org/abs/2111.12349>
* **Ref 235: Fine-Grained Action Segmentation via Attention Bottlenecks**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3394171.3413767>
* **Ref 236: Temporal Cross-Attention for Video Recognition**
  + **Paper:** <https://ieeexplore.ieee.org/document/9665045/>
* **Ref 237: Unsupervised Action Segmentation by Motion Clustering**
  + **Paper:** [https://ieeexplore.ieee.org/document/8953457/](https://www.google.com/search?q=https://ieeexplore.ieee.org/document/8953457/&authuser=1)
* **Ref 238: Spatio-Temporal Relational Networks for Multi-Agent Sports Analysis**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475682>
* **Ref 239: Semantic Aggregation in Temporal Proposal Networks**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475683>
* **Ref 240: Video Transformer with Learnable Token Sampling**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548412>
* **Ref 241: Recurrent Pose Memory Networks for Action Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475622>
* **Ref 242: Self-Supervised Video Pretraining via Temporal Cycle Consistency**
  + **Paper:** [https://dl.acm.org/doi/10.1145/3394171.3413704](https://www.google.com/search?q=https://dl.acm.org/doi/10.1145/3394171.3413704&authuser=1)
* **Ref 243: Event-Based Representation for Sports Analysis**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475352>
* **Ref 244: Transformer-Based Skill Assessment via Phase Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475623>
* **Ref 245: Fine-Grained Sport Action Localization with Attention Shift**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548375>
* **Ref 246: Graph-Based Tactical Activity Recognition**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3474085.3475353>
* **Ref 247: Temporal Attention-Augmented ConvNets for Video Recognition**
  + **Paper:** <https://ieeexplore.ieee.org/document/9098739/>
* **Ref 248: Scene-Text-Aware Sports Event Classification**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548392>
* **Ref 249: Semi-Supervised Multi-Label Action Recognition**
  + **Paper:** [https://dl.acm.org/doi/10.1145/3394171.3413988](https://www.google.com/search?q=https://dl.acm.org/doi/10.1145/3394171.3413988&authuser=1)
* **Ref 250: Pose-Based Transformer Models for Event Forecasting**
  + **Paper:** <https://dl.acm.org/doi/10.1145/3503161.3548385>