

Event Detection in Ice Hockey Using Game-Aware Representations: A Dataset and Baseline Study

Ken Nsiempba
University of Waterloo
kmnsiemp@waterloo.ca

Amir Nazemi
University of Waterloo
amir.nazemi@uwaterloo.ca

John Zelek
University of Waterloo
jzelek@uwaterloo.ca

David Clausi
University of Waterloo
dclausi@uwaterloo.ca

Abstract

Ice hockey occupies a significant cultural and economic position in Canada, yet analytical tools for understanding the game remain relatively underdeveloped. In particular, event detection is still performed manually, hindered by several challenges: the absence of publicly available, labeled ice hockey event datasets; the limited transferability of event-detection methods developed for other sports due to differing event semantics; and the lack of approaches that incorporate contextual priors, here referred to as game attributes, such as player locations or team structure.

Automating event detection through the integration of such game attributes has the potential to greatly enhance the analysis of broadcast footage by enabling efficient detection and review of key moments, improving player development pipelines, and supporting real-time highlight generation, especially in leagues without access to sophisticated tracking systems.

This work addresses these gaps by introducing a small-scale, fully annotated ice hockey event dataset, establishing baseline models for benchmarking, and comparing three distinct data representations: RGB frames, occupancy grids, and graph-based structures, some of which explicitly incorporate game attributes to improve detection performance.

1. Introduction

Event detection in video streams is a critical area within computer vision and machine learning, holding immense significance across diverse industries and research fields. An event here is defined as a temporally bounded occurrence. Event detection capabilities are pivotal in sectors such as surveillance [31, 39], healthcare [2], traffic moni-

toring [17], entertainment, and sports analytics [11, 23, 37].

Sports analytics represents a high-impact application of event detection technologies. The rapid growth of sports analytics is driven by substantial financial investment, significant public interest, and the increasing adoption of data-driven decision-making by professional sports teams [8]. Accurate event detection enables the extraction of detailed statistics, performance analysis, strategic insights, and automatic highlight generation, greatly enriching both player development and fan engagement.

Ice hockey, as Canada’s national sport, commands a significant cultural and economic presence, making it an ideal domain for applying advanced event detection systems. It is estimated that approximately \$2.6 billion circulates directly between communities in Canada each year as a result of hockey-related activities [29]. Despite substantial investments and public enthusiasm, hockey analytics remains underdeveloped compared to other major team sports [27]. Event detection in hockey is still done manually and could benefit greatly from being automated. Examples of ice hockey events include passes, shots or faceoffs [32].

1.1. Research gaps

The **main challenges** hindering the development of robust event detection pipelines for ice hockey are:

1. **Lack of datasets:** There are currently no publicly available, large-scale ice hockey event datasets that can be used to establish benchmarking baselines.
2. **Limited prior work:** Only a small number of models have been developed specifically for ice hockey event detection [32, 33].
3. **Poor cross-sport transferability:** Event detection methods designed for other sports do not transfer reliably to ice hockey due to differences in gameplay structure and types of event [37].
4. **Underuse of contextual priors:** Existing approaches do

not incorporate game attributes (e.g., player positions, team structure), despite their potential to significantly improve event discrimination.

1.2. Key contributions

This research provides the following contributions:

1. **A small-scale, fully supervised ice hockey event dataset**, including precise start and end frame annotations for multiple event classes.
2. **A benchmarking study of multiple data representations**, evaluating how RGB frames, occupancy grids, and graph-based structures perform on the proposed dataset.

2. Related works

The concept of an “event” in sports lacks a unified definition, with terms such as events, actions, and activities often used interchangeably across studies [6, 20]. Interpretations vary depending on the sport, dataset, and modeling perspective, which creates inconsistencies in how behaviors are categorized and evaluated.

Event detection pipelines generally differ along several dimensions, including input modality, data representation, feature extraction, and temporal modeling. Inputs may range from video [16] and audio [24] to symbolic or time-series data [28], while representations can include raw frames, player trajectories, occupancy grids, or graph-based multi-agent structures. Feature extraction methods span handcrafted descriptors [30] and deep learning models such as CNNs [19], ViTs [7], MambaVision [10], and GNNs [18]. Temporal reasoning is achieved through rule-based systems [22, 25, 34], probabilistic models [1, 14, 21, 26], reinforcement learning [4], or sequence models such as RNNs and LSTMs [3, 36].

In video analysis, event detection is commonly framed as temporal action localization (TAL) [35, 37], where pipelines are organized into one-stage, two-stage, or anchor-free architectures. Although modern TAL frameworks and foundation models provide strong general-purpose solutions, they often remain domain-agnostic and rely primarily on visual cues, which can limit performance in structured environments like sports.

Sports-specific event detection methods adapt these pipelines by incorporating domain knowledge such as player roles, spatial configurations, and object interactions. Advances in datasets for soccer [5], basketball [9], and volleyball [12] have enabled approaches combining tracking data, structured reasoning, temporal models, and graph-based representations. These studies demonstrate that leveraging contextual game information is critical for accurate event understanding.

In contrast, event detection research for ice hockey remains limited. Existing work largely relies on raw broad-

cast video and lacks the integration of structured game attributes. The scarcity of annotated hockey datasets further restricts methodological development and limits the direct transfer of approaches from other sports domains.

3. Methods

3.1. Data acquisition and preprocessing

Three event classes were defined: **Faceoff**, marking the restart of play around one of the nine circles (see Figure 2); **One-on-One**, including breakaways or situations where a forward faces only the last defender; and **Powerplay Offense**, representing sequences where the advantaged team executes offensive formations during a numerical imbalance.

These events were selected for their analytical relevance. From a **coaching and scouting** perspective, they provide insight into positioning, puck possession, and formation strategies [15]. From a **highlight-generation** standpoint, they correspond to moments of high engagement or clear structural transitions in play. Finally, from a **contextual representation** view, these events are closely tied to spatial and team dynamics, making them ideal for leveraging player locations and affiliations in event differentiation.

Raw hockey videos were provided by **Stathletes** and reformatted to match standard event-detection datasets [13]. Let the dataset be $\mathcal{D} = \{V_i\}_{i=1}^N$, where each untrimmed video

$$V_i = \{f_t \mid t = 1, \dots, T_i\}$$

is a sequence of frames f_t . For each event class $c \in \{\text{faceoff}, \text{1v1}, \text{pp_offense}\}$, temporal boundaries (s_i^c, e_i^c) (start/end frame indices) were annotated.

This process produced a collection of uniformly structured samples, each containing exactly one occurrence of *faceoff*, *1v1*, or *powerplay offense*. The resulting distribution of event instances across the training, validation, and test splits is illustrated in Figure 1, reflecting the natural class imbalance present in the original footage.

The dataset comprises a total of 140 video sequences, split into 121 training, 25 validation, and 29 test videos. Across all splits, the dataset includes 116 faceoff events and 24 one-on-one (1v1) events. Faceoffs represent the majority of the dataset (approximately 83%), reflecting their higher frequency in broadcast footage, while 1v1 events account for the remaining 17%. Figure 1 illustrates the distribution of event classes across the training, validation, and test splits.

To emulate untrimmed, in-the-wild inference, test videos were **elongated**. Given an elongation factor $\alpha \in [0, 1]$, new boundaries are

$$s_i'^c = \max(1, \lfloor s_i^c - \alpha(e_i^c - s_i^c) \rfloor)$$

$$e_i'^c = \min(T_i, \lceil e_i^c + \alpha(e_i^c - s_i^c) \rceil),$$

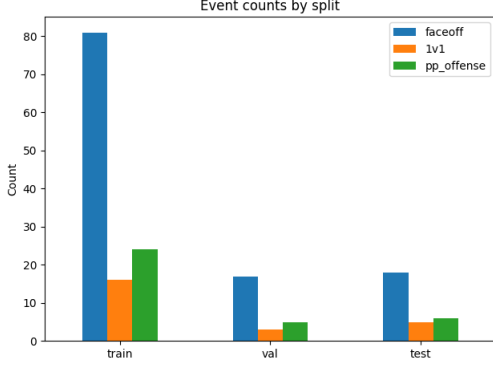


Figure 1. Distribution of Event Instances Across Dataset Splits



Figure 2. Frame showing a faceoff with the outlined puck. A faceoff is defined from the time the puck (in red) is dropped by the referee.

and frames outside $[s_i^c, e_i^c]$ but inside $[s_i'^c, e_i'^c]$ are labeled *Other*.

3.2. Representations and graph notation

Each frame f_t is mapped to a representation $\phi(f_t)$:

- **RGB**: raw image.
- **Occupancy grid**: $O_t \in \mathbb{R}^{H \times W \times 3}$, where one channel encodes player locations, one encodes a visibility mask, and the third is a zero-padded channel.
- **Graph**: $G_t = (V_t, E_t, X_t)$, where nodes $V_t = \{v_{t,n}\}_{n=1}^{N_t}$ are players in view, edges E_t connect visible player pairs (fully connected or proximity-based), and node features $X_t \in \mathbb{R}^{N_t \times d}$ encode homographic coordinates (e.g., $[p_x, p_y]$).

3.3. Frame-level classification pipeline

For all representations, a feature extractor F_θ first maps each frame representation $\phi(f_t)$ to a feature embedding

$$z_t = F_\theta(\phi(f_t)).$$

The embedding is then passed to a classifier h that produces class probabilities,

$$P(c | f_t) = h(z_t),$$

and the predicted label is obtained by

$$\hat{y}_t = \arg \max_{c \in \mathcal{C}} P(c | f_t),$$

$$\mathcal{C} = \{\text{faceoff}, \text{1v1}, \text{pp_offense}, \text{Other}\}.$$

We employ an image classifier (Mambavision [10]) for RGB and occupancy grids, and a GNN for graphs G_t . Per-frame predictions over the elongated interval $[s_i'^c, e_i'^c]$ are later grouped by temporal continuity to form event-level hypotheses.

3.4. Image dataset construction

Frames were randomly sampled to balance classes; images were mirrored vertically for augmentation, the scoreboard was masked, and an *Other* class was added for non-event frames.

4. Results

4.1. Frame-based image classification

The resulting confusion matrices for the frame-level classifiers are shown in Figure 3. A global classification accuracy of $A = 77.33\%$ was achieved, computed as

$$A = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = y_i), \quad (1)$$

where $\hat{y}_i$ and y_i denote the predicted and ground-truth labels for frame i , and N is the total number of frames.

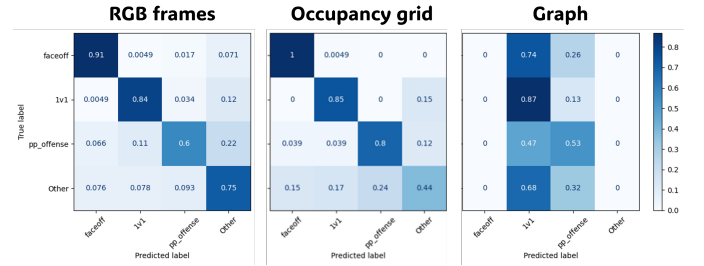


Figure 3. Confusion matrices obtained using different frame representations. (a) Occupancy-grid representation with MambaVision classifier; (b) Graph-based representation with GNN classifier.

Using the **occupancy grid representation**, the model achieved strong class separability. *Faceoffs* reached 100% recall, reflecting the clear spatial symmetry of players around the faceoff circles. *1v1* and *powerplay offense* achieved 85% and 80% recall, respectively, though both exhibited moderate confusion with the *Other* class. The lower 44% accuracy for *Other* suggests that neutral or transitional frames share spatial characteristics with structured events.

In contrast, the **graph-based GNN classifier** performed well for *1v1* (97% accuracy) but failed to identify *faceoffs*,

frequently misclassifying them as early-possession frames. This suggests that the model relies heavily on interaction-based cues while lacking explicit spatial context.

4.2. Event detection

For each test video, frame-level accuracy was computed as

$$A_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \mathbb{I}(\hat{y}_{i,t} = y_{i,t}), \quad (2)$$

where T_i is the number of frames in video i , $y_{i,t}$ and $\hat{y}_{i,t}$ denote the ground-truth and predicted labels for frame t , and $\mathbb{I}(\cdot)$ is the indicator function.

To evaluate performance at different thresholds, we report the proportion of videos whose accuracy exceeds a threshold τ :

$$S(\tau) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(A_i \geq \tau), \quad (3)$$

where N is the total number of test videos and $\tau \in \{0.3, 0.5, 0.7\}$.

Three representations (RGB, occupancy grid, and graph) were compared across $\tau = \{0.3, 0.5, 0.7\}$.

Threshold	0.3	0.5	0.7
RGB frames	0.483	0.276	0.138
Grid	0.621	0.483	0.000
Graph	0.207	0.172	0.000

Table 1. Event detection accuracy across thresholds τ .

Results show that accuracy decreases as the threshold τ increases for all representations. The occupancy grid consistently outperforms both RGB and graph-based approaches, suggesting that its structured encoding of spatial occupancy enhances temporal event localization. Graph representations yield the lowest scores, likely due to limited node attributes and the model’s sensitivity to incomplete player visibility, though they remain scalable to richer relational features.

For comparison, the TAL model **ActionFormer** [38] was used on trimmed test sequences. The model was trained on pre-extracted i3D features stored as `.numpy` files (feature dimension 512). Features are computed with a temporal stride of `feat_stride = 4` frames and a backbone clip length of `num_frames = 16` frames per feature chunk. Sequences are truncated/cropped during training using `trunc_thresh = 0.5` and a random crop ratio in `[0.9, 1.0]`, with a maximum sequence length of `max_seq_len = 2304` feature steps. The model is trained for `epochs = 30` with learning rate 1×10^{-4} , weight decay 0.05, and batch size 2. Training uses gradient clipping with `clip_grad_l2norm = 1.0`, classification prior probability `cls_prior_prob = 0.01`,

and center sampling (`center_sample=radius`) with radius 1.5. At inference, predictions are filtered with `voting_thresh = 0.7`, `pre_nms_topk = 2000`, `max_seg_num = 200`, and `min_score = 10^{-3}`, using multiclass NMS (`multiclass_nms=True`). In our label definition, both *one-on-one encounters* and *breakaways* are merged into the same *1v1* category. Results gave us an mAP of 0.0%. Indicating that the per-frame analysis might be a better event detection model for small scale datasets.

5. Conclusion

This study presented a pipeline for frame-based and sequence-level event detection in ice hockey videos, leveraging different data representations; RGB frames, occupancy grids, and player-interaction graphs. The results demonstrated that occupancy grids, by explicitly encoding spatial and visibility information, provide a better performance across thresholds, while graph-based representations, though currently limited by sparse node attributes, offer scalable potential for relational modeling.

Future work includes expanding the dataset to improve generalization, adopting advanced multi-scale TAL models to better capture temporal dynamics, and integrating a broader range of game attributes such as puck location, team IDs, and spatial context to enhance event discrimination. Models that leverage temporal dependencies across frames will also be developed, and hybrid architectures that fuse multiple data representations will be explored to improve contextual understanding and event continuity.

Acknowledgments

We would like to thank **Stathletes** for providing access to AHL broadcast videos and **IBET** fellowship for their support.

References

- [1] (PDF) Automatic Evaluation of Organized Basketball Activity using Bayesian Networks. 2
- [2] Menatalla Abououf, Shakti Singh, Rabeb Mizouni, and Hadi Otrouk. Explainable ai for event and anomaly detection and classification in healthcare monitoring systems. *IEEE Internet of Things Journal*, 11(2):3446–3457, 2023. 1
- [3] Timur Bagautdinov, Alexandre Alahi, Francois Fleuret, Pascal Fua, and Silvio Savarese. Social Scene Understanding: End-To-End Multi-Person Action Localization and Collective Activity Recognition. pages 4315–4324, 2017. 2
- [4] Pramod Chunduri, Jaeho Bang, Yao Lu, and Joy Arulraj. Zeus: Efficiently Localizing Actions in Videos using Reinforcement Learning. In *Proceedings of the 2022 International Conference on Management of Data*, pages 545–558, 2022. arXiv:2104.06142 [cs]. 2
- [5] Adrien Deliege, Anthony Cioppa, Silvio Giancola, Meisam J Seikavandi, Jacob V Dueholm, Kamal Nasrollahi, Bernard

- Ghanem, Thomas B Moeslund, and Marc Van Droogenbroeck. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4508–4519, 2021. 2
- [6] Junyi Dong, Qingze Huo, and Silvia Ferrari. A Holistic Approach for Role Inference and Action Anticipation in Human Teams. *ACM Transactions on Intelligent Systems and Technology*, 13(6):95:1–95:24, 2022. 2
- [7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021. arXiv:2010.11929 [cs]. 2
- [8] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective. *WIREs Data Mining and Knowledge Discovery*, 13(5):e1496, 2023. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1496>. 1
- [9] Xiaofan Gu, Xinwei Xue, and Feng Wang. Fine-grained action recognition on a novel basketball dataset. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2563–2567. IEEE, 2020. 2
- [10] Ali Hatamizadeh and Jan Kautz. MambaVision: A Hybrid Mamba-Transformer Vision Backbone, 2025. arXiv:2407.08083 [cs]. 2, 3
- [11] Jiabin He and Hsing-Kuo Pao. Multi-modal, Multi-labeled Sports Highlight Extraction. In *2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 181–186, 2020. ISSN: 2376-6824. 1
- [12] M Ibrahim, S Muralidharan, Z Deng, A Vahdat, and G Mori. Hierarchical deep temporal models for group activity recognition.(2016). 2
- [13] Haroon Idrees, Amir R. Zamir, Yu-Gang Jiang, Alex Gorbun, Ivan Laptev, Rahul Sukthankar, and Mubarak Shah. The THUMOS challenge on action recognition for videos “in the wild”. *Computer Vision and Image Understanding*, 155:1–23, 2017. 2
- [14] Stephen S Intille and Aaron F Bobick. A Framework for Recognizing Multi-Agent Action from Visual Evidence. 2
- [15] William Jiang, Ken Nsiempba, John Zelek, David Clausi, and Yuhao Chen. A success metric for individual player encounters in Ice Hockey videos. *Journal of Computational Vision and Imaging Systems*, 10(1):39–46, 2024. Number: 1. 2
- [16] H Jyothi, M Komala, and S Mallikarjunaswamy. A Comprehensive Survey on Technologies in Video-based Event Detection and Recognition Using Machine Learning and Deep Learning Techniques. In *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, pages 1–5, 2024. 2
- [17] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic monitoring and accident detection at intersections. *IEEE Transactions on Intelligent Transportation Systems*, 1(2):108–118, 2000. 1
- [18] Bharti Khemani, Shruti Patil, Ketan Kotecha, and Sudeep Tanwar. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, 11(1):18, 2024. 2
- [19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2
- [20] Fang Liu, Chunhua Tian, Jinzhong Wang, Youwei Jin, Luxiang Cui, and Ivan Lee. STAB-GCN: A Spatio-Temporal Attention-Based Graph Convolutional Network for Group Activity Recognition. *Mathematics*, 11(14):3074, 2023. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute. 2
- [21] Xiaohui Liu and Chin-Seng Chua. Multi-agent activity recognition using observation decomposed hidden Markov models. *Image and Vision Computing*, 24(2):166–175, 2006. 2
- [22] J. Lopes, L. Mota, S. Mota, José Torres, Rui Silva Moreira, Christophe Soares, Ivo Pereira, Feliz Gouveia, and Pedro Sobral. *Object and Event Detection Pipeline for Rink Hockey Games*. 2024. 2
- [23] Cise Midoglu, Saeed Shafiee Sabet, Mehdi Houshmand Sarkhoosh, Mohammad Majidi, Sushant Gautam, Håkon Maric Solberg, Tomas Kupka, and Pål Halvorsen. AI-Based Sports Highlight Generation for Social Media. In *Proceedings of the 3rd Mile-High Video Conference*, pages 7–13, New York, NY, USA, 2024. Association for Computing Machinery. 1
- [24] Sallauddin Mohammad and Suresh Kumar Sanampudi. Exploring current research trends in sound event detection: a systematic literature review. *Multimedia Tools and Applications*, 83(37):84699–84741, 2024. 2
- [25] Lia Morra, Francesco Manigrasso, Giuseppe Canto, Claudio Gianfrate, Enrico Guarino, and Fabrizio Lamberti. Slicing and Dicing Soccer: Automatic Detection of Complex Events from Spatio-Temporal Data. In *Image Analysis and Recognition*, pages 107–121, Cham, 2020. Springer International Publishing. 2
- [26] Shigeru Motoi, Toshie Misu, Yohei Nakada, Tomohiro Yazaki, Go Kobayashi, Takashi Matsumoto, and Nobuyuki Yagi. Bayesian event detection for sport games with hidden Markov model. *Pattern Analysis and Applications*, 15(1):59–72, 2012. 2
- [27] Banoth Thulasya Naik, Mohammad Farukh Hashmi, and Neeraj Dhanraj Bokde. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Applied Sciences*, 12(9):4429, 2022. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute. 1
- [28] Tingyu Qu, Ko Wai Mei, and Arnold Doray. A simple method to detect extreme events from financial time series data. *Machine Learning with Applications*, 10:100415, 2022. 2
- [29] Scotiabank. Ice hockey in canada: Summary and infographic. <https://www.scotiabank.com/>

[ca / common / pdf / Ice - Hockey - in - Canada - Summary-and-Infographic.pdf](#), 2020. Accessed: 2025-05-26. [1](#)

- [30] Grigorios Tsagkatakis, Mustafa Jaber, Panagiotis Tsakalides, Mustafa Jaber, and Panagiotis Tsakalides. Goal!! Event detection in sports video. *Electronic Imaging*, 29:15–20, 2017. Publisher: Society for Imaging Science and Technology. [2](#)
- [31] Fath U. Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq, and Sung Wook Baik. Violence Detection Using Spatiotemporal Features with 3D Convolutional Neural Network. *Sensors*, 19(11):2472, 2019. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [1](#)
- [32] Kanav Vats, Mehrnaz Fani, Pascale Walters, David A Clausi, and John Zelek. Event detection in coarsely annotated sports videos via parallel multi-receptive field 1d convolutions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 882–883, 2020. [1](#)
- [33] Kanav Vats, Mehrnaz Fani, David A Clausi, and John Zelek. Puck localization and multi-task event recognition in broadcast hockey videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4567–4575, 2021. [1](#)
- [34] Ferran Vidal-Codina, Nicolas Evans, Bahaeddine El Fakir, and Johsan Billingham. Automatic event detection in football using tracking data. *Sports Engineering*, 25(1):18, 2022. [2](#)
- [35] Binglu Wang, Yongqiang Zhao, Le Yang, Teng Long, and Xuelong Li. Temporal Action Localization in the Deep Learning Era: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2171–2190, 2024. [2](#)
- [36] Peng Wang, Fanwei Zeng, and Yuntao Qian. A survey on deep learning-based spatio-temporal action detection. *International Journal of Wavelets, Multiresolution and Information Processing*, 22(04):2350066, 2024. Publisher: World Scientific Publishing Co. [2](#)
- [37] Hao Xu, Arbind Agrahari Baniya, Sam Well, Mohamed Reda Bouadjeneq, Richard Dazeley, and Sunil Aryal. Action Spotting and Precise Event Detection in Sports: Datasets, Methods, and Challenges, 2025. arXiv:2505.03991 [cs]. [1](#), [2](#)
- [38] Chenlin Zhang, Jianxin Wu, and Yin Li. ActionFormer: Localizing Moments of Actions with Transformers, 2022. arXiv:2202.07925 [cs]. [4](#)
- [39] Wei Zhou, Xuanlin Min, Yiheng Zhao, Yiran Pang, and Jun Yi. A Multi-Scale Spatio-Temporal Network for Violence Behavior Detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(2):266–276, 2023. Conference Name: IEEE Transactions on Biometrics, Behavior, and Identity Science. [1](#)