Tracking the Untrackable: Failure Modes of Object Trackers on Ice Hockey Puck

Liam Salass, John Zelek, David Clausi

{liam.salass, jzelek, dclausi}@uwaterloo.ca

University of Waterloo, Waterloo, ON, Canada, N2L 3G1

Abstract

Tracking the hockey puck in broadcast video remains a difficult problem in sports analytics due to the puck's small size, high velocity, frequent occlusions, and heavy compression artifacts. In this paper, we benchmark seven state-of-the-art trackers—Siamese-FC, ARTrack, DiMP, SuperDiMP, PrDiMP, ToMP, and TaMOs—on a large broadcast hockey dataset. We analyze how existing tracking paradigms break down when applied to fast, small objects in cluttered environments. Our experiments reveal that approaches that rely on localized search regions are highly susceptible to irreversible drift, stick-locking, and blur-induced feature loss. In contrast, global frame-level reasoning (TaMOs) enables superior re-detection and stability across occlusions. However, even the best-performing trackers struggle with severe motion blur and extreme spatial aliasing. These findings expose fundamental limitations in current tracking assumptions and highlight the need for architectures that preserve fine-grained detail, are domain-specific, and reason globally.

1. Introduction

Ice hockey analytics has increasingly leveraged computer vision techniques [2, 3, 25, 27, 35, 38, 41–43]. Applications range from player detection and tracking [27, 43], pose estimation [3, 35], and jersey number recognition [2, 41] to higher-level analyses such as gameplay strategy understanding [25, 38, 42] and puck localization [18, 32, 39]. Recent advancements have improved single-frame puck detection; however, consistent puck tracking—establishing reliable temporal associations across frames—remains a challenge. Accurate temporal tracking is crucial for capturing consistent puck motion, maintaining puck identity through occlusions, and enabling deeper insights into team strategies and possession dynamics.

Robust puck tracking is central to any analytics pipeline that performs spatial-temporal reasoning [46]. Whether

Figure 1. Tracking failure example of multiple models. The **Red** bounding box is the ground truth.

assisting coaching systems [37], enhancing broadcasts, or supporting advanced performance analytics, tracking failures quickly cascade into incorrect strategic conclusions. While ball tracking has been well explored in sports like soccer [1, 4, 14, 15, 19, 21, 30, 44], volleyball [9, 10, 12, 13], and basketball [6–8], puck tracking remains underexplored due to challenges like occlusion, motion blur, small size, and broadcast compression [22, 26, 32, 34, 42, 48].

In this paper, we evaluate several off-the-shelf trackers—Siamese-FC, ARTrack, DiMP, PrDiMP, ToMP, SuperDiMP, and TaMOs—on puck tracking in broadcast hockey footage. Rather than introducing a novel model, we expose how standard tracking paradigms fundamentally fail under real-world hockey conditions. Our results underscore the need to rethink foundational tracking assumptions for sports like hockey.

2. Related Works

2.1. Visual Object Tracking

Tracking methods have undergone several architectural shifts over the past decade, each attempting to handle better appearance variation, background clutter, and long-term temporal uncertainty.

Early deep trackers were mostly Siamese networks, introduced by Fully-Convolutional Siamese Tracking (SiamFC) [11]. These models learn a generic similarity function between a template (the initial tracker object's appearance) and a search region (a small region in the next frame that may contain the tracked object), enabling efficient dense correlation and real-time tracking. However, because Siamese trackers rely solely on a fixed target template extracted in the first frame, they struggle with significant appearance changes, background distractors, and occlusions.

To overcome these limitations, a second wave of methods introduced discriminative online model prediction. DiMP [5] and SuperDiMP [16] embed an online optimization module that learns a target-specific classifier during tracking. This incorporates background examples directly into the learned model, dramatically improving robustness to distractors. Building on this, PrDiMP [17] leverages a probabilistic regression for tracking, predicting a full conditional density over the target state rather than a single confidence score. This alleviates issues with unknown states and addresses uncertainty during occlusion and blurring.

A parallel line of work introduced Transformer-based fusion and model prediction, enabling deeper global reasoning across template and search features. ToMP [23] replaces the optimization-based model predictor with a Transformer that directly predicts the weights of both the classifier and the bounding-box regressor. By removing the inductive bias of traditional optimization and learning, and by end-to-end model prediction, ToMP achieves stronger adaptivity to rapid appearance changes.

Most recently, TaMOs [24] extends this paradigm to multi-object tracking in generic scenarios. By sharing global frame-level features and predicting object-specific correlation filters in parallel, TaMOs eliminates the computational bottleneck of running multiple single object trackers independently. Its Transformer-based multi-object encoding enables robust handling of distractors and efficient re-detection across the entire frame.

The trajectory of modern tracking research shows a clear progression: Fixed-template similarity $\rightarrow$ online discriminative learning $\rightarrow$ probabilistic estimation $\rightarrow$ Transformer-based reasoning $\rightarrow$ global multi-object tracking. This evolution motivates our analysis of how these paradigms behave when applied to tracking tiny, fast-moving targets such as the hockey puck.

2.2. Physics-Guided and Motion-Aware Tracking.

Recent work on fast, small-object tracking has moved beyond appearance-based models by incorporating additional structural cues such as physics priors and motion-aware attention. Physics-Guided Fusion [36] stabilizes trajectories during erratic motion or occlusion by integrating kinematic constraints, collision dynamics, and depth cues, avoiding the simplistic assumption of smooth linear motion. Similarly, TrackNetV4 [28] introduces motion-attention maps and learnable motion prompts that highlight true object motion while suppressing background dynamics, improving detection under challenging broadcast conditions.

These approaches suggest that structured priors—whether physical or motion-based—can compensate for the ambiguity inherent in tracking tiny, fast-moving objects. However, their reliance on domain-specific assumptions (e.g., known physical parameters, consistent scene structure, reliable motion differencing) limits direct transfer to ice hockey broadcast video, where aliasing, irregular occlusions, and extreme camera motion are frequent. As a result, physics- or motion-informed trackers would require substantial adaptation to be effective for broadcast puck tracking.

2.3. Ice Hockey Puck Detection and Tracking

Puck detection and tracking has been addressed in several recent works in hockey videos [22, 26, 34, 42, 48]. These approaches typically combine deep learning-based localization, temporal reasoning, and contextual cues to maintain consistent puck trajectories over time.

Yang et al. [48] emphasize the limitations of earlier object detectors such as YOLOv3 [29] and Mask R-CNN [20] for puck tracking, citing high false-positive rates and difficulty maintaining temporal consistency due to their single-frame nature. To address this, they introduce a multi-headed deep learning architecture that explicitly leverages temporal features across nine consecutive frames to infer the puck's position for a single frame. This sequence-based inference provides short-term temporal associations, distinguishing their approach from static, frame-wise detection methods such as PLUCC[33].

Pidaparthy et al. [26] incorporate both spatial and motion cues by regressing the puck location using a deep neural network informed by estimated player positions and optical flow. Their method uses the predicted puck trajectory to dynamically reposition a virtual camera toward the region of play.

Li et al. [22] propose a two-state tracking framework that classifies puck motion into *controlled* and *free-moving* states. Their system integrates direct image-based matching when the puck is visible and motion estimation during occlusions, enabling short-term continuity in tracking. However, they acknowledge that visual ambiguity—such as reflections or players' skates—can still lead to identity switches or false trajectories.

Vats et al. [42] introduce *PuckNet*, a 3D convolutional neural network that predicts a spatio-temporal heatmap of the puck's approximate location. By leveraging video clips rather than individual frames, their approach captures temporal context and mitigates short-term occlusions. Nonetheless, the method remains limited by the puck's small size and rapid movement, often relying on contextual assumptions such as proximity to player clusters, which can reduce long-term tracking reliability.

Finally, Sarkhoosh et al. [34] explore puck localization with modern detection architectures, including Faster R-CNN [31] and YOLOv8 [40], noting that YOLOv8-X achieved the highest localization accuracy. While their approach enhances frame-level precision, it does not explicitly address temporal consistency—highlighting the continued need for robust temporal association in puck tracking.

3. Methodology

3.1. Dataset

We finetune and evaluate all trackers on a broadcast hockey dataset containing over 145,000 annotated frames across train, validation, and test splits. The data captures a wide range of broadcast artifacts relevant to real-world deployment, including motion blur, partial and full occlusions, compression noise, and rapid camera pans. Sequences vary substantially in length, with typical clips containing 16–40 visible-puck frames before an occlusion. Across all splits, the puck is visible in only ~68% of frames, reflecting the inherent difficulty of puck tracking in broadcast footage where the object frequently disappears behind sticks, skates, boards, and players.

3.2. Trackers

We benchmark the following methods with domain-specific fine-tuning:

- Siamese-FC [11]
- ARTrack [45]
- DiMP-50 [5]
- SuperDiMP-50 [16]
- ToMP-50 [23]
- PrDiMP-50 [17]
- TaMOs-Swin Base [24]

Trackers are re-initialized based on ground truth if they fail to predict the puck for 10 consecutive frames following evaluation practices in other papers[47].

3.3. Evaluation Metrics

- **AUC:** Area Under the Curve of the success plot (varying IoU threshold).
- **P@20:** Percentage of frames where the predicted bounding box center is within 20 pixels of the ground truth.
- **Avg Restarts / Seq:** Frequency of tracker resets per sequence due to total loss.

3.4. Experimental Setup

All experiments use 1280×720 broadcast footage. Model finetuning took place on 2 x A6000 GPUs, and the selected backbone for each network was the most extensive available pre-trained backbone. Backbones were also frozen during finetuning to prevent overfitting and stabilize training on our small dataset, while allowing later layers and heads to specialize. Lastly, all models were finetuned for 30 epochs with a step-wise learning rate schedule that decays the learning rate at 15 epochs and 25 with a decay factor $\gamma = 0.2$.

4. Results and Analysis

Table 1. **Performance comparison of puck tracking models in image coordinate space with tracker resets after 10 missed frames.** The table metrics include Area Under the Curve (AUC(%)), Precision (P@20(%)), and the average number of times the tracker required re-initialization with the ground truth label per sequence.

Method	AUC(%)	P@20(%)	Avg Restarts / Seq
Siamese-FC [11]	21.7	33.8	1.35
ToMP-50 [23]	25.38	46.11	1.12
ARTrack [45]	26.7	51.8	1.18
DiMP-50 [5]	27.05	40.61	1.29
Super DiMP-50 [16]	29.85	44.75	1.20
PrDiMP-50 [17]	**31.25**	47.75	1.12
TaMOs-Swin Base [24]	14.66	**55.65**	**0.64**

Across all evaluated methods, **PrDiMP** and **TaMOs** emerge as the strongest performers, though for fundamentally different reasons. As shown in Table 1, PrDiMP achieves the highest overall AUC, indicating that it maintains coarse overlap with the puck for longer stretches of time. In contrast, TaMOs attains the highest center precision (P@20) and the lowest restart frequency, suggesting that when it is correct, it is substantially more accurate and more stable.

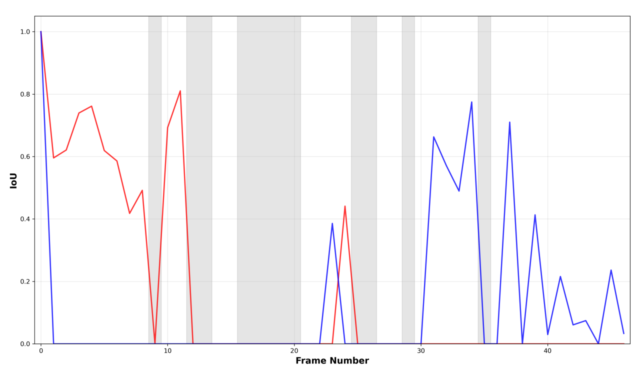

Figure 2. IoU over frame number comparing the best performing models, **PrDiMP** (red) and **TaMOs** (blue). Grey shaded regions indicate periods of occlusion where ground truth annotations are absent.

Global vs. Local Search Behavior. A key distinction between these models lies in how they process spatial context. TaMOs performs *global frame-level reasoning*: the Transformer backbone encodes the entire image at once and predicts object-specific filters in parallel. This global context allows TaMOs to re-detect the puck after occlusions and heavy motion blur more reliably than all other trackers, which depend on a localized *search region*. Because Siamese-FC, DiMP-family trackers, ToMP, and ARTrack restrict inference to a cropped neighborhood around the last known location, any tracking failure that pushes the true puck position outside this search space results in irreversible drift. This explains why TaMOs exhibits the fewest resets (0.64 per sequence), while template-based trackers fail to recover after losing the puck.

Figure 3. Tracking failure examples showing predictions from multiple trackers overlaid with colored bounding boxes: **Red** = Ground Truth, **Magenta** = DiMP, **Orange** = SuperDiMP, **Cyan** = ToMP, **Blue** = PrDiMP, **Green** = TaMOs, **Yellow** = SiamFC, **Purple** = ARTrack.

Occlusions and Recovery Dynamics. Figure 2 illustrates these behaviors directly. During occluded intervals (shaded regions), PrDiMP often maintains a weak but nonzero overlap, reflecting its probabilistic regression head that spreads uncertainty over a larger spatial region. However, the same uncertainty causes PrDiMP to accumulate small localization errors that eventually require reinitialization. TaMOs, conversely, tends to drop overlap quickly during occlusions—its predictions collapse when the puck is fully invisible—but once the puck reappears, the global frame encoding allows TaMOs to re-acquire it sharply, resulting in its much higher P@20 score.

Failure Modes: Drift, Stick-Locking, and Small-Object Ambiguity. Qualitative examples in Figure 3 highlight failure modes common across the evaluated trackers. The most severe issues occur in models like Siamese-FC, whose fixed-template matching causes them to "lock onto" visually similar distractors such as hockey sticks, skates, or patchy ice textures. When the puck passes behind a stick or merges with player clutter, these trackers frequently commit to the wrong object and never recover.

Additionally, many trackers struggle under *extreme motion blur*, where the puck becomes both aliased and elongated. Architecturally, this weakness is linked to the use of early convolutional layers with a stride of two, which discards fine-grained spatial detail that is critical for detecting a 3–5 pixel object in broadcast footage. Even high-performing modern architectures exhibit degraded confidence under these conditions, leading to sudden drift or delayed recovery.

5. Conclusion

Our evaluation demonstrates that standard object tracking paradigms break down when applied to the extreme conditions of broadcast hockey puck tracking. Methods that rely on a localized search window—such as Siamese-FC, DiMP, PrDiMP, SuperDiMP, ARTrack, and ToMP—struggle with irreversible drift once the puck leaves the search region, and are highly prone to stick-locking, distractor confusion, and failure under strong motion blur. Even architectures with probabilistic regression (PrDiMP) cannot reliably recover from extended occlusions or large displacements.

In contrast, TaMOs benefits from global frame-level reasoning, achieving the highest localization precision and the fewest restarts. Its Transformer-based global encoding enables sharper re-detection after occlusion and reduces catastrophic drift. However, TaMOs still suffers when the puck becomes heavily blurred or aliased, underscoring that global context alone is insufficient without fine-grained spatial detail.

Together, these findings reveal a central conclusion: *tiny, fast-moving objects violate the assumptions underlying most modern trackers*. Effective puck tracking will require new architectures that jointly preserve high-resolution detail, model rapid non-linear motion, and reason globally beyond a localized search window. Our analysis provides a foundation for future work aimed at designing trackers capable of operating under the extreme visual and temporal conditions present in ice hockey broadcast footage.

Acknowledgments

This work was supported by a grant with the Natural Sciences and Engineering Research Council (NSERC) partnered with Stathletes, Inc. Stathletes also provided the puck annotation data used in this research.

References

[1] Yasuo Ariki, Tetsuya Takiguchi, and Kazuki Yano. Digital camera work for soccer video production with event recognition and accurate ball tracking by switching search method. In *2008 IEEE International Conference on Multimedia and Expo*, pages 889–892, 2008. 1

[2] Bavesh Balaji, Jerrin Bright, Sirisha Rambhatla, Yuhao Chen, Alexander Wong, John Zelek, and David A Clausi. Domain-guided masked autoencoders for unique player identification. 1

[3] Bavesh Balaji, Jerrin Bright, Yuhao Chen, Sirisha Rambhatla, John Zelek, and David Clausi. Seeing beyond the crop: Using language priors for out-of-bounding box keypoint prediction. *Advances in Neural Information Processing Systems*, 37:102897–102918, 2024. 1

[4] Michael Beetz, Nico von Hoyningen-Huene, Bernhard Kirchlechner, Suat Gedikli, Francisco Siles, Murat Durus, and Martin Lames. Aspogamo: Automated sports games analysis models. *Int. J. Comput. Sci. Sport*, 8, 2009. 1

[5] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Learning discriminative model prediction for tracking. pages 6181–6190, 2019. 2, 3

[6] Bodhisattwa Chakraborty and Sukadev Meher. 2d trajectory-based position estimation and tracking of a ball in a basketball video. 2011. 1

[7] Bodhisattwa Chakraborty and Sukadev Meher. Real-time position estimation and tracking of a basketball. *2012 IEEE International Conference on Signal Processing, Computing and Control*, pages 1–6, 2012.

[8] Bodhisattwa Chakraborty and Sukadev Meher. A trajectory-based ball detection and tracking system with applications to shooting angle and velocity estimation in basketball videos. In *2013 Annual IEEE India Conference (INDICON)*, pages 1–6, 2013. 1

[9] Hua-Tsung Chen, Hsuan-Shen Chen, and Suh-Yin Lee. Physics-based ball tracking in volleyball videos with its applications to set type recognition and action detection. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, pages I–1097–I–1100, 2007. 1

[10] Hua-Tsung Chen, Wen-Jiin Tsai, Suh-Yin Lee, and Jen-Yu Yu. Ball tracking and 3d trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimedia Tools Appl.*, 60(3):641–667, 2012. 1

[11] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning, 2020. 2, 3

[12] Xina Cheng, Xizhou Zhuang, Yuan Wang, Masaaki Honda, and Takeshi Ikenaga. Particle filter with ball size adaptive tracking window and ball feature likelihood model for ball's 3d position tracking in volleyball analysis. pages 203–211, 2015. 1

[13] Xina Cheng, Masaaki Honda, Norikazu Ikoma, and Takeshi Ikenaga. Anti-occlusion observation model and automatic recovery for multi-view ball tracking in sports analysis. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1501–1505, 2016. 1

[14] Kyuhyoung Choi and Yongduek Seo. Probabilistic tracking of the soccer ball. pages 50–60, 2004. 1

[15] Kyuhyoung Choi and Yongduek Seo. Tracking soccer ball in tv broadcast video. pages 661–668, 2005. 1

[16] Seokeon Choi, Junhyun Lee, Yunsung Lee, and Alexander Hauptmann. Robust long-term object tracking via improved discriminative model prediction, 2020. 2, 3

[17] Martin Danelljan, Luc Van Gool, and Radu Timofte. Probabilistic regression for visual tracking, 2020. 2, 3

[18] Xin Duan. *Automatic determination of puck possession and location in broadcast hockey video*. PhD thesis, University of British Columbia, 2011. 1

[19] Xiao Han, Qi Wang, and Yongbin Wang. Ball tracking based on multiscale feature enhancement and cooperative trajectory matching. *Applied Sciences*, 14:1376, 2024. 1

[20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn, 2018. 2

[21] Marco Leo, N. Mosca, Paolo Spagnolo, and Pier Luigi Mazzeo. A semi-automatic system for ground truth generation of soccer video sequences. pages 559 – 564, 2009. 1

[22] Muyu Li, Henan Hu, and Hong Yan. Ice hockey puck tracking through broadcast video. *Neurocomputing*, 551:126484, 2023. 1, 2

[23] Christoph Mayer, Martin Danelljan, Goutam Bhat, Matthieu Paul, Danda Pani Paudel, Fisher Yu, and Luc Van Gool. Transforming model prediction for tracking. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8721–8730, 2022. 2, 3

[24] Christoph Mayer, Martin Danelljan, Ming-Hsuan Yang, Vittorio Ferrari, Luc Van Gool, and Alina Kuznetsova. Beyond sot: Tracking multiple generic objects at once. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6826–6836, 2024. 2, 3

[25] Ken Nsiempba, Amir Nazemi, David Clausi, and John Zelek. Leveraging player tracking for event detection in ice hockey. *Journal of Computational Vision and Imaging Systems*, 10 (1):69–74, 2024. 1

[26] Hemanth Pidaparthy and James H Elder. Keep your eye on the puck: Automatic hockey videography. In *Winter Conference on Applications in Computer Vision (WACV)*, 2019. 1, 2

[27] Harish Prakash, Jia Cheng Shang, Ken M Nsiempba, Yuhao Chen, David A Clausi, and John S Zelek. Multi player tracking in ice hockey with homographic projections. *arXiv preprint arXiv:2405.13397*, 2024. 1

[28] Arjun Raj, Lei Wang, and Tom Gedeon. Tracknetv4: Enhancing fast sports object tracking with motion attention maps, 2024. 2

[29] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018. 2

[30] Jinchang Ren, James Orwell, Graeme A. Jones, and Ming Xu. Tracking the soccer ball using multiple fixed cameras. *Computer Vision and Image Understanding*, 113(5): 633–642, 2009. Computer Vision Based Analysis in Sport Environments. 1

[31] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016. arXiv:1506.01497 [cs]. 3

[32] Liam Salass, Jerrin Bright, Amir Nazemi, Yuhao Chen, John Zelek, and David Clausi. Ice hockey puck localization using contextual cues. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6050–6059, 2025. 1

[33] Liam Salass, Jerrin Bright, Amir Nazemi, Yuhao Chen, John Zelek, and David Clausi. Ice hockey puck localization using contextual cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6106–6115, 2025. 2

[34] Mehdi Houshmand Sarkhoosh, Sayed Mohammad Majidi Dorcheh, Cise Midoglu, Saeed Shafiee Sabet, Tomas Kupka, Dag Johansen, Michael A. Riegler, and Pål Halvorsen. AI-Based Cropping of Ice Hockey Videos for Different Social Media Representations. *IEEE Access*, 12:118227–118249, 2024. Conference Name: IEEE Access. 1, 2, 3

[35] Marjan Shahi, David Clausi, and Alexander Wong. Goalienet: A multi-stage network for joint goalie, equipment, and net pose estimation in ice hockey. *arXiv preprint arXiv:2306.15853*, 2023. 1

[36] Prithvi Raj Singh, Raju Gottumukkala, Anthony S. Maida, Alan B. Barhorst, and Vijaya Gopu. Physics-guided fusion for robust 3d tracking of fast moving small objects, 2025. 2

[37] KINEXON Sports. What data-driven hockey software and puck data can do for coaches and players. https://kinexon-sports.com/blog/data-driven-hockey-software/, 2025. Accessed: 2025-03-10. 1

[38] Sijia Tian. *Group event recognition in ice hockey*. PhD thesis, University of British Columbia, 2018. 1

[39] Moumita Roy Tora, Jianhui Chen, and James J Little. Classification of puck possession events in ice hockey. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 147–154. IEEE, 2017. 1

[40] Rejin Varghese and Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pages 1–6, 2024. 3

[41] Kanav Vats, Mehrnaz Fani, David A Clausi, and John Zelek. Multi-task learning for jersey number recognition in ice hockey. In *Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports*, pages 11–15, 2021. 1

[42] 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, 2

[43] Kanav Vats, Pascale Walters, Mehrnaz Fani, David A Clausi, and John S Zelek. Player tracking and identification in ice hockey. *Expert systems with applications*, 213:119250, 2023. 1

[44] Jorge Armando Vicente-Martínez, Moisés-Vicente Márquez-Olivera, Abraham García-Aliaga, and Viridiana Hernández-Herrera. Adaptation of yolov7 and yolov7_tiny for soccer-ball multi-detection with deepsort for tracking by semi-supervised system. *Sensors (Basel, Switzerland)*, 23, 2023. 1

[45] Xing Wei, Yifan Bai, Yongchao Zheng, Dahu Shi, and Yihong Gong. Autoregressive visual tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9697–9706, 2023. 3

[46] Greg Wyshynski. Nhl brings advanced puck tracking stats to public, 2023. 1

[47] Mengjie Xu, Yitao Zhu, Haotian Jiang, Jiaming Li, Zhenrong Shen, Sheng Wang, Haolin Huang, Xinyu Wang, Qing Yang, Han Zhang, et al. Mitracker: Multi-view integration for visual object tracking. *arXiv preprint arXiv:2502.20111*, 2025. 3

[48] Xionghao Yang. Where is the puck? tiny and fast-moving object detection in videos. 2021. 1, 2