Evaluating Multi-Object Trackers for Collegiate Hockey

Henry Robb
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, New York
United States
robbh@rpi.edu

***Abstract*—Advanced analytics are critical in NCAA Division I ice hockey, however, the cost of proprietary software creates a significant barrier for collegiate programs. This project tries to address this gap by developing an automated tracking by detection computer vision pipeline. Ice hockey presents unique Multi-Object Tracking (MOT) challenges. These include identical team uniforms, several physical occlusions, and in our study, distortion from static overhead cameras. To overcome these domain shifts, a YOLOv8 object detector was fine tuned and integrated with four MOT architectures: ByteTrack, StrongSORT, HybridSORT, and BoT-SORT. Performance was tested against a human annotated ground truth data set using a Domain Adapted MOTA metric. The results show that global feature extractors hit a performance ceiling because of a lack of differentiating features among the uniforms, and this proves that tracking architecture is far more important than the Re-ID backbone. BoT-SORT achieved the highest overall accuracy (96.35% MOTA) by tracking bounding boxes' dimensions independently, allowing the tracker to successfully maintain player identities through severe occlusion and contact. The study concludes that AI models deployed in athletic domains must utilize geometric physics and spatial state estimation rather than traditional deep learning appearance extraction and this lays the groundwork for accessible, automated college sports analytics.**

I. Introduction

A. Background and Motivation

In areas as competitive as Division I NCAA ice hockey, the proper utilization of advanced analytics have become an essential component to team success and player development. While professional organizations like the the National Hockey League (NHL) have huge budgets to purchase and develop their own tracking software, college programs often lack the resources [1] [2]. Thus, there is a significant gap between the footage collected by college teams and the extraction of useful statistics. The option of manual annotation to record players' movements exists and is practiced to a certain extent, but this process is so costly as to take a human months to complete a single game. Therefore, developing an automatic and accessible computer vision pipeline is necessary to allow access to sports analytics for college programs.

B. Problem Statement

This project addresses the challenge of automating player tracking from hockey footage. Multi-Object Detection (MOT) in ice hockey has some unique challenges for tracking and detection algorithms. The greatest issue is that teammates wear nearly identical uniforms, so many traditional tracking features are ineffective when players are close to one another. This is exacerbated by the highly physical nature of the game. Players are constantly coming into contact or changing direction at high speeds in close proximity. Another issue is related to the angle of view of the camera. Because nearly no college team has a camera that able to remain directly above the play at all times, there will be distortion. It is going to muddle any tracker that is purely tracking based because it causes wild bounding box fluctuations depending on the players' postures and positions on the ice.

C. Objective and Scope

The primary objective of the project is to create a automated pipeline that can convert raw hockey footage into player coordinate data. This scope includes extracting frames, generating bounding boxes for players, and connecting these detections across time to maintain a consistent player identity. To assess the performance, this project uses an adapted Multiple Object Tracking Accuracy (MOTA) metric to evaluate the models against human annotated ground truth data.

D. Overview of Methodology

The project uses a supervised learning tracking by detection methodology. To begin the pipeline, a You Only Look Once (YOLO) object detection model is fine tuned using an open source annotated hockey database in order to maximize bounding box accuracy. Next, in order to overcome the tracking challenges of occlusion and camera angle, the YOLO detection are fed into MOT algorithms of various complexity. The evaluated models include ByteTrack for pure motion estimation, StrongSORT for deep learning appearance extraction, HybridSORT for tracklet confidence modeling, and BoT-SORT for advanced spatial state estimation.

II. Related Work

A. Evolution of Multi-Object Tracking

The tracking by detection approach has become ubiquitous in MOT literature across the past decade. Foundational work by Bewley et al. introduce Simple Online and Realtime Tracking (SORT) [3]. This showed that using a Kalman Filter for motion detection, alongside the Hungarian algorithm for bounding box association, achieved strong results in terms of both speed and accuracy. To build upon SORT, and attempt to alleviate the struggles it had with occlusions, Wojke et al. introduced DeepSORT [4]. DeepSORT integrated a deep learning Re-Identification (Re-ID) network to extract visual appearance features.

More recently, advancements on SORT have become far more specialized to attack more complex occlusions. Zhang et al. developed ByteTrack which did not use Re-ID [5]. It approached the problem by retaining low confidence detections and used them to bridge occlusion gaps purely utilizing motion estimates. Du et al., however, continued to improve upon the appearance based approach introduced by DeepSORT. They did so with StrongSORT by combining advanced motion estimation techniques with robust Re-Identification networks, such as OSNet [6]. Another attempt to handle occlusions resulted in Yang et al. creating HybridSORT [7]. This model utilizes Tracklet Confidence Modeling (TCM) to interpret weak cues, like confidence scores and height. Finally, BoT-SORT, proposed by Aharon et al., combined several advancements and introduced superior spatial state estimation to independently track bounding box dimensions [8].

B. Sports Analytics

While MOT algorithms are highly optimized, their traditional benchmark are pedestrian datasets, like MOT17, that feature uniquely clothed individuals [1] [8]. Therefore applying these models to sports footage does lead to some issues. Ice hockey, in particular, falls victim to several of the traditional tracking vulnerabilities. Teammates share identical uniform colors and logos, and this severely degrades the tracking power of Re-ID networks that were trained on pedestrians and their distinct clothing.

C. Addressing the Research Gap

Existing literature often evaluates modern MOT algorithms within highly controlled, usually pedestrian, environments. There is little existing analysis regarding how these specific architectural differences (pure motion, appearance Re-ID, TCM, and spatial state estimation) compare when faced with the chaotic college hockey environment. This project will also build upon the foundational CLEAR MOT metrics created by Bernardin and Stiefelhagen by simplifying one of their metrics to create a Domain Adapted MOTA [9]. This allows for a direct comparison of how effectively these modern trackers handle identical uniforms and severe occlusion in a real world sports analytics pipeline.

III. Methodology

A. System Architecture

This project uses the tracking by detection paradigm. This is a two stage approach that first identifies objects within individual frames, then associates these bounding boxes across time to form a continuous identification. This architecture comprising of multiple steps is helpful because it means that both components can individually be optimized. The ability to independently optimize each step has allowed for more pointed experiments in which different tracking architectures are tested while holding the detection inputs constant. The pipeline is implemented using the Ultralytics framework for object detection and the BoxMot library for MOT integration.

B. Datasets and Processing

To best meet the requirements of training, feature extraction, and evaluation, this project uses three distinct datasets:

- **HockeyAI Dataset [10]:** An open source collection of annotated hockey images used to fine tune the object detection model. This helped the object detection model adapt to domain specific classes like players and referees
- **RPI Hockey Footage:** This includes several periods of raw RPI Hockey footage, from both the Men's and Women's NCAA Division I teams. This footage was taken from a static camera that was positioned above the ice on the west side of the Houston Field House, RPI hockey's home ice. Some of this footage was manually annotated to create ground truths for the model evaluation
- **MSMT17 Dataset [11]:** This is a large multi scene pedestrian dataset used for transfer learning. Because it would be impossible to design an effective Re-ID network from scratch based on RPI hockey players with the current amount of data, this project uses pre-trained feature extractors that were trained on this dataset. Specifically, two Omni-Scale Networks (OSNet, *osnet_x0_25_msmt17.pt*), (OSNet, *osnet_x1_0_msmt17.pt*), and ResNet50 (*resnet50_msmt17.pt*) are systematically tested within the StrongSORT and BoT-SORT architecture to analyze the trade offs between Re-ID accuracy and processing speed.

C. Object Detection: YOLOv8

The first part of the pipeline uses YOLOv8 because of its impressive balance between mean Average Precision (mAP) and lightning fast speeds [12]. The base model was fine tuned on the HockeyAI dataset to recognize several classes (faceoff, goal, goaltender, player, etc.) so that it could reject false positives from the rink environment. During inference, the model was restricted to domain specific classes, like players and goalies, with a strict confidence threshold of 0.33. Additionally, Non-Maximum Suppression (NMS) with an Intersection over Union (IoU) threshold of 0.33 was applied to filter overlapping boxes before the data was passed downstream to tracking algorithms.

D. Multi-Object Tracking (MOT) Algorithms

To best evaluate solutions to the domain specific challenges of hockey, such as identical uniforms and occlusion, the YOLO detections were processed through four independent tracking algorithms of increasing complexity.

E. ByteTrack

ByteTrack serves as a pure motion based tracking baseline. Traditional trackers discard low confidence YOLO detections, and this results in fragmented trajectories during motion blur. However, ByteTrack incorporates these low confidence detections into its Kalman Filter association step. This allows it to bridge short term occlusions through only using spatial velocity [5]. This means that it is completely omitting Re-ID networks, and therefore, maximizing it computation efficiency.

F. StrongSORT

Despite being well designed, ByteTrack's architecture means that it will struggle with prolonged occlusions. StrongSORT introduces a deep learning Re-ID model to extract visual embeddings from bounding boxes. By calculating the cosine distance between the appearance features of current detections, and historical tracklets, StrongSORT tries to re-identify players as the emerge from occlusions [6].

G. HybridSORT

HybridSORT is implemented to use weak cues, and this allows it to overcome some of the issues that StrongSORT faces, like the uniformity of all the players' jerseys. Instead of relying only on appearance, HybridSORT uses Tracklet Confidence Modeling (TCM). It tracks the degradation of the YOLO confidence score that always occurs when a player is partially or totally occluded. It then uses this statistical drop as a cue to maintain the identity of the background player throughout the event [7].

H. BoT-SORT

The final tracker, BoT-SORT, was evaluated to test the limits of spatial state estimation. The algorithm is widely recognized for its Camera Motion Compensation (CMC), but this does not have as great of relevance to this project because of the static nature of the game footage being tested. Instead, its advantage for this dataset is in its modified Kalman Filter architecture. Most trackers estimate bounding box aspect ratios. These will vary wildly with the camera angle that this project's footage has as players change their position, move further and closer to the camera, and collide. BoT-SORT estimates the width and height of the bounding box directly. This results in far more stable coordinate predictions during both occlusions and rapidly changing scenes [8].

IV. Experimental Setup

A. Data Preprocessing and Spatial Alignment

To ensure both proper model training and evaluation, custom preprocessing pipelines were created for both the training and the results datasets.

First, to train the YOLOv8 object detector would require the proper HockeyAI dataset inputs. A custom script was developed to process the raw image frames and their corresponding annotations to ensure they were formatted to meet the strict architecture guidelines of Ultralytics. The data was then partitioned via an 80/20 train validation split. This led to 1680 training frames and 421 validation frames. The script automated the redistribution of these files into the required `images` and `labels` directory structure and generated the YAML file mapping the dataset's target classes.

Second, a spatial alignment pipeline was needed for model evalutation. Standard evaluation protocols for MOT rely on bounding box IoU to compare model predictions against ground truth. However, the collegiate ground truth data utilized in this study was manually created with (x, y) coordinate pairs representing the true center of the player's bounding box. To close this gap, a custom processing pipeline was created. Default YOLO tracking will usually export the bottom center coordinate of the bounding box. With the largely overhead view the proprietary data has, the difference between a player's torso and skate can be several pixels, and that would introduce a huge number of errors during evaluation. To correct this, the spatial output of all four tracking pipeline were processed to calculate the true center of their bounding boxes (C_x, C_y) where $C_x = (x_1 + x_2)/2$ and $C_y = (y_1 + y_2)/2$, ensuring alignment with the ground truth anchor points before scoring the outputs with MOTA.

B. Configurations and Hyperparameters

To make certain that the ablation study was rigorous, the detection variables were controlled across all four tracking experiments. The YOLOv8 feature extractor was configured with confidence and IoU for NMS thresholds of 0.33. To simplify the model and ensure a tight focus on the problem statement, the model was restricted to only predict three specific classes: players, goalies, and referees.

For the appearance based tracking models, StrongSORT, HybridSORT and BoT-SORT, three feature extractors were tested to compare Re-ID efficiency. All three loaded in weights that were trained on the MSMT17 dataset. There were two OSNets: *osnet_x0_25_msmt17.pt* acted as the lighter comparison, *osnet_x1_0_msmt17.pt* was the baseline. Additionally, ResNet50 (*resnet50_msmt17.pt*) acted as a heavier model comparison. This configuration allowed for a comparison of tracking accuracy (MOTA) versus computational speed (FPS) while keeping the YOLO detections constant.

C. Evaluation

To evaluate the proposed pipelines, a custom Domain Adapted MOTA was developed. Additionally, because the ground truth was a single coordinate point for each detection, rather than using a traditional IoU threshold, a Euclidean distance threshold was used instead.

For every frame, the scoring engine calculates the distance between the ground truth coordinate and the predicted bounding box centroids. A successful match, True Positive, is registered if the predicted centroid falls within 50 pixels of the ground truth point. If no prediction exists within this radius, a False Negative is registered. Very importantly, the engine is also tracking the Identity (ID) assigned to each player. If the player successfully locates the player, but assigns a different

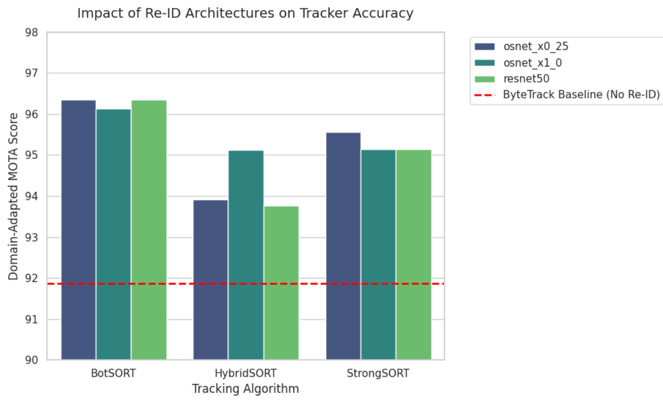

Fig. 1. Bar chart showing the custom MOTA scores for each combination of tracker and Re-ID model

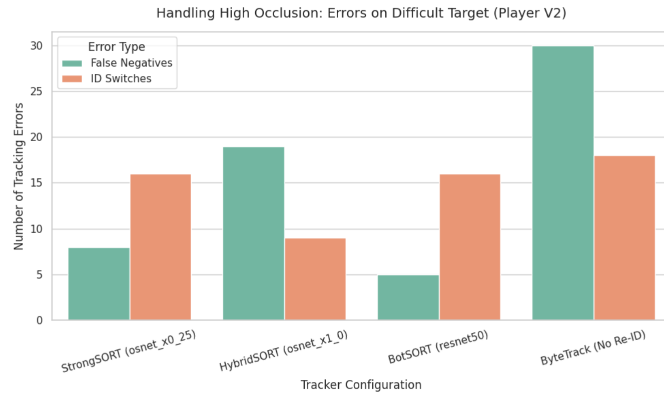

Fig. 2. Bar chart showing the false negative and ID switch errors for each tracker on Player V2

ID than the previous frame, which would indicate that it lost the player, then an Identity Switch ($IDSW$) is recorded.

The final performance of each model is calculated using the following variation of the MOTA formula where GT_{total} represents the total number of ground truth annotations for a specific player across an entire video sequence:

$$MOTA = 1 - \frac{FN + IDSW}{GT_{total}} \quad (1)$$

This formula purposefully penalized $IDSW$ very heavily. This will ensure that the metric will be a strong evaluation of how well each tracking algorithm handles the extreme occlusions that happen regularly within the sport of hockey.

V. Results

A. Architectural Impact Versus Re-Identification Backbones

Applying the custom MOTA to the collegiate footage provided quantitative insights into the effectiveness of appearance features versus using pure spatial tracking. As shown in Figure 1, the pure motion baseline, ByteTrack, achieved a MOTA of 91.85%. Despite being the lowest score among all of the models, it still was a very high accuracy baseline, and this proves that spatial tracking is effective even when used independently of appearance data.

When appearance data was introduced, the results pointed to the fact that the overall tracker architecture was a far more important factor for success than the choice of the Re-ID network. Changing the Re-ID network from an OSNet (*osnet_x1_0_msmt17.pt*) to a much more computationally heave ResNet50 (*resnet50_msmt17.pt*) within the same tracker lead to very minimal changes in the overall MOTA scores. For example, upgrading the BoT-SORT architecture from the lightest OSNet to ResNet50 had no change in performance as they both achieved a score of 96.35%. Additionally, the baseline OSNet was only marginally lower at a score of 96.13%.

B. High Occlusion Performance (Player V2)

To understand why the architectural differences drove the final scores, we looked at the performance on the most difficult target in the test footage, "Player V2", who experienced several severe occlusions. As shown in Figure 2, the different tracking architectures resulted in different balances of False Negatives (FN) and Identity Switches (IDSW).

ByteTrack struggled significantly with the occlusion, losing the player completely for 30 frames (FN) and swapping IDs 18 times. StrongSORT reduced total errors but still suffered 16 IDSW on this player. HybridSORT's reliance on weak cues drastically reduced ID Switches to a field leading 9, but at the cost of 19 False Negatives. In the end, BoT-SORT's state estimation allowed it to maintain the bounding box, even during physical overlap. This resulted in only 5 False Negatives and is what led to its superior MOTA score.

C. Inference Efficiency

Finally, these models were evaluated for computational efficiency using Frames Per Second as the metric. This allowed for a mapping of the tradeoff between speed versus accuracy, as visualized in Figure 3. As we expected, because of its lack of Re-ID, the baseline ByteTrack was by far the fastest model. It was able to process at 13.15 FPS. Also unsurprisingly, the older StrongSORT was the slowest with a rate of 4.58 FPS at its absolute quickest.

However, an anomaly emerged regarding the Re-ID network weights for HybridSORT and BoT-SORT. It was expected that the lightweight OSNet models would process significantly faster than the heavy ResNet50 architecture. Instead, for these two models, the ResNet50 backbone had the quickest processing speeds of their respective Re-ID configurations. With *resnet50_msmt17.pt* BoT-SORT achieved 7.53 FPS and HybridSORT achieved 7.92 FPS. However, with *osnet_x0_25_msmt17.pt*, the lightest model, they only processed at 6.68 and 6.67 FPS respectively.

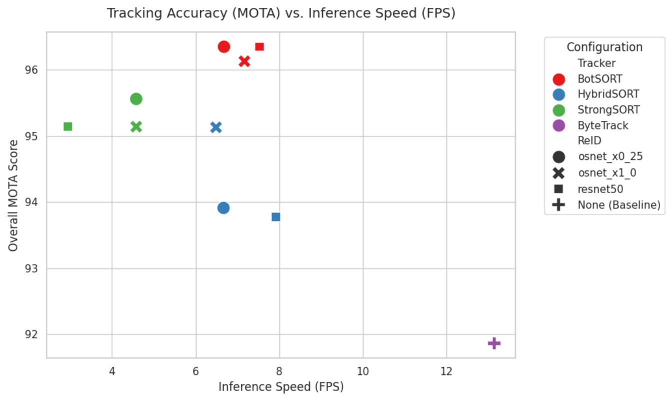

Fig. 3. Scatter plot demonstrating the difference of efficiency versus accuracy for each tracker and their possible Re-ID models

VI. Discussion

A. State Estimation Over Appearance

The central task of this project was create a successful, automated computer vision pipeline, and an important component of this was to see how heavily MOT algorithms should rely on visual appearance features when used in a sport environment. The results very clearly indicate that in domains where subjects wear identical uniforms, like ice hockey, tracking architecture, specifically spatial state estimation, is far more important than the complexity of the Re-Identification network.

Because hockey players on the same team share identical colors, logos, and helmet designs, global feature extractors hit a performance ceiling very early on. This explains why upgrading from a lightweight OSNet (2.2 million parameters) to a heavy ResNet50 (23.5 million parameters) resulted in minimal improvement of overall MOTA scores when looking at BoT-SORT and StrongSORT. The larger networks simply were not able to extract enough differentiating visual information from the identical jerseys to justify the increased size. Additionally, the fact that ByteTrack, which does not use any appearance features, achieved a pretty high MOTA of 91.86% further highlights that spatial geometry and velocity are the more critical tracking cues within the collegiate hockey environment.

B. Understanding Tracking Behavior During Occlusion

The isolated results on "Player V2" revealed the theoretical differences between the evaluated architectures. ByteTrack's reliance on pure motion led to its downfall in this instance. The player in the footage stopped to participate in a board battle, during which their velocity dropped to zero while also being occluded. This led to the Kalman filter losing its predictive momentum and the 30 False Negatives.

StrongSORT also struggled because its Re-ID module could not differentiate Player V2 from their teammate after the board battle. HybridSORT, however, showed the promise of Tracklet Confidence Modeling (TCM). By treating the natural drop in YOLO confidence as a weak cue that an occlusion was occurring, it successfully kept the algorithm from incorrectly merging the two players. This is how the HybridSORT achieved the best Identity Switches error rate of only 9.

Ultimately, BoT-SORT achieved the best results by treating the bounding box geometry differently. By abandoning the traditional aspect ratio estimation which is utilized by StrongSORT and ByteTrack, BoT-SORT tracked the width and the height of players independently. From a static overhead camera angle, like the one used in this experiment, a player's aspect ratio distorts wildly as the players bend and skate. This is especially the case during collisions where players can go from standing up straight to lying on the ice in only a few frames. BoT-SORT's independent dimensional tracking prevented the bounding boxes from collapsing during physical contact, yielding only 5 False Negatives and leading to the highest overall MOTA score.

C. Analyzing the Inference Efficiency and the Hardware Optimization Anomaly

In deploying these models for collegiate use, computational efficiency (FPS) is a critical metric. The baseline ByteTrack performed as anticipated by processing at over 13 FPS, due in large part to its lack of a secondary neural network.

Beyond this initial finding, the Re-ID efficiency results presented quite the anomaly: the huge ResNet50 backbone processed faster than the lightweight OSNet backbone when integrated with BoT-SORT and HybridSORT. Theoretically, OSNet having so many fewer parameters should have led to faster inference times. Thus, we have hypothesized that this inversion is being caused by hardware level optimizations on the NVIDIA T4 GPUs that were used in this project.

The ResNet architecture is a standard across the industry. Therefore, modern GPUs are hyper optimized to execute its specific convolutional layers [13]. OSNet, while smaller, relies on more complex multi stream residual blocks and these require non standard memory access patterns [14]. On a T4 GPU, these irregular operations will not run as efficiently, creating a memory bandwidth bottleneck that makes the theoretically lighter model process slower in practice [15]. Therefore, for deployment on standard cloud compute GPUs, ResNet50 provides a superior speed to accuracy ratio for these specific tracking pipelines.

VII. Conclusion

A. Summary of Contributions

The project successfully created an automated pipeline capable of extracting continuous player coordinate data from collegiate ice hockey footage. By developing a custom Domain Adapted MOTA metric, and a spatial alignment preprocessing pipeline, this study thoroughly evaluated four distinct tracking architectures. The evaluation was done under the constraints of a static overhead camera, and within the ice hockey environment. This meant that teams shared identical jerseys and there was prolific occlusion. The results demonstrated that BoT-SORT is the optimal tracking architecture for this environment as it peaked at a 96.35% MOTA. Its success was driven by its advanced spatial state estimation, specifically, its ability to track bounding box width and height independently. This allowed it to avoid the instability of traditional aspect ratio tracking during severe physical occlusions and collisions.

B. Implications for the Broader Field of AI

These findings have implications when looking at the consequences of deploying computer vision models across severe domain shifts. The plateaued performance of Re-Identification networks in this study showed a vulnerability in modern MOT architectures: algorithms that have been optimized for pedestrian environments will struggle when global appearance features become uniform. This research proves that in visually homogeneous domains, AI architectures must prioritize spatial state estimation and weak cue confidence modeling rather than traditional deep learning appearance extraction.

C. Future Research and Improvements

Future research will need to address several limitations found in this study in order to prepare the field for collegiate sports analytics that will rival professional leagues. First, to resolve the ever present challenge of Identity Switches during physical play, future work should focus on training custom Re-ID models on hockey players. These models will utilize localized features such as jersey numbers rather than global clothing colors. Next, overcoming data scarcity remains a huge challenge for training domain specific models. Because the ground truth data is painstaking to manually create, substantial time and effort will be required to build up a suitable supply of annotated collegiate footage. Finally, hardware upgrades present a practical path for pipeline improvement. Using cameras with higher resolutions, or introducing the ability to pan, could significantly reduce the top down geometric distortion that currently challenges spatial state estimation algorithms.

References

[1] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A comprehensive review of computer vision in sports: Open issues, future trends and research directions," *Applied Sciences*, vol. 12, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/9/4429

[2] National Hockey League, "Nhl edge website provides puck and player tracking data to fans," https://www.nhl.com/news/nhl-edge-launches-website-for-puck-and-player-tracking-data, 2023.

[3] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, p. 3464–3468. [Online]. Available: http://dx.doi.org/10.1109/ICIP.2016.7533003

[4] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017. [Online]. Available: https://arxiv.org/abs/1703.07402

[5] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," 2022. [Online]. Available: https://arxiv.org/abs/2110.06864

[6] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "Strongsort: Make deepsort great again," 2023. [Online]. Available: https://arxiv.org/abs/2202.13514

[7] M. Yang, G. Han, B. Yan, W. Zhang, J. Qi, H. Lu, and D. Wang, "Hybrid-sort: Weak cues matter for online multi-object tracking," 2024. [Online]. Available: https://arxiv.org/abs/2308.00783

[8] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," 2022. [Online]. Available: https://arxiv.org/abs/2206.14651

[9] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:13567980

[10] SimulaMet-HOST, "Hockeyai dataset," https://huggingface.co/datasets/SimulaMet-HOST/HockeyAI, 2024.

[11] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 79–88.

[12] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, p. 1680–1716, Nov. 2023. [Online]. Available: http://dx.doi.org/10.3390/make5040083

[13] NVIDIA Corporation, "Nvidia turing architecture in-depth," https://images.nvidia.com/aem-dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf, 2018.

[14] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," 2019. [Online]. Available: https://arxiv.org/abs/1905.00953

[15] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116–131.

Appendix A
Code Repository

All of the code for this project can be found on GitHub. The only component that is missing is the YOLOv8 weights, but there are details in the README.md on where to find those.

GitHub Repository: https://github.com/HenryRobb/CSCI-6170/tree/main/Project

Appendix B
Contributions

As the sole team member, I, Henry Robb, did all of the development and writing. Tom Morgan provided access to the RPI game footage for me to use.