Towards Efficient Online and Offline Tracking of Broadcast Ice Hockey

Axel Bogos [1] **Mathieu Péloquin** [1] **Mohamed Moosa** [1]

Abstract

In this study, we investigate the performance of the YOLOv8 single-stage object detection model in Multiple Object Tracking (MOT) within ice hockey. We integrate YOLOv8 both with online (SORT, DeepSORT) and offline, graph-based (MOT Neural Solver) tracking algorithms, evaluating on the MOT17 and McGill Hockey Tracking Dataset (MHTD) 2020. Our key finding is that the finetuned YOLOv8x model, combined with the Neural Solver, achieved above MOT17 benchmark tracking performance on MHTD with a HOTA score of 0.776, showcasing its suitability for dynamic and occlusion-prone environments. Additionally, YOLOv8m demonstrated effective tracking with greater efficiency. These results highlight the potential of single-stage models in specialized MOT applications.

1. Introduction

The field of Multiple Object Tracking (MOT) represents a signficant and continually evolving area in the realm of computer vision. Its applications are diverse, spanning from traffic or crowd monitoring, fall detection in elderly care homes and to the more dynamic and challenging domain of sports analytics. In recent years, the complexity of MOT tasks has increased substantially, driven by the need for higher precision and real-time processing capabilities in various scenarios. Our work delves into the application of MOT within the demanding context of ice hockey, a sport characterized by rapid player movements, frequent occlusions, and variable camera angles. These elements introduce significant challenges to traditional tracking methods, necessitating innovative approaches to ensure accurate and efficient player tracking.

*Equal contribution [1]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada. Correspondence to: Axel Bogos <axel.bogos@umontreal.ca>, Mathieu Péloquin <mathieu.peloquin.1@umontreal.ca>, Mohamed Moosa <mohamed.moosa@umontreal.ca>.

IFT 6269 project report - Université de Montréal

Our study aims to explore and evaluate a combination of online and offline state-of-the-art tracking algorithms that integrate PGM topics—SORT (Bewley et al., 2016), DeepSORT (Wojke et al., 2017), and the MOT Neural Solver (Brasó & Leal-Taixé, 2020)—integrated with the YOLOv8 single-stage object detection model (Jocher et al., 2023). To the best of our knowledge, it is the first evaluation of tracking frameworks combined with a single-stage detection model finetuned for ice-hockey.

We first review the pertinent literature, highlighting key developments in MOT and their relevance to sports tracking, particularly ice hockey. We then describe our methodologies, including data processing and the fine-tuning of object detection models. Subsequently, we present a comprehensive evaluation of the selected tracking algorithms using two distinct datasets: the benchmark MOT17 dataset (Milan et al., 2016) and the specialized McGill Hockey Tracking Dataset 2020 (Yingnan Zhao, 2020). Through this dual-dataset approach, we aim to provide a nuanced analysis of the algorithms' performance, from general tracking scenarios to the specific challenges of tracking in ice hockey. Our findings are intended to contribute valuable insights to the field of MOT and support the development of more robust and efficient tracking systems for sports.

2. Litterature Review & Related Work

The field of Multiple Object Tracking (MOT) spans a broad spectrum of methodologies, including both traditional algorithmic and deep learning-based techniques, each of which can be implemented in either online or offline modalities. Online tracking, essential for real-time applications, makes immediate decisions based on current video frames without access to future data. Offline tracking analyzes an entire sequence for more precise, global tracking but lacks real-time functionality. In this section, we explore significant contributions that align with our work, and we also consider comparable instances of tracking technology applied in sports settings.

2.1. Simple-Online-and-Realtime-Tracking (SORT)

Bewley et al. (Bewley et al., 2016) introduced the Simple-Online-and-Realtime-Tracking (SORT) algorithm, which uses a Kalman Filter for state estimation. This filter is instrumental in predicting the trajectories of moving objects, particularly in real-time contexts. The efficacy of the Kalman Filter lies in its ability to process sensor data that is typically noisy in dynamic environments. SORT further employs the Hungarian Algorithm (Kuhn, 1955) for associating object states over successive frames. Despite its efficiency, SORT exhibits limitations in scenarios characterized by high-speed, unpredictable movements or prolonged occlusions, conditions frequently encountered in the tracking of ice hockey players

2.2. DeepSORT

DeepSORT (Wojke et al., 2017) is an evolution of the SORT algorithm, integrating Convolutional Neural Networks (CNNs) for enhanced feature extraction. This advancement signifies a notable shift towards the incorporation of deep learning techniques in online MOT. The application of CNNs in DeepSORT enables more refined and robust tracking capabilities, particularly effective in scenarios characterized by frequent occlusions and rapid movement. DeepSORT's deep learning approach significantly improves upon the traditional SORT algorithm, particularly in its ability to maintain consistent extended tracking, reducing identity switches by 45% in rapid-movement scenarios.

2.3. MOT Neural Solver

The MOT Neural Solver (Brasó & Leal-Taixé, 2020) is an offline approach that combines graph theory with learning-based methods. This framework employs the network flow formulation of MOT, enabling it to operate effectively within a graph-based domain using Message Passing Networks (MPNs). By focusing on the data association step, the MOT Neural Solver extends the application of learning techniques in MOT beyond the conventional scope of feature extraction. This approach allows for a comprehensive processing of detection data, enabling the model to globally analyze and predict tracklets. A key aspect of this approach is its graphical formulation of the tracking problem. The MOT Neural Solver treats tracking as an undirected graph $G = (V, E)$, where $V = \{1, 2, ..., n\}$ represents a set of n nodes corresponding to n detections across all video frames. In this model, the edge set E encompasses connections between every pair of detections, thus allowing for the recovery of trajectories. The primary objective is to partition this graph into disjoint components, with each component representing a distinct trajectory T_i. Node and edge embeddings are computed using a CNN for the affinity calculation between detections d_i. Subsequently, the model employs a graph message passing algorithm responsible for classifying whether an edge in the graph signifies that the connected nodes (detections) are part of the same player trajectory. The efficacy of the MOT Neural Solver, as evidenced by improvements in both MOTA and IDF1 scores across various benchmarks, underscores its capability to adeptly manage complex challenges in MOT, especially in scenarios demanding an intricate understanding of object relationships and dynamics.

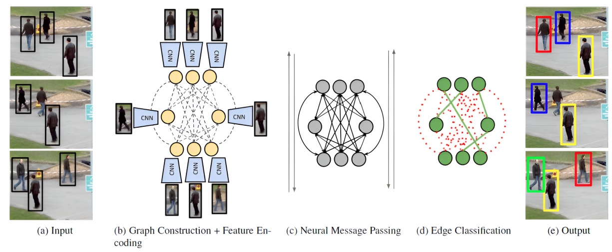

Figure 1. Overview of the Neural Solver. Fig 1 of Brasó et al, 2020

2.4. YOLOv8

In the domain of object detection, the YOLO (You Only Look Once) framework represents a family of single-stage models, prioritizing efficient real-time processing while maintaining accuracy; its most recent state-of-the-art iteration is YOLOv8 (Jocher et al., 2023). Unlike two-stage models such as R-CNN (Girshick et al., 2014) or Fast R-CNN (Girshick, 2015) which separate object localization and recognition into two detection stages, single-stage models jointly optimize the region proposal and bounding-box extraction, facilitating faster processing speeds crucial for real-time applications. To the best of our knowledge, the usage of YOLOv8 in the context of ice hockey tracking evaluation across multiple framework is novel. Its usage underscores the potential of YOLOv8 for detecting and tracking fast-moving objects under conditions of frequent occlusion, a common challenge in sports analytics and MOT.

2.5. Related Work in Sports

The application of Multiple Object Tracking (MOT) in sports has seen considerable growth, particularly in disciplines such as basketball and football (soccer). Notable works in these areas include Lu and Wei's approach to player tracking in basketball (Lu et al., 2013) and Acuna et al.'s research on real-time detection and tracking in football (Acuna, 2017). Additionally, the development of sports-specific benchmark datasets, such as SoccerNet-Tracking (Cioppa et al., 2022) and SportsMOT (Cui et al., 2023), which include sequences from basketball, football, and volleyball, have been instrumental in advancing this field. We further discuss ice hockey-specific datasets in Section 3. In the context of ice hockey, Vats et al. (Vats et al., 2021) have proposed a comprehensive player tracking and team identification pipeline. They reported optimal results in an offline setting using the Neural Solver coupled with a ResNet50 Feature Pyramid Network (FPN) (Lin et al., 2017).

3. Data

3.1. MOT17 Dataset

To ensure a comprehensive evaluation, our project uses the Multiple Object Tracking 2017 (MOT17) dataset (Milan et al., 2016) as an initial benchmark to validate our use of various trackers and evaluation pipeline. Recognized as a foundational dataset, MOT17 encompasses a wide range of scenarios, each captured in varied environments. The dataset consists of annotated tracklets predominantly from crowded scenes, presenting diverse challenges for tracking algorithms. It includes sequences with varying degrees of occlusion and instances of abrupt object movement making it an fairly strong benchmark for assessing the performance of MOT frameworks. Using MOT17 allows for an inital analysis of the chosen tracking algorithms and validation of our implementation of the evaluation script and data conversion distinct from the unique challenges of ice-hockey, as MOT17 benchmark results are known for the proposed trackers and thus we may directly validate our experimental results.

3.2. McGill Hockey Tracking Dataset 2020

The McGill Hockey Tracking Dataset (MHTD) 2020 (Yingnan Zhao, 2020) comprises an extensive collection of annotated tracklets, each associated with bounding boxes delineating the positions of players on the rink. The annotated tracklets reflect the high-velocity movements, abrupt directional changes, and frequent occlusions typical of ice hockey. Another feature of the dataset is its inclusion of varying camera angles across tracklets. This variability in perspective is instrumental in testing the adaptability and effectiveness of our detection and tracking algorithms. The MHTD provides ground truth data in a format similar but not identical to the MOT17 dataset. Each frame in the dataset is annotated with a set of bounding boxes, linked to an individual player given a unique ID in each tracklet. This slightly differs from the MOT challenge, as the MOT17 dataset assigns a new ID for individual object re-entering the frame in a different tracklet.

Each tracklet is provided as a HD mp4 video file, with half being 30 FPS and half being 60 FPS, which are the two typical frame rates for broadcast sports. In total, the MHTD dataset is made up 25 HD video clips, jointly having a total of 82,305 frames, 2056 individual tracklets and 632,785 annotated bounding boxes. As detailed in Section 5, we use the MHTD both in finetuning object detection models as well as evaluating tracking frameworks.

4. Metrics

Our evaluation of Multiple Object Tracking (MOT) frameworks incorporates multiple metrics to ensure a comprehensive assessment of tracking performance. These include Multiple Object Tracking Accuracy (MOTA), High-level Object Tracking Accuracy (HOTA), and Identity F1 Score (idf1). MOTA assesses overall tracking accuracy, HOTA measures the balance between detection and association quality, and idf1 focuses on identity maintenance across frames. Together, these metrics provide a robust, multidimensional view of tracking effectiveness, capturing both per-frame precision and temporal consistency.

4.1. Multiple Object Tracking Accuracy (MOTA)

MOTA is a comprehensive metric which served until recently as the main MOT challenge metric. It incorporates False Positives (FP), False Negatives (FN), and Identity Switches (IDS). MOTA is calculated as:

$$\text{MOTA} = 1 - \frac{FP + FN + IDS}{GT}$$

Here, GT represents the ground truth count of object identities in the dataset. MOTA scores range from -∞ to 1, with higher values indicating superior tracking performance. This metric effectively captures the overall accuracy of a tracking system, penalizing for misidentifications and continuity errors in object tracking.

4.2. High-level Object Tracking Accuracy (HOTA)

HOTA, currently the official MOT challenge metric, is a metric that extends beyond conventional tracking measures by integrating high-level attributes, notably localization accuracy (Loc-IoU). Loc-IoU is computed as the ratio of the intersection (overlap) between the Ground Truth Detection and the aggregate area encompassed by both the Ground Truth and Predicted Detections.

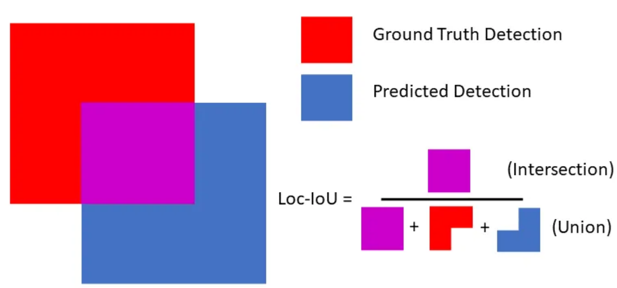

Figure 2. Illustration of HOTA Localization accuracy

HOTA synthesizes Localization Accuracy (LocA), Detection Quality (Det-IoU), and Association Accuracy (Ass-IoU) to produce the overall HOTA score. The formula for HOTA is given by:

$$\text{HOTA} = \sqrt{\frac{\sum_{c \in TP} Ass-IoU(c)}{|TP| + |FN| + |FP|}}$$

This metric provides a nuanced evaluation of tracking systems, balancing spatial accuracy and the continuity and consistency of tracking across time.

4.3. Identity F1 Score (idf1)

While MOTA and HOTA offer insights into overall tracking accuracy, the Identity F1 Score (idf1) is puts emphasis on maintaining object identities throughout a tracking sequence. The idf1 score is calculated using True Positives (TP), False Positives (FP), and False Negatives (FN), as per the following formula:

$$\text{idf1} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

This metric is particularly important in contexts like ice hockey, where preserving the consistent identities of players across long tracklets is critical. idf1 effectively quantifies the precision and recall of identity assignments, highlighting a tracking system's ability to accurately follow objects without identity switches or mismatches.

4.4. TrackEval

TrackEval (Jonathon Luiten, 2020) is the standard evalaution library for numerous object tracking challenges, including MOT (Milan et al., 2016). We have used it directly to evaluate our results on MOT17 from SORT, DeepSORT and the Neural Solver as well as adapting the code-base to evaluate the MHTD dataset. TrackEval allows us to obtain tracking metrics for multiple trackers paired with multiple object detection models, as shown in Section 6.

5. Methods

We integrate state-of-the-art tracking algorithms—SORT (Bewley et al., 2016), DeepSORT (Wojke et al., 2017), and the MOT Neural Solver (Brasó & Leal-Taixé, 2020)—with the YOLOv8 single-stage object detection model (Jocher et al., 2023). These methodologies encompass both online and offline tracking approaches within the tracking by detection (TBD) framework. The TBD framework consists of 3 main steps: (1) detecting objects frame-by-frame, (2) calculataing some affinity metric between detections across frames, (3) infering object ids across frames to produce tracklets. Concretely, for a whole video sequence, the YOLOv8 model produces a set of detections $D = \{d_1, \ldots, d_n\}$ where each detection is parametrized by $\{f_i, x_i, y_i, w_i, h_i, c_i\}$, where f_i is the frame number, x_i and y_i the top-left corner of a bounding box, w_i and h_i its width and height and c_i the model's confidence in the detection. Affinity calculation between detections d_i follows and may use simple features such as Intersection Over Union (IOU) or features extracted using deep neural networks. Finally, a set of tracklets $T = \{t_1, \ldots t_m\}$ which best explains D where t_i are time-ordered observations is determined at the final inference step, involving either the filtering for SORT and deepSORT, or the graphical formulation for the Neural Solver, as described in Section 2.

5.1. Data Processing

As a benchmark evaluation dataset, MOT17 (Milan et al., 2016) required no additional processing. MHTD required some additional processing to convert it to the MOT challenge format as well as a format suitable for object detection training. For each .mp4 sequence, all frames were extracted and saved as JPG files with Open-CV (Bradski, 2000). The clip level ground truth files we converted to distinct per-frame annotations file following the YOLO annotation format in order to train the YOLOv8. We have kept the proposed train-test split proposed by the authors, which defines these splits at the game level, keeping 22 clips in the training set and 3 clips of distinct other games in the test set. However, 1 training clip missed its corresponding ground truth annotation files, bringing the usable number of clips in the training set to 21.

5.2. Object Detection Finetuning

As remarked in (Bewley et al., 2016), the quality of detections is instrumental to the performance of tracking, therefore justifying the use of 2-stage detection models. However, performance by YOLO-family pretrained on COCO (Lin et al., 2015) has significantly improved in the last few years (Terven & Cordova-Esparza, 2023), notably with YOLOv7 (Wang et al., 2022), YOLO-Nas (Aharon et al., 2021) and YOLOv8 (Jocher et al., 2023). As such, we wished to evaluate tracking on ice-hockey both in a zero-shot single-stage object detection setting, as well as using a finetuned single-stage detection model. To that effect, we converted the MHTD dataset from an MOT-format to the standard, COCO-inspired object detection training format. We have evaluated both non-finetuned and finetuned versions of YOLOv8 of multiple sizes, namely YOLOv8n (3.2M parameters), YOLOv8s (11.2M), YOLOv8m (25.9M), YOLOv8l (43.7M) and YOLOv8x (68.2) with results shown in Table 1. All training was done locally on an Nvidia RTX 4080, with the backbone fully unfrozen and default training augmentations.

5.3. Evaluating Trackers

For each of the 3 tracker framworks being evaluated, we forked and adapted the original source code - each fork is available here. The main working repository, beside tracker framework forks can be found here. For older tracker implementation such as SORT, some changes were necessary to adapt them to the current MOT format, as well as comptability issues with current versions of numpy and upgrading from Tensorflow 1.X to Tensorflow 2.X in the case of deepSORT. Their respective output were then transferred to a local version of TrackEval (Jonathon Luiten, 2020), adapted for MHTD in order to have a standard benchmark of their performance on both MOT17 and MHTD.

6. Results & Discussion

‘

6.1. Results

Table 1. Comparison of Object Detection Test Metrics on MHTD

Model	**mAP50**		**mAP50-95**	
	Non-finetuned	Finetuned	Non-finetuned	Finetuned
YOLOv8n	0.740	0.874	0.329	0.563
YOLOv8s	0.833	0.883	0.380	0.585
YOLOv8m	0.866	0.885	0.415	0.5979
YOLOv8l	0.881	0.888	0.434	**0.602**
YOLOv8x	**0.885**	**0.894**	**0.443**	0.5995

Table 2. Tracker Performance on MOT17 Dataset

	SORT			**deepSORT**			**Neural Solver**		
Detection Model	M	H	I	M	H	I	M	H	I
DPM	0.265	0.340	0.311	0.313	0.404	0.391	0.591	0.643	0.661
FRCNN	0.479	0.486	0.492	0.492	0.542	0.541	0.646	0.688	0.700
SDP	**0.572**	**0.537**	**0.536**	**0.616**	**0.613**	**0.610**	**0.691**	**0.713**	**0.717**
Average	0.448	0.476	0.458	0.482	0.542	0.537	0.644	0.697	0.712

Note: M = MOTA, H = HOTA, I = IDF1

Table 3. Tracker Performance on MHTD Dataset

	SORT			**deepSORT**			**Neural Solver**		
Detection Model	M	H	I	M	H	I	M	H	I
YOLOv8n	-0.019	0.275	0.189	-0.183	0.354	0.258	-0.208	0.402	0.293
YOLOv8n*	0.601	0.529	0.417	0.517	0.585	0.485	0.586	0.664	0.550
YOLOv8m	-0.022	0.372	0.269	-0.207	0.418	0.311	-0.235	0.474	0.353
YOLOv8m*	0.788	0.634	0.533	0.752	0.668	0.577	0.853	0.758	0.655
YOLOv8x	0.015	0.394	0.290	-0.159	0.439	0.334	-0.180	0.498	0.379
YOLOv8x*	**0.828**	**0.658**	**0.558**	**0.797**	**0.684**	**0.590**	**0.904**	**0.776**	**0.669**

*Note: * = finetuned, M = MOTA, H = HOTA, I = IDF1*

6.2. Discussion

The results of our study demonstrate the efficacy of single-stage detectors, particularly in the context of the challenging dynamics presented by hockey. Notably, the finetuned YOLOv8m model, despite its relatively lower complexity (25.9M parameters, 78.8B FLOPS) (Jocher et al., 2023), significantly surpassing the performance of the more intricate FRCNN with a ResNet50 backbone (166M parameters, 166B FLOPS) (Girshick, 2015) across all evaluated metrics and tracking frameworks. This finding contributes a novel perspective to the ongoing discourse in object detection and tracking, underscoring the potential of less complex models to achieve, and in certain cases, exceed the accuracy of more complex architectures.

A critical factor in the observed performance enhancements is the finetuning of the single-stage detectors. This improvement underscores the importance of model adaptability to specific domain features, particularly when the pre-training dataset (COCO in this case) does not fully encapsulate the target domain's characteristics. We see that despite relatively small improvements in mAP, finetuning single-stage models dramatically impacts tracking metrics.

Our analysis also sheds light on the evolving landscape of online and offline tracking methodologies. The comparable HOTA performance difference between YOLOv8n* with deepSORT and YOLOv8x* with deepSORT (0.099) and YOLOv8x* with deepSORT and YOLOv8x* with Neural Solver (0.092) suggests online tracking combined with relatively large but fast single-stage detector might rival the traditionally superior offline methods. This observation opens up avenues for future research, particularly in exploring the scalability and adaptability of online tracking frameworks when paired with increasingly sophisticated single-stage detection models.

However, the specificity of the MHTD dataset to hockey implies that the findings might not directly translate to other sports or dynamic environments without further validation. Future research should thus aim to explore a broader range of environments and conditions to ascertain the generalizability of these results, as well as conduct more extensive comaprative study such as testing more extensively two-stage detectors on MHTD and single-stage detectors on MOT17. This was left as future work due to time and computation constraints.

The ability to accurately and efficiently track objects in dynamic settings, as evidenced by the performance of the YOLOv8 models, holds promise for developing more sophisticated and responsive analytics tools in various real-world applications.

7. Conclusion

This work presents a comparative study of Multiple Object Tracking (MOT) in ice hockey using the McGill Hockey Tracking Dataset (MHTD) (Yingnan Zhao, 2020) and tracking methods grounded with PGM content, demonstrating the effectiveness of single-stage detectors when integrated with advanced filter-based tracking algorithms such as SORT (Bewley et al., 2016), deepSORT (Wojke et al., 2017) and graph-based ones such as the MOT Neural Solver (Brasó & Leal-Taixé, 2020). We achieved our best result with a fine-tuned YOLOv8x model and the neural solver with a HOTA score of 0.776, showcasing its robustness in a challenging tracking environment. Additionally, the YOLOv8m model stands out for its efficiency and relative simplicity, surpassing typical two-stage detectors such as FRCNN despite having fewer parameters and lower computational complexity. These outcomes highlight the potential of single-stage detectors in sports analytics and beyond, particularly when fine-tuned for specific domains. The study encourages further exploration of MOT applications in diverse and dynamic settings, building upon the foundations laid by our investigations in ice hockey tracking.

References

Acuna, D. Towards real-time detection and tracking of basketball players using deep neural networks. 2017. URL https://api.semanticscholar.org/CorpusID:31248790.

Aharon, S., Louis-Dupont, Ofri Masad, Yurkova, K., Lotem Fridman, Lkdci, Khvedchenya, E., Rubin, R., Bagrov, N., Tymchenko, B., Keren, T., Zhilko, A., and Eran-Deci. Super-gradients, 2021. URL https://zenodo.org/record/7789328.

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, September 2016. doi: 10.1109/icip.2016.7533003. URL http://dx.doi.org/10.1109/ICIP.2016.7533003.

Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

Brasó, G. and Leal-Taixé, L. Learning a neural solver for multiple object tracking, 2020.

Cioppa, A., Giancola, S., Deliege, A., Kang, L., Zhou, X., Cheng, Z., Ghanem, B., and Droogenbroeck, M. V. Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos, 2022.

Cui, Y., Zeng, C., Zhao, X., Yang, Y., Wu, G., and Wang, L. Sportsmot: A large multi-object tracking dataset in multiple sports scenes, 04 2023.

Girshick, R. Fast r-cnn, 2015.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation, 2014.

Jocher, G., Chaurasia, A., and Qiu, J. YOLO by Ultralytics, January 2023. URL https://github.com/ultralytics/ultralytics.

Jonathon Luiten, A. H. Trackeval. https://github.com/JonathonLuiten/TrackEval, 2020.

Kuhn, H. W. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955. doi: https://doi.org/10.1002/nav.3800020109. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800020109.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. Microsoft coco: Common objects in context, 2015.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

Lu, W.-L., Ting, J.-A., Little, J. J., and Murphy, K. P. Learning to track and identify players from broadcast sports videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1704–1716, 2013. doi: 10.1109/TPAMI.2012.242.

Milan, A., Leal-Taixé, L., Reid, I., Roth, S., and Schindler, K. MOT16: A benchmark for multi-object tracking. *arXiv:1603.00831 [cs]*, March 2016. URL http://arxiv.org/abs/1603.00831. arXiv: 1603.00831.

Terven, J. and Cordova-Esparza, D. A comprehensive review of yolo: From yolov1 and beyond, 2023.

Vats, K., Walters, P., Fani, M., Clausi, D. A., and Zelek, J. Player tracking and identification in ice hockey, 2021.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.

Wojke, N., Bewley, A., and Paulus, D. Simple online and realtime tracking with a deep association metric, 2017.

Yingnan Zhao, Zihui Li, K. C. A method for tracking hockey players by exploiting multiple detections and omni-scale appearance features. *Project Report*, 2020.