Article

Object and Event Detection Pipeline for Rink Hockey Games

Jorge Miguel Lopes [1], Luis Paulo Mota [1], Samuel Marques Mota [1], José Manuel Torres [1,2,*], Rui Silva Moreira [1,2], Christophe Soares [1,2], Ivo Pereira [1], Feliz Ribeiro Gouveia [1,2] and Pedro Sobral [1,2]

[1] ISUS Unit, Faculty of Science and Technology, University Fernando Pessoa, 4249-004 Porto, Portugal; 38607@ufp.edu.pt (J.M.L.); 38186@ufp.edu.pt (L.P.M.); 36726@ufp.edu.pt (S.M.M.); rmoreira@ufp.edu.pt (R.S.M.); csoares@ufp.edu.pt (C.S.); ivopereira@ufp.edu.pt (I.P.); fribeiro@ufp.edu.pt (F.R.G.); pmsobral@ufp.edu.pt (P.S.)

[2] Artificial Intelligence and Computer Science Laboratory, LIACC, University of Porto, 4100-000 Porto, Portugal

[*] Correspondence: jtorres@ufp.edu.pt

Abstract: All types of sports are potential application scenarios for automatic and real-time visual object and event detection. In rink hockey, the popular roller skate variant of team hockey, it is of great interest to automatically track player movements, positions, and sticks, and also to make other judgments, such as being able to locate the ball. In this work, we present a real-time pipeline consisting of an object detection model specifically designed for rink hockey games, followed by a knowledge-based event detection module. Even in the presence of occlusions and fast movements, our deep learning object detection model effectively identifies and tracks important visual elements in real time, such as: ball, players, sticks, referees, crowd, goalkeeper, and goal. Using a curated dataset consisting of a collection of rink hockey videos containing 2525 annotated frames, we trained and evaluated the algorithm's performance and compared it to state-of-the-art object detection techniques. Our object detection model, based on YOLOv7, presents a global accuracy of 80% and, according to our results, good performance in terms of accuracy and speed, making it a good choice for rink hockey applications. In our initial tests, the event detection module successfully detected an important event type in rink hockey games, namely, the occurrence of penalties.

Keywords: deep learning; rink hockey; object and event detection; artificial intelligence; computer vision; YOLOv7; team sports

1. Introduction

Rink hockey, also known as quad hockey or roller hockey, is a sport that is both thrilling and fast-paced and is one that captivates players and spectators alike. The game is played on a rink with roller skates, combining elements of ice hockey and roller skating to create an exhilarating game. Historically, this sport has been particularly popular in the following four countries: Portugal, Spain, Italy, and Argentina. As the sport continues to evolve and gain popularity, deep learning and other artificial intelligence techniques have emerged as valuable tools to enhance various aspects of rink hockey, including game visualization, referee work, and training.

In the context of rink hockey, deep learning techniques can be applied to enhance the visualization of the game, thereby facilitating more accurate and detailed analysis of player movements, strategies, and game dynamics. Computer vision algorithms, when deployed in conjunction with strategically placed cameras, can be used to capture high-resolution footage of the rink. This footage is then processed using deep learning models to track the positions and actions of players, the ball, and even referees in real time.

This enhanced game visualization offers several benefits. Firstly, it provides a valuable tool for coaches, allowing them to analyze and strategize more effectively. Such data can be used to identify patterns in player movements, study tactical decisions, and gain

insights into opponent behavior. By analyzing the data collected through deep learning techniques, coaches can make informed decisions and develop customized training programs to enhance their players' performance.

Moreover, the application of deep learning-based game visualization has the potential to transform the role of referees in rink hockey. In this context, referees can play a pivotal role in ensuring fair play and enforcing the rules of the game. However, due to the fast-paced nature of rink hockey, it can be challenging for referees to keep up with all the action and make accurate decisions in real time. The incorporation of deep learning techniques enables referees to receive real-time assistance through computer vision systems. Such systems are capable of tracking players, identifying fouls or infractions, and even providing instant feedback to referees. This reduces the potential for human error and improves the overall fairness of the game.

In addition to their application in the field of game visualization and referee work, these automatic techniques can also significantly impact player training in rink hockey. Training in this sport necessitates the acquisition of a combination of technical abilities, physical conditioning, and tactical comprehension. The utilization of deep learning algorithms enables the creation of personalized and optimized training programs, which are designed to meet the specific needs of individual players. The data collected from game visualization can be utilized to identify strengths and weaknesses, monitor progress, and provide targeted feedback to players. This information can be used by coaches to design training sessions that address specific areas for improvement and enhance overall performance.

Furthermore, the application of deep learning techniques can facilitate the identification of potential injury risks through the analysis of player movements. By monitoring the performance data of players, coaches and medical staff are able to identify signs of fatigue, monitor load management, and make informed decisions to minimize the risk of injuries.

This paper proposes a real-time pipeline consisting of an object detection model specifically designed for rink hockey games, followed by a knowledge-based event detection module. Given the limited availability of developed content relating to this sport, we have decided to create a dataset and test the effectiveness of deep learning and computer vision improvements in the real world. Consequently, object detection, specifically employing the YOLOv7 algorithm, can be used to track the movement of the hockey ball and players on the rink, providing coaches, players, and fans with valuable information. Our paper also examines the advantages of using this approach for object and event detection in rink hockey, as well as its potential impact on the sport itself.

The remainder of the paper is organized as follows. Section 2 presents a discussion of related work. Section 3 presents our approach for the automatic analysis of rink hockey games, while Section 4 presents the evaluation and results. Finally, Section 5 presents the conclusions and outlines future work.

2. Computer Vision Applied to Hockey Sports

Rink hockey and ice hockey are two team sports that share numerous similarities, yet there are also notable differences between them. One of the most significant distinctions between the two sports is the type of equipment employed. Rink hockey is typically played on a dry surface, such as a concrete rink or a sports court, and the playing area is rectangular (Figure 1 right). In contrast, ice hockey is played on an ice rink, which is typically oval in shape (Figure 1 left). The field material is a significant factor influencing the variations in movement patterns and techniques employed by players, as they must adapt to different surfaces and levels of friction and stability. The participants utilize sticks to propel a hard ball/puck into the opposing team's goal, as the game is played on a hard, smooth surface, such as a basketball court or roller rink. The faster pace of play in roller hockey is a consequence of the players' greater mobility on roller skates. Furthermore, the regulations of the two sports diverge in certain respects. For example, in standard conditions on the playing rink, each team in rink hockey is composed of one goalkeeper and

four field players. In ice hockey, each team typically comprises six players. Furthermore, the offside rule is frequently employed in ice hockey, in contrast to rink hockey.

(**a**) Ice hockey players.

(**b**) Rink hockey players.

Figure 1. Ice hockey vs. rink hockey (images: Wikimedia Commons).

The latest innovations in object detection techniques have resulted in considerable increases in both the accuracy and speed of object detection [1]. Convolutional neural networks (CNNs) represent one of the most extensively used techniques for object detection, having obtained state-of-the-art results on diverse object detection benchmarks [2]. One of the primary challenges in rink hockey object detection is the substantial diversity in the appearance of objects due to factors such as lighting conditions, player uniforms, and background noise. Researchers have proposed a number of approaches to address this challenge, including the use of multiple CNNs with distinct architectures [3] and the use of data augmentation techniques to increase the diversity of the training data [4].

In rink hockey, these techniques can be employed for a multitude of purposes, including player tracking, ball tracking, and game analysis. Unfortunately, to date, there has been little work done on object detection in rink hockey. This encompasses both the generation of a dataset and the implementation of deep learning techniques for the associated predictions. However, in relation to ice hockey, there is a plethora of content available, including tracking and identification of players, characteristics of the playing field, and match analysis [5–7].

In [5], a system for automatically tracking and identifying players in NHL streamed videos was proposed. The system comprised three components: player tracking, team identification, and player identification. The authors tested five state-of-the-art tracking algorithms on the hockey player tracking dataset [8–12]. The most effective tracking performance was achieved by utilizing the MOT Neural Solver tracking model [10], which had been re-trained on the hockey dataset with a Multi-Object Tracking Accuracy (MOTA) score of 94.5%. Furthermore, for the purpose of identifying teams, the away team jerseys were grouped into a single class, whereas the home team jerseys were grouped according to their color. In the case of the team identification dataset, a CNN was trained, achieving an accuracy of 97% on the test set. Furthermore, a novel player identification model was presented that employed a temporal one-dimensional convolutional network to identify players from sequences of player bounding boxes. Utilizing the available National Hockey League (NHL) game roster data, the player identification model achieved an accuracy of 83% in player identification.

In [6], the authors developed a cascaded convolutional neural network (CNN) model comprising two phases for the detection of ice hockey players. In the same work, the jersey color of each of the detected players was extracted in order to determine their team affiliations. A filter was incorporated into the proposed model to exclude distracting information, such as the audience and sideline advertising bars, thereby refining the detection of the targeted players. This resulted in accurate detection with a precision of 98.75% and a recall of 94.11% for individual players and an average accuracy of 93.05% for team classification using a dataset of collected images from the 2018 Winter Olympics. The authors argued that the custom-built dataset enabled their player detection model

to achieve the best results when compared to some state-of-the-art approaches such as YOLOv3 [13], Faster R-CNN [14], and CornerNet [15].

Another challenge in rink hockey object detection is the need for real-time performance given the fast-paced nature of the sport and the requirement for object detection updates to be made with regularity. Researchers have proposed a number of solutions to this problem, including the use of lightweight CNN architectures [16] and the implementation of the object detection pipeline on specialized hardware like graphics processing units (GPUs) [17] or field-programmable gate arrays (FPGAs) [18].

In [16], a CNN with three branches and a classification network with four cascades was presented. The cascaded networks were initially trained using labeled image patches and subsequently applied to an entire image by employing a dilation testing strategy. The authors claimed that their approach achieved state-of-the-art accuracy on three types of games (basketball, soccer, and ice hockey) with 1000 fewer parameters than those of CNNs that were adapted from general object detection networks such as Faster R-CNN.

The work presented in [17] proposed an approach to locate the puck, which is the key component of an ice hockey game. One of the principal challenges is the puck's small size. The motion blur caused by the puck's rapid movement, the occlusion between the puck and other objects, and the visual noise (such as the advertisements on the rink) presented additional difficulties for the project. The author proposed a two-stage model for the detection of minute objects. The initial stage involved a two-dimensional CNN that summarized the representation of each frame. In the second stage, the fusion of the per-frame representations was fed into a three-dimensional CNN in order to decode the video's temporal information. The proposed approach achieved a precision of 90.8% and a recall of 86.7%. The F1 score reached 88.7%, which is higher than the performance of YOLOv3 [13] (F1 score = 0.685) and Mask RCNN [19] (F1 score = 0.749).

In [18], a hardware accelerator was presented that implemented a YOLO CNN for real-time object detection with high throughput and power efficiency. The parameters of the implemented YOLO were retrained and also quantized using binary weight and flexible low-bit activation. The binary weight technique enabled the storage of the entire network model in block RAMs of a field-programmable gate array (FPGA). This approach resulted in a reduction in off-chip accesses, thereby achieving a performance boost. All convolutional layers have been pipelined in order to optimize hardware utilization. Additionally, the input images were delivered and processed in a sequential manner to the hardware accelerator. Similarly, the output of the previous layer was transmitted in a sequential manner to the subsequent layer. The intermediate data were fully reused across layers, thereby eliminating memory accesses from the outside. A reduction in DRAM accesses resulted in a reduction in power consumption. Furthermore, the authors argued that the fully parameterized convolutional layers facilitated straightforward scaling of the network. Finally, each convolution layer was mapped to a dedicated block of hardware, thereby enabling the network to outperform other solution designs in terms of energy efficiency and performance.

In [20], a people tracking algorithm tailored for soccer players was presented. The algorithm was designed to cope with a wide range of scenarios, encompassing varying light conditions, high frame rates, and real-time processing. The algorithm employed background subtraction for object segmentation, with the addition of a novel method utilizing pixel energy evaluation to address the challenges of handling moving objects and lighting fluctuations during background modeling. Subsequently, an unsupervised clustering algorithm was employed to classify the detected objects, addressing challenges associated with blob splitting and merging. A stochastic approach, which employs maximum a posteriori probability (MAP), was proposed for the purpose of human tracking. The algorithm initially assessed geometric information regarding blob overlapping and subsequently employed color feature classification to track players and resolve blob merging scenarios. Experimental validations conducted on extensive soccer image sequences across various weather and lighting conditions have demonstrated the effectiveness of the algorithm.

Table 1 provides a summary of the articles discussed in this section, accompanied by a comparison of their respective characteristics. To the best of our knowledge, our approach represents the only solution specifically designed and tested for rink hockey. This sport presents certain distinctive characteristics, and it benefits from having a customized approach. Our approach is suitable for real-time processing, even when utilizing a low-cost processing unit. It is worth noting that, in our tests, we prioritized temporal performance. The YOLOv7 profile was employed as the baseline model for the object detector stage, with a considerable number of parameters (36.9 M parameters, as illustrated in Table 1 of [21]). In our pipeline, we employ object detection as a method to track the semantically meaningful objects in the visual scene. This approach became possible due to the high quality of object detectors that have recently become available.

Table 1. Comparative analysis between different computer vision systems.

	Sport	Dataset	Detection Type	Performance
Vats [5]	Ice Hockey	84 NHL Games	PT (CNN)	83% (Acc.)
Guo [6]	Ice Hockey	6 2018 Winter Olympics Games	PT (CNN)	93.05% (Acc.)
Lu [16]	Ice Hockey Soccer Basketball	2 Ice Hockey Games 2 Soccer Games APIDIS SPIROUDOME	OD (CNN)	94.6% (Acc.)
Yang [17]	Ice Hockey	5 NHL Games	OD (CNN)	88.7% (F1)
Mazzeo [20]	Soccer	CAVIAR	BSOS, PT	94.7% (Acc. [1])
Ours	Rink Hockey	2525 Rink Hockey Game Frames [22]	OD, ED (YOLOv7)	80% (Acc.)

Object Detection (OD); Player tracking (PT); Event Detection (ED); Background Subtraction-based Object Segmentation (BSOS). [1] Combined Acc. in detection of two players in video sequence of CAVIAR dataset according to Table 3 in [20].

3. Automatic Pipeline for Object and Event Detection in Rink Hockey Games

One major contribution of this work is the specification, implementation, and testing of an automatic pipeline capable of detecting and tracking players, the ball, and other in-game objects in real time during fast-paced rink hockey games, as well as some important events, as can be seen in Figure 2.
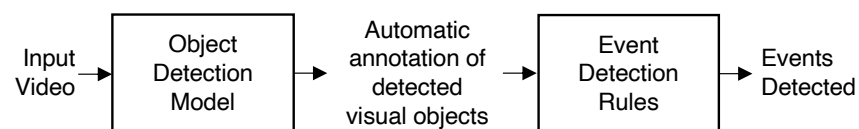

Figure 2. Pipeline for real-time detection of hockey objects and events.

That said, the pipeline was designed to meet the following requirements:

- Object detection capabilities for player tracking, ball tracking, and video analysis.
- Real-time performance for updating object detections regularly during fast-paced games.
- Ability to handle diversity in the appearance of objects due to factors such as lighting conditions, player uniforms, and background noise.
- High accuracy in object detection.
- High speed response in object detection.
- Compatibility with specialized hardware such as GPUs.
- Adaptability to different environments and lighting conditions.

- Ability to detect and represent major events based on the stream of visual objects.

3.1. Dataset Organization

A major contribution of this work is the organization of the dataset. Our rink hockey dataset was annotated with the aid of a web-based computer vision annotation tool that helps to label video and images. The selected solution, Roboflow, is a popular computer vision development platform that facilitates data collection, preprocessing, and model training. This framework allows easy access to public datasets as well as the ability for users to upload their own custom data [23]. Additionally, it supports various annotation formats, including JSON, CSV, and XML.

The dataset was divided into three subsets: training (70%), validation (20%), and testing (10%). The dataset included objects labeled with seven classes: ball, player, stick, referee, crowd, goalkeeper, and goal. The main goal of this annotation process was to produce a dataset that can be used to train object detection models, such as YOLOv7 [24], for transfer learning.

From the total of 2525 annotated frames/images (1764 training, 501 validation, and 260 testing), the image augmentation technique was applied to the training images which, by applying horizontal flipping, cropping with a zoom up to 20%, and changing the brightness changing between −15% (darker) and +15% (brighter), resulted in a total of 5292 images for training. Beyond that, through these and other possible modifications, we artificially increased the size of the dataset. This technique is used not only to increase the size of the dataset and hence the robustness and generalization of the model, but also to try to avoid model overfitting, where a trained model performs better on trained data but poorly on unseen data. In addition to the previously mentioned image augmentation in the training set, we also used the default data augmentation techniques configured in the training process of YOLOv7. Of these techniques, we emphasize mosaic, which allows for the combination of several images to create a single training image with a mosaic appearance [25].
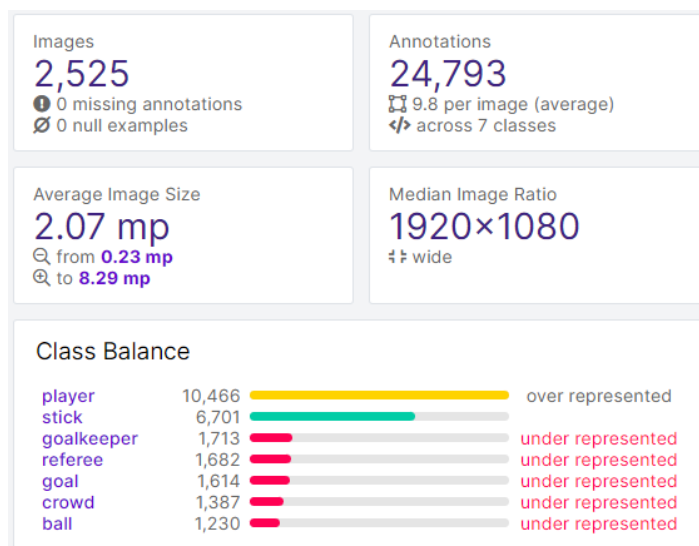

Figure 3. Dataset health check.

As Figure 3 shows, the dataset is slightly unbalanced, with the player and stick classes being the most represented. This is not surprising, since these are the most visible objects in all images.

The resulting dataset is available online [22]. It can definitely be useful for a wide range of applications, including player tracking and video analysis of rink hockey games. Being publicly available, this dataset will serve as a valuable resource for other researchers and practitioners within the field of computer vision and sports analysis.

3.2. Object Detection Model

3.2.1. YOLOv7

YOLO (You Only Look Once) [24] is a state-of-the-art object detection system that is highly valuable in the domain of computer vision. In contrast to traditional object detection systems that run a classifier on various parts of the image independently, YOLO innovatively applies a single neural network to the entire image at once. The image is divided into regions, and then the system predicts bounding boxes and probabilities for each region. As such, YOLO significantly enhances the efficiency and speed of real-time object detection, making it a pivotal asset in a wide range of applications. Specifically in the context of hockey, YOLO can be effectively utilized to track the movements of players and the puck in the arena. This can furnish extensive insights into the analysis of gameplay, performance review of players, and the development of strategic gameplay. It is noteworthy that YOLO is not a standalone tool, but rather, it refers to a type of neural network architecture and algorithm. It is commonly implemented using comprehensive deep learning libraries, such as TensorFlow or PyTorch.

The authors in [21] present another version of YOLO, the YOLOv7. Its base architecture and version are significantly better than prior YOLO versions, such as YOLOv4 and YOLOv5, and it has reached state-of-the-art performance on numerous object detection benchmarks, such as the COCO dataset [26]. The use of a "bag-of-freebies" is one key feature of YOLOv7. This refers to a set of training strategies that do not require the usage of additional computational resources and may thus be utilized to increase the performance of the model without significantly increasing its complexity. Data augmentation, network architecture changes, and multiscale training are examples of these strategies. By using these techniques, this version is able to detect objects at many scales, resulting in better results on object detection tasks while still running in real time on standard hardware. Overall, apart from its high precision object detection abilities, YOLOv7 is also known for its efficiency and speed, delivering outstanding performance without losing speed or simplicity, making it a valuable tool for a wide range of applications. Typically a visual object detection solution, such as YOLOv7, offers several model options with increasing degrees of complexity, that is, with an increasing number of parameters. In the case of YOLOv7, the models range from YOLOv7-tiny to YOLOv7-E6E. There is a trade-off between speed and accuracy. The simpler, the faster and the more complex, the more accurate. Figure 4 illustrates a comparison between YOLOv7 and other important visual object detectors, showing its better performance. These characteristics make YOLOv7 an appropriate base model for a real-time application, such as the solution proposed in this article, and justify its use in the tests we carried out.
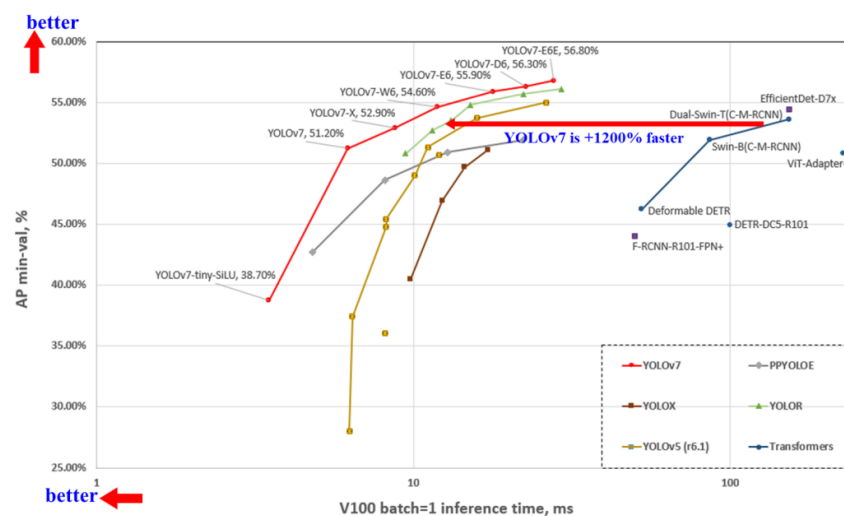

Figure 4. Comparing YOLOv7 with other object detectors (source: [21]).

3.2.2. Object Detection for Rink Hockey

In this section, we present the technical details of the developed (https://github.com/LuisMota1999/ProjetoVC, accessed on 19 January 2023) object detection system for rink hockey utilizing YOLOv7. To facilitate the training of the model, we constructed a dataset composed of various rink hockey scenarios, which were carefully annotated to include different perspectives, lighting conditions, data augmentation, and object scales. The YOLOv7 architecture was implemented using the PyTorch framework [27], and the pre-trained weights were fine-tuned on the dataset we were training. From the several YOLOv7 pre-trained weight ranges available (see Figure 4) we selected, in our tests, the *YOLOv7, AP min-val* = 51.20%, a good compromise between complexity and real-time performance (https://github.com/WongKinYiu/yolov7/tree/main, accessed on 19 January 2023). Our experiments demonstrated that the model achieved high accuracy in detecting rink hockey objects such as the crowd, goalkeeper, and players. For real-time object detection, we used multiple online videos from professional rink hockey games [28–37] to acquire the feed and processed the frames using the trained YOLOv7 model. Our results demonstrate the ability of our model to detect visual objects in the video with low latency, thereby making it suitable for real-time applications. Additionally, we leveraged CUDA and cuDNN [38] for GPU acceleration, which significantly improved the processing speed of the model. The conjunction with various technologies such as CUDA, OpenCV, and PyTorch has enabled the development of a robust and efficient object detection system for rink hockey.

3.3. Event Detection Rules Module

Event detection follows object detection in the pipeline shown in Figure 2. The proposed specification of an event detection module, whose input is a timeline stream of detected visual objects, or the output of the object detection model, can comprise up to four stages, as follows:

1. Filtering: due to the noisy nature of the object detection model, a filtering operator, such as a windowed spatio-temporal median filter, can be used as a pre-processing stage before the application of a rule-based detector.
2. Model-based Rule Detection: rule-based system that can detect predefined types of events.
3. Representation: the language or taxonomy used to represent the events. Event calculus is a formal method of event representation [39].
4. Revision: can be used to revise or update represented event beliefs due to new information available in the stream of detected visual objects.

This module follows the object detection module and, consequently, depends on it according to the directed acyclic graph (DAG) presented in Figure 2. In the preliminary tests performed, a proof-of-concept approach was adopted for this module, and the detection of a unique type of event was considered for this article.

4. Experiments and Evaluation

4.1. Object Detection

In our experiments, we trained a YOLOv7-based model for object detection using footage acquired, essentially, from professional rink hockey games. We performed both qualitative and quantitative evaluations to assess the performance of this model.

The evaluation process consisted of applying YOLOv7 to detect objects of interest in multiple video frames from rink hockey games [40,41] and comparing the results to manually annotated ground truth labels. We used precision (p), recall (r), and F1 score (F_1), the harmonic mean of precision and recall, as the three quantitative metrics to evaluate the performance of the system, according to the following equations:

$$p = \frac{TP}{TP + FP} \tag{1}$$

$$r = \frac{TP}{TP + FN} \tag{2}$$

$$F_1 = 2 \times \frac{p \times r}{p + r}. \tag{3}$$

The model was trained using 100 epochs and a batch size of 32, with the system being implemented on a Windows desktop with a Zotac Gaming GeForce RTX 3090 Trinity OC 24 GB GDDR6X GPU and 64 GB of RAM, using the Darknet framework in Python 3.7.0.

When testing on frames not seen in the training phase (present in the test subset), the overall performance was quite good (around 80%).

The results of the evaluation are presented in the following figures. Figure 5 presents a multi-frame inferred plot of the test subset, showing the labels that the trained model detected. Figure 6 represents the trade-off between precision and recall in the classification model. As can be seen from the graph, the precision–recall values are higher for the player, goalkeeper, and goal classes, as we had more annotated images for these classes. However, the precision–recall for the ball is lower, as we have fewer annotated images for this class.

Figure 5. Multi-frame (4 × 4 = 16 frames) inferred plot of the test subset. The upper-left frame is at the coordinate (row, col) = (1, 1).
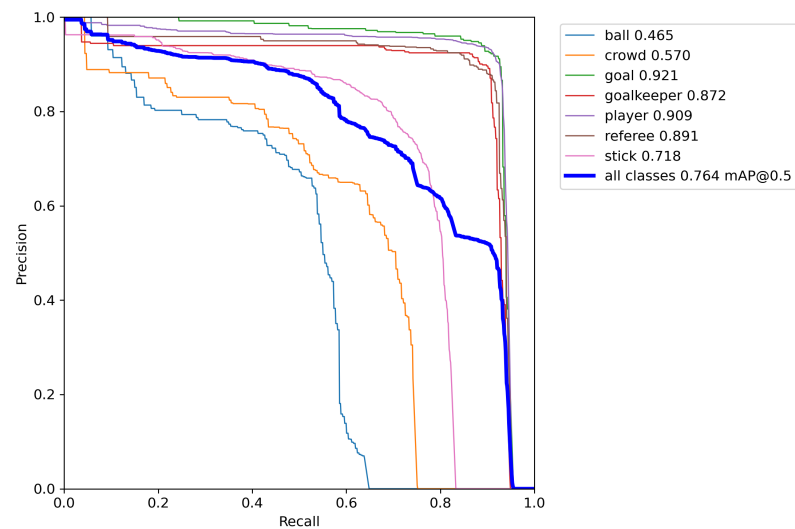

Figure 6. Precision–recall curves.

It is important to note that, in our evaluation, we only considered detections with an objective confidence score above 0.3. This threshold was chosen to balance the trade-off between detecting more objects while maintaining a slightly higher precision.

Also, as seen in Figure 6, our trained model got a score of 0.764 mAP@0.5 and, in this context, 0.764 represents the mAP score, i.e., the mean average precision (the average of AP over all detected classes, calculated for each class separately) ranging between 0 and 1. The parameter 0.5 represents the IoU (intersection over union) selected threshold; in other words, the model is able to detect objects with an overlap of at least 50% with the ground truth. Having said that, in the context of this work, a mAP score of 0.764 is a fairly good score for object detection models.

In follow-up to what was previously said, we can also use the F_1 score to evaluate the model performance. The F_1 score is a metric that is calculated as the harmonic mean of precision and recall. This metric also has a range of 0 to 1, with 1 representing perfect precision and recall and 0 representing the worst possible value. Beyond that, precision reflects the number of samples that were predicted as positive and are actually positive (ground truth), and recall measures the number of actual positive samples that were predicted correctly as positive. As shown in the Figure 7, the model has a maximum F_1 score of 0.79 at 0.240. Analyzing this graph reveals that there is a large plateau with $F_1 > 0.75$ approximately parallel to the confidence axis, which can be considered a very good indicator of the model behavior.
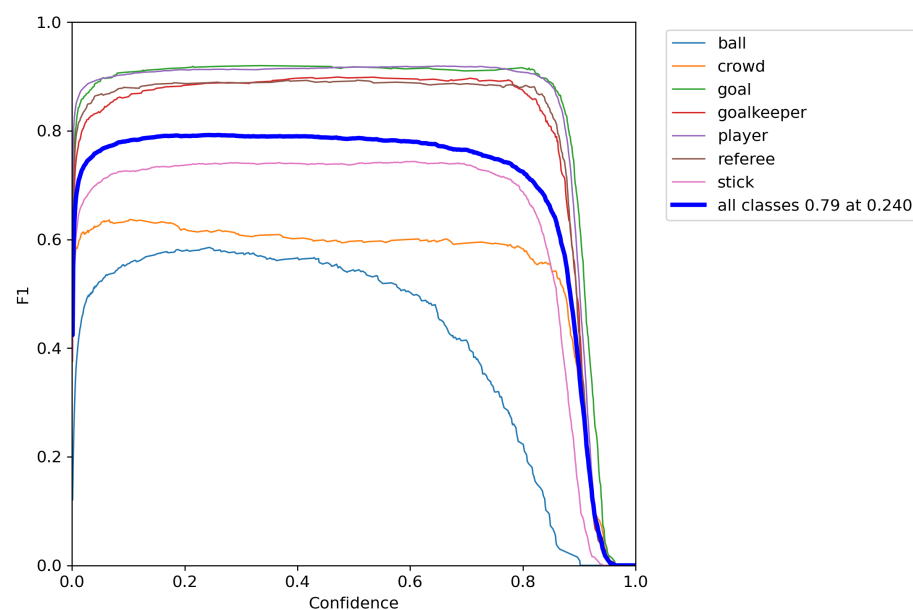

Figure 7. F_1 Curve.

The metrics F_1 score, precision, and recall are all derived from the values of the confusion matrix, which is another useful tool for understanding the performance of a classification model, since it shows where the model is making mistakes and how it performs on different classes. In this case, as can be seen in the normalized confusion matrix obtained in Figure 8, when evaluating the test subset. In other words, on the images that were not seen by the model while training, we got an accuracy of 80%. This figure also provides the information that the ball and crowd classes were the ones with the worst results.
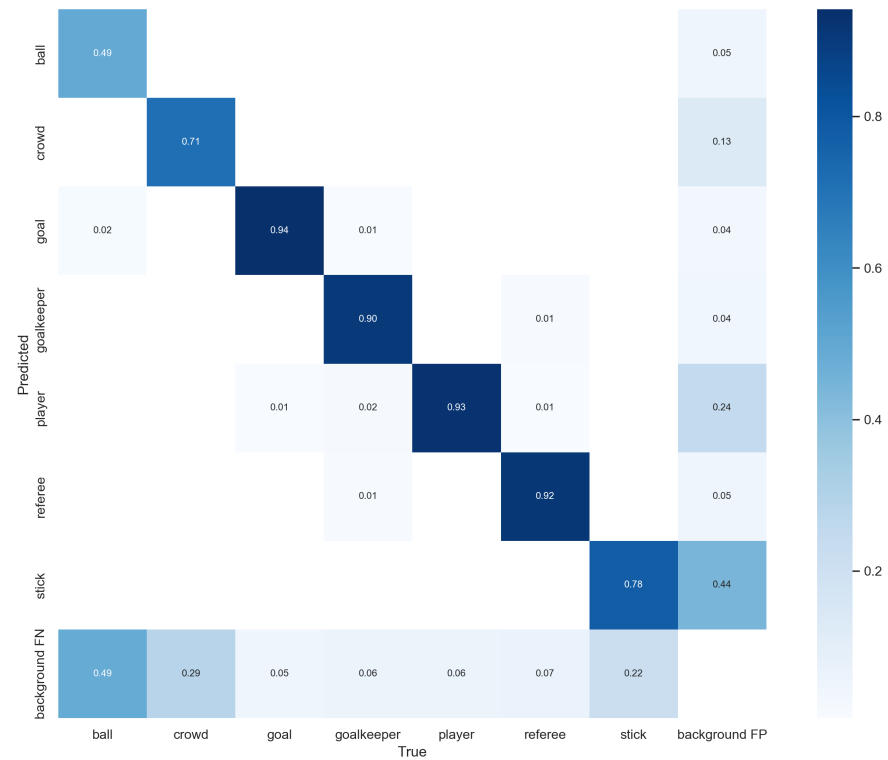
Figure 8. Confusion matrix.

4.2. Event Detection

For the rule production system used for event detection, one rule, illustrated in Algorithm 1, was tested to detect the occurrence of penalties or direct free hits, according to the existing official rink hockey regulations [42].

Algorithm 1 Function Implementing Rule for Direct Free Hit or Penalty Detection

---

inputs: *current_time*, the current time; Δt, temporal window duration for event searching
returns: *EVENT_DIRECT_FREE_HIT_OR_PENALTY* if detects the event or *FALSE* if not
   function DIRECT_FREE_HIT_OR_PENALTY_RULE(*current_time*, Δt)
      $OL \leftarrow$ UPDATE_OBJECTS_DETECTED() $\triangleright$ Gets the automatically annotated detected list of visual objects, updated and timestamped
      $t_1 \leftarrow current\_time - \Delta t$
      $t_2 \leftarrow current\_time$
      if $T(isReferee(OL) \wedge isGoal(OL) \wedge isGoalkeeper(OL) \wedge countPlayer(OL) == TRUE, t_1, t_2)$ **then**
         return *EVENT_DIRECT_FREE_HIT_OR_PENALTY*
      else
         return *FALSE*
      end if
   end function

This rule considers several predicates. The predicate $T(expression, t_1, t_2)$ is true if *expression* is true for the temporal interval between t_1 and t_2. Remaining predicates are auto-explicated and have, as their argument, the updated list *LO* of timestamped and automatically detected visual objects. Predicates starting with the prefix *is* mean the presence of at least one visual object in that frame for that class. Predicates starting with the prefix *count* return the total number of visual objects in that frame for that class.

According to the current regulations [42], the player taking the penalty or the direct free hit has, after the indication of the main referee, a maximum of five seconds to start

the execution with the ball stopped. Thus, in this rule, the temporal window duration, $\Delta t = (t_2 - t_1)$, should be appropriately chosen. In our qualitative analysis, we set $\Delta t = 4$ s. In the quantitative preliminary tests, we annotated a total of 23 penalty images: 11 from the training set; 7 from the validation set; and 5 from the test set. Figure 9 illustrates an example of a penalty detected by our rule where six distinct visual object instances are considered:

- Predicate $isReferee(LO)$ is satisfied by the visual object of class referee;
- Predicate $isGoal(LO)$ is satisfied by the visual object of class goal;
- Predicate $isGoalkeeper(LO)$ is satisfied by the visual object of class goalkeeper;
- Predicate $countPlayer(LO) == 1$ is satisfied by the presence of exactly one instance of class player per frame;
- Predicate $T(f, t_1, t_2)$ is satisfied by the truthfulness of f, where $f = isReferee(LO) \wedge isGoal(LO) \wedge isGoalkeeper(LO) \wedge countPlayer(LO) == 1$ in at least one frame within the time interval $[t_1, t_2]$ and according to the test criteria adopted.

The penalty and direct free hit sanction executions are probably the most important ones in rink hockey. Thus, their detection is rather important as an event. Figure 5 (row, col) = (3, 4) illustrates an example of a penalty event during the game that was correctly detected by this rule.

Figure 9. Example of event detection of a penalty (sources: [22,32]).

5. Conclusions

The purpose of our investigation is to advance the field of research in object and event detection in rink hockey sports. This has the potential to significantly enhance the analysis and comprehension of this sport, which is characterized by fast-paced gameplay.

The use of object detection techniques has been instrumental in the tracking of objects in rink hockey videos. Nevertheless, the diversity in the appearance of objects and the need for real-time performance remain significant challenges in this field.

This research employed a state-of-the-art object detection methodology for rink hockey, utilizing the YOLOv7 algorithm in conjunction with a knowledge-based event detection module. This approach has demonstrated good object detection performance in terms of precision, even in the presence of occlusions and rapid motion.

For future work, it is essential to enlarge the dataset with more annotated data in order to balance the dataset and compare it with other algorithms. It is important to keep testing and comparing with other methods and techniques. Finally, it should be noted that this work could be practically applied in collaboration with roller hockey clubs and hockey federations. This would enable the development and testing of this approach, as well as the continuous evolution of this sport.

References

1. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
2. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
3. Srivastava, S.; Divekar, A.V.; Anilkumar, C.; Naik, I.; Kulkarni, V.; Pattabiraman, V. Comparative analysis of deep learning image detection algorithms. *J. Big Data* **2021**, *8*, 66. [CrossRef]
4. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
5. Vats, K.; Walters, P.; Fani, M.; Clausi, D.A.; Zelek, J.S. Player tracking and identification in ice hockey. *Expert Syst. Appl.* **2023**, *213*, 119250. [CrossRef]
6. Guo, T.; Tao, K.; Hu, Q.; Shen, Y. Detection of ice hockey players and teams via a two-phase cascaded CNN model. *IEEE Access* **2020**, *8*, 195062–195073. [CrossRef]
7. Sousa, T.; Sarmento, H.; Harper, L.D.; Valente-Dos-Santos, J.; Vaz, V. Match analysis in rink hockey: A systematic review. *Hum. Mov.* **2022**, *23*, 33–48. [CrossRef]
8. Bergmann, P.; Meinhardt, T.; Leal-Taixe, L. Tracking without bells and whistles. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 941–951.
9. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: New York, NY, USA, 2016; pp. 3464–3468.
10. Brasó, G.; Leal-Taixé, L. Learning a neural solver for multiple object tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6247–6257.
11. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017 ; IEEE: New York, NY, USA, 2017; pp. 3645–3649.
12. Zhang, Y.; Wang, C.; Wang, X.; Zeng, W.; Liu, W. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *Int. J. Comput. Vis.* **2021**, *129*, 3069–3087. [CrossRef]
13. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [CrossRef] [PubMed]
15. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
16. Lu, K.; Chen, J.; Little, J.J.; He, H. Lightweight convolutional neural networks for player detection and classification. *Comput. Vis. Image Underst.* **2018**, *172*, 77–87. [CrossRef]
17. Yang, X. Where Is the Puck? Tiny and Fast-Moving Object Detection in Videos. Master's Thesis, McGill University, Montréal, QC, Canada, 2021.
18. Nguyen, D.T.; Nguyen, T.N.; Kim, H.; Lee, H.J. A high-throughput and power-efficient FPGA implementation of YOLO CNN for object detection. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2019**, *27*, 1861–1873. [CrossRef]
19. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
20. Mazzeo, P.L.; Spagnolo, P.; Leo, M.; D'Orazio, T. Visual Players Detection and Tracking in Soccer Matches. In Proceedings of the 2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, Santa Fe, NM, USA, 1–3 September 2008; pp. 326–333. [CrossRef]
21. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.

22. Lopes, J.M.; Mota, L.P.; Mota, S.M.; Torres, J.M. Roller Hockey Dataset. 2024. Available online: https://universe.roboflow.com/visao-computacional/roller-hockey (accessed on 1 September 2023).
23. Roboflow. Roboflow Universe. 2023. Available online: https://universe.roboflow.com (accessed on 1 May 2023).
24. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. CrossRef]
25. Chris Hughes and Bernat Puig Camps. YOLOv7: A Deep Dive into the Current State-of-the-Art for Object Detection, Everything You Need to Know to Use YOLOv7 in Custom Training Scripts. 2022. Available online: https://towardsdatascience.com/yolov7-a-deep-dive-into-the-current-state-of-the-art-for-object-detection-ce3ffedeeaeb (accessed on 1 September 2023).
26. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 740–755.
27. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the NIPS-W, 2017, Long Beach, CA, USA, 4–9 December 2017.
28. Porto, F. Vitória ÉPICA do Hóquei em Patins do FC Porto (Resumo). 2022. Available online: https://youtu.be/JRe0iIi55Ig (accessed on 19 January 2023).
29. D'hoquei, S. Resum del GSH Trissino vs. Credit Agricole Sarzana. 2020. Available online: https://youtu.be/k3cFN_Y2prc (accessed on 19 January 2023).
30. D'hoquei, S. Highlights Portugal vs. France. 2022. Available online: https://youtu.be/Jbjut3NvCGg (accessed on 19 January 2023).
31. TV, F. Hóquei em Patins: Física × CRIAR-T—25 de Abril 18h30. 2022. Available online: https://youtu.be/YZMHSuskcCg (accessed on 19 January 2023).
32. Euro Rink Hockey TV. FC Porto × SL Benfica. 2018. Available online: https://youtu.be/jnFRrurson0 (accessed on 19 January 2023).
33. Euro Rink Hockey TV. FC Porto (PT) × Sporting CP (PT). 2018. Available online: https://youtu.be/uIsaU7ME8h4 (accessed on 19 January 2023).
34. FC Barcelona. [ESP] LIGA EUROPEA Hockey Patines: FC Barcelona Lassa—FC Oporto (3-1). 2017. Available online: https://youtu.be/zGXZaarw5BE (accessed on 19 January 2023).
35. Euro Rink Hockey TV. Euroleague—Barça (SP) × SL Benfica (PT). 2020. Available online: https://youtu.be/yTe1Hv9OV-w (accessed on 19 January 2023).
36. Euro Rink Hockey TV. BENFICA-BARCELONA, 1st Semi-Final of Rink Hockey Euroleague 2015–2016. Played at Pavilhão. 2016. Available online: https://youtu.be/hTNZm1CGww4 (accessed on 19 January 2023).
37. Euro Rink Hockey TV. Match #249—Spain × Portugal [HD]. 2021. Available online: https://youtu.be/fKDWzskg2pE (accessed on 19 January 2023).
38. NVIDIA. NVIDIA CUDA Deep Neural Network Library (cuDNN). 2023. Available online: https://developer.nvidia.com/cudnn (accessed on 1 September 2023).
39. Shanahan, M. The Event Calculus Explained. *Artif. Intell. Lnai* **2000**, *1600*, 409–430. [CrossRef]
40. D'hoquei, S. Highlights SL Benfica vs. Sporting CP. 2022. Available online: https://youtu.be/CS4P8hkM8vk (accessed on 19 January 2023).
41. Euro Rink Hockey TV. Extended Highlights—Match #243—France × Portugal [HD]. 2022. Available online: https://youtu.be/S0fSwSclAsE (accessed on 19 January 2023).
42. Rink Hockey Technical Commission. *Rink Hockey Official Regulation*; Technical Report; World Skate: Montpellier, France, 2020.