



## Article

# Enhancing Handball Analytics with Computer Vision and Machine Learning: An Exploratory Experiment

Mostafa Farahat <sup>1</sup>, Hassan Soubra <sup>2,\*</sup>, Donatien Koulla Moulla <sup>3,\*</sup> and Alain Abran <sup>4</sup>

- <sup>1</sup> Computational Engineering: Medical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany; farahatt.mostafa@gmail.com
- <sup>2</sup> Department of Computer Engineering, Ecole Centrale d'Electronique-ECE Lyon, 24 rue Salomon Reinach, 69007 Lyon, France
- <sup>3</sup> Centre for Augmented Intelligence and Data Science, School of Computing, University of South Africa, Johannesburg 1709, South Africa
- <sup>4</sup> Department of Software Engineering and Information Technology, Ecole de Technologie Supérieure, Montréal, QC H3C 1K3, Canada; alain.abran@etsmtl.ca
- \* Correspondence: hsoubra@ece.fr (H.S.); moulldk@unisa.ac.za (D.K.M.)

## Abstract

Recent advancements in artificial intelligence (AI) have strengthened the interaction between sports and digital technologies. However, unlike widely studied sports such as football and basketball, handball has received limited attention from the scientific community, despite its fast-paced nature and strategic importance. This study focuses on object detection in handball and targets key entities, such as players, referees, goalkeepers, and the ball. A comprehensive dataset was created through a collaborative annotation process, consisting of annotated images extracted from real handball games. The YOLOv8 model was then trained and evaluated on this dataset to assess its effectiveness in entity recognition. The proposed approach achieved an object detection accuracy of 86.8% on a relatively small held-out test set, providing an indicative first benchmark for the application of state-of-the-art machine learning models to handball. To the best of our knowledge, the dataset generated in this study is the first comprehensive collection of annotated handball images, providing a valuable resource for further research. By bridging sports analytics and computer vision, this study contributes to the advancement of performance assessment in handball. These exploratory results suggest potential directions for future real-time systems and practical applications, such as improved understanding of player performance, team dynamics, and strategic decision-making.

**Keywords:** computer vision; machine learning; sports technology; internet of things; handball



Academic Editors: Rui Yu, Sooyeon Lee, Syed Masum Billah and John M. Carroll

Received: 16 February 2026  
Revised: 7 April 2026  
Accepted: 8 April 2026  
Published: 10 April 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

In recent years, the convergence of sports and digital technology has opened new opportunities for innovation, particularly through the integration of the Internet of Things (IoT), machine learning (ML), and computer vision. IoT refers to an interconnected network of physical devices embedded with sensors, software, and connectivity, enabling them to collect and exchange data [1]. IoT devices, such as connected cameras, generate continuous streams of visual data that enable advanced analyses of sporting events.

ML, a subset of artificial intelligence (AI), provides algorithms that can automatically learn and make predictions from data without explicit programming. By leveraging vast

amounts of data, ML algorithms can uncover hidden patterns and provide valuable insights. In computer vision applications, ML plays a crucial role in extracting features from images and videos to accomplish visual tasks, such as object detection and recognition. Computer vision systems can analyze this data to identify key entities, such as players, referees, and equipment, with accuracy in real time.

Despite progress in sports technology, there is still a significant lack of publicly available datasets tailored specifically for handball. This sport is characterized by speed and strategic complexity. The lack of resources limits the development of automated systems for performance analysis, tactical evaluation, and refereeing support. Furthermore, existing challenges, such as low camera frame rates and visual complexity in crowded scenes (where players cluster tightly near the goal and mutual occlusion reduces detection reliability), complicate accurate player and ball detection.

This study addresses these gaps by creating a comprehensive dataset of annotated handball images and developing tools to facilitate the automated analysis of handball matches. The novelty of this work lies in its focus on handball, a relatively underexplored sport in sports analytics, and in providing open resources for future research and development.

The main contributions of this study are:

1. The creation and release of a publicly available annotated image dataset for handball.
2. The development of tools for automated analysis of handball matches.
3. A discussion of key technological challenges in automated handball analysis.
4. A foundation for future research on handball analytics and sports technology.

The practical significance of this study extends to multiple stakeholders. Coaches can use the system to identify tactical patterns and enhance decision-making. Players can review individual performance to recognize strengths and weaknesses. Referees can benefit from support in detecting fouls and rule violations. The generation of real-time statistics and visualizations can enhance the viewing experience for fans and broadcasters. More broadly, the motivation of this study reflects the recognition that human capabilities in sports performance have reached a plateau, which has fueled increasing interest in the potential of technology to elevate training and analysis [2]. The rapid growth of sports technology research in recent years, as indicated by [3], further underscores this trend.

The remainder of this paper is organized as follows. Section 2 explores the background literature. The background for sports and handball in particular is discussed, along with the key technologies that support the study. Section 3 presents and discusses the approach used in this study, along with an overview of the driving factors behind our decisions. Section 4 explains our implementation of the proposed approach and the experimental setup. Section 5 discusses the results. Section 6 concludes with a synthesis of the main results, identified limitations, and perspectives for future work.

## 2. Background and Literature Review

This section provides the background for the study across three themes: handball as a sport and its current analytical landscape; the key enabling technologies (Internet of Things, machine learning, and computer vision) that underpin automated sports analysis; and the existing literature on automated analysis in handball and related sports, identifying the gaps this study addresses.

### 2.1. Sports and Handball

In a world where sedentary lifestyles have become the norm, sports continue to hold a special place in human society, offering physical well-being, joy, and excitement. The integration of digital technology in sports has enabled significant advances in athlete

training, performance monitoring, and spectator engagement. Smart sport training (SST), which combines wearables, sensors, Internet of Things (IoT) devices, and intelligent data analysis tools, is increasingly used to optimize athlete training while minimizing physical strain [3]. Beyond athletes, technological innovations such as augmented reality (AR) and real-time tracking enhance the experience of spectators and broadcasters.

Although developments in SST have been widely applied in sports such as football and basketball, handball (a sport characterized by speed, intensity, and tactical complexity) remains relatively underexplored in the context of SST. Existing technologies, such as sensor-based tracking and video-assisted replay systems, have demonstrated their value in enhancing decision-making and entertainment in other team sports. However, similar applications in handball are limited.

Handball has a long history, dating back to early versions in ancient Rome and the Middle Ages, with the first international games organized in the 1920s and 1930s [4]. Today, the sport has more than seven million registered players worldwide, engaging in fast-paced matches of two 30 min halves with teams of six field players and one goalkeeper. Despite its popularity, performance analysis in handball still relies heavily on manual observation. These traditional methods are limited by human error, the high tempo of play, and the restricted scope of data collection. Consequently, automation has emerged as a crucial step for advancing data-driven performance evaluation, tactical analysis, and refereeing support.

## 2.2. Key Technologies for Handball Analysis Automation

Recent advances in sports technology suggest that automated handball analysis can be achieved by combining IoT devices, ML, and computer vision techniques. These technologies address challenges such as high game speed, frequent ball occlusion, and the need for real-time insights. The following technologies can be applied to handball analysis:

- (a) IoT-based approaches:
  - Sensor-based solutions: Companies such as Kinexon employ Ultra-Wideband (UWB) and Inertial Measurement Unit (IMU) sensors for tracking player movements [5]. The effectiveness of such systems in sports analytics has been demonstrated by Foina et al. [6].
  - Camera-based solutions: Event-based cameras such as Prophesee [7], multi-camera tracking systems [8], and stereo cameras, such as Zed2 [9], provide rich visual data for match analysis.
- (b) Machine Learning (ML): ML techniques enable automatic learning from large datasets to recognize patterns, predict player performance, and assess injury risks. In the context of handball, ML can be applied to model player trajectories, predict goal-scoring probabilities, and optimize tactical strategies.
- (c) Computer Vision: Computer vision methods, often combined with ML, analyze images and videos to detect and track key entities such as players, referees, and the ball. Techniques proposed by Schrapf et al. [8] illustrate the potential of computer vision in approximating trajectories and identifying game events with high accuracy.
- (d) Integration of Technologies: The fusion of IoT, ML, and computer vision creates a comprehensive framework for automated handball analysis. Camera-based systems provide visual streams, sensor data capture ball velocity and movement, and ML models synthesize these inputs to generate actionable insights. Together, these technologies enable real-time decision support for coaches and referees, tactical evaluations for teams, and enhanced experiences for fans and broadcasters.

### 2.3. Literature Review

#### 2.3.1. Existing Studies in Handball and Related Sports

In contrast to widely studied sports, such as soccer, research in automated handball analysis remains limited. Nevertheless, several approaches from related sports provide valuable insights that can be adapted. Foina et al. [6] employed Radio Frequency Identification (RFID) technology, achieving a positional accuracy of 1–2 m, although this precision remains insufficient for the fast dynamics of handball. Computer vision approaches, such as those demonstrated by Schrapf et al. [8], enable more accurate player tracking and trajectory estimation. Vallance et al. [10] explored internal factors for injury prediction, offering a perspective relevant for performance monitoring. In the context of action recognition, Host et al. [11] applied Long Short-Term Memory (LSTM) networks, illustrating the potential of deep learning for event detection in sports videos. Finally, the dataset and taxonomy introduced by Biermann et al. [12] represent an important step toward standardized resources, which are essential for further research in this field.

#### 2.3.2. Insights and Limitations

Existing studies have underscored several important insights. First, computer vision has shown promising results for player tracking and activity recognition [8,11]. Second, the predictive modeling of injuries [10] highlights the potential of data-driven methods beyond performance analysis. However, most existing approaches focus on sports other than handball, and the results are not directly transferable because of differences in gameplay speed, rules, and tactical complexity. Furthermore, while Biermann et al. [12] provided a dataset and taxonomy, the field continues to suffer from a lack of publicly available data, which hinders reproducibility and large-scale benchmarking.

#### 2.3.3. Gaps and Challenges

Three recurring challenges have been identified from the literature:

1. Lack of datasets: Publicly available datasets for handball are rare, limiting the development of robust and generalizable models [12].
2. Disconnect between research and practice: A gap persists between researchers and practitioners, as highlighted in interviews with coaches and players. This reduces the practical implementation of technological solutions.
3. Technical limitations: Sensor-based systems face issues such as missing data and noise [6], whereas computer vision approaches struggle with occlusion, crowded scenes, and the rapid pace of handball matches [8]. Researchers often need to design custom tools, which is time-consuming and error-prone.

In summary, existing research demonstrates the potential of IoT, computer vision, and machine learning for sports analytics; however, significant gaps remain in the context of handball. The field lacks large-scale annotated datasets, practical tools, and reliable systems capable of handling the dynamics of sport. This study aims to address these challenges by focusing on dataset creation, tool development, and the application of stereo camera technology to enhance automation and effectiveness in handball analysis. While the broader context of this work situates automated handball analysis within a vision-enhanced IoT ecosystem, the technical contributions of this study focus specifically on computer vision and machine learning: the creation of an annotated handball image dataset and the benchmarking of state-of-the-art object detection models. IoT hardware considerations, including camera selection and multi-sensor integration, are discussed as contextual motivators and directions for future system deployment rather than as core implemented components.

### 3. Approach

To address some of the limitations discussed above, this research focused on three main parts that aimed to improve the quality and applicability of sports analytics. These parts were:

1. Reducing the gap between coaches/players and researchers by conducting interviews with both groups to understand their needs, preferences, and challenges in using sports analytics.
2. Data gathering and training of a machine learning model by collecting video data, extracting frames from them, and applying computer vision and deep learning techniques to extract relevant features and patterns.
3. Brief review of camera technologies by exploring the current state-of-the-art and future trends in camera systems and sensors for sports analytics.

These three parts formed the basis for the mixed methods approach adopted in this study, which combined qualitative and quantitative data to answer the research question.

This study adopts a mixed-methods approach to advance sports analytics for handball. The methodology integrates:

1. qualitative interviews with coaches and players to capture practical needs and challenges;
2. dataset development and training of computer vision models for object detection and;
3. exploration of camera technologies to support depth-based analysis.

Prior to the interviews, all participants were informed about the study's purpose, and verbal informed consent was obtained from them prior to their inclusion in the study. This approach was selected because of the minimal-risk nature of the study. All procedures were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of South Africa (reference number # 2022/CSET/SOC/008).

#### 3.1. Coach and Player Interviews

To bridge the gap between researchers and practitioners, we conducted semi-structured interviews with three professional handball coaches, one basketball player, and one handball analyst. The inclusion of a basketball player provided cross-sport insights for comparison. Participants were selected through purposive sampling based on their professional experience. The coaching and analyst perspectives were prioritized as they represent the primary intended end-users of a performance analytics system. Engaging handball players directly is identified as a priority for future work.

Interviews were guided by open-ended questions focusing on:

- current practices in performance analysis;
- challenges in data collection and interpretation, and;
- desired analytics tools and outputs.

Each interview lasted approximately 45 min and was audio-recorded with consent. The transcripts were analyzed using thematic coding, following Braun and Clarke's six-step methodology, to identify recurring themes.

Key findings included:

- Data collection: Coaches reported that manual data gathering is time-consuming and often discourages systematic analysis. Nevertheless, imperfect data were perceived as more valuable than no data.
- Psychological aspects: Coaches emphasized that player psychology strongly influences performance, requiring analytics that account for contextual and personal factors.
- Team-level over individual metrics: Data on team formations, collective effort, and tactical execution were considered more useful than individual statistics.

- Data quality gap: Compared with sports such as basketball (NBA), handball analytics suffer from limited reliability and poor-quality datasets.

These findings directly shaped several technical decisions in this study:

1. Multi-class annotation design: coaches’ emphasis on the need to simultaneously observe all match participants motivated annotating four entity classes (player, goalkeeper, referee, handball) rather than a simpler single-class approach.
2. Open dataset priority: interviewees’ identification of data scarcity as a key barrier.
3. Transparent per-class reporting: the practitioners’ pragmatic view that imperfect data are more valuable than no data informed our decision to report per-class true positive rates honestly, including the low handball detection rate.
4. Acknowledged gap: The focus on team-level and collective metrics highlighted that frame-level object detection is necessary but not a sufficient step; this gap is explicitly framed as a known limitation and future research direction in Sections 5.5 and 6.2.

### 3.2. Dataset Development and Computer Vision Models

This section describes how we gathered the dataset and set up the infrastructure to use the data for ML models. The quantitative component of this study involved two main activities: the creation of a manually annotated dataset of handball match images, supported by custom tooling developed to assist annotators; and the training and evaluation of state-of-the-art object detection models on this dataset. Our approach aimed to reduce the per-annotator workload through tool-assisted workflows and the crowdsourced distribution of annotation tasks, rather than replacing manual annotation entirely. The system comprised three stages: data extraction, annotation, and supervised learning. Figure 1 shows the system architecture. The following subsections explain each component and the techniques used.

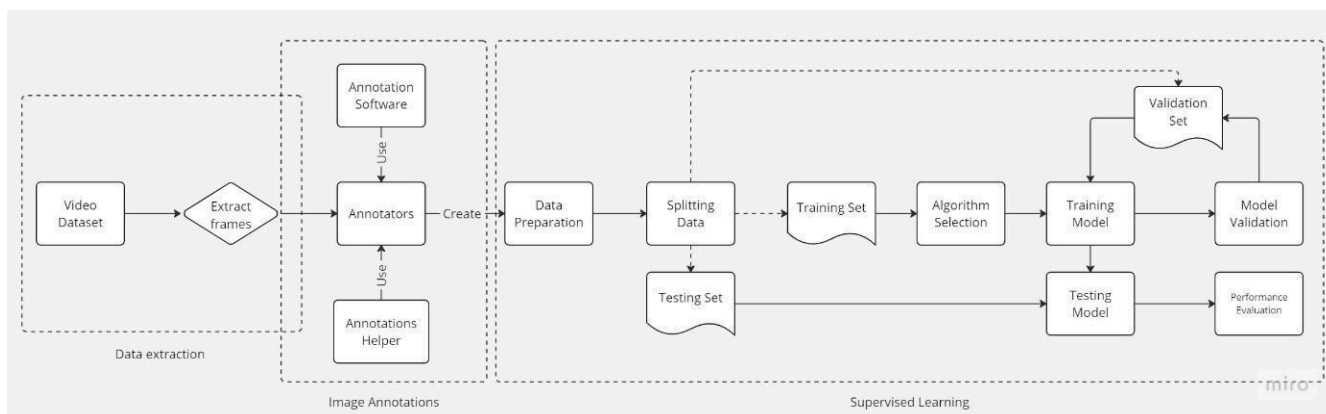
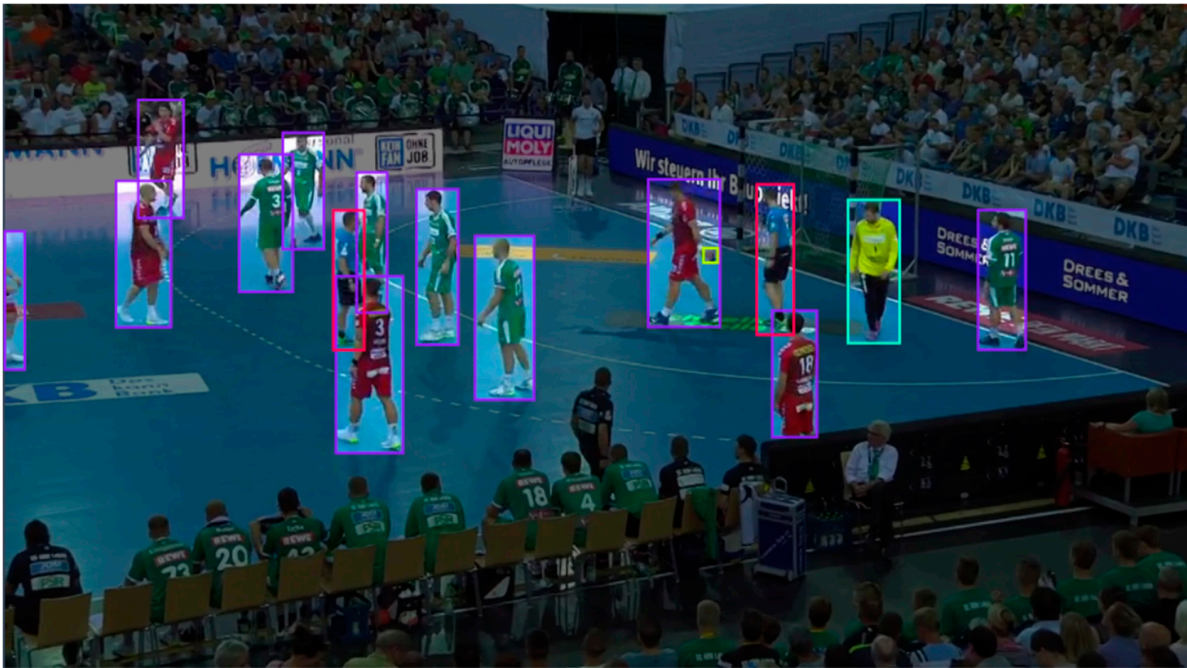


Figure 1. System Overview.

#### 3.2.1. Data Extraction and Annotation

We extracted images from video footage of handball matches using a custom script. We annotated the images by drawing bounding boxes around four target objects: a handball, player, goalkeeper, and referee. Figure 2 shows an example of the annotations.

Bounding boxes are color-coded by entity class (grey: player, green: goalkeeper, red: referee, yellow: handball). Frames were selected to illustrate varying levels of detection difficulty, including clear multi-player scenes, motion blur, occlusion, and partial concealment of the ball.



**Figure 2.** Example annotated frame from the handball dataset.

We annotated 2500 images across four object categories: player, goalkeeper, referee, and ball. Annotation was performed collaboratively using the Roboflow platform (Roboflow Inc., Des Moines, IA, USA), which enabled bounding box labeling and dataset version control. A crowdsourcing approach was adopted by recruiting volunteers through email and social media and distributing annotation tasks to improve coverage and reduce bias. The final dataset was split into 70% training, 15% validation, and 15% testing to ensure reproducibility in the model evaluation.

### 3.2.2. Model Selection and Training

We trained a computer vision model on an annotated dataset using supervised learning. Supervised learning is a method of learning from labelled examples and making predictions or classifications based on input data. The model learns to detect objects and their bounding boxes by fine-tuning its parameters.

Figures 3 and 4 show the overview and architecture of a convolutional neural network (CNN) respectively, which is a type of neural network used for computer vision.

We tested three computer vision models on our dataset:

1. YOLOv8 [13]: A state-of-the-art object detection model that builds upon the original YOLO [14].
2. YOLONAS [15]: A new model that promises better scores with reduced inference time.
3. InternImage (version InternImage-L, OpenGVLab, Shanghai, China) [16]: A CNN that incorporates a transformer-like layer to refine object detections and achieves the highest mAP score on the COCO dataset.

These models were selected based on their performance on benchmark datasets, relevance to the task of object detection in sports analytics, and suitability for real-time applications. YOLOv8 was chosen for its state-of-the-art accuracy and efficiency, whereas YOLONAS was included for its potential to improve performance with faster inference times. InternImage was selected for its high accuracy and ability to refine detections using a transformer-like layer, which is particularly relevant for complex scenes with multiple objects and occlusions.

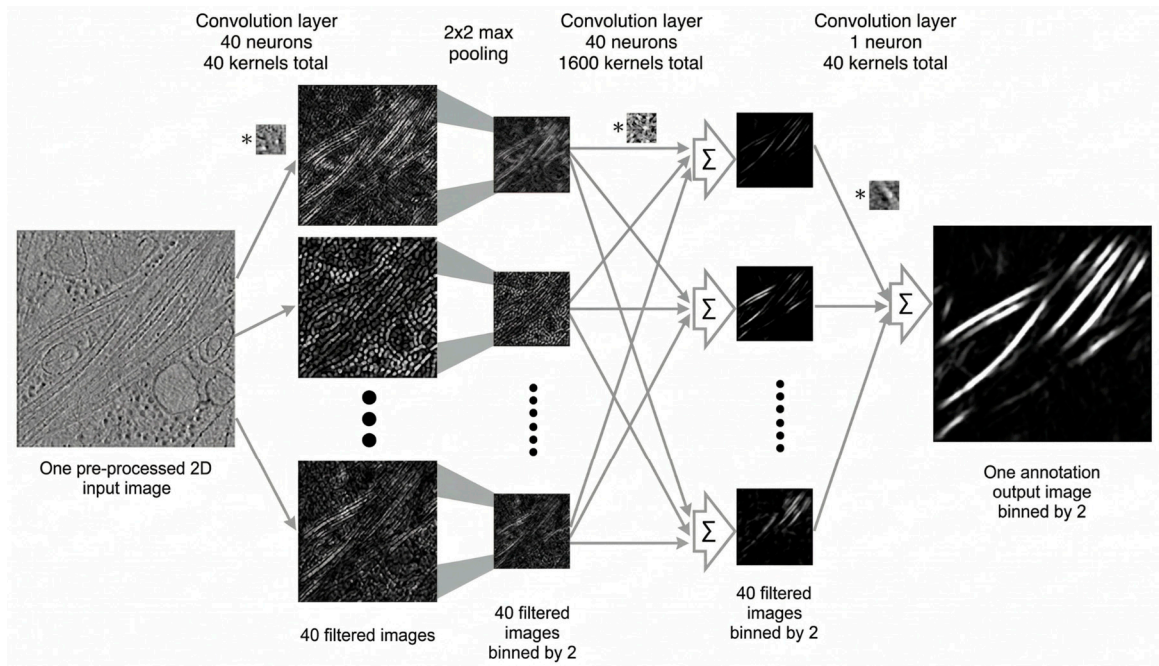


Figure 3. Overview of Convolutional Layer (\* indicates convolution operation).

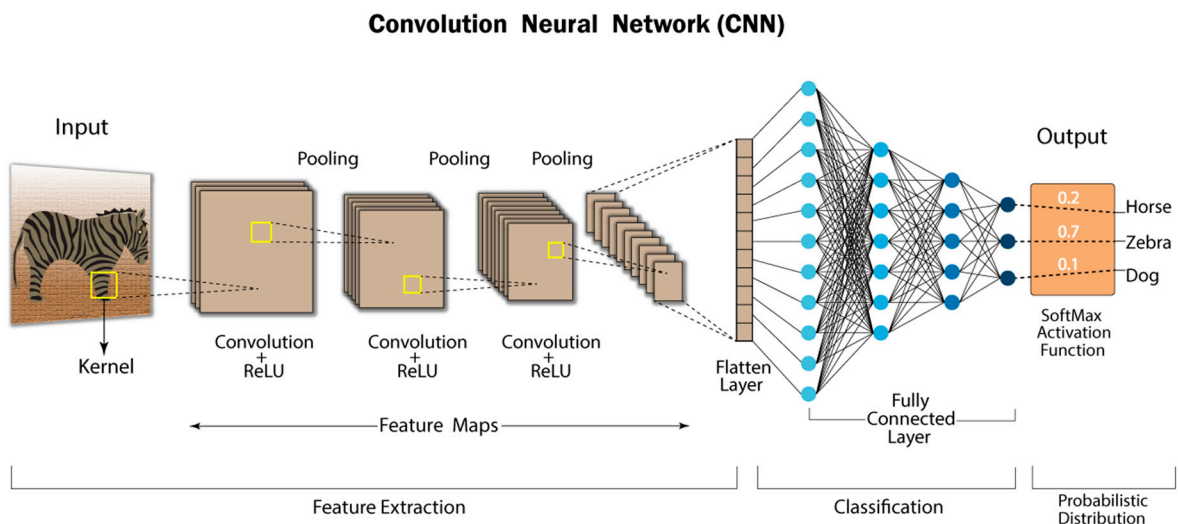


Figure 4. Full Architecture of a Standard CNN.

However, with pre-trained weights, these models struggled to identify handball and all players accurately, often detecting irrelevant objects, such as chairs and spectators. These limitations rendered them unsuitable for precise identification in match scenarios. To overcome these limitations, we fine-tuned the models to focus solely on detecting target objects with greater accuracy.

### 3.3. Cameras Technology

To explore hardware improvements for handball analytics, we explored and tested two emerging camera technologies:

- Event-based cameras: Event-based cameras capture pixel-level intensity changes asynchronously, rather than recording full frames. This enables efficient motion tracking with high temporal resolution. Prophesee [7] proposed such sensors that mimic human retinal activity. However, owing to their high cost and limited availability, event-based cameras were not used in this study but are considered promising for future research.

- Stereo cameras: Stereo cameras use two cameras to estimate depth information from a scene. We used a Zed 2 Camera 2, which is a passive stereo camera that can estimate depths up to 40 m. It also has a comprehensive Software Development Kit (SDK) that includes neural networks, Robot Operating System (ROS) integration, and an object detector. Stereo cameras enable the 3D reconstruction of a game.

For full-court coverage, we propose a four-camera setup positioned in each corner of the handball court. This configuration enables the 3D reconstruction of player and ball trajectories, reduces occlusion issues, and provides synchronized depth-enhanced analytics.

### 3.4. Justification of the Approach and Limitations

The justification of the approach used in this study includes:

- Mixed methods: Interviews ensured practical relevance, whereas computer vision experiments contributed technical rigor.
- Dataset focus: Creating a public dataset to address the scarcity of handball-specific data identified in prior studies.
- Model selection: YOLOv8, YOLONAS, and InternImage were chosen for their balance of accuracy, speed, and robustness to occlusion.
- Cameras: Stereo vision was prioritized owing to its affordability and accessibility compared with event-based sensors.

Limitations include a small interview sample size (no handball players were interviewed); a restricted effective training dataset (119–508 images per experimental phase out of 2500 total annotated); absence of k-fold cross-validation, statistical significance testing, and confidence interval reporting; a small 72-image test set that limits the statistical robustness of performance estimates; and a limited testing environment for stereo cameras. The reported metrics should therefore be interpreted as indicative benchmarks rather than definitive performance claims. Addressing these statistical limitations through k-fold cross-validation, multi-run experiments with variance reporting, and testing on independent external datasets is identified as direction for future work. Nevertheless, the methodological design provides a reproducible framework for future work in handball analytics.

## 4. Implementation

In this section, we describe the implementation of the approach and ideas and the details of the systems that were developed to address the aforementioned research gaps.

### 4.1. Data Wrangling

We used the EIGD dataset [12], which consists of 25 videos, each 5 min long, recorded at 30 frames per second (FPS), resulting in approximately 225,000 frames.

We randomly sampled 100 frames from each video, yielding a dataset of 2500 images. This number was chosen to balance annotation feasibility (given volunteer availability, quality control overhead, and review time) with the goal of producing a dataset large enough for meaningful model training experiments. It is important to note that at the time of the main experiments, only a subset of images had completed the full annotation and quality review process: 119 images were fully reviewed for the first experimental phase, and 508 for the second. The remainder were pending review or flagged as low quality. This staged availability explains the discrepancy between the total dataset size and the training set sizes reported in Section 5.3. We named each image <video\_name>@<frame\_number>.jpg and stored the metadata of the extracted frames in JSON files. This allowed us to track the original frames and account for different frame rates.

We obtained a dataset of images for annotation and shared the code in a public repository [17]. We plan to add a command-line tool with various options to make the data wrangling process more flexible and user-friendly.

#### 4.2. Tools and Novel Software

Roboflow [18] was employed to annotate the images from the EIGD dataset [12], which consists of 25 videos, each 5 min long, at 30 FPS. A box was drawn around each object of interest, such as a handball, player, referee, and goalkeeper, to make the dataset suitable for training an object detection model. To track the ball in blurry frames, the Annotations Helper Tool was developed, a novel web application that plays back the last 5 s of the video leading up to a specific frame. The tool utilizes the metadata generated by Frozen Video, a Python script (version 3.9, Python Software Foundation, Wilmington, DE, USA) that randomly sampled 100 frames from each video and named them <video\_name>@<frame\_number>.jpg. The Annotations Helper Tool and Frozen Video facilitated the annotation process and prepared the data for analysis. The Annotations Helper Tool can be accessed at [https://tofyliion.github.io/annotations\\_helper](https://tofyliion.github.io/annotations_helper) (accessed on 7 April 2026). The tool is available as Open Source code [19].

#### 4.3. Volunteering Campaign for Data Annotation

We launched a crowdsourcing campaign to recruit volunteers for annotating the images from the EIGD dataset [12] using Roboflow [18]. We created a Google Form to provide an overview of the project and received 70 registrations. We sent a follow-up form and a tutorial video to explain the annotation process and invited 35 sign-ups on Roboflow. We sent progress updates and recognition emails to active contributors every two weeks.

However, we observed a decline in annotation productivity after one month. To address this, we created a progress email with a new model and a Discord server to foster community and collaboration. These measures increased annotation output and engagement. We also developed a Google Sheet to track the annotations submitted by each volunteer and manage job assignments. These strategies enabled us to establish a strong community of contributors for the project.

#### 4.4. Models and Setup

We chose three computer vision models to train on our handball dataset: YOLOv8, YOLO-NAS, and InternImage. Each model has its own advantages and challenges in terms of accuracy, latency, and complexity. We used different hardware resources and virtual environments to train each model according to its specifications. In this subsection, we describe the details of our model selection and setup processes.

##### 4.4.1. Hardware Availability

Our personal setup consists of a laptop with an NVIDIA RTX2070 mobile graphics processing unit (GPU) (Nvidia Corporation, Santa Clara, CA, USA), which has 8 GB of video random access memory (VRAM). This allows us to train models faster than using a central processing unit (CPU); however, it also limits the maximum model size and batch size that can be used. Additionally, we have access to a lab with two NVIDIA RTX3090 GPUs, each with 24 GB of VRAM. This provides us with more flexibility and power to train larger and more complex models.

##### 4.4.2. Model Selection

We selected three models based on their performance and suitability for object detection tasks. The models are:

- YOLOv8 [13]: A popular and fast object detection model developed by Ultralytics, which incorporates various enhancements over previous YOLO versions.
- YOLO-NAS: A variant of YOLO that uses neural architecture search (NAS) to automatically design a customized model architecture for object detection tasks.
- InternImage [16]: A state-of-the-art computer vision model that employs a transformer-like layer on top of a traditional CNN.

#### 4.4.3. Models Training

We trained each model using dedicated Jupyter Notebook files and separate virtual environments using Anaconda (python 3.9). The training details for each model are as follows: YOLOv8: We trained the YOLOv8 model for 300 epochs with an image size of  $640 \times 640$  pixels and a batch size of six images per batch using an NVIDIA RTX2070 mobile GPU. YOLO-NAS: We trained the YOLO-NAS model for 400 epochs with an image size of  $640 \times 640$  pixels and a batch size of four images per batch using an NVIDIA RTX2070 mobile GPU. InternImage: We trained the InternImage model for 300 epochs with an image size of  $1330 \times 800$  pixels and a batch size of four images per batch using one NVIDIA RTX3090 GPU.

#### 4.5. Training Evaluation

After training several models, we need some metrics to be able to quickly compare the different models and estimate how well they can perform in real-match scenarios. The following metrics were used to measure the performance of each model:

- Intersection over Union (IoU);
- Confusion Matrix: especially the true-positive score for each category;
- Mean Average Precision (mAP) Score.

## 5. Results and Evaluation

In this section, we discuss the progress of the volunteering campaign and the results of each model, which were evaluated using the metrics proposed in the previous section. We also compare the performances of different models and interpret the data.

#### 5.1. Research Questions

This study aimed to address the following research questions:

1. How can we create a comprehensive and annotated dataset of handball images to facilitate automated analysis?
2. What are the challenges and opportunities associated with using different computer vision models for object detection in handball?
3. What are the prospects and current limitations of integrating advanced camera technologies, such as event-based and stereo cameras, into a handball analytics pipeline?

The results presented in the following subsections address these research questions as follows: RQ1 is addressed in Section 5.2 (dataset and campaign results); RQ2 is addressed in Sections 5.3 and 5.4 (model performance and comparison); and RQ3 is addressed at the discussion level in Section 3.3 and revisited in Section 6.2, as full quantitative benchmarking of advanced cameras was beyond the experimental scope of this study.

#### 5.2. Volunteering Campaign

Our campaign successfully contributed to the annotation of the full dataset of 2500 images. Of these, approximately 1500 images were annotated through the volunteer crowdsourcing campaign, and the remaining 1000 were completed by the core

research team. This achievement reflects the community's collective efforts and the contributors' dedication.

### 5.3. Models Performance

We tested the object detection models on the dataset and obtained interesting results, revealing distinct patterns across different technologies. These results provide valuable insights for future research. In this section, we present the mAP scores and the confusion matrix of each model based on our tests. We used a specific portion of the dataset with 72 images of  $640 \times 640$  resolution. This corresponds to the 15% test partition of the 508-image annotated pool, consistent with the 70/15/15 train/validation/test split described in Section 3.2.1 for model assessment. These images were not part of the training process for any model. This separation ensures that the models are evaluated on unseen data, providing a more accurate measure of their generalization ability and preventing overfitting to the training set. It is important to note that the evaluation was conducted on a held-out test set of 72 images that were not used in the training. However, given the small dataset size, the absence of k-fold cross-validation, and the use of a single experimental run, the reported metrics should be interpreted as indicative rather than definitive. They represent a first benchmark for handball object detection rather than a production-ready performance claim.

We analyzed the model performance and examined the confusion matrix to understand the strengths and weaknesses of each model. This evaluation helped us select the most suitable models for our object detection tasks.

#### 5.3.1. YOLOv8 Control Model

We evaluated the YOLOv8x model, which was trained on the COCO dataset, without fine-tuning it on our custom dataset. This model was tested only for handball detection, as the COCO dataset has 80 generic classes and identifies players, goalkeepers, and referees as "Person". The COCO dataset has a "Sports Ball" class, which we compared to the "Handball" annotations from our original dataset.

The YOLOv8 control model failed to detect the handball in any of the test or training images. This indicates that the pre-trained model without fine-tuning could not detect the handball in our dataset; hence, the handball true-positive was 0%. We could not calculate an mAP score for the control model because we could not evaluate the player, referee, and goalkeeper detections.

#### 5.3.2. YOLOv8 Evaluation

We trained several versions of the YOLOv8 model, using the YOLOv8X model as a pre-trained base. We tested three versions with different training data and resolutions. Figures 5–7 show the confusion matrix for each version.

We trained the first version (YOLOv8-v1) on 119 images (the number of fully quality-reviewed images available at the start of the experiments) augmented by five times, with a resolution of  $640 \times 640$ . We trained the second version (YOLOv8-v2) on 508 images (reflecting the expanded pool of quality-reviewed images following a second annotation campaign push) augmented by five times, with the same resolution. The full 2500-image dataset was not used in the experiments as annotation quality review was ongoing; the 119 and 508 figures represent the images that had completed the full review pipeline at each experimental phase. We trained the third version (YOLOv8-v3) on the same data as version 2, but with a resolution of  $1024 \times 576$ . The training time ranged from 4 to 24 h on an RTX2070 mobile GPU. The mAP score increased from 0.847 to 0.868, and the handball true-positive increased from 0.39 to 0.51 as we increased the training data size. However, increasing the resolution did not have a significant effect on the performance.

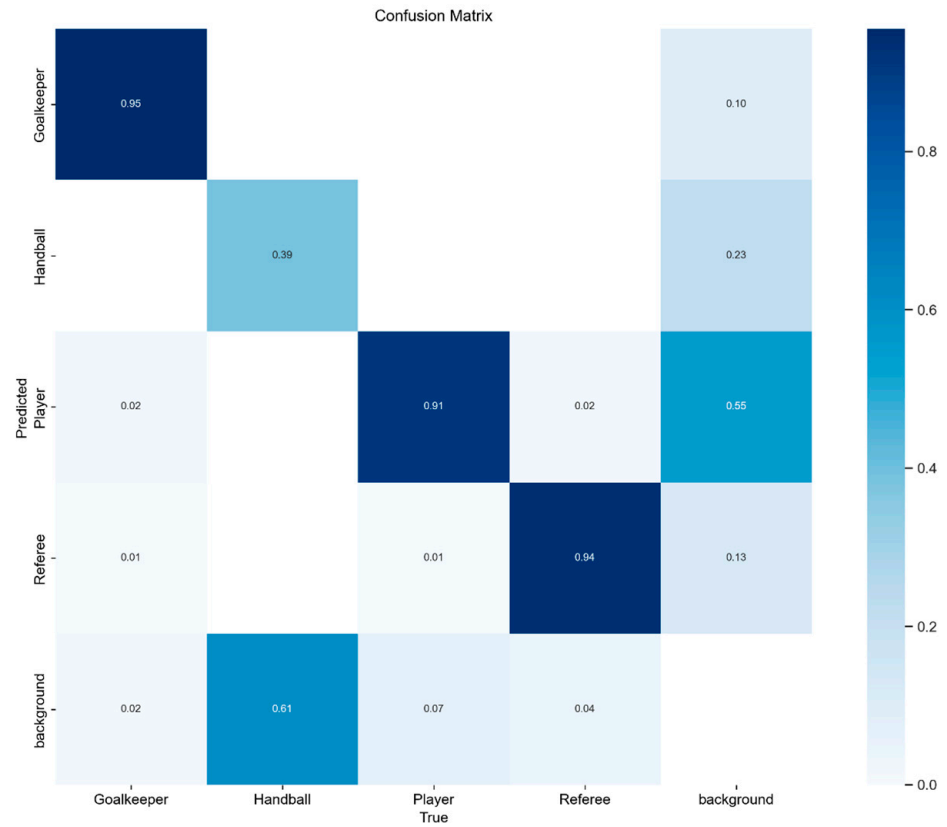


Figure 5. Confusion Matrix for YOLOv8-v1.

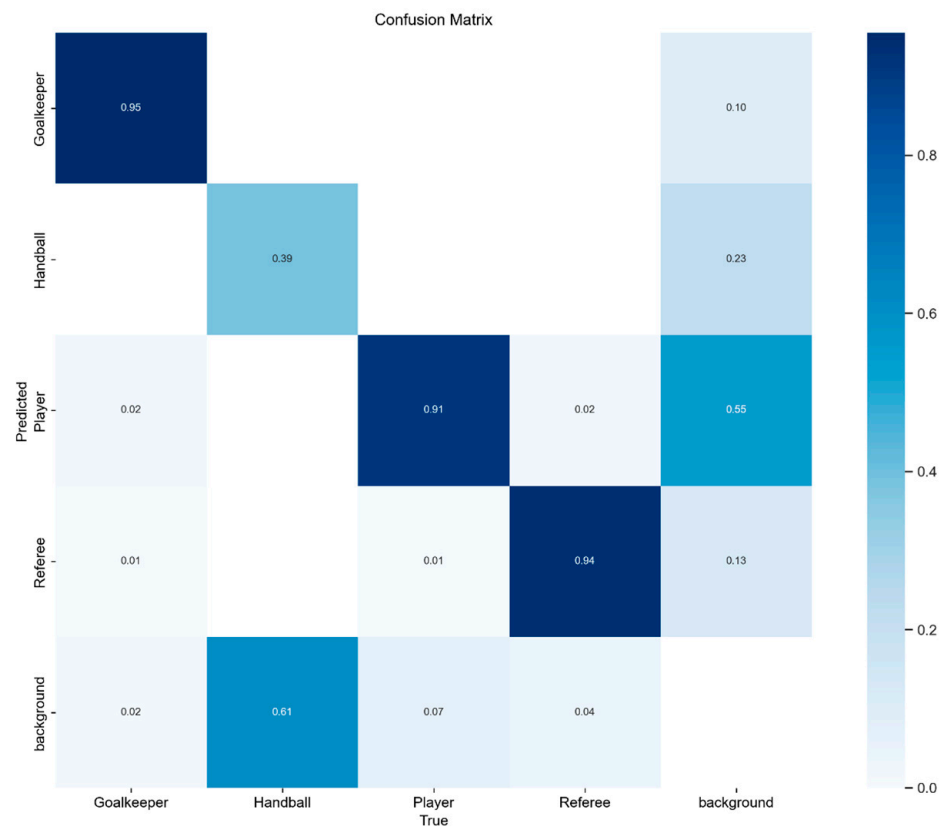


Figure 6. Confusion Matrix for YOLOv8-v2.

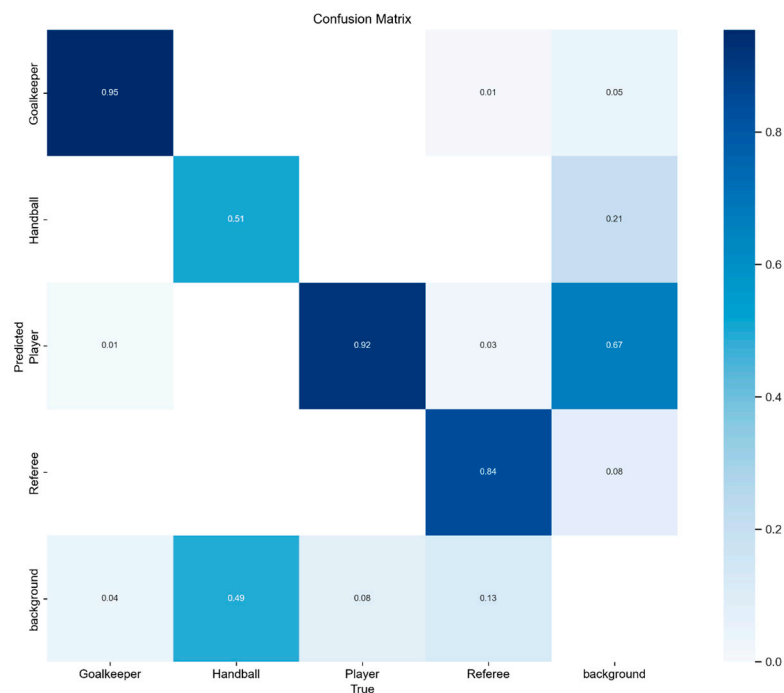


Figure 7. Confusion Matrix for YOLOv8-v3.

### 5.3.3. YOLO-NAS Evaluation

We fine-tuned YOLO-NAS-L, a fast-inference model, on 119 training images augmented by five times. Training took approximately 6 h on an RTX2070 mobile GPU. The model achieved an mAP score of 0.629 and a handball true-positive of 0.14, which are low compared to those of other models. Figure 8 shows the confusion matrix for the fine-tuned YOLO-NAS model.

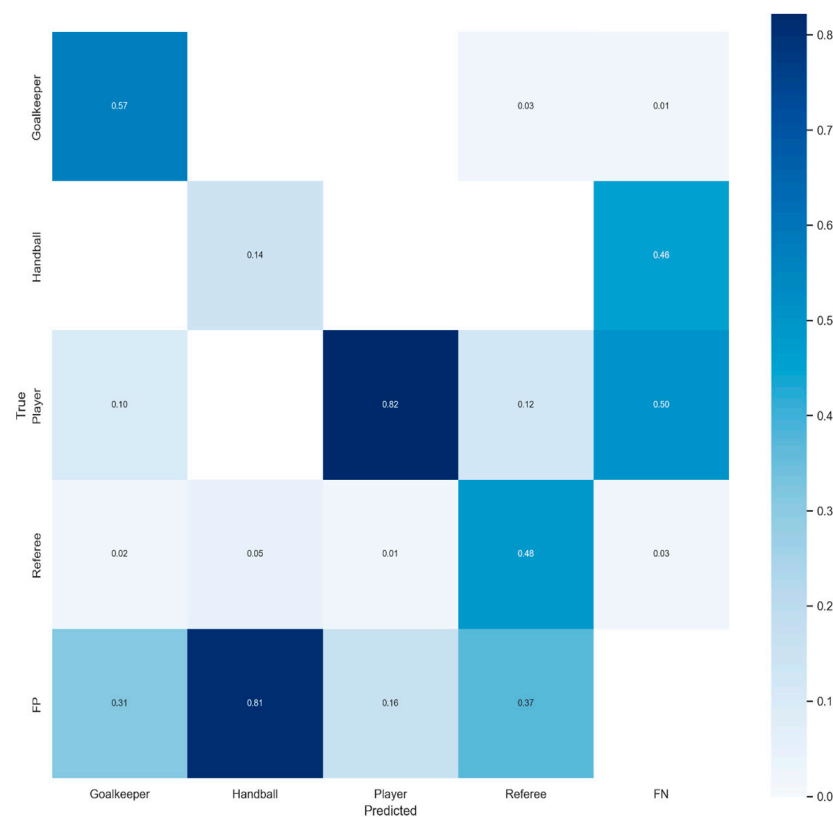


Figure 8. Confusion Matrix for fine-tuned YOLO-NAS.

### 5.3.4. InternImage Evaluation

We fine-tuned InternImage, a transformer-inspired model, on 119 images with a resolution of  $640 \times 640$  pixels. Training took approximately 33 h on an RTX3090 desktop GPU. It achieved an mAP score of 0.718 and a handball true-positive of 0.45, which are higher than those of YOLO-NAS-L but lower than those of YOLOv8. The confusion matrix is shown in Figure 9.

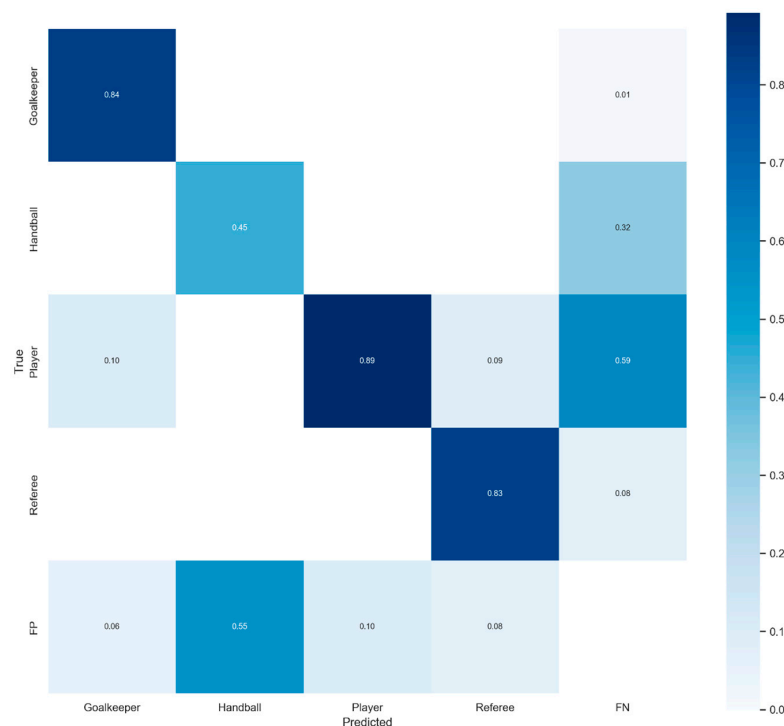


Figure 9. Confusion Matrix for fine-tuned InternImage.

### 5.4. Models Comparison

We compared the performances of YOLOv8, YOLO-NAS-L, and InternImage on our custom dataset, as shown in Table 1.

Table 1. Comparison between the different computer vision models.

Model	Resolution	Count	mAP@50	Handball	Player	Referee	Goalkeeper	FPS
Control	$640 \times 640$	-	-	0.00	-	-	-	-
YOLOv8v1	$640 \times 640$	119	<b>0.847</b>	0.39	<b>0.91</b>	<b>0.94</b>	<b>0.95</b>	15.6
YOLONAS	$640 \times 640$	119	0.6249	0.14	0.82	0.48	0.57	<b>185.5</b>
InternImage	$640 \times 640$	119	0.718	<b>0.45</b>	0.89	0.83	0.84	2.3

YOLOv8 is an efficient and accurate model, YOLO-NAS-L is a fast inference model, and InternImage is a transformer-inspired model. We tested different versions of YOLOv8 with different training data and resolutions. We tested YOLO-NAS-L and InternImage with 119 training images, augmented five times.

We observed that YOLOv8 achieved the highest mAP score and true-positive rate for handball among the models, particularly with more training data. These results demonstrate the superior performance of YOLOv8 on various object detection tasks. The efficiency of YOLOv8 may be attributed to its improved architecture and training process, which incorporates advanced techniques such as cross-stage partial connections and Mish activation. However, increasing the resolution did not significantly improve performance, suggesting

that the model may have reached its optimal performance level with the initial resolution, or that additional training data may be required to leverage the increase in input features.

YOLO-NAS-L achieved the lowest mAP score and handball true-positive rate, which could be due to a lack of fine-tuning or data. This finding is consistent with previous research on neural architecture search (NAS) methods, which often require large datasets and extensive training to achieve optimal performance [15]. NAS methods automate the process of designing neural network architectures by exploring a vast search space of possible architectures and selecting the best performers for a given task. However, the effectiveness of NAS algorithms is heavily influenced by factors such as the search space, choice of hyperparameters, and preprocessing steps [20]. In our case, the lower accuracy of YOLO-NAS-L may be attributed to the limitations of the search space and architecture parameters used in the NAS process. Further tuning of these parameters could improve performance.

InternImage achieved scores similar to those of YOLOv8 with less training data, revealing the potential of dynamic sparse kernel CNNs, which are inspired by vision transformers. This result supports the findings of [16], which highlighted the ability of transformer-inspired models to achieve high accuracy with limited training data. The strong performance of InternImage may be attributed to the dynamic sparse kernel, which allows the model to focus on the most relevant parts of the image and capture long-range dependencies between objects. However, this model also had longer training and inference times, which could limit its applicability in real-time systems. This limitation is a common challenge for transformer-based models, as they often require more computational resources and time than CNN-based models [21]. Future research could explore techniques to optimize the efficiency of dynamic sparse kernel CNNs or investigate their applications in scenarios where real-time performance is not critical, such as post-game analysis or player tracking over longer periods.

The models also exhibited challenges in distinguishing between players, goalkeepers, and referees, as well as detecting small or occluded balls. These could be addressed by fine-tuning the model parameters, adding more diverse and annotated data, or applying post-processing techniques. Further enhancements can include several camera angles and better/different sensors.

Table 2 compares different YOLO models.

**Table 2.** Comparison between the different YOLO models.

Model	Resolution	Count	mAP@50	Handball	Player	Referee	Goalkeeper	FPS
Control	640 × 640	-	-	0.00	-	-	-	-
YOLOv8v1	640 × 640	119	0.847	0.39	0.91	0.94	0.95	<b>15.6</b>
YOLOv8v2	640 × 640	508	0.866	<b>0.51</b>	<b>0.97</b>	<b>0.95</b>	<b>0.96</b>	<b>15.6</b>
YOLOv8v3	1024 × 576	508	<b>0.868</b>	<b>0.51</b>	0.92	0.84	0.95	7.4

In Table 2, resolution refers to the resolution of the training images. All images predicted by the models were automatically resized before being entered into the network. Count refers to the number of unique training images used by each model.

Handball, player, referee, and goalkeeper all refer to the true positives of each entity. The maximum is 1. More details can be found in the confusion matrix for each model’s results. The higher the number, the better the model is at identifying this entity.

The FPS column refers to the expected number of frames per second that can be generated if the model is run on a video. The average frame rate of a video is 30; therefore, if the model can run at 30 fps, it can be used in real time. The higher the FPS, the faster the model identifies entities in a video. This refers to the FPS of the isolated model

inference as tested on our NVIDIA RTX2070 environment, rather than an end-to-end pipeline evaluation.

### 5.5. Results Takeaways

The results show that YOLOv8 provides a good balance between accuracy and speed for object detection tasks. However, it has not yet been tested in production environments. InternImage is a promising model that can achieve high scores with limited training data; however, it requires more hardware resources and time. YOLO-NAS-L is not a viable option for our dataset at the moment; however, it could be retested with more data in the future.

A quantitative breakdown of the error modes, derived from the YOLOv8-v3 confusion matrix (Figure 7), reveals distinct challenges across classes. The primary bottleneck remains handball detection, which suffers from a high False Negative (FN) rate of 49% (true handballs completely undetected and classified as background). This quantitatively reinforces the severe impact of motion blur and occlusion. Furthermore, 21% of the model's background false positive (FP) errors were incorrect handball predictions, indicating that the model struggles to differentiate the ball from visually similar background artifacts, such as shoes or court markings. Conversely, human entities showed much lower FN rates (players at 8%, goalkeepers at 4%), though minor inter-class confusion occurred, such as 3% of Referees being misclassified as Players.

Although the detection rates for players, goalkeepers, and referees are sufficiently high for exploratory use, the handball true positive rate of approximately 0.51 remains below the threshold required for reliable real-world or real-time deployment. This finding underscores the inherent difficulty of detecting a small, fast-moving, and frequently occluded object using static frame-level detection alone. The primary failure modes identified include frames with extreme ball motion blur, partial occlusion by players or court elements, and frames in which the ball is near the goal net with a similar background texture.

Figure 10 presents representative success and failure cases for handball detection under challenging visual conditions, where the ball is difficult to discern even for human annotators. The top frame demonstrates a successful detection despite complex background elements. Conversely, the bottom frame illustrates a typical failure mode (a false negative) where the ball is partially occluded during a catch, and its color visually blends with the player's jersey, preventing accurate detection by the model.

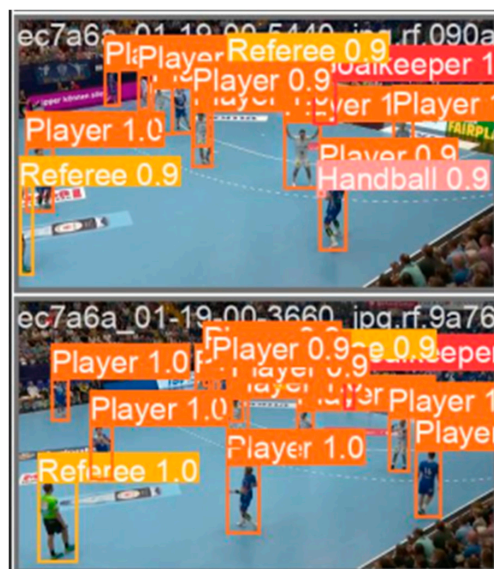


Figure 10. 2 Examples of success (top) and failure (bottom) with challenging conditions.

## 6. Conclusions and Future Work

### 6.1. Conclusions

This study contributes to the advancement of object detection in handball matches by implementing and testing three computer vision models. We implemented and evaluated three state-of-the-art models, achieving notable improvements in detecting entities, particularly handball. These contributions establish a foundation for further research on real-time sports analytics systems.

This study addressed key challenges in object detection for handball, including limited datasets and model adaptability. By developing an annotated dataset and training models to identify players, referees, goalkeepers, and the ball, we demonstrated the feasibility of automated analysis and its potential real-world applications:

- **Real-time performance tracking:** The developed models can serve as a foundation for integration into camera systems to track player movements and ball trajectories during live matches, providing coaches and analysts with immediate feedback on player performance and team strategies. Achieving reliable real-time deployment will require temporal modeling, tracking pipelines, and substantially expanded and validated datasets.
- **Post-game analysis:** The annotated dataset and trained models lay a foundation for future systems that analyze recorded matches, enabling coaches to identify tactical patterns, evaluate player decisions, and develop targeted training programs.
- **Referee assistance:** Object detection capabilities, once sufficiently accurate and validated, can assist referees in making more accurate decisions, particularly in fast-paced or complex situations where it may be difficult to track all players and the ball simultaneously.
- **Automated event tagging:** The developed framework represents an early-stage foundation that can be extended to automatically identify events from a pre-recorded match or in real-time.

In addition to its technical achievements, the study highlights the significance of automation in enhancing decision-making in handball. Coaches gain reliable statistics for tactical planning, players benefit from objective feedback on performance, and referees receive assistance in rule enforcement. Collectively, these insights contribute to the modernization and professionalization of handball analytics.

This study also revealed areas where further improvements are needed, particularly in data quality, model robustness, and the inclusion of advanced techniques, such as temporal modeling. These findings inform the roadmap for continued exploration in this domain.

While this study provides a vital first step, substantial improvements in dataset scale, temporal modeling, tracking, statistical validation, and latency optimization are required before these models can be reliably deployed in live, real-world match environments.

### 6.2. Future Work

We improved the object detection of handball entities, especially handball. This provides a basis for future research, such as object tracking. However, we face some limitations, such as:

Although this study achieved promising results, some limitations remain. Addressing the following points will be critical for advancing the automation of handball analytics:

- **Statistical validation:** The current results are based on a single experimental run evaluated on a relatively small held-out test set of 72 images, without k-fold cross-validation or multi-run variance reporting. Future work should prioritize k-fold cross-validation (e.g., 5-fold or 10-fold) to obtain more reliable performance estimates and

confidence intervals; multi-run experiments with different random seeds to quantify result variability; and evaluation on an independent external handball dataset.

- Camera technology: Although stereo cameras were tested, other emerging technologies such as event-based cameras have not been fully explored. Future studies should investigate their integration to capture finer motion details and improve model performance.
- Coaches' and players' feedback: Interviews with coaches and analysts provided valuable perspectives; however, the sample size was limited. Expanding this engagement with a broader and more diverse group of stakeholders will ensure that the developed systems address practical needs more effectively.
- Dataset expansion: The annotated dataset created in this study is a first step. Expanding it with more match recordings, varied camera angles, female players, and action-specific annotations will improve generalizability and enable more complex analyses, such as tactical recognition.
- Computer vision models: Our focus was primarily on detection accuracy rather than inference speed. Future research should prioritize model optimization, including hyperparameter tuning, lightweight architectures, and the use of temporal or depth information for faster and more efficient detections.
- Object tracking: Object tracking is a natural progression from object detection. Assigning consistent IDs to entities over time will enable the analysis of player movement patterns, ball trajectories, and tactical formations. Our dataset provides a foundation for advancing research in this direction. Future iterations will incorporate temporal information such as multi-frame smoothing, optical flow, or dedicated tracking algorithms, to leverage ball trajectory continuity and substantially improve detection reliability. In addition, we plan to report performance variability across multiple runs to strengthen the scientific validity.
- Future work will also include full-match latency benchmarking under realistic streaming conditions, accounting for preprocessing overhead, I/O bottlenecks, and hardware variability, to provide a more credible assessment of the real-time deployment potential.
- In future work, we plan to conduct a more detailed error analysis in a more in-depth manner. Moreover, we plan to incorporate attention mechanisms or instance segmentation to help differentiate overlapping entities in future model iterations.

This study provides preliminary, exploratory evidence for the applicability of computer vision and machine learning to handball, a sport that has received limited attention in sports analytics research. By creating a new dataset, evaluating state-of-the-art models, and exploring advanced camera technologies, we laid the groundwork for automated handball analysis. The outlined future work provides a clear path forward for researchers aiming to transform performance evaluation, coaching strategies, and refereeing decisions in this dynamic sport.

**Author Contributions:** Conceptualization, M.F. and H.S.; methodology, M.F. and H.S.; software, M.F.; validation, H.S., D.K.M. and A.A.; formal analysis, M.F.; investigation, M.F., H.S. and D.K.M.; resources, M.F.; data curation, M.F.; writing—original draft preparation, M.F., H.S. and D.K.M.; writing—review and editing, M.F., H.S., D.K.M. and A.A.; visualization, M.F., H.S. and D.K.M.; supervision, H.S., D.K.M. and A.A.; project administration, H.S. and D.K.M.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data used to support the findings of this study are available from the Roboflow platform at <https://universe.roboflow.com/handball-ai-training/handball-detection> (accessed on 7 April 2026). The dataset is publicly released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license to encourage reproducibility and future research. The underlying video frames are derived from the EIGD dataset [12]; researchers should verify the EIGD source terms of use for any intended commercial applications, as the CC BY 4.0 licence applies specifically to the annotations and associated metadata.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. IEEE Standard P2413/D0.4.6; IEEE Standard for an Architectural Framework for the Internet of Things (IoT). IEEE: New York, NY, USA, 2019.
2. Marck, A.; Antero-Jacquemin, J.; Berthelot, G.; Saulière, G.; Jancovici, J.M.; Masson-Delmotte, V.; Gilles, B.; Spedding, M.; Le Bourg, É.; Toussaint, J.F. Are We Reaching the Limits of *Homo sapiens*? *Front. Physiol.* **2017**, *8*, 812. [CrossRef]
3. Rajšp, A.; Fister, I., Jr. A systematic literature review of intelligent data analysis methods for smart sport training. *Appl. Sci.* **2020**, *10*, 3013. [CrossRef]
4. BBC. Handball-Factfile. Available online: <https://www.bbc.co.uk/bitesize/guides/zxspfrd/revision/1> (accessed on 30 June 2023).
5. Kinexon. Kinexon Handball Homepage. Available online: <https://kinexon-sports.com/sports/handball> (accessed on 30 June 2023).
6. Foina, A.G.; Badia, R.M.; El-Deeb, A.; Ramirez-Fernandez, F.J. Player Tracker—a tool to analyze sport players using RFID. In *Proceedings of the 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Mannheim, Germany, 29 March–2 April 2010*; IEEE: New York, NY, USA, 2010; pp. 772–775.
7. Prophesee. Prophesee Homepage. Available online: <https://www.prophesee.ai> (accessed on 30 June 2023).
8. Schrapf, N.; Alsaied, S.; Tilp, M. Tactical interaction of offensive and defensive teams in team handball analysed by artificial neural networks. *Math. Comput. Model. Dyn. Syst.* **2017**, *23*, 363–371. [CrossRef]
9. Labs, S. Zed 2 Camera Homepage. Available online: <https://www.stereolabs.com/zed-2> (accessed on 30 June 2023).
10. Vallance, E.; Sutton-Charani, N.; Imoussaten, A.; Montmain, J.; Perrey, S. Combining internal-and external-training-loads to predict non-contact injuries in soccer. *Appl. Sci.* **2020**, *10*, 5261. [CrossRef]
11. Host, K.; Ivasic-Kos, M.; Pobar, M. Action recognition in handball scenes. In *Intelligent Computing: Proceedings of the 2021 Computing Conference*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 645–656.
12. Biermann, H.; Theiner, J.; Bassek, M.; Raabe, D.; Memmert, D.; Ewerth, R. A unified taxonomy and multimodal dataset for events in invasion games. In *Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports, Chengdu, China, 20 October 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–10.
13. Jocher, G.; Chaurasia, A.; Qiu, J. *Ultralytics YOLOv8, version 8.0.0*; Ultralytics: Los Angeles, CA, USA, 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 7 April 2026).
14. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
15. TofyLion. GitHub Repository for Annotations Helper Tool. Available online: [https://github.com/tofyion/annotations\\_helper](https://github.com/tofyion/annotations_helper) (accessed on 30 June 2023).
16. Wang, W.; Dai, J.; Chen, Z.; Huang, Z.; Li, Z.; Zhu, X.; Hu, X.; Lu, T.; Lu, L.; Li, H.; et al. InternImage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023*; pp. 14408–14419.
17. TofyLion. GitHub Repository for Tool That Extracts Random Frames from Videos. Available online: <https://github.com/tofyion/frozen-video> (accessed on 30 June 2023).
18. Roboflow. Roboflow Home. Available online: <https://roboflow.com> (accessed on 30 June 2023).
19. Aharon, S.; Louis-Dupont; Masad, O.; Yurkova, K.; Fridman, L.; Lkdci; Khvedchenya, E.; Rubin, R.; Bagrov, N.; Tymchenko, B.; et al. *Super-Gradients*, version 3.0.8; Zenodo: Geneva, Switzerland, 2021. [CrossRef]
20. Kyriakides, G.; Margaritis, K.G. An Introduction to Neural Architecture Search for Convolutional Networks. *arXiv* **2020**, arXiv:2005.11074. [CrossRef]
21. Khan, S.H.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 200. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.