# Event-based High-speed Ball Detection in Sports Video

Takuya Nakabayashi*
nakka0204@keio.jp
Keio University
Yokohama, Kanagawa, Japan

Akimasa Kondo*
akms.k1223@keio.jp
Keio University
Yokohama, Kanagawa, Japan

Kyota Higa
k-higa@nec.com
NEC Corporation
Kawasaki, Kanagawa, Japan

Andreu Girbau
agirbau@nii.ac.jp
National Institute of Informatics
Chiyoda-ku, Tokyo, Japan

Shin'ichi Satoh
satoh@nii.ac.jp
National Institute of Informatics
Chiyoda-ku, Tokyo, Japan

Hideo Saito
hs@keio.jp
Keio University
Yokohama, Kanagawa, Japan

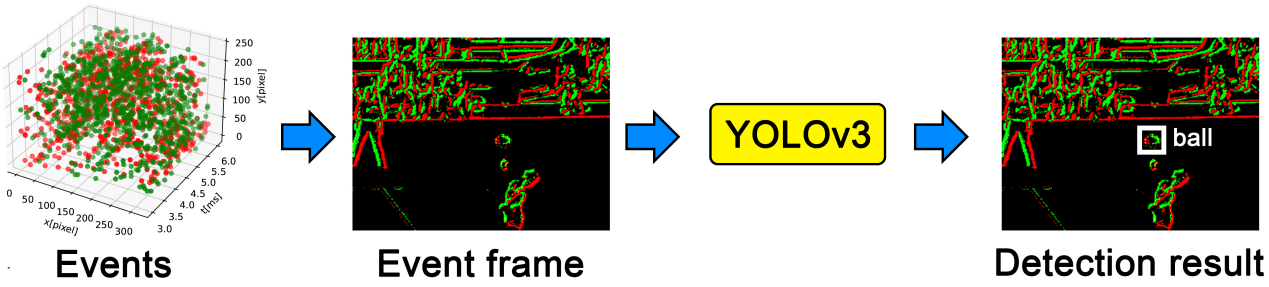


**Figure 1: Overview of the proposed method**

## ABSTRACT

Ball detection in sports, particularly in fast-paced games like volleyball, where the ball is constantly in high motion, presents a significant challenge for game analysis and automated sports broadcasting. Conventional camera-based ball detection faces issues, such as motion blur, in high-speed ball movement scenes. To address these challenges, we propose a deep learning-based method for detecting balls using event cameras. Event cameras, also known as dynamic vision sensors, operate differently from traditional cameras. Instead of capturing frames at fixed intervals, they record individual pixel-level luminance changes, referred to as events. This unique feature enables event cameras to provide precise temporal information with low latency. Our proposed method transforms sparse events into an image format, enabling the use of current deep-learning architectures for object detection. Given the limited amount of events available for training an object detector, we generate synthetic events from RGB frames. This approach reduces the need for extensive annotation and ensures sufficient data availability. Experimental results confirm that our proposed method can detect balls that are undetectable in RGB frames and outperform existing methods that utilize event-based ball detection. Moreover, we conducted tests to verify our method's ability to detect balls in real events, not just synthetic ones. These results demonstrate that our proposed method opens up new possibilities in sports ball detection.

## CCS CONCEPTS

• **Computing methodologies** → **Object detection**.

## KEYWORDS

ball detection, event camera, synthetic datasets

**ACM Reference Format:**
Takuya Nakabayashi, Akimasa Kondo, Kyota Higa, Andreu Girbau, Shin'ichi Satoh, and Hideo Saito. 2023. Event-based High-speed Ball Detection in Sports Video. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports (MMSports '23), October 29, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3606038.3616164

## 1 INTRODUCTION

Sports video analysis using computer vision technology is used in a variety of situations. For example, VAR (Video Assistant Referee) in soccer and Hawkeye in tennis support human referees and play a role in reducing misjudgments. Automatic highlight generation, which extracts the best scenes in a game from long sports videos, is also being researched for practical application.

In ball games such as soccer, table tennis, and volleyball, the detection and tracking of the ball's position is particularly crucial. Accurately capturing the ball's position during a game can aid in

*Both authors contributed equally to this research.


*MMSports '23, October 29, 2023, Ottawa, ON, Canada*

ACM ISBN 979-8-4007-0269-3/23/10…$15.00
https://doi.org/10.1145/3606038.3616164

analyzing player performance and providing support to human referees.

In the field of computer vision, several techniques have been proposed to automatically detect ball positions in images. One widely known technique is to estimate the ball position by detecting circles in an image using the Hough Transform [5]. Recently, a method for estimating ball positions using deep learning-based object detection has also been proposed [7, 16, 20–22].

A common drawback of these methods is the difficulty in detecting the position of a fast-moving ball. This challenge arises because, when a fast-moving ball is captured using a normal camera, motion blur is generated around the ball. This motion blur obscures the ball's position in the image, making it challenging to accurately capture its exact position. This difficulty is particularly pronounced in sports videos where the ball is often moving at high speeds, further complicating its detection.

A new device expected to address limitations of conventional computer vision technology is the event camera. An event camera captures images in a fundamentally different manner than a normal camera. While a normal camera acquires synchronous luminance data from all pixel positions and outputs it as an image, an event camera captures asynchronous luminance changes at each pixel position and outputs them as events. Due to their exceptionally high temporal resolution compared to normal cameras, event cameras excel at capturing the movement of fast-moving objects.

This paper proposes a method for detecting the position of a ball in sports videos using an event camera. We employ a deep learning-based object detection model for recognizing the ball.

We introduce a new dataset by generating events from regular videos captured during the game. Since preparing a large number of datasets of sports matches using event cameras can be labor-intensive, this problem was addressed by repurposing datasets captured by conventional RGB cameras.

The contributions to this study are as follows.

- We propose a ball detection method for sports video analysis using an event camera. The utilization of an event camera allows for the detection of the position of a fast-moving ball, which is challenging to capture using conventional frame-based methods.
- We propose a method to create a dataset for ball detection in sports scenes by generating events from video captured by common cameras. The proposed method solves the problem of preparing a large number of datasets captured by event cameras by allowing the reuse of existing datasets consisting of RGB frames.

## 2 RELATED WORK

### 2.1 Object Detection in Sports Video Analysis

Video analysis plays a significant role in sports analytics, and with the advancements in computer vision, it has become a popular research area across various sports. Recent developments in deep learning techniques have further enhanced the capabilities of video analysis, allowing for more advanced and detailed analysis. One of the prominent tasks in computer vision is object detection, which involves identifying the position and category of specific objects in images.

In the context of sports video analysis, object detection is utilized to extract players and other relevant objects for a wide range of applications. For example, Liu et al. proposed a method that groups detected players and sports-related objects using object detection techniques [9]. Similarly, Martin et al. and Nonaka et al. employed object detection to determine the positions of players in rugby game footage, specifically when assessing tackle risks [11, 13].

In ball sports, such as volleyball, detecting and tracking the ball's position is of paramount importance. The ball's position serves as a crucial indicator of the game situation and significantly influences player actions and strategic decisions. This study focuses on the task of ball detection in volleyball, aiming to develop effective techniques for accurately identifying the ball in volleyball game footage.

### 2.2 Ball Detection with RGB image

The challenge of detecting sports balls from RGB images has been explored through a variety of deep learning techniques. Generally, these methods are divided into two categories: segmentation approaches and object detection approaches.

In the segmentation approach, there is a proposed method designed to detect basketballs [21]. The method uses a convolutional neural network (CNN) to generate a heat map showing the location of the ball, effectively isolating the regions where the ball is most likely to be present.

In contrast, object detection strategies range from traditional sliding window methods to novel object detector. The sliding window approach[7, 16] involves scanning the image and subsequently classifying these regions. Although effective, this approach can be computationally demanding and slow owing to the extensive number of regions requiring scanning. To overcome the limitations of the sliding window approach, ball detection methods using object detectors have been proposed as a more effective alternative. For example, techniques for detecting golf balls[22] employ predictions of the ball's position from prior frames, which allows the object detection process to focus on restricted areas. This method significantly enhances both speed and efficiency. Another recent innovation in soccer ball detection[20] leverages semi-supervised learning to construct precise ball detectors with minimal data.

Nonetheless, these techniques fundamentally depend on RGB images and thus exhibit inherent limitations. For instance, when the ball's motion is too rapid, detection becomes problematic due to motion blur. To resolve these challenges, our research suggests the utilization of data obtained from event-based cameras. This innovative solution can address issues of rapid and blurred motion, facilitating more robust ball detection in sports scenarios.

### 2.3 Ball Detection with Event Camera

In the realm of event-based ball detection, a method has been proposed by Glover and Bartolozzi[4]. This technique extracts circles with high precision from events. This is accomplished by imposing geometric constraints on the conventional Hough transform and adjusting the process to fit the event-driven setting.

Nevertheless, a fundamental assumption of the method by Glover and Bartolozzi is that events are predominantly concentrated at the ball's contours. While this approach is efficient for detecting

objects with uniform or smooth surfaces, it struggles when applied to textured entities such as volleyballs. Specifically, for a volleyball, events also occur within the ball due to its textured surface, which contradicts the foundational premise of the method by Glover and Bartolozzi. Consequently, the method by Glover and Bartolozzi proves insufficient for the effective detection of volleyballs.

To address this challenge, we employ a deep learning-based object detector. This approach allows for the robust detection of balls.

# 3 METHOD

## 3.1 Event Camera

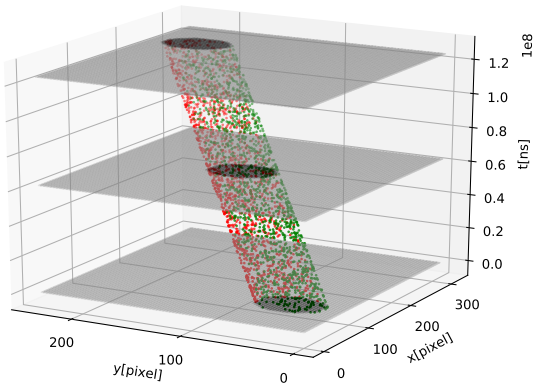


**Figure 2: Events. The green dots indicate $p = 1$ events, while the red dots indicate $p = -1$ events.**

An event camera captures pixel luminance changes as events. Each event comprises three elements: the position $(x, y)$ where the luminance change occurred, the timestamp $t$, and the polarity $p$ ($p = \pm 1$ for increasing or decreasing intensity). Figure 2 illustrates the events that occur when the black circle is in constant velocity linear motion in the {x, y, t} space. As events are more likely to occur near edges with significant spatial luminance gradients, it is evident from the figure that the events closely follow the trajectory of the black circle.

Since events are sparse data, directly handling them in a CNN poses challenges. To address this, in this study, events are transformed into event frames using the method described in Section 3.2, and subsequent inference is conducted on these event frames.

## 3.2 Preprocessing

In this paper, events are converted into event frames[17, 18] before network inference. An event frame represents an accumulation of events that occurred within a specific time interval, based on their coordinates and polarity. First, events occurring within a time interval of $n$ milliseconds is retrieved. Next, the normalized count of events that occurred at pixel position $(x, y)$ in the color channel, based on polarity, is used as the pixel value in the event frame.

## 3.3 Ball Detection from Events

The event frames created by Section 3.2 are used as input for ball detection by our object detector. In this work, we use YOLOv3[15] as our object detector. YOLOv3 is a deep learning model for object detection used by many methods[1, 12]. It detects objects at various scales by dividing the input image into multiple feature maps with different resolutions. YOLOv3 predicts objects from feature maps at three different scales and supports a wide range of object sizes. It also uses anchor boxes to estimate the position and size of objects and handles objects with different aspect ratios. By using non-maximum suppression (NMS), YOLOv3 eliminates duplicate detections and provides reliable results.

# 4 DATASET

Four datasets were created and used for experimentation in this study: RGB, EventStatic, EventDynamic, and EventReal-world. The details and attributes of each dataset are presented in Tables 1 and 2.

## 4.1 RGB dataset

The RGB dataset generates from high-definition videos of volleyball matches. The videos are recorded at a resolution of $1920 \times 1080$ pixels and a frame rate of 60 frames per second (FPS), using a fixed camera position to maintain a consistent perspective. It includes annotations for ball trajectories associated with five specific actions; serving, digging, setting, spiking, and blocking. Each action is represented in 15 sequences, further divided into training, validation, and test sets with a distribution of 10, 2, and 3 sequences respectively.

## 4.2 EventStatic dataset

The EventStatic dataset consists of events generated from frames composited based on the RGB dataset. EventStatic indicates that the scene is assumed to be captured by a fixed event camera. The following is an explanation of how the dataset is synthesized.

As part of our approach, we generate 3D models of balls, incorporating various rotational patterns, using volleyball-like texture. This procedure is designed to capture the dynamic behavior of balls under diverse conditions.

In the subsequent step, we generate synthetic frames for each sequence. Initially, we create images where the ball is artificially removed by leveraging the data associated with the ball's position within the image. We seek out adjacent frames where there is no overlap with the ball and substitute the pixels corresponding to the ball with pixels from these frames.

Next, a frame interpolation model based on deep learning[6] is employed to interpolate the background image without the ball. The position of the ball is also interpolated using a cubic spline[2] to match the count of the interpolated frames. By employing interpolation algorithms to determine the ball's position, we can calculate a bounding box tailored to the ball's size, thereby obviating the need for annotations on interpolated frames and significantly reducing the volume of annotations.

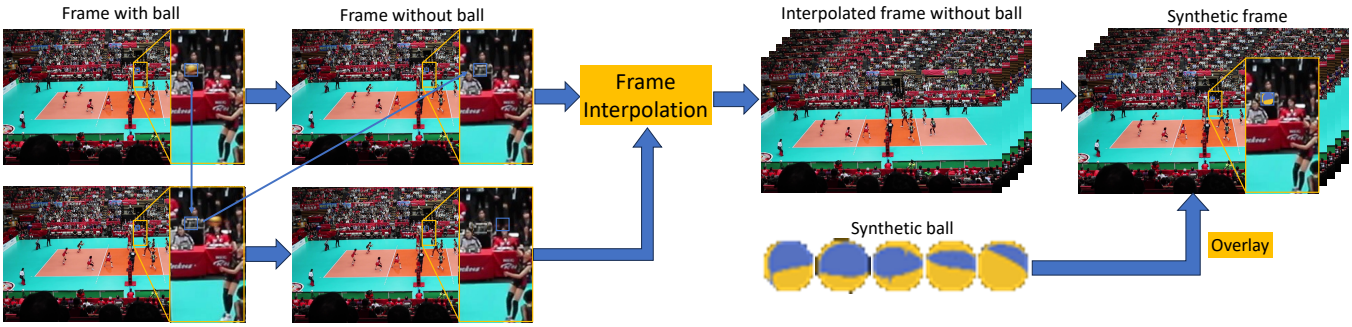


**Figure 3: Overview of frame synthesis**

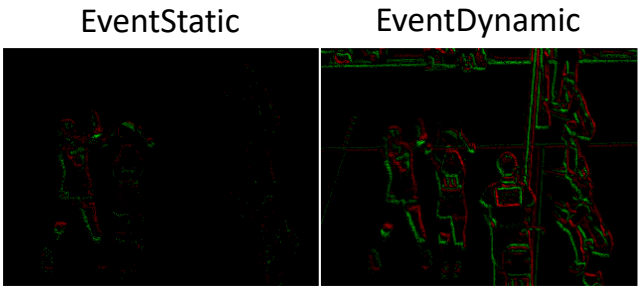


**Figure 4: Comparison of frame images between EventStatic and EventDynamic. EventStatic shows cropped images of the corresponding region.**

After completing the interpolation, we overlay the image of the 3D-modeled ball at the corresponding positions onto the interpolated frames. Figure 3 illustrates the method for generating synthetic RGB frames up tp this point.

The EventStatic dataset is synthesized by inputting the group of frames generated in this manner into ESIM[14] to generate events. ESIM is one of the event camera simulators, which generates events by inputting video frames. These events that occurred within intervals of 3 milliseconds are then converted into event frames using the method described in Section 3.2.

## 4.3 EventDynamic dataset

The EventDynamic dataset contains events that occur in a scene captured by a camera moving to track a ball. It is created by cropping the area around the ball from an image interpolated to a high frame rate using the method described in Section 4.2. These cropped images are then fed into ESIM, and the resulting events occurring within 3 millisecond intervals are converted into an event frame. Figure 4 illustrates a comparison between the frames of EventStatic and EventDynamic and reveals significant variations in the number of events.

## 4.4 EventReal-world dataset

To evaluate the real-world performance of our proposed method, we created EventReal-world dataset. This dataset consists of events captured using a DAVIS346 MONO camera, firmly mounted on a tripod, depicting a player performing a volleyball spike. This events is complemented by ground truth data representing the position of the volleyball. The dataset contains 10 sequences featuring different spike scenes. Event frames within this dataset were generated from events that occurred within intervals of 3 milliseconds.

**Table 1: Number of frames in each dataset**

| | Training | Validation | Test |
|---|---|---|---|
| RGB | 1058 | 157 | 310 |
| EventStatic | 17571 | 2651 | 5201 |
| EventDynamic | 17571 | 2651 | 5201 |
| EventReal-world | 285 | 26 | 55 |

**Table 2: Attributes of our datasets**

| | Event | Synthetic | Camera motion |
|---|---|---|---|
| RGB | | | |
| EventStatic | ✓ | ✓ | |
| EventDynamic | ✓ | ✓ | ✓ |
| EventReal-world | ✓ | | |

# 5 EXPERIMENTS

## 5.1 RGB vs. Events

In this experiments, we compare the outcomes from RGB dataset with those from EventStatic dataset. The main objective is to demonstrate the effectiveness of object detection in event frames and our synthetic dataset creation method.

### 5.1.1 *Experimental Settings.*

***Datasets:*** We evaluated our model using two datasets: RGB dataset and EventStatic dataset. These datasets are described in Sections 4.1 and 4.2 respectively. Using these diverse datasets, we provide a comprehensive and comparative performance evaluation, highlighting the robustness and adaptability of our proposed method.

***Comparison methods:*** Since the purpose of this experiment is to focus on the differences in the datasets, we only use our model as the comparison method.

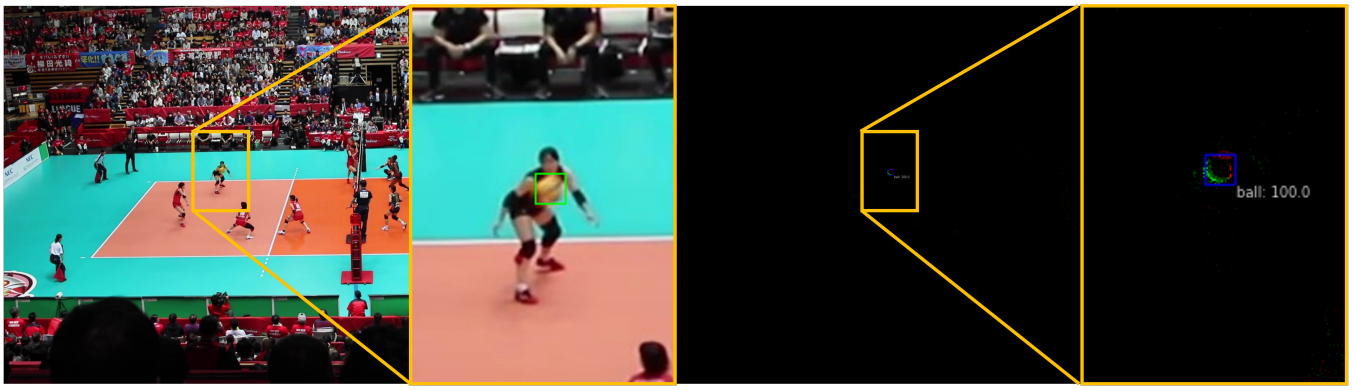


**Figure 5: Comparative analysis of object detection results in an RGB frame and an EventStatic frame. The green bounding box in the left RGB frame represents the ground truth, while the blue bounding box in the right EventStatic frame illustrates the inferred bounding box and its corresponding object detection score. The areas highlighted with yellow frames in both images indicate the regions that have been zoomed in for a more detailed examination.**

**Table 3: Comparison results between our method and the benchmark object detectors**

| Method | DatasetType | AP | $AP_{50}$ | $AP_{75}$ |
|---|---|---|---|---|
| Ours | RGB | 68.2 | 98.9 | 79.3 |
| Faster-RCNN | EventStatic | 85.4 | 99.0 | 98.8 |
| FCOS | EventStatic | 81.7 | 99.0 | 96.2 |
| TOOD(Anchor-based) | EventStatic | 83.7 | 99.0 | 97.8 |
| TOOD(Anchor-free) | EventStatic | 83.9 | 99.0 | 97.9 |
| Ours | EventStatic | **86.9** | **99.0** | **98.9** |
| Faster-RCNN | EventDynamic | 72.3 | 97.9 | 89.3 |
| FCOS | EventDynamic | 74.5 | 98.7 | 93.7 |
| TOOD(Anchor-based) | EventDynamic | 70.2 | 97.9 | 87.8 |
| TOOD(Anchor-free) | EventDynamic | 67.3 | 97.4 | 85.3 |
| Ours | EventDynamic | **76.5** | **99.0** | **96.6** |

***Implementaion details:*** In our study, we employed AdamW[10] as the optimizer, with a learning rate and weight decay both set to 0.0001. AdamW provided better generalization capability and stable training, while the chosen parameters promoted balance between convergence speed and model performance. Learning rate adjustments during training were managed by a combination of LinearLR and MultiStepLR schedulers. LinearLR linearly decreased the learning rate over time, while MultiStepLR made nuanced lr reductions at specified intervals. A batch size of 2 was used.

#### 5.1.2 *Results.*

***Quantitative evaluation:*** We evaluated the effectiveness of our proposed method using two datasets: the original RGB dataset and the synthetically generated EventStatic dataset. We utilized key performance indicators including Average Precision (AP), $AP_{50}$, and $AP_{75}$.

The results, as outlined in Table 3, indicate that our method showed significant improvement when applied to EventStatic dataset rather than RGB dataset. The result on the RGB dataset showed that our method achieved an AP of 68.2, $AP_{50}$ of 98.9, and $AP_{75}$ of 79.3. In contrast, the performance of our method on the EventStatic dataset was notably higher, with an AP of 86.9 and nearly perfect scores for $AP_{50}$ and $AP_{75}$ (99.0 and 98.9 respectively).

These findings underscore the effectiveness of our method in generating events from RGB frames and its adaptability to event camera datasets. Therefore, our approach offers a viable solution to the existing problem of data scarcity in the field of event-driven vision.

***Qualitative evaluation:*** Figure 5 presents inference results from both RGB and EventStatic modalities.

During the inference process, while utilizing the RGB image, our model encountered difficulties in accurately detecting the position of the ball. As a result, we were unable to obtain visualized results, due to the model's inability to allocate high scores to the ball's detection process. This problem stemmed primarily from two factors: the motion blur associated with the ball's movement, and the blending of the ball's color with the background, thereby hindering the model's ability to distinguish the ball effectively.

In contrast, in the corresponding EventStatic frame, the model showed superior performance. The unique characteristics of events, which capture luminance changes in the scene, facilitated the effective extraction of events related to the ball's movement. Consequently, the model was able to infer the ball's position accurately with high scores.

These findings suggest that Event-based methods could offer a valuable alternative for successful object detection in dynamic scenarios where RGB-based models face difficulties.

## 5.2 EventStatic vs. EventDynamic

In this experiment, the primary objective is to establish the effectiveness of our proposed method by comparing its performance against other existing techniques within the field. Furthermore, we aim to demonstrate that our method maintains its robustness and accuracy, even in dynamic scene conditions.

#### 5.2.1 *Experimental Settings.*

**Table 4: Comparison results between our method and the method by Glover and Bartolozzi**

| Method | DatasetType | RMSE of $x_{center}$ | RMSE of $y_{center}$ | RMSE of radius |
|---|---|---|---|---|
| Glover and Bartolozzi | EventStatic | 216.79 | 128.04 | 15.88 |
| Ours | EventStatic | **0.83** | **2.80** | **0.05** |
| Glover and Bartolozzi | EventDynamic | 99.23 | 85.53 | 31.64 |
| Ours | EventDynamic | **4.34** | **2.68** | **0.05** |
| Glover and Bartolozzi | EventReal-world | 112.32 | 117.85 | 22.00 |
| Ours | EventReal-world | **1.36** | **1.18** | **0.46** |

***Datasets:*** This experiment utilizes two datasets: the EventStatic dataset and the EventDynamic dataset. As discussed in Section 4.2, the EventStatic dataset serves as a controlled baseline, facilitating the assessment of our algorithms' performance in stable environments.

Detailed in Section 4.3, the EventDynamic dataset is generated using a method that encompasses the influence of camera movement on event generation. This distinct characteristic enables the evaluation of our algorithm's robustness under dynamic conditions, thereby simulating real-world scenarios. The composition of this dataset is comparable to that of the EventStatic dataset.

***Comparison methods:*** In this section, we present a comprehensive comparative study of various object detection techniques applied to both the EventStatic and EventDynamic datasets. Specifically, we contrast our approach with several other object detectors, including Faster R-CNN, FCOS anchor-free[19], TOOD anchor-based and anchor-free[3], and the conventional Hough transform-based method proposed by Glover and Bartolozzi[4], which is used for event-based ball detection.

Faster R-CNN employs a distinct two-stage mechanism. Initially, it isolates regions of interest (RoIs) using a Region Proposal Network (RPN). Following this, it deploys a second network for precise object detection. The FCOS anchor-free method, released around the same time as our proposed YOLOv3, eliminates the use of anchors and instead opts for a more direct bounding box prediction mechanism. TOOD, also known as Task-aligned One-stage Object Detection, offers both anchor-based and anchor-free variations. Unlike conventional methods that predict object location and class separately, TOOD enhances object detection accuracy by predicting object location and class concurrently. The method proposed by Glover and Bartolozzi enhances the detection of circles by incorporating geometric constraints into the Hough transform that receives events as an input.

For a fair and unbiased evaluation, we use the widely accepted MSCOCO[8] evaluation metric, Average Precision (AP), for the object detector assessment. However, for the comparison with the method by Glover and Bartolozzi, we employ the Root Mean Square Error (RMSE) metric instead of Average Precision (AP). This choice arises from the necessity to gauge the average discrepancy between the predicted and actual geometric parameters, such as the center or radius of a ball. The use of RMSE allows us to effectively measure and compare the precision of both the traditional Hough transform method and our proposed YOLOv3-based approach.

***Implementaion details:*** All object detectors were trained following the same experimental setup detailed in Section 5.1.1, specifically in the paragraph entitled Implementation details.

#### *5.2.2 Results.*

***Quantitative evaluation:*** The results from our extensive experiments, including both EventStatic and EventDynamic scenarios, are displayed in Table 3.

In the EventStatic scenario, wherein the camera position remains constant, our methodology exhibited remarkable superiority over several highly regarded methods, including FCOS, TOOD in both its anchor-based and anchor-free cases, and Faster-RCNN. In the context of Average Precision (AP), our method stood out, registering a leading score of 86.9, which is significantly higher than the values achieved by FCOS (81.7), TOOD anchor-based (83.7), TOOD anchor-free (83.9), and Faster-RCNN (85.4). These AP results represent prominent gains of 5.2, 3.2, 3.0, and 1.5 points over these comparison methods, respectively.

In the challenging EventDynamic scenario, characterized by dynamic changes in the camera viewpoint, our method displayed remarkable resilience, achieving an Average Precision (AP) of 76.5. This outperforms other object detectors such as Faster-RCNN (72.3), FCOS (74.5), and TOOD (both anchor-based and anchor-free variants, at 70.2 and 67.3 respectively).

These results indicate the robustness of our method over other models under diverse and challenging scenarios. Furthermore, these results suggest that the method has great potential for practical application in dynamical, real-world situations.

In addition, we evaluated the performance of the proposed method in the EventStatic and EventDynamic datasets using the root mean square error (RMSE) of the center coordinates $(x, y)$ and radius, and compared the performance of our method to the method by Glover and Bartolozzi. The results are shown in Table 4. In the EventStatic dataset, our method surpasses the method by Glover and Bartolozzi reducing RMSE for $x_{\text{center}}$, $y_{\text{center}}$, and radius from 216.79, 128.04, and 15.88 respectively, to 0.83, 2.80, and 0.05. This demonstrates our method's superior precision for both object localization and size estimation. In the EventDynamic dataset, our method also outperforms the method by Glover and Bartolozzi, with RMSE for $x_{\text{center}}$, $y_{\text{center}}$, and radius decreasing from 99.23, 85.53, and 31.64 to 4.34, 2.68, and 0.05 respectively.

In conclusion, our method consistently offers improved accuracy over the method by Glover and Bartolozzi for both object localization and size estimation, in both static and dynamic environments, demonstrating its robustness and precision.

## 5.3 Real-World data

In this experiment, we aim to illustrate the effectiveness of our object detector using event frames based on results obtained from real-world events captured by an event camera rather than synthetic events.

### 5.3.1 *Experimental Settings.*

***Datasets:*** To evaluate the real-world performance of our proposed method, we created our own dataset, EventReal-world dataset. This dataset is described in detail in Section 4.4.

***Comparison methods:*** We conduct a comparative analysis of our proposed method in comparison to method using events to detect balls, the method by Glover and Bartolozzi[4]. This method often encounters difficulties in maintaining robust performance. In contrast, our proposed method leverages a learning-based approach specifically optimized for asynchronous event processing, thereby significantly enhancing adaptability and performance. We utilized the Root Mean Square Error (RMSE) for assessing the comparative performance.

***Implementaion details:*** The experimental setup of the model used in this experiment is the same as described in Section 5.1.1, specifically in the paragraph entitled Implementation details.

### 5.3.2 *Results.*

***Quantitative evaluation:*** We evaluated the performance of our proposed method against the method by Glover and Bartolozzi[4] using real-world event-based datasets.

Table 4 presents the RMSE values for the center coordinates ($x_{center}$, $y_{center}$) and the radius for both the method by Glover and Bartolozzi and ours.

Our method significantly outperforms the method by Glover and Bartolozzi, demonstrating a substantial reduction in RMSE for both $x_{center}$ and $y_{center}$, as well as the radius. This substantial improvement underlines our method's superior capability in accurately localizing and sizing objects in event-based datasets.

These results underscore the potential applicability of our object detection model to various datasets and object types.

***Qualitative evaluation:*** Figure 6 shows the results of visualizing the inference outcomes of various methods contrasted against the ground truth, using sequential gray-scale images and corresponding event frames.

Our proposed technique demonstrates superior performance in the precise detection of a ball's location, exhibiting close alignment with the ground truth. However, the method by Glover and Bartolozzi displays significant false detections, erroneously identifying substantial circles at the extremities of feet and hands as balls.

These findings underscore the superior accuracy exhibited by our method, characterized by a lower incidence of false detections and a heightened precision in identifying target objects compared to the method by Glover and Bartolozzi.

## 6 CONCLUSION

This paper presents a method for detecting the position of a ball in sports videos using an event camera. The proposed method leverages the high temporal resolution of the event camera to accurately estimate the position of a fast-moving ball, a task challenging for conventional image-based methods. Given the scarcity and limited availability of event datasets, we propose a solution that repurposes an existing dataset comprising RGB frames for training the method. Experimental results show that the proposed method outperforms other methods in accurately detecting the ball's position. While this paper demonstrates the effectiveness of the proposed method specifically for volleyball, it has the potential to be adapted to other sports by modifying the training data accordingly.

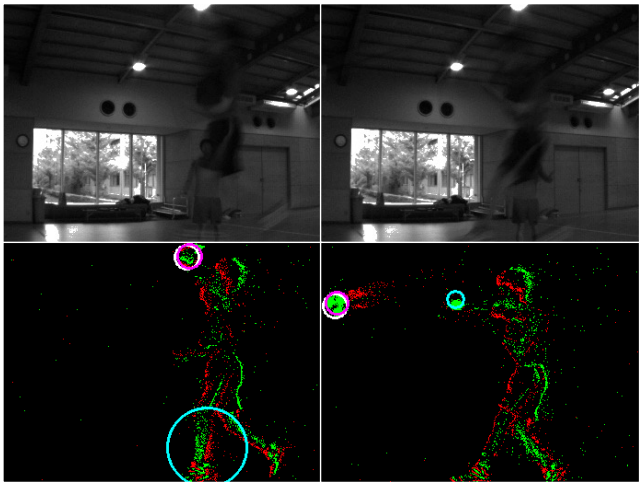

**Figure 6: Comparative analysis of ball detection results. Ground Truth is depicted by white lines, the method by Glover and Bartolozzi by cyan, and our method by magenta lines.**

## ACKNOWLEDGEMENT

This work was partially supported by JSPS KAKENHI Grant Number JP23H03422.

## REFERENCES

[1] Thulasya Banoth and Mohammad Farukh Hashmi. 2022. YOLOv3-SORT: detection and tracking player/ball in soccer sport. *Journal of Electronic Imaging* 32 (03 2022). https://doi.org/10.1117/1.JEI.32.1.011003

[2] S.A. Dyer and J.S. Dyer. 2001. Cubic-spline interpolation. 1. *IEEE Instrumentation & Measurement Magazine* 4, 1 (2001), 44–46. https://doi.org/10.1109/5289.911175

[3] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. [n. d.]. TOOD: Task-aligned One-stage Object Detection. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* ([n. d.]).

[4] Arren Glover and Chiara Bartolozzi. 2016. Event-driven ball detection and gaze fixation in clutter. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2203–2208.

[5] P. V. C. Hough. 1962. Method and means for recognizing complex patterns. U.S. Patent 3 069 654.

[6] Xin Jin, Longhai Wu, Jie Chen, Youxin Chen, Jayoon Koo, and Cheul-hee Hahm. 2023. A Unified Pyramid Recurrent Network for Video Frame Interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1578–1587.

[7] P R Kamble, A G Keskar, and K M Bhurchandi. 2019. A deep learning ball tracking system in soccer videos. *Opto-Electron. Rev.* 27, 1 (March 2019), 58–69.

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*. Springer International Publishing, 740–755.

[9] Yang Liu, Luiz G. Hafemann, Michael Jamieson, and Mehrsan Javan. 2021. Detecting and Matching Related Objects With One Proposal Multiple Predictions. In

*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 4520–4527.

[10] Ilya Loshchilov and Frank Hutter. 2017. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

[11] Zubair Martin, Sharief Hendricks, and Amir Patel. 2021. Automated Tackle Injury Risk Assessment in Contact-Based Sports - A Rugby Union Example. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 4594–4603.

[12] Banoth Thulasya Naik and Md Farukh Hashmi. 2022. YOLOv3-SORT: detection and tracking player/ball in soccer sport. *JEI* 32, 1 (March 2022), 011003.

[13] Naoki Nonaka, Ryo Fujihira, Monami Nishio, Hidetaka Murakami, Takuya Tajima, Mutsuo Yamada, Akira Maeda, and Jun Seita. 2022. End-to-end high-risk tackle detection system for rugby. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (New Orleans, LA, USA). IEEE.

[14] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. 2018. ESIM: an Open Event Camera Simulator. *Conf. on Robotics Learning (CoRL)* (Oct. 2018).

[15] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *ArXiv* abs/1804.02767 (2018).

[16] Vito Renó, Nicola Mosca, Roberto Marani, Massimiliano Nitti, Tiziana D'orazio, and Ettore Stella. 2018. Convolutional Neural Networks Based Ball Detection in Tennis Games. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2018), 1839–18396.

[17] Nitin J. Sanket, Chethan M. Parameshwara, Chahat Deep Singh, Ashwin V. Kuruttukulam, Cornelia Fermüller, Davide Scaramuzza, and Yiannis Aloimonos. 2020. EVDodgeNet: Deep Dynamic Obstacle Dodging with Event Cameras. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 10651–10657. https://doi.org/10.1109/ICRA40945.2020.9196877

[18] Binyi su, Lei Yu, and Wen Yang. 2020. Event-Based High Frame-Rate Video Reconstruction With A Novel Cycle-Event Network. In *2020 IEEE International Conference on Image Processing (ICIP)*. 86–90. https://doi.org/10.1109/ICIP40778.2020.9191114

[19] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. [n. d.]. FCOS: Fully Convolutional One-Stage Object Detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* ([n. d.]).

[20] Renaud Vandeghen, Anthony Cioppa, and Marc Van Droogenbroeck. 2022. Semi-supervised training to improve player and ball detection in soccer. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (New Orleans, LA, USA). IEEE.

[21] Gabriel Van Zandycke and Christophe De Vleeschouwer. 2019. Real-time CNN-based Segmentation Architecture for Ball Detection in a Single View Setup. In *Proceedings of the 2nd International ACM Workshop on Multimedia Content Analysis in Sports*.

[22] Xiaohan Zhang, Tianxiao Zhang, Yiju Yang, Zongbo Wang, and Guanghui Wang. 2020. Real-time Golf Ball Detection and Tracking Based on Convolutional Neural Networks. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2808–2813.