

# BlurBall: Joint Ball and Motion Blur Estimation for Table Tennis Ball Tracking

Thomas Gossard\*, Filip Radovic, Andreas Ziegler, Andreas Zell  
 University of Tuebingen  
 Sand 1 Tuebingen 72076, Germany  
 thomas.gossard@uni-tuebingen.de

## Abstract

*Motion blur reduces the clarity of fast-moving objects, posing challenges for detection systems, especially in racket sports, where balls often appear as streaks rather than distinct points. Existing labeling conventions mark the ball at the leading edge of the blur, introducing asymmetry and ignoring valuable motion cues correlated with velocity. This paper introduces a new labeling strategy that places the ball at the center of the blur streak and explicitly annotates blur attributes. Using this convention, we release a new table tennis ball detection dataset. We demonstrate that this labeling approach consistently enhances detection performance across various models. Furthermore, we introduce BlurBall, a model that jointly estimates ball position and motion blur attributes. By incorporating attention mechanisms such as Squeeze-and-Excitation over multi-frame inputs, we achieve state-of-the-art results in ball detection. Leveraging blur not only improves detection accuracy but also enables more reliable trajectory prediction, benefiting real-time sports analytics. Project page: <https://cogsys-tuebingen.github.io/blurball/>*

## 1. INTRODUCTION

Sports analytics has become a vital component in understanding performance, refining strategies, and enhancing athlete evaluation. Extracting meaningful insights from raw game footage has traditionally been both time-consuming and costly. Nevertheless, the benefits, ranging from improved performance metrics to deeper tactical understanding, are substantial. Recent advances in computer vision now enable automatic extraction of key game elements such as ball trajectories, player poses, and even racket orientation [24]. These capabilities not only support retrospective analysis but also enable predictive applications, for example, forecasting ball bounce positions from serve strokes [43], or identifying the characteristics of a winning



Figure 1. Motion blur frequently appears in broadcast footage but is typically disregarded. Yet, it offers valuable cues for estimating the ball’s velocity. The blue cross denotes the classical labeling approach, which introduces asymmetry and ambiguity to the detection task. We propose a refined annotation strategy: relabel the ball center to correspond to the middle of the blur (red cross) and include a directional blur label (green line) to capture motion information better.

rally [23]. Such techniques also open the door to automated umpiring systems based solely on video footage [41].

Moreover, this data is particularly promising for training table tennis robots. Despite remarkable progress [6, 15, 19, 35], incorporating human-like behavior remains challenging due to the lack of diverse, high-quality datasets. Ideally, the abundance of online match recordings could be leveraged to train robust, human-informed policies. For example, [8] introduced a 20-hour dataset collected from real matches and demonstrated its utility in simulation, leading to improved ball return prediction and enhanced robotic performance.

These applications rely on accurate and robust ball tracking, an essential prerequisite for intelligent analysis, forecasting, and autonomous behavior in high-speed sports environments. However, achieving this is far from trivial, especially in racket sports like table tennis where the ball generally moves at extreme speeds. The high speed of the ball often results in motion blur, transforming its appearance from a distinct dot into a streak of varying length. This

\*The work was partially funded by Sony AI

motion blur not only complicates the detection process but also raises the question of how to define the ball’s position accurately when the blur is present.

Traditionally, the convention for racket sports [12, 32] has been to define the ball’s position as the leading edge (or front) of the blur streak as shown in Figure 1. However, this approach introduces ambiguities, such as determining which end of the streak is the front, and results in a non-symmetric representation that can complicate both detection and tracking. A common workaround has been to provide the model with multiple sequential frames, allowing it to infer motion implicitly. Although often dismissed as visual noise, motion blur inherently captures information about the ball’s velocity and direction of motion. When properly leveraged, this visual cue becomes a powerful signal, enhancing both the robustness of ball tracking and the accuracy of trajectory prediction.

In this paper, we address the challenges of detecting fast-moving balls in high-speed sports like table tennis by explicitly modeling motion blur, a commonly ignored but informative visual cue. Table tennis balls can reach speeds of up to 35m/s, making precise localization difficult, especially in blurred frames. To address these challenges, we present the following contributions:

- **BlurBall:** A model that jointly predicts ball position and motion blur, achieving state-of-the-art performance.
- **Blur-aware labeling:** A new annotation scheme that defines ball position at the blur center and encodes blur attributes, improving accuracy and resolving ambiguities.
- **Table tennis ball dataset:** We extend the field of sport ball detection and tracking to table tennis.

## 2. RELATED WORK

**Ball detection and tracking** have traditionally been approached using classical computer vision techniques, such as the use of color filtering, background subtraction, and blob detection [1, 13, 33, 35, 41]. While these methods can be effective in controlled environments, they often struggle with variations in lighting, camera angles, and background clutter, requiring extensive fine-tuning to maintain performance. Post-processing techniques can improve their robustness by enforcing physical constraints, such as filtering detections based on physically plausible trajectories [33]. However, these rule-based approaches remain sensitive to environmental changes and do not generalize well across different sports settings.

With the advancement of deep learning, neural network-based models have demonstrated greater robustness to

dynamic environments and varying illumination conditions [47]. Some approaches repurpose general object detectors like Detectron2 [2] and YOLOv4 [18], fine-tuning them on in-house sports datasets. However, these models are not ideal for small, fast-moving objects like sports balls, which often appear as single instances per frame and can be heavily blurred. To address these limitations, several specialized approaches have emerged that treat ball detection as a heatmap regression problem. DeepBall [16] was an early example using fully convolutional networks to generate ball-centered heatmaps, demonstrating effectiveness in soccer videos. BallSeg [44] is a modified version of IC-Net [46] applied to basketball videos. TrackNet [12], developed for tennis, introduced the use of Gaussian heatmaps to represent the probability distribution of ball locations. Both previous methods introduced multiple consecutive input frames for implicit background subtraction and used the ball dynamics for filtering. TrackNetV2 [32] extended this approach with a U-Net backbone [26] and adopted a multiple-input multiple-output (MIMO) design to jointly detect ball positions across several frames, improving robustness and inference speed. Monotrack [22] further enhanced detection by adding residual connections and replacing the focal loss with a weighted combination of the Dice loss and the binary cross-entropy loss to better handle the extreme class imbalance caused by the ball’s small size. TrackNetV3 [5] addressed occlusion by incorporating a trajectory rectification module and an inpainting mechanism to recover missing detections. The Widely Applicable Strong Baseline (WASB) [34], switched from a U-Net to an HRNet [38] backbone for improved results. Another key insight from it is that reducing the step size in the MIMO setup, effectively oversampling by shifting input windows by one frame instead of three, leads to notably better detection performance, albeit at the cost of slower inference.

**Attention Mechanisms.** All the previously mentioned methods rely on CNNs, but transformers [36] have proven highly effective for vision tasks such as segmentation and object detection, particularly with the introduction of Vision Transformers (ViTs) [7, 36]. While applying a full transformer architecture may be excessive for the task of detecting a single small object like a ball, hybrid CNN-transformer models have emerged to capture temporal dependencies more effectively. TrackNetV4 [25] builds on top of TrackNetV3 and introduces motion attention maps and a motion-aware fusion mechanism. For table tennis, Li et al. [20] proposed a detector that integrates a transformer module with a CNN to leverage temporal context. However, their dataset and code are not publicly available, which currently limits reproducibility.

As a lighter alternative, attention mechanisms within CNNs have gained traction for improving feature representation and detection accuracy. While these mechanisms typ-

ically operate across channels, in the context of multi-frame inputs, as used in ball detectors—channel wise attention can implicitly act as temporal attention, helping the model better capture motion cues and leverage temporal information. We mostly focus on the following mechanisms. The Squeeze-and-Excitation (SE) block [11] enhances performance by adaptively recalibrating channel-wise features. It generates a channel attention vector by applying global average pooling, followed by two fully connected layers with a ReLU activation, allowing the network to emphasize informative channels and suppress less relevant ones. Efficient Channel Attention (ECA) [39] simplifies this process by removing the dimensionality reduction step and replacing the fully connected layers with a lightweight 1D convolution, enabling efficient local cross-channel interaction. Coordinate Attention (CA) [10] extends channel attention by embedding spatial information through directional pooling along the horizontal and vertical axes. This allows the network to capture both content and location. We propose to integrate these attention mechanisms to improve the ball detection performance. Other attention mechanisms like Convolutional Block Attention Module (CBAM) [42], Global Context (GC) blocks [3], and non-local blocks [40] have also demonstrated improvements in object detection by combining channel and spatial attention or modeling long-range dependencies. However, due to their higher computational overhead and focus on broader contextual modeling, we leave their integration for future work and concentrate on methods most compatible with real-time, high-precision detection in high-speed sports scenarios.

**Racket sport datasets.** The OpenTTGames dataset [37] is a publicly available resource for table tennis ball detection, collected under highly controlled conditions with high frame rates (120 fps), 1080p resolution, close-up views, and stable lighting. However, such ideal conditions are rarely found in real-world broadcasts, where lower frame rates, lower resolutions, and inconsistent lighting make detection significantly more challenging. Additionally, the dataset suffers from unbalanced camera color calibration, resulting in unnatural tones that hinder generalization to real-world footage. Datasets for fast-moving objects (FMO) also include some table tennis footage [17, 29], but with fewer than ten rallies they are insufficient for training; the same holds for the synthetic VOT-FMO dataset [31]. Ball detection datasets exist for other sports such as badminton [32] and tennis [12]. Among them, the badminton shuttlecock is visually most similar to a table tennis ball, and models trained on badminton data [34] can detect table tennis balls to some extent. However, we observed that their detection accuracy on table tennis footage is significantly lower than the reported performance for badminton, highlighting the domain gap between the two tasks. This highlights the necessity of a dedicated dataset for table tennis.

**Blur Estimation.** Research on fast-moving objects has shown that motion blur can be explicitly modeled and exploited for trajectory reconstruction. Rozumnyi et al. [29] introduced the first dataset and tracker for FMOs, where objects appear as blurred streaks. This line of work was extended with Tracking by Deblurring [17, 27, 28], which combines blind deblurring and image matting to recover intra-frame trajectories from blur. Further extensions enabled 3D reconstruction, using object size as a depth cue [30, 45]. Angular velocity estimation becomes also possible by finding the 3D rotation that best aligns successive blurred appearances [30]. More recently, FMODetect [31] proposed a learning-based detector trained on synthetic FMOs, achieving more efficient detection, trajectory estimation, and appearance recovery. However, most of these methods require labor-intensive labeling, such as high-speed camera annotations [17] or manual blur masks [29]. In contrast, for volleyball, Chao et al. [4] attempted to exploit blur implicitly by constructing velocity heatmaps from consecutive frames, but without explicitly modeling the underlying blur properties. Unlike prior FMO methods, we assume blur follows a straight line, which simplifies dataset labeling and allows us to annotate a large number of diverse videos for training our model.

### 3. METHOD

Traditional sports ball detectors are typically designed to output only the ball’s position, overlooking other valuable information available in a single frame, such as player positions and the layout of the playing field. Among these, motion blur provides direct insight into the ball’s velocity. Motion blur occurs when an object moves significantly during the camera’s exposure time, causing its appearance to stretch into a streak. The magnitude of the blur is determined by the ball’s velocity and the exposure time: shorter exposures produce minimal blur, while longer exposures result in more pronounced streaks.

In this section, we first introduce our newly collected and labeled table tennis ball detection dataset, which follows our proposed labeling convention incorporating motion blur information. We then describe how this dataset is used to train BlurBall, our model designed to jointly estimate both the ball’s position and its associated motion blur properties.

#### 3.1. Dataset

A comprehensive table tennis dataset was created using footage from 26 different online recordings, spanning both amateur and professional games. 64,119 frames were gathered in total. In an effort to enhance the diversity of the dataset, we varied the point of view, table color schemes, lighting, and general environmental conditions as much as possible, as shown in Figure 2. All videos were recorded from fixed, static camera positions.

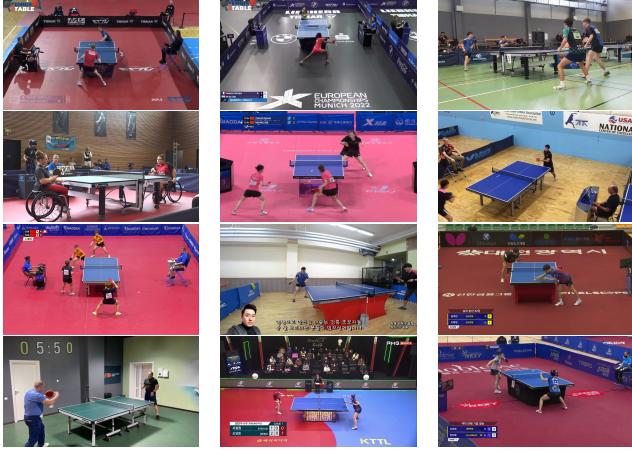


Figure 2. Example scenes from our dataset, showcasing a diverse range of contexts to ensure comprehensive coverage.

	Games	Clips	Frames	disp. [px]	Blur ratio
Train	22	363	51423	$18.7 \pm 16.4$	0.58
Test	4	80	12696	$20.1 \pm 12.1$	0.77

Table 1. Table tennis dataset description. disp. is the mean ball displacement between frames. Blur ratio is the fraction of observations where the ball has motion blur.

Previous racket sport datasets [12, 32] annotate the ball position at the leading edge of the motion blur streak. In contrast, we propose a new labeling convention in which the ball position is defined as the midpoint of the blur streak. Furthermore, we extend the labeling to include additional information about the blur streak, specifically its length and orientation as shown in Figure 3. For each frame, we manually generated labels for the ball’s position, orientation, and blur length, represented as  $[p_b, \theta, l]$ , where  $p_b = (x_b, y_b)$  denotes the ball center. These labels were obtained by manually drawing a line along the blur streak. This is more time-effective compared to drawing the blur streak mask for each frame [29] or labelling all the frames from high-speed cameras [17]. Approximating the blur streak as a straight line is justified, since the short exposure time ensures that the ball trajectory does not deviate significantly during image capture, except in specific rare cases as mentioned in Section 4.6. When motion blur is present, we define the angle  $\theta$  and the half-length  $l$  of the blur. The two extremities of the blur streak are then calculated as follows:

$$\begin{aligned} p_1 &= p_b + (l \cos \theta, l \sin \theta) \in \mathbb{R}^2, \\ p_2 &= p_b - (l \cos \theta, l \sin \theta) \in \mathbb{R}^2 \end{aligned} \quad (1)$$

For comparison, we additionally include the traditional leading-edge labels i.e. the front of the ball blur streak.

Table 1 contains statistics that describe the generated dataset. We observe motion blur in 62% of the frames. The

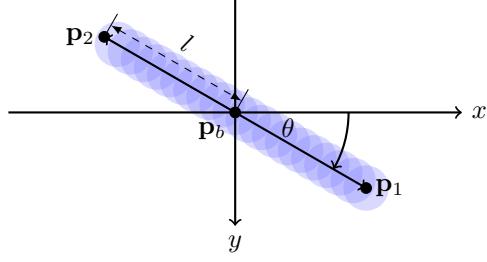


Figure 3. **Motion blur labeling schematic.** Conventional labels mark the front blur edge  $p_1$ , which may be confused with  $p_2$  without motion context. We propose labeling the blur center  $p_b$  with its half-length  $l$  and orientation  $\theta$ .

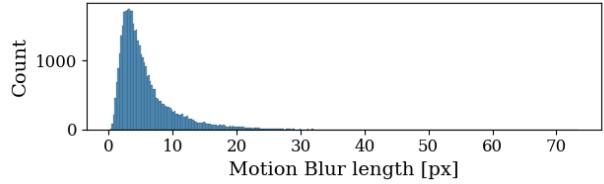


Figure 4. Distribution of the half-lengths  $l$  of the motion blur streak in our dataset. The maximum recorded value was 73 pixels.

distribution of motion blur lengths is shown in Figure 4, where we can see that the blur length is primarily concentrated around  $l = 5$  pixels, although it can extend up to 73 pixels in some rare scenarios.

In addition to ball positions, we provide camera calibration data for each game, enabling 3D reconstruction of the ball trajectory [9, 14]. Specifically, we compute the camera’s extrinsic parameters, rotation  $R$  and translation  $T$ , relative to the world frame, which is defined by the table tennis table.

### 3.2. Ball Detector

Among existing ball detection and tracking methods, WASB [34] showed strong performance for tennis and badminton. In WASB, an HRNet backbone generates heatmaps of likely ball locations, from which connected components above a threshold  $\delta = 0.5$  are selected as candidates. A tracker then refines these candidates using tracklets from previous frames, ensuring temporal consistency and suppressing false positives. We therefore adopt WASB as the basis of our model. However, to effectively incorporate motion blur information, we introduced modifications to its training procedure.

Our key modification lies in the GT heatmap design. In WASB, a binary map is generated using the following equation:

$$y_{\mathbf{p}}^{bin} = \begin{cases} 1 & \text{if } \|\mathbf{p} - \mathbf{p}^{GT}\| \leq d \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where  $\mathbf{p}^{GT}$  denotes the ground truth ball position,  $y_{\mathbf{p}}^{bin}$  represents the value of the GT map at location  $\mathbf{p}$ , and  $d$  is a distance threshold defining the radius of the disk. In contrast, we redefine the ground truth (GT) heatmap to encompass the entire motion-blurred region of the ball as follows:

$$y_{\mathbf{p}}^{bin} = \begin{cases} 1 & \text{if } \exists \mathbf{p}' \in [\mathbf{p}_1^{GT}, \mathbf{p}_2^{GT}] \text{ s.t. } \|\mathbf{p} - \mathbf{p}'\| \leq d \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

In this formulation, all pixels within the blur streak—ranging from the detected start point  $\mathbf{p}_1$  to the endpoint  $\mathbf{p}_2$ —are assigned a value of 1, effectively capturing the full extent of the motion blur of the ball. This adaptation ensures that the model learns to localize not only the ball’s center but also the full extent of its motion blur.

To further refine the learning process, we adapted the Hard-to-Localize Sample Mining (HLSM) of WASB [34] to our modified GT heatmap. The original formulation of the GT heatmap is:

$$y_{\mathbf{p}}^{real} = \begin{cases} \min\left(C \exp\left(-\frac{\|\mathbf{p} - \mathbf{p}^{GT}\|^2}{d^2}\right), 1\right) & \text{if } \|\mathbf{p} - \mathbf{p}^{GT}\| \leq d \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $y_{\mathbf{p}}^{real}$  is the value of the real-valued GT map at  $\mathbf{p}$  and  $C$  is determined so that the non-zero minimum value is set to a predefined value  $c_{min}$ . Similarly to the binary map, we extend the real-valued heatmap to encompass the blur with:

$$y_{\mathbf{p}}^{real} = \begin{cases} \max_{\mathbf{p}' \in [\mathbf{p}_1^{GT}, \mathbf{p}_2^{GT}]} \min\left(C \exp\left(-\frac{\|\mathbf{p} - \mathbf{p}'\|^2}{d^2}\right), 1\right) & \text{if } \exists \mathbf{p}' \in [\mathbf{p}_1^{GT}, \mathbf{p}_2^{GT}] \text{ s.t. } \|\mathbf{p} - \mathbf{p}'\| \leq d \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The network was trained using the quality focal loss [21], which helps the model focus on challenging regions where ball localization is difficult. For inference, the center of the blur is calculated in a manner similar to WASB [34], where it is obtained as the weighted mean of the heatmap. This ensures that the center is accurately located based on the intensity of the detected blur.

To further improve the HRNet, we incorporate the attention mechanisms SE, CA, and ECA, previously discussed in Section 2. While SE and ECA improve channel-wise feature discrimination, capturing temporal cues in our multi-frame input with minimal computational overhead, CA additionally encodes positional information, enabling more precise and spatially aware attention. As shown in Section 4, SE yields the most consistent performance improvement, and we choose this as our attention mechanism. We refer to our model based on HRNet with SE and trained for joint ball detection and blur estimation as **BlurBall**. The training data and output predictions of BlurBall, including both ball localization and blur information, are visualized in Figure 5.

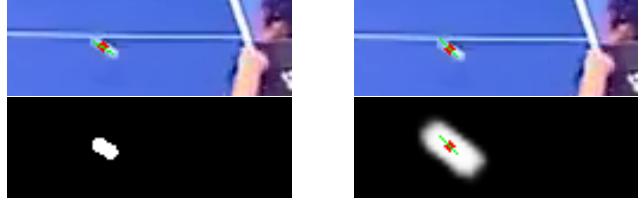


Figure 5. Left: Label and target heatmap. Right: Model prediction overlay with estimated ball center (red) and blur (green).

### 3.3. Estimating motion blur

Once the trained BlurBall model produces a heatmap highlighting ball pixels, including those affected by motion blur, the next step is to process this heatmap to extract the blur parameters: orientation and length. To isolate the blur region, we first apply a thresholding operation to the predicted heatmap. By default, pixels with heatmap values over  $\delta$  are considered part of the blur region. To determine the direction of motion blur, we apply Principal Component Analysis (PCA) to the set of pixel coordinates within the extracted mask. PCA identifies the principal axis  $\mathbf{B}$ , which corresponds to the eigenvector associated with the largest eigenvalue. This axis represents the primary direction of elongation in the blur streak and serves as an estimate of the ball’s motion direction during the camera’s exposure time. The orientation angle  $\theta$  of the blur is then computed as:

$$\theta = \arctan\left(\frac{B_y}{B_x}\right), \quad (6)$$

which gives the blur direction relative to the horizontal axis.

To determine the blur length, we project all pixel coordinates  $\mathbf{p}$  from the mask onto the principal axis. The range of these projections represents the total extent of the blur along its dominant direction.

$$l = \frac{\max_{\mathbf{p} \in \text{mask}}(\mathbf{B} \cdot \mathbf{p}) - \min_{\mathbf{p} \in \text{mask}}(\mathbf{B} \cdot \mathbf{p})}{2} \quad (7)$$

By extracting these two parameters, orientation ( $\theta$ ) and length ( $l$ ), we effectively recover the motion blur characteristics, which provide direct insight into the velocity of the ball.

In [32, 34], the confidence score is calculated as the sum of the heatmap values within the detected blob region. However, since our goal is to detect motion blur rather than the precise ball position, this approach tends to favor longer blur streaks, as the summed value naturally increases with streak length rather than signal strength. To address this bias, BlurBall instead uses the mean value of the heatmap blob as the confidence score, making it more robust to variations in blur extent.

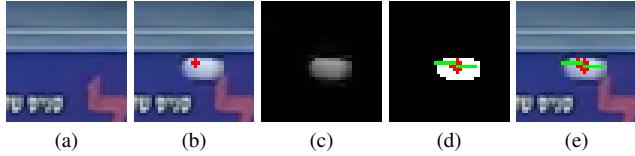


Figure 6. Baseline Method for Motion Blur Estimation. From left to right: (a) Frame  $n - 2$ , (b) Frame  $n$  with the estimated ball position, (c) Absolute difference between frames  $n$  and  $n - 2$ , (d) Binary mask obtained after thresholding, highlighting the estimated blur region, and (e) Frame  $n$  with the detected blur overlayed.

## 4. EXPERIMENTS

### 4.1. Blur estimation baseline

Existing ball detectors can be leveraged to extract the ball’s blur streak by incorporating additional standard computer vision techniques. To establish a baseline for comparison, we implemented a method that detects motion blur using traditional image processing techniques, similar to [29]. First, we define a small region of interest (ROI) centered around the detected ball position. The input RGB image is then converted to grayscale, and we compute the frame difference to highlight regions of motion. To suppress noise, a thresholding operation is applied, retaining only significant intensity changes. The resulting binary mask might contain multiple connected components, among which we identify the one closest to the ROI center as the blur streak corresponding to the ball’s motion. An example of this process is illustrated in Figure 6. Although this approach is not perfect, it could also be used to generate new labels using our convention for existing sports ball datasets. While this approach provides a straightforward way to estimate motion blur, it is inherently limited to static cameras and is less robust than BlurBall in dynamic environments.

### 4.2. Training

BlurBall was trained from scratch for 30 epochs using Adam. We experimented with pretraining the network with the badminton dataset [32], which is the closest looking sport to table tennis, but this did not yield any improvements. The batch size was set to 8 and the ball detector takes as input images rescaled to 288 x 512. We set  $d$  to 2.5 and  $c_{min}$  to 0.7 and start running HLSM at the beginning of the 20th epoch, similarly to the original WASB implementation. Experiments were conducted on an NVIDIA GeForce RTX 2080 Ti.

### 4.3. Ball Tracking

The effect of different labeling conventions on the performance of ball detectors was investigated. Table 2 shows

<sup>1</sup><https://github.com/Chang-Chia-Chi/TrackNet-Badminton-Tracking-tensorflow2>

the performance of several sports ball detectors trained with two distinct labeling strategies: one where the ball position is labeled at the front edge of the blur streak (Front), and another where the ball’s position is defined at the midpoint of the blur (Mid.). We used a fixed distance threshold of  $\tau = 4 \text{ px}$  for the benchmark, as motivated in [34]. WASB (1-step) achieves the best performance across all metrics except recall. TrackNetV3 attains near-perfect recall, primarily due to its trajectory inpainting and rectification module. Using the midpoint convention consistently improves detection performance across all models. These results indicate that ball detectors perform better when the motion blur streak is symmetric around the ball label, leading to more accurate localization. However, we did not notice any improvement in training speed with regard to the labeling convention used. Given these findings, we recommend adopting this midpoint labeling convention for ball detectors, particularly in fast-paced sports where motion blur is common. This approach enhances the robustness and accuracy of detection systems in such dynamic environments.

As shown in Table 3, incorporating blur information into WASB with a 3-step setting slightly decreases the F1 score and accuracy, yet it improves the AP. In contrast, for the 1-step setting, adding blur information consistently enhances all metrics, suggesting that using the blur from adjacent frames helps infer the ball’s likely location in the center frame more accurately.

Introducing attention mechanisms improves performance across all metrics compared to the blur-augmented baseline. Among them, the SE block yields the most consistent improvements. Its ability to globally recalibrate channel-wise feature responses allows it to emphasize both motion and appearance features. In comparison, CA incorporates positional information through coordinate attention, which can be beneficial for larger or spatially consistent objects. However, in our case where the target is small, fast-moving, and lacks stable spatial patterns, CA’s added complexity and use of per-axis pooling and normalization may introduce noise, particularly in the 3-step setup. ECA offers a more lightweight alternative by using local 1D convolutions without dimensionality reduction, but this comes at the cost of only modeling short-range dependencies. In practice, it lacks the global context modeling needed for robust motion blur interpretation. While CA and SE achieve similar performance in the 1-step setting, SE is both slightly faster and more robust across settings. For these reasons, we select SE as the attention mechanism for BlurBall.

Although WASB achieves the highest accuracy in the 1-step setting, its precision drops compared to the 3-step version. This suggests that the model is overconfident in the central frame predictions and that the default threshold of  $\delta = 0.5$  is suboptimal. As shown in Figure 7, the threshold has little influence for 3-step inference, but for 1-step

Model	F1		Acc		AP		Recall		Precision		#Params (M)	FPS
	Front	Mid.										
DeepBall [16]	39.36	57.14	26.21	41.36	42.96	61.40	86.86	91.52	25.45	41.54	0.1	65
DeepBall-Large	52.40	71.72	36.93	56.66	55.42	75.03	89.82	94.12	36.99	57.94	1.1	51
BallSeg [44]	85.09	89.01	75.47	81.33	75.58	81.56	63.67	83.56	27.15	35.94	12.7	40
TrackNetV2 [32]	90.37	91.61	83.25	85.27	85.07	86.88	89.79	88.38	90.96	95.09	11.3	85
ResTrackNetV2 <sup>1</sup>	87.03	91.81	78.11	85.61	75.53	86.58	86.76	89.02	87.90	94.78	1.2	112
Monotrack [22]	93.60	94.97	88.55	90.89	87.97	91.26	93.13	94.93	94.07	<b>96.36</b>	2.9	<b>115</b>
TrackNetV3 [5]	93.74	95.93	88.35	92.28	N.A.*	N.A.*	<b>99.64</b>	<b>99.66</b>	88.50	92.48	17.8	77
WASB (steps=3) [34]	94.23	95.77	89.62	92.26	91.82	94.66	93.89	95.23	<b>94.57</b>	96.31	1.5	95
WASB (steps=1) [34]	<b>95.58</b>	<b>96.00</b>	<b>91.90</b>	<b>92.56</b>	<b>95.10</b>	<b>96.85</b>	97.08	97.77	94.15	94.50	1.5	42

Table 2. Comparison of different models’ performance depending on the labeling convention. \* AP is not available for TrackNetV3 because the rectification module doesn’t provide a confidence value.

Model	F1		Acc		AP		MAE l [px]		MAE $\theta$ [deg]		#Params (M)	FPS for 3 step
	1 step	3 steps	1 step	3 steps	1 step	3 steps	1 step	3 steps	1 step	3 steps		
WASB [34] (baseline)	96.00	95.77	92.56	92.26	96.85	94.66	3.1±3.4*	3.0±3.2*	13.5±23.2*	14.8±24.6*	1.48	95
+ Blur label	96.28	95.54	93.05	91.81	97.59	95.40	1.5±1.2	1.4±1.2	6.5±18.2	7.2±20.1	1.48	95
+ ECA	96.35	95.60	93.18	91.92	97.91	95.67	1.8±1.3	1.8±1.3	<b>6.4±18.5</b>	7.0±19.8	1.48	79
+ CA	96.50	95.76	93.48	92.23	97.24	94.82	1.4±1.2	1.4±1.2	6.5±18.6	<b>6.7±18.8</b>	1.50	72
+ SE → BlurBall (ours)	96.52	<b>96.16</b>	93.47	<b>92.89</b>	<b>98.23</b>	<b>96.72</b>	1.5±1.2	1.6±1.2	6.5±18.9	6.9±19.5	1.49	79
BlurBall with $\delta = 0.7$	<b>97.17</b>	96.12	<b>94.75</b>	<b>92.89</b>	97.34	94.80	<b>1.2±1.1</b>	<b>1.2±1.1</b>	6.8±18.9	7.3±19.9	1.50	72

Table 3. Performance comparison starting from the WASB baseline, showing the effect of replacing the position-only heatmap with blur-aware heatmaps and evaluating different attention mechanisms. BlurBall, using SE attention, achieves the best overall performance. \*WASB blur masks are obtained via background subtraction. Bold values indicate the best result in each column.

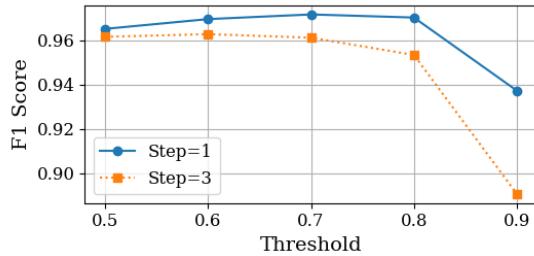


Figure 7. F1 score for BlurBall with different threshold values  $\delta$  on the test set.

the optimal value shifts to  $\delta = 0.7$ . With this adjustment, BlurBall obtains the best F1-score and accuracy (Table 3), though at the cost of a slight decrease in AP and blur estimation quality.

#### 4.4. Blur Estimation

We evaluate and compare the accuracy of BlurBall and WASB in Table 3. The motion blur estimation performance is assessed using the Mean Absolute Error (MAE) for both blur length  $l$  and blur angle  $\theta$ . For small blur lengths, particularly when  $l \leq 3$ , the estimation of the blur angle becomes highly sensitive to noise, rendering the measurements unreliable. To ensure a robust evaluation, the MAE for the blur

angle is computed only for predictions where the estimated blur length exceeds 3 pixels. Under this constraint, BlurBall proves significantly more accurate, achieving nearly twice the precision of WASB combined with the baseline blur estimation. Among the attention mechanisms tested, SE and CA improve blur estimation, whereas ECA does not offer any notable benefit.

#### 4.5. Ball Trajectory Prediction

In this subsection, we demonstrate that motion blur provides valuable cues for predicting the ball’s trajectory. The ball’s 3D trajectory can be reasonably approximated using a second-degree polynomial. This is based on the assumption that the forces acting on the ball—gravity, drag, and Magnus force—remain relatively constant while the ball is airborne. Consequently, the 2D projection of the trajectory in the image plane can also be modeled as a second-degree polynomial. We represent the ball’s image coordinates over time as  $(P_x(t), P_y(t))$ , where  $P_x$  and  $P_y$  are quadratic polynomials with respect to time.

In this context, we show that incorporating motion blur information improves trajectory prediction while the ball is in flight. Motion blur is directly linked to the ball’s speed and camera exposure time, allowing us to relate the observed blur length  $l$  and angle  $\theta$  to the derivative of the tra-

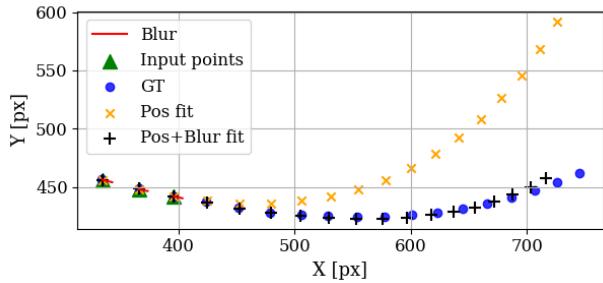


Figure 8. Trajectory prediction comparison using position-only vs. position+blur fitting. Polynomials are fitted using the first three observations to predict the ball’s trajectory. Predictions are compared to the ground truth (GT) positions. Red lines indicate observed motion blur.

jectory as follows:

$$\dot{P}_x(t) = \frac{l \cos(\theta)}{t_{\exp}}, \quad \dot{P}_y(t) = \frac{l \sin(\theta)}{t_{\exp}} \quad (8)$$

For evaluation, we use BlurBall to infer both the ball position and motion blur. To predict the ball’s trajectory, we fit a second-degree polynomial using only the first 3 frames of each sequence—the minimum required to estimate a quadratic curve. Trajectory fitting based solely on position is performed using a standard least-squares method. For position + blur fitting, we minimize the following cost functions using the Nelder-Mead algorithm to obtain the polynomial coefficients and  $t_{\exp}$ :

$$J_x = \frac{1}{3} \sum_{k=0}^2 \|P_x(t_k) - \hat{x}_k\|^2 + 0.2 \left\| \dot{P}_x(t_k) - \frac{l_k \cos(\theta_k)}{t_{\exp}} \right\|^2 \quad (9)$$

$$J_y = \frac{1}{3} \sum_{k=0}^2 \|P_y(t_k) - \hat{y}_k\|^2 + 0.2 \left\| \dot{P}_y(t_k) - \frac{l_k \sin(\theta_k)}{t_{\exp}} \right\|^2 \quad (10)$$

Here,  $(\hat{x}_k, \hat{y}_k)$  are the detected positions,  $l_k$  is the estimated blur length, and  $\theta_k$  is the blur angle for each frame  $k$ . We weight the blur loss with a coefficient of 0.2. After fitting the polynomial, we compare the predicted trajectory to the ground truth positions for the remainder of the ball’s flight. This evaluation is performed on 95 manually segmented trajectories from the test set. An example of the predicted trajectory is shown in Figure 8.

We obtain a MAE of  $84.4 \pm 136.6$  pixels for the trajectory fitting using position only, and a significantly lower MAE of  $53.0 \pm 87.1$  pixels when incorporating both position and blur information. The notable reduction in both mean and variance of the prediction error demonstrates that motion blur encodes meaningful velocity information. Incorporating these cues leads to more accurate and consistent

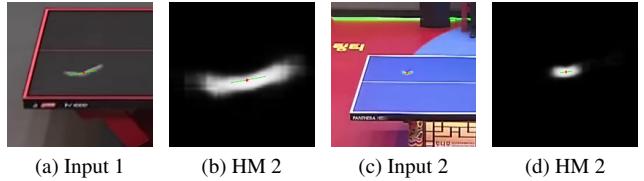


Figure 9. Frames and corresponding inferred heatmaps where the bounce is exactly captured by the camera.

trajectory estimates, particularly in fast-motion scenarios or when only a few frames are available.

#### 4.6. Limitations

BlurBall has certain limitations. One key assumption in our approach is that motion blur appears as a single linear streak. However, in rare cases, such as when the exact moment of a ball bounce is captured, this assumption no longer holds, as illustrated in Figure 9. The ball is still detected, but the estimated position is inaccurate, as the detector takes the center of a non-convex blob, distorted due to the bounce. Still, the outputted heatmap seems to have captured the change in direction to some degree, as shown by the bend in the heatmap blob. Also, although the model can work for videos with non-static cameras, we observed that it is more prone to misdetections. This introduces potential challenges in accurately estimating the ball’s motion in such cases.

While our current model achieves state-of-the-art ball detection for table tennis, some failure cases remain and highlight areas for future improvement. A common source of false positives involves white objects being misidentified as the ball. This typically includes shoes, hands, or logos on uniforms that resemble the ball in size, shape, or motion. To reduce false positives, future work could include more context from the whole scene to help tell the ball apart from similar-looking objects.

## 5. CONCLUSION

In this paper, we introduced a novel labeling convention for sports ball detectors that explicitly incorporates motion blur, significantly improving detection performance across all models tested on our newly introduced table tennis dataset. We also presented BlurBall, a detector capable of jointly estimating ball position and motion blur, offering valuable cues about the ball’s velocity and enabling more accurate trajectory prediction. Our model achieves competitive performance by integrating attention mechanisms, which benefit from multi-frame inputs. Our approach enables more precise and robust tracking, with strong potential for broader applications in high-speed sports analytics and real-world deployments.

## References

- [1] M. Archana and M. Kalaisevi Geetha. Object Detection and Tracking Based on Trajectory in Broadcast Tennis Video. *Procedia Computer Science*, 58:225–232, Jan. 2015. 2
- [2] J. Calandre, R. Péteri, L. Mascarilla, and B. Tremblais. Extraction and analysis of 3D kinematic parameters of Table Tennis ball from a single camera. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9468–9475, Milan, Italy, Jan. 2021. IEEE. 2
- [3] Yue Cao, Jiarui Xu, Stephen Lin, Fangyun Wei, and Han Hu. Global context networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):6881–6895, 2020. 3
- [4] V. Chao, H. Nguyen, A. Jamsrandorj, Y. Oo, K. Mun, H. Park, S. Park, and J. Kim. Tracking the Blur: Accurate Ball Trajectory Detection in Broadcast Sports Videos. In *Proceedings of the 7th ACM International Workshop on Multimedia Content Analysis in Sports*, pages 41–49, Melbourne VIC Australia, Oct. 2024. ACM. 3
- [5] Yu-Jou Chen and Yu-Shuen Wang. Tracknetv3: Enhancing shuttlecock tracking with augmentations and trajectory rectification. In *Proceedings of the 5th ACM International Conference on Multimedia in Asia, MMAAsia ’23*, New York, NY, USA, 2024. Association for Computing Machinery. 2, 7
- [6] David B D’Ambrosio, Saminda Wishwajith Abeiyruwan, Laura Graesser, Atil Iscen, Heni Ben Amor, Alex Bewley, Barney Reed, Krista Reymann, Leila Takayama, Yuval Tassa, et al. Achieving human level competitive robot table tennis. In *7th Robot Learning Workshop: Towards Robots with Human-Level Abilities*, 2024. 1
- [7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021. 2
- [8] Daniel Etaat, Dvij Kalaria, Nima Rahamanian, and S Shankar Sastry. Latte-mv: Learning to anticipate table tennis hits from monocular videos. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7115–7124, 2025. 1
- [9] Thomas Gossard, Andreas Ziegler, and Andreas Zell. Tt3d: Table tennis 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025. 4
- [10] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13708–13717, 2021. 3
- [11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 3
- [12] Y. Huang, I. Liao, C. Chen, T. İk, and W. Peng. TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, Sept. 2019. 2, 3, 4
- [13] C. Hung. A Study of Automatic and Real-Time Table Tennis Fault Serve Detection System. *Sports*, 6(4):158, Nov. 2018. 2
- [14] Daniel Kienzle, Robin Schön, Rainer Lienhart, and Shin’Ichi Satoh. Towards ball spin and trajectory analysis in table tennis broadcast videos via physically grounded synthetic-to-real transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025. 4
- [15] J. Kober, K. Muelling, O. Kroemer, C.H. Lampert, B. Schoelkopf, and J. Peters. Movement templates for learning of hitting and batting. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2010. 1
- [16] J. Komorowski, G. Kurzejamski, and G. Sarwas. Deepball: Deep neural-network ball detector. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 5: VISAPP*, pages 297–304. INSTICC, SciTePress, 2019. 2, 7, 11
- [17] Jan Kotera, Denys Rozumnyi, Filip Šroubek, and Jiří Matas. Intra-frame object tracking by deblurring. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2300–2309, 2019. 3, 4
- [18] K. Kulkarni, R. Jamadagni, J. Paul, and S. Shenoy. Table Tennis Stroke Detection and Recognition Using Ball Trajectory Data. *SSRN Electronic Journal*, Jan. 2022. 2
- [19] Asai Kyohei, Nakayama Masamune, and Yase Satoshi. The ping pong robot to return a ball precisely trajectory prediction and racket control for spinning balls. 2019. 1
- [20] W. Li, X. Liu, K. An, C. Qin, and Y. Cheng. Table Tennis Track Detection Based on Temporal Feature Multiplexing Network. *Sensors*, 23(3):1726, Jan. 2023. 2
- [21] X. Li, W. Wang, X. Hu, J. Li, J. Tang, and J. Yang. Generalized Focal Loss V2: Learning Reliable Localization Quality Estimation for Dense Object Detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11627–11636, Nashville, TN, USA, June 2021. IEEE. 5
- [22] P. Liu and J. Wang. MonoTrack: Shuttle trajectory reconstruction from monocular badminton video. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3512–3521, New Orleans, LA, USA, June 2022. IEEE. 2, 7, 11
- [23] Katharina Muelling, Abdeslam Boularias, Betty Mohler, Bernhard Schölkopf, and Jan Peters. Learning strategies in table tennis using inverse reinforcement learning. *Biological Cybernetics*, 108(5):603–619, Oct. 2014. 1
- [24] Banoth Thulasya Naik, Mohammad Farukh Hashmi, and Neeraj Dhanraj Bokde. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Applied Sciences*, 12(9):4429, Jan. 2022. 1
- [25] Arjun Raj, Lei Wang, and Tom Gedeon. TrackNetV4: Enhancing Fast Sports Object Tracking with Motion Attention Maps, Sept. 2024. 2
- [26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation.

- In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 2
- [27] Denys Rozumnyi, Jan Kotera, Filip Šroubek, and Jiří Matas. Non-causal tracking by deblatting. In Gernot A. Fink, Simone Frintrop, and Xiaoyi Jiang, editors, *Pattern Recognition*, pages 122–135, Cham, 2019. Springer International Publishing. 3
- [28] Denys Rozumnyi, Jan Kotera, Filip Šroubek, and Jiří Matas. Tracking by Deblatting. *International Journal of Computer Vision*, 129(9):2583–2604, Sept. 2021. 3
- [29] Denys Rozumnyi, Jan Kotera, Filip Šroubek, Lukas Novotny, and Jiri Matas. The World of Fast Moving Objects . In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4838–4846, Los Alamitos, CA, USA, July 2017. IEEE Computer Society. 3, 4, 6
- [30] Denys Rozumnyi, Jan Kotera, Filip Šroubek, and Jiří Matas. Sub-frame appearance and 6d pose estimation of fast moving objects. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6777–6785, 2020. 3
- [31] Denys Rozumnyi, Jiří Matas, Filip Šroubek, Marc Pollefeys, and Martin R. Oswald. Fmodetect: Robust detection of fast moving objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3541–3549, October 2021. 3
- [32] N. Sun, Y. Lin, S. Chuang, T. Hsu, D. Yu, H. Chung, and T. Ik. TrackNetV2: Efficient Shuttlecock Tracking Network. In *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pages 86–91, Dec. 2020. 2, 3, 4, 5, 6, 7, 11
- [33] S. Tamaki and H. Saito. Reconstruction of 3D Trajectories for Performance Analysis in Table Tennis. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1019–1026, June 2013. 2
- [34] S. Tarashima, M. Haq, Y. Wang, and N. Tagawa. Widely applicable strong baseline for sports ball detection and tracking. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 2, 3, 4, 5, 6, 7, 11
- [35] J. Tebbe, Y. Gao, M. Sastre-Rienietz, and A. Zell. A table tennis robot system using an industrial kuka robot arm. In *Pattern Recognition: 40th German Conference, GCPR 2018, Stuttgart, Germany, October 9-12, 2018, Proceedings 40*, pages 33–45. Springer, 2019. 1, 2
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 2
- [37] R. Voeikov, N. Falaleev, and R. Baikulov. TTNet: Real-time temporal and spatial video analysis of table tennis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3866–3874, Seattle, WA, USA, June 2020. IEEE. 3
- [38] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3349–3364, 2021. 2
- [39] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020. 3
- [40] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018. 3
- [41] P. Wong, H. Myint, L. Dooley, and A. Hopgood. A multi-view automatic table tennis umpiring framework. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, page 175433712311714, Apr. 2023. 1, 2
- [42] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 3–19, Cham, 2018. Springer International Publishing. 3
- [43] Erwin Wu, Florian Perteneder, and Hideki Koike. Real-time Table Tennis Forecasting System based on Long Short-term Pose Prediction Network. In *SIGGRAPH Asia 2019 Posters*, SA ’19, pages 1–2, New York, NY, USA, Nov. 2019. Association for Computing Machinery. 1
- [44] G. Van Zandycke and C. De Vleeschouwer. Real-time CNN-based Segmentation Architecture for Ball Detection in a Single View Setup. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports, MMSports ’19*, pages 51–58, New York, NY, USA, Oct. 2019. Association for Computing Machinery. 2, 7, 11
- [45] W. Zhang, Y. Zhang, Y. Zhao, and B. Zhang. Recognizing and Recovering Ball Motion Based on Low-Frame-Rate Monocular Camera. *Applied Sciences*, 13(3):1513, Jan. 2023. 3
- [46] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In *ECCV*, 2018. 2
- [47] Y. Zhao, J. Wu, Y. Zhu, H. Yu, and R. Xiong. A learning framework towards real-time detection and localization of a ball for robotic table tennis system. In *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 97–102, Okinawa, July 2017. IEEE. 2

# BlurBall: a ball detector with blur estimation

## Supplementary Material

### Dataset

The dataset was generated from publicly available online videos under a Creative Commons license to ensure compliance with copyright regulations. We deliberately covered a wide range of scenarios, including different playing conditions, camera angles, and lighting variations, to improve model robustness. Orange table tennis balls, although officially approved for competition, are not included. This omission is unlikely to affect performance, as white balls are overwhelmingly preferred and almost exclusively used in professional matches.

The dataset is publicly available at: <https://cogsys-tuebingen.github.io/blurball/>. Each video is accompanied by a CSV file with the following structure:

Frame	Visibility	X	Y	$\theta$	$l$
000049	1	581.62	295.26	-152.5	2.8
000050	1	572.98	292.86	171.8	2.1

Table 4. CSV description. Each row contains the ball position ( $X, Y$ ), blur orientation ( $\theta$ ), and half-length ( $l$ ).

Angles are given in degrees and follow the convention illustrated in Figure 3.

### Camera calibration

We provide camera calibration for each table tennis match in the dataset. Specifically, we include the focal length  $f$  and the camera extrinsics: rotation vector  $r$  and translation vector  $T$ . The world frame is defined by the table geometry, with the table’s length aligned to the  $Y$ -axis and the surface normal aligned to the  $Z$ -axis, as shown in Figure 10.

This choice is motivated by the precise, known geometry of the table, which makes it a reliable calibration target. We manually annotate keypoints such as the four table corners, the midline–backline intersection, and the intersection between the net and the table’s side edge (Figure 10). The camera pose is then estimated by minimizing the reprojection error via a standard PnP optimization.

Due to the limited and near-coplanar set of keypoints, distortion coefficients  $d_d$  and the optical center cannot be reliably estimated. We therefore assume an ideal pinhole model, with intrinsic parameters reduced to the focal length  $f$ .

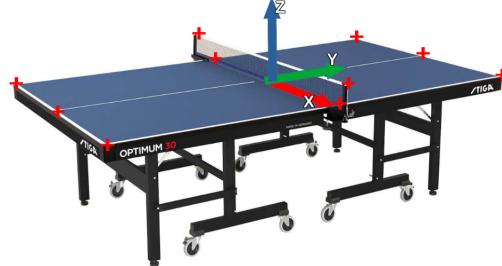


Figure 10. Annotated table keypoints used for camera calibration and definition of the world frame.

### Model training

We followed the training setup of WASB [34] for consistency across models. Each model was trained for 30 epochs with an appropriate loss function and optimizer:

- **DeepBall** [16] and **DeepBall-large**: BCE loss, Adam optimizer, learning rate  $3 \times 10^{-4}$ .
- **BallSeg** [44]: focal loss ( $\gamma = 2$ ), Adam optimizer, learning rate  $3 \times 10^{-4}$ .
- **TrackNetV2** [32] and **ResTrackNetV2**: focal loss, AdaDelta optimizer, learning rate 1.0.
- **MonoTrack** [22]: Combo loss, AdaDelta optimizer, learning rate 1.0.
- **WASB** [34]: quality focal loss, Adam optimizer, learning rate  $3 \times 10^{-4}$ .

This ensures fair comparison, with each model optimized using strategies suited to its architecture.

### BlurBall: further insights

#### Impact of the threshold value

To evaluate the influence of the detection threshold in our BlurBall model, we plot the Precision-Recall (PR) curve in Figure 11. Overall, the 1-step variant achieves a better balance than the 3-step variant. However, for identical threshold values, the 1-step detector consistently yields higher recall but lower precision, indicating that its middle-frame predictions tend to be overconfident. This trade-off motivates the choice of a threshold value of  $\delta = 0.7$ , which we adopt for subsequent tracking experiments in Section 4.3.

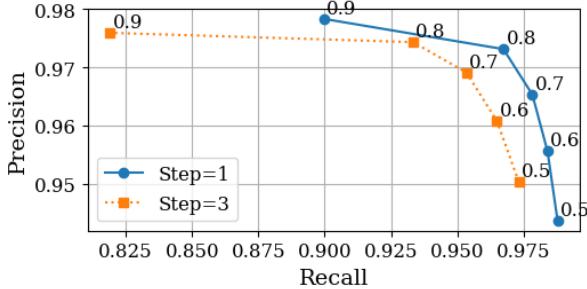


Figure 11. PR curves for BlurBall at different confidence thresholds  $\delta$ .

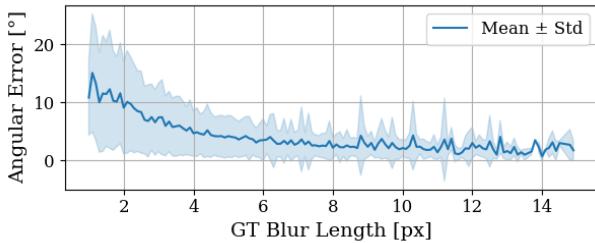


Figure 12. Angular estimation error as a function of ground-truth blur length.



Figure 13. Examples of BlurBall failures. FN1 and FN2: missed detections. FP1: false positive on a hand.

## Blur prediction

Figure 12 shows the relationship between blur length and angle estimation error. As expected, longer blur streaks yield more accurate angle estimation. For  $l > 3$ , the angular error remains below  $10^\circ$ , validating its use in downstream tasks such as trajectory prediction (Section 4.5).

## Failure cases

Figure 13 shows typical failure cases. In FN1, the ball briefly appears between two occluded frames, leading to a missed detection. In FN2, the ball overlaps with a moving player’s body, a scenario difficult even for human observers. In FP1, a hand is incorrectly detected as the ball.

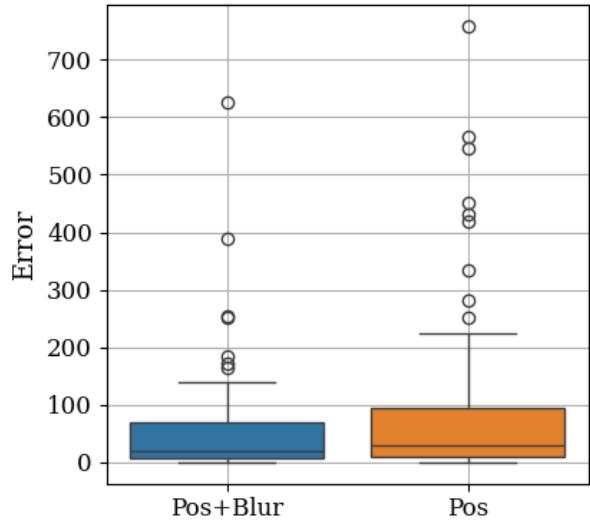


Figure 14. Full box plot of trajectory prediction errors for Pos and Pos+Blur, including outliers.

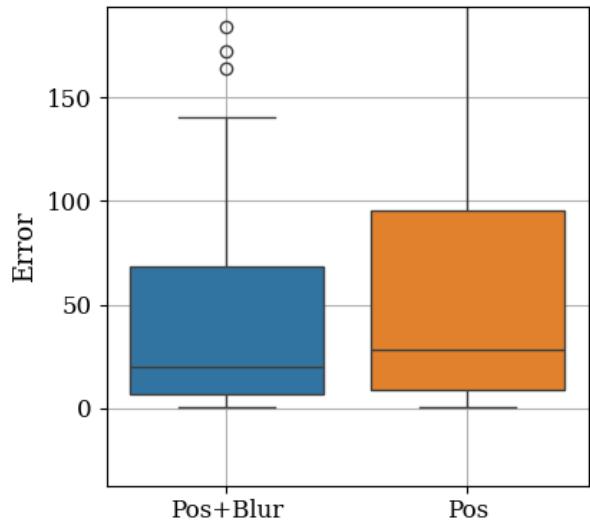


Figure 15. Zoomed-in box plot of trajectory prediction errors.

## Trajectory prediction benchmark

We benchmark trajectory prediction using position-only (Pos) versus position+blur (Pos+Blur) fitting. Figure 14 and Figure 15 report full and zoomed-in box plots.

As shown in Figure 15, Pos+Blur achieves a lower median error (19.9 px vs. 28.4 px) and a narrower interquartile range, indicating greater consistency. The full distribution (Figure 14) also reveals fewer extreme outliers and a shorter upper whisker (140.2 px vs. 224.6 px), confirming improved robustness in difficult cases. Overall, incorporat-

ing blur leads to more accurate and stable trajectory predictions, especially under sparse or noisy observations.