Widely Applicable Strong Baseline for Sports Ball Detection & Tracking

Shuhei Tarashima Muhammad Abdul Haq Yushan Wang Norio Tagawa

NTT Communications
Tokyo Metropolitan Univ.

BMVC 2023

TL; DR

- We propose **a new SBDT baseline, WASB.**
- We introduce a new evaluation protocol using **5 SBDT datasets** from different sports. **6 SOTA methods** are (re-)implemented for fair comparison.
- Experiments show that WASB substantially outperforms SBDT SOTAs on all the datasets.

Sports Ball Detection & Tracking (SBDT)

Input: a (sports) video clip, **Output**: a (x, y)-coordinate of a sports ball (if visible) for each frame

Dataset & Codebase

github

- SBDT datasets from 5 different sport categories:
  - Volleyball and Basketball are newly introduced by us
  - for Soccer and Basketball, new annotations are provided

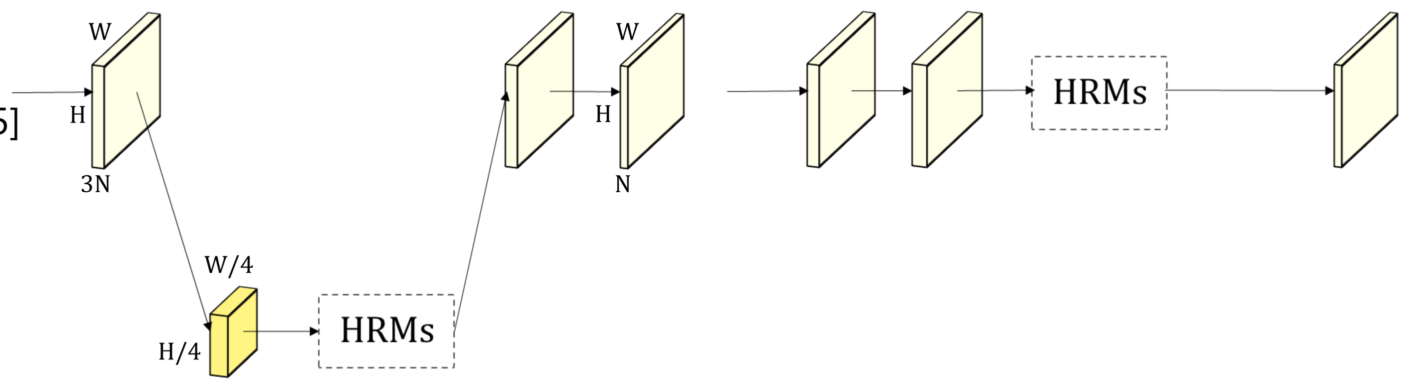
	resolution	FPS	Train games	Train clips	Train frames	Train disp.[pixel]	Test games	Test clips	Test frames	Test disp.
Soccer [19]	1920 × 1080	25	1	4	11994	10.4 ± 10.0	1	2	5999	15.7 ± 13.0
Tennis [32]	1280 × 720	30	7	65	14160	15.3 ± 13.0	3	30	5675	13.6 ± 10.2
Badminton [75]	1280 × 720	30	26	172	78558	11.8 ± 12.2	3	29	12656	12.5 ± 12.9
Volleyball	1280 × 720	N/A	39	3493	143213	14.4 ± 11.4	16	1337	54817	15.1 ± 11.5
Basketball	1920 × 1080	N/A	70	3392	244224	33.7 ± 21.8	11	432	31104	33.9 ± 21.4

- 6 SOTA SBDT methods, 2 of which (★) are minorly updated by us
  - DeepBall [1], DeepBall-Large★, BallSeg [2], TrackNetV2 [3], ResTrackNetV2★, MonoTrack [4]

Widely Applicable Strong Baseline (WASB)

1. High-Resolution Feature Extraction Model

- High-Resolution Modules (HRMs) of small HRNet [5]
- **Stem without strides** to feed higher-resolution features to HRMs
- Multi-In Multi-Out (MIMO) design ($N = 3$)



Original stem design in HRNet **Proposed stem design**

2. Position-Aware Model Training

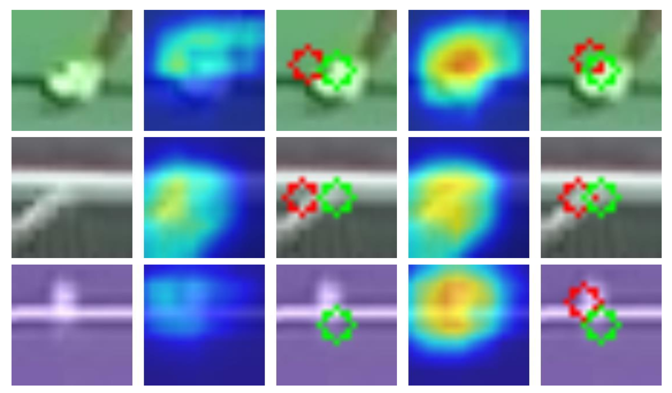
- Train a model that predicts heatmaps representing ball positions
- **Focal-loss** [6] with **binary** ground truth (GT) during the first T epochs
- **Quality focal loss** [7] with **real-valued** GT during remaining T' epochs

0	0	0	0	0	0	0
0	0	1	1	1	0	0
0	1	1	1	1	1	0
0	1	1	1	1	1	0
0	1	1	1	1	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Binary GT

0	0	0	0	0	0	0
0	0	0.73	0.86	0.73	0	0
0	0.73	1	1	1	0.73	0
0	0.86	1	1	1	0.86	0
0	0.73	1	1	1	0.73	0
0	0	0.73	0.86	0.73	0	0
0	0	0	0	0	0	0

Real-valued GT



after T epochs after $T + T'$ epochs

3. A Bunch of Tricks during Inference

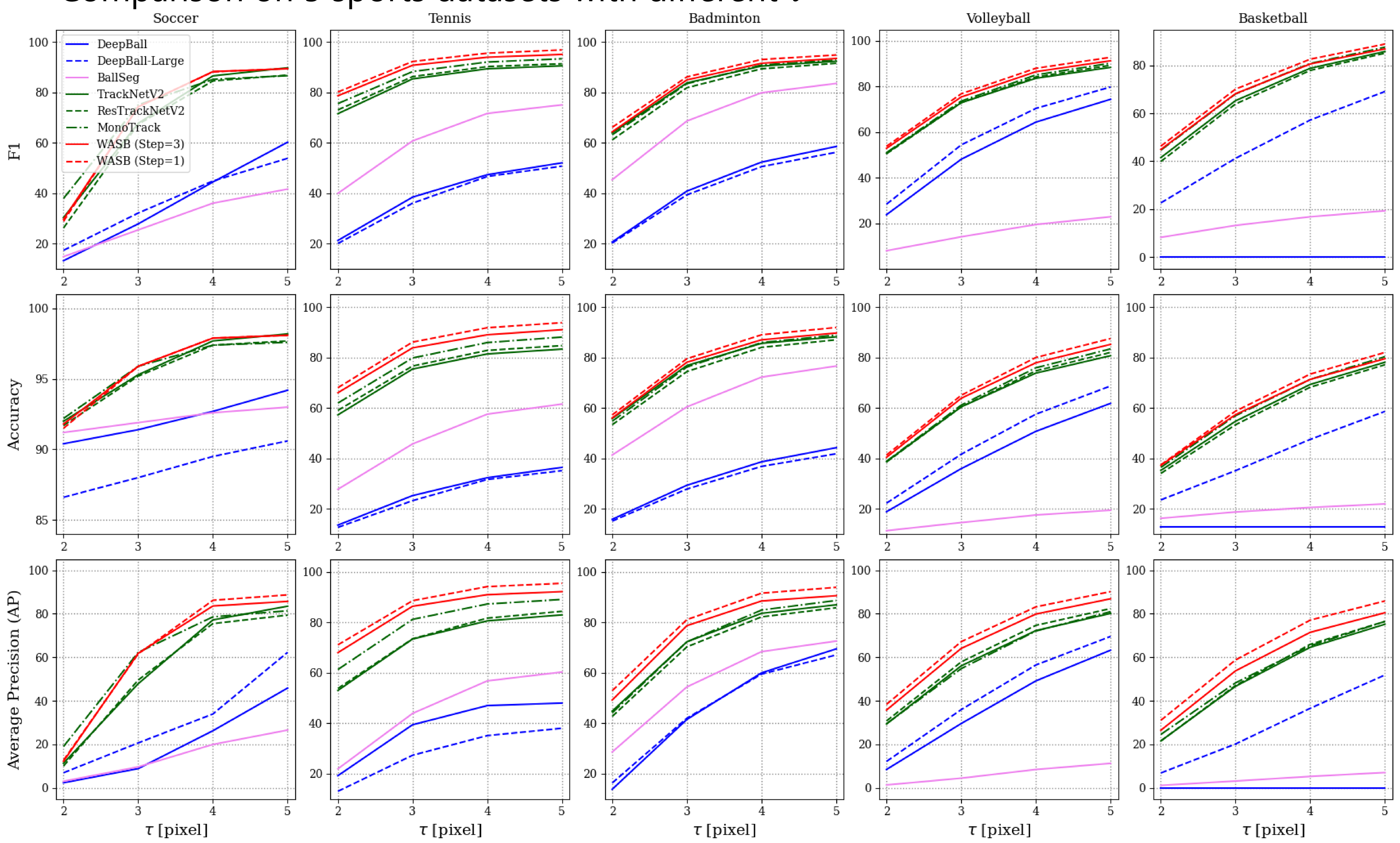
- prediction of each ball position (i.e., (x, y)-coordinate) as **a center of heatmap values in a detected blob**
- **online tracking with local motion model** to take long-term temporal consistency into account
- **oversampling the same image in different MIMO combinations** to produce diverse detection candidates

Results

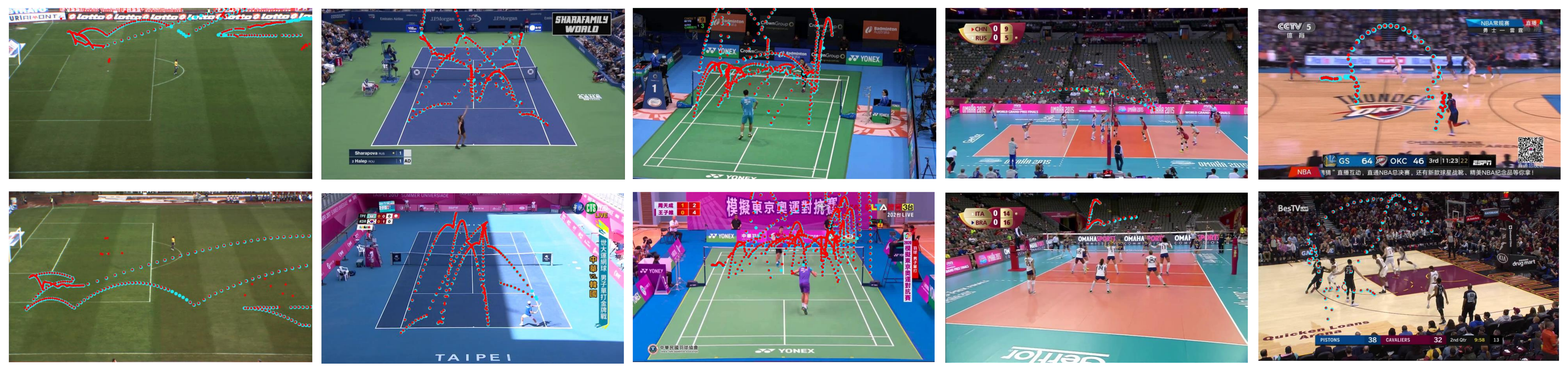
- Comparison on 5 datasets from different sports (distance threshold $\tau = 4$[pixel])

	# param.	Soccer F1 ↑	Soccer Acc. ↑	Soccer AP ↑	Soccer FPS ↑	Tennis F1	Tennis Acc.	Tennis AP	Tennis FPS	Badminton F1	Badminton Acc.	Badminton AP	Badminton FPS	Volleyball F1	Volleyball Acc.	Volleyball AP	Volleyball FPS	Basketball F1	Basketball Acc.	Basketball AP	Basketball FPS
DeepBall [40, 41]	0.1M	44.5	92.7	26.3	44.6	47.4	32.3	47.0	52.1	52.4	38.6	60.0	57.1	64.4	50.7	49.2	21.1	0.0	12.9	0.0	30.3
DeepBall-Large	1.0M	44.9	89.5	34.0	42.0	46.7	31.6	35.1	47.7	50.6	36.8	59.5	53.0	70.4	57.5	56.5	21.1	57.2	47.5	36.6	30.9
BallSeg [80]	12.7M	36.1	92.6	20.0	64.8	71.7	57.5	56.8	62.7	79.9	72.2	68.4	75.0	19.5	17.5	8.5	18.2	16.8	20.5	5.3	29.5
TrackNetV2 [75]	11.3M	86.6	97.7	77.2	66.0	89.4	81.4	80.6	55.3	90.5	85.6	83.6	77.0	83.6	73.8	72.3	17.6	78.8	69.3	64.6	28.0
ResTrackNetV2	1.2M	84.6	97.4	75.5	56.2	90.3	82.8	81.7	59.0	89.4	84.0	82.2	71.3	84.2	74.7	74.7	28.6	77.9	68.2	66.0	38.2
MonoTrack [50]	2.9M	85.2	97.4	78.6	58.0	92.1	85.9	87.3	64.1	90.9	85.9	84.9	75.5	85.1	75.9	72.1	19.7	80.8	71.3	65.3	32.1
WASB (Ours, Step=3)	1.5M	88.3	97.9	83.6	55.7	94.0	89.0	91.0	58.2	91.6	87.0	88.5	70.4	86.5	77.9	79.9	18.0	80.6	71.3	71.5	30.2
WASB (Ours, Step=1)	1.5M	88.2	97.9	86.2	23.6	95.6	91.8	94.2	35.2	93.1	89.0	91.6	34.3	88.0	80.0	83.2	15.8	82.6	73.4	77.1	22.3

- Comparison on 5 sports datasets with different τ



- Qualitative results (blue: ground truth, red: precition)



[1] DeepBall: Deep Neural-Network Ball Detector, in VISAPP, 2019.
[2] Real-time CNN-based Segmentation Architecture for Ball Detection in a Single View Setup, in ACM MM Workshops, 2019.
[3] TrackNetV2: Efficient Shuttlecock Tracking Network, in ICPAI, 2020.
[4] MonoTrack: Shuttle Trajectory Reconstruction from Monocular Badminton Video, in CVPRW, 2022.
[5] Deep High-Resolution Representation Learning for Visual Recognition, in TPAMI, 2020.
[6] Focal Loss for Dense Object Detection, in ICCV, 2017.
[7] Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection, in NeurIPS, 2020.