

Broadcast2Pitch: Game State Reconstruction from Unconstrained Soccer Videos

Supplementary Material

In this supplementary document, we provide further analysis of our framework. Sec. 7 details the ablation study for the IDATR module’s key hyperparameter, τ_{gap} . Sec. 8 provides a detailed breakdown of the computational cost of each major module in our proposed GSR framework.

7. IDATR Hyperparameters

To assess the sensitivity of the framework to the τ_{gap} , we perform an ablation study by varying its value and measuring the resulting GS-HOTA, GS-DetA, GS-AssA, and IDF1 scores in Table 8. The highest performance is observed at $\tau_{gap} = 25$, while the lowest score occurs at $\tau_{gap} = 5$. Overall, the variation in GS-HOTA across different τ_{gap} values remain marginal, indicating that the framework is relatively robust to the choice of τ_{gap} . Nonetheless, the results suggest that preserving tracklet continuity—even in the presence of short-term occlusions or missing detections—yields more stable and coherent tracking performance than aggressively fragmenting tracklets.

τ_{gap}	GS-HOTA \uparrow	GS-DetA \uparrow	GS-AssA \uparrow	IDF1 \uparrow
5	60.40	47.49	76.81	63.00
10	61.08	48.22	77.38	63.65
15	61.23	48.24	77.71	63.84
20	61.41	48.40	77.93	64.11
25	61.48	48.47	78.00	64.20
30	61.43	48.34	78.08	64.18
35	61.36	48.31	77.96	64.12
40	61.22	48.19	77.78	63.86

Table 8. Ablation study on the effect of τ_{gap} on IDATR performance in our GSR framework.

8. Computational Cost Analysis

Our GSR pipeline consists of five primary modules: (1) detection and tracking, (2) sports field registration via keypoint and line detection and homography estimation, (3) athlete identification using a vision-language model, (4) tracklet refinement (IDATR), and (5) post-processing. Table 9 reports the per-module computational cost in terms of parameter count (Params) and average inference speed (FPS), measured on a single NVIDIA RTX 4090 GPU with a batch size of 1 in inference time.

The primary computational bottleneck in our framework is the Athlete Identification module. To make the

Module	Model	Params	FPS
Detection Tracking + ReID	YOLOX-X	99.1M	12.1
	DeepEIoU + OSNet	2.2M	
Sports Field Registration	EfficientNet-Attention U-Net	36.9M	182
Athlete Identification	CLIP (ViT-L/14)	428.8M	8.3
	LLaMA-3.2-Vision	11B	1.4
IDATR	—	—	952.4
Post-processing	—	—	3225.8

Table 9. Computational cost analysis of each module in our Broadcast2Pitch framework (tested on an RTX 4090 GPU).

deployment of the LLaMA-3.2-Vision model feasible on a single NVIDIA RTX 4090 GPU, we leverage the Unsloth library [11] to load the model in a memory-efficient 4-bit quantized format. This optimization is critical for fitting the full pipeline within the 24 GB VRAM limit, enabling state-of-the-art accuracy while remaining compatible with consumer-grade hardware.

Despite this optimization, LLaMA-3.2-Vision runs at only 1.4 FPS, making it suitable for high-accuracy offline analysis. In contrast, CLIP (ViT-L/14) offers a significantly more efficient alternative, running at 8.3 FPS. These results present two practical deployment modes: (1) a high-accuracy offline mode using LLaMA-3.2-Vision for maximum precision in post-match analysis, and (2) a high-throughput analytics mode using CLIP, which offers a strong balance of accuracy and speed for faster processing.