Universität Augsburg University

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 National Institute of Informatics

Towards Ball Spin and Trajectory Analysis in Table Tennis Broadcast Videos via Physically Grounded Synthetic-to-Real Transfer

Daniel Kienzle, Robin Schön, Rainer Lienhart, Shin'ichi Satoh

11th International Workshop on Computer Vision in Sports (CVSports) at CVPR 2025

June 12th, 2025

Motivation

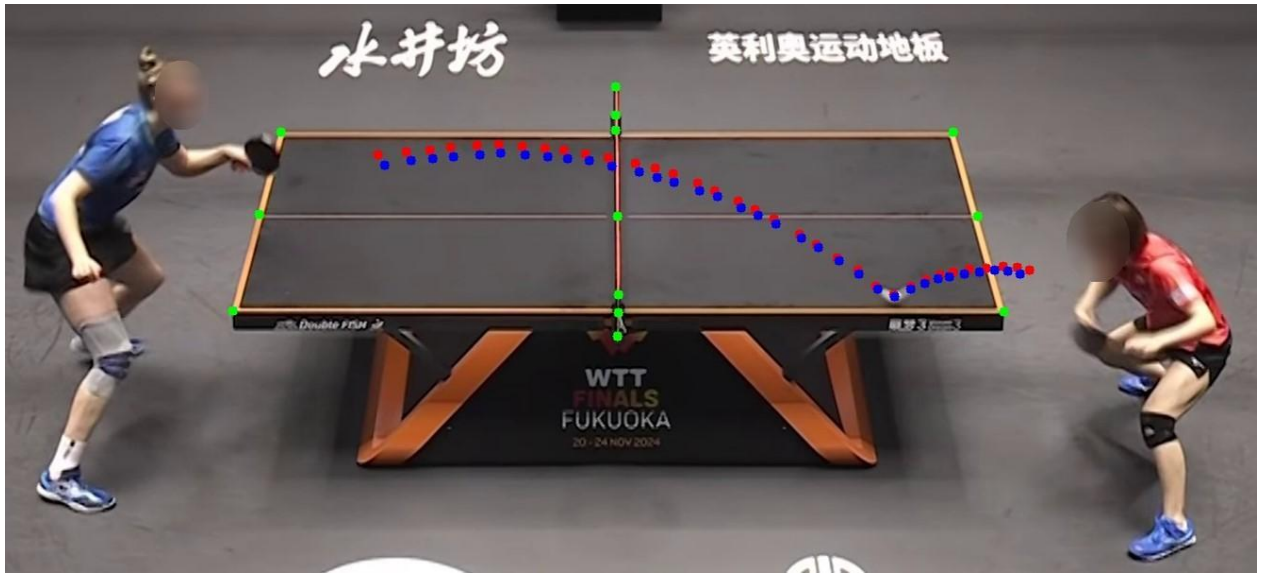
The Unseen Game Changer: Ball's 3D trajectory and spin is crucial for analysis

Goal: Train network to predict **3D trajectory** and **initial spin** in **real broadcast videos**

Problem: Ground truth is not available

➔ We train on **synthetic data only**

➔ **Zero-Shot Generalization**



Comparison of predicted and annotated trajectory

Methodology

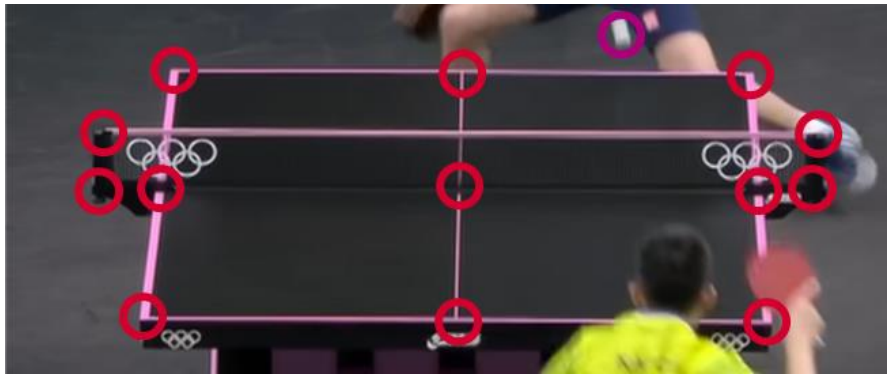
Overview

2D to 3D uplifting similar to 3D human pose estimation

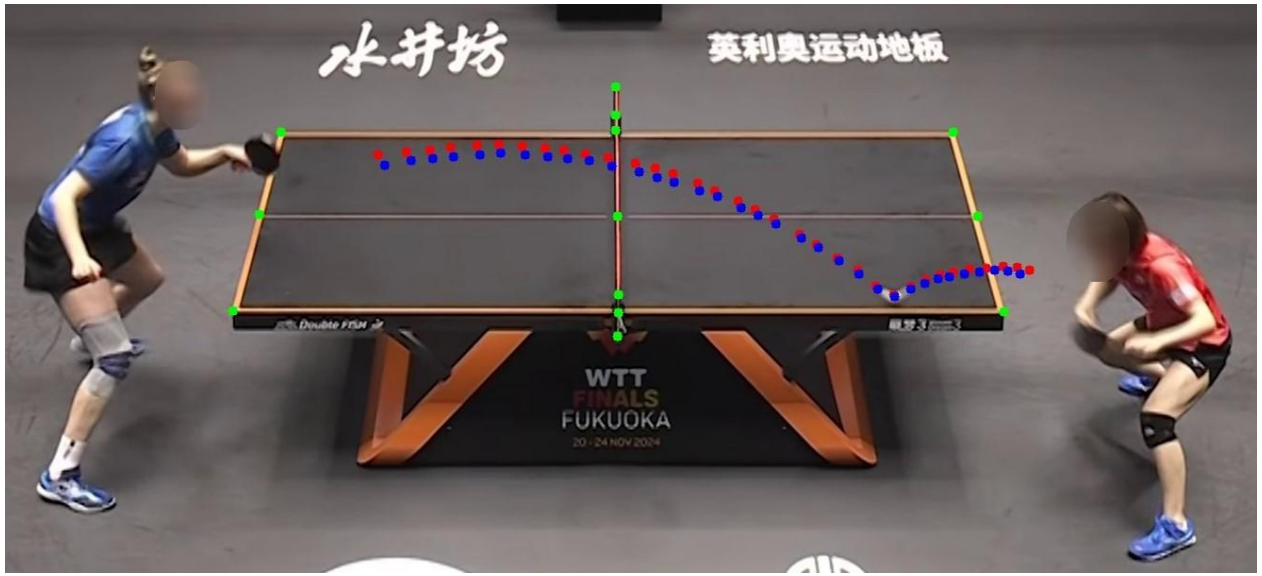
Input: 2D trajectory + 2D table keypoints

→ 2D keypoints in the image are “for free”

Output: 3D trajectory + initial spin



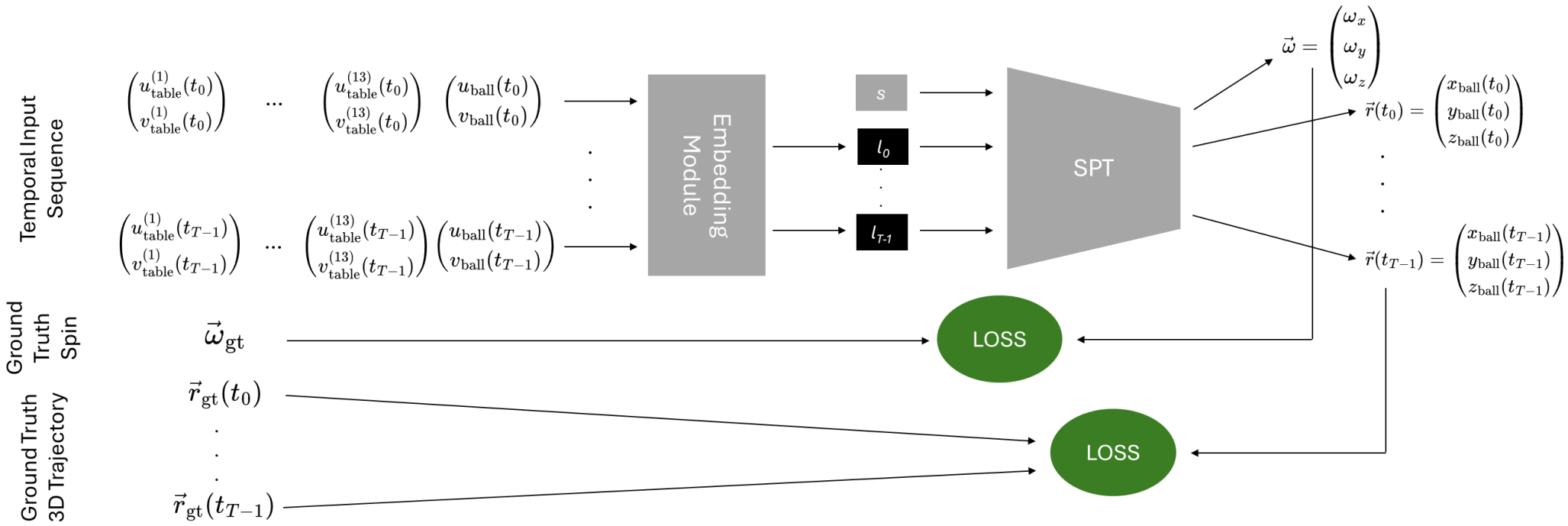
13 table keypoints + ball position



Comparison of predicted and annotated trajectory

Methodology

Pipeline



- For each frame at time t_i :
  → Extract 2D ball coordinates and table keypoints
  → Embed into a location token l_i
- Prepend learnable token s : $\{s, l_0, ..., l_{T-1}\}$

- Process sequence with **Spin Prediction Transformer (SPT)**
  → 3D position $\vec{r}(t_i)$ for each time t_i
  → Initial ball spin $\vec{\omega}$

➔ **Fully supervised** training on **synthetic ground truth**

Methodology

Generalization: Bridging the Reality Gap with Minimal Effort

- **Smart Data Representation**: 2D Annotations are enough

  → No visual gap in the representation

- **Leveraging Physics**: Realistic Trajectories via MuJoCo [1]

  → Physics of table tennis is well-understood

- **Simple Yet Effective Augmentations**

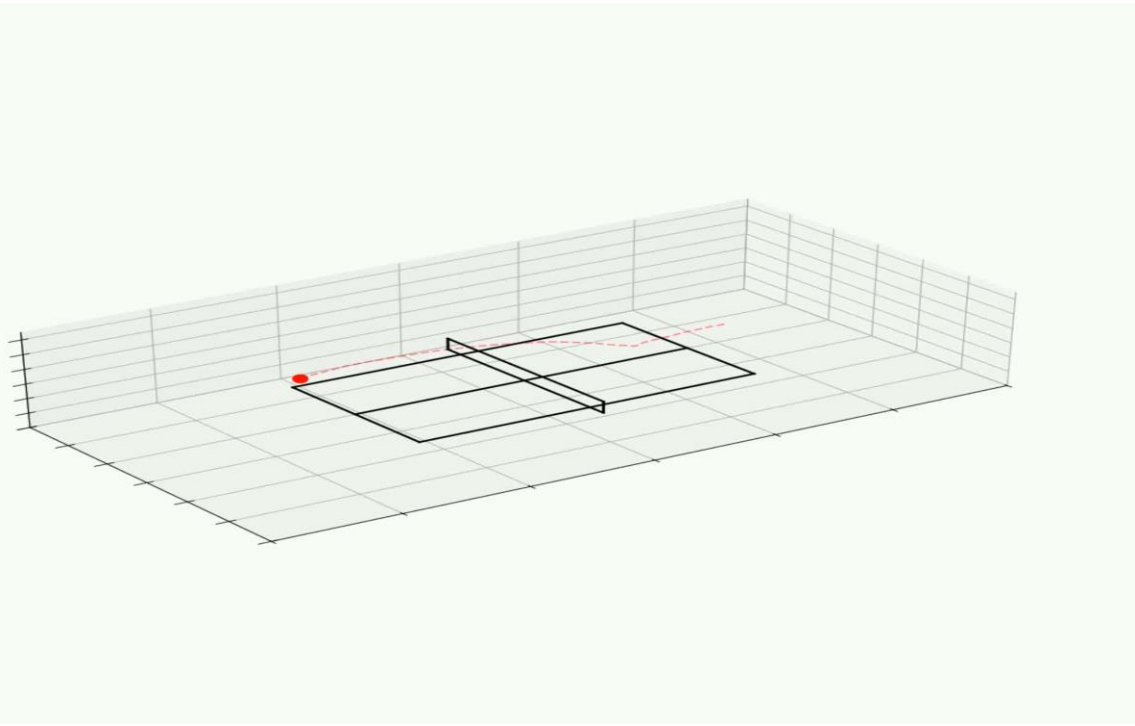
  → Motion Blur, Detection Noise, Early Trajectory Cutoffs

➔ **Simulated and real trajectories become indistinguishable**

[1]: D. B. D'Ambrosio, S. W. Abeyruwan, L. Graesser, et.al., "Achieving Human Level Competitive Robot Table Tennis", 7th Robot Learning Workshop: Towards Robots with Human-Level Abilities, 2025

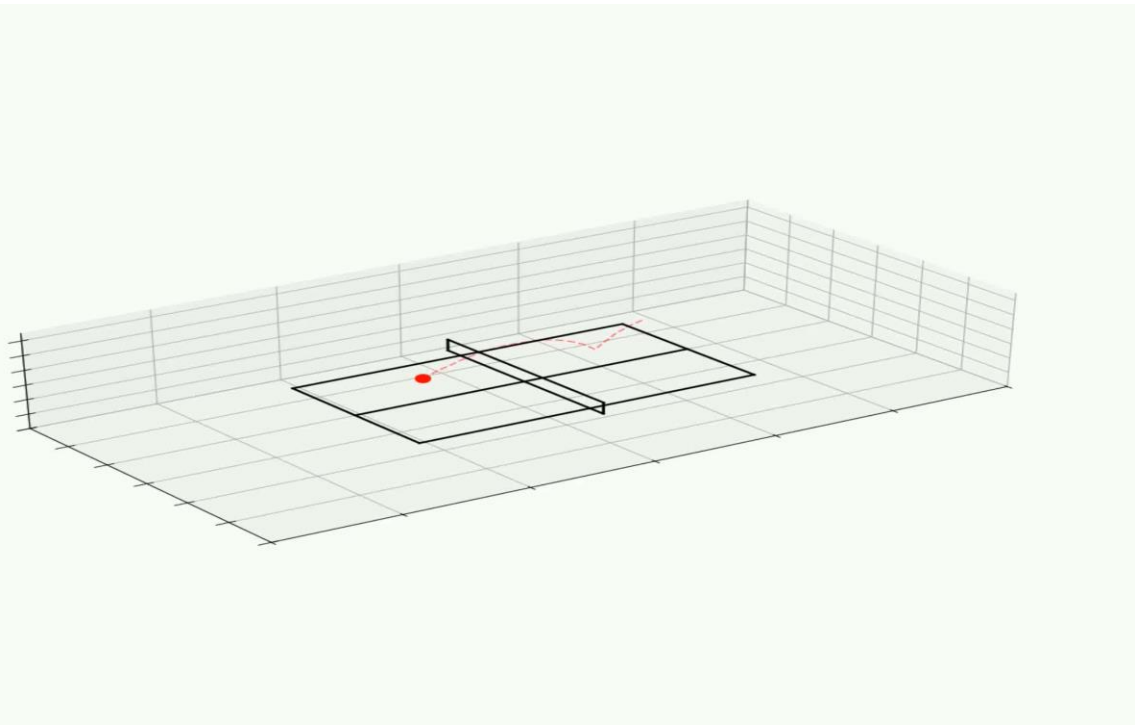
Results

Evaluation on real broadcast videos: Topspin



Results

Evaluation on real broadcast videos: Backspin



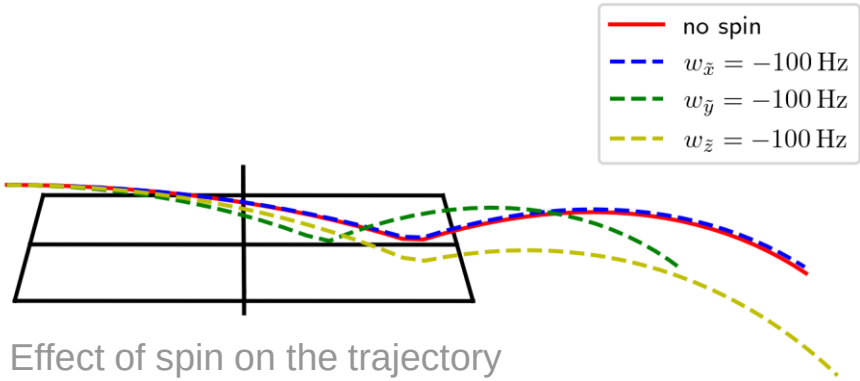
What‘s more

Enjoy the paper

- Physics analysis & coordinate system discussion
- Architecture discussion

  → Spin Prediction Transformer (SPT)

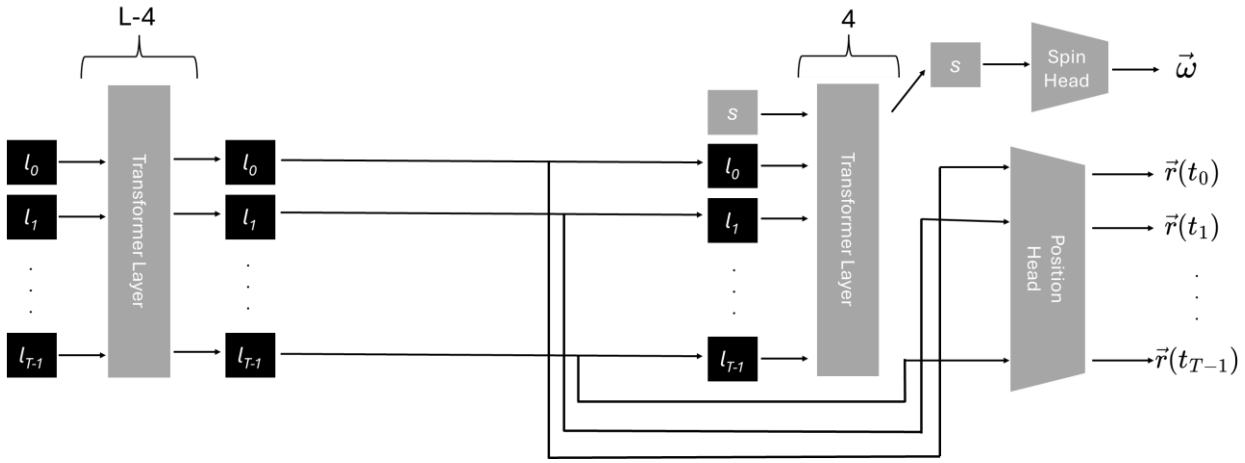
  → Embedding module
- Extensive evaluation & ablation study



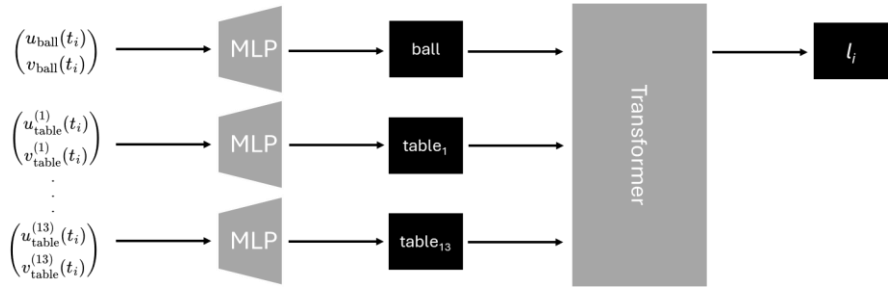
Effect of spin on the trajectory

spin		trajectory	
$acc \uparrow$	$F_1 \uparrow$	$\Delta\vec{r}_{\mathrm{img}} \downarrow$	rel. $\Delta\vec{r}_{\mathrm{img}} \downarrow$
92.0 %	0.917	5.6 px	0.19 %

Results of best model



SPT architecture



Dynamic embedding

Conclusion

Recap of main contributions

- **First learning-based approach** for spin & trajectory estimation in table tennis broadcast videos
- Nearly perfect **zero-shot generalization**
- Novel uplifting **architecture**
- **Synthetic training trajectories** & **real evaluation videos**

Thank you very much!

See you at the poster session

Daniel Kienzle
Chair of Machine Learning and Computer Vision
University of Augsburg, Germany
daniel.kienzle@uni-a.de
https://www.uni-augsburg.de/en/fakultaet/fai/informatik/prof/mmc/team/kienzle/