DANIEL KIENZLE, KATJA LUDWIG, JULIAN LORENZ, SHIN'ICHI SATOH, RAINER LIENHART

University of Augsburg
Centre for Advanced Analytics and Predictive Sciences

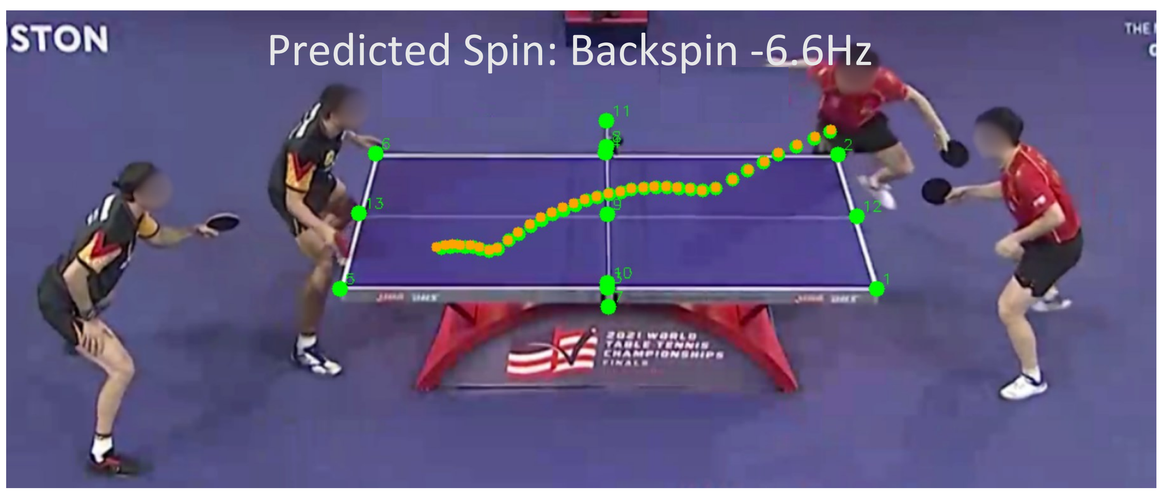
大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

Uplifting Table Tennis: A Robust, Real-World Application for 3D Trajectory and Spin Estimation

MOTIVATION

3D ball trajectory & initial spin are key to gameplay analytics
→ Improve training, extract statistics, enable virtual replay

- **Goal**: Predict **3D Ball Trajectory** & **Initial Spin**
- **Challenge**: No 3D ground truth in real videos
- **Solution**: Implement **Two-Stage Pipeline**

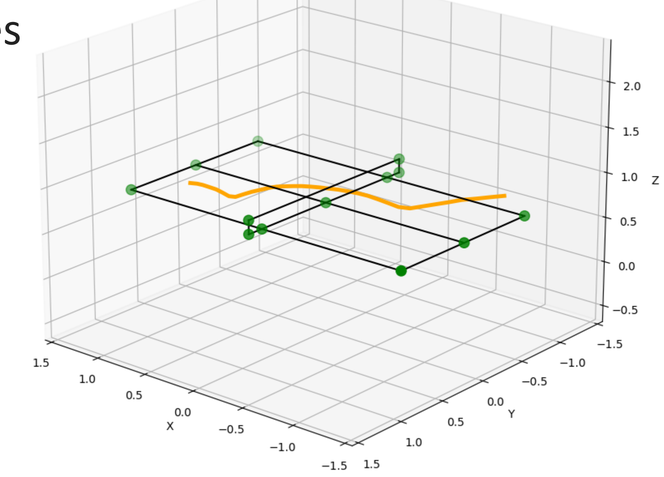


2D detections (green) and reprojected 3D predictions (orange)

FROM CONCEPT TO PRACTICAL APPLICATION

Two-Stage Pipeline:

- **Front-End:** Detections in the video frames
  Video frames → 2D ball trajectory & 2D table keypoints
- **Back-End:** Uplifting approach [1]
  2D trajectory → 3D ball trajectory & initial spin
- **Core Challenge**: Combining Front- and Back-end
  → Imperfect 2D detections
  → Varying framerates



Predicted 3D trajectory

A ROBUST TWO-STAGE PIPELINE

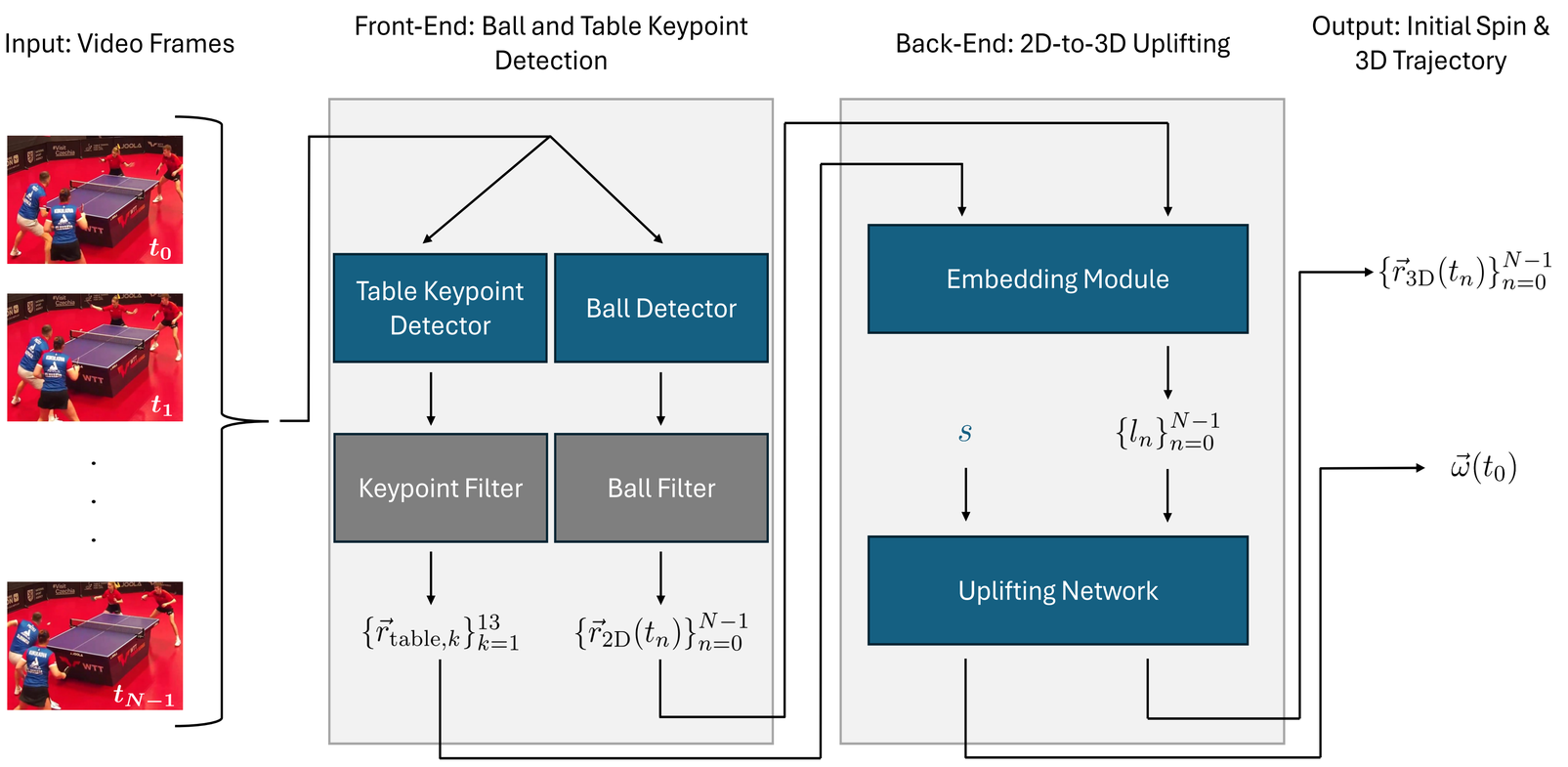
Challenge: No **3D Ground Truth**
Video → 3D trajectory & spin is impossible

Solution: Introduce **Two-Stage pipeline**

- Train **Front-End** with **Real 2D Annotations**
- Train **Back-End** with **Synthetic Data**

Contributions:

- High-performance **Detectors** utilizing the Segformer++ architecture [2]
- Tailored **Filters** removing false positives
- Robust **Uplifting Network**
  → Zero-shot generalization
  → Deal with noisy & missing detections

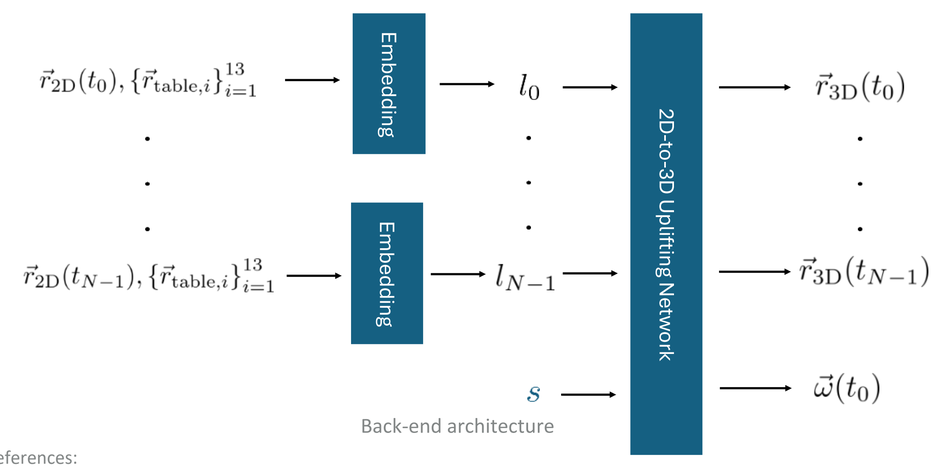


Overview of the full pipeline

2D-TO-3D UPLIFTING NETWORK

Trained solely on **Synthetic Data**
→Smart data representation
→No synthetic-to-real gap
→Zero-shot generalization

We adjust the architecture from [1] to varying framerates and real-world imperfections



Back-end architecture

RESULTS

Model	#Params	Input Res.	FPS ↑	ACC@2px ↑	ACC@5px ↑	ACC@10px ↑
Segformer++ (B0)	$3.7 \cdot 10^6$	1920×1088	26	43.2 %	**86.8 %**	**94.4 %**
Segformer++ (B2)	$24.7 \cdot 10^6$	1600×896	19	**54.3 %**	85.3 %	93.0 %
WASB (HRNet Small)	$1.5 \cdot 10^6$	1280×704	17	41.1 %	83.8 %	89.3 %
VitPose (ViT Small)	$25.3 \cdot 10^6$	1152×640	19	30.0 %	68.5 %	79.7 %

Ball Detection with different architectures

Model	#Params	Input Res.	FPS ↑	ACC@2px ↑	ACC@5px ↑	ACC@10px ↑
Segformer++ (B0)	$3.7 \cdot 10^6$	1920×1088	26	**75.0 %**	85.9 %	90.3 %
Segformer++ (B2)	$24.7 \cdot 10^6$	1600×896	18	**75.0 %**	87.1 %	**91.8 %**
WASB (HRNet Small)	$1.5 \cdot 10^6$	1280×704	16	72.4 %	**87.4 %**	91.3 %
VitPose (ViT Small)	$25.9 \cdot 10^6$	1152×640	19	38.0 %	50.3 %	52.1 %

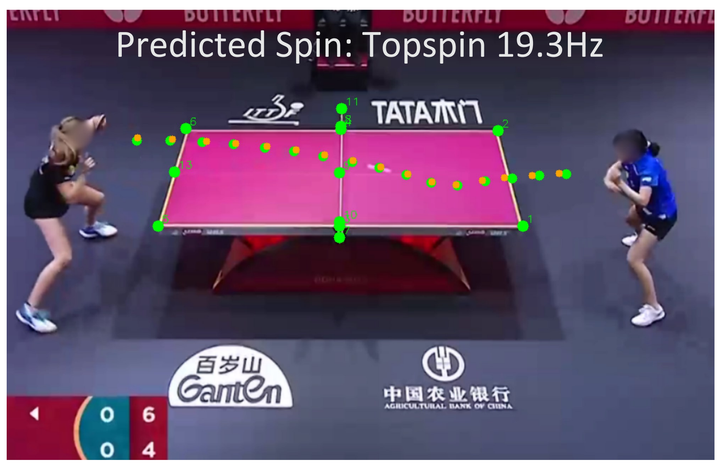
Table Keypoint Detection with different architectures

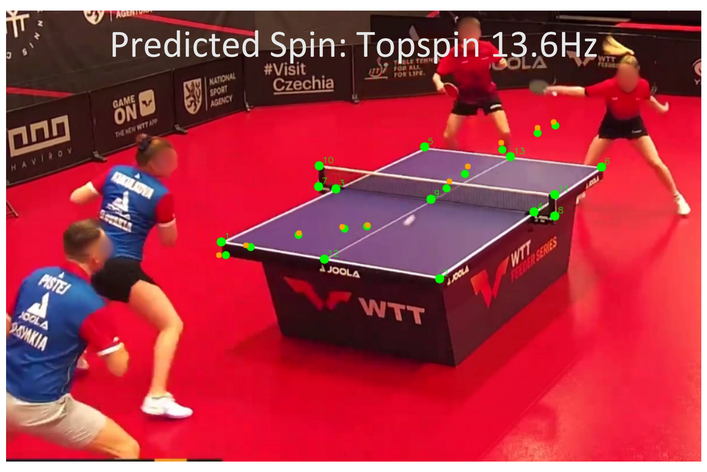
Dataset	Table: m2DRE ↓	Ball: m2DRE ↓	Spin: ACC ↑	Spin: F_1 ↑
TTHQ	2.72 ± 5.71 px	12.28 ± 10.84 px	89.5 %	0.900
TTST	5.75 ± 10.26 px	9.41 ± 16.90 px	97.1 %	0.974

Evaluation of the full pipeline

Method	Transforms: Half FPS	Transforms: Miss. Det.	Metrics: *ACC* ↑	Metrics: F_1 ↑	Metrics: m2DRE ↓
Kienzle et al. [22]	×	×	97.1 %	0.970	2.98 px
Mixed	×	×	**100.0 %**	**1.000**	**2.49 px**
Ours	×	×	97.1 %	0.970	3.43 px
Kienzle et al. [22]	✓	×	76.5 %	0.731	**2.71 px**
Mixed	✓	×	79.4 %	0.770	3.13 px
Ours	✓	×	**100.0 %**	**1.000**	3.54 px
Kienzle et al. [22]	×	✓	88.2 %	0.876	24.15 px
Mixed	×	✓	**97.1 %**	**0.970**	**5.45 px**
Ours	×	✓	**97.1 %**	**0.970**	5.56 px
Kienzle et al. [22]	✓	✓	67.7 %	0.598	23.54 px
Mixed	✓	✓	70.6 %	0.646	5.99 px
Ours	✓	✓	**97.1 %**	**0.970**	**5.75 px**

Back-end architectures under the influence of real-world imperfections





2D detections (green) and reprojected 3D predictions (orange)

References:

- [1]: D. Kienzle, R. Schön, R. Lienhart and S. Satoh, "Towards Ball Spin and Trajectory Analysis in Table Tennis Broadcast Videos via Physically Grounded Synthetic-to-Real Transfer", Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) Workshops, 2025
- [2]: D. Kienzle, M. Kantonis, R. Schön and R. Lienhart, "Segformer++: Efficient Token-Merging Strategies for High-Resolution Semantic Segmentation", IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR), 2024