Spatiotemporal Self-Attention Based Biomimetic Carnivora Visual System for Table Tennis Trajectory Prediction

Li, Zhanran (School: Beijing National Day School)

Instantly forecasting a table tennis ball's trajectory—including spin and landing spot—before it crosses the net while matching elite athlete performance, is a key challenge in spatial intelligence. This study employs a spatiotemporal Transformer inspired by carnivorous vision systems, integrating aerodynamics and self-attention time-window entropy analysis. Unlike previous multi-camera tracking systems, this approach predicts long-horizon future motion features in 3D space from monocular first-person 2D image sequences, offering a novel solution to this problem. Using a self-built 2D video dataset, the model exploits Transformer spatiotemporal self-attention with three key improvements: Biomimetic Carnivora Visual System:Landing-spot prediction accuracy improved by 5.76% with a 42.33% reduction in computational complexity. Integrated Aerodynamic Analysis:Based on the biomimetic enhancements, landing prediction accuracy improved by an additional 3.28%. Combined with self-attention time-window entropy analysis:Spin accuracy improved by 1.59%, with a 10.63% reduction in complexity. The model predicts post-net spin and landing spot using only pre-net videos, achieving 82.99% spin accuracy (+8.89% over average sports academy students) and 93.54% landing accuracy (+4.35% over top athletes). Unlike OMRON and Google DeepMind's real-time stereo vision systems or post-match recognition, this study achieves true prediction—like a real table tennis player, it determines spin and landing spots using only pre-net video, enabling robots to prepare before the ball lands. With shorter video input and lower computational demands, this approach paves the way for low-cost table tennis robots and broader applications in 3D motion prediction from 2D video.

Awards Won: