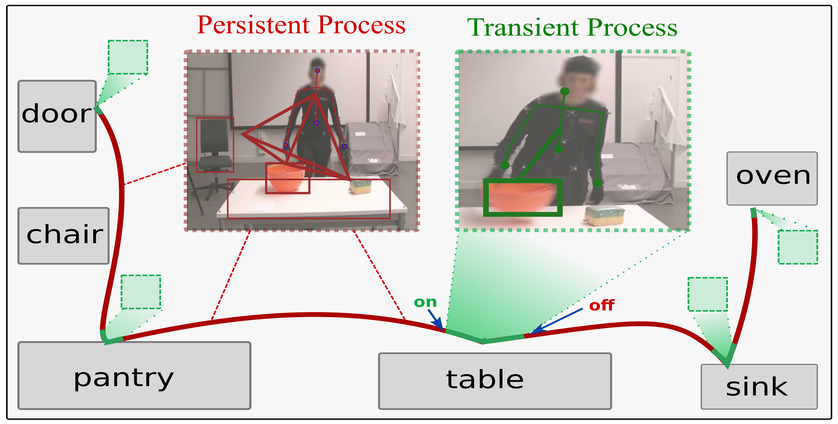
Persistent-Transient Duality: A Multi-mechanism Approach for Modeling Human-Object Interaction

Hung Tran[1], Vuong Le[2], Svetha Venkatesh[1], Truyen Tran[1]
Applied Artificial Intelligence Institute (A2I2), Deakin University, Australia[1], Amazon[2]

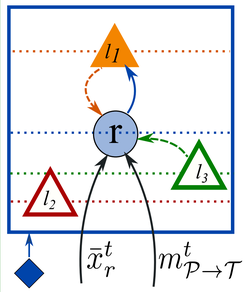
ICCV23 PARIS

Motivation



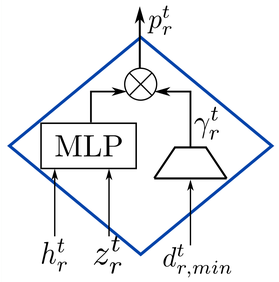
- During HOI activities, humans need to switch between two modes:
  - global navigation plan (red): continuous throughout the activity
  - local human-object interaction (green): occurs intermittently.
- It's key to model the two modes differently and have a switching mechanism between them to accurately represent the activity.

Transient Channel

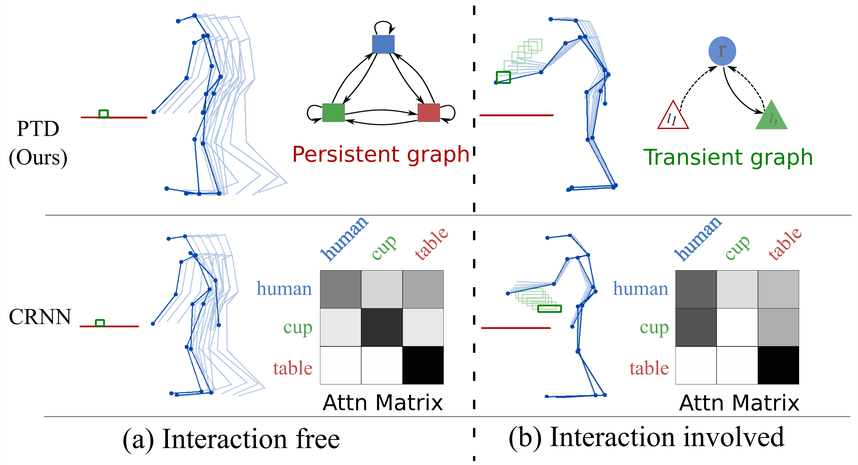


- Reflect the human's perspective during an interaction via:
  - Egocentric Structure
  - Egocentric Represntation
  - Egocentric Inference

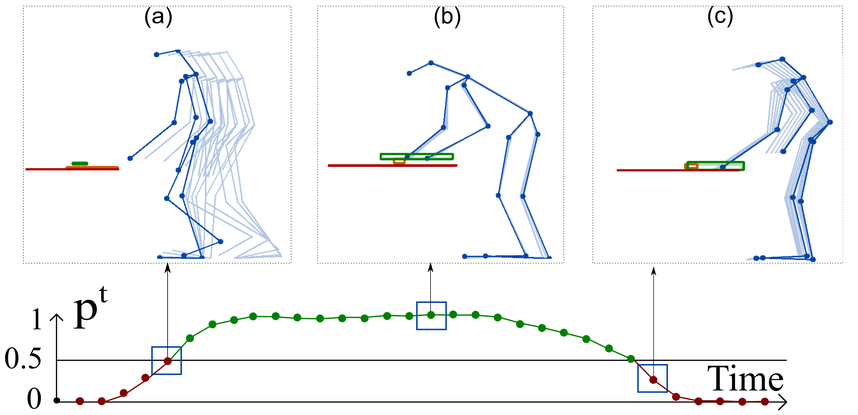
Transient Switch



- Switch on/off Transient channels based on two factors:
  - Spatial evidence
  - Temporal tendency

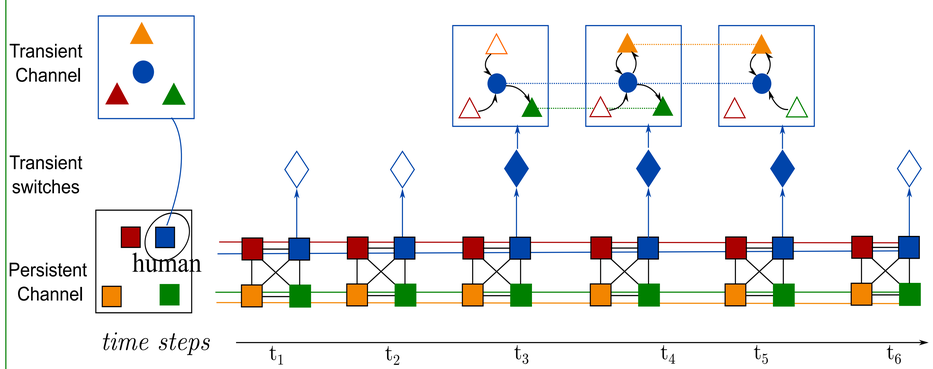


- Qualitative performance: Utilizing a dedicated transient graph yield better HOI modeling compared to SOTA approach (CRNN).



- The Transient Switch (a) anticipate the the begining, (b) stay stable during, and (c) anticipate the end of the interaction.

Persistent-Transient Duality Network



- A parent-child neural network models both global navigation (persistent channel) and local human-object interaction (transient channel).
- The transient channel is designed to reflect the human's unique perspective while interacting with objects. It's life-cycle is controled by a Transient Switch.

Results and Findings

	Human	Obj
Zero-Velocity	176.45	128.6
Running avg. 2	183.95	133.3
GRU [35]	102.86 ± 1.4	119.64 ± 1.6
STS-GCN [40]	101.36 ± 2.4	-
Motion-Mixer [6]	87.35 ± 1.2	-
CRNN-OPM [8]	99.01 ± 1.1	87.52 ± 1.6
CRNN-OPM-LI [8]	95.96 ± 1.7	74.27 ± 1.3
PTD (Ours)	**85.53 ± 0.9**	**70.69 ± 0.5**

- Quantitative performance: Set new SOTA results in both human and object future prediction.

Summary

- Modeling HOI activities using a multi-mechanism neural network with mode switching capability.
- The mode switching mechanism can be learned from data and can reflect the interaction accurately
- State-of-the-art results on two motion prediction datasets
- Generic modeling, applicable to other problems (e.g. pedestrian trajectory prediction)