X-VoE: Measuring eXplanatory Violation of Expectation in Physical Events

Bo Dai[1,2], Linge Wang[3], Baoxiong Jia[2], Zeyu Zhang[2], Song-Chun Zhu[1,2,3], Chi Zhang[2]✉, Yixin Zhu[4]✉
[1] School of Intelligence Science and Technology, Peking University [2] Beijing Institute for General Artificial Intelligence
[3] Department of Automation, Tsinghua University [4] Institute for Artificial Intelligence, Peking University

ICCV23 PARIS
https://yzhu.io/publication/intuitive2023iccv/

Motivation

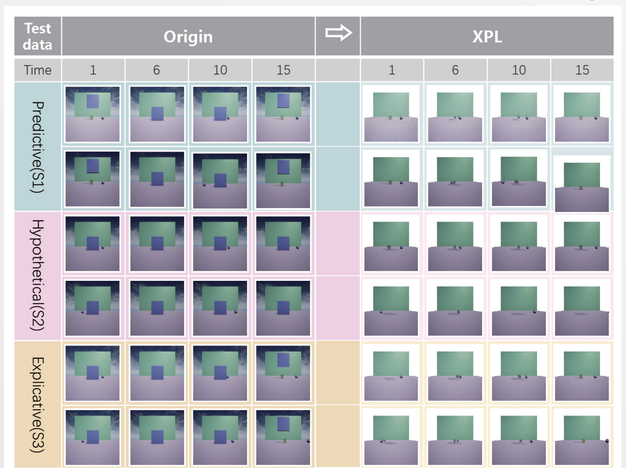
Explanatory Violation of Expectation

	Origin video	Surprise: w-o explain		Surprise: w- explain	Explaining result	Description
predictive (S1)		○	=	○		• Determine the conformity of videos to physical laws.
		!	=	!		
hypothetical (S2)		!	≠	○		• Reason about multiple possibilities in scenarios.
		○	=	○		
explicative (S3)		!	≠	○		• Make informed judgments about physical violations.
		○	≠	!		

- Even as infants, humans are capable of adapting to all three settings effortlessly.
- Existing works typically prioritize prediction without explanation and only work for specific settings.
- X-VoE as a new standard designed to emphasize predictive and interpretive capabilities.

Achievement

Visualization of Three Settings

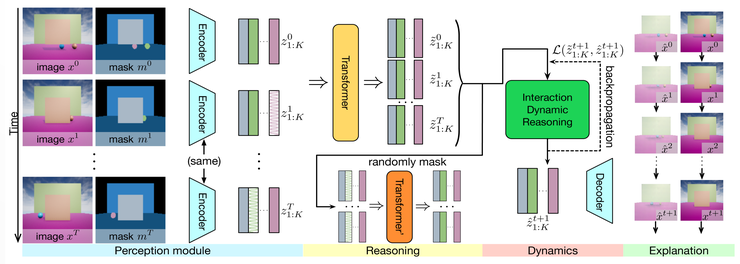


Our research is focused on two key aspects:

- First, we reason about possible representations of occluded objects based on intuitive physics.
- Second, we complete VoE judgments based on the reasoned object representations.

Method

Learning Process of XPL

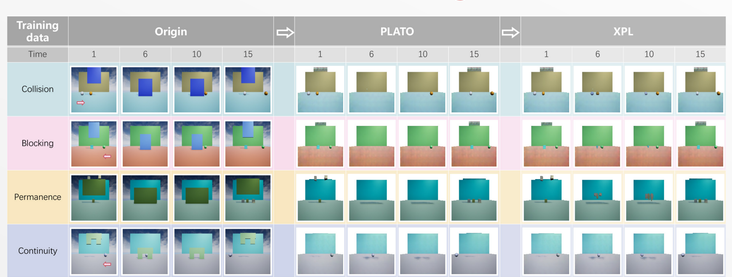


We propose XPL inspired by developmental psychology theories.

- Perception module encodes objects in picture information into vectors using a Component Variational Autoencoder.
- Reasoning module uses transformer to reason about partially occluded objects, avoiding inaccurate information on the physical dynamics model.
- The dynamics module uses the reasoned object information to train a predictive model via a self-supervised approach.
- We visualize explanation results by decoding it and removing the occluding baffle for better visibility.

Explanation

Visualization for Training Dataset

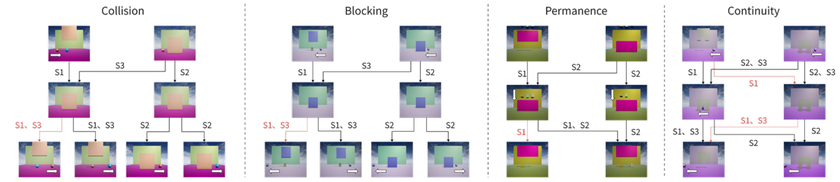


In order to train dynamic modules to learn intuitive physics, it is crucial to consider the presence of occluded objects. The lack of information about these objects can negatively impact the accuracy of the dynamic module in learning intuitive physics.

- The PLATO method, which is used to learn intuitive physics, relies on incorrect information about the representation of occluded objects.
- The XPL method uses more accurate information about the representation of occluded objects, which is obtained through a reasoning process.

Experiment

Four Scenarios of Testing Dataset

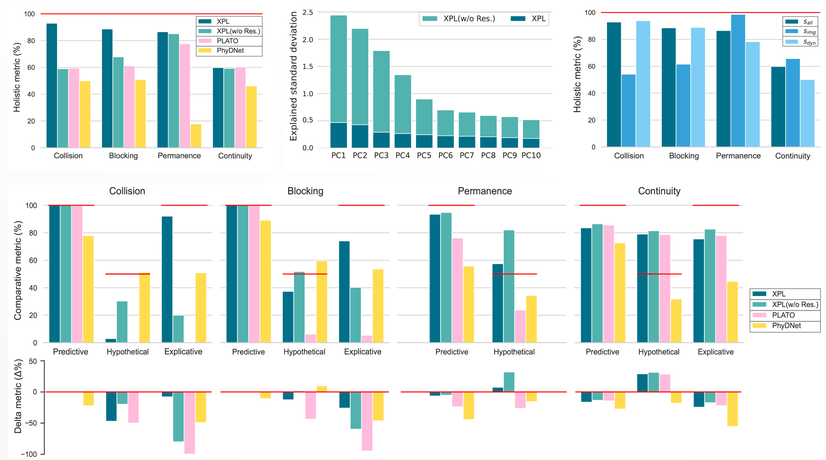


Our X-VoE dataset has four scenarios: ball collision, ball blocking, object permanence, and object continuity, based on three setup modes.

- Predictive setting has a compliant process in black and an unphysical process in red.
- Hypothetical setting has two physically consistent processes in black.
- Explicative setting also has a compliant process and an unphysical process.

Results

Two Metrics for Testing Dataset



Our model's accuracy is based on comparing scores of two videos using surprise scores to identify compliance and violations of intuitive physics laws. We use two metrics, holistic and comparative, to validate our model from different perspectives.

- Holistic metric compares compliant and non-compliant videos in different scenarios. An ideal score is 100%. Our model performs well in the first three scenarios and not too bad in the last one.
- Comparative metric shows that Simple AI predicts simple tasks, but complex tasks need explanatory capabilities for reasoning.