# *Let Me Help You!* Neuro-Symbolic Short-Context Action Anticipation

General Robotics

Carnegie Mellon University Robotics Institute

Sarthak Bhagat*', Samuel Li', Joseph Campbell', Yaqi Xie', Katia Sycara', Simon Stepputtis'

*General Robotics, 'Carnegie Mellon University

IEEE ROBOTICS & AUTOMATION LETTERS
A PUBLICATION OF THE IEEE ROBOTICS & AUTOMATION SOCIETY

## Problem Statement

How can robots quickly and accurately predict human intentions from limited observations to enhance collaborative tasks, such as assistive cooking?

## Knowledge Guided Action Anticipation

### A. Extracting Domain Knowledge

**KG Construction:** We build a knowledge graph with **object nodes** (e.g., *knife*) and **affordance nodes** (e.g., *cuttable*), using off-the-shelf object detectors for initial concept extraction from video frames.

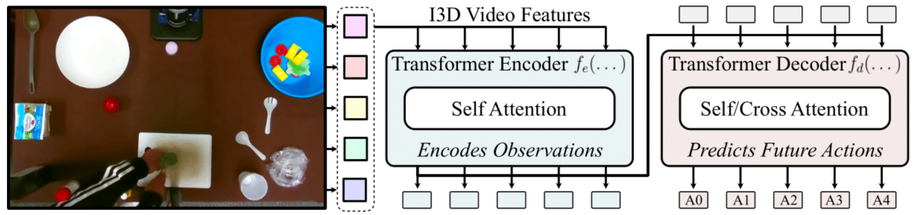
**Concept Expansion (CGS):** CGS iteratively updates relevant concepts using a **Propagation Network** (GATv2) and **Importance Network**, producing a latent representation of objects and their affordances for action anticipation.

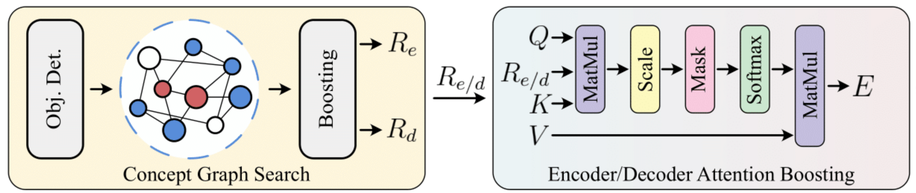
### B. Action Anticipation with Domain Knowledge

**Transformer Architecture:** We adapt a transformer-based pipeline for action anticipation by modifying the **attention mechanism** to integrate domain knowledge, improving contextual reasoning.

**Encoder** extracts visual features from video frames using I3D, producing frame embeddings.
**Decoder** predicts future actions and their durations using these embeddings and learnable action queries.

**Prediction and Confidence:** We estimate action confidence using the **negative entropy** of the softmax distribution over predicted actions.




### C. Knowledge Guided Attention

We introduce a **knowledge-guided attention mechanism** that uses a rectification matrix derived from domain knowledge to boost or attenuate attention between features, enhancing contextual predictions in action anticipation.

$$\text{KG-Attn}_{e/d}(\mathbf{Q},\mathbf{K},\mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{R}_{e/d}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

## Action Anticipation Results

- Outperforms the current state-of-the-art in long-term action anticipation using short context in all the metrics on the 50Salads dataset and on nine out of the ten metrics we used on the Breakfast dataset.
- On the MoC metric, outperforms the baseline by up to 9% on 50Salads and 1% on Breakfast.

## Human Robot Collaboration Results

| Approach | Finetuning | Confidence | Success | | MoC | |
|---|---|---|---|---|---|---|
| | | | $\alpha = 5\%$ | $\alpha = 10\%$ | $\alpha = 5\%$ | $\alpha = 10\%$ |
| Autoregressive | | | 13.0 | 17.4 | 6.2 | 7.4 |
| Autoregressive | ✓ | | 27.3 | 36.4 | 8.9 | 12.2 |
| FUTR | | | 16.6 | 20.8 | 6.7 | 9.2 |
| NeSCA | | | 19.2 | 23.1 | 6.9 | 9.9 |
| NeSCA | ✓ | | 33.7 | 41.8 | 12.4 | 18.1 |
| NeSCA (Full) | | | 35.2 | 43.6 | 14.4 | 20.2 |
| NeSCA | ✓ | ✓ | 42.8 | 50.1 | - | - |

## Contributions

We introduce a knowledge-guided action anticipation method that leverages object affordances to enhance short-horizon prediction accuracy in human-robot collaboration tasks.

## Takeaways

Our approach improves prediction speed and accuracy, enabling more responsive and effective robotic assistance in collaborative tasks like salad preparation.

**Paper, Code and Dataset**

**Acknowledgements:** We would like to acknowledge the support from DARPA under grant FA8750-23-2-1015, AFOSR under grants FA9550-18-1-0251 and FA9550-18-1-0097, and ARL under grant W911NF-19-2-0146 and W911NF-2320007.