Why Not Use Your Textbook? Knowledge-Enhanced Procedure Planning of Instructional Videos

Kumaranage Ravindu Yasas Nagasinghe, Honglu Zhou, Malitha Gunawardhana, Martin Renqiang Min, Daniel Harari, Muhammad Haris Khan

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

NEC Laboratories America

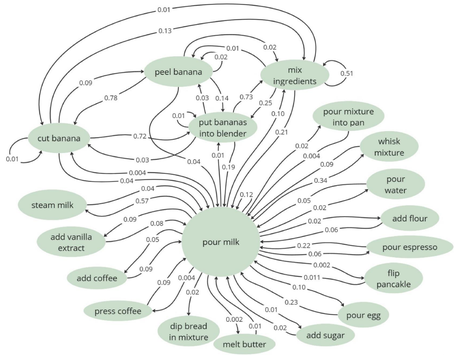
CVPR SEATTLE, WA JUNE 17-21, 2024

Contributions

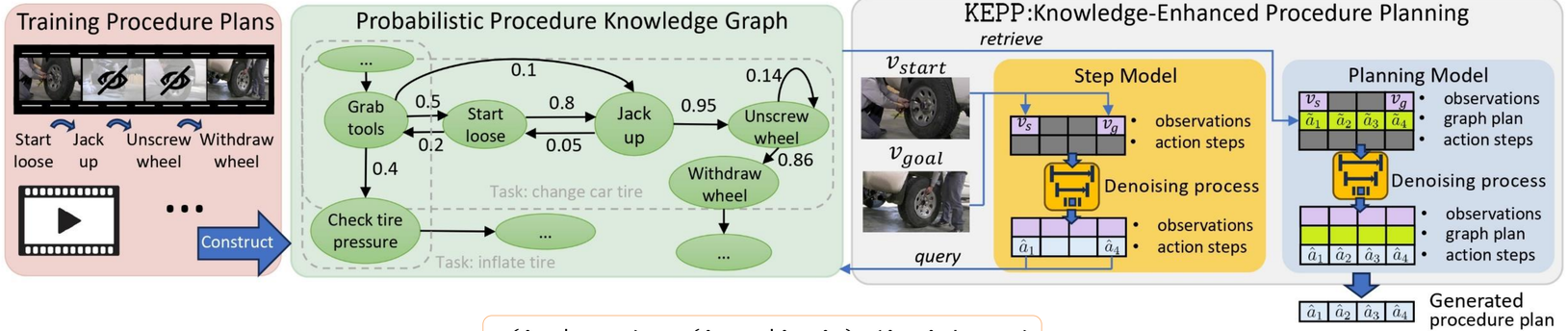
- We propose KEPP, a Knowledge-Enhanced Procedure Planning system for instructional videos that leverages rich procedural knowledge from a probabilistic procedural knowledge graph (P^2KG).
- Requires a minimal number of annotations for supervision.
- Decompose the problem in procedure planning of instructional videos:
  - predicting the initial and final steps from the start and end visuals, and then creating a plan using procedural knowledge retrieved based on these predicted steps.
- Experimental evaluations on three datasets, under settings of varying complexity, reveal that KEPP attains state-of-the-art results in procedure planning.

Introduction

- Problem Statement : Generate procedural plans with minimal supervision considering causal constraints in the sequencing of steps and the variability inherent in multiple feasible plans.
- Solution: Infuse procedure planning with comprehensive procedural knowledge, derived from training procedure plans and structured as a directed weighted graph.



Method



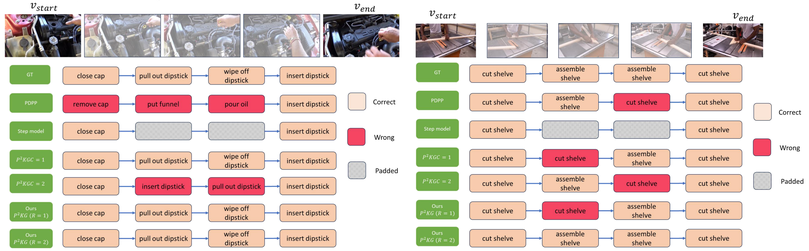
- Decompose procedure planning problem: $p(\hat{a}_{1:T}|v_s, v_g) = p(\hat{a}_{2:T-1}|\hat{a}_1, \hat{a}_T)p(\hat{a}_1, \hat{a}_T|v_s, v_g)$
- Harnessing a Probabilistic Procedural Knowledge Graph (P^2KG): $p(\hat{a}_{1:T}|v_s, v_g) = p(\hat{a}_{1:T}|\tilde{a}_{1:T}, v_s, v_g)p(\tilde{a}_{1:T}|\hat{a}_1, \hat{a}_T)p(\hat{a}_1, \hat{a}_T|v_s, v_g)$
- Adopting conditioned projected diffusion Model as the Step Model and Planning Model

$$\begin{bmatrix} \hat{v}_1 & \hat{v}_2 & \dots & \hat{v}_{T-1} & \hat{v}_T \\ \hat{a}_1 & \hat{a}_2 & \dots & \hat{a}_{T-1} & \hat{a}_T \end{bmatrix} \xrightarrow{\text{Projection}} \begin{bmatrix} v_s & 0 & \dots & 0 & v_g \\ \hat{a}_1 & 0 & \dots & 0 & \hat{a}_T \end{bmatrix} \text{ Step model}$$

$$\begin{bmatrix} v_s & 0 & \dots & 0 & v_g \\ \tilde{a}_1 & \tilde{a}_2 & \dots & \tilde{a}_{T-1} & \tilde{a}_T \\ a_1 & a_2 & \dots & a_{T-1} & a_T \end{bmatrix} \text{ Planning model}$$

Results

- '$P^2KGC = 1$' and '$P^2KGC = 2$' indicates the first and second paths obtained from the probabilistic procedure knowledge graph, respectively.



Conclusion

- KEPP employs a P^2KG, sourced from the training domain, effectively serving as a 'textbook' for procedure planning.
- KEPP delivers top-tier performance while necessitating only a minimal amount of supervision.

Models	T = 3	T = 4	T=5	T=6
PDPP	26.38	18.69	13.22	7.49
Skip Plan	28.85	15.56	8.55	5.12
Ours (R=1)	33.34	20.38	**13.25**	8.09
Ours (R=2)	**33.38**	**21.02**	12.74	**9.23**

SR comparison on CrossTask

SCAN ME