United We Stand, Divided We Fall: UnityGraph for Unsupervised Procedure Learning from Videos

Siddhant Bansal, Chetan Arora, C.V. Jawahar

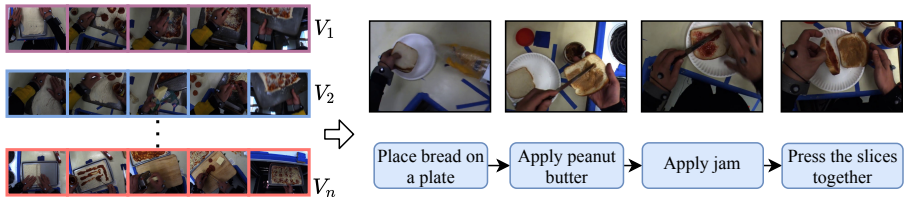
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

CVIT

INDIAN INSTITUTE OF TECHNOLOGY DELHI

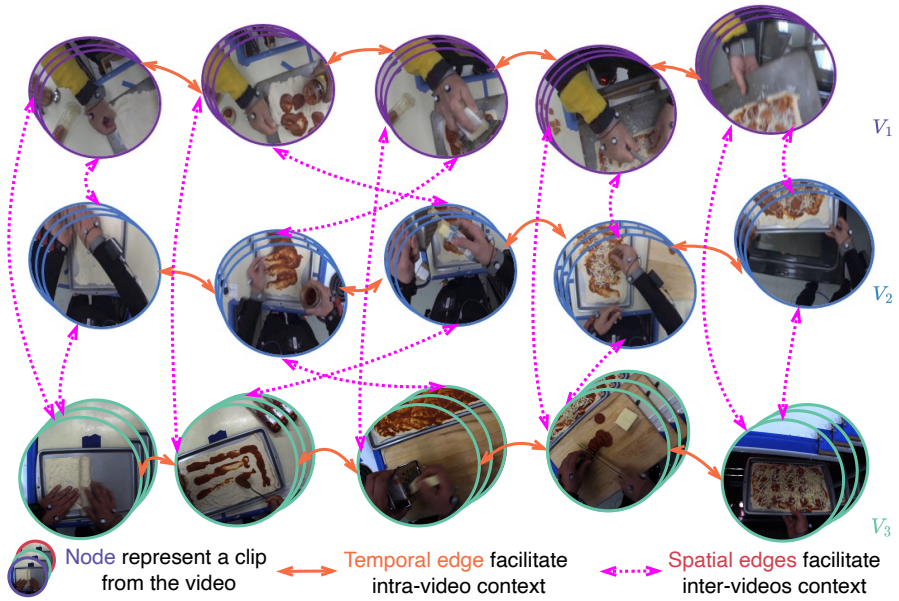
JAN 4-8 WACV 2024 WAIKOLOA HAWAII

What is Procedure Learning?



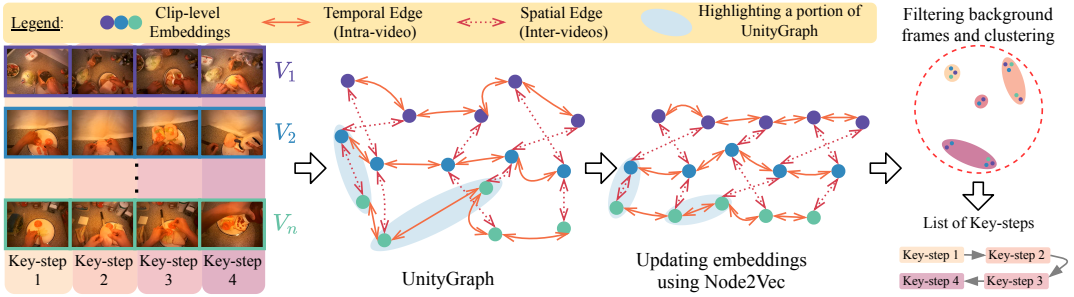
Given a few task's videos, the aim is to identify the key-steps required to perform the task.

UnityGraph: Using Graphs for Video Representation



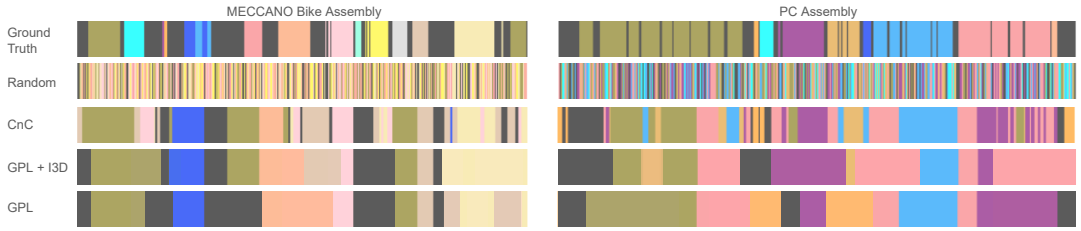
- UnityGraph facilitates procedure learning by creating a unified representation of an arbitrary number of videos from the same category.
- The nodes represent a clip from the video. Further, the temporal edges connect temporally close frames, allowing **intra-video context**, whereas the spatial edges connect semantically similar frames across the videos, enabling **inter-videos context**.

Graph-based Procedure Learning (GPL) Framework



- From multiple videos of the same task, we create **UnityGraph**.
- Using the **Node2Vec** algorithm, we exploit the structure of UnityGraph to **enhance the node embeddings in an unsupervised manner.**
- We cluster the embeddings using **KMeans** and filter the background frames to obtain the key-steps.

Results and Analysis



Qualitative Results for MECCANO and PC Assembly

	EgoProceL											
	CMU-MMAC		EGTEA G.		MECCANO		EPIC-Tents		PC Assembly		PC Disassembly	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
Random	15.7	5.9	15.3	4.6	13.4	5.3	14.1	6.5	15.1	7.2	15.3	7.1
CnC [3]	22.7	11.1	21.7	9.5	18.1	7.8	17.2	8.3	25.1	12.8	27.0	14.8
GPL-2D (*ours*)	21.8	11.7	23.6	14.3	18.0	8.4	17.4	8.5	24.0	12.6	**27.4**	**15.9**
UG-I3D (*ours*)	28.4	15.6	25.3	14.7	18.3	8.0	16.8	8.2	22.0	11.7	24.2	13.8
GPL (*ours*)	**31.7**	**17.9**	**27.1**	**16.0**	**20.7**	**10.0**	**19.8**	**9.1**	**27.5**	**15.2**	26.7	15.2

Qualitative Results on EgoProceL. Here GPL has the best results this highlights the effectiveness of the proposed **UnityGraph** framework for procedure learning.