

VideoGraph

Recognizing Minutes-Long Human Activities



Nouredien Hussein



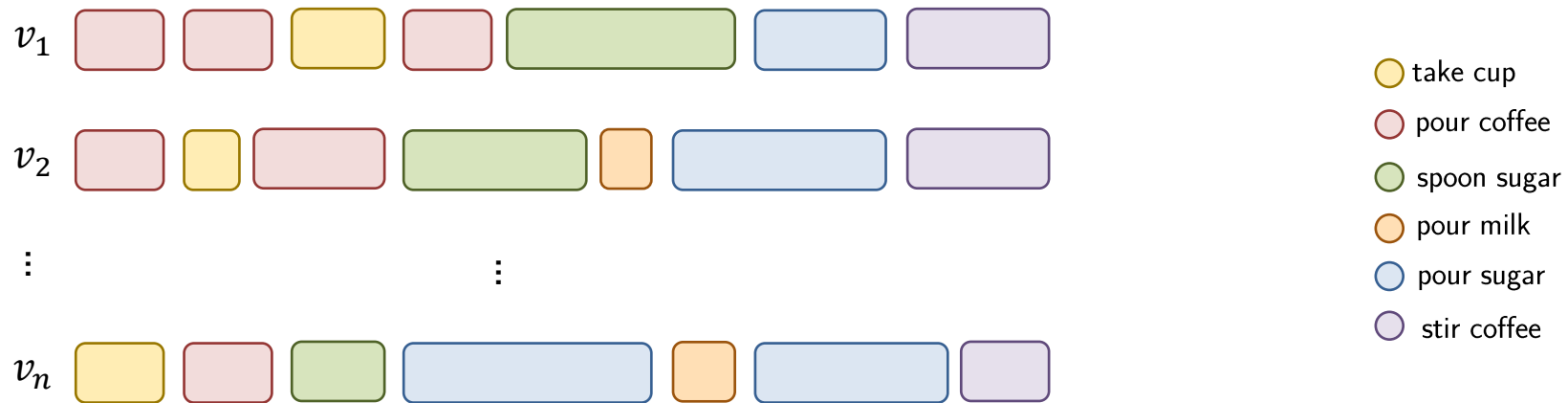
Efstratios Gavves



Arnold Smeulders

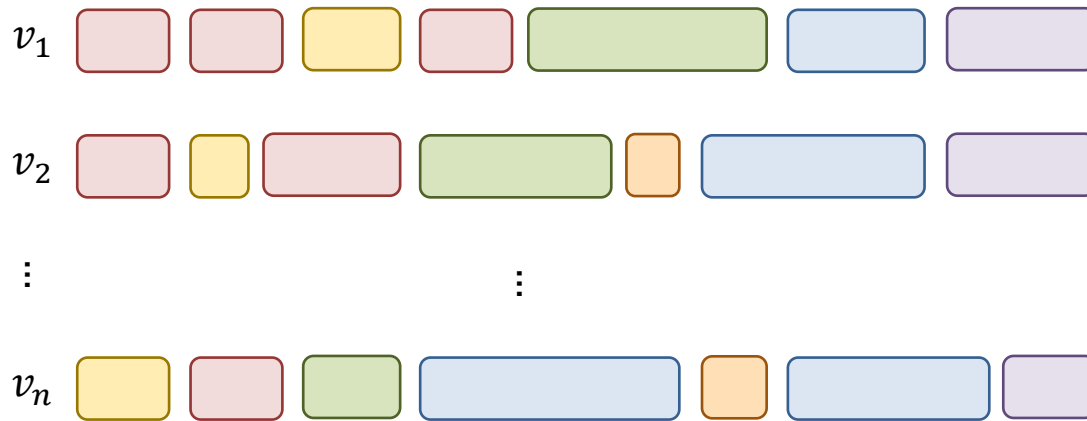
Problem

Graph-Based Representation

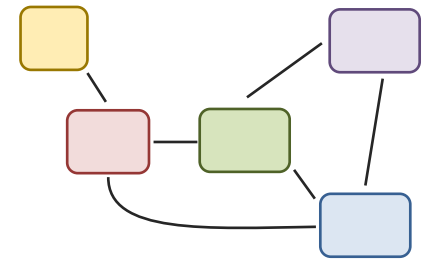
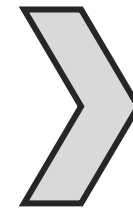


Video Examples of "Preparing Coffee"

Graph-Based Representation



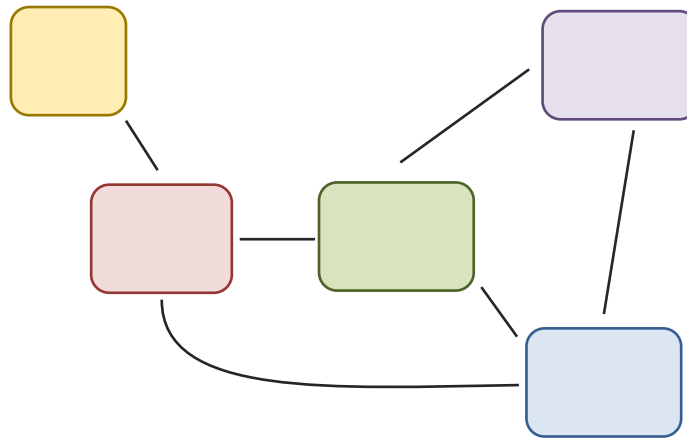
Video Examples of "Preparing Coffee"



Graph-based Representation

Graph-Based Representation

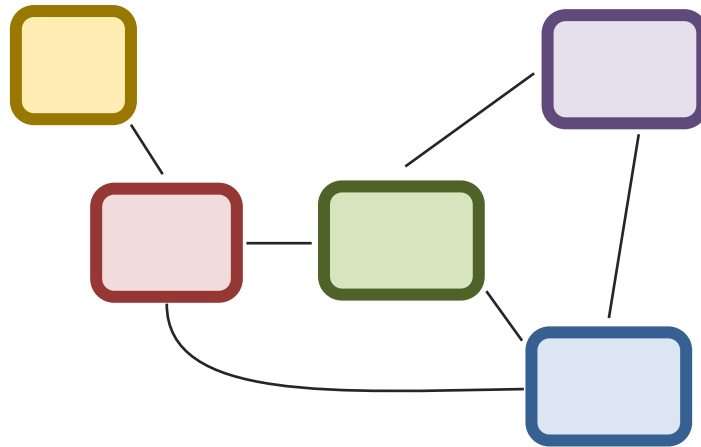
- Graph-based representation



Graph-based Representation

Graph Nodes

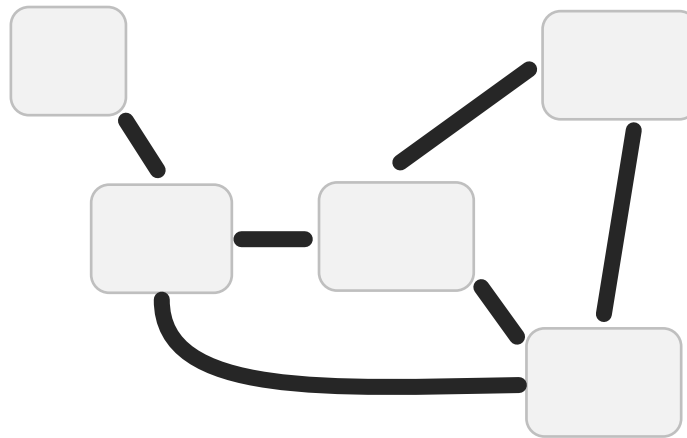
- Graph-based representation
 - Nodes: represent key one-actions in the activity



Graph-based Representation

Graph Edges

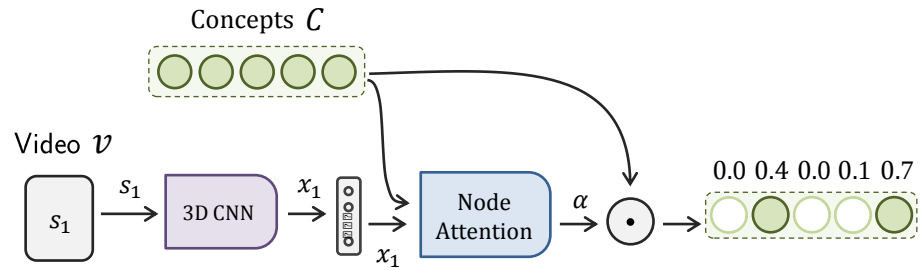
- Graph-based representation
 - Nodes: represent key one-actions in the activity
 - Edges: encode the temporal relationship between one-actions



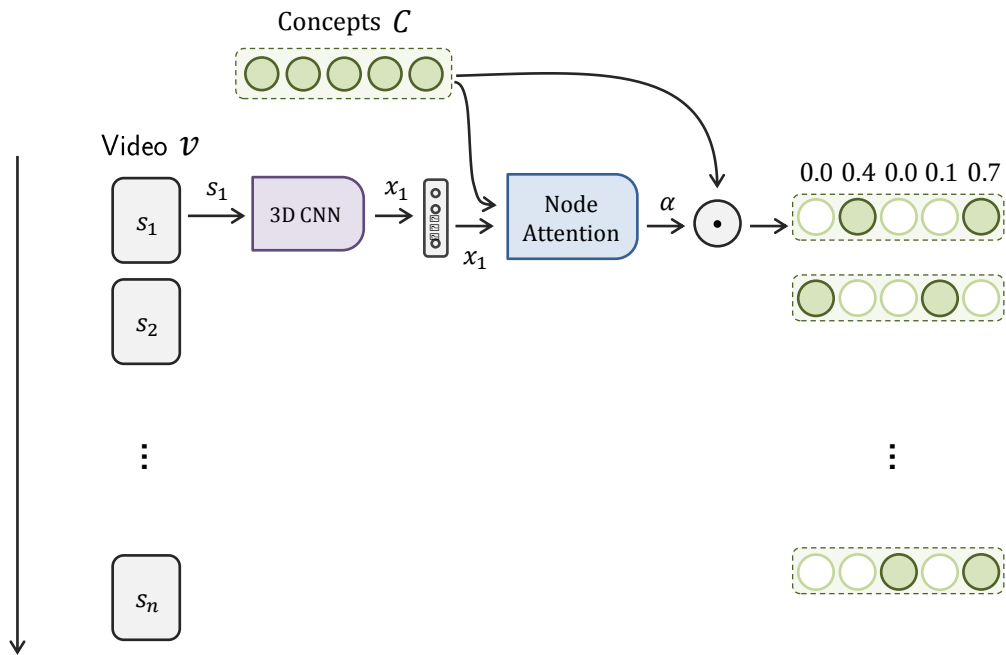
Graph-based Representation

Method

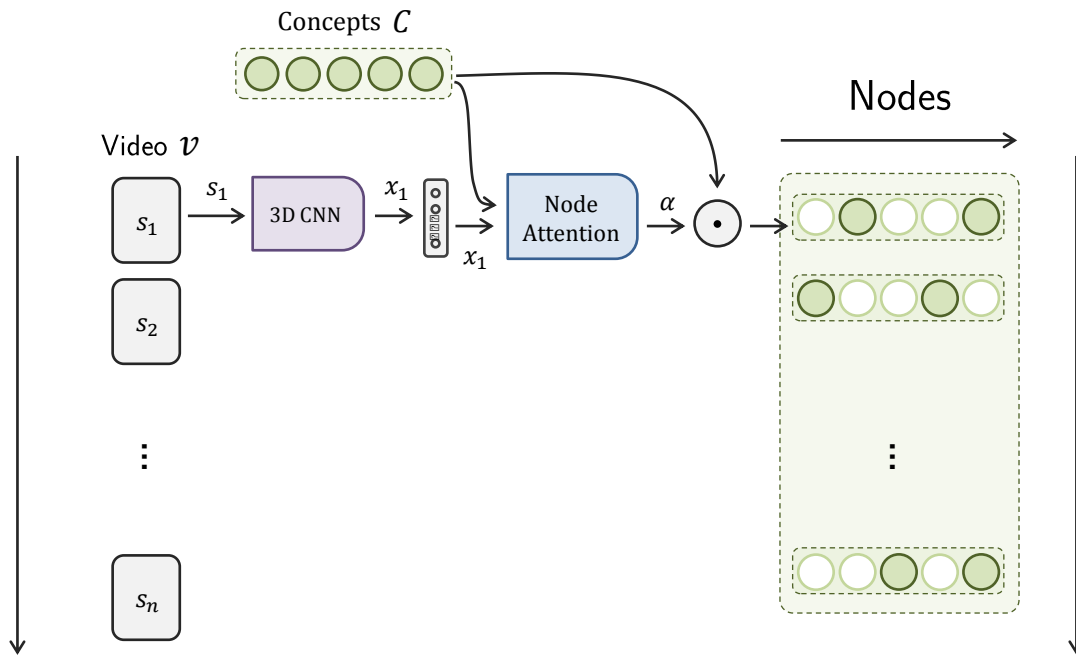
Node Attention



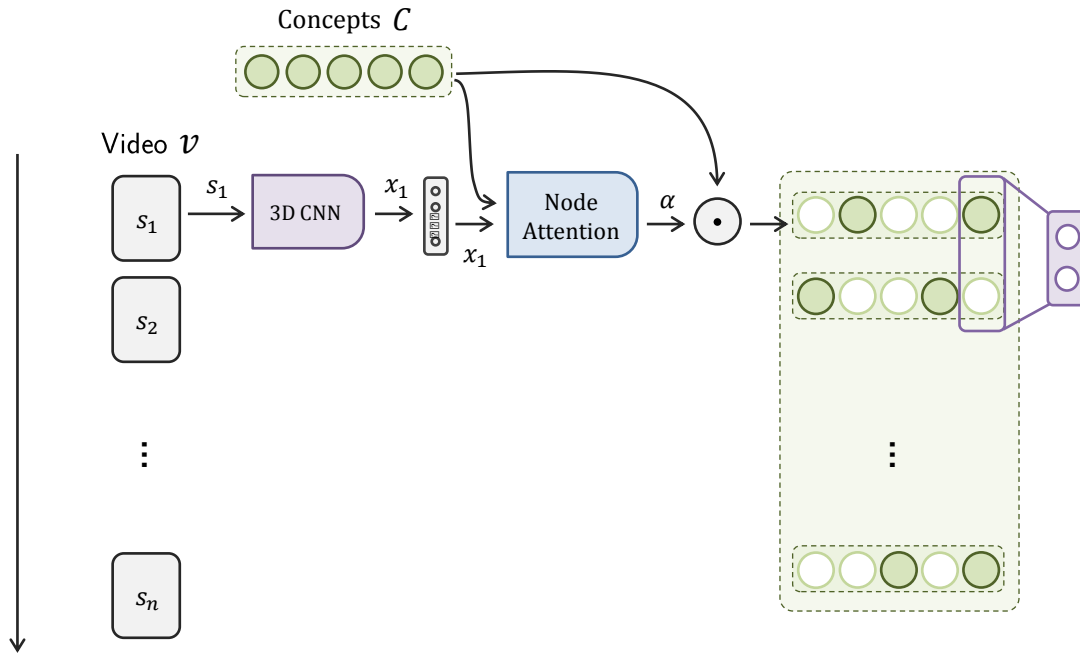
Node Attention



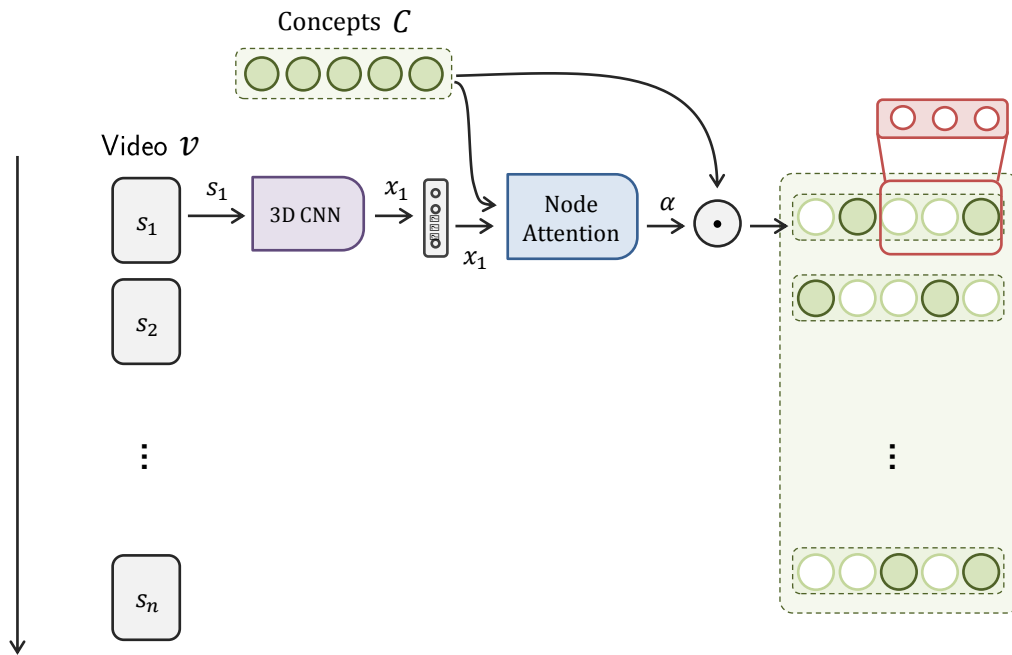
Node Attention



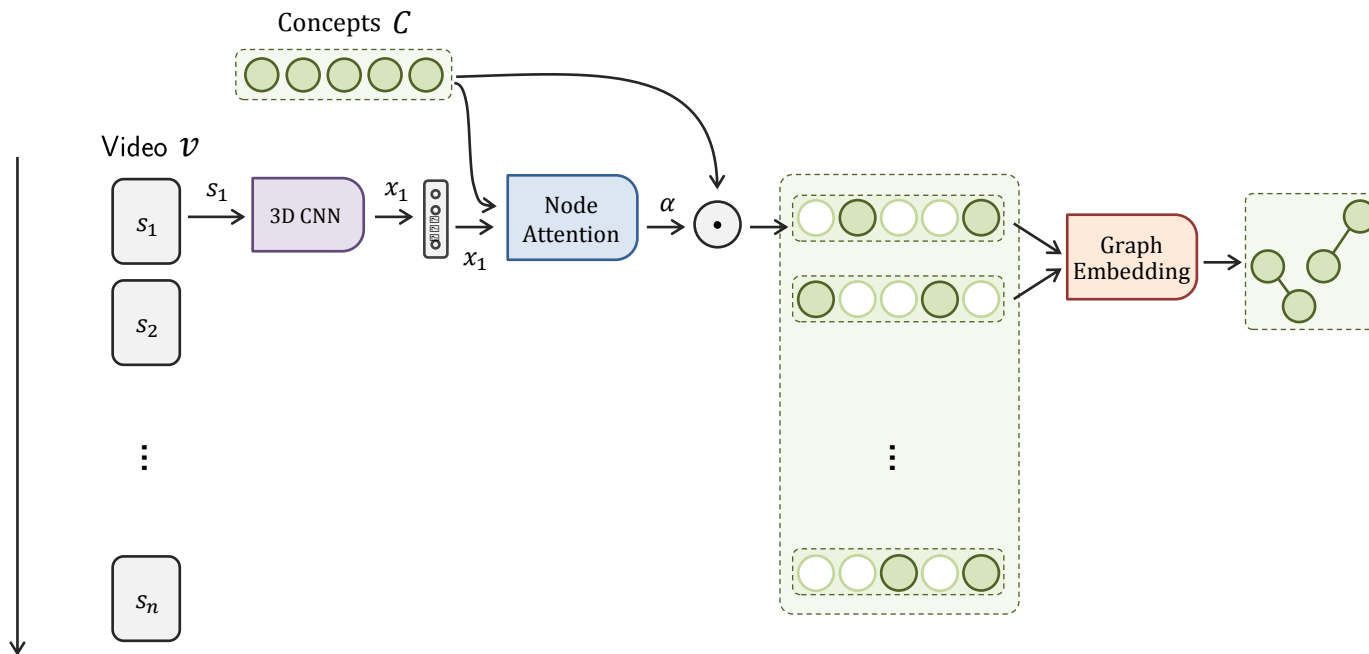
Node Attention



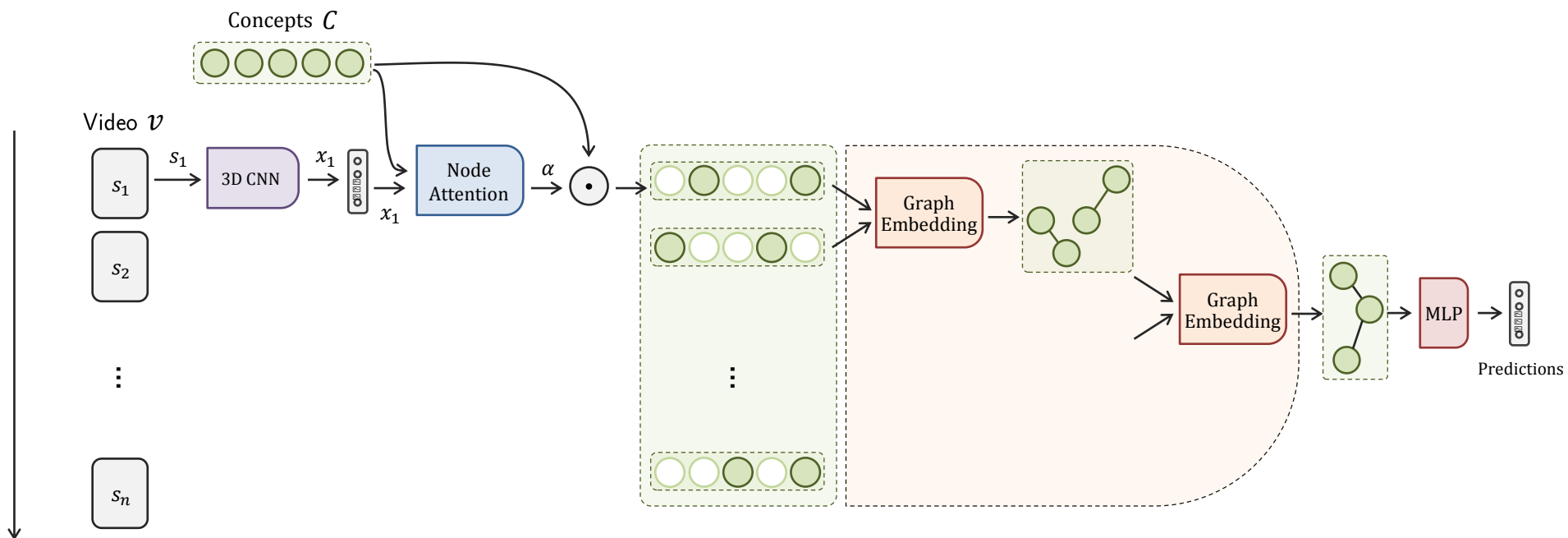
Node Attention



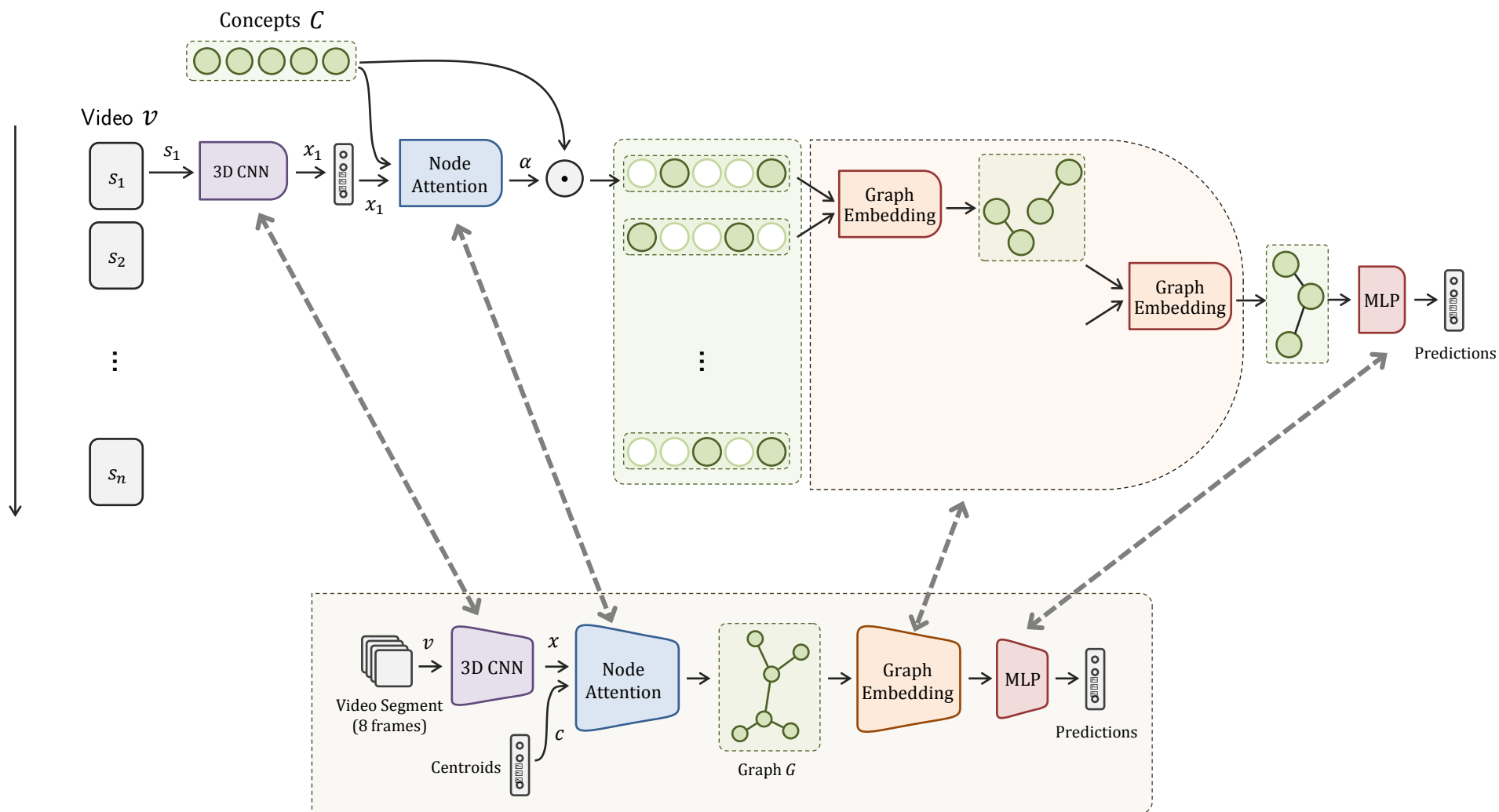
Graph Embedding



Graph Embedding



VideoGraph



Results

Related Methods

Method	Minutes-long	Temporal Pattern
Orderless	✓	
Order-ware		✓
VideoGraph	✓	✓

Related Methods

Related Methods

Method	Minutes-long	Temporal Pattern
Orderless ActionVLAD	✓	
Order-ware I3D		✓
VideoGraph	✓	✓

Related Methods

Performance Comparison

Method	Modality	mAP (%)
Two-stream [17]	RGB + Flow	18.6
Two-stream + LSTM [17]	RGB + Flow	17.8
ActionVLAD [5]	RGB + iDT	21.0
Temporal Fields [17]	RGB + Flow	22.4
Temporal Relations [23]	RGB	25.2
ResNet-152 [61]	RGB	22.8
ResNet-152 + Timeception [2]	RGB	31.6
I3D [9]	RGB	32.9
I3D + ActionVLAD [5]	RGB	35.4
I3D + Timeception [2]	RGB	37.2
I3D + VideoGraph	RGB	37.8

Method	Breakfast Acc. (%)	Breakfast mAP (%)
ResNet-152 [61]	41.13	32.65
ResNet-152 + ActionVLAD [5]	55.49	47.12
ResNet-152 + Timeception [2]	57.75	48.47
ResNet-152 + VideoGraph	59.12	49.38
I3D [9]	47.05	58.61
I3D + ActionVLAD [5]	60.20	65.48
I3D + Timeception [2]	61.82	67.07
I3D + VideoGraph	63.14	69.45

Method	EPIC-Kitchens mAP (%)
I3D [9]	48.86
I3D + ActionVLAD [5]	51.45
I3D + Timeception [2]	55.46
I3D + VideoGraph	55.32

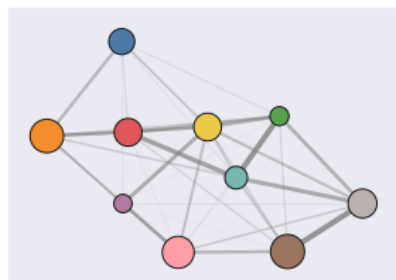
Graph Visualization



(a) Making Cereals



(b) Preparing Coffee



(c) Frying Eggs



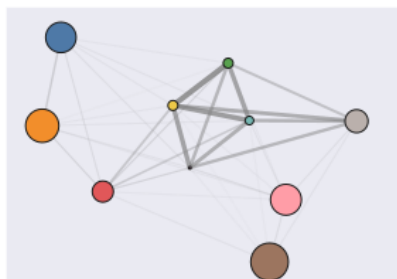
(d) Making Juice



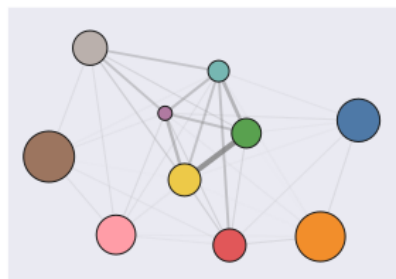
(e) Preparing Milk



(f) Making Pancake



(g) Making Salat



(h) Making Sandwich



(i) Making Scrambled Egg



(j) Preparing Tea

● cereal, ● pan, ● eggs, ● sandwich, ● kettle, and ● foodbox

Node Visualization



● cereal, ● pan, ● eggs, ● sandwich, ● kettle, and ● foodbox