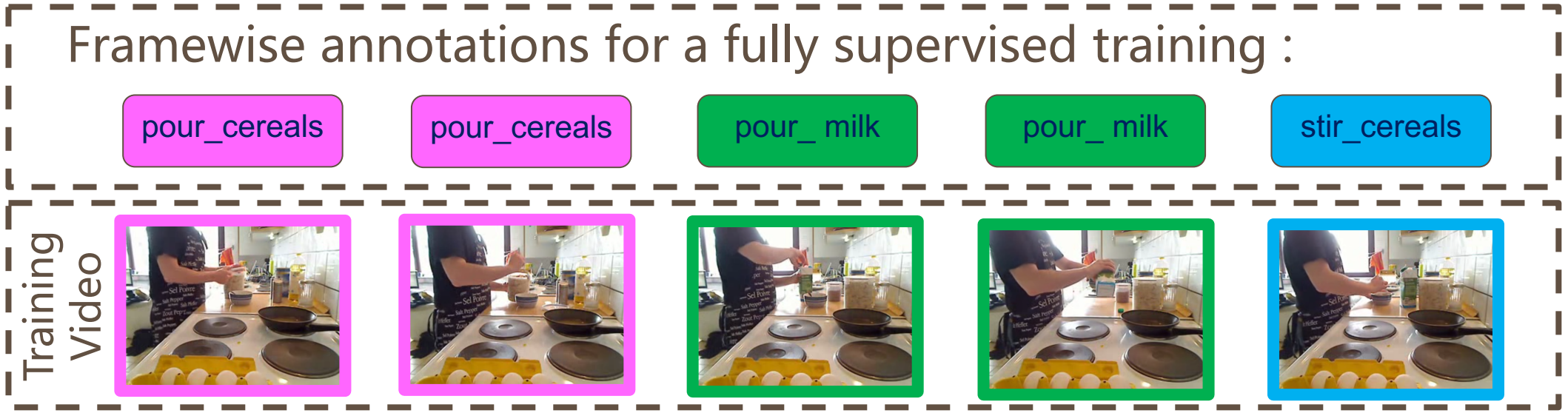
Weakly Supervised Energy-Based Learning for Action Segmentation

Jun Li, Peng Lei, Sinisa Todorovic
Oregon State University

Paper ID:1820

Problem Statement

Given an untrimmed video, label every frame with action classes under weakly supervised training

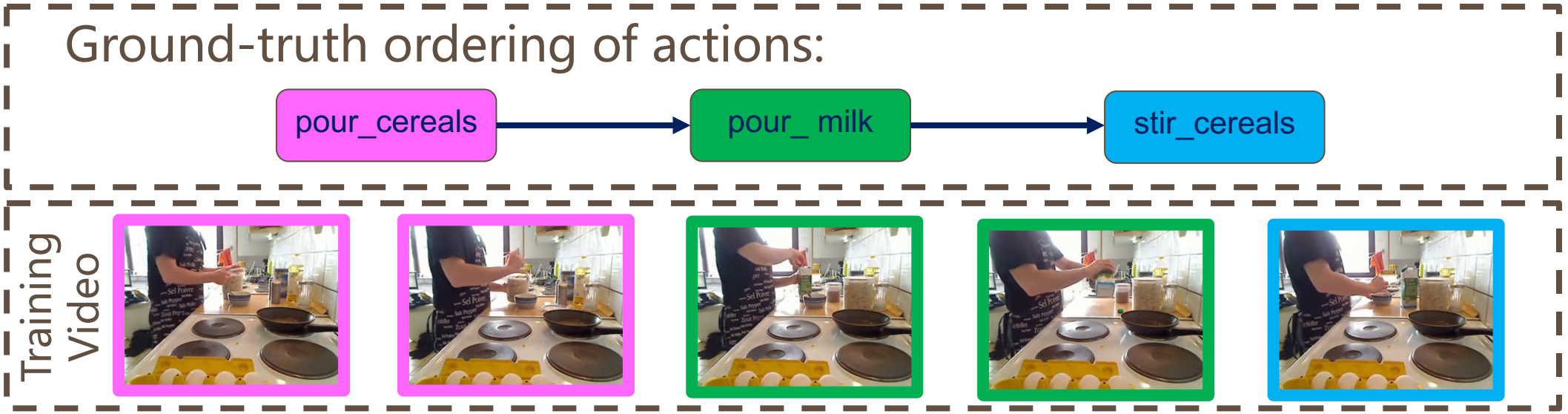


Problem Statement

Given an untrimmed video, label every frame with action classes

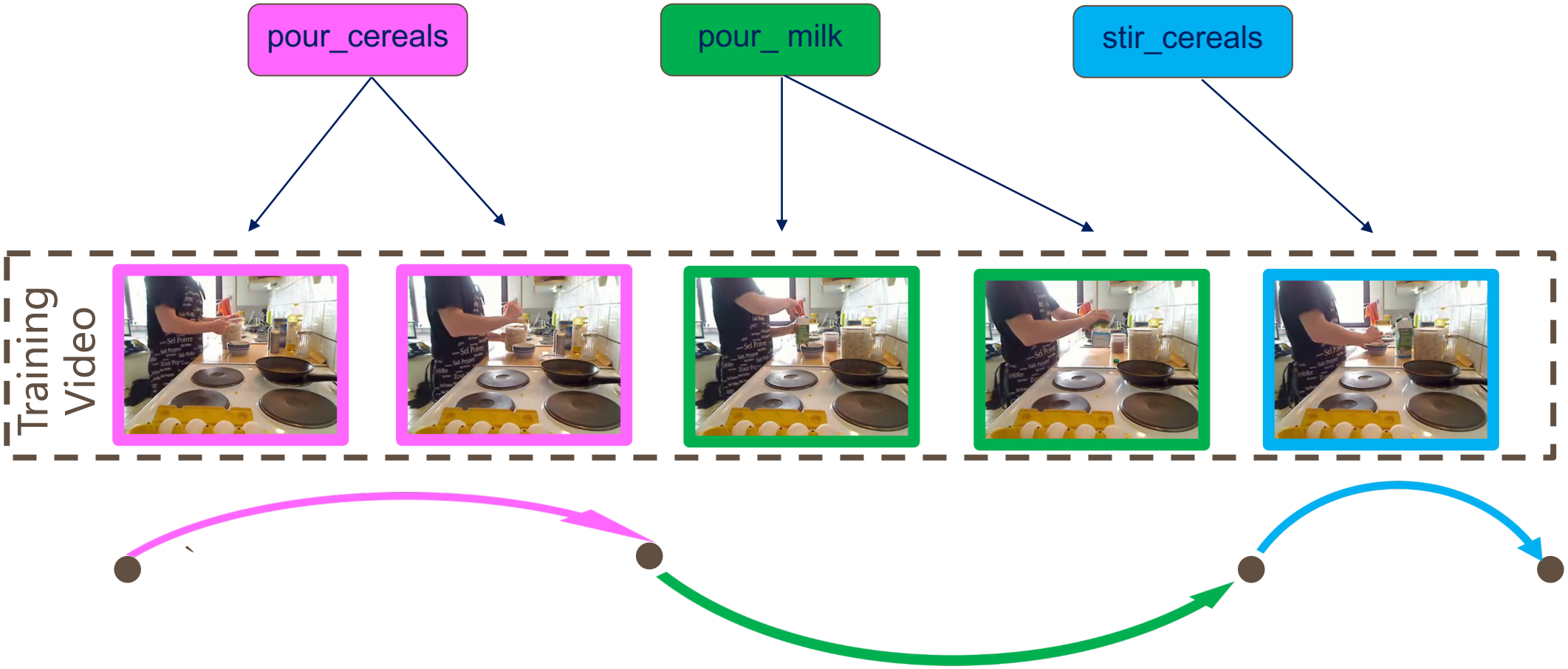
under weakly supervised training –

annotations in training are only temporal orderings of actions



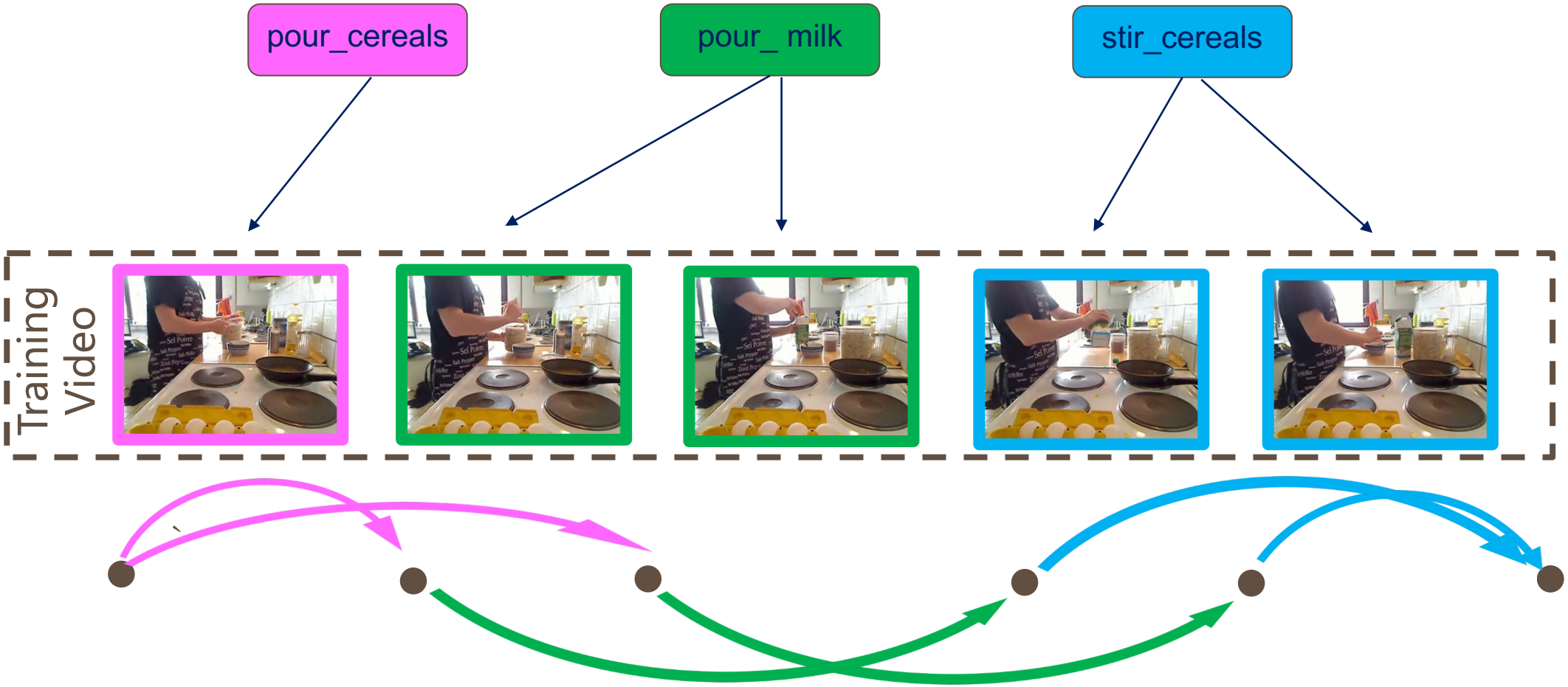
Challenge

Multiple **legal** segmentations that respect the ground-truth

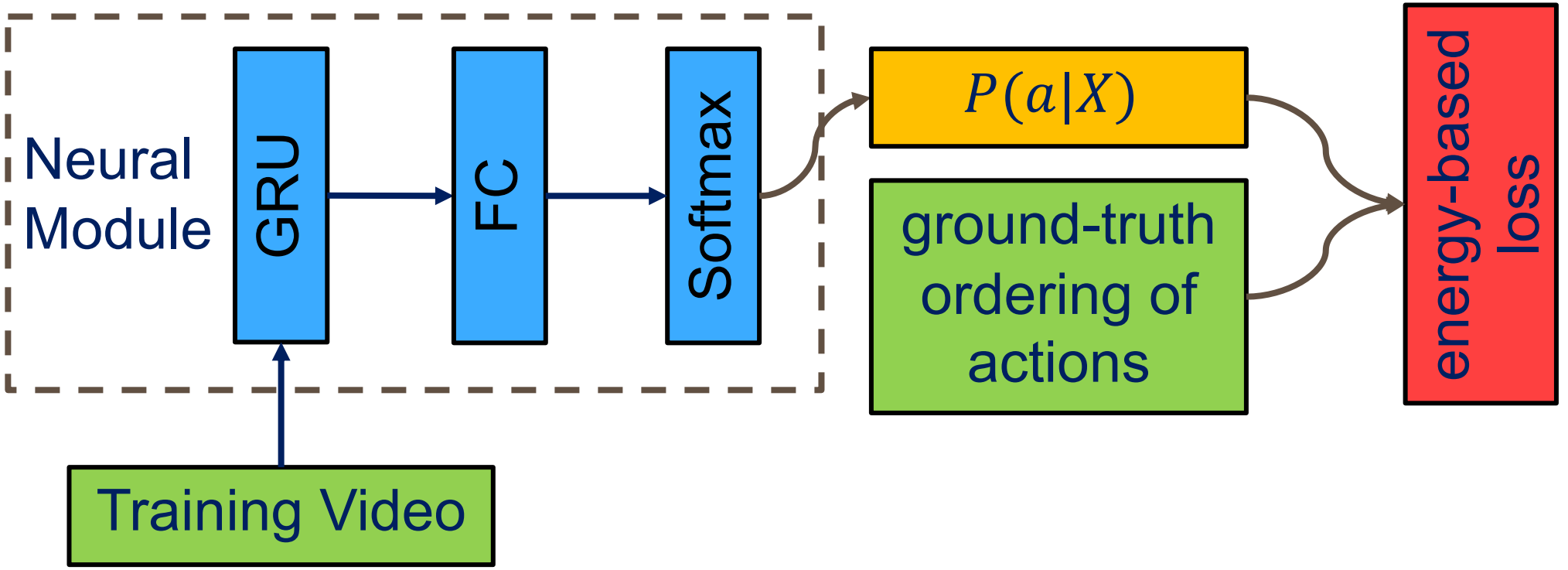


Challenge

Multiple **legal** segmentations that respect the ground-truth



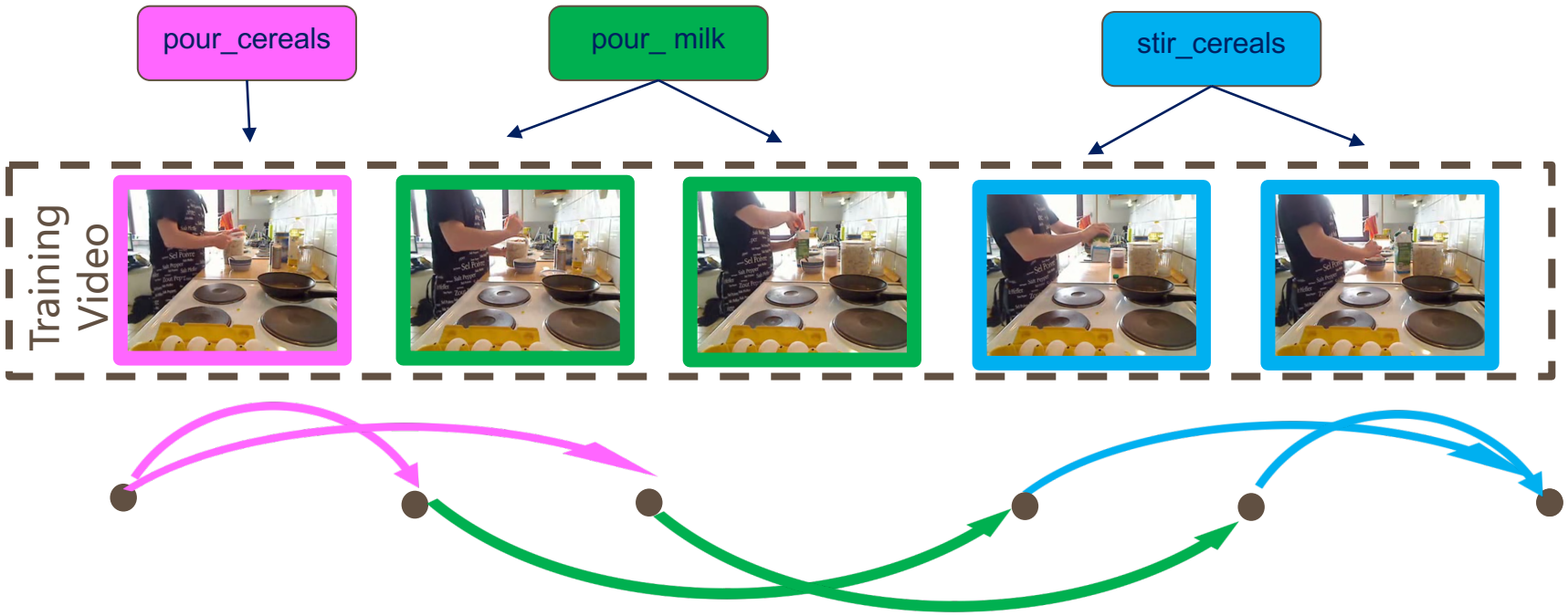
Overview of Our Training



Our Training

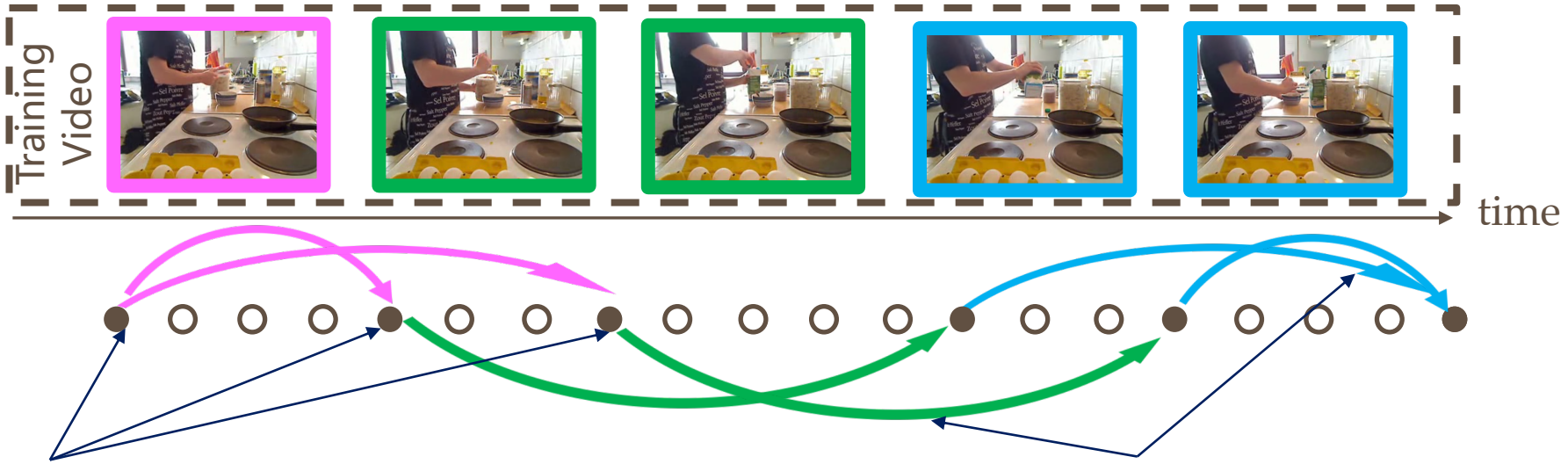
1. For every video, account for all legal and illegal video segmentations

2. Estimate an energy-based loss of all segmentations for learning



Segmentation Graph

Training video → Segmentation graph → Paths = Candidate segmentations



nodes = cuts

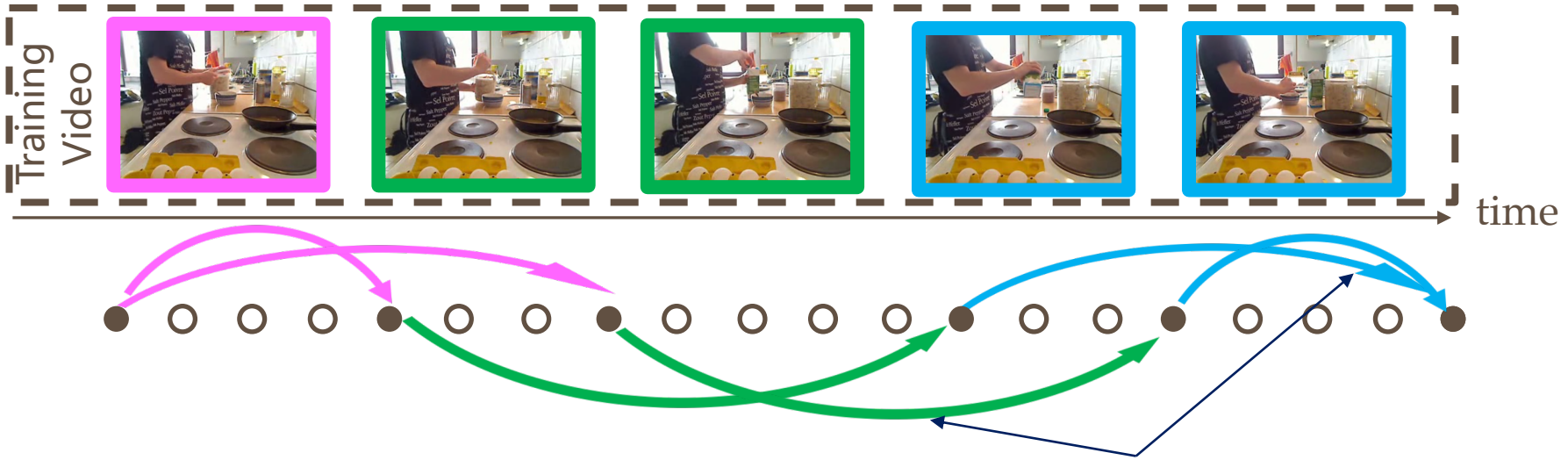
edges = segments between every pair of cuts

The cuts are initially estimated using the Viterbi algorithm,

but later allowed to move in a local temporal neighborhood.

Segmentation Graph

Training video → Segmentation graph → Paths = Candidate segmentations



edge weights:

edges = segments between every pair of cuts

$$w_{ii'}(a) = \sum_{t \in (i,i')} -\log p(a|x_t)$$

GRU's softmax score for action a at frame t

Energy-based Loss for Learning

energy of path π → $E_\pi = \sum_{(i,i')\in\pi} w_{ii'}(a)$ ← weights of edges along path π

Energy-based Loss for Learning

loss

$$L = \text{logadd}\{E_\pi : \pi \in \text{valid paths}\} - \alpha\, \text{logadd}\{E_\pi : \pi \in \text{invalid paths}\}$$

total energy of all valid paths

total energy of all in valid paths

Energy-based Loss for Learning

loss

$$L = \text{logadd}\{E_\pi : \pi \in \text{valid paths}\} - \alpha\, \text{logadd}\{E_\pi : \pi \in \text{invalid paths}\}$$

total energy of all valid paths

total energy of all in valid paths

where: $\text{logadd}(u, v) = -\log(\exp(-u) + \exp(-v))$

Energy-based Loss for Learning

Our key contribution:

Use logadd for efficient computation of the total energy

$$L = \text{logadd}\{E_\pi : \pi \in \text{valid paths}\} - \alpha\, \text{logadd}\{E_\pi : \pi \in \text{invalid paths}\}$$

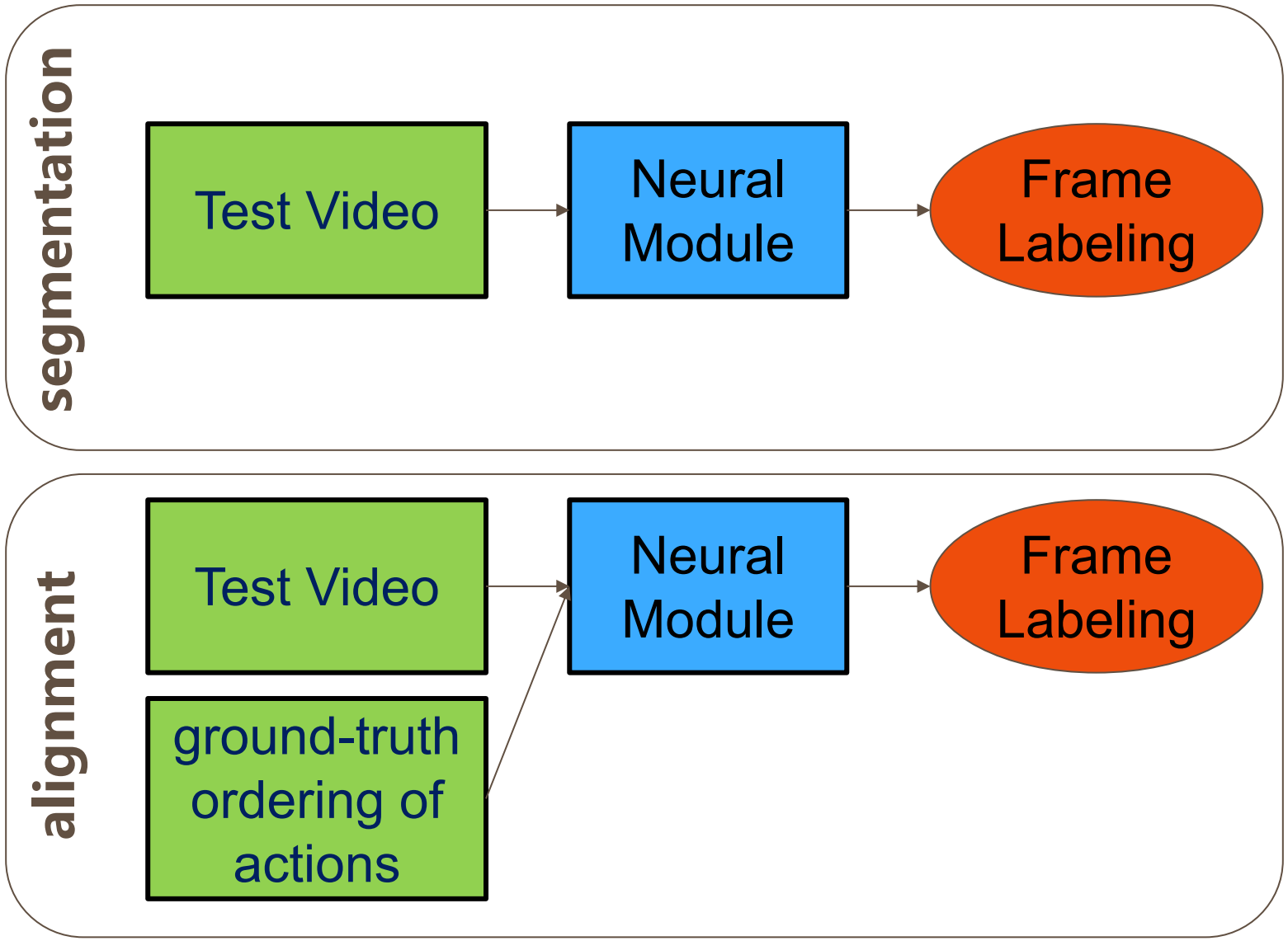
Energy-based Loss for Learning

Our key contribution:

$$L = \text{logadd}\{E_\pi : \pi \in \text{valid paths}\} - \alpha\, \text{logadd}\{E_\pi : \pi \in \text{invalid paths}\}$$

the logadd allows for a **recursive computation** of the total energy

Experiments: Action Segmentation & Alignment



Action Segmentation: Framewise Accuracy

Dataset	We outperform the state of the art by
Breakfast	**4.7%**
Hollywood Ext.	**11.4%**

Action Alignment: Intersection over Detection

Dataset	We outperform the state of the art by
Breakfast	**7.6%**
Hollywood Ext.	**2.0%**

Qualitative Results

We may miss the true start and end of some actions, but our action detect is generally good.

A sample test video *P03_stereo01_P03_milk* from Breakfast dataset.