# Anchor-Constrained Viterbi for Set-Supervised Action Segmentation

Jun Li, Sinisa Todorovic

Oregon State University

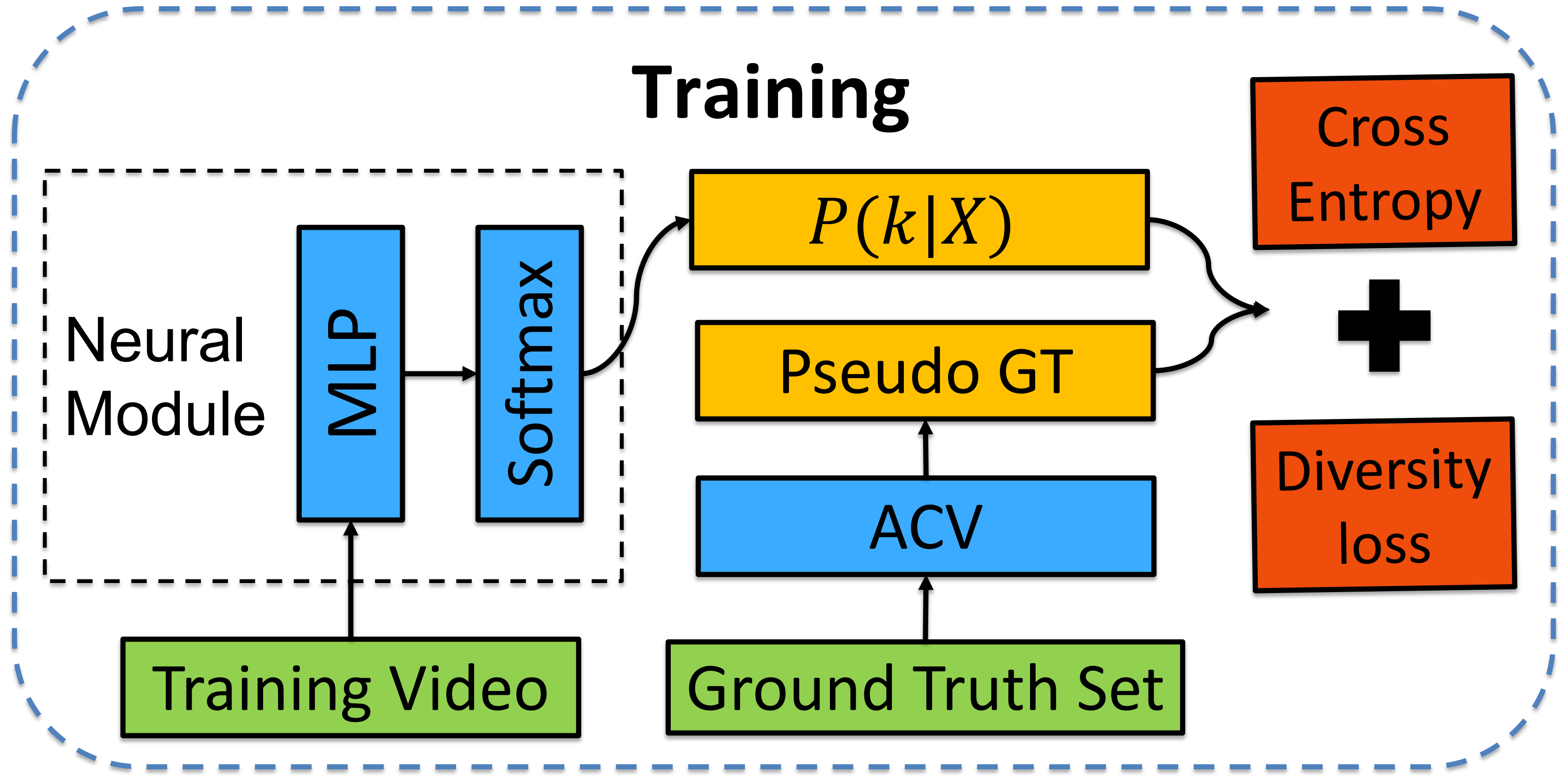
CVPR VIRTUAL JUNE 19-25

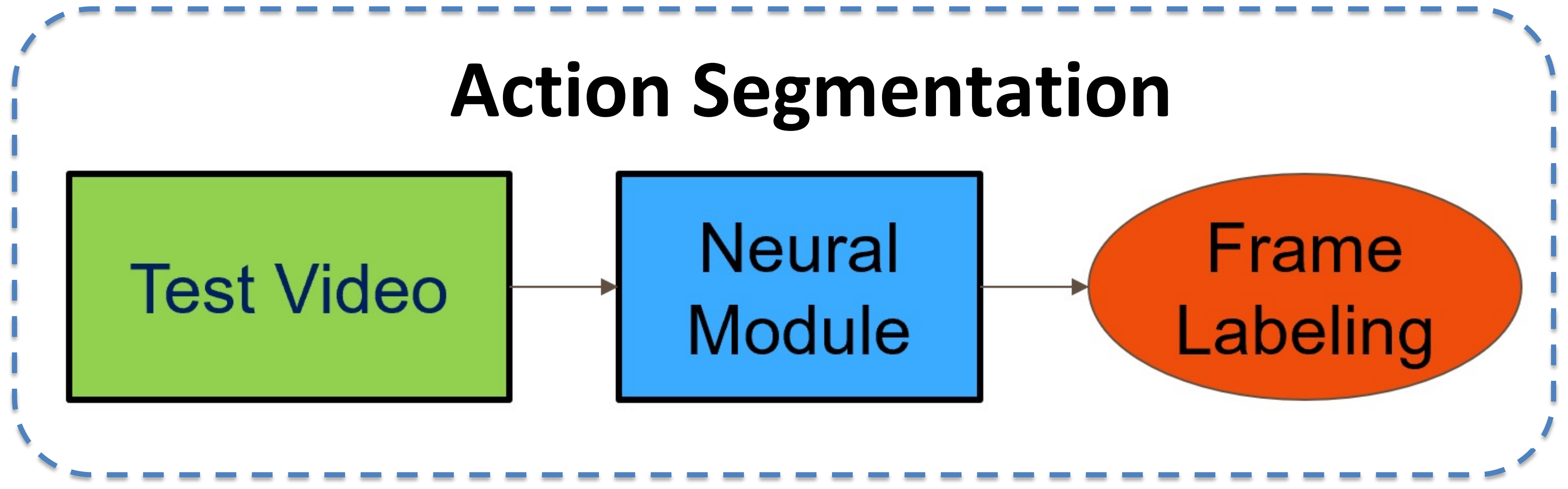
## Problem:

Predict frame labels, when the ground truth in training is limited and specifies only a set of actions present, without their temporal ordering and temporal extents.

## Key Ideas for Set-Supervised Training:

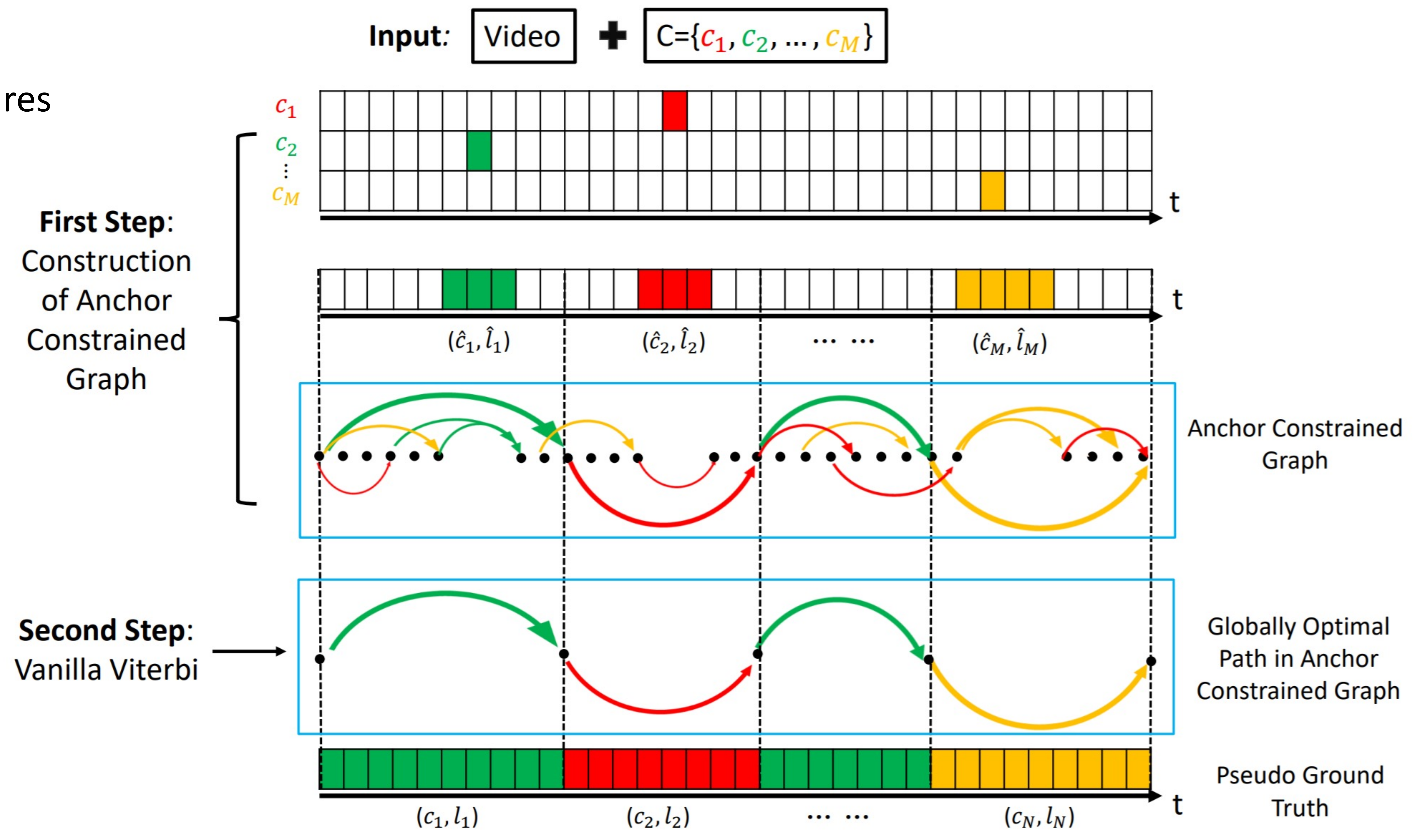
- Use ACV to generate framewise pseudo ground truth
- Train a frame labeler based on the pseudo ground truth
- Regularize learning with the Diversity loss of frame features





## Anchor-Constrained Viterbi

- Goal: Find an optimal segmentation of the training video based on a given ground-truth set of actions.
- This is an NP-hard problem.
- Our solution: A globally optimal path on the anchor-constrained graph.


Input: Video + C={$c_1, c_2, \ldots, c_M$}
$c_1$
$c_2$
$c_M$
t
First Step: Construction of Anchor Constrained Graph
$(\hat{c}_1, \hat{l}_1)$
$(\hat{c}_2, \hat{l}_2)$
… …
$(\hat{c}_M, \hat{l}_M)$
Anchor Constrained Graph
Second Step: Vanilla Viterbi
Globally Optimal Path in Anchor Constrained Graph
Pseudo Ground Truth
$(c_1, l_1)$
$(c_2, l_2)$
$(c_N, l_N)$

## Regularization with the Diversity Loss

- For every class, compute its saliency scores for all temporal frames.
- For every pair of action saliency scores, minimize their cosine distance to diversify their temporal saliency.

## Results

| Model | Breakfast<br>(*Mof*) | Cooking2<br>(*midpoint*) | Holl.Ext<br>(*IoD*) |
|---|---|---|---|
| (Set-supervised) | | | |
| Action Set [27] | 23.3 | 10.6 | 9.3 |
| SCT [7]our features | 26.6 | 14.3 | 17.7 |
| SCV [20] | 30.2 | 14.5 | 17.7 |
| Our ACV | **33.4** | **15.5** | **20.9** |
| (Transcript-supervised) | | | |
| OCDC [2] | 8.9 | - | - |
| HTK [14] | 25.9 | 20.0 | 8.6 |
| CTC [8] | 21.8 | - | - |
| ECTC [8] | 27.7 | - | - |
| HMM+RNN [26] | 33.3 | - | 11.9 |
| TCFPN [5] | 38.4 | - | 18.3 |
| NN-Viterbi [28] | 43.0 | - | - |
| D3TW [3] | 45.7 | - | - |
| CDFL [19] | 50.2 | - | 25.8 |

## Acknowledgement.

DARPA XAI N66001-17-2-4029, DARPA MCS N66001-19-2-4035