D^3TW: Discriminative Differentiable Dynamic Time Warping for Weakly Supervised Action Alignment and Segmentation

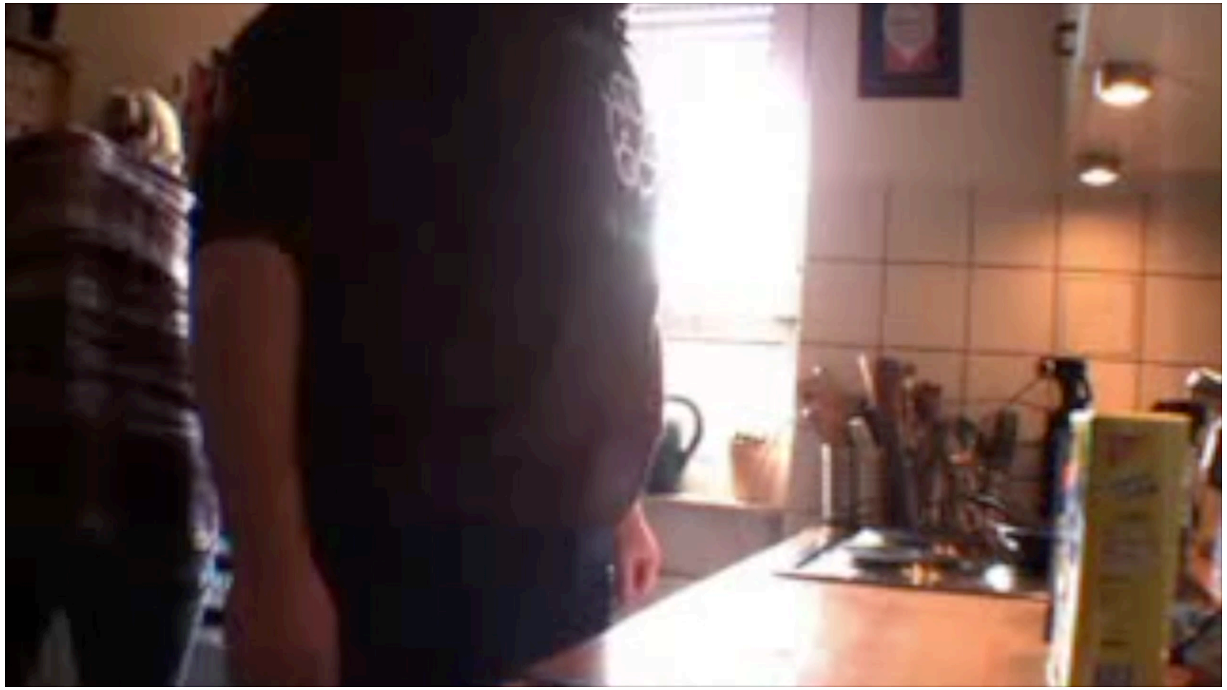
Chien-Yi Chang, De-An Huang, Yanan Sui, Li Fei-Fei, Juan Carlos Niebles

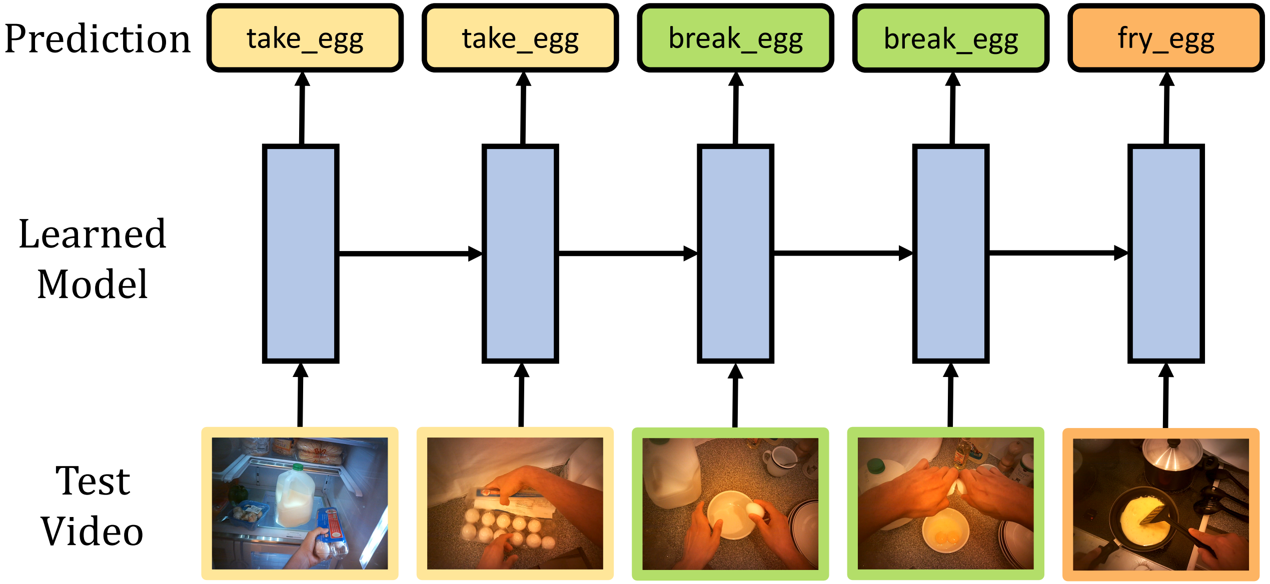
STANFORD VISION & LEARNING

Stanford ARTIFICIAL INTELLIGENCE

TOYOTA RESEARCH INSTITUTE

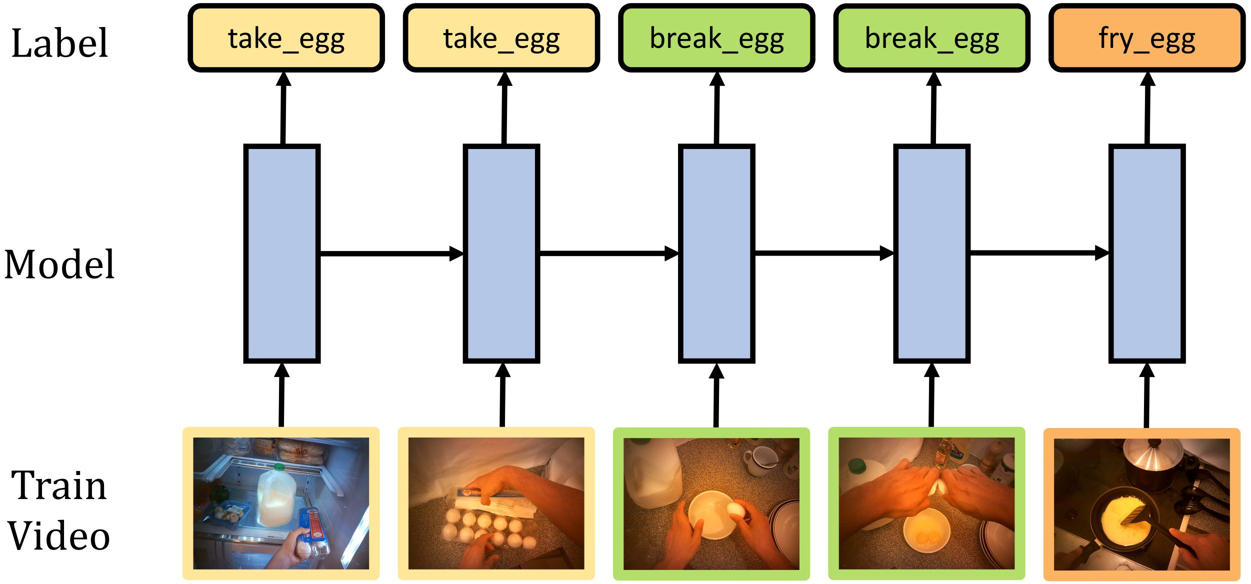
Temporal Action Segmentation





[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

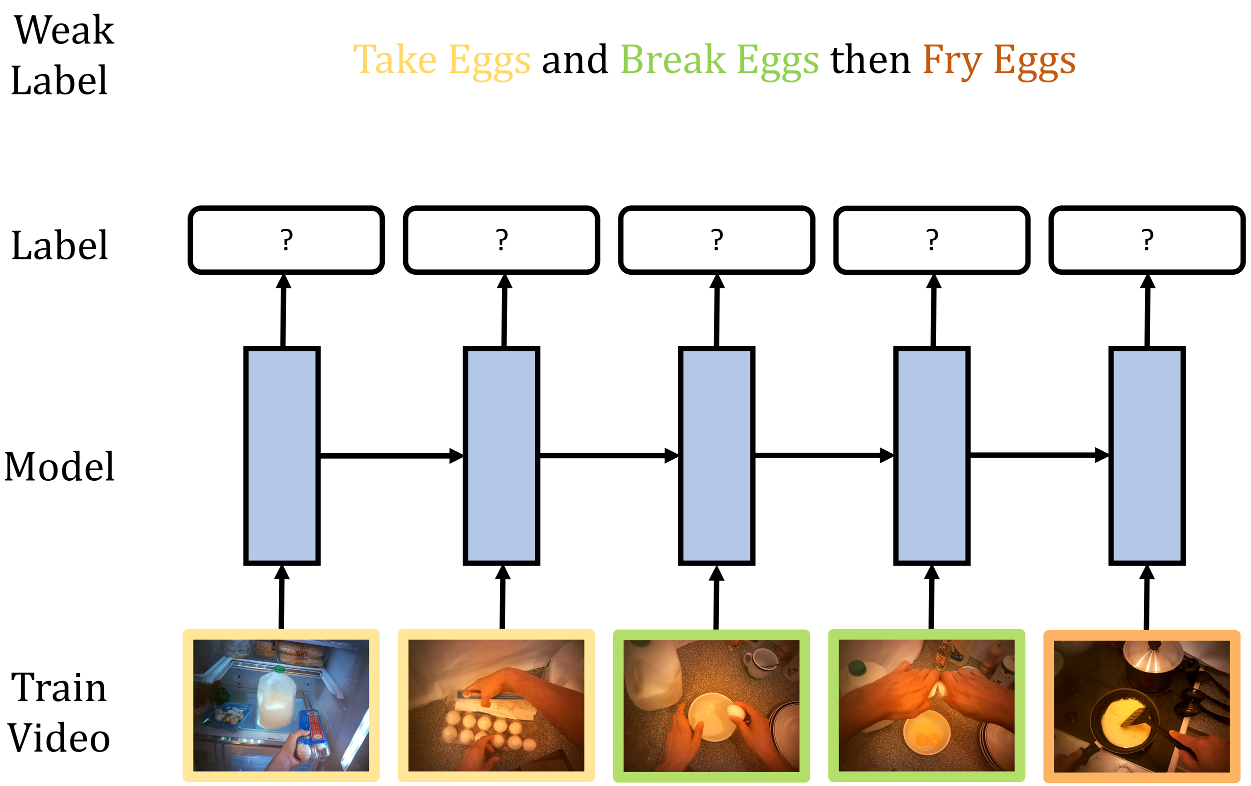
Fully Supervised Learning



- Requires many training videos with per frame action labels
- Expensive to annotate!

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Weakly-Supervised Learning



- Only use action ordering
- Easy to obtain from closed captions

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contributions

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

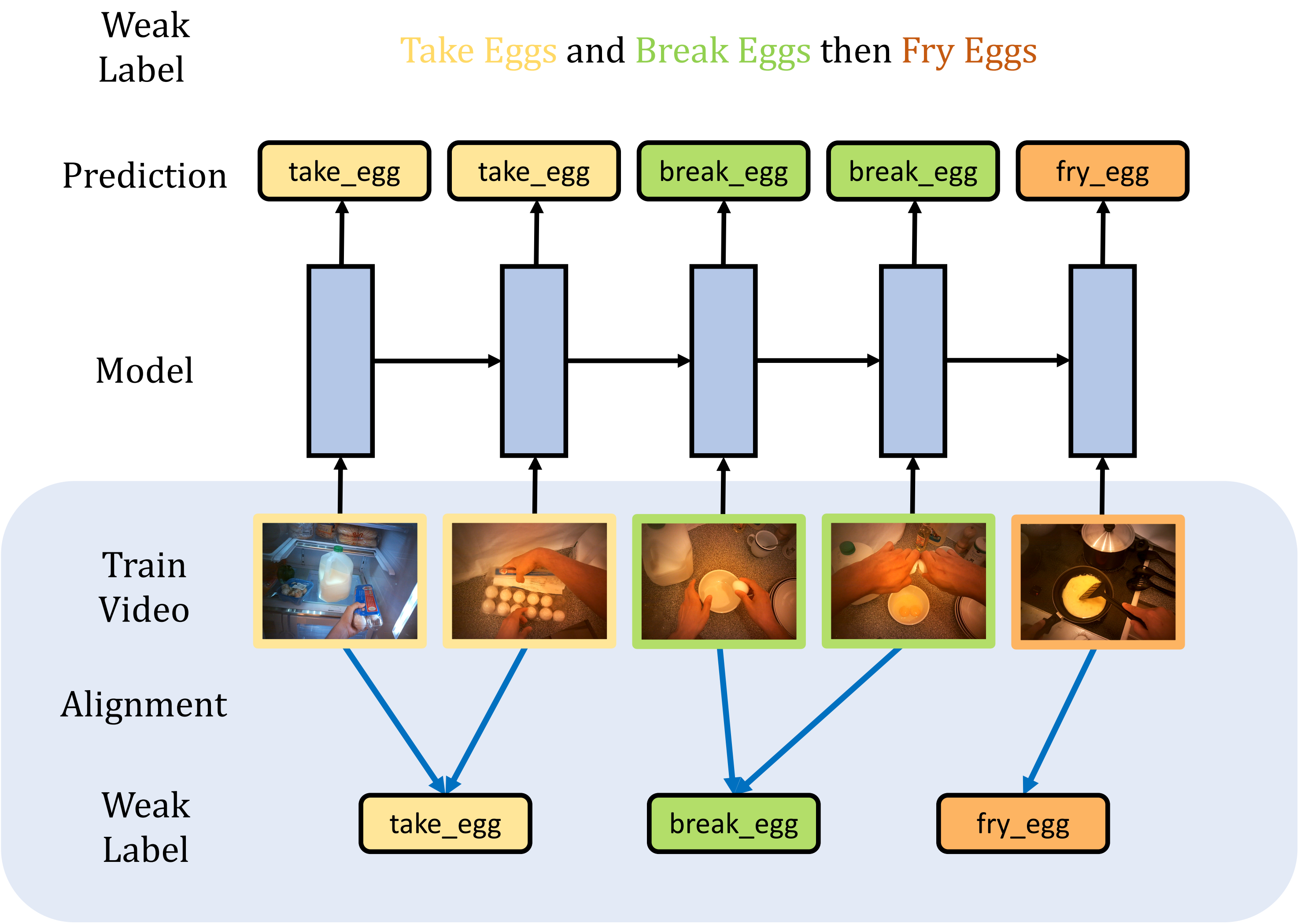
#2 Apply continuous relaxation to make our model end-to-end differentiable

#3 Propose the first discriminative model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

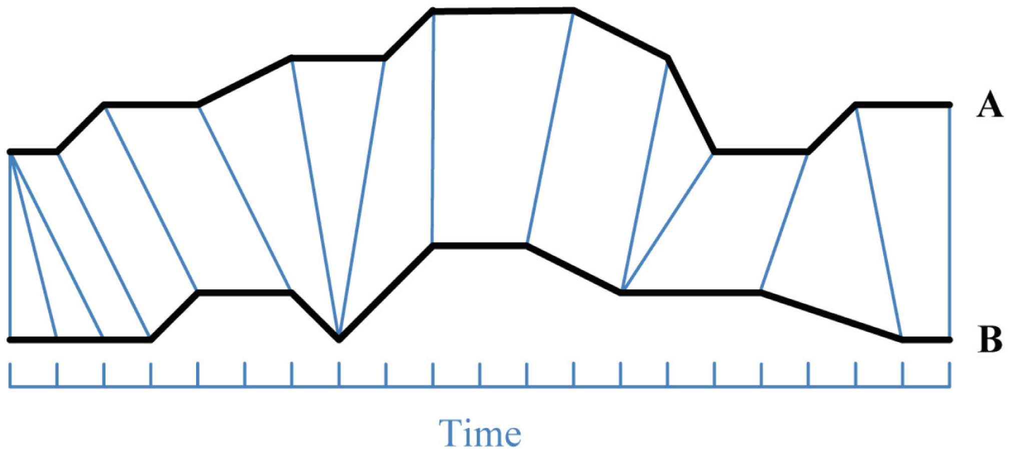
Key Contribution #1

- Train temporal action segmentation model as alignment

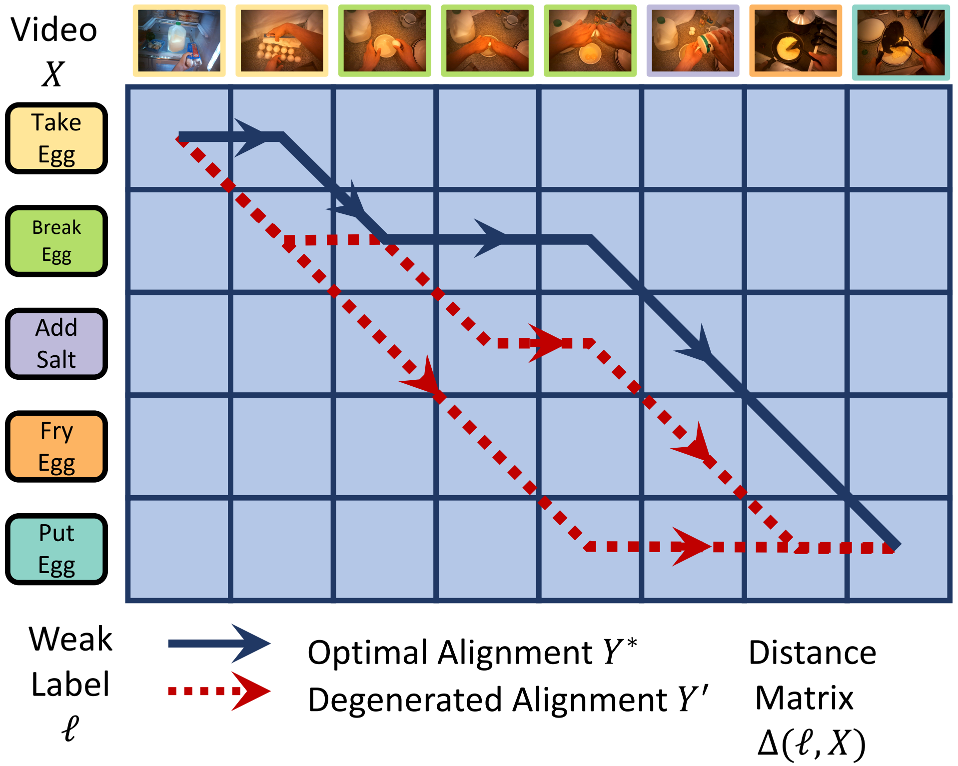


Key Contribution #1

- Solve the alignment problem with modified **Dynamic Time Warping** (DTW)



Classical DTW



Our DTW

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contributions

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end differentiable

#3 Propose the first discriminative model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contributions

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first discriminative model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contribution #2

- Continuous relaxation: $$min_{\gamma}\{a_1, \dots, a_n\} = -\gamma log \sum_{i=1}^{n} e^{-\frac{a_i}{\gamma}}, \gamma > 0$$

Data

3	2	7	4

Regular Min Operator

$min(\cdot) = 2$

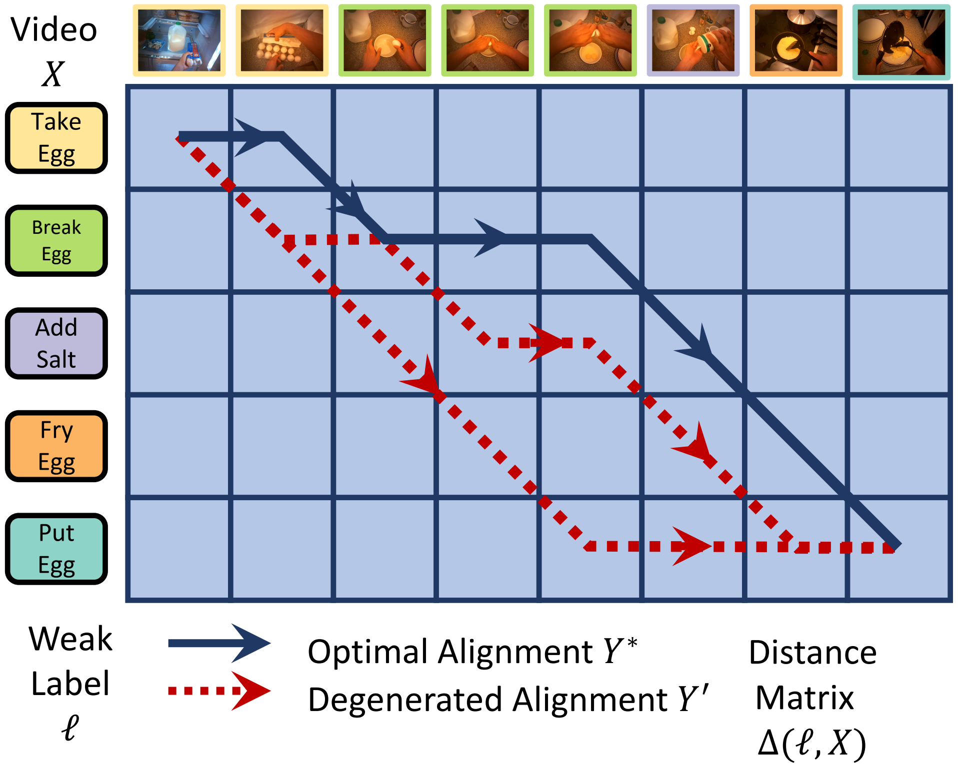
Gradient

0	1	0	0

Continuous Relaxation

$min_{\gamma}(\cdot) = 1.99$

.01	.98	0	.01



[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contributions

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first discriminative model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contributions

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

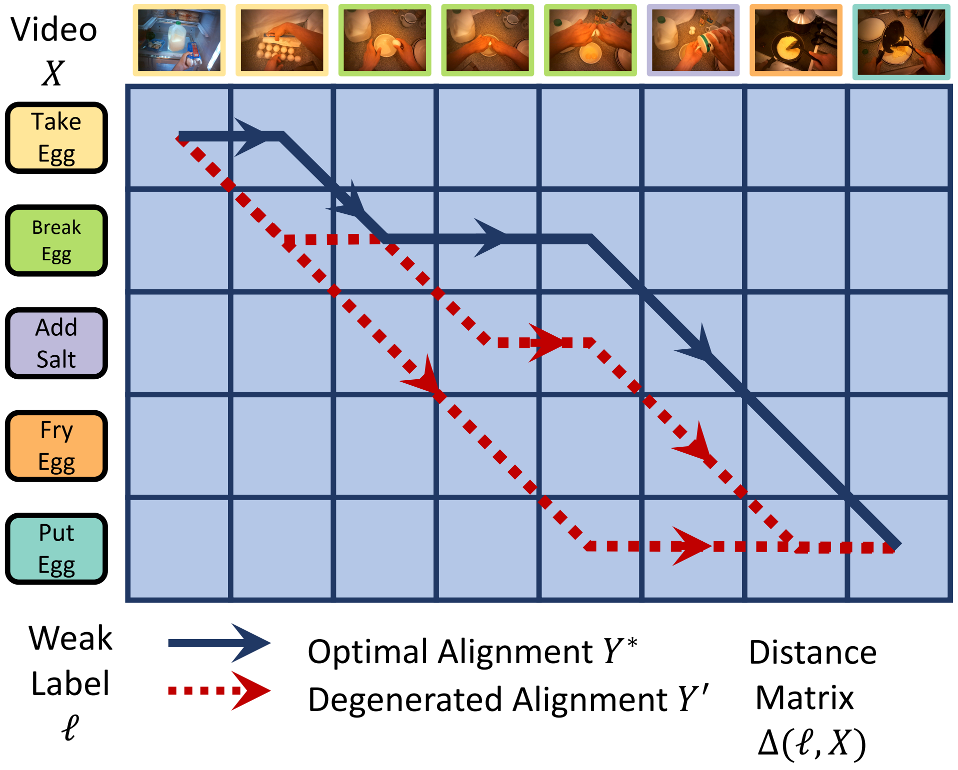
#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first **discriminative** model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contribution #3

- Design a loss function with only weak supervision



- Full Supervision:
  - Known ground truth alignment $\hat{Y}$
  - Straightforward loss function

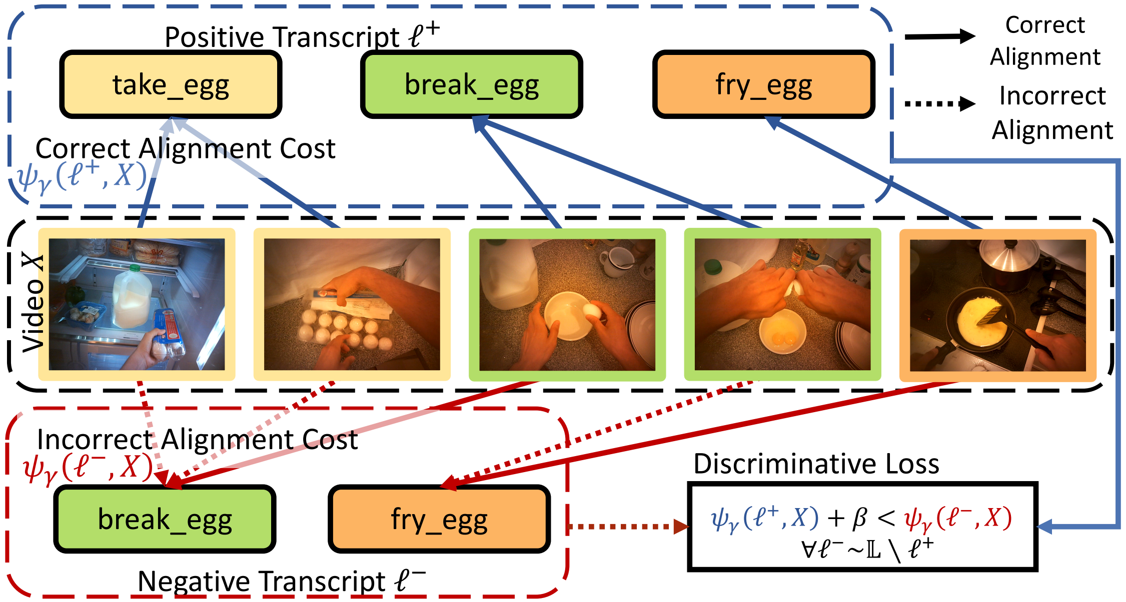
$$CE(Y^*, \hat{Y})$$

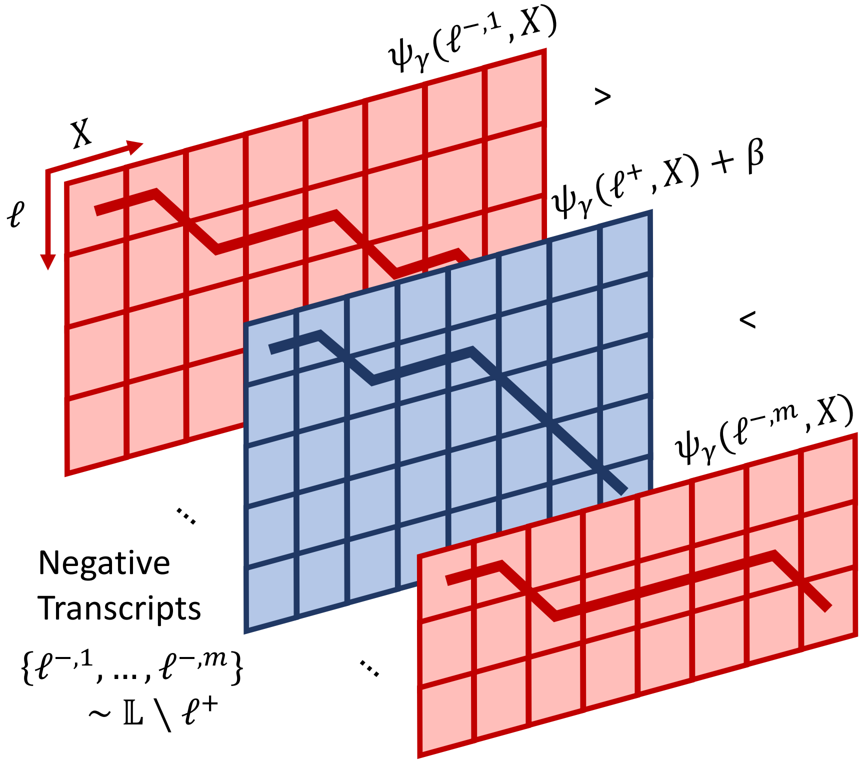
- Weak Supervision:
  - $\hat{Y}$ is unknown
  - Previous work resorts to generating pseudo $\hat{Y}$

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Key Contribution #3

- **Discriminative** loss: $\psi_\gamma(\ell^+, X) + \beta < \psi_\gamma(\ell^-, X), \quad \forall \ell^- \sim \mathbb{L} \setminus \ell^+$





[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first **discriminative** model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first **discriminative** model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first **discriminative** model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

#3 Propose the first **discriminative** model for weak ordering supervision

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

#1 Pose temporal action segmentation as **dynamic** alignment between two sequences

#2 Apply continuous relaxation to make our model end-to-end **differentiable**

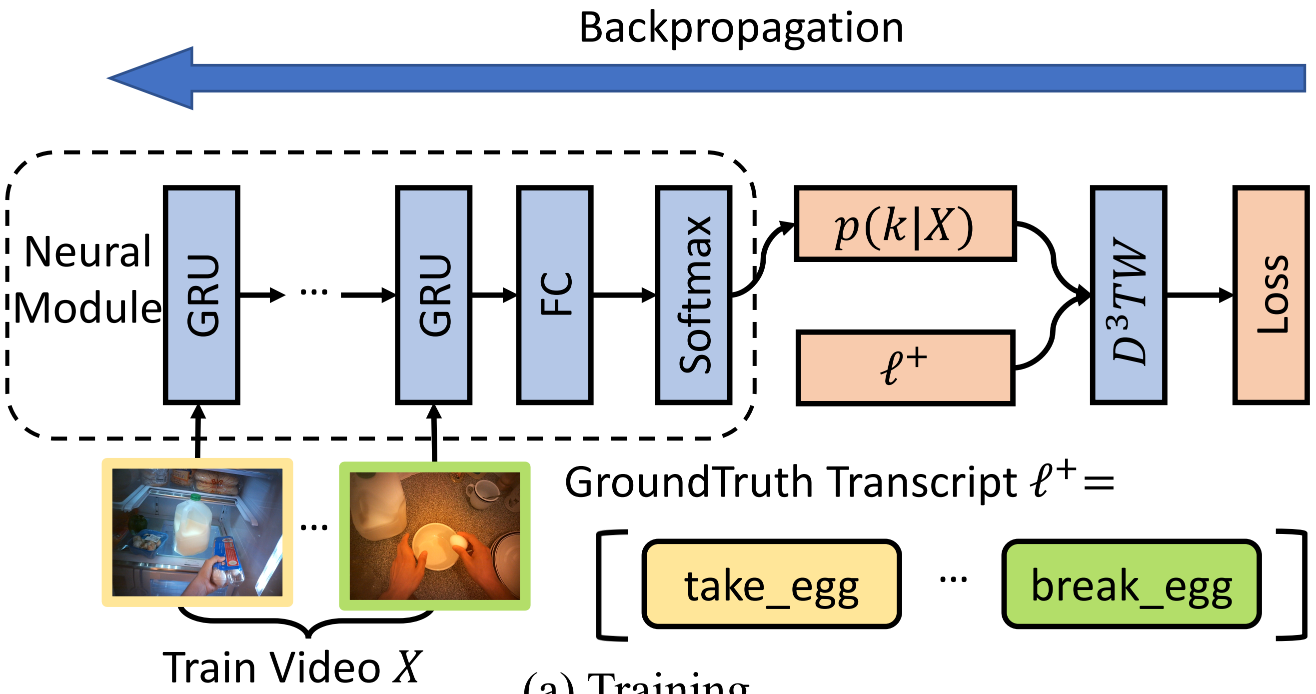
#3 Propose the first **discriminative** model for weak ordering supervision

D^3TW: **Discriminative Differentiable Dynamic** Time Warping for Weakly Supervised Action Alignment and Segmentation

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

D^3TW : Summary

- D^3TW, a differentiable layer that
  - captures regularities in the input sequences
  - imposes prior structure on the output as alignment



(a) Training

[Chang et al. D3TW:Discriminative Differentiable Dynamic Time Warping for Weakly-Supervised Action Alignment and Segmentation. CVPR 2019]

Model Evaluation – Temporal Action Segmentation

	Breakfast		Hollywood	
	Facc.	Uacc.	Facc.	Uacc.
ECTC[1]	27.7	35.6	-	-
GRU reest.[2]	33.3	-	-	-
TCFPN[3]	38.4	-	28.7	-
NN-Viterbi[4]	43.0	-	26.2	25.5

- Breakfast Actions
  - 3,600,000 frames
  - 48 action classes
  - ~ 6.9 action instances per video

- Hollywood Extended
  - 800,000 frames
  - 16 classes
  - ~ 2.5 action instances per video

[1] Huang et al. ECCV 2016
[2] Richard et al. CVPR 2017
[3] Ding et al. CVPR 2018
[4] Richard et al. CVPR 2018

Model Evaluation – Temporal Action Segmentation

	Breakfast		Hollywood	
	Facc.	Uacc.	Facc.	Uacc.
ECTC[1]	27.7	35.6	-	-
GRU reest.[2]	33.3	-	-	-
TCFPN[3]	38.4	-	28.7	-
NN-Viterbi[4]	43.0	-	26.2	25.5
Ours Dynamic	34.9	36.1	25.9	24.3

[1] Huang et al. ECCV 2016
[2] Richard et al. CVPR 2017
[3] Ding et al. CVPR 2018
[4] Richard et al. CVPR 2018

Model Evaluation – Temporal Action Segmentation

	Breakfast		Hollywood	
	Facc.	Uacc.	Facc.	Uacc.
ECTC[1]	27.7	35.6	-	-
GRU reest.[2]	33.3	-	-	-
TCFPN[3]	38.4	-	28.7	-
NN-Viterbi[4]	43.0	-	26.2	25.5
Ours Dynamic	34.9	36.1	25.9	24.3
Ours Differentiable Dynamic	38.0	38.4	30.0	28.3

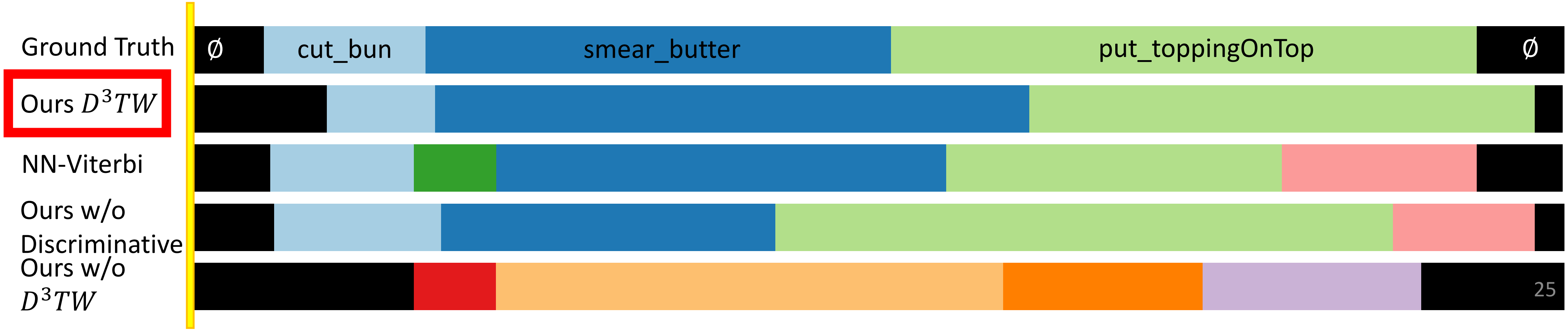
[1] Huang et al. ECCV 2016
[2] Richard et al. CVPR 2017
[3] Ding et al. CVPR 2018
[4] Richard et al. CVPR 2018

Model Evaluation – Temporal Action Segmentation

	Breakfast		Hollywood	
	Facc.	Uacc.	Facc.	Uacc.
ECTC[1]	27.7	35.6	-	-
GRU reest.[2]	33.3	-	-	-
TCFPN[3]	38.4	-	28.7	-
NN-Viterbi[4]	43.0	-	26.2	25.5
Ours Dynamic	34.9	36.1	25.9	24.3
Ours Differentiable Dynamic	38.0	38.4	30.0	28.3
Ours Discriminative Differentiable Dynamic	**45.7**	**47.4**	**33.6**	**30.5**

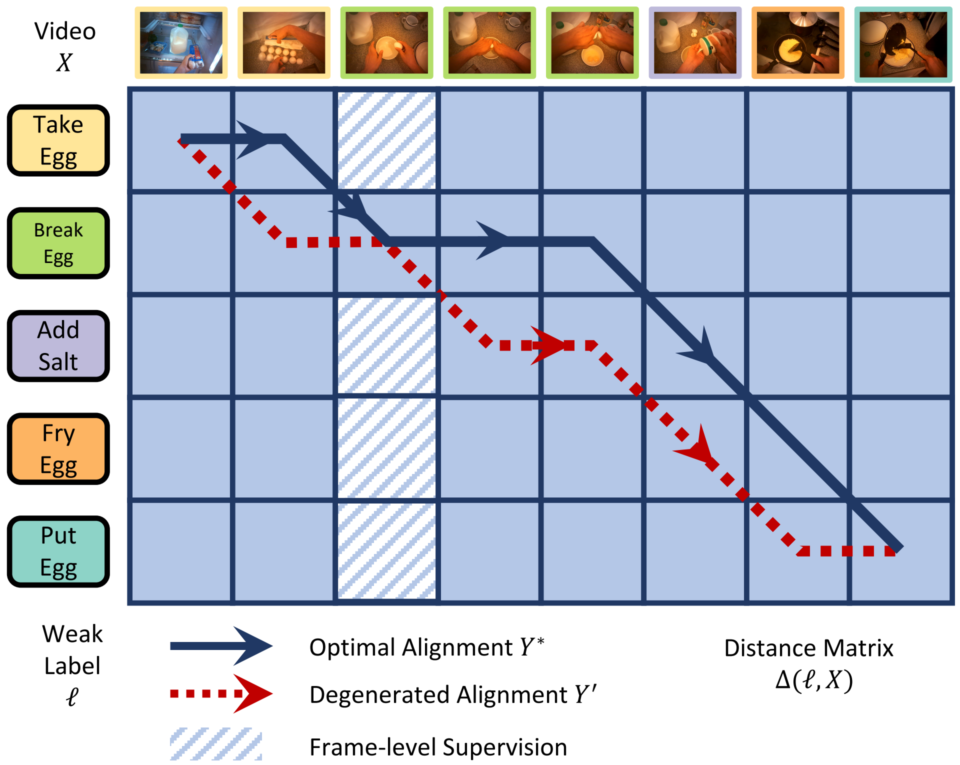
[1] Huang et al. ECCV 2016
[2] Richard et al. CVPR 2017
[3] Ding et al. CVPR 2018
[4] Richard et al. CVPR 2018

Qualitative Results of Temporal Action Segmentation

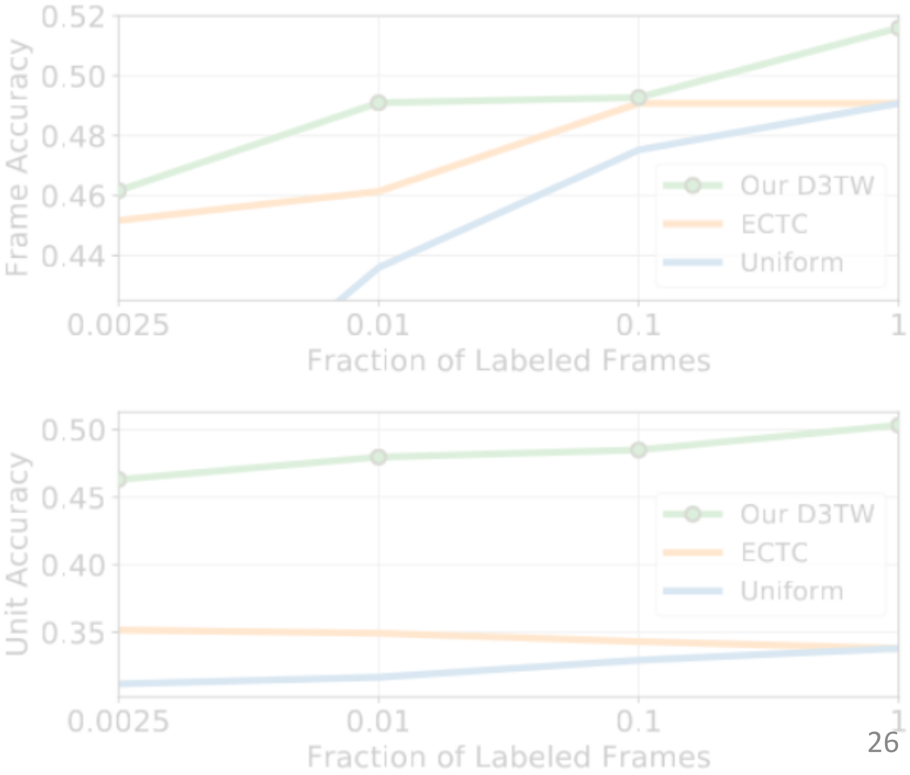


Semi-Supervised Learning with Our Framework

- Using semi-supervision by imposing path constraints



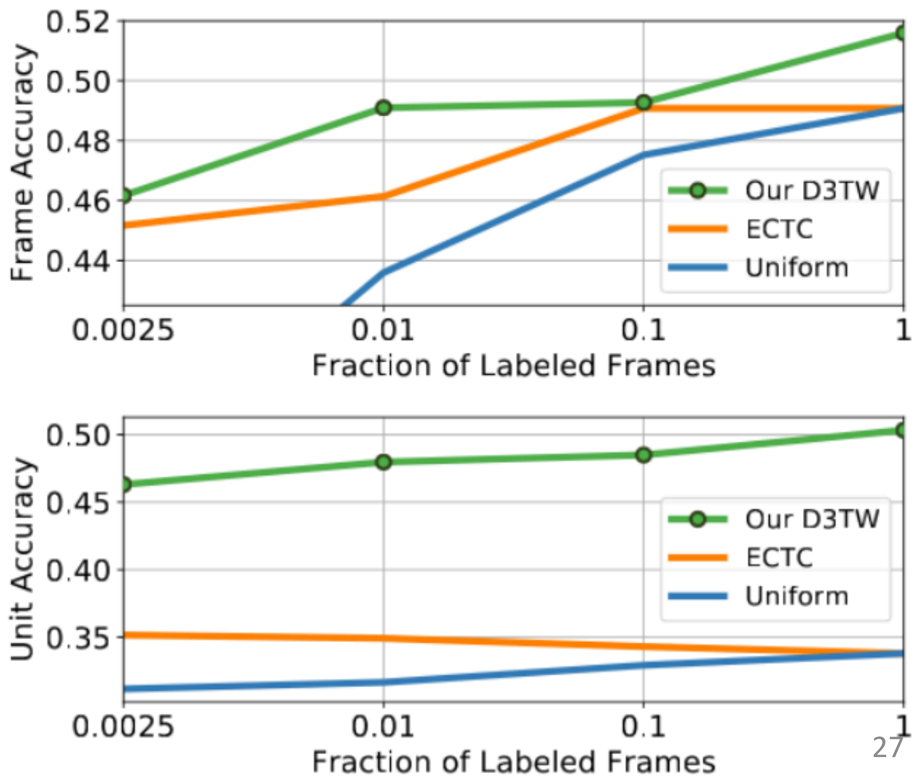
- Model performance compared with previous work



Semi-Supervised Learning with Our Framework

- Using semi-supervision by imposing path constraints

- Model performance compared with previous work



Thank You!