Domain Adaptation with Contextual Foreground Attention for Action Segmentation

Presenter: Ti-Haui Song

Advisor: Professor Li-Chen Fu

Date: 2021/12/22

國立臺灣大學
National Taiwan University

Outline

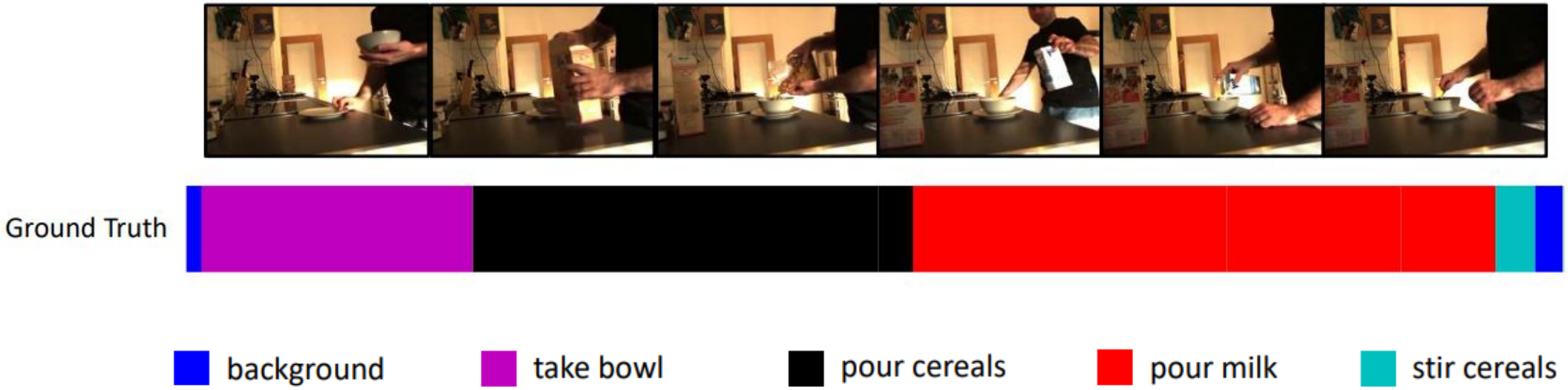
- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

Introduction – Background

- Action Segmentation
  - Segment videos by time, predicting an action class for each segment.



Introduction – Background

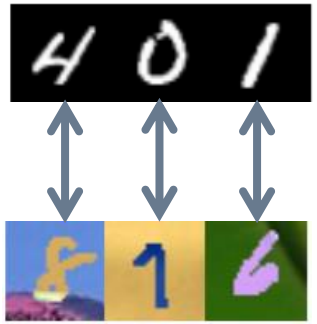
- Action Segmentation
  - Segment videos by time, predict an action class for each segment.

- Spatio-temporal variations
  - Different people may perform the same activity in various ways and environments.
  - Activities differ in space and time.
  - Leads to **domain discrepancy**.

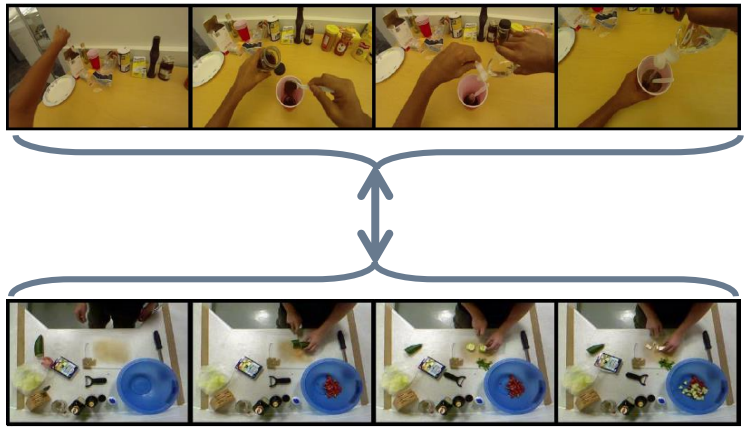
Introduction – Challenge

- Existing domain adaptation approaches are mainly focus on images.
  - Ignore the **temporal relations** of actions.

Image



Video



- Video contains a lot of **background frames** which are not supposed to do the domain adaptation.

Introduction – Contribution

- Propose the **Contextual Foreground Attention** for domain adaptation of action segmentation.
- Design an **attention mechnism** to capture the temporal relations of actions.
- Utilize **foreground labels** to filter out the irrelavent information.
- Outperm the existing frame-wise DA approaches and achieve **SOTA** in GTEA dataset.

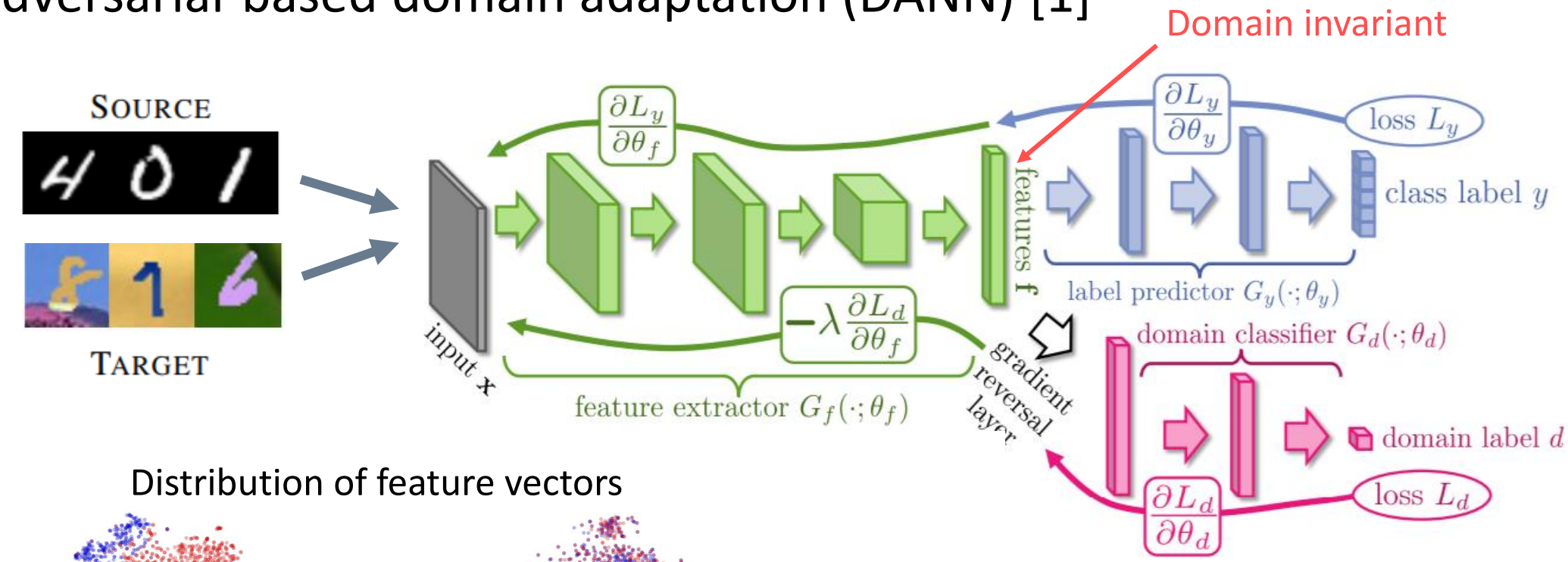
Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

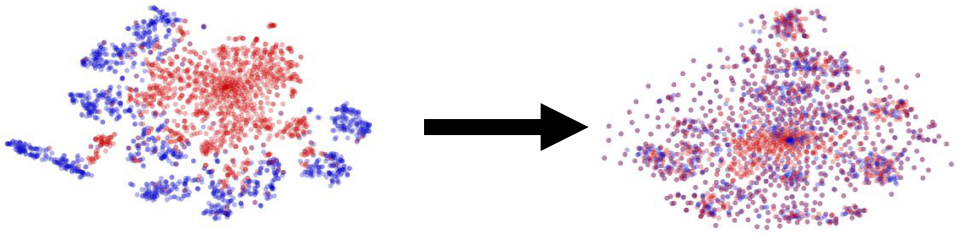
[1] Domain-adversarial training of neural networks. In: JMLR (2016)

Related work

- Adversarial-based domain adaptation (DANN) [1]



Distribution of feature vectors



[2] Human Action Recognition across Datasets by Foreground-weighted Histogram Decomposition. In: CVPR (2014)

Related work

- Foreground-weighted representation for action recognition [2]
  - Their experiments show that **background information is so discriminative** that the model **'learns the dataset'** rather than the action.

STIP Sampling	UCF Sports	UCF Youtube
Foreground only	71.92%	59.80%
Background only	73.97%	55.27%
Dense	75.34%	60.60%

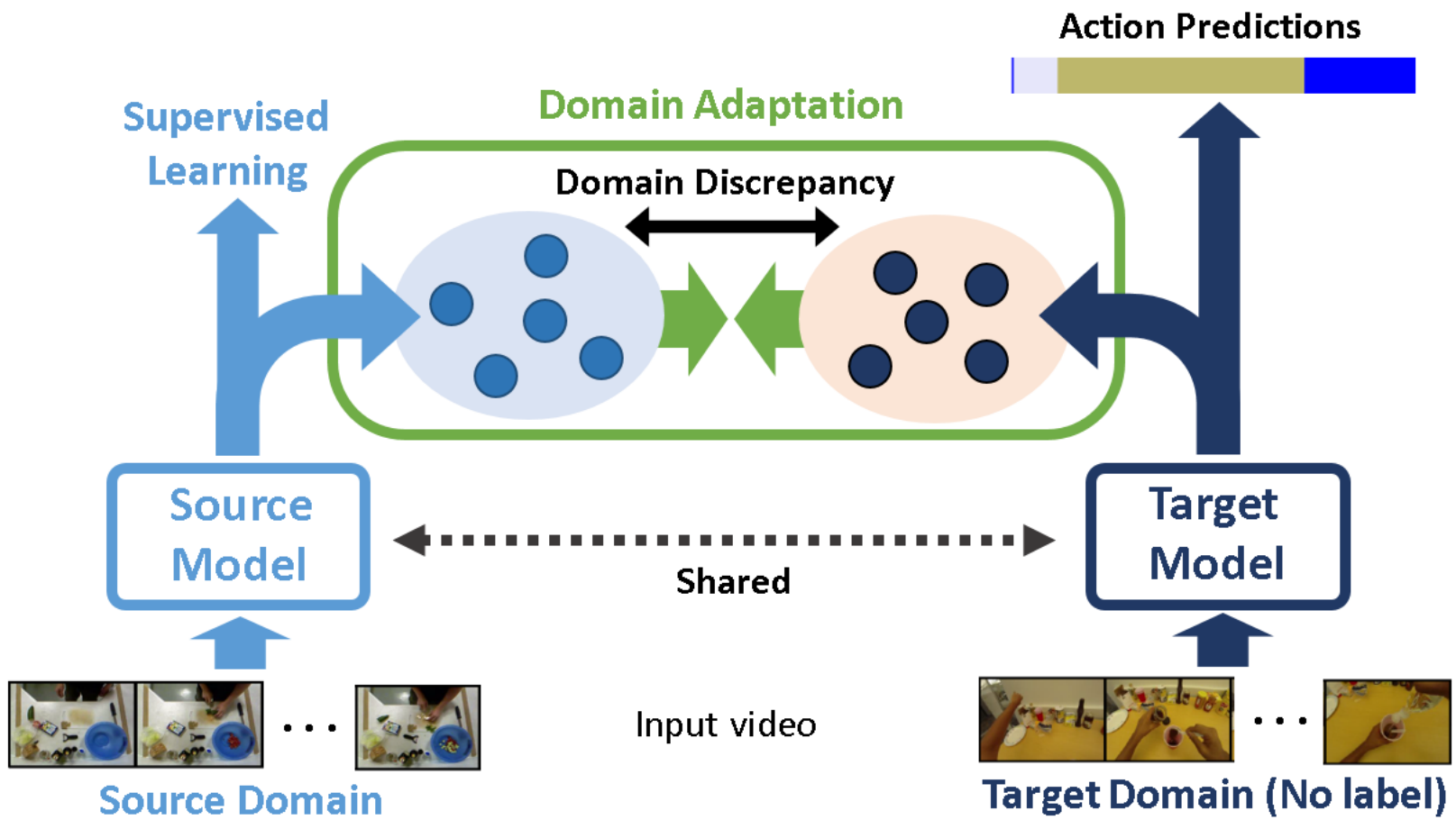
  - Attend on foreground part **spatially**.

Biking Basketball

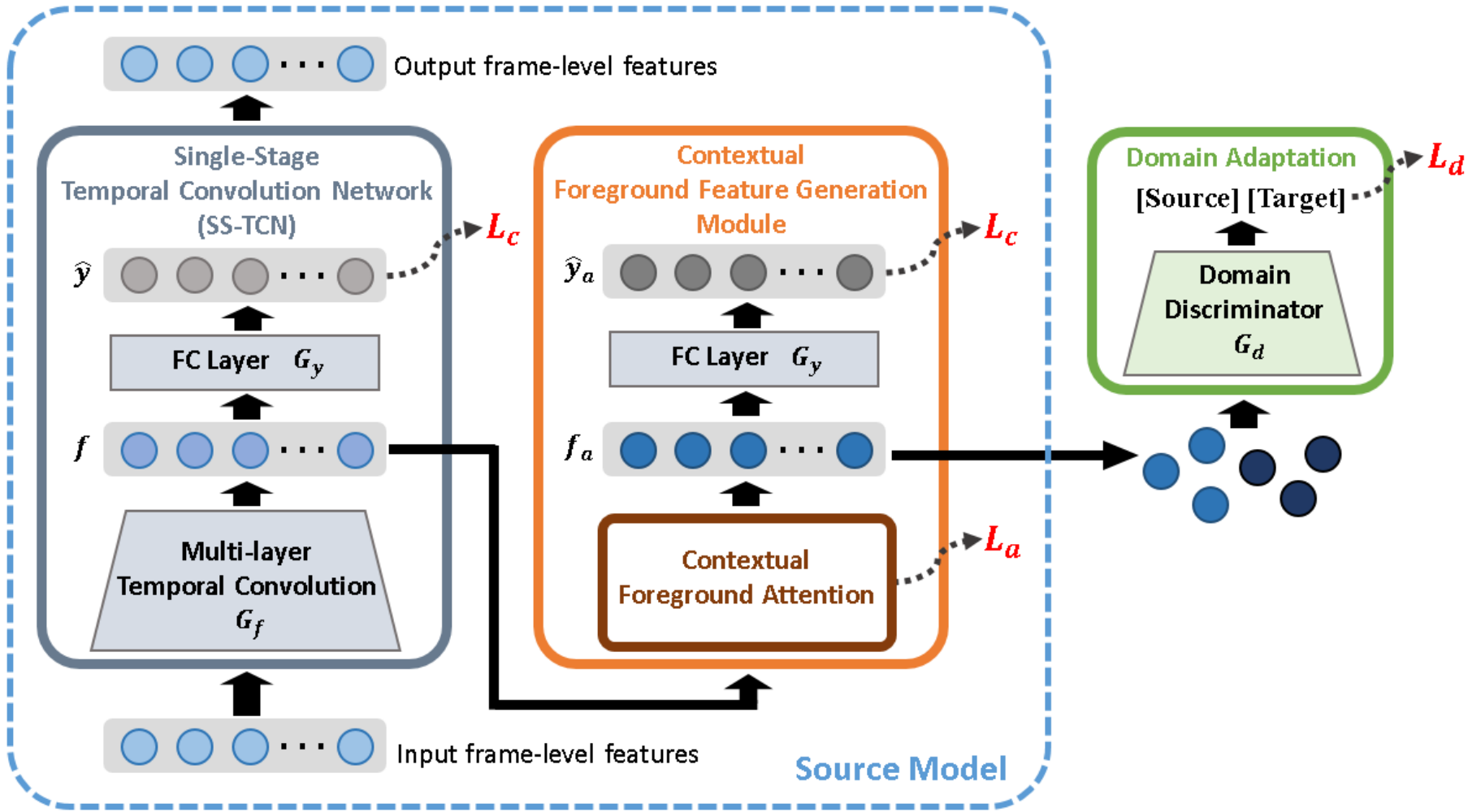
Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

System Overview

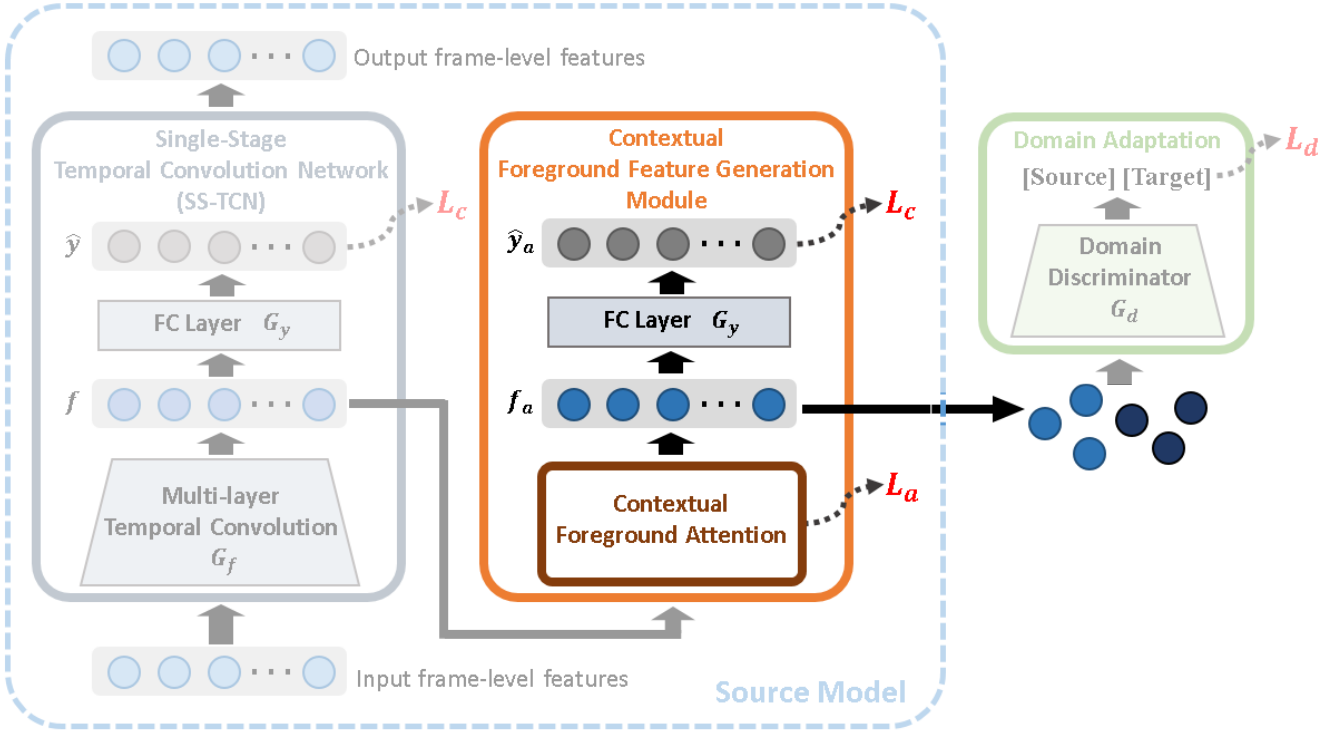


System Overview



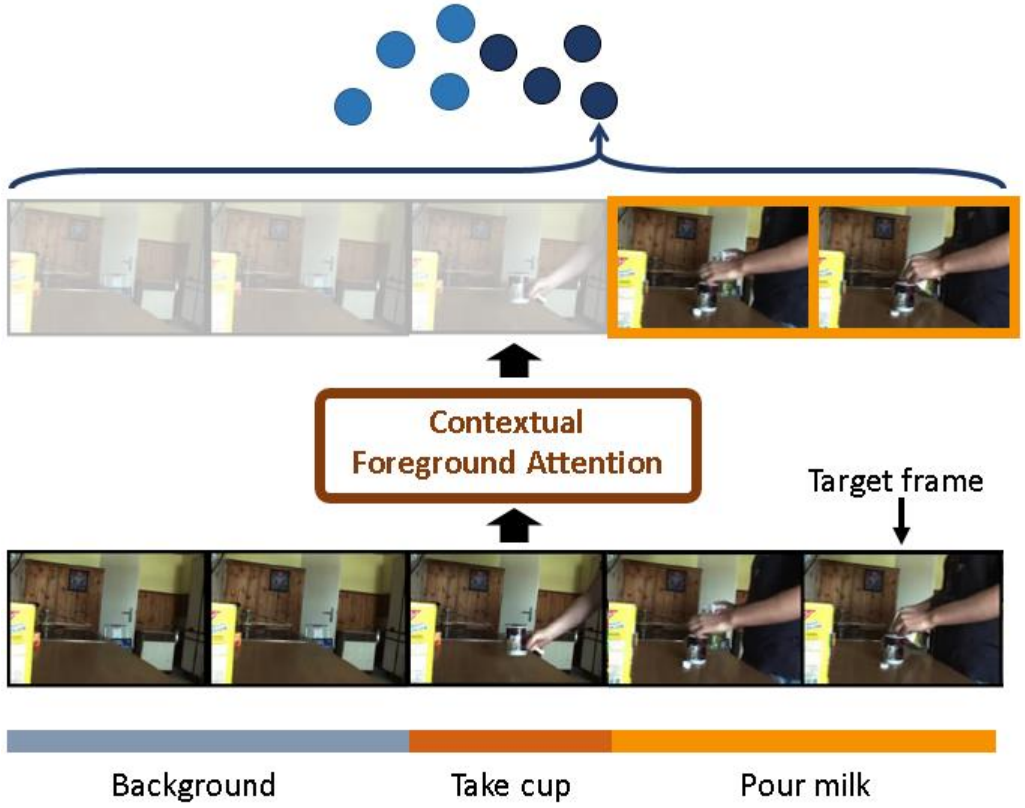
Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work



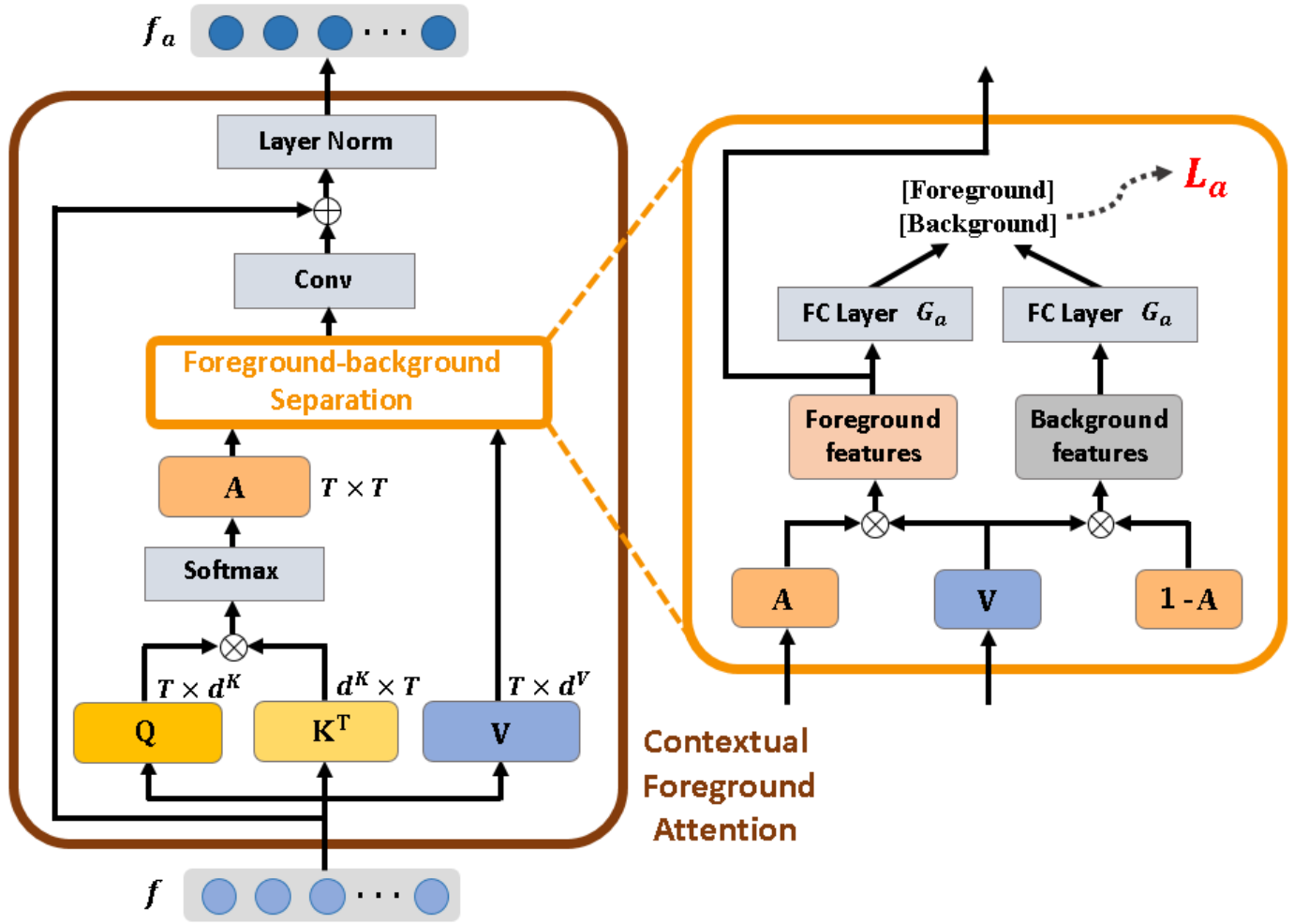
Methodology – Contextual Foreground Attention (1/2)

- Only the **foreground frames** should be aligned between different domains.
- Utilize **foreground labels** to learn the foreground attention.
- Not only attend on foreground, but also on the **related frames**, which captures the temporal context.



Methodology – Contextual Foreground Attention (2/2)

- **Foreground-background Separation module** can force the model to attend on foreground.
- f_a will also **need to be classified into correct actions** to guarantee the temporal context.

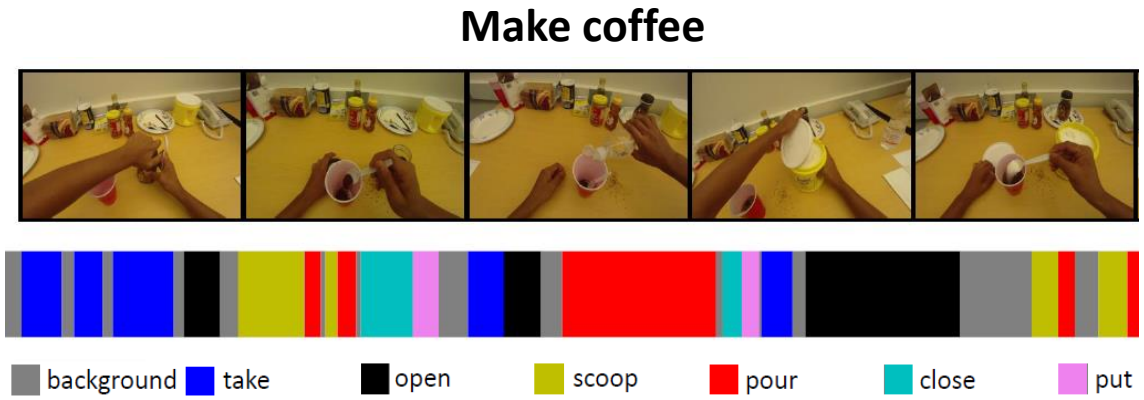


Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

Experiment – Datasets and Evaluation Metrics

- Datasets: GTEA
  - Separate the training and valid sets by different people.
  - Training set: **source domain**, validation set: **target domain**.



- Evaluation metrics
  - Frame-wise accuracy (Acc)
  - Segmental edit score
  - Segmental F1 score

} Emphasize on the **ordering of actions**

Experiment – Results

- Ablation study

FBS: Foreground-background Separation

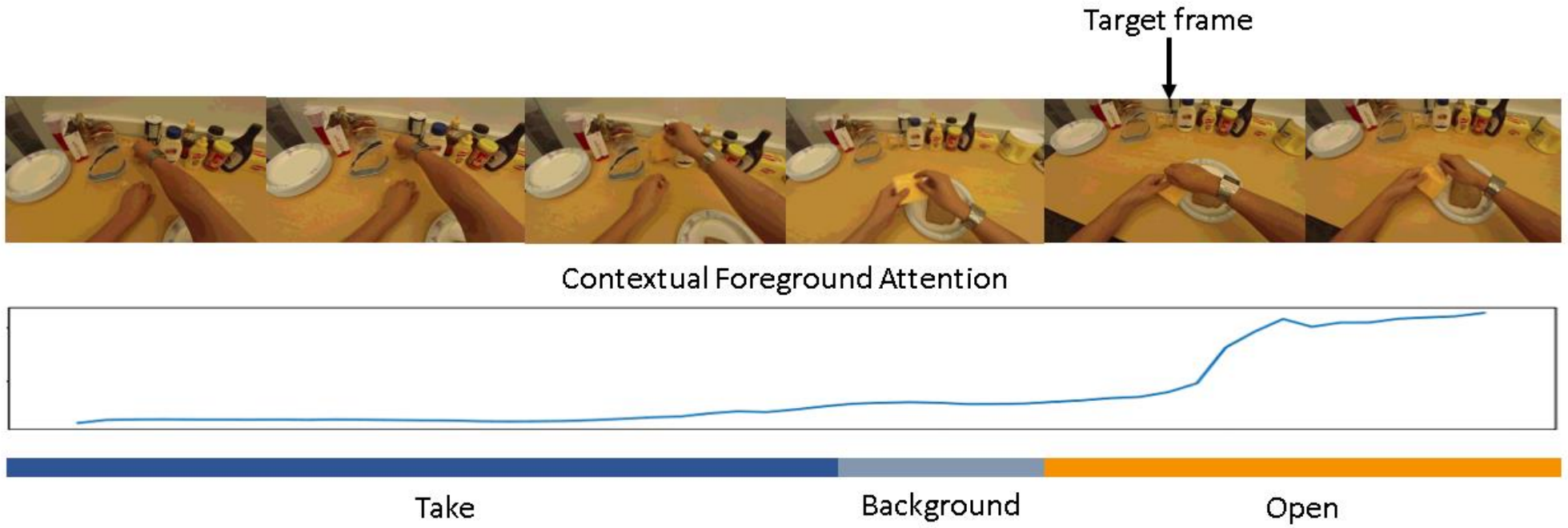
	F1@{10, 25, 70}			Edit	Acc
Source only (MS-TCN)	86.5	83.6	71.9	81.3	76.5
Frame-wise DA	89.6	87.9	74.4	84.5	**80.1**
Ours (w/o FBS)	88.6	87.1	74.9	83.5	79.6
Ours	**91.2**	**89.4**	**78.5**	**87.0**	80.0

- Comparison with other action segmentation methods

	F1@{10, 25, 70}			Edit	Acc
Source only (MS-TCN)	86.5	83.6	71.9	81.3	76.5
ASRF [2]	89.4	87.8	**79.8**	83.7	77.3
ASFormer [3]	90.1	88.8	79.2	84.6	79.7
SSTDA [4]	90.0	89.1	78.0	86.2	79.8
Ours	**91.2**	**89.4**	78.5	**87.0**	**80.0**

Experiment – Visulaization

- Vizualization of Contexual Foreground Attention



Outline

- Introduction
- Related work
- System Overview
- Methodology
- Experiment
- Conclusion & Future work

Introduction – Conclusion & Future work

- Propose the **Contextual Foreground Attention** for domain adaptation of action segmentation.
- The proposed **attention mechnism** is able to capture the temporal relations of actions and focus on foreground information.
- Outperm the existing frame-wise DA approaches and achieve **SOTA** in GTEA dataset.
- Attept to introduce additional modal (e.q, optical flow) to better attend on foreground information spatially and temporally.