Georgia Tech

OLIVES @GeorgiaTech

ghassanalregib.info/

Baidu USA

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Action Segmentation with Joint Self-Supervised Temporal Domain Adaptation





Min-Hung Chen[1*] Baopu Li[2] Yingze Bao[2] Ghassan AlRegib[1] Zsolt Kira[1]

[1]Georgia Institute of Technology [2]Baidu USA

June 17, 2020

[Paper] https://arxiv.org/abs/2003.02824

[Project] https://minhungchen.netlify.app/project/cdas

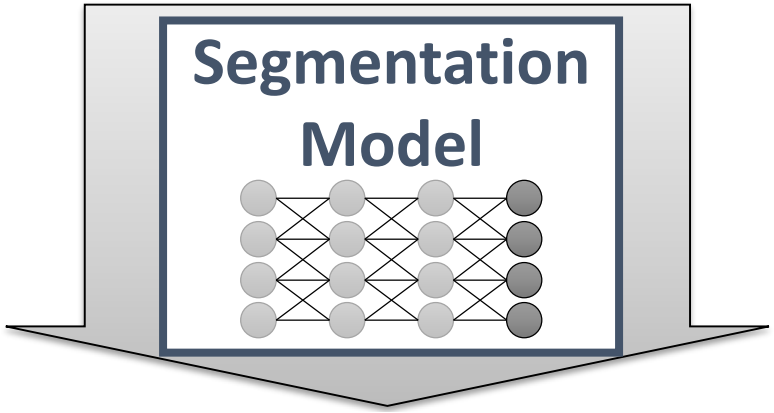
*Work done during an internship at Baidu USA

Action Segmentation

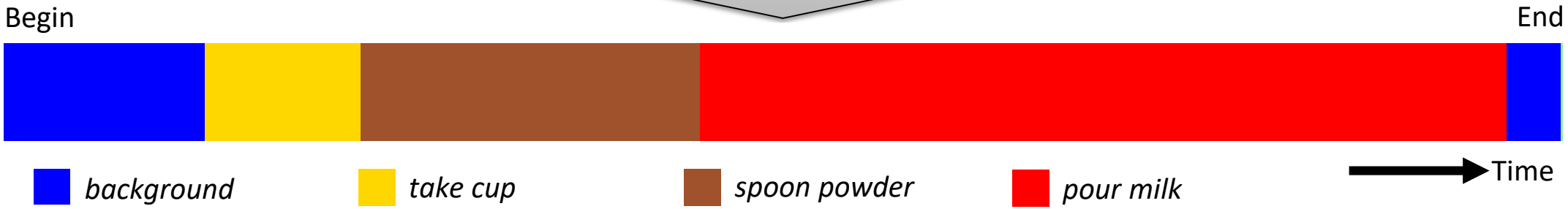
Action segmentation = Action Recognition + Temporal Segmentation

Input Video

Make milk



Output Predictions



Action Segmentation

Source Videos

Labels

Unlabeled Videos

Fully Supervised Learning

Source Model

Standard Action Segmentation

Focus on architecture design with fully-supervised learning

Challenge

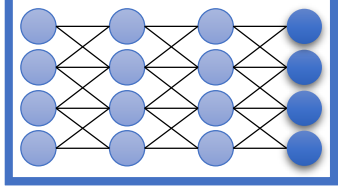
Source Videos

Labels

Unlabeled Videos

Fully Supervised Learning

Source Model



Standard Action Segmentation

Focus on architecture design with fully-supervised learning

Target Videos

Unlabeled Videos

Source model may not work on target videos

Source vs. Target

Different people perform the same action in different styles

Adapting Source Model

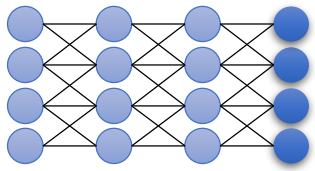
Source Videos

Labels

Unlabeled Videos

Fully Supervised Learning

Source Model



Goal

Adapt the source model without additional labels

Target Videos

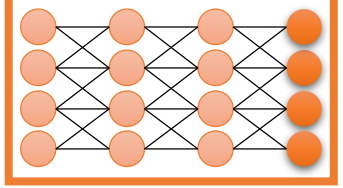
Unlabeled Videos

Labels

Annotating data is time consuming!

Model Adaptation

Target Model



Goal

Adapt the source model with unlabeled data

Source Videos

Labels

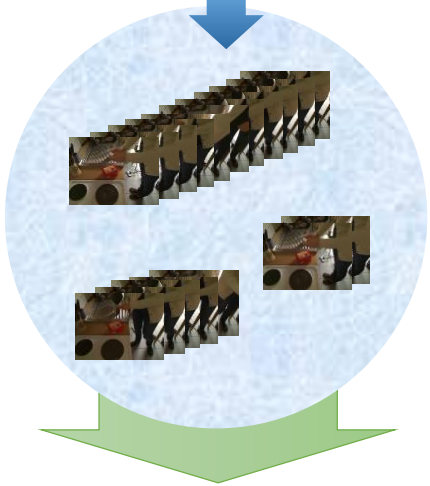
Unlabeled Videos

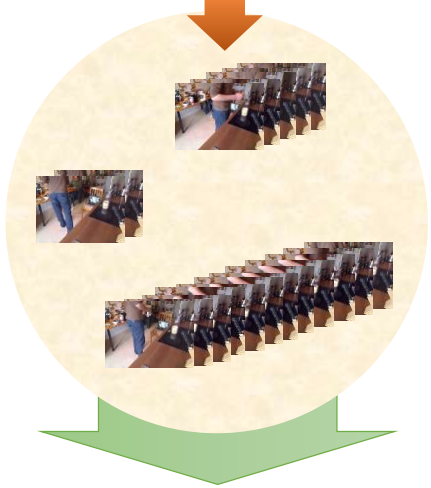
Target Videos

Unlabeled Videos

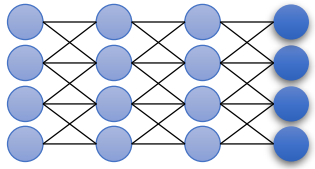
Labels

Fully Supervised Learning





Source Model



Previous Works

Adversarial-based Domain Adaptation

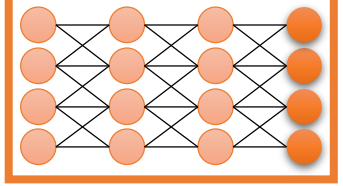
Not consider dependencies between adapted features

Self-Supervised Learning

Not consider cross-domain discrepancy

Model Adaptation

Target Model



Temporal Domain Permutation

Previous Works

- Predict temporal orders
- Predict binary domains

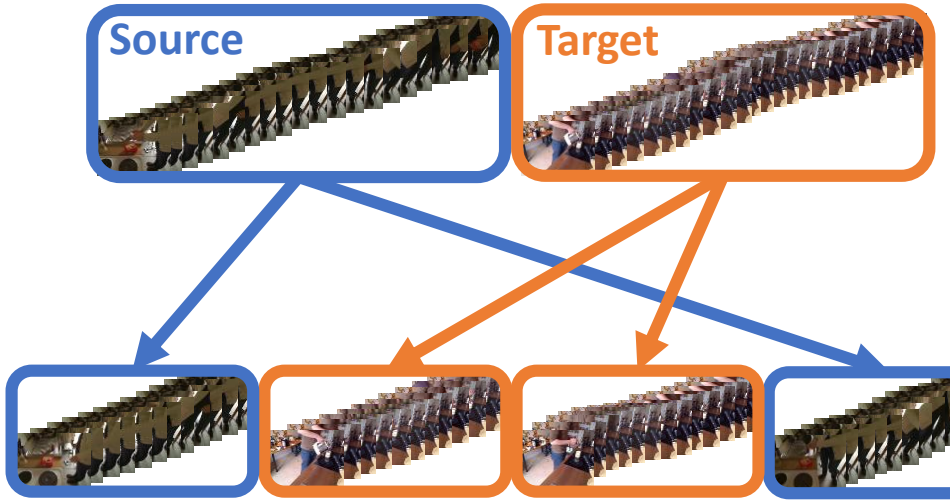
Intuition

DA for classification ➔ domain classification

DA for segmentation ➔ domain segmentation

Our Method

Predict temporal permutations of domains



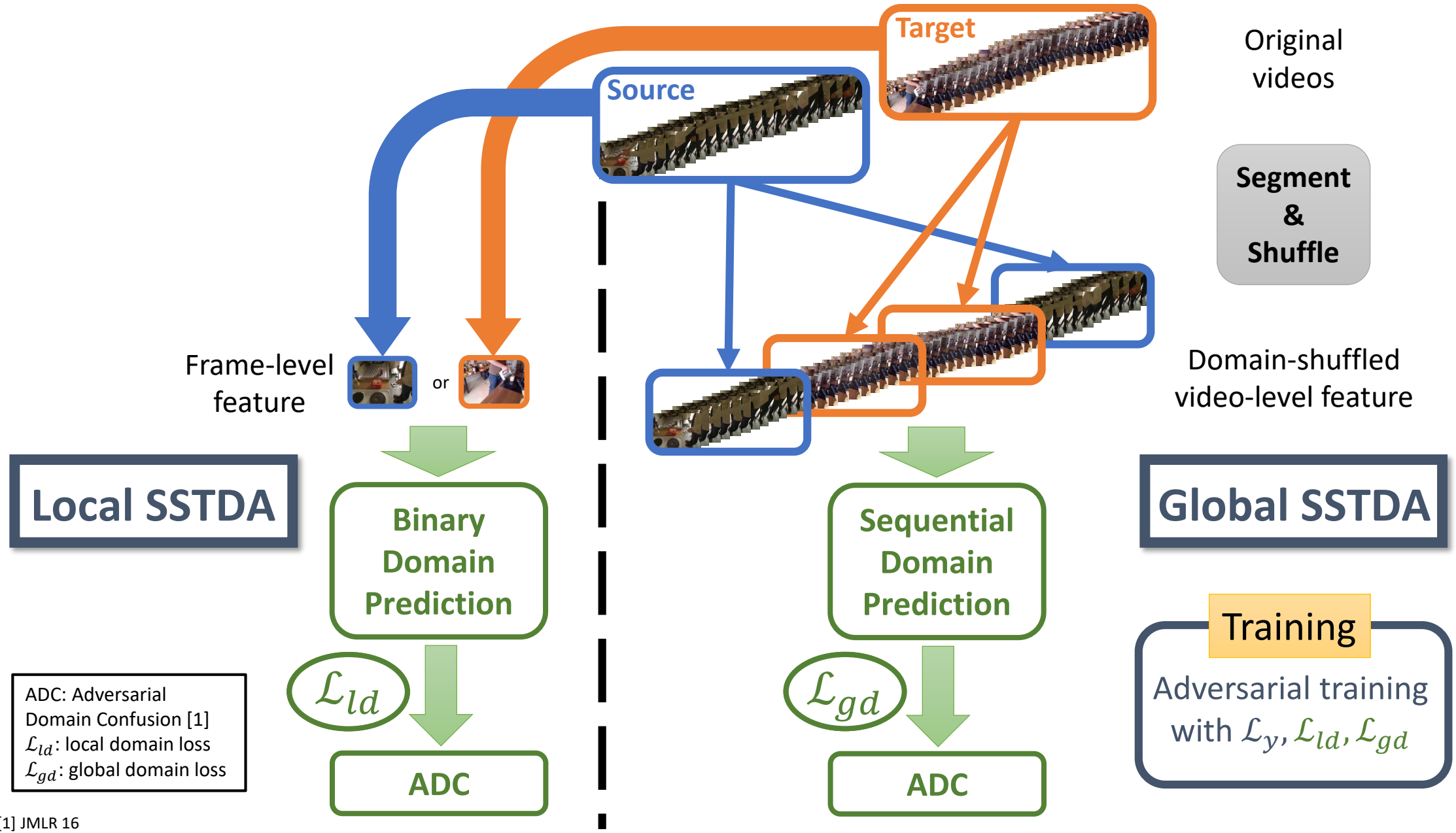
Segment & Shuffle

Domain Permutation Prediction

[0,0,1,1] **[0,1,1,0]** [1,1,0,0]
[0,1,0,1] [1,0,1,0] [1,0,0,1]

$\mathcal{L}_{gd}$

Self-Supervised Temporal Domain Adaptation (SSTDA)



[1] JMLR 16

Our Approach: SSTDA

Experimental Results

50Salads [1]

Source-only: results from directly running the official released code of MS-TCN [2]

50Salads	F1@10	F1@25	F1@50	Edit score
Source-only [2]	75.4	73.4	65.2	68.9
Local SSTDA	79.2	77.8	70.3	72.0
SSTDA	**83.0**	**81.5**	**73.8**	**75.8**
SSTDA (65%)	77.7	75.0	66.2	69.3

Effectively exploit unlabeled target videos for action segmentation

[1] UbiComp 13, [2] CVPR 19

Comparison: Unlabeled Target Videos

50Salads	F1@10	F1@25	F1@50	Edit score
Source-only	75.4	73.4	65.2	68.9
VCOP [1]	75.8	73.8	65.9	68.4
DANN [2]	79.2	77.8	70.3	72.0
JAN [3]	80.9	79.4	72.4	73.5
MADA [4]	79.6	77.4	70.0	72.4
MSTN [5]	79.3	77.6	71.5	72.1
MCD [6]	78.2	75.5	67.1	70.8
SWD [7]	78.2	76.2	67.4	71.6
SSTDA	**83.0**	**81.5**	**73.8**	**75.8**

Jointly adapt domains with multiple temporal scales can better address discrepancy problems for videos

[1] CVPR 19, [2] JMLR 16, [3] ICML 17, [4] AAAI 18, [5] ICML 18, [6] CVPR 2018, [7] CVPR 19

Visualization: 50Salads

Ground Truth

Expectation

Only highlight the difference from the ground truth

MS-TCN [1]

Local SSTDA

SSTDA

action start
peel cucumber
cut cucumber
place cucumber into bowl
cut tomato
place tomato into bowl
cut lettuce
place lettuce into bowl
cut cheese
place cheese into bowl
mix ingredients
add oil
add vinegar
mix dressing
add pepper
serve salad onto plate
add dressing
action end

[1] CVPR 19

Summary

- Goal: adapt action segmentation models using unlabeled videos
- Approach: **Self-Supervised Temporal Domain Adaptation (SSTDA)**
  - Perform domain adaptation for multiple temporal scales
  - Learn feature representations with domain-invariant temporal dynamics
- Outperform other self-supervised methods and image-based DA methods
- Improve action segmentation by large margins using unlabeled target videos

Paper

Project

Code

Poster: #93 @ Session 2.4
Date: June 17 (Wed.)
Q&A Time: 16 - 18 & 04 - 06

[Paper] https://arxiv.org/abs/2003.02824
[Project] https://minhungchen.netlify.app/project/cdas
[Code] https://github.com/cmhungsteve/SSTDA