Università degli Studi di Padova
Dipartimento di Ingegneria dell'Informazione

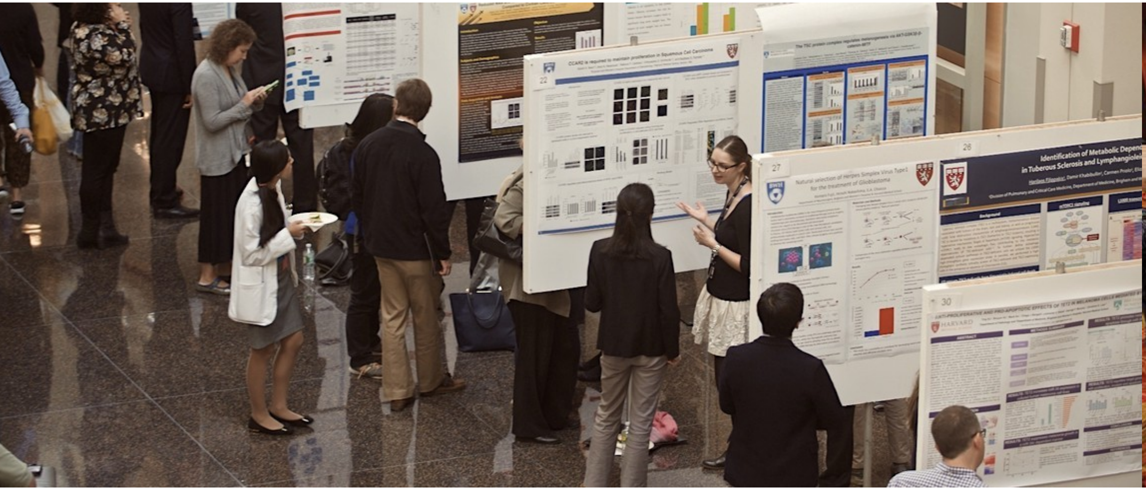
IAS-Lab
Intelligent Autonomous Systems Laboratory

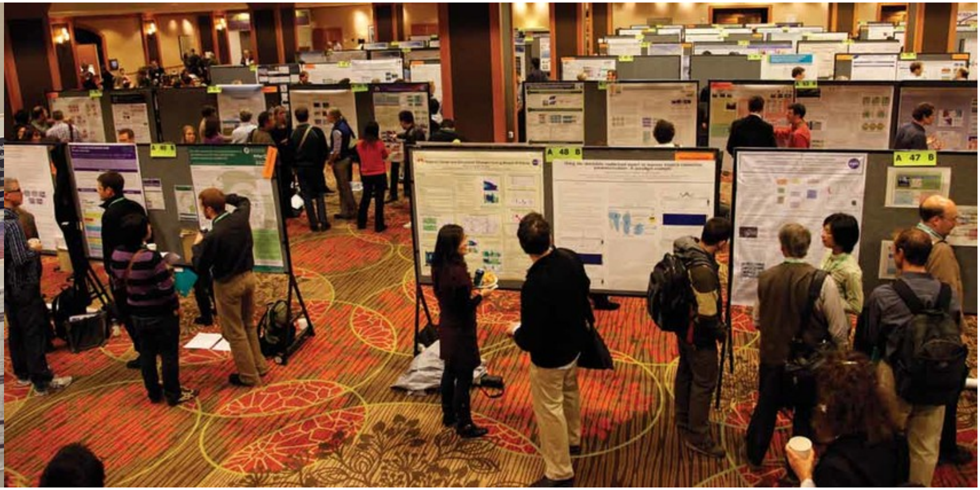
3D Data Processing

Poster Session

Alberto Pretto

Poster Session





Poster Session

Groups: 2 students

When: June 11 2024, h.12:30 – 14:30

Where: in the corridor in front of the exit door of Ve classroom (third floor of DEI/G)

➡ **Day and location to be confirmed!**

Papers

We will shortly publish a selection of recent papers about:

- Stereo/multi-view stereo
- 3D reconstruction/Structure from Motion
- 3D SLAM/Visual SLAM
- Local/global 3D features
- Point cloud preocessing/semantic segmentation
- 3D object pose estimation
- 3D Shape registration
- 3D sensors
- Neural fileds
- ...

You will have a couple of days to checkout them, then each paper will be assigned to only one group with a "first come-first served" policy

Selection via moodle, we will communicate the time/date shortly

Poster Session

- The poster session will be **public**, so you will have the opportunity to present your poster to other interested students, PhD students, or researchers from DEI.
- A poster presentation is typically 6-7 minutes long (around 3 minutes each student), and it is more informal and interactive compared with the classical oral presentations.
- All the posters are set up at once, and each presenter is expected to stand with their poster for the entirety of the session to answer questions from attenders.

Poster Evaluation

My collaborators and I will review each poster, listening to your presentation and asking questions.

Poster Layout

Format: The poster should be prepared as a single face, landscape or portrait, A0 sheet, and saved in **PDF** format.

Tools: You may use Microsoft Office Power Point, LibreOffice Impress, Google Slides, ...

Poster Submission

- Through the e-learning website (a submission link will be shared soon) by **June 9 h. 24**. I will print the posters you submitted, so on the day of presentation I'll bring the posters
- You may also bring your poster on June 11, but in this case you will have to pay for the printing of the poster ☻...

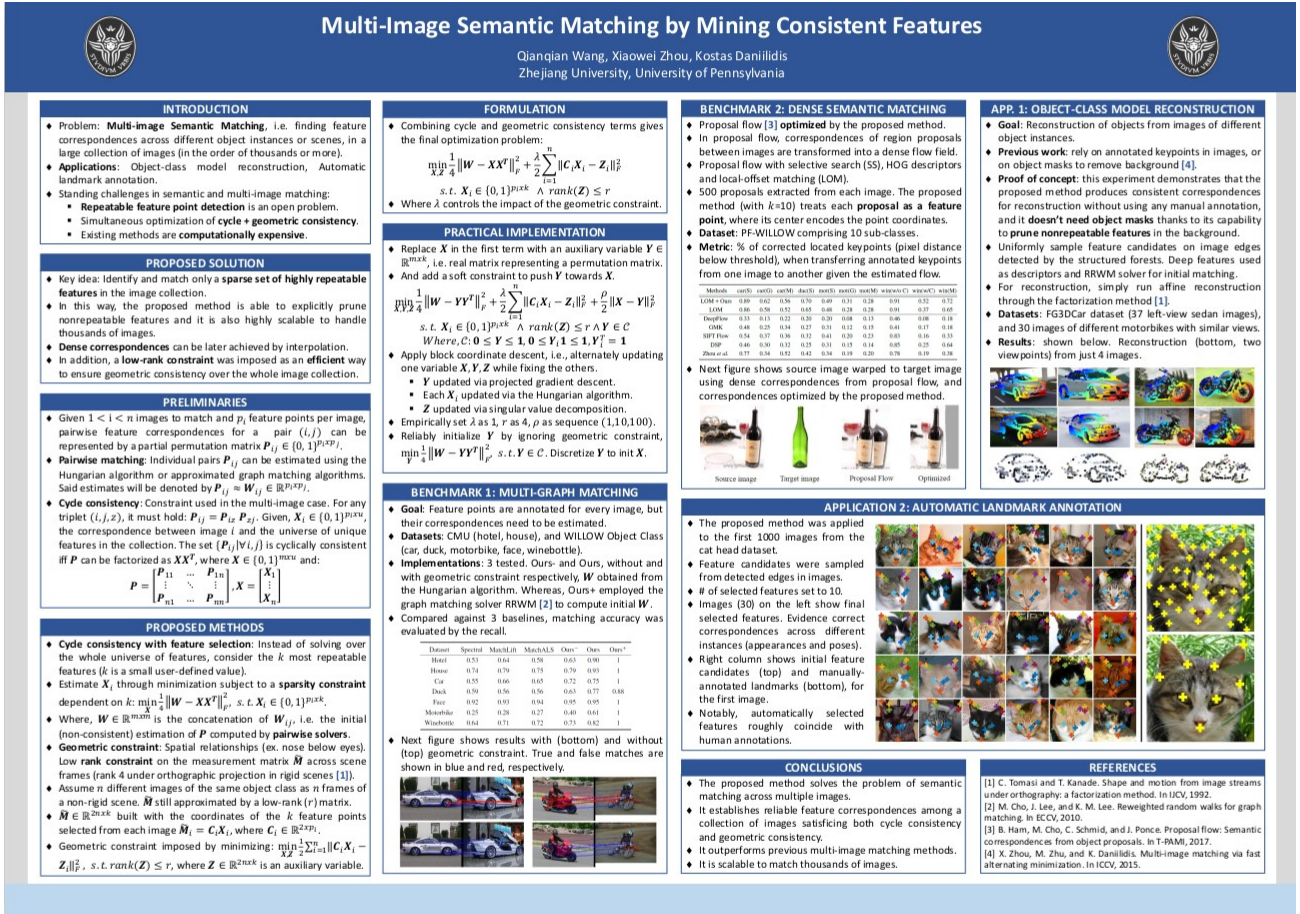
What is a Research Poster?

- Posters should summarize the paper **concisely** and **attractively** to help advertise it and generate discussion.
- The poster should report **brief and clear text blocks** mixed with easy to understand **tables, graphs, pictures**.
- The poster should support the presentation and the interaction with the audience.

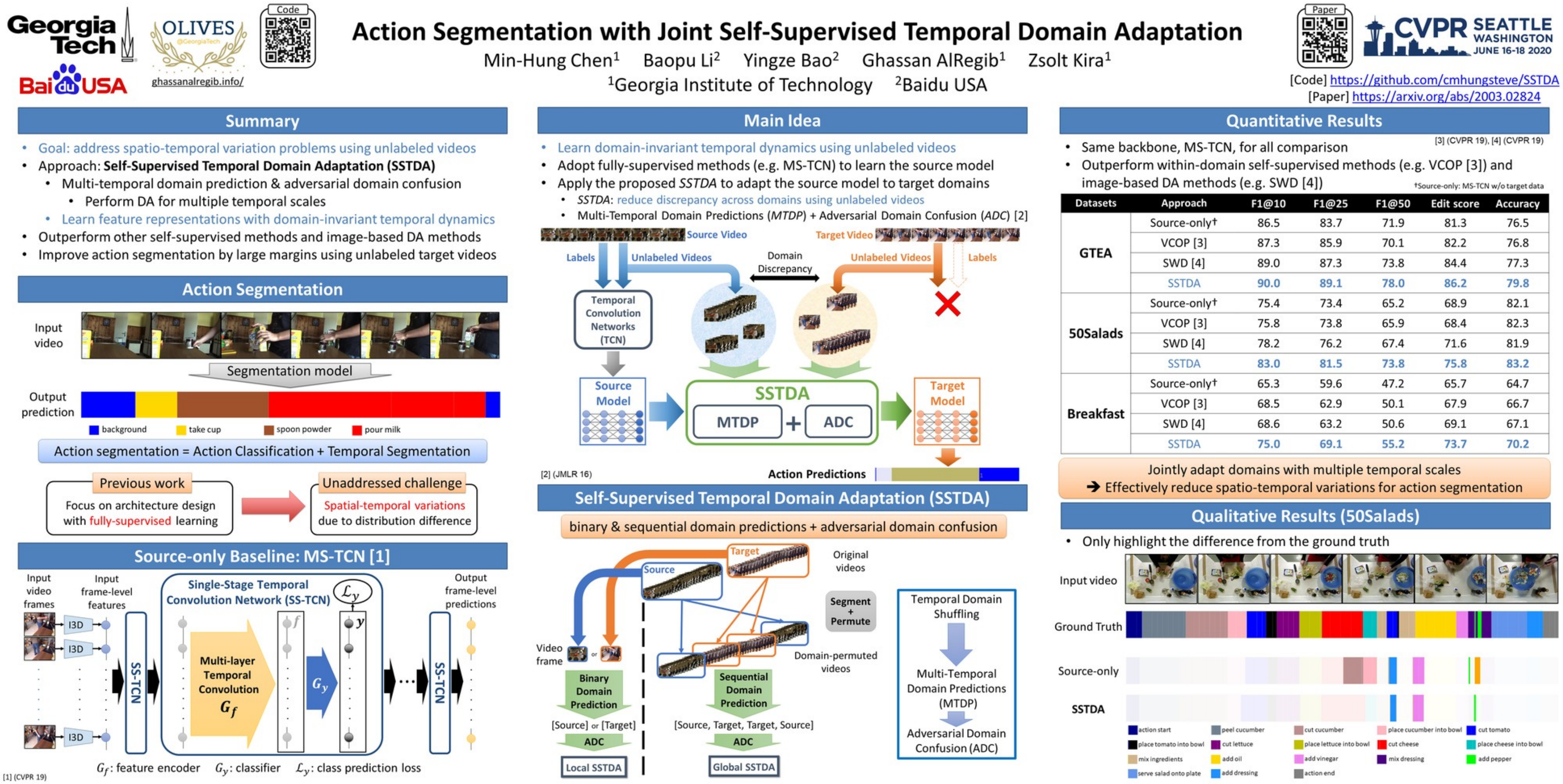
What Makes a Good Poster?

- Important information should be readable from 3 m
- Word count: 500 to 1000 words
- Text should be clear and direct → subject, verb and object complement!
- Use bullets, numbering, and boxes.
- Consistent, clean and attractive layout → **Effective use of graphics, color and fonts**.

"Bad" Poster


Multi-Image Semantic Matching by Mining Consistent Features
Qianqian Wang, Xiaowei Zhou, Kostas Daniilidis
Zhejiang University, University of Pennsylvania
INTRODUCTION
PROPOSED SOLUTION
PRELIMINARIES
PROPOSED METHODS
FORMULATION
PRACTICAL IMPLEMENTATION
BENCHMARK 1: MULTI-GRAPH MATCHING
BENCHMARK 2: DENSE SEMANTIC MATCHING
APP. 1: OBJECT-CLASS MODEL RECONSTRUCTION
APPLICATION 2: AUTOMATIC LANDMARK ANNOTATION
CONCLUSIONS
REFERENCES

Better Poster



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

Georgia Tech · Baidu USA · OLIVES ghassanalregib.info/ · Code · Paper · CVPR Seattle Washington June 16-18 2020

[Code] https://github.com/cmhungsteve/SSTDA
[Paper] https://arxiv.org/abs/2003.02824

Summary

- Goal: address spatio-temporal variation problems using unlabeled videos
- Approach: **Self-Supervised Temporal Domain Adaptation (SSTDA)**
  - Multi-temporal domain prediction & adversarial domain confusion
    - Perform DA 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

Action Segmentation

Input video
Segmentation model
Output prediction
background · take cup · spoon powder · pour milk

Action segmentation = Action Classification + Temporal Segmentation

Previous work: Focus on architecture design with fully-supervised learning
Unaddressed challenge: Spatial-temporal variations due to distribution difference

Source-only Baseline: MS-TCN [1]

Input video frames · Input frame-level features · I3D · SS-TCN · Single-Stage Temporal Convolution Network (SS-TCN) · Multi-layer Temporal Convolution G_f · G_y · $\mathcal{L}_y$ · Output frame-level predictions

G_f: feature encoder G_y: classifier $\mathcal{L}_y$: class prediction loss

[1] (CVPR 19)

Main Idea

- Learn domain-invariant temporal dynamics using unlabeled videos
- Adopt fully-supervised methods (e.g. MS-TCN) to learn the source model
- Apply the proposed *SSTDA* to adapt the source model to target domains
  - *SSTDA*: reduce discrepancy across domains using unlabeled videos
  - Multi-Temporal Domain Predictions (*MTDP*) + Adversarial Domain Confusion (*ADC*) [2]

Source Video · Target Video · Labels · Unlabeled Videos · Domain Discrepancy · Temporal Convolution Networks (TCN) · Source Model · SSTDA · MTDP + ADC · Target Model · Action Predictions

[2] (JMLR 16)

Self-Supervised Temporal Domain Adaptation (SSTDA)

binary & sequential domain predictions + adversarial domain confusion

Source · Target · Original videos · Segment + Permute · Domain-permuted videos · Video frame · Binary Domain Prediction · [Source] or [Target] · ADC · Local SSTDA · Sequential Domain Prediction · [Source, Target, Target, Source] · ADC · Global SSTDA

Temporal Domain Shuffling → Multi-Temporal Domain Predictions (MTDP) → Adversarial Domain Confusion (ADC)

Quantitative Results

- Same backbone, MS-TCN, for all comparison
- Outperform within-domain self-supervised methods (e.g. VCOP [3]) and image-based DA methods (e.g. SWD [4])

[3] (CVPR 19), [4] (CVPR 19) †Source-only: MS-TCN w/o target data

Datasets	Approach	F1@10	F1@25	F1@50	Edit score	Accuracy
GTEA	Source-only†	86.5	83.7	71.9	81.3	76.5
	VCOP [3]	87.3	85.9	70.1	82.2	76.8
	SWD [4]	89.0	87.3	73.8	84.4	77.3
	SSTDA	90.0	89.1	78.0	86.2	79.8
50Salads	Source-only†	75.4	73.4	65.2	68.9	82.1
	VCOP [3]	75.8	73.8	65.9	68.4	82.3
	SWD [4]	78.2	76.2	67.4	71.6	81.9
	SSTDA	83.0	81.5	73.8	75.8	83.2
Breakfast	Source-only†	65.3	59.6	47.2	65.7	64.7
	VCOP [3]	68.5	62.9	50.1	67.9	66.7
	SWD [4]	68.6	63.2	50.6	69.1	67.1
	SSTDA	75.0	69.1	55.2	73.7	70.2

Jointly adapt domains with multiple temporal scales
➔ Effectively reduce spatio-temporal variations for action segmentation

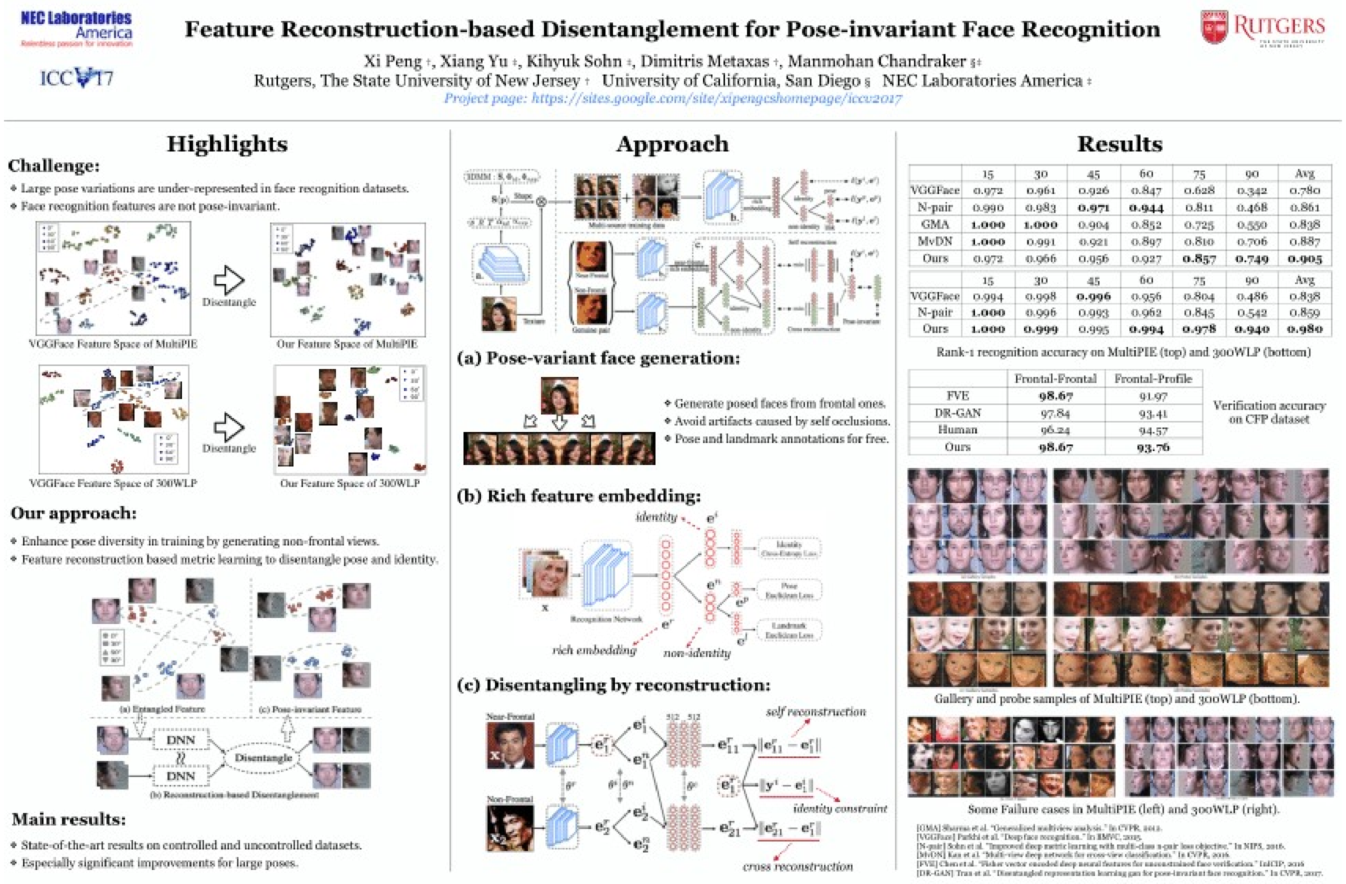
Qualitative Results (50Salads)

- Only highlight the difference from the ground truth

Input video · Ground Truth · Source-only · 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

Better Poster



	15	30	45	60	75	90	Avg
VGGFace	0.972	0.961	0.926	0.847	0.628	0.342	0.780
N-pair	0.990	0.983	0.971	0.944	0.811	0.468	0.861
GMA	1.000	1.000	0.904	0.852	0.725	0.550	0.838
MvDN	1.000	0.991	0.921	0.897	0.810	0.706	0.887
Ours	0.972	0.966	0.956	0.927	0.857	0.749	0.905

	15	30	45	60	75	90	Avg
VGGFace	0.994	0.998	0.996	0.956	0.804	0.486	0.838
N-pair	1.000	0.996	0.993	0.962	0.845	0.542	0.859
Ours	1.000	0.999	0.995	0.994	0.978	0.940	0.980

	Frontal-Frontal	Frontal-Profile
FVE	98.67	91.97
DR-GAN	97.84	93.41
Human	96.24	94.57
Ours	98.67	93.76

Better Poster