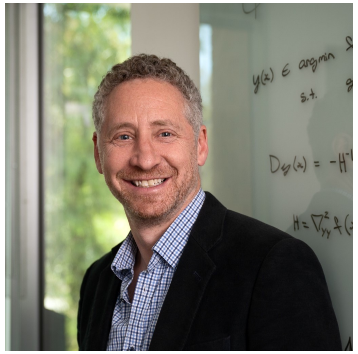
Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation



Ming Xu

Stephen Gould

CVPR SEATTLE, WA JUNE 17-21, 2024

seeing machines

Australian National University

Temporal Action Segmentation (TAS)

- 1D (temporal) segmentation task
- Assign video ***frames*** to one of K action classes.

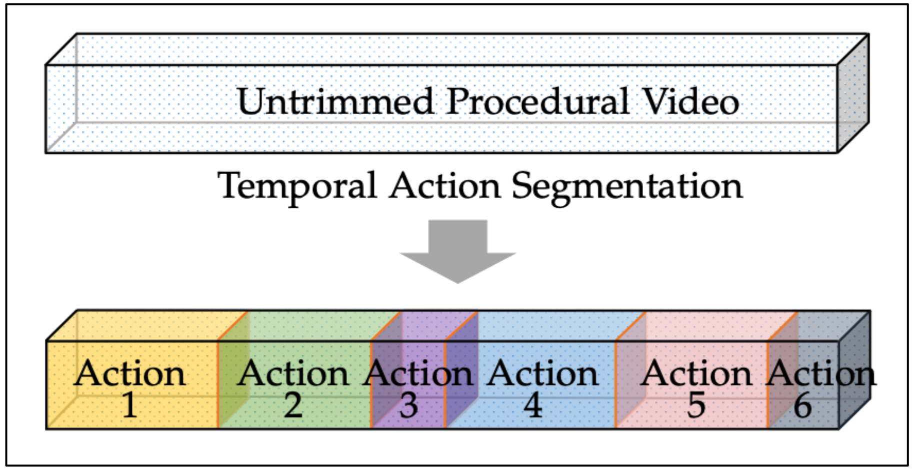


Figure courtesy Ding et al., 2023

Example: Instructional video of CPR

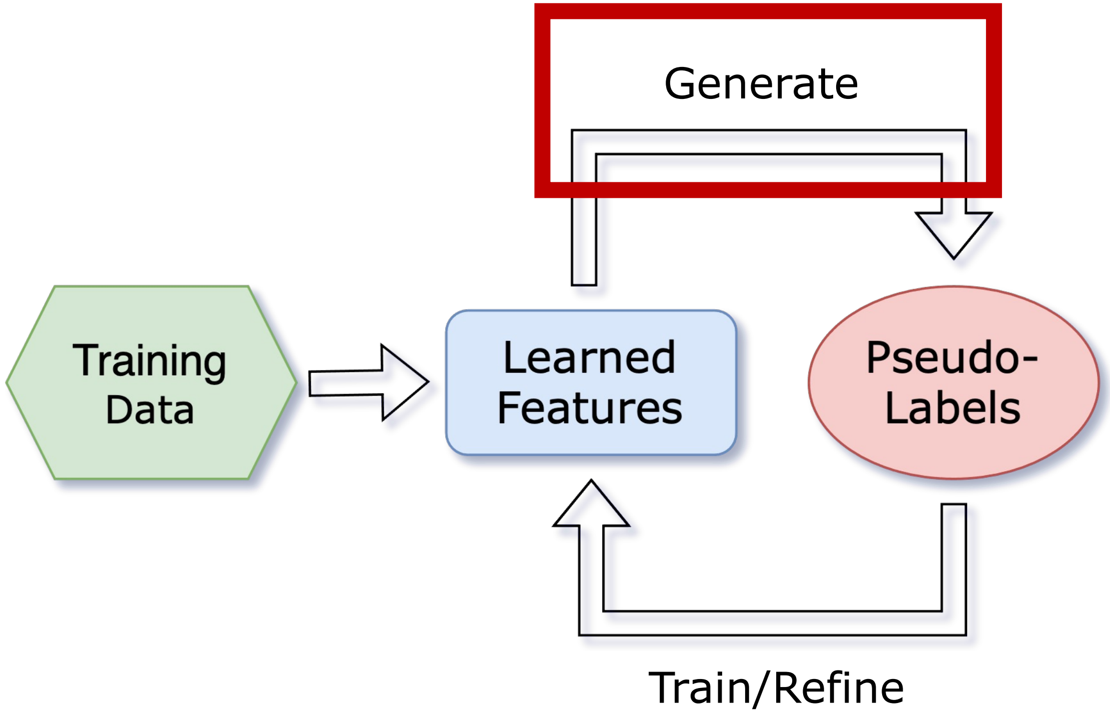
Figure courtesy Kumar et al., 2022

- ***Long-form:*** can be several minutes long
- ***Multi-stage*** assembly/instructional videos

Unsupervised Learning: Simultaneous Learning and Clustering

Jointly learn ***representations*** and ***labels*** from a dataset.

- Alternate between label generation and learning.
- ***Interpretation:*** clustering as auxiliary task.

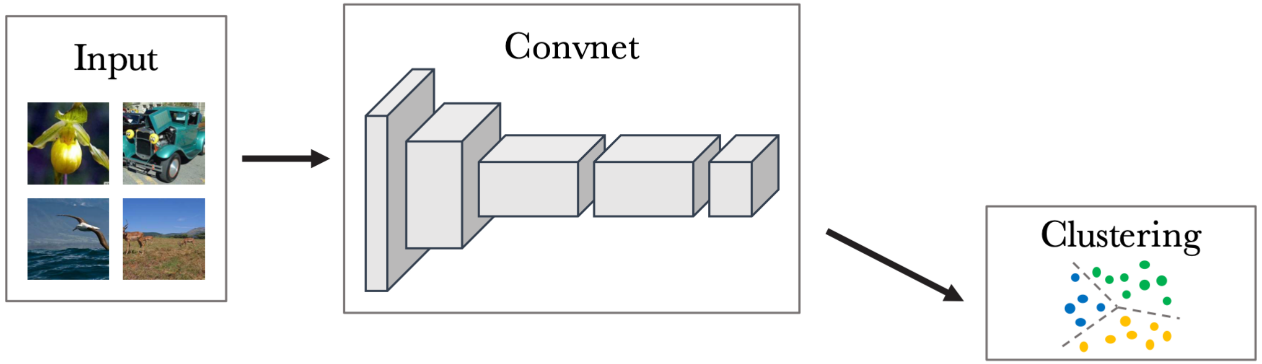


Example: Image Classification

Australian National University

Example: DeepClustering (Caron et al., 2018)

Figure courtesy Caron et al., 2018

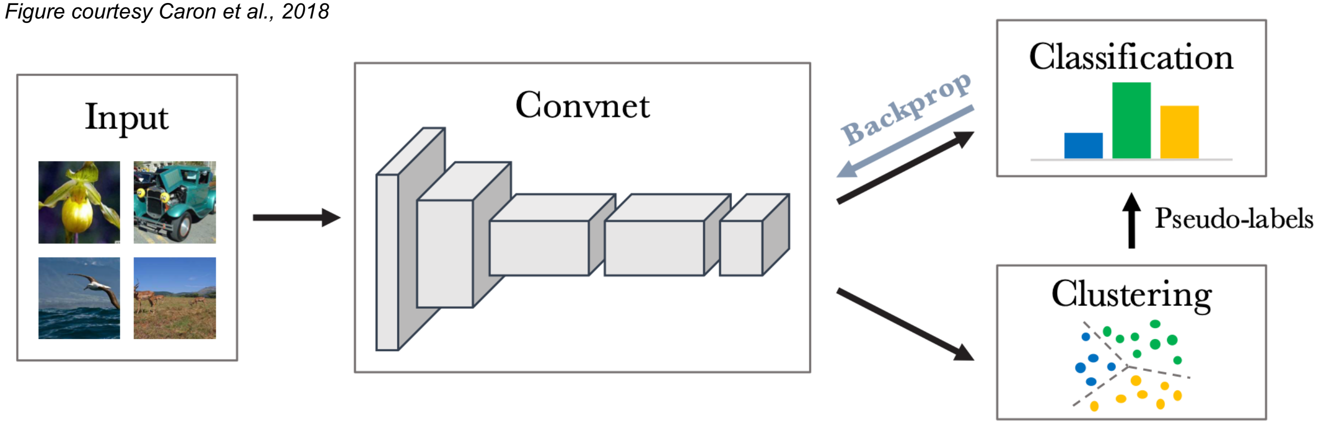


- K-means clustering to learned representations

Example: Image Classification

Example: DeepClustering (Caron et al., 2018)

Figure courtesy Caron et al., 2018

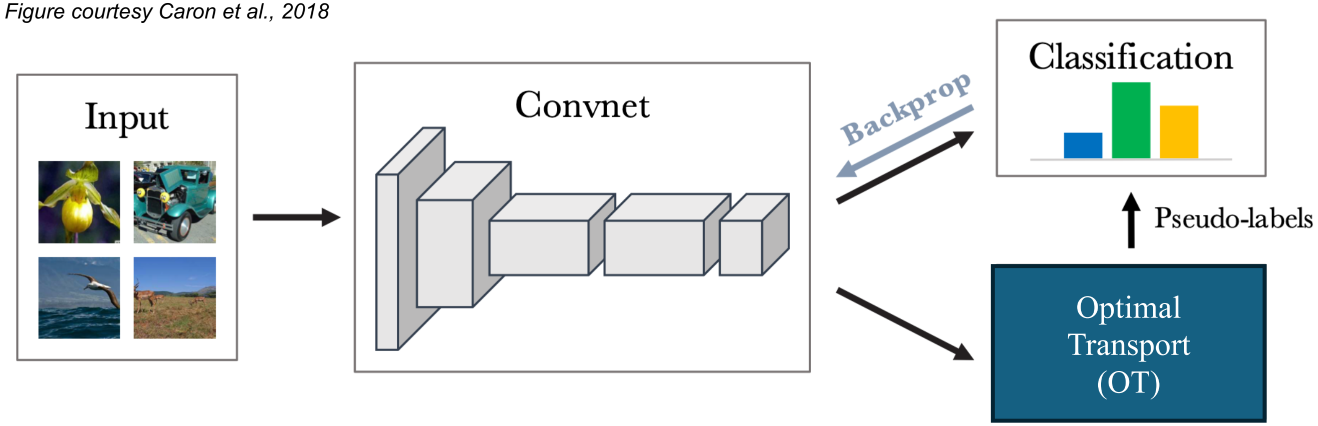


- ➢ K-means clustering to learned representations
- ➢ Cluster assignments become pseudo-labels

Example: Image Classification

Example: SeLA (Asano et al., 2020)

Figure courtesy Caron et al., 2018



➢ **Idea:** Use ***optimal transport (OT)*** for label generation!

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{aligned} \underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad & \langle \boxed{\mathbf{C},} \mathbf{T}\rangle, \\ \text{subject to} \quad & \mathbf{T}\mathbf{1}_K = \tfrac{1}{N}\mathbf{1}_N, \\ & \mathbf{T}^\top\mathbf{1}_N = \tfrac{1}{K}\mathbf{1}_K, \end{aligned}$$

Label assignment cost

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{aligned} \underset{\mathbf{T} \in \mathbb{R}_+^{N \times K}}{\text{minimize}} \quad & \langle \boxed{\mathbf{C},} \mathbf{T} \rangle, \\ \text{subject to} \quad & \mathbf{T}\mathbf{1}_K = \tfrac{1}{N}\mathbf{1}_N, \\ & \mathbf{T}^\top \mathbf{1}_N = \tfrac{1}{K}\mathbf{1}_K, \end{aligned}$$

Label assignment cost,

e.g., negative of the logits from the FC layer.

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{aligned} \underset{\mathbf{T} \in \mathbb{R}_+^{N \times K}}{\text{minimize}} \quad & \langle \mathbf{C}, \mathbf{T} \rangle, \\ \text{subject to} \quad & \boxed{\begin{aligned} \mathbf{T}\mathbf{1}_K &= \tfrac{1}{N}\mathbf{1}_N, \\ \mathbf{T}^\top \mathbf{1}_N &= \tfrac{1}{K}\mathbf{1}_K, \end{aligned}} \end{aligned}$$

Balanced assignment / equipartition constraint

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{aligned} \underset{\mathbf{T} \in \mathbb{R}_{+}^{N \times K}}{\text{minimize}} \quad & \langle \mathbf{C}, \mathbf{T} \rangle, \\ \text{subject to} \quad & \boxed{\mathbf{T} \mathbf{1}_K = \frac{1}{N} \mathbf{1}_N,} \\ & \mathbf{T}^\top \mathbf{1}_N = \frac{1}{K} \mathbf{1}_K, \end{aligned}$$

All images must be labelled

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{aligned} \underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad & \langle \mathbf{C}, \mathbf{T} \rangle, \\ \text{subject to} \quad & \mathbf{T}\mathbf{1}_K = \tfrac{1}{N}\mathbf{1}_N, \\ & \boxed{\mathbf{T}^\top \mathbf{1}_N = \tfrac{1}{K}\mathbf{1}_K,} \end{aligned}$$

Pseudo-labels must be evenly spread across clusters, i.e., N/K labels per cluster

Optimal Transport for Pseudo-labels

Australian National University

For N training images and K clusters/classes, solve

$$\begin{array}{ll}\underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} & \langle \mathbf{C}, \mathbf{T} \rangle, \\ \text{subject to} & \mathbf{T}\mathbf{1}_K = \frac{1}{N}\mathbf{1}_N, \\ & \boxed{\mathbf{T}^\top \mathbf{1}_N = \frac{1}{K}\mathbf{1}_K,}\end{array}$$

Pseudo-labels must be evenly spread across clusters,
prevents collapse!

Optimal Transport for Pseudo-labels

Australian National University

Works well for *image classification* datasets with

- ***unstructured*** image collections and,
- ***balanced*** ground truth class annotations

Optimal Transport for Pseudo-labels

Australian National University

Remark: Sinkhorn-Knopp for entropy regularised OT

- $O(NK)$ complexity per iteration
- Amenable to GPU computation (few lines of PyTorch)
- Fast convergence in practice

Does This Work for Temporal Action Segmentation?

Isn't This Just an Image Dataset?

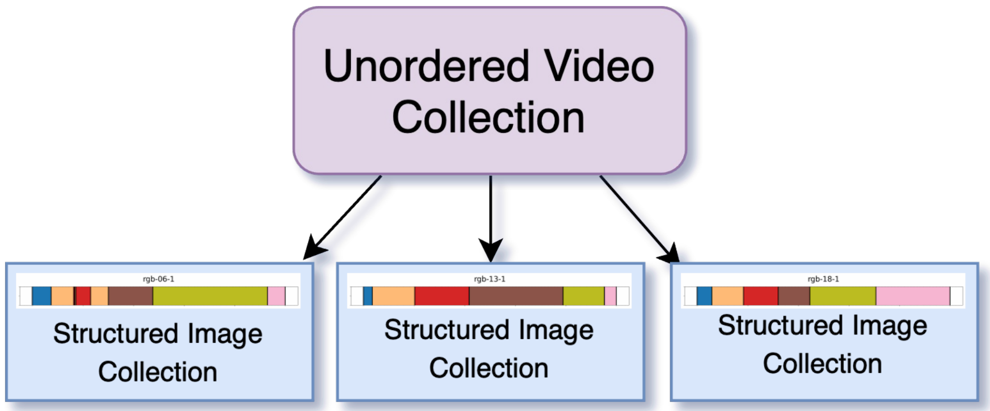
Australian National University

We still have a collection of images… what has changed?

Image Classification

Unordered Image Collection

Temporal Action Segmentation



"Standard" optimal transport has **no understanding of structure**!

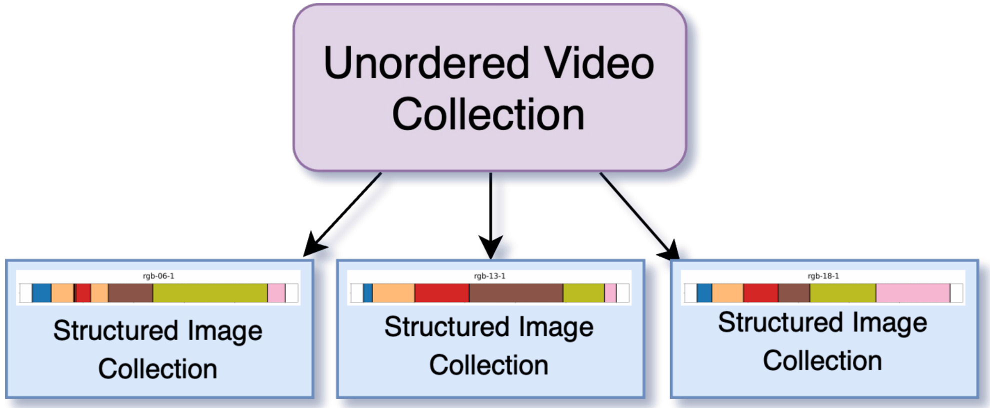
Isn’t This Just an Image Dataset?

We still have a collection of images… what has changed?

Image Classification

Unordered Image Collection

Temporal Action Segmentation



i.e., temporal consistency!

Long-tail Class Distributions

e.g., Breakfast dataset

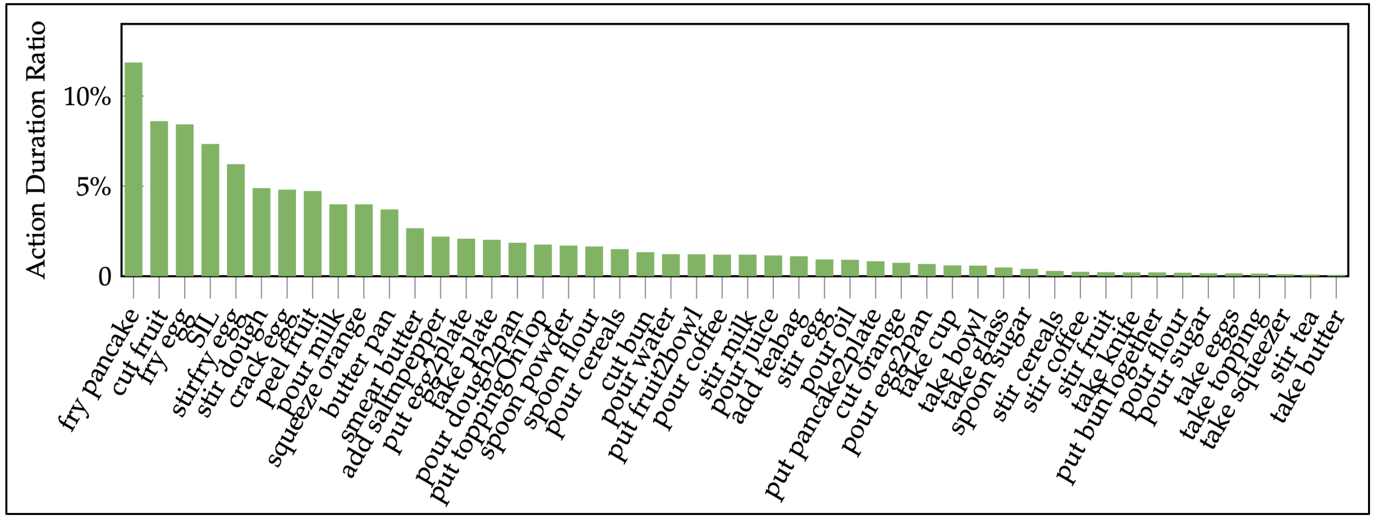


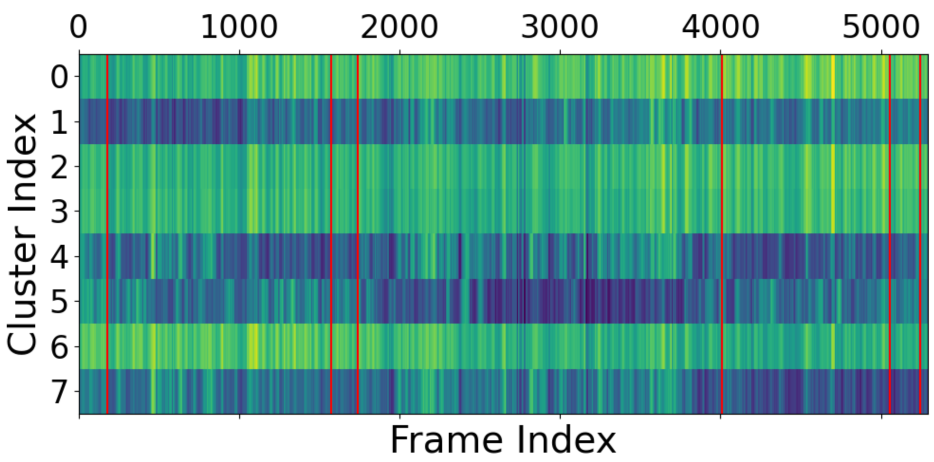
Figure courtesy Ding et al., 2023[1]

Difficult to curate balanced classes

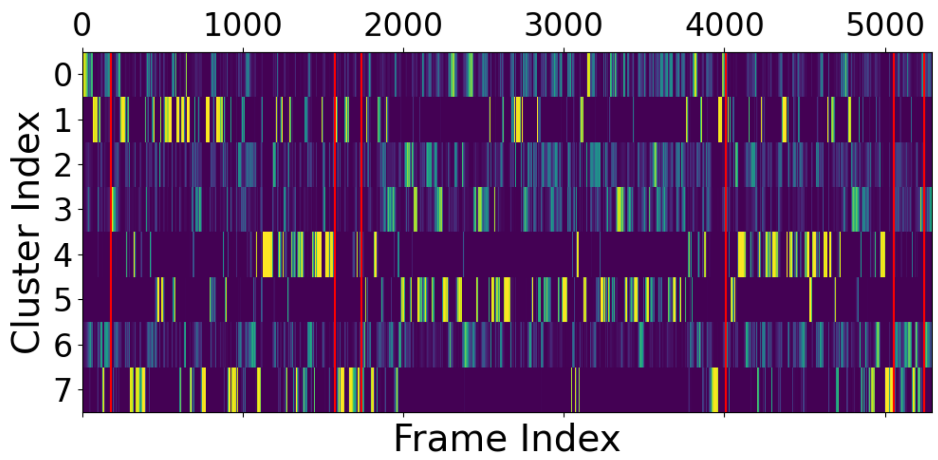
[1] Ding et al. Temporal Action Segmentation: An Analysis of Modern Techniques. IEEE TPAMI, 2023.

Standard Optimal Transport for Videos: Let's Try!

Label assignment costs ($\mathbf{C}$)



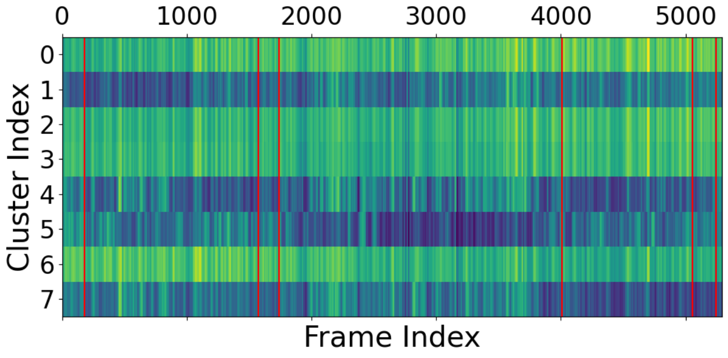
OT pseudo-labels ($\mathbf{T}$)



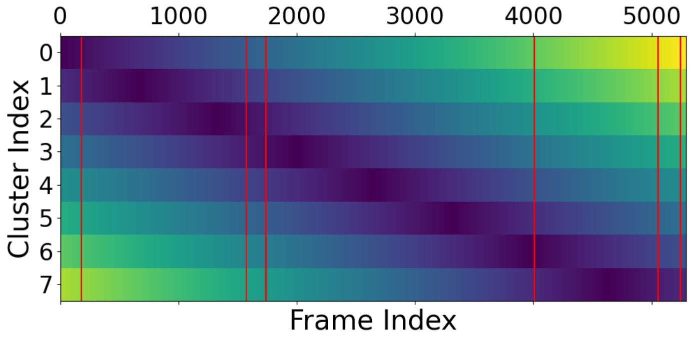
- Temporal consistency ✖
- Long-tail class distribution ✖

A Regularisation Approach

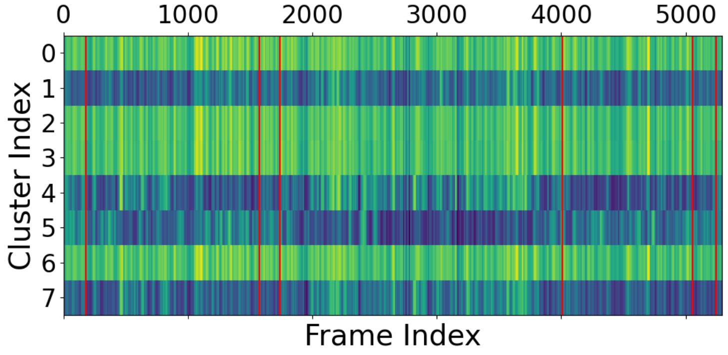
Label assignment costs ($\mathbf{C}$)



Temporal Regularisation ($\mathbf{R}$)



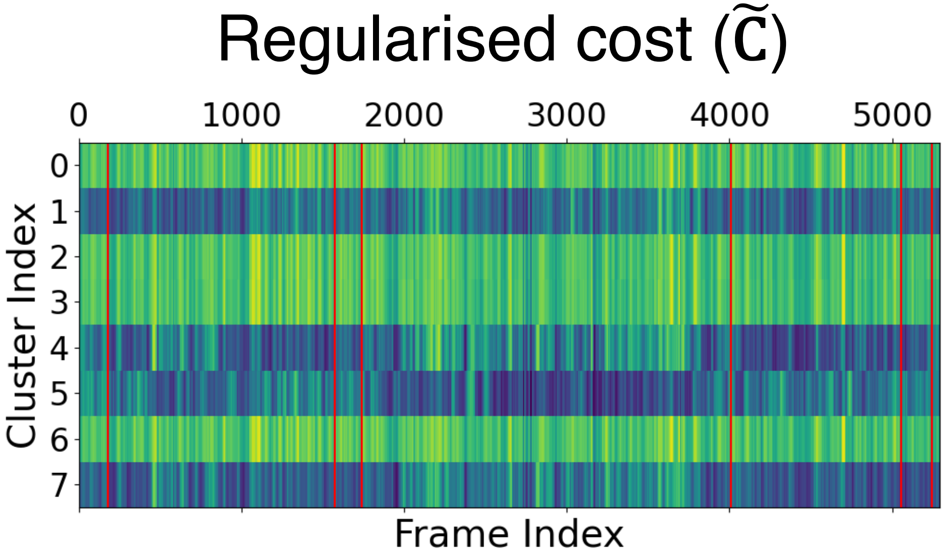
Regularised cost ($\widetilde{\mathbf{C}}$)

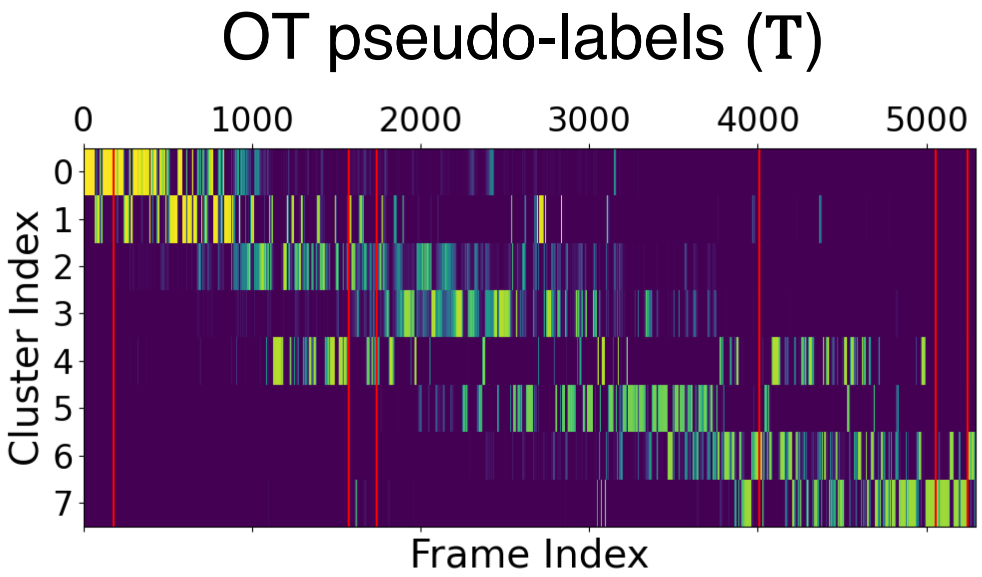


Kumar et al. Unsupervised Action Segmentation by Joint Representation Learning and Online Clustering. CVPR 2022

A Regularisation Approach

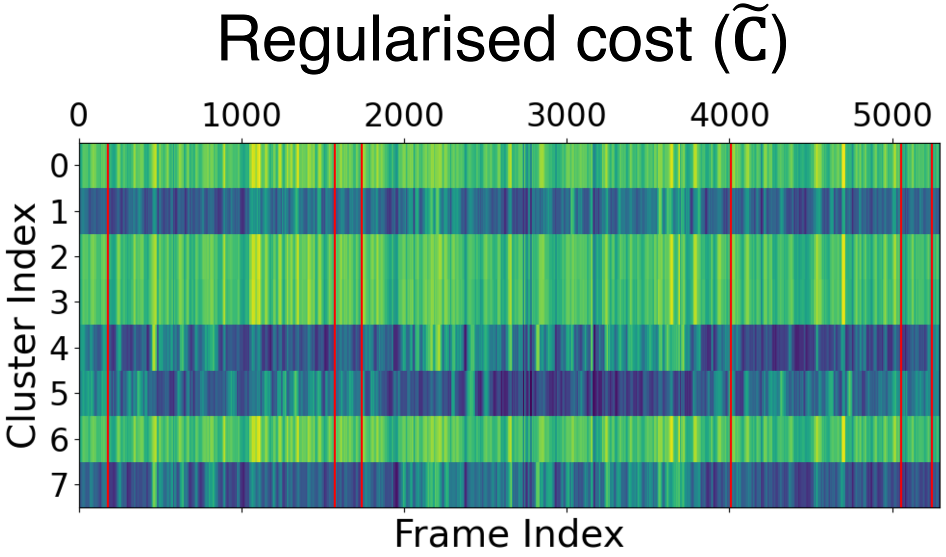
Australian National University

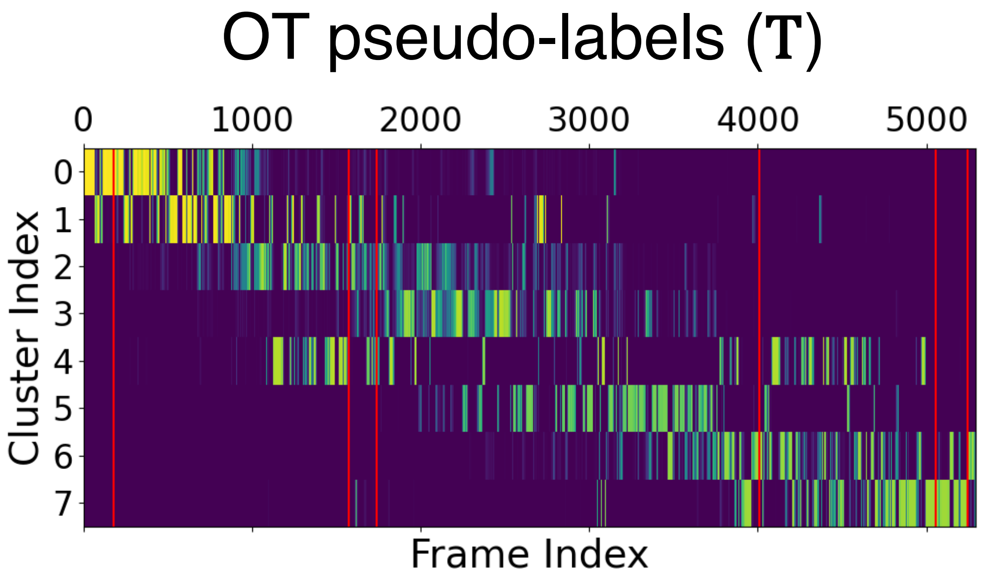




- ➢ Temporal consistency ?
- ➢ Long-tail class distribution ✖

A Regularisation Approach





Also assumes actions always ***occur in the same order***!

Core Methodology

Our Approach: Use ***Non***-standard OT!

Avoid “standard” OT, use ***structured optimal transport***.

We use an *unbalanced*, *fused Gromov-Wasserstein* formulation

- Temporal consistency
- Long-tail class distributions

Our Approach: Use ***Non***-standard OT!

Australian National University

Avoid “standard” OT, use ***structured optimal transport***.

We use an *unbalanced*, *fused Gromov-Wasserstein* formulation

- Temporal consistency → **Gromov-Wasserstein**
- Long-tail class distributions → **unbalanced transport**

Gromov-Wasserstein for Encoding Structural Priors

A (relatively) general formulation for (discrete) GW problems:

$$\begin{aligned} \underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad & \sum_{\substack{i,k\in[N]\\ j,l\in[K]}} L(C^v_{ik}, C^a_{jl}) T_{ij} T_{kl}, \\ \text{s.t.} \quad & \mathbf{T}\mathbf{1}_K = \frac{1}{N}\mathbf{1}_N, \\ & \mathbf{T}^\top \mathbf{1}_N = \frac{1}{K}\mathbf{1}_K, \end{aligned}$$

- Cost matrices $\mathbf{C}^v \in \mathbb{R}^{N\times N}$ and $\mathbf{C}^a \in \mathbb{R}^{K\times K}$
- "Loss" function between cost matrix elements L: $\mathbb{R}\times\mathbb{R}\to\mathbb{R}$

Gromov-Wasserstein for Encoding Structural Priors

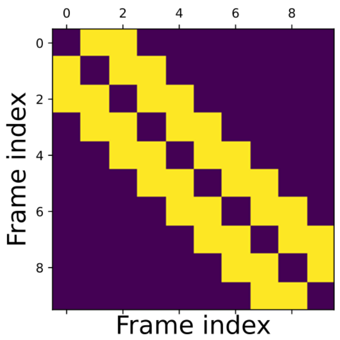
A (relatively) general formulation for (discrete) GW problems:

$$\begin{aligned} \underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad & \textstyle\sum_{\substack{i,k\in[N] \\ j,l\in[K]}} L(C^v_{ik}, C^a_{jl}) T_{ij} T_{kl}, \\ \text{s.t.} \quad & \mathbf{T}\mathbf{1}_K = \tfrac{1}{N}\mathbf{1}_N, \\ & \mathbf{T}^\top \mathbf{1}_N = \tfrac{1}{K}\mathbf{1}_K, \end{aligned}$$

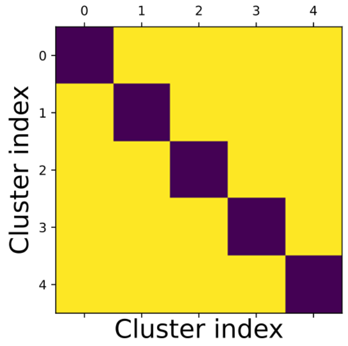
- ***Quadratic*** instead of linear objective (non-convex)
- ***Quadratic*** term allows us to encode **structural priors**

Gromov-Wasserstein for Encoding Structural Priors

For a **single video** with N frames and K action classes,



$$\mathbf{C}^v_{ik} := \begin{cases} 0 & i = k \\ 1/r & |i-k| \le Nr \ \ \& \ \ i \neq k \\ 0 & |i-k| > Nr \end{cases}, \quad \text{and} \quad \mathbf{C}^a_{jl} := \begin{cases} 0 & j = l \\ 1 & \text{otherwise} \end{cases}.$$

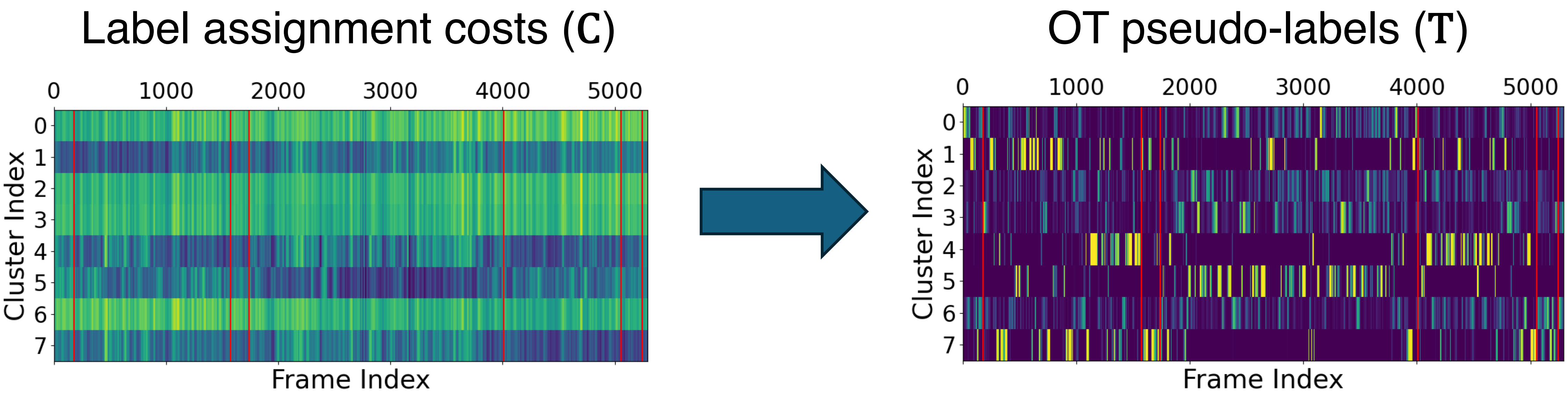


- Let $L(a,b) := ab$ and let $0 < r < 1$ be a *temporal radius* parameter.
- **Remark:** Objective function is ***simplified*** to $\langle \mathtt{C}^{\mathtt{v}}\mathtt{T}\mathtt{C}^{\mathtt{a}}, \mathtt{T} \rangle$.

Intuition: Labelling adjacent frames to different clusters incurs a cost

Effect of the Structural Prior

From this….

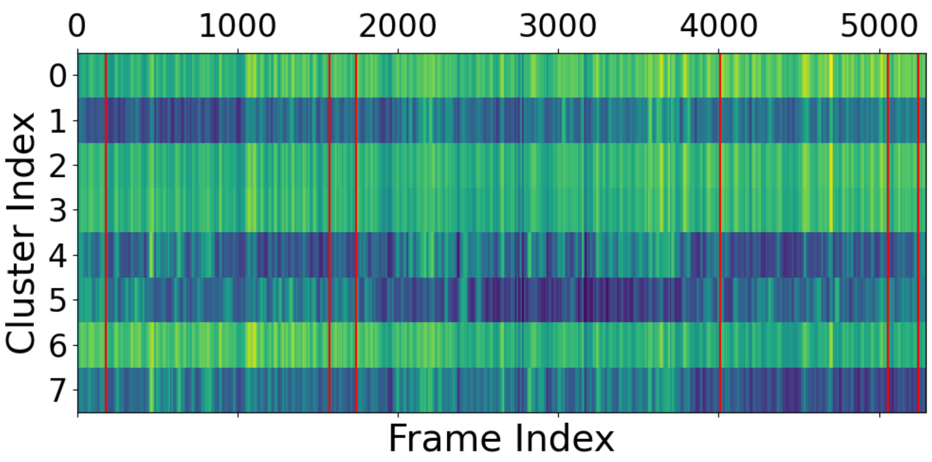


Effect of the Structural Prior

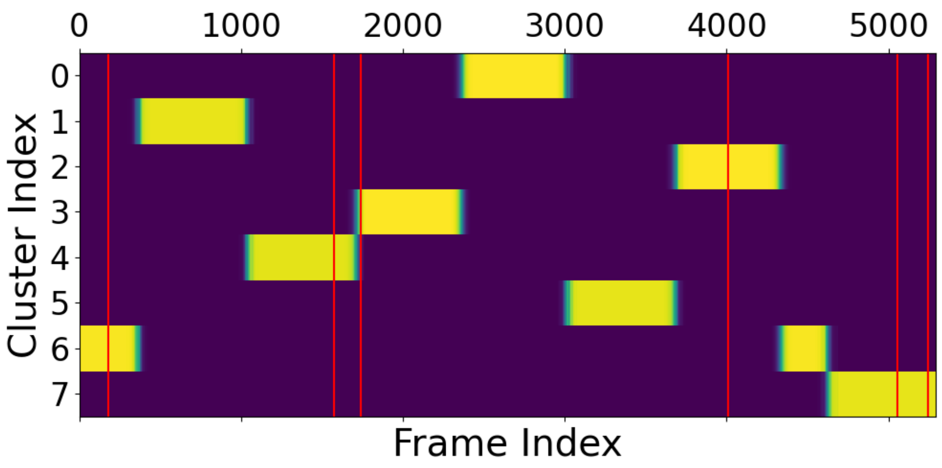
Australian National University

to this!

Label assignment costs ($\mathbf{C}$)



GW OT pseudo-labels ($\mathbf{T}$)



- Labels are still balanced however…

Unbalanced Transport for Long-tail Class Distributions

For standard optimal transport, **replace constraints…**

$$
\begin{aligned}
\underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad & \langle \mathbf{C}, \mathbf{T} \rangle, \\
\text{subject to} \quad & \mathbf{T}\mathbf{1}_K = \tfrac{1}{N}\mathbf{1}_N, \\
& \boxed{\mathbf{T}^\top \mathbf{1}_N = \tfrac{1}{K}\mathbf{1}_K,}
\end{aligned}
$$

Unbalanced Transport for Long-tail Class Distributions

For standard optimal transport, **with a penalty!**

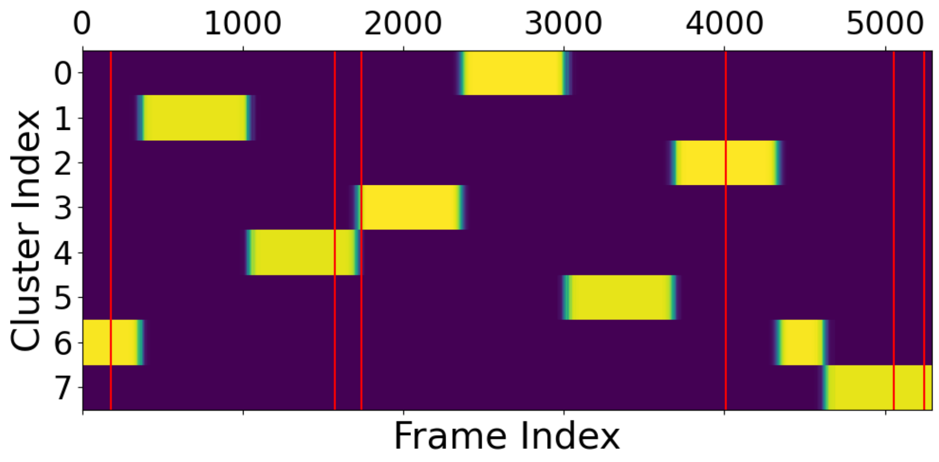
$$\begin{aligned} \underset{\mathbf{T} \in \mathbb{R}_+^{N \times K}}{\text{minimize}} \quad & \langle \mathbf{C}, \mathbf{T} \rangle \boxed{+ \lambda \mathrm{D}_{\mathrm{KL}}(\mathbf{T}^\top \mathbf{1}_N \| \tfrac{1}{K} \mathbf{1}_K),} \\ \text{subject to} \quad & \mathbf{T} \mathbf{1}_K = \tfrac{1}{N} \mathbf{1}_N \end{aligned}$$

- Adapt parameter $\lambda > 0$ to reflect the level of class imbalance
- We use the KL-divergence, but **other options are possible**

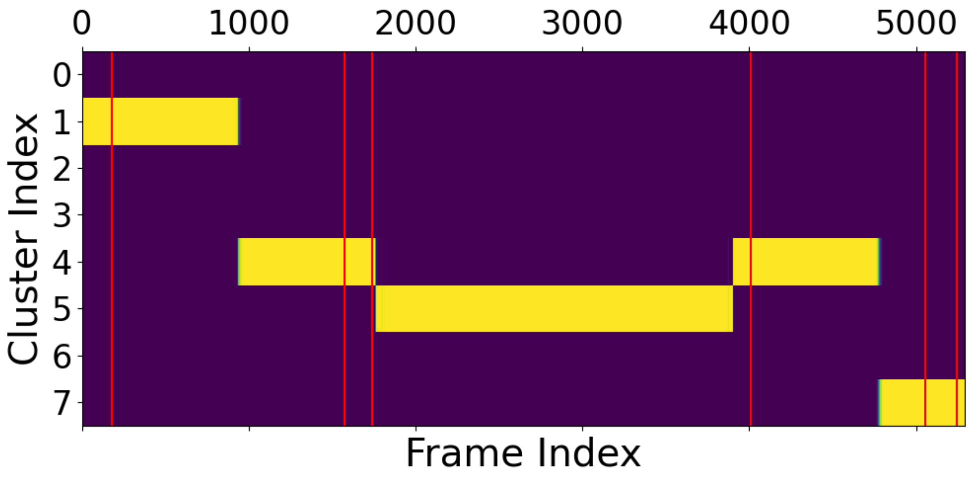
Unbalanced Transport for Long-tail Class Distributions

Australian National University

GW OT pseudo-labels ($\mathbf{T}$)



ASOT pseudo-labels ($\mathbf{T}$)



- Temporal consistency
- Long-tail class distribution

Action Segmentation Optimal Transport (ASOT)

Our final, ASOT formulation solves the problem

$$\begin{aligned}\underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}}\quad & \underbrace{\boxed{\alpha\langle\mathbf{C}^v\mathbf{T}\mathbf{C}^a,\mathbf{T}\rangle}}_{\text{temp. consist.}} + \underbrace{\boxed{(1-\alpha)\langle\mathbf{C},\mathbf{T}\rangle}}_{\text{learned repn.}} + \underbrace{\boxed{\lambda\mathrm{D}_{\mathrm{KL}}(\mathbf{T}^\top\mathbf{1}_N\|\tfrac{1}{K}\mathbf{1}_K)}}_{\text{long-tail class distn.}}, \\ \text{subject to}\quad & \mathbf{T}\mathbf{1}_K=\tfrac{1}{N}\mathbf{1}_N\end{aligned}$$

where $\alpha \in [0,1]$ is the relative weighting of the structure term

Unbalanced, fused Gromov-Wasserstein problem!

Action Segmentation Optimal Transport (ASOT)

Our final, ASOT formulation solves the problem

$$\underset{\mathbf{T}\in\mathbb{R}_+^{N\times K}}{\text{minimize}} \quad \underbrace{\boxed{\alpha\langle \mathbf{C}^v\mathbf{T}\mathbf{C}^a, \mathbf{T}\rangle}}_{\text{temp. consist.}} + \underbrace{\boxed{(1-\alpha)\langle \mathbf{C}, \mathbf{T}\rangle}}_{\text{learned repn.}} + \underbrace{\boxed{\lambda \mathrm{D}_{\mathrm{KL}}(\mathbf{T}^\top \mathbf{1}_N \| \tfrac{1}{K}\mathbf{1}_K),}}_{\text{long-tail class distn.}}$$

$$\text{subject to} \quad \mathbf{T}\mathbf{1}_K = \frac{1}{N}\mathbf{1}_N$$

where $\alpha \in [0,1]$ is the relative weighting of the structure term

Unbalanced, fused ***Gromov-Wasserstein*** problem!

ASOT is Efficient

Australian National University

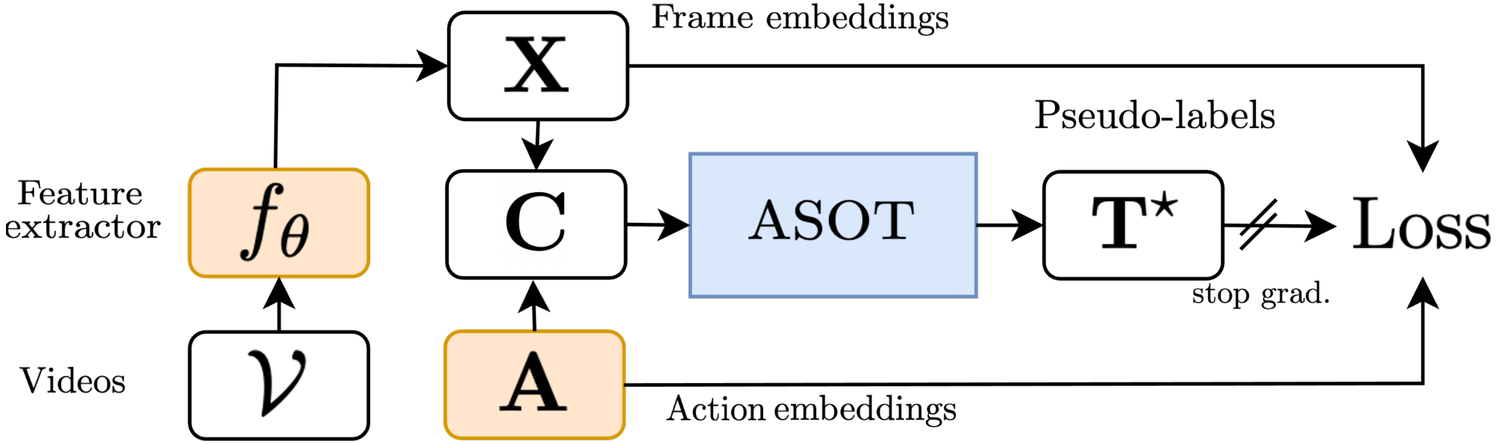
ASOT is solved using ***projected mirror descent***

- Each iteration has complexity $O(NK)$
- Still amenable to GPUs (and simple PyTorch code)
- ***24.1ms*** for $N = 16k$ frames (~9 mins of video) and $K = 19$ classes on single RTX 4090

Experimental Results

Unsupervised Temporal Action Segmentation: Training Pipeline

Simple self-training pipeline w/ ASOT pseudo-labels



- Raw data is frame features, not images
- Simple MLP frame feature encoder (random init.)
- Pseudo-labels generated "online", i.e., per batch

ASOT

Australian National University

SOTA!

Australian National University

State-of-the-art Comparison

Australian National University

Metrics: Mean-over-frames accuracy (MoF), segmental F1-score (F1), framewise mean intersection-over-union (mIoU)

	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
CTE[1]	41.8 / 26.4 / -	39.0 / 28.3 / -	30.2 / - / -	35.5 / - / -	47.6 / 44.9 / -
TOT[2]	47.5 / 31.0 / -	40.6 / 30.0 / -	31.8 / - / -	47.4 / 42.8 / -	56.3 / 51.7 / -
UFSA[3]	52.1 / 38.0 / -	49.6 / 32.4 / -	36.7 / 30.4 / -	55.8 / 50.3 / -	65.4 / 63.0 / -
ASOT (Ours)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**

Table: State-of-the-art comparison results. For all evaluation metrics, higher is better.

➢ **6-26%** improvements to MoF accuracy compared to SOTA

[1]Kukleva et al. Unsupervised Learning of Action Classes With Continuous Temporal Embedding. CVPR 2019.
[2]Kumar et al. Unsupervised Action Segmentation by Joint Representation Learning and Online Clustering. CVPR 2022
[3]Tran et al. Permutation-Aware Action Segmentation via Unsupervised Frame-to-Segment Alignment. WACV 2024

State-of-the-art Comparison

Australian National University

Metrics: Mean-over-frames accuracy (MoF), segmental F1-score (F1), framewise mean intersection-over-union (mIoU)

	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
CTE[1]	41.8 / 26.4 / -	39.0 / 28.3 / -	30.2 / - / -	35.5 / - / -	47.6 / 44.9 / -
TOT[2]	47.5 / 31.0 / -	40.6 / 30.0 / -	31.8 / - / -	47.4 / 42.8 / -	56.3 / 51.7 / -
UFSA[3]	52.1 / 38.0 / -	49.6 / 32.4 / -	36.7 / 30.4 / -	55.8 / 50.3 / -	65.4 / 63.0 / -
ASOT (Ours)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**

Table: State-of-the-art comparison results. For all evaluation metrics, higher is better.

➢ UFSA and TOT use (standard) optimal transport

[1]Kukleva et al. Unsupervised Learning of Action Classes With Continuous Temporal Embedding. CVPR 2019.
[2]Kumar et al. Unsupervised Action Segmentation by Joint Representation Learning and Online Clustering. CVPR 2022
[3]Tran et al. Permutation-Aware Action Segmentation via Unsupervised Frame-to-Segment Alignment. WACV 2024

State-of-the-art Comparison

Metrics: Mean-over-frames accuracy (MoF), segmental F1-score (F1), framewise mean intersection-over-union (mIoU)

	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
CTE[1]	41.8 / 26.4 / -	39.0 / 28.3 / -	30.2 / - / -	35.5 / - / -	47.6 / 44.9 / -
TOT[2]	47.5 / 31.0 / -	40.6 / 30.0 / -	31.8 / - / -	47.4 / 42.8 / -	56.3 / 51.7 / -
UFSA[3]	52.1 / 38.0 / -	49.6 / 32.4 / -	36.7 / 30.4 / -	55.8 / 50.3 / -	65.4 / 63.0 / -
ASOT (Ours)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**

Table: State-of-the-art comparison results. For all evaluation metrics, higher is better.

- UFSA has a complex, multi-stage transformer architecture

[1]Kukleva et al. Unsupervised Learning of Action Classes With Continuous Temporal Embedding. CVPR 2019.
[2]Kumar et al. Unsupervised Action Segmentation by Joint Representation Learning and Online Clustering. CVPR 2022
[3]Tran et al. Permutation-Aware Action Segmentation via Unsupervised Frame-to-Segment Alignment. WACV 2024

Unsupervised Temporal Action Segmentation: Ablation Study

Australian National University

	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
ASOT (full)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**
Balanced OT	29.7 / 29.3 / 17.8	39.4 / 31.4 / 14.6	39.7 / 39.8 / 25.3	35.7 / 41.8 / 24.9	56.5 / 72.7 / 37.8
No GW	34.4 / 25.9 / 14.4	41.1 / 24.9 / 11.7	29.0 / 22.6 / 14.3	35.1 / 38.6 / 22.5	49.5 / 49.4 / 30.4

Table: Ablation study results, effects are not additive.

Unsupervised Temporal Action Segmentation: Ablation Study

	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
ASOT (full)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**
Balanced OT	29.7 / 29.3 / 17.8	39.4 / 31.4 / 14.6	39.7 / 39.8 / 25.3	35.7 / 41.8 / 24.9	56.5 / 72.7 / 37.8
No GW	34.4 / 25.9 / 14.4	41.1 / 24.9 / 11.7	29.0 / 22.6 / 14.3	35.1 / 38.6 / 22.5	49.5 / 49.4 / 30.4

Table: Ablation study results, effects are not additive.

- Unbalanced transport important with dominant action classes (Breakfast vs Desktop Assembly)

Unsupervised Temporal Action Segmentation: Ablation Study

Australian National University

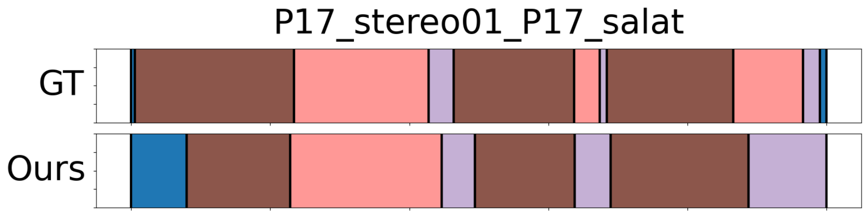
	Breakfast	YouTube Instr.	50 Salads (Mid)	50 Salads (Eval)	Desktop Ass.
	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU	MoF / F1 / mIoU
ASOT (full)	**56.1 / 38.3 / 18.6**	**52.9 / 35.1 / 24.7**	**46.2 / 37.4 / 24.9**	**59.3 / 53.6 / 30.1**	**70.4 / 68.0 / 45.9**
Balanced OT	29.7 / 29.3 / 17.8	39.4 / 31.4 / 14.6	39.7 / 39.8 / 25.3	35.7 / 41.8 / 24.9	56.5 / 72.7 / 37.8
No GW	34.4 / 25.9 / 14.4	41.1 / 24.9 / 11.7	29.0 / 22.6 / 14.3	35.1 / 38.6 / 22.5	49.5 / 49.4 / 30.4

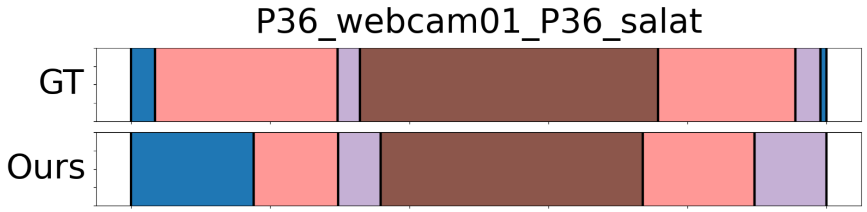
Table: Ablation study results, effects are not additive.

- Structural prior is important across the board

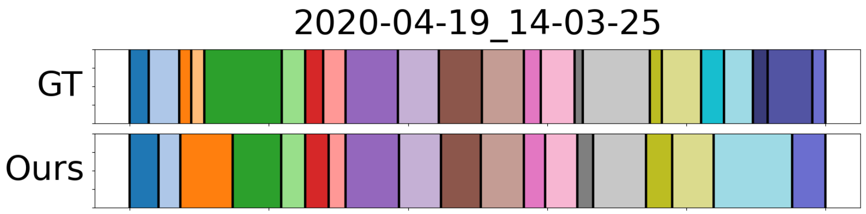
Qualitative Examples

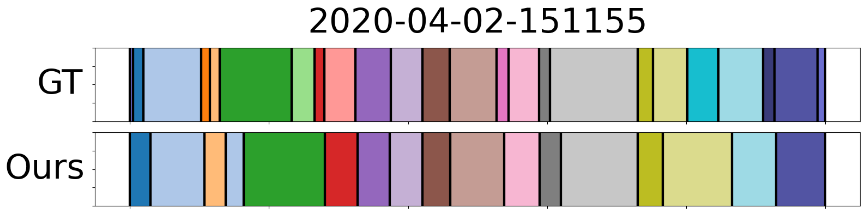
Breakfast





Desktop Assembly





Order variations and repeated actions!

Discussion and Future Work

Broader Impact: Other Settings

Australian National University

Pseudo-labels are ubiquitous

- Semi/weakly-supervised learning (and other variants)
- Unsupervised domain adaptation

Broader Impact: Other Applications

Image segmentation

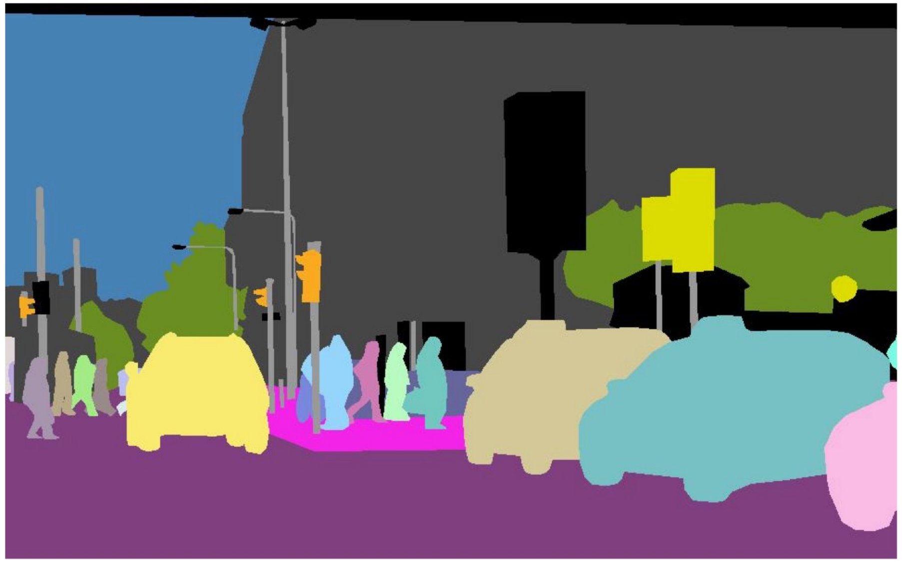


Figure courtesy Kirillov et al., 2019[1]

Monocular depth

Figure courtesy Ranftl et al., 2020[2]

[1]Kirillov et al. Panoptic Segmentation. CVPR 2019.
[2]Ranftl et al. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. PAMI 2020.

Broader Impact: Other Applications

Local feature extractors/matchers

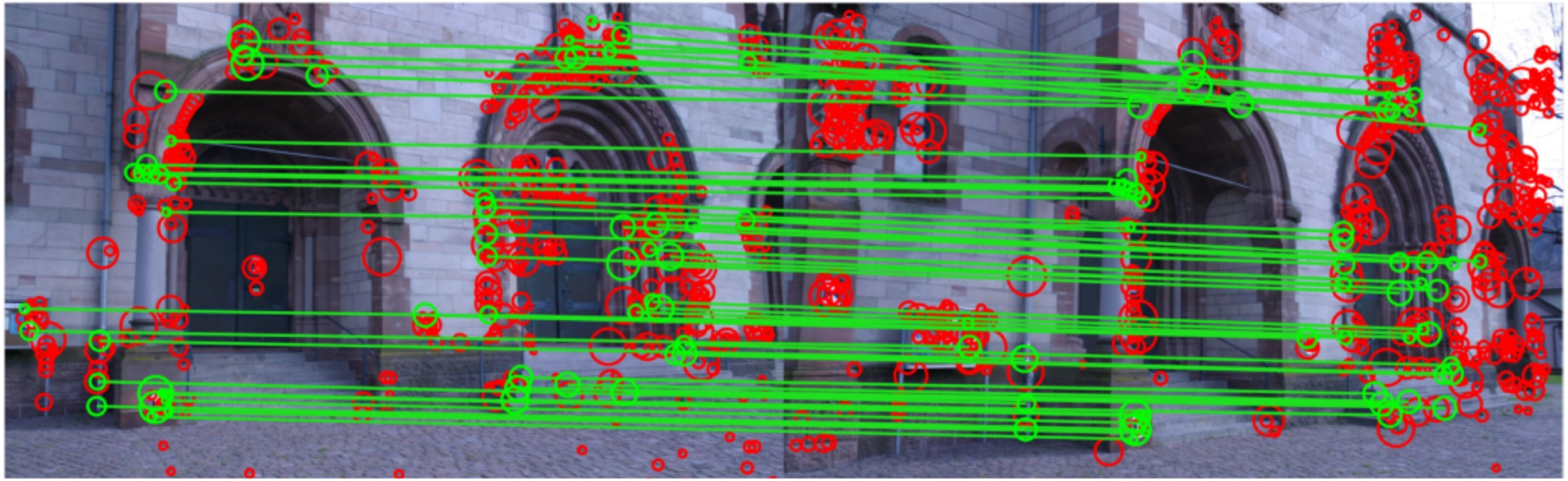


Figure courtesy Yi et al., 2016

Yi et al. LIFT: Learned Invariant Feature Transform. ECCV 2016.

Australian National University

Theoretical Understanding

Recent theoretical developments for self-training (ICLR 2021)

Published as a conference paper at ICLR 2021

THEORETICAL ANALYSIS OF SELF-TRAINING WITH DEEP NETWORKS ON UNLABELED DATA

Colin Wei & Kendrick Shen & Yining Chen & Tengyu Ma
Department of Computer Science
Stanford University
Stanford, CA 94305, USA
{colinwei,kshen6,cynnjjs,tengyuma}@stanford.edu

ABSTRACT

Self-training algorithms, which train a model to fit pseudolabels predicted by another previously-learned model, have been very successful for learning with unlabeled data using neural networks. However, the current theoretical understanding of self-training only applies to linear models. This work provides a unified theo-

How does OT (and ASOT) pseudo-labelling fit into this framework?

Thank You!

Poster Session 4
Arch 4A-E @
5:00 p.m. -6:30 p.m.
Poster #400

Link to paper!

CVPR SEATTLE, WA JUNE 17-21, 2024

seeing machines

Australian National University