

Adaptive Latent Trajectory Anchoring for Action Segmentation Dataset Condensation

Artheme Gauthier-Villars^{1,2*}, Guodong Ding^{1*}, and Angela Yao¹

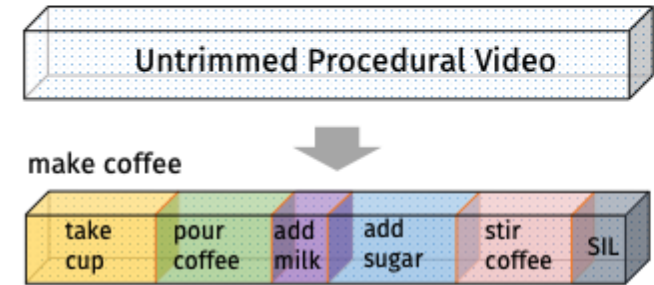
¹National University of Singapore ²ETH Zurich



Task & Motivation

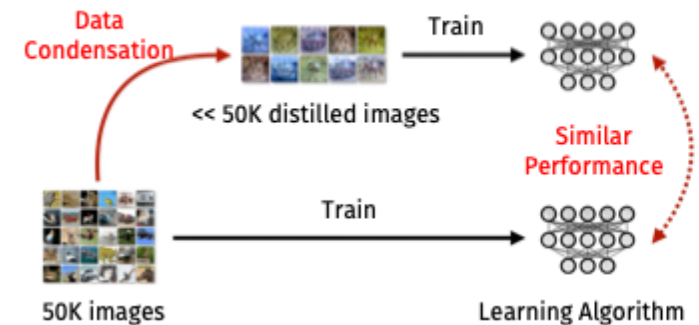
Temporal Action Segmentation

- Temporally segment untrimmed procedural videos and assign frame-wise semantic labels



Dataset Condensation

- Synthesizes a small set that trains models as effectively as the full dataset

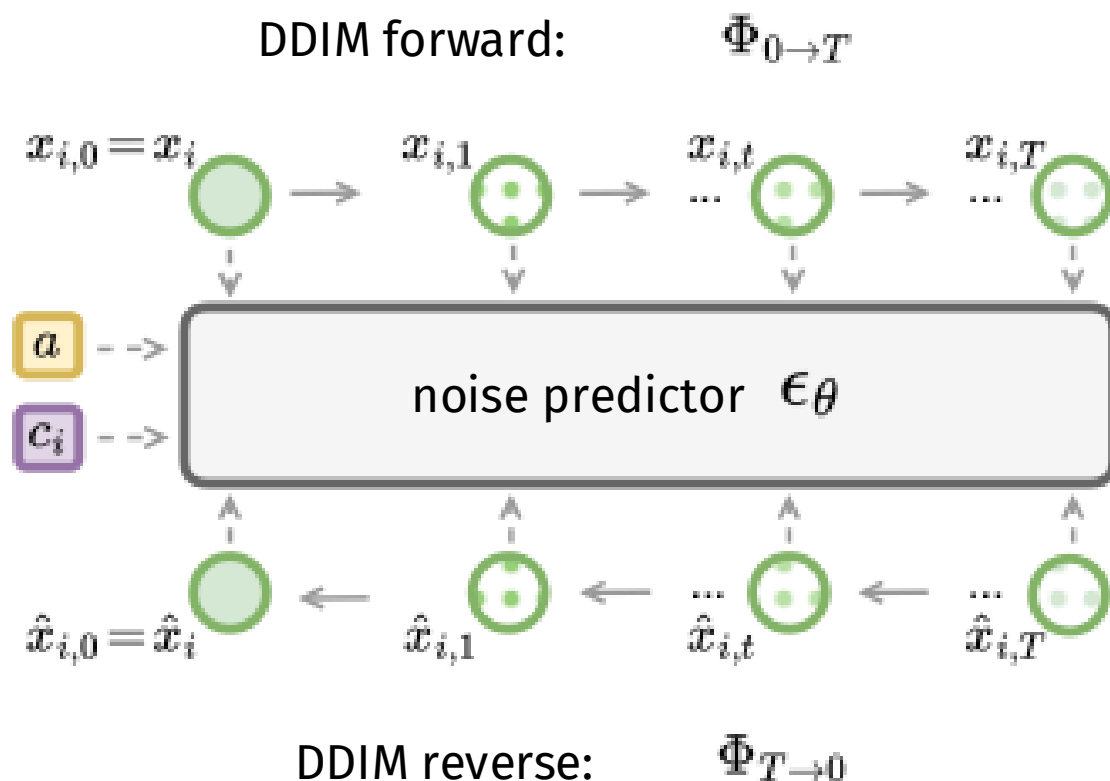


Challenges

- How to *preserve fine-grained temporal dynamics* under aggressive condensation?
- How to *distribute condensation budget* across segments of varying complexity?
- How to *efficiently achieve high-fidelity reconstruction*?

Approach

1) Action Modeling with Diffusion Model



Model actions with Diffusion conditioned on:

- action label a
- temporal coherence c_i
 $c_i = (i - 1)/(\ell - 1)$ and $c_i \in [0, 1]$

Loss function

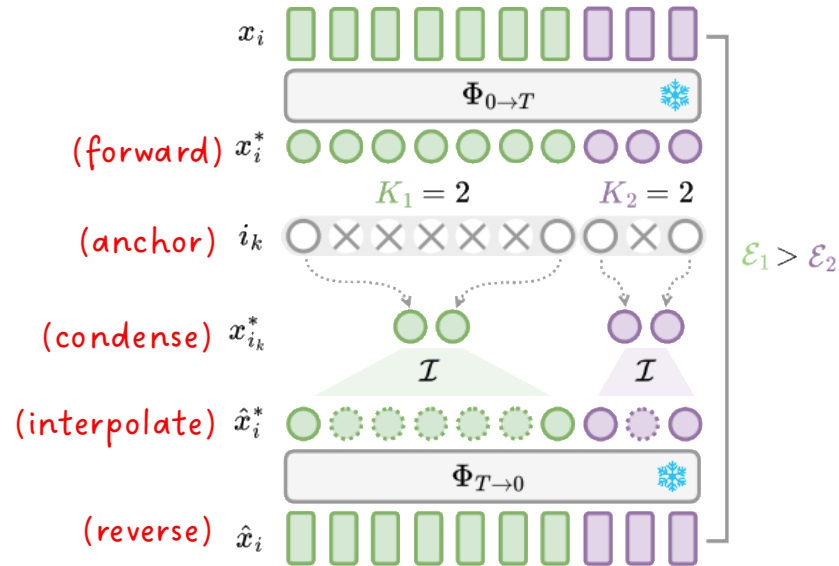
$$\mathcal{L}_{act} = \mathbb{E}_{x_i, a, c_i, t} \left[|\epsilon - \epsilon_\theta(x_{i,t}, a, c_i, t)|^2 \right]$$

Approach

2) Adaptive Latent Trajectory Anchoring

Anchor Initialization:

$$i_k = \lceil 1 + \frac{k-1}{K-1}(\ell-1) \rceil, \quad k = 1, \dots, K$$

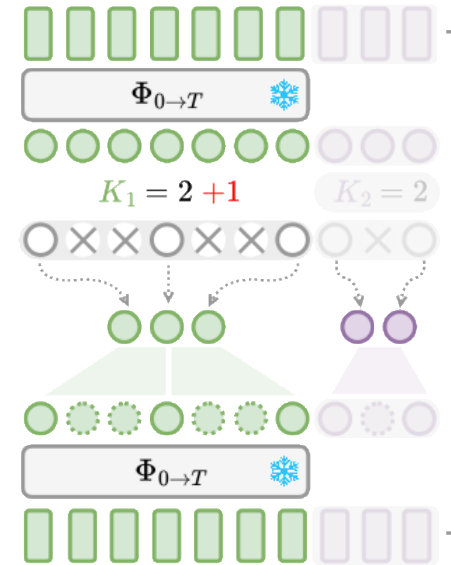


$$x_j^* = \mathcal{I}(x_{i_k}^*, x_{i_{k+1}}^*; \lambda_j) \quad \lambda_j = \frac{c_j - c_{i_k}}{c_{i_{k+1}} - c_{i_k}}$$

Adaptive Allocation:

$$\mathcal{E}_n = \mathbb{E}_{x_i \in s_n} [|x_i - \Phi_{T \rightarrow 0}(x_i^*, a, c_i)|^2]$$

Reconstruction
Difficulty



$$n^* = \arg \max_n \mathcal{E}_n, \quad K_{n^*} \leftarrow K_{n^*} + 1$$

Approach

3) Decoding & Training

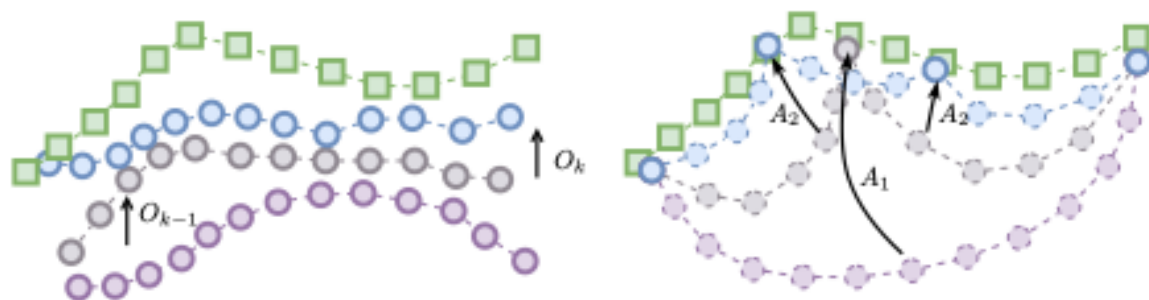
Decoding from latent anchors:

$$\hat{x}_i = \Phi_{T \rightarrow 0}(\mathcal{I}(x_{i_k}^*, x_{i_{k+1}}^*; \lambda_i), a, c_i)$$

Identical TAS loss for learning:

$$\mathcal{L}_{\text{tas}} = \mathcal{L}_{\text{cls}}(\hat{x}, y) + \lambda \mathcal{L}_{\text{sm}}(\hat{x})$$

4) Computational Complexity



GNI

Ours

Method	Complexity	Typical Setting	Operation
GNI	$\mathcal{O}(\text{NS})$	S=10,000	Latent Optimization
Ours [†]	$\mathcal{O}(\text{NT})$	T=50	DDIM Mapping
Ours	$\mathcal{O}(\text{NKT})$	K=8, T=50	DDIM + Anchor Allocation

N: # segments S: # optimization steps T: # diffusion steps K: # of anchors per segment

More efficient condensation than optimization-based GNI

Results

Performance

		GTEA (256MB)				50Salads (4.7GB)				Breakfast (27.4GB)			
	$\hat{K}$	Acc	Edit	F1@10/25/50		Acc	Edit	F1@10/25/50		Acc	Edit	F1@10/25/50	
MSTCN [10]													
Mean	1	69.4	67.5	72.6/68.6/52.8		70.0	46.7	54.2/49.5/40.2		48.0	33.2	29.6/25.5/17.6	
Intp.	2	61.1	69.6	74.6/64.4/46.8		49.9	53.9	54.4/49.6/34.6		33.6	42.9	38.7/31.8/19.3	
Coreset	1	60.3	63.4	65.4/59.2/44.2		69.6	62.5	65.7/63.1/56.6		50.2	41.3	36.7/31.6/22.3	
GNI [5]	8	66.7	71.9	73.9/68.9/45.9		73.5	43.8	49.0/46.6/38.6		37.9	44.5	38.1/33.5/24.7	
GNI [5]	64	66.6	66.9	74.2/69.9/48.8		72.8	44.2	49.3/45.0/40.2		47.8	54.0	46.9/41.9/30.9	
Ours [†]	8	71.2	80.6	80.5/80.5/58.6		72.7	65.1	71.9/68.6/57.9		54.8	67.2	63.8/57.6/45.1	
Ours	8	72.1	79.1	85.9/78.0/ 61.4		70.1	62.3	68.8/65.3/53.5		63.4	65.6	63.8/58.4/46.1	
Original	ℓ	72.7	73.8	79.9/73.7/59.4		74.6	60.0	66.9/64.7/56.2		67.9	67.9	67.7/61.8/49.5	
ASFormer [31]													
Mean	1	70.7	73.3	78.2/76.9/67.3		62.2	40.7	47.9/42.2/33.4		51.8	46.5	44.9/39.8/28.1	
Intp.	2	64.1	72.7	76.5/71.9/54.9		54.8	48.2	55.4/49.5/31.9		43.1	48.3	47.3/40.1/25.6	
Coreset	1	67.2	69.4	73.6/72.3/55.3		66.8	45.3	53.6/48.4/38.2		49.9	51.0	47.1/41.6/30.2	
GNI [5]	8	71.6	76.4	80.8/79.4/63.4		70.5	51.4	61.6/57.6/47.3		62.8	66.3	65.8/59.4/47.2	
GNI [5]	64	70.8	73.5	81.2/77.8/62.1		72.1	49.9	59.7/56.1/46.2		61.1	64.6	63.5/58.4/45.9	
Ours [†]	8	73.7	81.5	84.3/81.5/71.8		75.1	64.9	71.7/69.8/57.6		65.3	71.5	70.1/65.2/48.4	
Ours	8	73.4	80.5	85.2/83.1/73.9		71.2	60.7	69.2/64.9/54.4		68.0	71.3	71.4/66.3/52.9	
Original	ℓ	74.3	77.4	82.9/80.1/73.1		77.1	67.6	77.1/73.1/61.7		70.7	72.0	73.6/69.0/56.1	

- Significant improvements over baselines
- Close performance to the original features

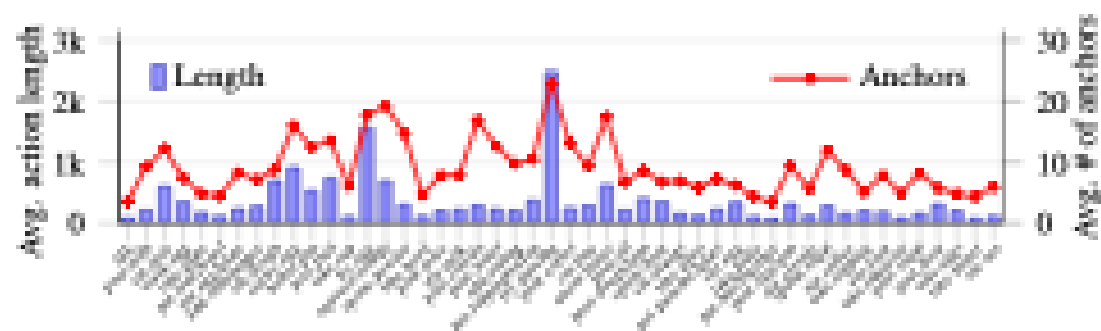
Results

Anchor Budget

$\bar{K}$	ρ	Ours [†]			Ours		
		Acc	Edit	F1@10/25/50	Acc	Edit	F1@10/25/50
2	0.06	64.4	79.5	78.2/68.4/51.1	-	-	- / - / -
4	0.12	68.6	81.9	84.8/76.5/57.6	70.5	78.8	81.4/72.8/59.3
6	0.18	71.2	79.4	85.9/76.3/57.8	72.5	80.0	84.8/76.8/58.7
8	0.24	71.2	80.6	86.5/80.5/58.6	72.1	79.1	85.9/78.0/61.4
10	0.29	73.3	78.4	84.7/78.8/63.5	73.8	77.2	83.3/75.5/61.5
12	0.32	71.5	81.2	82.8/75.5/61.5	71.8	75.5	82.4/76.7/60.2

- Increasing the anchor budget generally improves the performance
- The gains saturate beyond a moderate budget
- Adaptive allocation is more effective than uniform anchoring under smaller budgets

Adaptive Anchor Distribution



- Adaptive allocation is independent of the action frequency
- Longer actions tend to receive more anchors, reflecting more need for finer grained temporal coverage

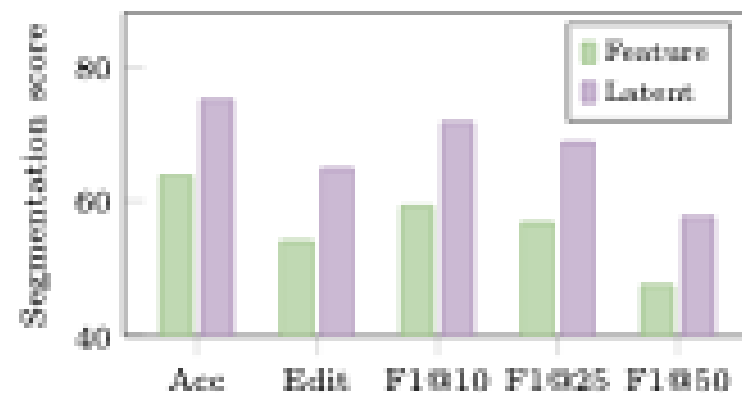
Results

Choice of Action Models

	K	ρ	Acc	Edit	F1@10/25/50
GNI [5]	8	0.03	66.7	71.9	73.9/68.9/45.9
	64	0.24	66.6	66.9	74.2/69.9/48.8
Ours [†]	8	0.24	71.2	80.6	86.5/80.5/58.6
Ours	8	0.24	72.1	79.1	85.9/78.0/61.4

- GNI saturates with more latent codes, indicating an expressiveness bottleneck.

Interpolation Space



- Latent-space interpolation consistently outperform feature-space interpolation.

Takeaways

Fine-grained Temporal Dynamics

Latent trajectory anchoring preserves fine-grained temporal dynamics through interpolation.

Condensation Budget Allocation

Adaptive anchor allocation distributes capacity according to segment reconstruction difficulty.

Efficient High-fidelity Reconstruction

Deterministic DDIM mapping enables high-fidelity reconstruction without iterative optimization.