# Temporal Prototype Alignment

## for Frozen-Feature Dataset Distillation

Rui Ding[1] Jing Hu[1] Mei Chen[1] Xi Wu[1] Hualin Zhou[3] Fan Wu[2] Kehua Guo[2*] Tao Gu[4]

[1] Chengdu University of Information Technology [2] Central South University [3] Macquarie University [4] Shanghai Jiao Tong University * Corresponding author

NeurIPS 2026

Project page

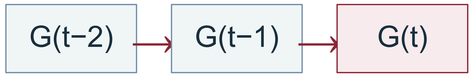
## 1 Why temporal reuse?

LGM stabilizes classifier gradients by averaging ten augmented views per step through frozen encoders. Processing these views together increases memory and I/O costs.

### LGM: evidence across space



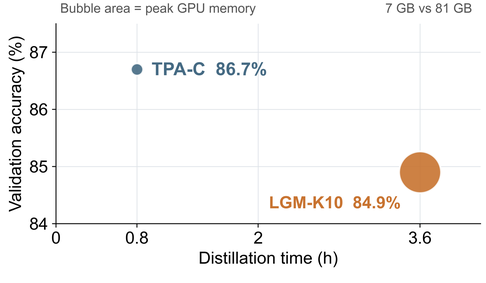

### TPA: evidence across time




### One fresh view + a running prototype

Fixed-size moments reuse gradient history across steps instead of concurrent views.

## Memory and accuracy

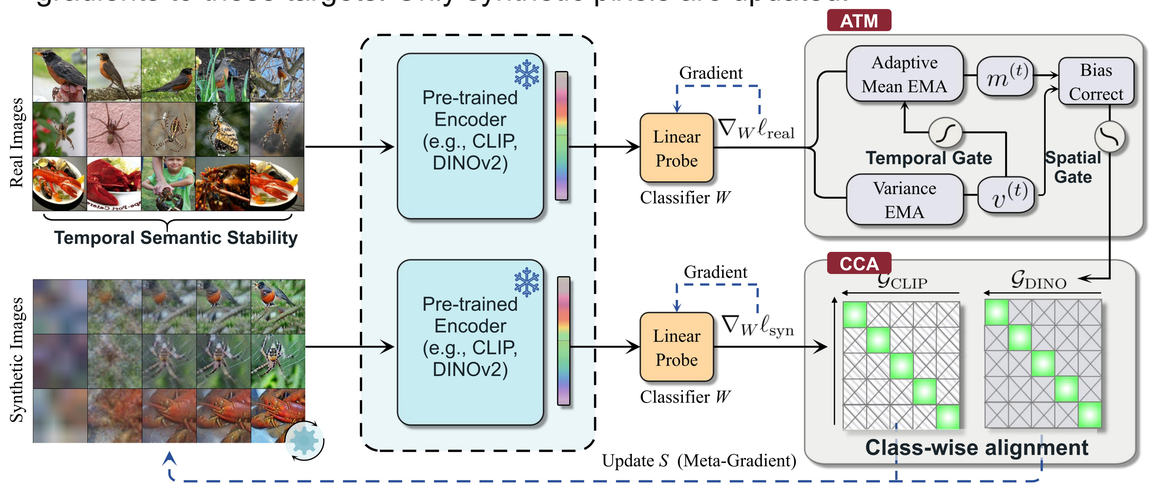



**81 → 7 GB** **3.6 → 0.8 h** **84.9 → 86.7%**

CLIP, ImageNet-100, IPC = 1.
LGM: 4 × RTX 4090. TPA: 1 × RTX 4090.

## 2 How temporal prototypes guide synthesis

Real-data gradients build temporal prototypes. Frozen teachers align synthetic gradients to these targets. Only synthetic pixels are updated.




Upper path: temporal prototypes. Lower path: teacher-local class alignment.

### Adaptive Temporal Momentum (ATM)

One view per step updates first and second moments. A **reliability-dependent gate** controls temporal updates. **Inverse-RMS weighting** builds the prototype, with bias correction for zero initialization. Real and synthetic batches share a freshly sampled classifier head.

### Contrastive Consensus Alignment (CCA)

CCA matches same-class gradients and separates different classes **within each teacher**. CLIP, DINOv2 and MoCo-v3 jointly optimize **one shared image set**. Each teacher retains its own head and prototypes, so feature dimensions can differ.

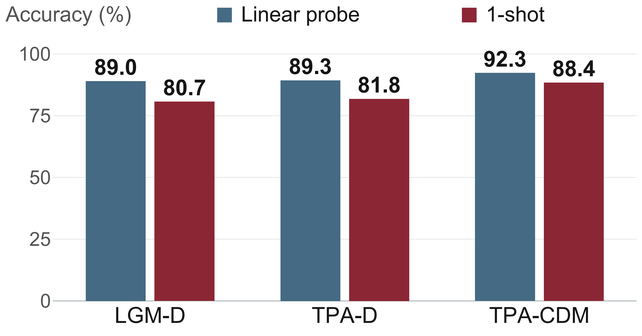
## 3 Cross-encoder transfer

ImageNet-100, one image per class (IPC = 1). One distilled set is evaluated on nine unseen encoders with linear probing and training-free 1-shot inference.

**+3.3 pp** Linear probing

**+7.7 pp** 1-shot inference

Gain over LGM-D · mean of nine unseen encoders




TPA-CDM reaches **92.3%** linear-probe and **88.4%** 1-shot accuracy across nine unseen encoders. Multi-teacher synthesis improves both transfer metrics over LGM-D.

### What does the distilled set look like?



Selected images from the paper's TPA-CDM set.