

■ ICML 2026 · SEOUL, SOUTH KOREA

CHESS

Chebyshev Spectral Synthesis for Trajectory Condensation

A function-first paradigm — condensing continuous-time sensor trajectories on a low-rank, piecewise-Chebyshev manifold, rather than on discrete pixels.



Ruituo Wu · Hongyu Zhang · Qiang Wang · Jiawei Du · Wei Cui · Ce Zhu · Bing Li
UESTC · Shanghai Jiao Tong University · A*STAR (CFAR / IHPC / I²R)

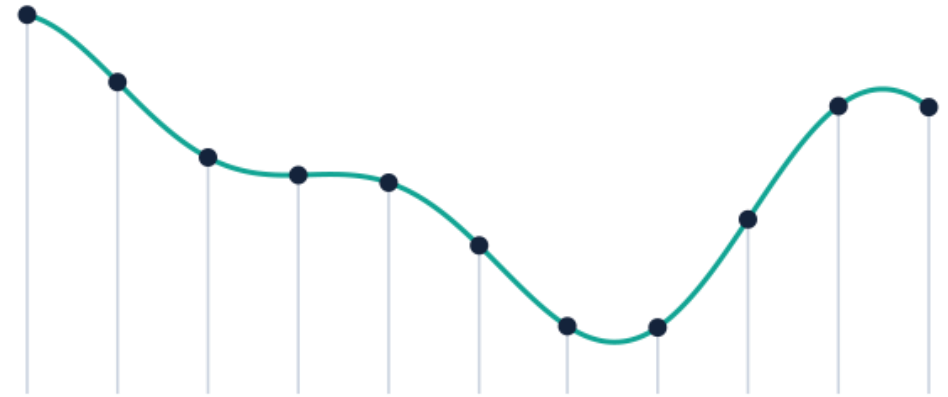


■ MOTIVATION

Learning from Continuous-Time Trajectories

Sensor data are discrete samples of underlying continuous physical processes.

- WiFi CSI, wearable IMUs, and biomedical signals encode dynamics that frame-based views cannot capture.
- Deep models need large, densely-sampled datasets — costly to collect, store, and transmit on edge devices.
- The same process is observed under different sampling rates, sensor layouts, and downstream architectures.

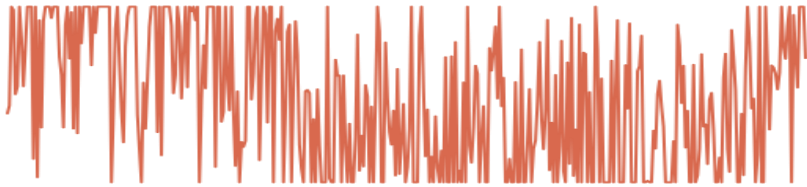


continuous trajectory (teal) vs discrete observations (dots)

The challenge. Condense trajectories compactly while staying reusable across architectures and unseen sampling rates.

Discrete Optimization Breaks Continuous Signals

Discrete pixel-wise optimization



high-frequency, non-physical spikes

CHES – synthesis on a function manifold



smooth, physically meaningful

Sample-wise parameterization drifts from a structured function space into arbitrary high-dimensional Euclidean space. Ill-posedness arises along three axes:

01 Temporal

Unconstrained inter-sample behavior violates physical continuity, producing degenerate high-frequency oscillations.

02 Spatial

Independent channel modeling ignores cross-channel coupling, inducing architecture-specific artifacts.

03 Sampling

Binding the representation to a fixed grid prevents reuse under unseen resolutions.

What Effective Trajectory Condensation Needs

01

Compactness

Extreme data reduction while preserving task-relevant information.

02

Cross-architecture reusability

Compressed data generalizes across different downstream models.

03

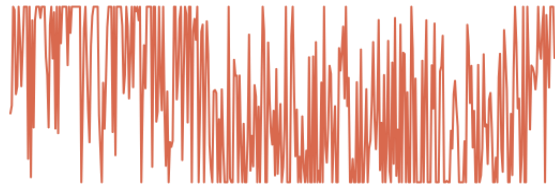
Resolution-agnostic generalization

Reuse under unseen temporal sampling rates, with no redistillation.

These goals favor representations as structured functions — low intrinsic complexity, decoupled from any architecture or sampling grid.

A Function-First Paradigm Shift

OPTIMIZE DISCRETE POINTS



each time step is an independent variable



MODEL CONTINUOUS TRAJECTORIES



synthesis lives on a structured function manifold

Anchor synthesis in a physically meaningful manifold — **low-rank in space, piecewise-Chebyshev in time.**

CHES: Three Tightly-Coupled Components

01

Low-Rank Spatial Modeling

Restrict multivariate trajectories to a shared low-dimensional subspace, capturing cross-channel coupling.

02

Piecewise Chebyshev Temporal Fitting

Parameterize trajectories with continuous polynomial bases — explicit smoothness, suppressed spikes.

03

Continuous Trajectory Resampling

Decouple from any fixed grid; resample analytically at arbitrary resolutions, with no redistillation.

Real datasets



Low-rank + Chebyshev parameters



Differentiable reconstruction



Distribution matching (frozen expert)

■ COMPONENT 1

Low-Rank Spatial Modeling

Channels are not independent — they are correlated projections of the same latent process.

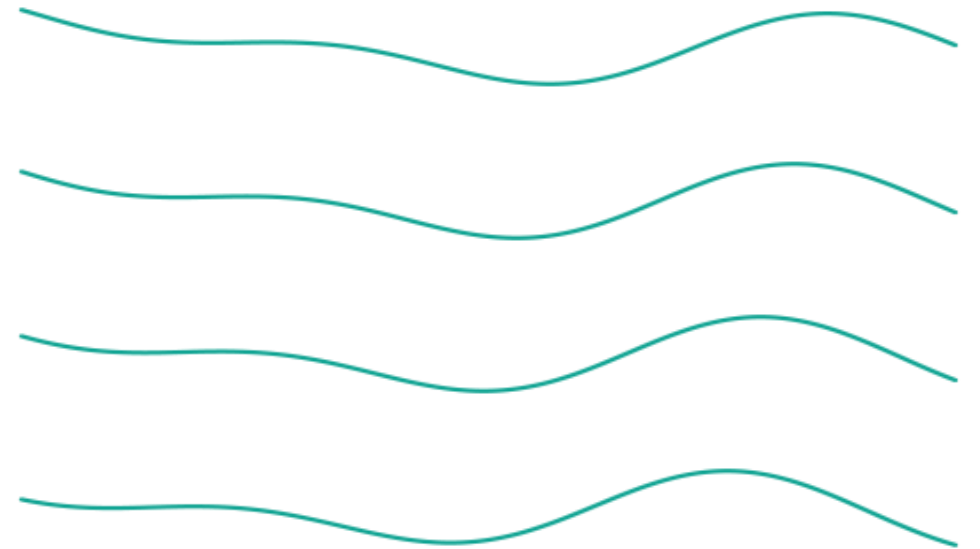
- The signal matrix has strong linear dependence and lies near a low-dimensional subspace.
- Constraining the rank R enforces coherent channel activations — preventing uncorrelated noise.

EXPLICIT LOW-RANK FACTORIZATION

$$\mathbf{X}_{\text{syn}} = \mathbf{U}_{\text{time}} \mathbf{\Sigma} \mathbf{V}_{\text{space}}^T$$

$U_{\text{time}} \in \mathbb{R}^{N \times R}$ (temporal) · $V_{\text{space}} \in \mathbb{R}^{S \times R}$ (spatial) · Σ : singular values

30 sensor channels — highly synchronized



Effective DoF is low → **MeR uses R = 5 for 30 channels.**

Piecewise Chebyshev Temporal Fitting

Stone-Weierstrass theorem

Any continuous trajectory can be uniformly approximated by polynomials — a theoretically guaranteed hypothesis space.

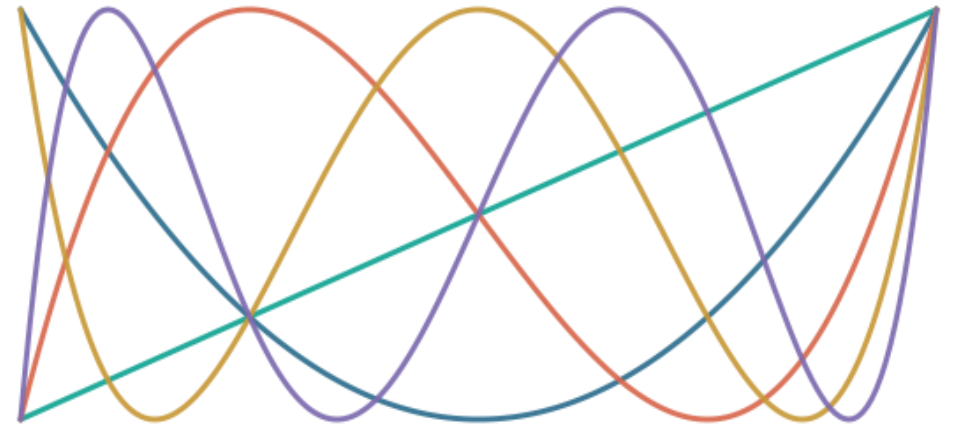
Chebyshev minimax property

The flattest basis: smallest maximum amplitude → implicitly minimizes derivatives, suppressing high-frequency spikes and Runge oscillations.

$$U_{\text{time}} = [C_1 B \cdots C_M B]_{\mathcal{T}}$$

N split into M segments of length L; learn coefficients C_m on a fixed Chebyshev basis B.

Chebyshev polynomials $T_1 \dots T_5$ on $[-1, 1]$



A structural denoiser within each segment — expressive, yet smooth.

Provable Smoothness & Stability

PROPOSITION 1 · DERIVATIVE CONTROL

$$\|u'_r(t)\| \leq (2 / \Delta T) \cdot D^2 \cdot \|u_r(t)\|$$

A hard constraint on the signal's rate of change, scaling with D^2 inside each segment — via the Markov Brothers' Inequality.

Lemma 1 — preserved under linear combinations.

Smoothness extends from the basis to the whole synthesized sample X_{syn} .

Piecewise is essential.

A single global polynomial makes degree D scale with length — the D^2 bound explodes and loses its power.

Empirically verified. On UCI-HAR the NCFM baseline's derivatives reach $\approx \pm 1.0$ — nearly 50× the natural sensor noise ($\approx \pm 0.02$). CHESS yields sharp, zero-centered derivatives, echoing the Minimum-Jerk Principle of motor control.

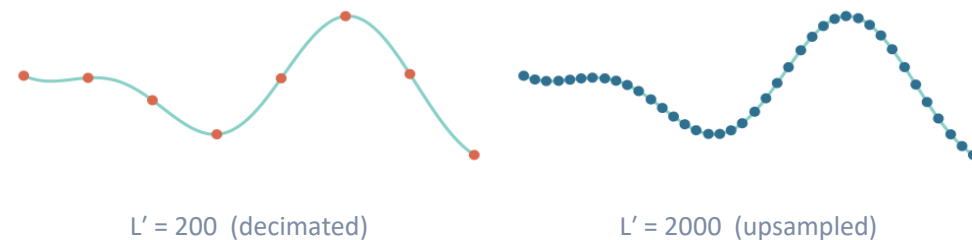
Zero-Shot Resolution Adaptation

Conventional condensation produces fixed tensors — adapting to a new sampling rate needs lossy interpolation.

CHES stores continuous trajectory coefficients C_m . To target a new length L' , re-evaluate the Chebyshev basis B' and reconstruct analytically — no redistillation.

$$U'_{\text{time}} = [C_1 B' \cdots C_M B']_{\text{T}}$$

Same trajectory, any grid

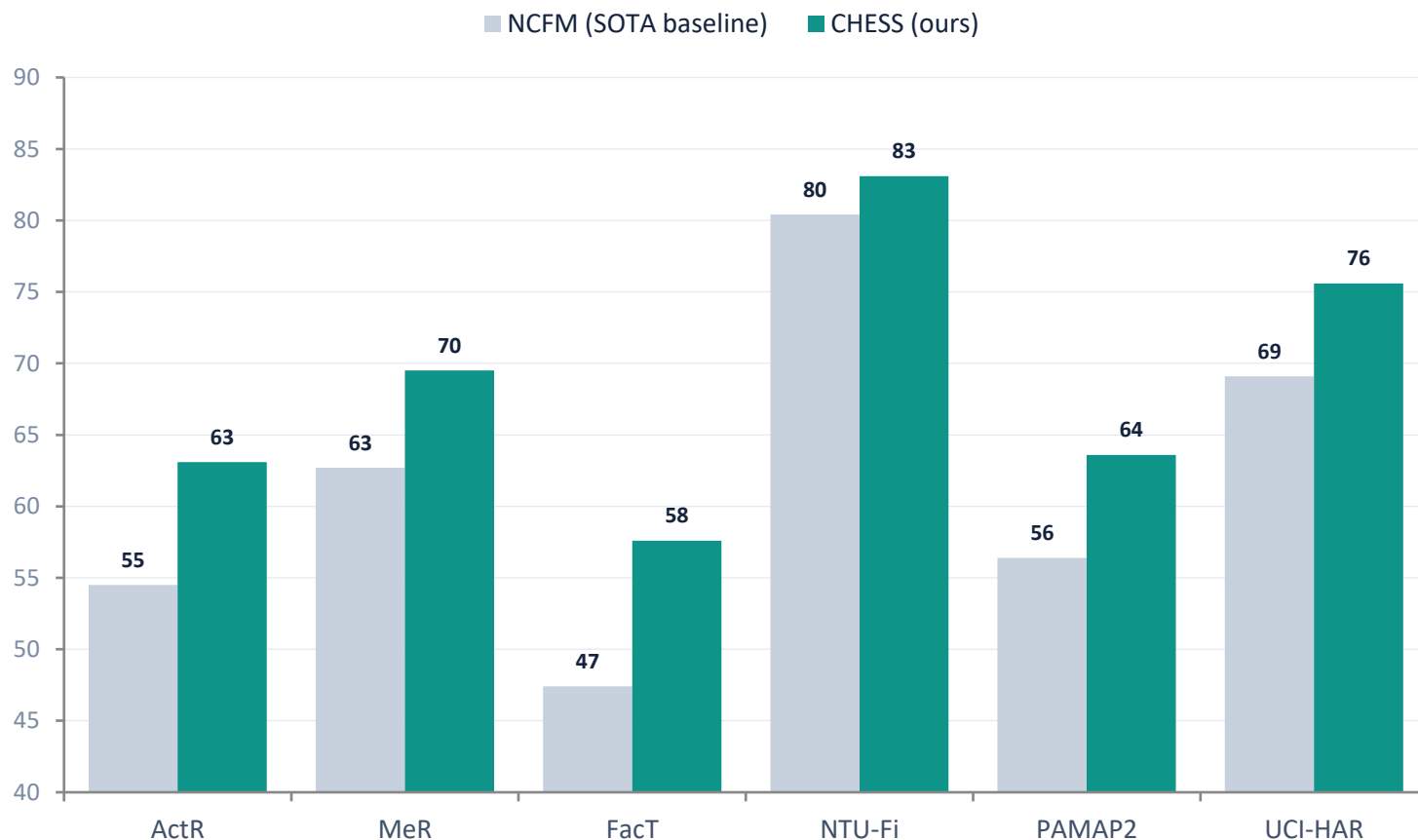


“Distill once, deploy anywhere.”

MeR upsampling to $L = 2000$: **CHES 75.8%** vs **DANCE 62.3%** — a +12.5% gain over interpolated baselines.

RESULTS

State-of-the-Art Across Six Sensor Testbeds



133.3×

compression per sample on high-rate signals

+6.4%

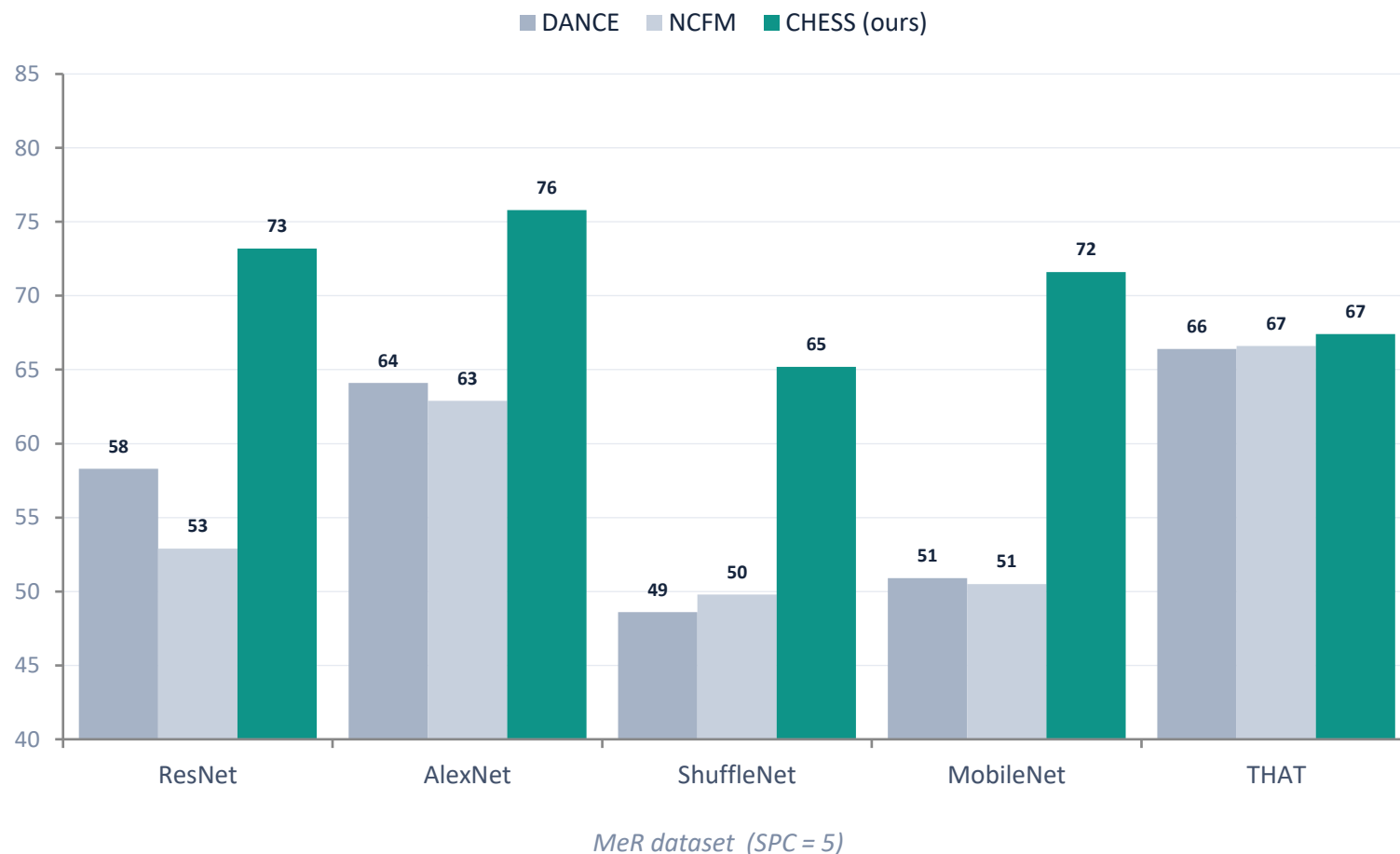
average accuracy over the strongest baseline

+11.2%

on ActR (SPC = 5) vs the runner-up

RESULTS

Robust Cross-Architecture Generalization



+16.6%

over DANCE on ShuffleNet — where baselines collapse below 50%.

67.4%

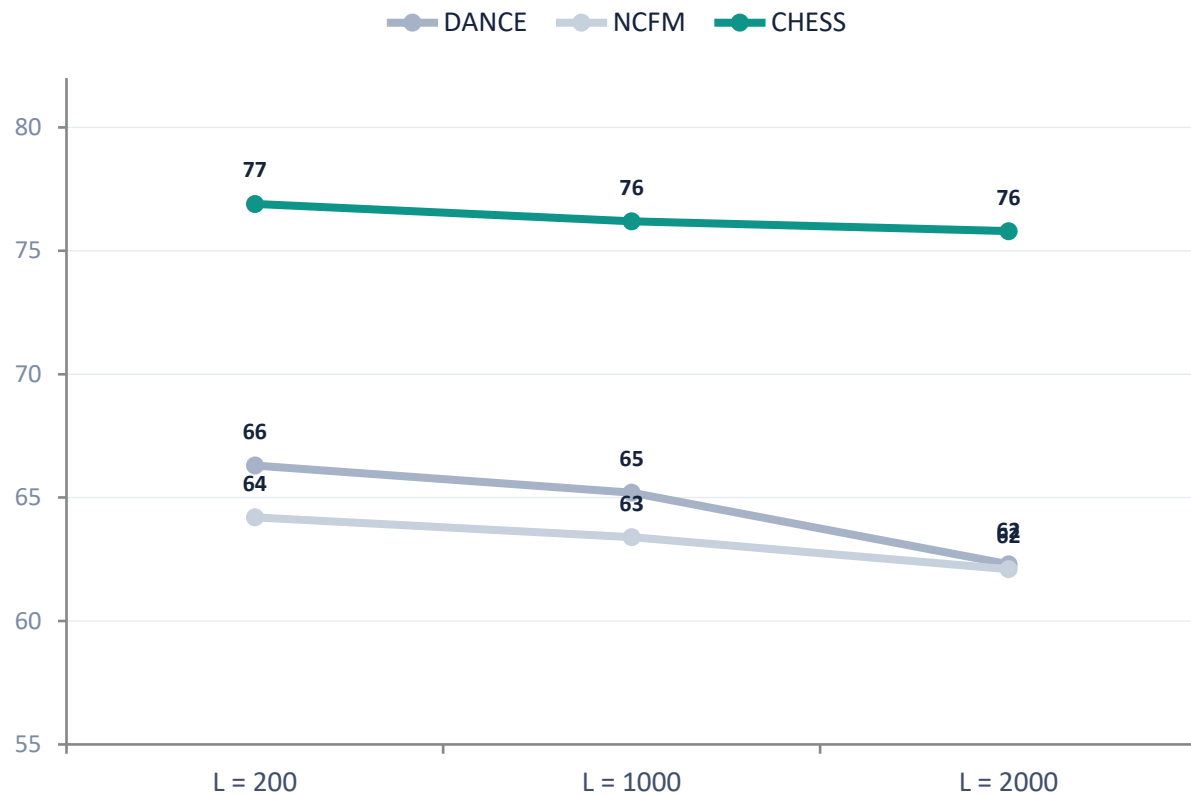
highest accuracy on the Transformer backbone (THAT).

Ablations

w/o Chebyshev: MobileNet drops > 9%. **w/o low-rank:** consistent degradation.

Resolution Transfer & What Matters

Zero-shot resolution adaptation – MeR (distill @ L = 1000)



Baselines degrade with interpolation; CHESS stays flat across scales.

Initialization is decisive

Random init fails (~14% = chance). A least-squares projection onto the manifold gives a physically-grounded warm start.

Orthogonality matters

Chebyshev / Legendre / Hermite decorrelate coefficients → > 73% on ResNet. Plain monomials collapse from collinearity.

Low degree wins

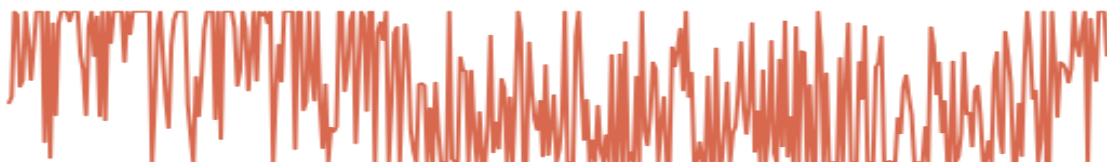
Optimal degree $D \in [2, 3]$; higher degrees invite the Runge phenomenon. Gains come from structure, not capacity.

■ WHY IT WORKS

Smooth, Faithful – and Efficient

Distilled samples (MeR, first channels)

Baselines (DC / MTT / NCFM)



CHESS (ours)



Faithful coverage. In feature (t-SNE) space, distilled samples sit within high-density regions of the real data — preserving discriminative structure.

No efficiency penalty

Despite on-the-fly continuous decoding, compact parameters speed up training.

-5.0%

training time on MeR

-2.3%

training time on UCI-HAR

+0.3-2.2%

memory overhead
(negligible)

CHESS in One Slide

01

Function-first, not pixel-first

Optimize continuous trajectories on a physically meaningful manifold instead of discrete samples.

02

Low-rank space + Chebyshev time

Cross-channel coherence and piecewise polynomial smoothness — with provable derivative bounds.

03

Distill once, deploy anywhere

Analytical resampling enables zero-shot adaptation to unseen resolutions; up to 133× compression.

