

Less is More

Efficient Time Series Dataset Condensation via Two-fold Modal Matching

Hao Miao, Ziqiao Lui, Yan Zhao, Chenjuan Guo, Bin Yang, Kai Zheng, Christian S. Jensen

February 2025

Proceedings of the VLDB Endowment, Volume 18, Issue 2

Presented by **Andreas Gottschalk Krath**

1. Introduction



1.1 Motivation

IoT

- More devices
- Time series

1.1 Motivation

IoT

- More devices
- Time series

Data Volume

- Too much data
- Edge computing
- What can we do?
 - Less data \rightarrow Same value

1.1 Motivation

IoT

- More devices
- Time series

Data Volume

- Too much data
- Edge computing
- What can we do?
 - Less data \rightarrow Same value

Existing Methods

- Coreset construction
- Issues
 - Not optimal
 - Downstream tasks

1.2 Problem

Challenge 1

- Effectiveness
 - Bi-level optimization
 - Capture relevant information
- Generalization
 - Different networks

1.2 Problem

Challenge 1

- Effectiveness
 - Bi-level optimization
 - Capture relevant information
- Generalization
 - Different networks

Solution

- Curriculum Training Trajectory Matching (CT²D)
- Expert trajectories
 - Based on original dataset

1.2 Problem

Challenge 2

- Complex Temporal Dependencies
 - Existing methods focus on image
 - Time series requires temporal analysis

1.2 Problem

Challenge 2

- Complex Temporal Dependencies
 - Existing methods focus on image
 - Time series requires temporal analysis

Solution

- Time Series Feature Extraction (TSFE)
 - Stacked TSOoperators
- Decomposition-Driven Frequency Matching (DDFM)
 - Analyze intermediate TSFE
 - Match to original data

1.3 Contributions

Dataset Condensation

- Novel time series dataset condensation
- TSFE
- DDFM
- CT²M

2. Methodology

2.1 Overview

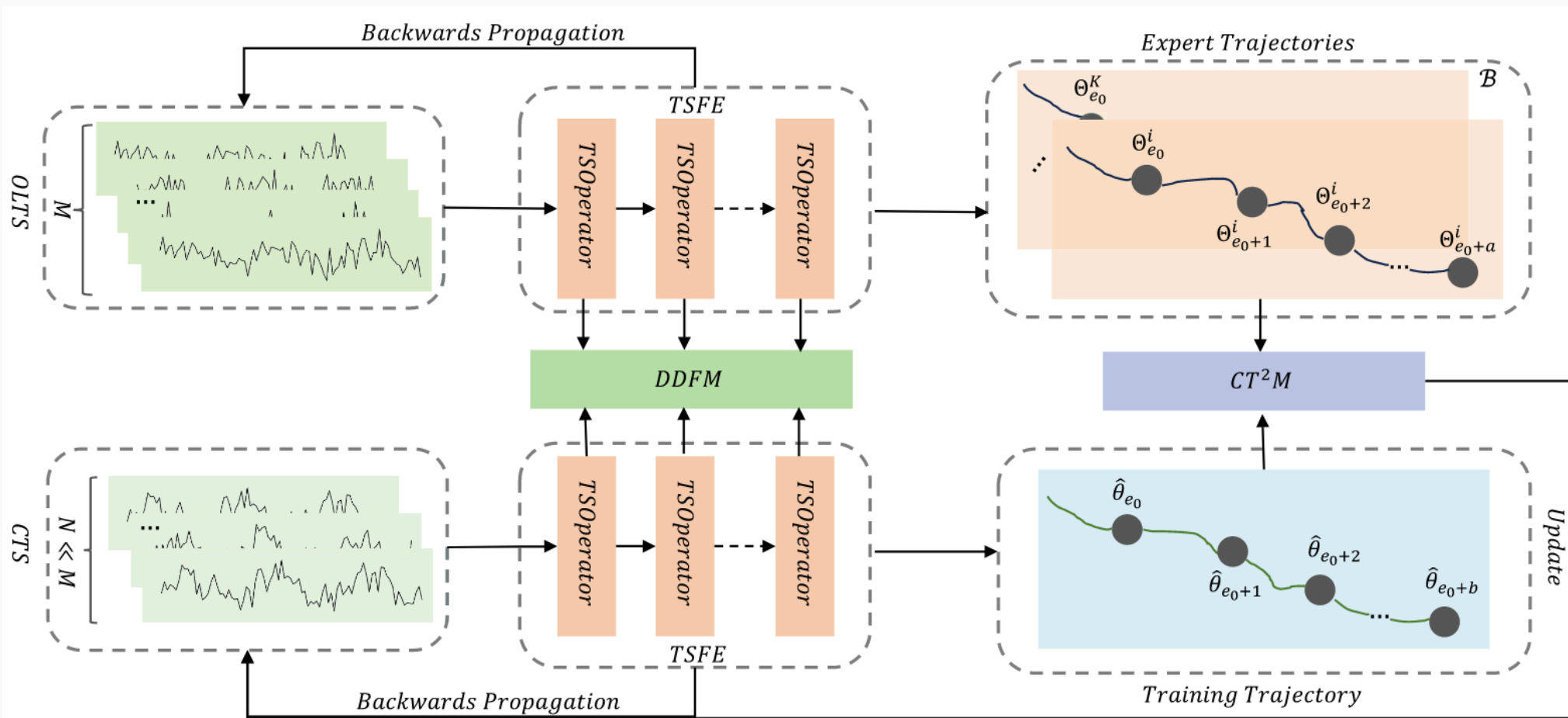


Figure 2: TimeDC Framework Overview

2.2 Time Series Feature Extraction

Architecture

- Channel independent mechanism
 - Split variables
- Patching
 - Combine data
- Stacked TSOOperators
 - Self-attention
 - Fully connected
 - Normalization
- Concatenate to combine variables

Algorithm 1: Time Series Feature Extraction

Input: a batch of time series: $T_{input} \in \mathcal{B} \times n \times C$; the number of TSOOperators: N_{op} .

Output: extracted features: h .

```
1  $T_{input}^1, \dots, T_{input}^c, \dots, T_{input}^C \leftarrow$  Separate  $T_{input}$  into  $C$ 
   univariate channel-independent time series with
   Equation 3;
2 Split each univariate time series into patches;
3 for  $1 \leq c \leq C$  do
4    $h_c^0 \leftarrow T_{input}^c$ ;
5   for  $0 < j < N_{op}$  do
6     Feature extraction with Equation 4;
7      $h_c^j \leftarrow \text{TSOperator}(h_c^{j-1})$ ;
8    $h_c \leftarrow h_c^{N_{op}-1}$ ;
9  $h \leftarrow$  Concatenate  $\{h_c\}_{c=1}^C$  along the channel;
10 return  $h$ .
```

$$h^j = \text{TSOperator}(h^{j-1}) = \text{Norm}(\text{FC}(\text{MultiHead}(h^{j-1}))), \quad (4)$$

2.3 Decomposition-Driven Frequency Matching

Architecture

- Works on intermediate TSFE output
- Splits representation
 - Trends (long-term)
 - Seasonality (the rest)
- Cosine similarity to match
- Produces a “data-related” loss term
 - Relevant for back propagation

$$h_{TRE}^j = AvgPool(Padding(h^j)), h_{SEA}^j = h^j - h_{TRE}^j, \quad (6)$$

$$\begin{aligned} \cos(h_{TRE_{\mathcal{T}}}^j, h_{TRE_S}^j) &= \frac{h_{TRE_{\mathcal{T}}}^j}{\|h_{TRE_{\mathcal{T}}}^j\|_2} \cdot \frac{h_{TRE_S}^j}{\|h_{TRE_S}^j\|_2} \\ \cos(h_{SEA_{\mathcal{T}}}^j, h_{SEA_S}^j) &= \frac{h_{SEA_{\mathcal{T}}}^j}{\|h_{SEA_{\mathcal{T}}}^j\|_2} \cdot \frac{h_{SEA_S}^j}{\|h_{SEA_S}^j\|_2}, \end{aligned} \quad (8)$$

2.4 Curriculum Training Trajectory Matching

Architecture

- Expert buffer and current model
 - Current model is based on expert
- Train current model a epochs
- Compare cosine for parameter movement
- Rank based on similarity
 - Easy \rightarrow Hard
- Compute trajectory matching loss

Algorithm 2: Curriculum Training Trajectory Query and Matching

Input: A buffer \mathcal{B} with a set of trajectories pre-trained on the original TS dataset \mathcal{T} parameterized by $\{\Theta_{\mathcal{T}}^k\}_{k=1}^K$; current model parameters $\tilde{\theta}^S$ on S .

Output: Trajectory matching loss L_{tmm} .

```
1 Distance list  $DT \leftarrow []$ ;  
2 Pre-update  $\tilde{\theta}^S$  for  $a$ -steps with Equation 13;  
3  $\tilde{\theta}_{e_0+a} \leftarrow \tilde{\theta}_{e_0} - \sum_{s=1}^a (\alpha \nabla \mathcal{L}(f_{\tilde{\theta}^S}, S))$ ;  
4 for  $\Theta_{\mathcal{T}}^k \in \{\Theta_{\mathcal{T}}^k\}_{k=1}^K$  do  
5   Compute the distance  $dis_k$  between  $\tilde{\theta}_{e_0}^a$  and  $\Theta_{\mathcal{T}}^k$  with  
   Equation 15;  
6    $dis_k \leftarrow -D(\tilde{\theta}_{e=e_0}^a, \theta_{e=e_0}^k)$ ;  
7    $DT \leftarrow (k, dis_k)$ ;  
8 Rank  $DT$  in a descending order;  
9  $\beta \leftarrow 0$   
10 while  $\beta < K$  do  
11    $k \leftarrow DT[\beta][0]$   
12    $L_{tmm} \leftarrow$  Sample trajectory  $\Theta_{\mathcal{T}}^k$  and match the training  
   trajectory according to Equation 11;  
13    $\beta \leftarrow \beta + 1$ ;  
14 return  $L_{tmm}$ 
```

2.5 Loss

Loss

- Task specific loss
 - How good is the data?
- Frequency related loss
 - How similar is the data?
- Trajectory matching loss
 - How does data affect training?

$$L_{\text{all}} = \mathcal{L} + L_{\text{Fre}} + L_{\text{tmm}} \quad (16)$$

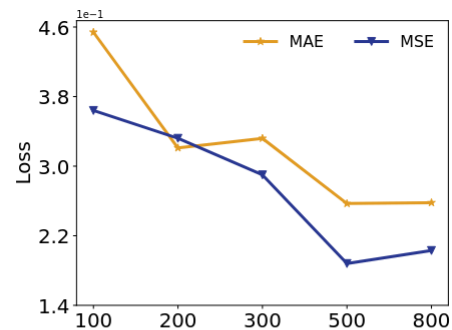
3. Experiments

3.1 Dataset Condensation

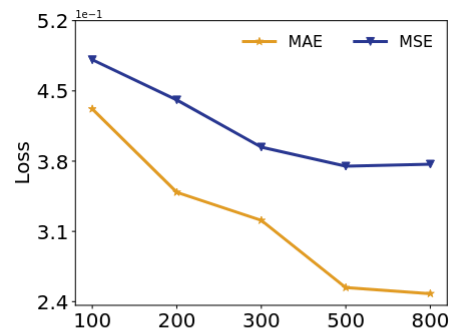
Table 1: Overall Performance Comparison on Seven Datasets

Baseline		Random		K-Center		Herding		DC		MTT		TimeDC		Whole Dataset	
Dataset	<i>PL</i>	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Weather	96	0.731	1.256	0.452	0.687	0.478	0.677	0.361	0.514	<u>0.295</u>	<u>0.244</u>	0.257	0.188	0.239	0.182
	192	0.786	1.302	0.487	0.723	0.512	0.688	0.413	0.527	<u>0.344</u>	<u>0.301</u>	0.285	0.247	0.261	0.195
	336	0.794	1.311	0.524	0.756	0.554	0.712	0.444	0.567	<u>0.368</u>	<u>0.328</u>	0.330	0.287	0.282	0.241
Traffic	96	0.675	1.125	0.503	0.576	0.483	0.554	0.375	0.603	<u>0.279</u>	<u>0.403</u>	0.254	0.375	0.247	0.337
	192	0.712	1.144	0.514	0.604	0.517	0.606	0.432	0.633	<u>0.336</u>	<u>0.442</u>	0.297	0.405	0.265	0.338
	336	0.729	1.117	0.523	0.611	0.553	0.654	0.449	0.676	<u>0.355</u>	<u>0.471</u>	0.312	0.423	0.297	0.360
Electricity	96	0.421	0.669	0.448	0.583	0.501	0.592	0.376	0.513	<u>0.296</u>	<u>0.283</u>	0.274	0.267	0.252	0.268
	192	0.450	0.743	0.476	0.601	0.534	0.628	0.419	0.532	<u>0.315</u>	<u>0.337</u>	0.285	0.294	0.239	0.255
	336	0.491	0.853	0.506	0.622	0.569	0.477	0.436	0.544	<u>0.339</u>	<u>0.356</u>	0.304	0.322	0.271	0.285
ETTh1	96	0.523	0.745	0.554	0.698	0.536	0.656	0.503	<u>0.442</u>	<u>0.456</u>	0.464	0.413	0.401	0.354	0.386
	192	0.557	0.786	0.578	0.722	0.589	0.698	0.552	0.508	<u>0.504</u>	<u>0.471</u>	0.436	0.428	0.362	0.355
	336	0.588	0.802	0.604	0.745	0.603	0.723	0.556	0.513	<u>0.498</u>	<u>0.464</u>	0.447	0.431	0.409	0.387
ETTh2	96	0.487	0.655	0.589	0.711	0.521	0.589	0.463	0.524	<u>0.388</u>	<u>0.342</u>	0.368	0.271	0.324	0.255
	192	0.509	0.673	0.605	0.732	0.553	0.621	0.488	0.536	<u>0.416</u>	<u>0.384</u>	0.389	0.302	0.332	0.257
	336	0.524	0.689	0.623	0.744	0.564	0.640	0.505	0.540	<u>0.435</u>	<u>0.455</u>	0.411	0.334	0.376	0.296
ETTm1	96	0.743	1.124	0.525	0.492	0.607	0.554	0.603	0.665	<u>0.512</u>	<u>0.453</u>	0.503	0.442	0.453	0.403
	192	0.764	1.245	0.566	0.510	0.628	0.571	0.597	0.647	<u>0.563</u>	<u>0.501</u>	0.512	0.465	0.464	0.432
	336	0.801	1.128	0.571	0.523	0.644	0.582	0.624	0.668	<u>0.552</u>	<u>0.488</u>	0.500	0.483	0.477	0.455
ETTm2	96	0.664	0.795	0.486	0.623	0.524	0.558	0.472	0.535	<u>0.376</u>	<u>0.421</u>	0.354	0.391	0.347	0.381
	192	0.687	0.804	0.512	0.643	0.549	0.583	0.488	0.567	<u>0.453</u>	<u>0.479</u>	0.401	0.421	0.358	0.403
	336	0.702	0.823	0.558	0.661	0.598	0.624	0.493	0.556	<u>0.473</u>	<u>0.523</u>	0.453	0.474	0.406	0.435

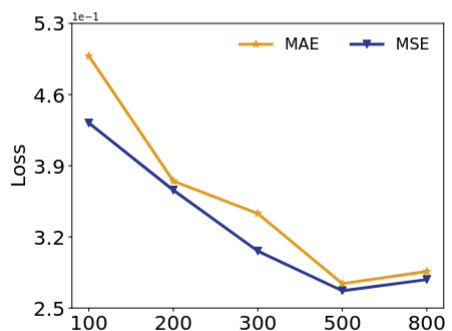
3.2 Size of Condensed Dataset



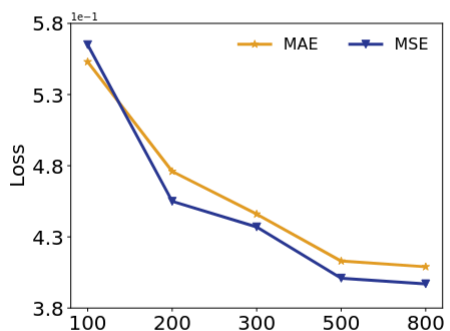
(a) Weather



(b) Traffic



(c) Electricity



(d) ETTh1

Figure 5: Effect of the Size of Condensed TS Dataset on Four Datasets ($PL = 96$)

3.3 Ablation Study

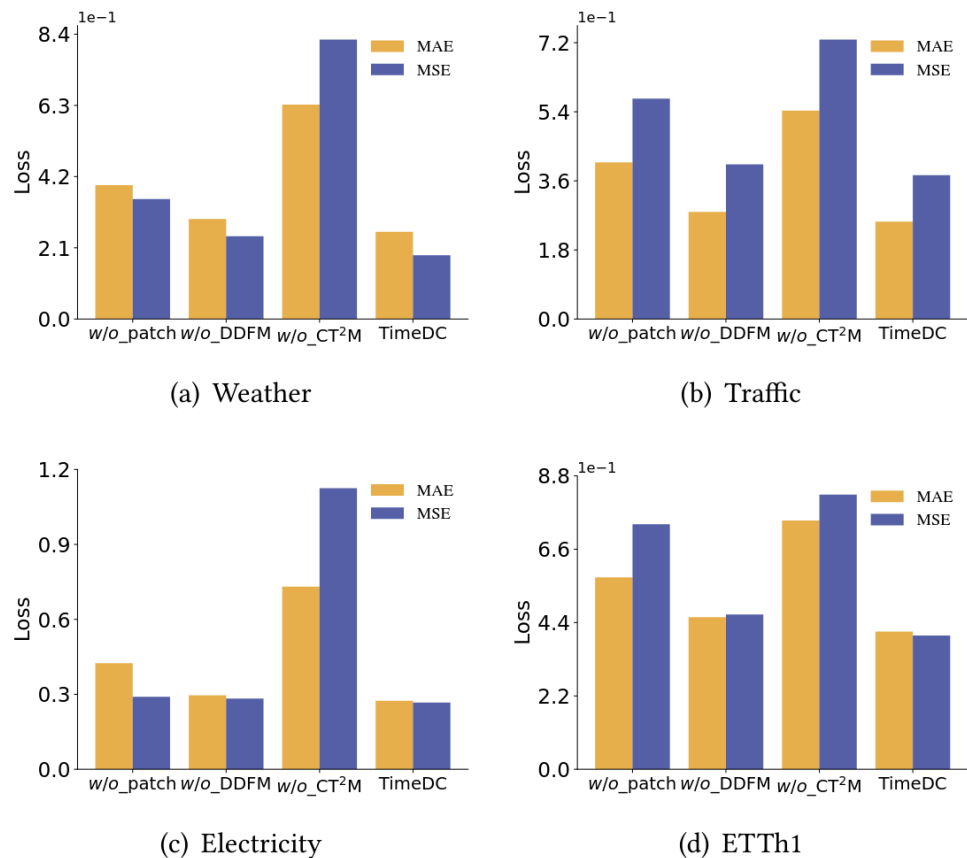


Figure 6: Performance of TimeDC and Its Variants on Four Datasets ($PL = 96$)

3.4 Condensed Dataset Generalization

Table 2: Cross-Architecture Performance Comparison									
Method	Metric	PL	Weather	Traffic	Electricity	ETT			
						ETTh1	ETTh2	ETTm1	ETTm2
TimeDC	MAE	96	0.257	0.254	0.274	0.413	0.368	0.503	0.354
		192	0.285	0.297	0.285	0.436	0.389	0.512	0.401
	MSE	96	0.188	0.375	0.267	0.401	0.271	0.442	0.391
		192	0.247	0.405	0.294	0.428	0.302	0.465	0.421
Autoformer	MAE	96	<u>0.312</u>	<u>0.370</u>	<u>0.343</u>	<u>0.453</u>	<u>0.473</u>	<u>0.548</u>	<u>0.342</u>
		192	<u>0.381</u>	<u>0.385</u>	<u>0.355</u>	<u>0.478</u>	<u>0.491</u>	<u>0.550</u>	<u>0.334</u>
	MSE	96	<u>0.255</u>	0.597	<u>0.236</u>	<u>0.465</u>	<u>0.412</u>	<u>0.542</u>	<u>0.265</u>
		192	<u>0.334</u>	0.613	<u>0.264</u>	<u>0.493</u>	<u>0.488</u>	<u>0.532</u>	<u>0.287</u>
Informer	MAE	96	0.423	0.430	0.428	0.773	0.842	0.576	0.552
		192	0.482	0.476	0.446	0.788	0.954	0.597	0.532
	MSE	96	0.354	0.643	0.253	0.992	1.032	0.624	0.402
		192	0.478	0.710	0.271	0.987	1.055	0.653	0.432
Transformer	MAE	96	0.389	0.412	0.398	0.632	0.506	0.563	0.555
		192	0.588	0.431	0.422	0.612	0.513	0.588	0.576
	MSE	96	0.344	<u>0.578</u>	0.267	0.785	0.579	0.615	0.479
		192	0.524	<u>0.567</u>	0.288	0.732	0.542	0.643	0.455

4. Critique

4.1 Terminology

TimeDC

- Condensation framework
- TSFE prediction architecture
- Interchangeable
 - Understand from context

4.1 Terminology

Table 1: Overall Performance Comparison on Seven Datasets

Baseline		Random		K-Center		Herding		DC		MTT		TimeDC		Whole Dataset	
Dataset	PL	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Weather	96	0.731	1.256	0.452	0.687	0.478	0.677	0.361	0.514	<u>0.295</u>	<u>0.244</u>	0.257	0.188	<i>0.239</i>	<i>0.182</i>
	192	0.786	1.302	0.487	0.723	0.512	0.688	0.413	0.527	<u>0.344</u>	<u>0.301</u>	0.285	0.247	<i>0.261</i>	<i>0.195</i>
	336	0.794	1.311	0.524	0.756	0.554	0.712	0.444	0.567	<u>0.368</u>	<u>0.328</u>	0.330	0.287	<i>0.282</i>	<i>0.241</i>
Traffic	96	0.675	1.125	0.503	0.576	0.483	0.554	0.375	0.603	<u>0.279</u>	<u>0.403</u>	0.254	0.375	<i>0.247</i>	<i>0.337</i>
	192	0.712	1.144	0.514	0.604	0.517	0.606	0.432	0.633	<u>0.336</u>	<u>0.442</u>	0.297	0.405	<i>0.265</i>	<i>0.338</i>
	336	0.729	1.117	0.523	0.611	0.553	0.654	0.449	0.676	<u>0.355</u>	<u>0.471</u>	0.312	0.423	<i>0.297</i>	<i>0.360</i>
Electricity	96	0.421	0.669	0.448	0.583	0.501	0.592	0.376	0.513	<u>0.296</u>	<u>0.283</u>	0.274	0.267	<i>0.252</i>	<i>0.268</i>
	192	0.450	0.743	0.476	0.601	0.534	0.628	0.419	0.532	<u>0.315</u>	<u>0.337</u>	0.285	0.294	<i>0.239</i>	<i>0.255</i>
	336	0.491	0.853	0.506	0.622	0.569	0.477	0.436	0.544	<u>0.339</u>	<u>0.356</u>	0.304	0.322	<i>0.271</i>	<i>0.285</i>
ETTh1	96	0.523	0.745	0.554	0.698	0.536	0.656	0.503	<u>0.442</u>	<u>0.456</u>	0.464	0.413	0.401	<i>0.354</i>	<i>0.386</i>
	192	0.557	0.786	0.578	0.722	0.589	0.698	0.552	0.508	<u>0.504</u>	<u>0.471</u>	0.436	0.428	<i>0.362</i>	<i>0.355</i>
	336	0.588	0.802	0.604	0.745	0.603	0.723	0.556	0.513	<u>0.498</u>	<u>0.464</u>	0.447	0.431	<i>0.409</i>	<i>0.387</i>
ETTh2	96	0.487	0.655	0.589	0.711	0.521	0.589	0.463	0.524	<u>0.388</u>	<u>0.342</u>	0.368	0.271	<i>0.324</i>	<i>0.255</i>
	192	0.509	0.673	0.605	0.732	0.553	0.621	0.488	0.536	<u>0.416</u>	<u>0.384</u>	0.389	0.302	<i>0.332</i>	<i>0.257</i>
	336	0.524	0.689	0.623	0.744	0.564	0.640	0.505	0.540	<u>0.435</u>	<u>0.455</u>	0.411	0.334	<i>0.376</i>	<i>0.296</i>
ETThm1	96	0.743	1.124	0.525	0.492	0.607	0.554	0.603	0.665	<u>0.512</u>	<u>0.453</u>	0.503	0.442	<i>0.453</i>	<i>0.403</i>
	192	0.764	1.245	0.566	0.510	0.628	0.571	0.597	0.647	<u>0.563</u>	<u>0.501</u>	0.512	0.465	<i>0.464</i>	<i>0.432</i>
	336	0.801	1.128	0.571	0.523	0.644	0.582	0.624	0.668	<u>0.552</u>	<u>0.488</u>	0.500	0.483	<i>0.477</i>	<i>0.455</i>
ETThm2	96	0.664	0.795	0.486	0.623	0.524	0.558	0.472	0.535	<u>0.376</u>	<u>0.421</u>	0.354	0.391	<i>0.347</i>	<i>0.381</i>
	192	0.687	0.804	0.512	0.643	0.549	0.583	0.488	0.567	<u>0.453</u>	<u>0.479</u>	0.401	0.421	<i>0.358</i>	<i>0.403</i>
	336	0.702	0.823	0.558	0.661	0.598	0.624	0.493	0.556	<u>0.473</u>	<u>0.523</u>	0.453	0.474	<i>0.406</i>	<i>0.435</i>

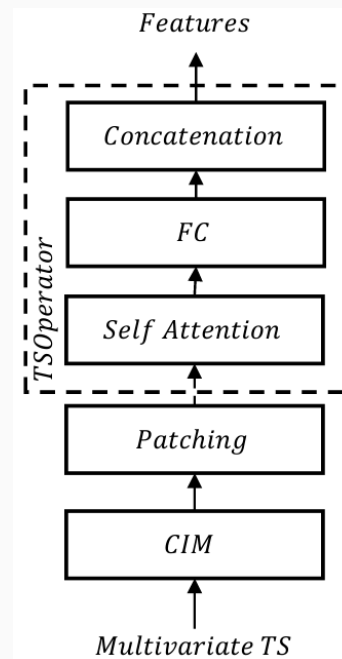
Table 2: Cross-Architecture Performance Comparison

Method	Metric	PL	Weather	Traffic	Electricity	ETT			
						ETTh1	ETTh2	ETThm1	ETThm2
TimeDC	MAE	96	0.257	0.254	0.274	0.413	0.368	0.503	0.354
		192	0.285	0.297	0.285	0.436	0.389	0.512	0.401
	MSE	96	0.188	0.375	0.267	0.401	0.271	0.442	0.391
		192	0.247	0.405	0.294	0.428	0.302	0.465	0.421
Autoformer	MAE	96	<u>0.312</u>	<u>0.370</u>	<u>0.343</u>	<u>0.453</u>	<u>0.473</u>	<u>0.548</u>	<u>0.342</u>
		192	<u>0.381</u>	<u>0.385</u>	<u>0.355</u>	<u>0.478</u>	<u>0.491</u>	<u>0.550</u>	<u>0.334</u>
	MSE	96	<u>0.255</u>	<u>0.597</u>	<u>0.236</u>	<u>0.465</u>	<u>0.412</u>	<u>0.542</u>	<u>0.265</u>
		192	<u>0.334</u>	0.613	<u>0.264</u>	<u>0.493</u>	<u>0.488</u>	<u>0.532</u>	<u>0.287</u>
Informer	MAE	96	0.423	0.430	0.428	0.773	0.842	0.576	0.552
		192	0.482	0.476	0.446	0.788	0.954	0.597	0.532
	MSE	96	0.354	0.643	0.253	0.992	1.032	0.624	0.402
		192	0.478	0.710	0.271	0.987	1.055	0.653	0.432
Transformer	MAE	96	0.389	0.412	0.398	0.632	0.506	0.563	0.555
		192	0.588	0.431	0.422	0.612	0.513	0.588	0.576
	MSE	96	0.344	<u>0.578</u>	0.267	0.785	0.579	0.615	0.479
		192	0.524	<u>0.567</u>	0.288	0.732	0.542	0.643	0.455

4.2 Method

TSOperator Structure

- Figure 3
 - Self-attention
 - Fully connected
 - Concatenation
- Equation 4 + Algorithm 1
 - Self-attention
 - Fully connected
 - Normalization



$$h^j = \text{TSOperator}(h^{j-1}) = \text{Norm}(\text{FC}(\text{MultiHead}(h^{j-1}))), \quad (4)$$

4.2 Method

DDFM

- No details on comparison
 - Full condensed dataset?
 - Single time series from original?
 - Average similarity?

4.2 Method

CT²M

- Unclear how loss works
 - Accumulated?
 - Sequential backpropagation?
- Why order similarity?

Algorithm 2: Curriculum Training Trajectory Query and Matching

Input: A buffer \mathcal{B} with a set of trajectories pre-trained on the original TS dataset \mathcal{T} parameterized by $\{\Theta_{\mathcal{T}}^k\}_{k=1}^K$; current model parameters $\tilde{\theta}^S$ on \mathcal{S} .

Output: Trajectory matching loss L_{tmm} .

```
1 Distance list  $DT \leftarrow []$ ;  
2 Pre-update  $\tilde{\theta}^S$  for  $a$ -steps with Equation 13;  
3  $\tilde{\theta}_{e_0+a} \leftarrow \tilde{\theta}_{e_0} - \sum_{s=1}^a (\alpha \nabla \mathcal{L}(f_{\tilde{\theta}^S}, \mathcal{S}))$ ;  
4 for  $\Theta_{\mathcal{T}}^k \in \{\Theta_{\mathcal{T}}^k\}_{k=1}^K$  do  
5   Compute the distance  $dis_k$  between  $\tilde{\theta}|_{e_0}^a$  and  $\Theta_{\mathcal{T}}^k$  with  
   Equation 15;  
6    $dis_k \leftarrow -D(\tilde{\theta}|_{e=e_0}^a, \theta^k|_{e=e_0})$ ;  
7    $DT \leftarrow (k, dis_k)$ ;  
8 Rank  $DT$  in a descending order;  
9  $\beta \leftarrow 0$   
10 while  $\beta < K$  do  
11    $k \leftarrow DT[\beta][0]$   
12    $L_{tmm} \leftarrow$  Sample trajectory  $\Theta_{\mathcal{T}}^k$  and match the training  
   trajectory according to Equation 11;  
13    $\beta \leftarrow \beta + 1$ ;  
14 return  $L_{tmm}$ 
```

4.3 Experiments and Results

Dataset Condensation Performance (Table 1)

- All datasets
- 96, 192, 336 PL

Generalization Performance (Table 2)

- All datasets
- 96, 192 PL
 - No explanation for 336 missing

Dynamic Memory Cost (Table 3)

- Only 4 datasets
- Only ETTh1 as sub-dataset

4.3 Experiments and Results

Condensed Dataset Size (Figure 5)

- Only 4 datasets
- Only ETTh1 as sub-dataset
- Only 96 PL
 - No explanation
- Same for ablation study (Figure 6)

Training Time for Dataset Condensation

- Only 4 datasets (same as others)
- 96, 192, 336 PL again

5. Project & Course

5. Project & Course

Current Project

- Kind of opposites
 - We have little data, would like more
- Explore how “dense” our data is
 - New usecase for TimeDC

5. Project & Course

Current Project

- Kind of opposites
 - We have little data, would like more
- Explore how “dense” our data is
 - New usecase for TimeDC

Previous Projects

- Trajectory simplification
- Not synthesized data
- Also focused on use case
 - Query-driven trajectory simplification

5. Project & Course

Course

- Memformer
- Three shared authors
- Time series forecasting (long-term)
- Patching
- Channel independence

6. Evaluation

6. Evaluation

Idea

- Logical methodology
 - Data loss (DDFM)
 - Training loss (CT²M)
 - Usecase loss (downstream prediction task)
- Focus on efficiency
 - Faster training
 - Less data
 - Lower memory usage

6. Evaluation

Execution

- Inconsistencies
 - TSOoperator structure
 - TimeDC terminology
- Missing details
 - Loss in CT²M
- Cherry-picked results
 - Datasets
 - Forecasting horizons
 - Bold in Table 2
- Resource constrained environments
 - Edge computing?
 - Embedded?
 - Moved computation to offline