OliO: ODE-based Linear Transition Operator for Self-Supervised Time Series Forecasting

Kwangryeol Park
Graduate School of Artificial Intelligence
Ulsan National Institute of Science and Technology
Ulsan, South Korea
pkr7098@unist.ac.kr

Seulki Lee*
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
seulki.lee@kaist.ac.kr

Abstract

Self-supervised representation learning has advanced time series forecasting by capturing robust features from unlabeled data. However, existing contrastive and masking-based methods often struggle to explicitly model the underlying temporal flow, either by treating time lag as noise or by disrupting the inherent temporal dynamics essential for forecasting. To address these limitations, we propose OliO (**O**DE-based **Li**near Transition **O**perator), a plug-and-play self-supervised learning method that learns bidirectional temporal flows as a continuous and structurally consistent dynamical system. OliO introduces a transition operator derived from linear Ordinary Differential Equations (ODEs) to align latent representations across arbitrary time shifts. By imposing a strictly upper triangular constraint on the transition matrix, we mathematically ensure numerical stability and enforce a robust polynomial boundary that effectively captures long-term dependencies while preventing the exponential divergence typically found in unconstrained ODE-based models. This structural constraint induces a temporally coherent representation space that preserves the underlying flow essential for accurate forecasting. Extensive experiments across various backbone architectures and benchmarks demonstrate that OliO achieves significant performance gains, providing Mean Squared Error (MSE) reductions of up to 5.75% over the most competitive grid-searched SOTA baselines. The code is available at https://github.com/eai-lab/OliO.

1 Introduction

Time series forecasting plays a crucial role in various real-world applications, such as energy consumption planning, traffic analysis, healthcare monitoring, and financial market prediction [31, 34]. However, building accurate forecasting models remains a significant challenge due to the complex temporal variations and high-dimensional nature of massive time series data. Recently, to address these challenges, self-supervised learning (SSL), which learns meaningful representations from unlabeled data, has been actively adopted in the time series domain. Representative approaches include contrastive learning [32, 7, 30], which learns representations by regulating the similarity

*Corresponding Author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

of related temporal views, and masked modeling [13, 5, 6], which learns through reconstruction of masked observations.

A central question for forecasting is how the temporal displacement between historical windows should be represented during self-supervised pre-training. Two overlapping windows may share many observations while representing different temporal states. Their temporal displacement specifies both how far apart the windows are and which window precedes the other. Contrastive and masked-modeling objectives provide supervision through representation similarity or observation reconstruction. SoftCLT [12] additionally uses temporal distance to weight similarity, but does not explicitly learn a transition conditioned on the signed lag between the two windows. This motivates an objective that explicitly learns where a historical representation should move under the pair's actual signed temporal displacement.

To address this question, we propose OliO (**O**DE-based **Li**near Transition **O**perator), a plug-and-play self-supervised framework for learning signed, lag-conditioned, bidirectional temporal transitions. Rather than directly aligning representations from different time points, OliO first transitions one representation from its original time point to the other before performing alignment. Specifically, given two representations independently encoded from historical windows separated by a lag $\tau > 0$, OliO models the relationship between the two representations using an ODE-derived linear transition operator [18, 25, 20]. It then transitions the earlier representation to the later time point using $+\tau$ and aligns it with the later representation. In the reverse direction, OliO similarly transitions the later representation using $-\tau$ and aligns it with the earlier representation.

OliO introduces the projection and transition modules only during pre-training. Both modules are discarded for downstream forecasting, leaving the forecasting architecture of the original backbone unchanged. We evaluate the framework across multiple backbones under a shared-backbone protocol and examine representation quality and transferability through frozen-backbone probing and zero-shot forecasting.

The main contributions of this paper are as follows:

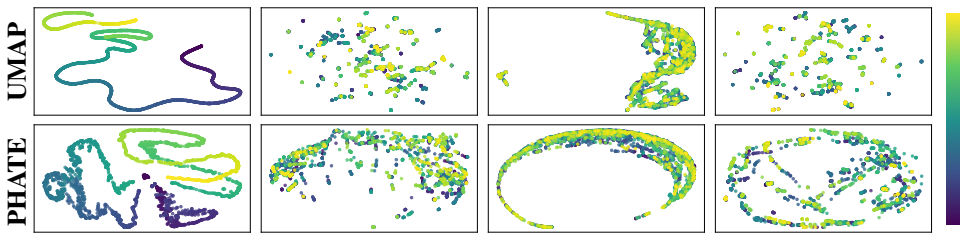
- **Signed temporal-transition pre-training:** We formulate a forecasting-oriented self-supervised objective that learns bidirectional transitions between independently encoded representations of historical windows, conditioned on the magnitude and direction of their actual temporal displacement.
- **Finite-polynomial transition structure:** We parameterize bidirectional transitions between representations of historical windows using a strictly upper-triangular ODE generator. The generator's nilpotency ensures that these transitions are exactly expressible as finite polynomials in temporal lag.
- **Model-agnostic representation learning:** OliO operates on the final-layer representations of forecasting backbones, allowing it to be applied during pre-training without modifying their internal architectures.

2 Related work

2.1 Self-supervised learning for time series

Contrastive learning and masked modeling provide effective supervision for learning time-series representations without forecasting labels. Contrastive methods such as TS2Vec [32], TS-TCC [7], and CoST [30] regulate similarity between related temporal views while separating negative pairs. Masked-modeling methods such as Ti-MAE [13] and SimMTM [5] learn representations through observation reconstruction. These objectives capture useful temporal structure, but do not explicitly parameterize how one historical representation should transition toward another under their actual signed temporal displacement.

Non-contrastive methods such as SimSiam [3] and BYOL [8] learn by predicting a target representation from another view, with gradients stopped through the target branch. OliO builds on this alignment mechanism. The distinction lies in the input and role of the prediction module: conventional view-alignment heads receive neither the actual temporal separation of two historical windows nor its direction. In contrast, OliO takes the pair-specific signed lag as an input and explicitly incorporates it into the transition module's formulation to transition the source representation before



(a) OliO (b) SimMTM (c) TS2Vec (d) SoftCLT

Figure 1: Visual comparison of the pseudo output h from the pretrained iTransformer [15] backbone using UMAP [17] (top row) and PHATE [21] (bottom row) on the Weather dataset [31]. The color gradient represents the progression of time. While baseline methods (SimMTM [5], TS2Vec [32], SoftCLT [12]) yield scattered or tightly clustered results that collapse temporal order, OliO maintains a continuous trajectory, effectively capturing the underlying temporal flow.

alignment. Thus, its temporal supervision comes from the actual displacement between the windows, in addition to prediction across the two branches.

Time-aware methods incorporate temporal relations more directly. SoftCLT [12] uses temporal distance to weight a contrastive similarity objective. The distance controls the strength of attraction between representations, rather than a transformation toward a lag-conditioned destination; exchanging temporal order leaves the unsigned distance unchanged. TimeSiam [6] incorporates temporally separated subseries and past-to-current reconstruction. It therefore exploits temporal context, while supervising observation reconstruction rather than a pair-specific signed transition between independently encoded historical representations. OliO instead conditions a learnable bidirectional transition on the actual signed lag and aligns each transitioned representation with the independently encoded representation of the other window. This makes temporal displacement determine a representation-level transition, rather than only the strength of similarity or the context for reconstructing observations.

Fig. 1 provides a qualitative comparison of temporal organization in representations learned on Weather [31]. Each point corresponds to a historical window, and its color encodes the window's time index. In SimMTM and SoftCLT, points from different time indices are scattered throughout the embedding, yielding fragmented and intermixed color patterns. TS2Vec forms a denser structure, but colors associated with different time indices remain mixed within it. In contrast, OliO exhibits a smoother progression of colors along its trajectory, suggesting a more temporally ordered organization of the visualized representations. These visualizations complement the forecasting evaluation; additional backbone and dataset visualizations appear in App. A.

2.2 Relation to latent dynamics and operator learning

Latent dynamics and operator-learning methods learn representations for trajectory generation [24], forecasting [28], and evolution-operator estimation and analysis [26]. OliO draws on latent-transition modeling to define a forecasting-oriented pre-training objective: the actual signed lag between independently encoded historical windows supervises bidirectional representation transitions. Its transition module is used to train forecasting representations and is then discarded together with the projection head; downstream prediction is performed by the original backbone. OliO therefore uses established latent-dynamics tools to construct signed temporal supervision for forecasting representations.

3 Methodology

3.1 Problem formulation and notation

Fig. 2 illustrates the OliO framework. We describe the construction using a univariate time series for simplicity; the same formulation applies to multivariate windows. Scalars are denoted by lowercase letters, vectors by bold lowercase letters, and matrices by uppercase letters. Here, t is absolute

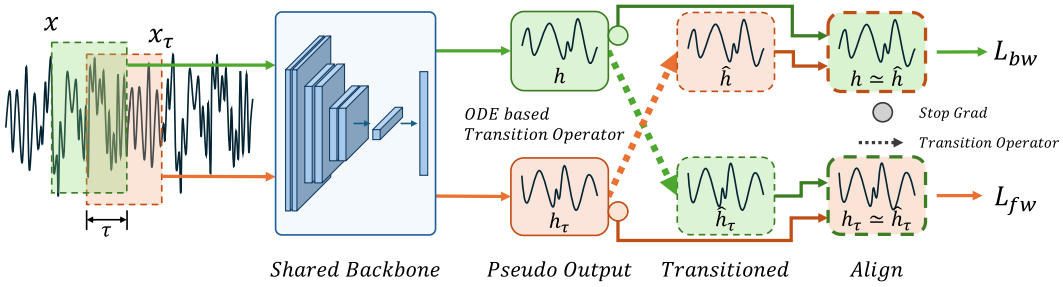


Figure 2: Overall architecture of the OliO framework. (a) Input windows x and x_τ are encoded by a shared backbone into pseudo-outputs. (b) The forward shift operator aligns h to h_τ via the matrix exponential $e^{\tau K}$, modeling the transition from past to future. (c) Conversely, the backward shift operator restores the temporal flow from future to past, ensuring bidirectional transition consistency.

time and τ the nonnegative separation between window starts. Subscripts θ and ϕ denote network parameters, and fw and bw denote the forward and backward directions.

Consider a discretely sampled time series and a historical window $\boldsymbol{x}^{(t)} = [x_t, x_{t+1}, \ldots, x_{t+l-1}]^\top \in \mathbb{R}^l$ of length l. A second window $\boldsymbol{x}^{(t+\tau)}$ has the same length and begins τ time steps later. During standard pre-training, $\tau \in \mathbb{N}$ satisfies $0 < \tau < l$, so the two windows overlap. We use $+\tau$ for the transition from the earlier window to the later one and $-\tau$ for the reverse direction.

A shared forecasting backbone f_θ independently encodes the two windows into final-layer representations $\boldsymbol{r}^{(t)}, \boldsymbol{r}^{(t+\tau)} \in \mathbb{R}^d$. A shared projection head g_ϕ, implemented as a two-layer MLP with ReLU activation, maps them into the pseudo-output space [2], producing $\boldsymbol{h}^{(t)} = g_\phi(\boldsymbol{r}^{(t)})$ and $\boldsymbol{h}^{(t+\tau)} = g_\phi(\boldsymbol{r}^{(t+\tau)})$ in $\mathbb{R}^{d_h}$. OliO transitions each source pseudo-output according to the signed lag before aligning it with the independently encoded pseudo-output of the other window. In each direction, the target remains untransformed.

3.2 ODE-based transition operator

We derive the matrix transition from a homogeneous linear ordinary differential equation:

$$\frac{d\boldsymbol{h}}{d\tau} = K\boldsymbol{h}, \tag{1}$$

where $K \in \mathbb{R}^{d_h \times d_h}$ is a learnable generator. The corresponding homogeneous transition of a source pseudo-output $\boldsymbol{h}$ over a lag τ is $e^{\tau K}\boldsymbol{h}$, providing a closed-form transition without an iterative nonlinear ODE solver. The ODE describes latent transport without assuming that observations follow a linear or continuous-time process. Training evaluates it at integer lags.

OliO augments this matrix transition with a shared learnable offset $\boldsymbol{v} \in \mathbb{R}^{d_h}$, scaled by the signed lag. The forward and backward transitions are

$$\hat{\boldsymbol{h}}^{(t+\tau)} = \mathcal{T}_{+\tau}(\boldsymbol{h}^{(t)}) = e^{\tau K}\boldsymbol{h}^{(t)} + \tau\boldsymbol{v}, \tag{2}$$

$$\hat{\boldsymbol{h}}^{(t)} = \mathcal{T}_{-\tau}(\boldsymbol{h}^{(t+\tau)}) = e^{-\tau K}\boldsymbol{h}^{(t+\tau)} - \tau\boldsymbol{v}. \tag{3}$$

Both directions share the same generator K and offset $\boldsymbol{v}$; reversing the temporal direction changes the sign of the lag in both terms. The matrix component is ODE-derived, while the shared signed offset supplies additional lag-dependent displacement. The transitioned source $\hat{\boldsymbol{h}}^{(t+\tau)}$ is aligned with the later target $\boldsymbol{h}^{(t+\tau)}$, and $\hat{\boldsymbol{h}}^{(t)}$ is aligned with the earlier target $\boldsymbol{h}^{(t)}$.

3.3 Strictly upper-triangular constraint

We constrain the generator to be strictly upper triangular, with $k_{ij} = 0$ for $i \geq j$. This makes K nilpotent: there exists a smallest positive integer $q \leq d_h$ such that $K^q = 0$. Nilpotency provides an exact finite-polynomial realization of the matrix transition in both temporal directions.

Theorem 1 (Finite-polynomial transition structure)**.** *Let $K \in \mathbb{R}^{d_h \times d_h}$ be strictly upper triangular with nilpotency index q. For every nonnegative integer lag $\tau \in \mathbb{N}$,*

$$e^{\pm \tau K} = \sum_{j=0}^{q-1} \frac{(\pm\tau)^j K^j}{j!}. \tag{4}$$

The matrix $e^{\pm\tau K}$ is upper triangular with unit diagonal. For any fixed $\boldsymbol{h}$ and $\boldsymbol{v}$, the complete transition $\mathcal{T}_{\pm\tau}(\boldsymbol{h}) = e^{\pm\tau K}\boldsymbol{h} \pm \tau\boldsymbol{v}$ is a vector-valued polynomial in τ of degree at most $\max\{q-1, 1\}$.

The finite expansion follows because every power K^j with $j \geq q$ vanishes; App. B provides the derivation. The expansion can contain higher-order terms while excluding exponential dependence on lag. Temporal direction is specified by the sign of $\pm\tau$, whereas triangularity determines the algebraic structure across latent coordinates. Polynomial dependence alone guarantees neither bounded magnitudes nor global distance monotonicity.

3.4 Computational complexity analysis

The finite expansion in Eq. (4) can be evaluated directly on a representation vector. Retaining terms through order p requires p matrix–vector products, giving a cost of $O(pd_h^2)$, plus $O(d_h)$ for the additive offset. App. B.4 gives the matrix- and vector-wise evaluations using Horner's method.

For exact evaluation, $p = q-1$, so the vector-wise cost is $O((q-1)d_h^2 + d_h)$ and remains cubic in the worst case $q = d_h$. Using a fixed lower order $p < q-1$ reduces computation but approximates the complete transition. The main OliO formulation uses the complete transition; lower-order approximations are evaluated separately in Sec. 4.4. Their accuracy and measured transition-computation speed reflect a trade-off between retained polynomial terms and computational cost. Since the projection and transition modules are used only during pre-training, downstream forecasting retains the inference architecture and cost of the original backbone.

3.5 Self-supervised objective

OliO learns the transition relation by aligning each transitioned pseudo-output with its independently encoded target. This objective uses the observed historical windows as supervision without mining negative pairs. Following non-contrastive representation alignment [3, 8, 23], we apply stop-gradient, denoted by $sg(\cdot)$, to the target in each direction. The directional losses are

$$\mathcal{L}_{fw} = 1 - \mathrm{CosSim}\Big(sg(\boldsymbol{h}^{(t+\tau)}), \hat{\boldsymbol{h}}^{(t+\tau)}\Big), \tag{5}$$

$$\mathcal{L}_{bw} = 1 - \mathrm{CosSim}\Big(sg(\boldsymbol{h}^{(t)}), \hat{\boldsymbol{h}}^{(t)}\Big). \tag{6}$$

The total objective averages the two directions:

$$\mathcal{L}_{\mathrm{OliO}} = \frac{1}{2}\left(\mathcal{L}_{fw} + \mathcal{L}_{bw}\right). \tag{7}$$

Each window serves as a source in one direction and a stopped-gradient target in the other. Stop-gradient breaks gradient symmetry and empirically mitigates representation collapse [3, 8, 23].

Through this objective, the backbone and projection head learn representations whose transitions are conditioned on the actual temporal displacement between the windows. After pre-training, the projection head and transition module are discarded, and downstream forecasting proceeds through the original backbone and its forecasting head.

4 Experiments

We evaluate forecasting performance, representation quality, and transferability, then test the signed-lag supervision and transition structure through controlled ablations.

4.1 Experimental setup

Datasets and backbones. We use ETTh1, ETTh2, ETTm1, ETTm2, Weather, and Electricity (ECL) [31], with prediction horizons $\{96, 192, 336, 720\}$. We evaluate three forecasting backbones:

Table 1: Long-term forecasting results for PatchTST [22] and iTransformer [15] backbones. Comparison between random initialization (Rand.), OliO (ours), and state-of-the-art SSL methods. TS2Vec [32] is denoted as TS2V, SimMTM [5] as MTM, SoftCLT [12] as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined.

Dataset		PatchTST							iTransformer						
		Rand.	**OliO**	TS2V	MTM	SCLT	Siam	URL	Rand.	**OliO**	TS2V	MTM	SCLT	Siam	URL
ETTh1	96	.3735	.3716	.3708	.3985	.3754	.3871	.3708	**.3903**	.3969	.4027	.3993	.4040	.3914	.3924
	192	.4094	.4082	.4080	.4299	.4095	.4192	**.4062**	.4448	.4442	.4507	.4490	.4540	**.4408**	.4448
	336	.4343	**.4252**	.4311	.4473	.4357	.4380	.4329	.4884	.4865	.4875	.4927	.4932	**.4824**	.4868
	720	.4582	**.4356**	.4483	.4560	.4589	.4547	.4520	.5120	.4919	.4978	.5043	.5203	**.4812**	.5016
ETTh2	96	.2833	**.2769**	.2778	.2783	.2802	.2808	.2773	**.2988**	.3002	.3079	.3181	.3016	.3050	.3037
	192	.3638	**.3392**	.3406	.3431	.3441	.3440	.3410	**.3785**	.3811	.3923	.4040	.3871	.3794	.3891
	336	.3918	**.3599**	.3606	.3642	.3643	.3676	.3601	.4250	.4230	.4333	.4496	.4271	**.4180**	.4313
	720	.3986	.3934	**.3910**	.3949	.3960	.3925	.3926	.4357	**.4271**	.4373	.4386	.4412	.4289	.4398
ETTm1	96	.2859	**.2830**	.2860	.2839	.2917	.2948	.2837	.3456	**.3291**	.3367	.3341	.3368	.3435	.3315
	192	.3266	**.3242**	.3279	.3255	.3331	.3363	.3252	.3831	**.3711**	.3777	.3746	.3764	.3837	.3732
	336	.3642	.3605	.3668	**.3592**	.3692	.3672	.3627	.4205	**.4064**	.4138	.4083	.4097	.4175	.4124
	720	.4140	**.4060**	.4162	.4120	.4154	.4165	.4126	.4896	**.4730**	.4844	.4784	.4805	.4839	.4812
ETTm2	96	.1660	**.1636**	.1669	.1725	.1694	.1692	.1658	.1857	**.1812**	.1901	.1836	.1875	.1826	.1838
	192	.2243	**.2189**	.2220	.2278	.2249	.2279	.2210	.2528	**.2486**	.2548	.2488	.2539	.2536	.2517
	336	.2770	**.2727**	.2760	.2835	.2803	.2802	.2752	.3146	.3105	.3181	**.3095**	.3173	.3203	.3134
	720	.3670	**.3644**	.3684	.3706	.3722	.3672	.3701	.4138	**.4088**	.4184	.4131	.4151	.4191	.4129
Weather	96	.1529	**.1507**	.1615	.1646	.1540	.1511	.1528	.1776	**.1756**	.2139	.2041	.2156	.1778	.1834
	192	.1959	**.1948**	.2036	.2043	.1975	.1949	.1976	.2263	**.2256**	.2591	.2499	.2570	.2262	.2314
	336	.2478	**.2453**	.2528	.2549	.2466	.2467	.2467	.2829	.2825	.3075	.3008	.3085	**.2819**	.2865
	720	.3221	**.3193**	.3242	.3253	.3201	.3207	.3216	.3590	.3587	.3774	.3706	.3798	**.3574**	.3623
ECL	96	.1356	.1297	.1314	.1387	.1325	.1300	**.1287**	.1503	**.1497**	.2611	.3656	.1646	.2849	.1508
	192	.1515	.1476	.1481	.1541	.1493	.1481	**.1471**	**.1649**	.1671	.2646	.3655	.1789	.2908	.1659
	336	.1682	**.1633**	.1653	.1707	.1656	.1672	.1638	.1815	.1800	.2880	.3800	.1982	.3050	**.1797**
	720	.2064	**.2001**	.2013	.2026	.2017	.2042	.2123	.2131	**.2107**	.3216	.4121	.2394	.3138	.2114
Top1 count		0	**18**	1	1	0	0	3	4	**12**	0	1	0	6	1
Top2 count		1	3	6	1	2	3	9	2	9	0	3	0	5	5

PatchTST [22], which uses temporal patches; iTransformer [15], which attends across variable tokens; and the multimodal TimeVLM [33]. Dataset statistics appear in App. C.

Baselines and evaluation. We compare random initialization with TS2Vec [32], SimMTM [5], SoftCLT [12], TimeSiam [6], TimesURL [14], and OliO using the same forecasting backbones. These comparisons evaluate SSL objectives under shared-backbone adaptation; temporal loss components requiring a timestamp axis are omitted when that axis is unavailable, as documented in App. D. Following TimeSiam, we fully finetune the pretrained backbone using L_2 loss. Input lengths follow the backbone configurations, and standard OliO pre-training samples overlapping window pairs. We retain the baseline-specific hyperparameter searches in App. D. Configurations are selected by validation MSE, fixed, and evaluated on the test split with seeds 2023, 2024, and 2025. Test performance is excluded from model and hyperparameter selection. Unless stated otherwise, results report three-seed mean MSE; App. G provides per-setting MSE and MAE with sample standard deviations. Few-shot results using 1%, 5%, and 10% of the training data are provided in App. F.

4.2 Long-term forecasting

Tables 1 and 2 report forecasting MSE for the three backbones, and Fig. 3 summarizes relative gains across horizons. Against random initialization, OliO reduces MSE by up to 8.1%, 4.8%, and 9.2% on PatchTST, iTransformer, and TimeVLM, respectively. On PatchTST, it ranks first in 18 of 24 settings. The improvement varies by dataset and backbone; some SSL methods also underperform random initialization, showing that pre-training benefits depend on the downstream setting.

Table 2: Long-term forecasting results for TimeVLM [33]. TS2Vec is denoted as TS2V, SimMTM as MTM, SoftCLT as SCLT, TimeSiam as Siam, and TimesURL as URL. Best results are in **bold** and the second best are underlined.

Dataset		TimeVLM						
		Rand.	**OliO**	TS2V	MTM	SCLT	Siam	URL
ETTh1	96	.3704	.3669	.3666	.3666	**.3660**	.3779	.3672
	192	.4074	.4029	.4040	.4038	**.4017**	.4074	.4027
	336	.4659	.4232	.4287	.4341	.4232	.4285	.4250
	720	.4981	.4587	.4634	.4930	**.4556**	.4577	.4596
ETTh2	96	.2818	.2755	.2766	.2862	.2762	.2777	**.2754**
	192	.3497	**.3359**	.3369	.3409	.3377	.3409	.3368
	336	.3767	**.3613**	.3621	.3664	.3624	.3713	.3625
	720	.4117	.4009	.4069	.4073	.4119	.4218	**.3977**
ETTm1	96	.2908	.2895	.2896	.2899	.2903	.2940	.2895
	192	.3354	**.3271**	.3310	.3348	.3308	.3331	.3309
	336	.3676	**.3593**	.3617	.3663	.3615	.3619	.3605
	720	.4209	.4096	.4102	.4104	.4102	.4165	**.4081**
ETTm2	96	.1690	.1674	.1673	.1691	.1688	.1678	.1673
	192	.2251	.2216	.2233	.2238	**.2211**	.2227	.2229
	336	.2745	.2733	.2702	.2728	**.2699**	.2716	.2712
	720	.3676	**.3521**	.3540	.3581	.3544	.3590	.3540
Weather	96	.1495	.1490	.1490	.1492	**.1488**	.1520	.1491
	192	.1937	.1937	.1938	.1946	.1942	.1956	.1938
	336	.2445	**.2433**	.2444	.2451	.2453	.2453	.2445
	720	.3180	.3148	**.3147**	.3163	.3170	.3151	.3149
ECL	96	.1353	.1319	**.1315**	.1318	.1325	.1346	.1363
	192	.1513	.1483	**.1478**	.1487	.1481	.1533	.1534
	336	.1673	**.1643**	.1670	.1665	.1651	.1692	.1727
	720	.2080	**.2044**	.2091	.2080	.2077	.2080	.2061
Top1 count		0	**8**	3	0	6	0	3
Top2 count		1	8	4	1	4	1	6

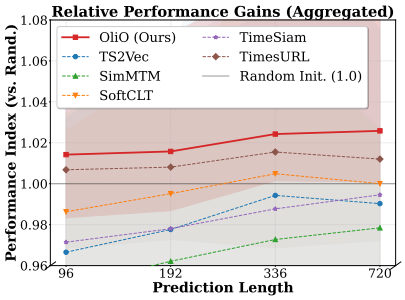


Figure 3: Relative performance improvement of our proposed method, OliO, compared to state-of-the-art SSL baselines. The performance index is defined as $MSE_{\text{Rand}}/MSE_{\text{Meth.}}$ across varying prediction lengths, where a value of 1.0 represents the random initialization baseline, and **higher values indicate better performance**. Solid lines indicate the mean index aggregated across three backbone architectures (PatchTST, iTransformer, TimeVLM) and all datasets, with shaded regions representing the min-max performance envelope. Note the truncated Y-axis $[0.96, 1.08]$ for focused visualization of performance gains.

Effect size and seed variation. TimesURL is the strongest baseline by aggregate MSE. Using unrounded three-seed means, OliO outperforms it in 58/72 backbone–dataset–horizon settings, with mean, median, and maximum relative MSE improvements of 0.88%, 0.55%, and 5.75%. Averaging the four horizons within each dataset–backbone group, OliO wins 17/18 groups; a two-sided exact Wilcoxon signed-rank test gives $W = 3$, $p = 3.81 \times 10^{-5}$. It also wins all six dataset-level averages. These grouped comparisons reduce dependence among horizon-level results. Table 3 summarizes full-finetuning performance and seed variation alongside frozen-backbone results. Each entry $\bar{m} \pm \bar{s}$ averages the per-setting three-seed means and sample standard deviations, respectively; $\bar{s}$ describes typical seed variation, rather than the standard deviation of an aggregated score. Further statistical comparisons are in App. G.1.

Table 3: OliO–TimesURL comparison under full finetuning and frozen-backbone linear probing.

	Full finetuning		Frozen backbone	
Backbone	TimesURL	OliO	TimesURL	OliO
PatchTST	0.2925 ± 0.0007	$\mathbf{0.2898 \pm 0.0007}$	$\mathbf{0.2994 \pm 0.0011}$	0.2997 ± 0.0008
iTransformer	0.3300 ± 0.0017	$\mathbf{0.3262 \pm 0.0022}$	0.3412 ± 0.0007	$\mathbf{0.3369 \pm 0.0007}$
TimeVLM	0.2918 ± 0.0018	$\mathbf{0.2906 \pm 0.0014}$	–	–

4.3 Representation quality and transferability

Frozen-backbone linear probing. To isolate representation quality from full finetuning, we freeze the pretrained backbone and train an identical linear forecasting head. Across PatchTST and iTransformer, six datasets, and four horizons (48 settings), OliO outperforms TimesURL in 35/48 settings and 8/12 dataset–backbone averages. Overall MSE decreases from 0.3203 to 0.3183. As Table 3 shows, iTransformer improves from 0.3412 to 0.3369, while PatchTST is comparable (0.2994 versus 0.2997). Thus, the aggregate representation-quality gain does not imply an improvement for every backbone. Dataset-level means and seed variation appear in App. E.

Zero-shot forecasting. We additionally use LightGTS-4M [29] with its original architecture, forecasting settings, and pre-training protocol. A single pretrained checkpoint is evaluated directly on eight downstream datasets disjoint from pre-training, without downstream parameter updates or checkpoint selection. Table 4 compares OliO, TimesURL, and the default mask-reconstruction objective over 32 dataset–horizon settings. OliO achieves average MSE 0.3034, compared with 0.3114 for mask reconstruction and 0.3219 for TimesURL. Relative to mask reconstruction, average MAE decreases from 0.3469 to 0.3391. The MSE gains include Electricity, Solar, and Traffic, while Weather favors mask reconstruction. These results extend the evaluation beyond full finetuning and complement the frozen-backbone comparison. Complete horizon-level results are in App. E.2.

Table 4: Zero-shot forecasting results with the LightGTS-4M time-series foundation model [29]. Mask Recon. denotes its default mask-reconstruction pre-training. Values average four forecasting horizons. Best results are in **bold**, and tied best values are underlined. Lower MSE is better.

Dataset	TimesURL	Mask Recon.	OliO
ETTh1	0.4096	0.3958	**0.3937**
ETTh2	0.3501	0.3474	**0.3420**
ETTm1	0.3461	0.3335	**0.3329**
ETTm2	0.2566	0.2461	0.2461
Weather	0.2357	**0.2134**	0.2182
ECL	0.2556	0.2476	**0.2325**
Solar	0.1679	0.1712	**0.1612**
Traffic	0.5540	0.5364	**0.5010**
Average	0.3219	0.3114	**0.3034**

4.4 Ablation studies

Actual signed temporal displacement. Table 5 tests which aspects of the signed lag provide useful supervision on PatchTST and iTransformer across 48 settings. Each control changes only the supplied lag relative to the original formulation. *Fixed* retains the correct direction but replaces the pair-specific magnitude with a constant. Its 1.10% higher MSE indicates that the actual separation carries useful information beyond temporal order. *Random Magnitude* retains the correct sign but supplies a mismatched magnitude. The 2.57% increase shows that varying the conditioning value alone does not recover the benefit of matching it to the actual window separation. *Random Direction* preserves the actual magnitude and opposite signs between branches, but randomizes their correspondence to earlier–later order. Its 1.82% increase isolates the value of assigning signs according to the true temporal direction, beyond simply using opposite signs in the two branches. *Unsigned* supplies the correct magnitude with a positive sign in both branches. Its 4.86% increase, the largest among these controls, supports distinguishing forward from backward transport. Together, the controls support conditioning the transition on both the actual separation and its temporal direction. With the transition module and alignment objective held fixed, these results indicate that the actual temporal relation between windows provides useful supervision for forecasting.

Exponential transport and additive bias. Table 6 examines the contributions of matrix transport and the additive offset over the same 48 settings. *Bias-only* removes matrix transport, applying the same displacement to every representation at a given signed lag. Its 3.83% higher MSE supports allowing the displacement to depend on the source representation. *K-only* retains exponential transport but removes the offset. The 1.40% increase indicates that a learned translation complements the multiplicative transformation of the source state; neither component alone matches the complete

transition. *MLP* tests whether a generic nonlinear mapping can replace the explicit transition structure. Its 3.13% increase supports the proposed structured parameterization over this nonlinear alternative. *Spec-K* replaces the strictly upper-triangular generator with a full generator using spectral normalization. It remains numerically stable but has 1.18% higher MSE. Thus, the performance gap persists when both variants train without numerical divergence, supporting a contribution of the generator structure beyond avoiding numerical failure. Dataset-level results are in App. I.4.

Table 5: Effect of signed temporal displacement over 48 settings.

Variant	Avg. MSE	Δ MSE	Wins
Original	**0.3080**	–	–
Fixed	0.3114	+1.10%	40/48
Random Direction	0.3136	+1.82%	40/48
Random Magnitude	0.3159	+2.57%	41/48
Unsigned	0.3230	+4.86%	45/48

Table 6: Ablation of alternative transition parameterizations over 48 settings.

Variant	Avg. MSE	Δ MSE	Wins
OliO	**0.3080**	–	–
Spec-K	0.3116	+1.18%	39/48
K-only	0.3123	+1.40%	42/48
MLP	0.3176	+3.13%	45/48
Bias-only	0.3198	+3.83%	47/48

Bidirectionality, approximation order, and efficiency. Table 7 evaluates the original three-backbone, six-dataset setting (72 horizon-level configurations). Bidirectional learning achieves average MSE 0.3022, versus 0.3240 forward-only and 0.3245 backward-only, and improves all six dataset averages. This supports supervising both directions without requiring exact inverse composition. Table 8 compares the generator constraint and retained order p. The unconstrained full generator diverges in these experiments, whereas low-order polynomial approximations remain finite but can lose accuracy, especially on ECL. Strict upper triangularity guarantees a finite expansion, rather than bounded magnitudes for arbitrary lags. Fig. 4 shows up to $3.8\times$ faster transition computation with lower orders during pre-training; downstream inference uses the unchanged forecasting backbone. App. I provides backbone-specific results and orders $p = 1, \ldots, 5$.

Table 7: **Effect of bidirectional modeling.** Average MSE across all backbones.

Dataset	**Bi**	Fwd	Bwd
ETTh1	**.4260**	.4485	.4465
ETTh2	**.3562**	.3625	.3618
ETTm1	**.3616**	.3938	.3930
ETTm2	**.2653**	.2714	.2717
Weather	**.2378**	.2495	.2452
ECL	**.1664**	.2185	.2285

Table 8: **Structural constraints and order p.** Evaluation (MSE) of transition matrix constraints and Taylor orders.

Dataset	Full	$p = 1$	$p = 2$	$p = 3$	**Exact**
ETTh1	NaN	.4434	.4315	.4315	**.4260**
ETTh2	NaN	.3661	.3624	.3606	**.3562**
ETTm1	NaN	.3752	.3695	.3686	**.3616**
ETTm2	NaN	.2674	.2661	.2678	**.2653**
Weather	NaN	.2470	.2480	.2382	**.2378**
ECL	NaN	.2137	.2063	.2029	**.1664**

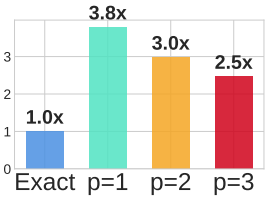


Figure 4: **Approximation speedup.** Latency improvement relative to full matrix exponential.

4.5 Overlap controls and finite-lag behavior

Shared observations. Varying overlap produces no monotonic relationship with MSE. Comparing 100% and 0% overlap across six representative dataset–backbone settings yields differences within 0.7% in five cases; 0% overlap improves over the model without OliO in five cases. These endpoints change both overlap and lag. To isolate shared observations, Table 9 instead holds the window pair, signed lag, 50% mask budget, backbone, training schedule, and hyperparameters fixed. We zero either the shared half of the second window or an equal-sized unique half. The aggregate MSE gap is only 0.22%, with 23 versus 25 lower-MSE settings. Thus, shared observations alone do not explain the gains under this controlled intervention. Full sweeps and endpoint results are in App. I.1; masking details are in App. I.3.

Higher-order terms and distance ordering. Table 10 compares $e^{\tau K}$ with $I + \tau K$ using relative Frobenius error. On learned ETTh1 and ETTh2 operators, the discrepancy increases with lag and reaches 1.004 and 0.430 at $\tau = 64$, respectively. Hence, the learned finite-lag transport does not generally reduce to its first-order term. Separately, checkpoint measurements over adjacent lags $k = 0, \ldots, 7$ show complete distance ordering for 88%–100% of evaluated representations, with positive mean and median squared-displacement increments at every adjacent lag. This is empirical

Table 9: Matched masking at 50% overlap, averaged over six datasets and four horizons. "Wins": Overlap-Mask / Unique-Mask. Lower MSE is better.

Backbone	Overlap mask	Unique mask	Gap	Wins
PatchTST	0.29818	**0.29767**	+0.17%	9 / 15
iTransformer	0.35022	**0.34932**	+0.26%	14 / 10
Overall	0.32420	**0.32349**	+0.22%	23 / 25

Table 10: Relative approximation error between the complete transition and its first-order approximation.

Lag τ	ETTh1	ETTh2
1	0.018	0.000
20	0.857	0.104
40	0.965	0.258
64	1.004	0.430

evidence over the evaluated range, rather than a global monotonicity guarantee. App. I.5 gives the approximation definition and additional lags; App. I.6 details the checkpoint diagnostic.

5 Future work and discussion

The signed-lag controls (Table 5) favor conditioning transitions on the actual temporal separation and direction. The overlap sweeps and matched masking control (Table 9) suggest that shared observations alone do not explain the forecasting gains. Together, these findings support learning temporal change by aligning window representations through transitions matched to their observed temporal relation.

The ODE determines OliO's homogeneous matrix transition, while the shared signed offset is parameterized separately. The complete affine map need not be the flow of an autonomous affine ODE. Each lag-specific affine map is invertible, but the opposite-lag transition need not be its inverse. Bidirectional alignment therefore supervises both temporal directions without imposing exact inverse composition. The projection and transition modules provide pre-training supervision and are discarded for downstream forecasting.

Koopman-based methods [11, 16] motivate representing nonlinear observations in coordinates suited to linear transport. OliO uses a nonlinear projection head for this purpose, but a finite-dimensional learned projection does not guarantee exact linearization of arbitrary dynamics. Its nilpotent generator models finite-lag polynomial transients rather than sustained periodic orbits. A single state-independent generator can also be restrictive when transitions vary across states or regimes; state-dependent transport is a possible extension.

Beyond forecasting, signed-lag supervision may be useful for other time-series tasks. SoftCLT [12] has demonstrated the value of distance-aware representations for classification, motivating further evaluation of OliO for structural-break detection and activity classification on benchmarks such as UCR [4] and UEA [1]. Such transfer remains an empirical question rather than a consequence of the finite-polynomial transition theorem.

6 Conclusion

We introduced OliO, a self-supervised pre-training framework that learns bidirectional transitions between representations of historical windows using their actual signed temporal displacement. Its transition combines an ODE-derived matrix exponential with a shared offset scaled by the signed lag. A strictly upper-triangular generator makes the transition exactly expressible as a finite polynomial in lag in both temporal directions. Across three forecasting backbones and six datasets, OliO improves on TimesURL, the strongest SSL baseline by aggregate MSE, in 58 of 72 settings, with up to 5.75% lower MSE. Frozen-backbone probing and zero-shot evaluation provide additional evidence of representation quality and transferability. Controlled ablations support using the actual lag magnitude and direction, combining exponential transport with an additive offset, and supervising both temporal directions. The projection and transition modules are discarded after pre-training, so downstream forecasting retains the original backbone's architecture and inference cost. These findings support using observed temporal displacement to guide representation learning for forecasting.

Acknowledgments and Disclosure of Funding

This work was supported by the Technology Innovation Program (Project No. RS-2026-25618509 and RS-2026-25620133) funded by the Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea) and the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2026-25617720, AI Star Fellowship Support Program).

References

[1] Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*, 2018.

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 13–18 Jul 2020. URL `https://proceedings.mlr.press/v119/chen20j.html`.

[3] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 15745–15753. IEEE, 2021.

[4] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

[5] Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm: A simple pre-training framework for masked time-series modeling. volume 36, pages 29996–30025, 2023.

[6] Jiaxiang Dong, Haixu Wu, Yuxuan Wang, Yunzhong Qiu, Li Zhang, Jianmin Wang, and Mingsheng Long. Timesiam: A pre-training framework for siamese time-series modeling. *arXiv preprint arXiv:2402.02475*, 2024.

[7] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. 2021.

[8] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[9] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, 2008.

[10] William George Horner. A new method of solving numerical equations of all orders, by continuous approximation. *Philosophical Transactions of the Royal Society of London*, 109: 308–335, 1819.

[11] B. O. Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931. doi: 10.1073/pnas.17.5.315. URL `https://www.pnas.org/doi/abs/10.1073/pnas.17.5.315`.

[12] Seunghan Lee, Taeyoung Park, and Kibok Lee. Soft contrastive learning for time series. In *International Conference on Learning Representations*, volume 2024, pages 46815–46839, 2024.

[13] Zhe Li, Zhongwen Rao, Lujia Pan, Pengyun Wang, and Zenglin Xu. Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint arXiv:2301.08871*, 2023.

[14] Jiexi Liu and Songcan Chen. Timesurl: self-supervised contrastive learning for universal time series representation learning. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i12.29299. URL https://doi.org/10.1609/aaai.v38i12.29299.

[15] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

[16] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):4950, 2018.

[17] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[18] Jose Marie Antonio Miñoza. Spike: Sparse koopman regularization for physics-informed neural networks. *arXiv preprint arXiv:2601.10282*, 2026.

[19] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.*, 20(4):801–836, October 1978. ISSN 0036-1445. doi: 10.1137/1020098. URL https://doi.org/10.1137/1020098.

[20] Arnab Mondal, Siba Smarak Panigrahi, Sai Rajeswar, Kaleem Siddiqi, and Siamak Ravanbakhsh. Efficient dynamics modeling in interactive environments with koopman theory. In *International Conference on Learning Representations*, volume 2024, pages 54691–54714, 2024.

[21] Kevin R Moon, David Van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37 (12):1482–1492, 2019.

[22] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

[23] Kwangryeol Park, Jaeho Kim, and Seulki Lee. Alio: Output alignment matters in long-term time series forecasting. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 119563–119608. Curran Associates, Inc., 2025. doi: 10.52202/085713-3992. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/ad52f9c2b232a719963a620dcf88fbe5-Paper-Conference.pdf.

[24] Yulia Rubanova, Ricky T. Q. Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/42a6845a557bef704ad8ac9cb4461d43-Paper.pdf.

[25] Nithin Somasekharan, Yadi Cao, and Shaowu Pan. Deep koopman sensing. *arXiv preprint arXiv:2609.15762*, 2026.

[26] Giacomo Turri, Luigi Bonati, Kai Zhu, Massimiliano Pontil, and Pietro Novelli. Self-supervised evolution operator learning for high-dimensional dynamical systems. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Ku3kLJle7Q.

[27] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[28] Rui Wang, Yihe Dong, Sercan O Arik, and Rose Yu. Koopman neural operator forecaster for time-series with temporal distributional shifts. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kUmdmHxK5N.

[29] Yihang Wang, Yuying Qiu, Peng Chen, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. Lightgts: A lightweight general time series forecasting model. *arXiv preprint arXiv:2506.06005*, 2025.

[30] Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. 2022.

[31] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

[32] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 8980–8987, 2022.

[33] Siru Zhong, Weilin Ruan, Ming Jin, Huan Li, Qingsong Wen, and Yuxuan Liang. Time-vlm: Exploring multimodal vision-language models for augmented time series forecasting. 2025. URL https://openreview.net/forum?id=b5h60xQnzM.

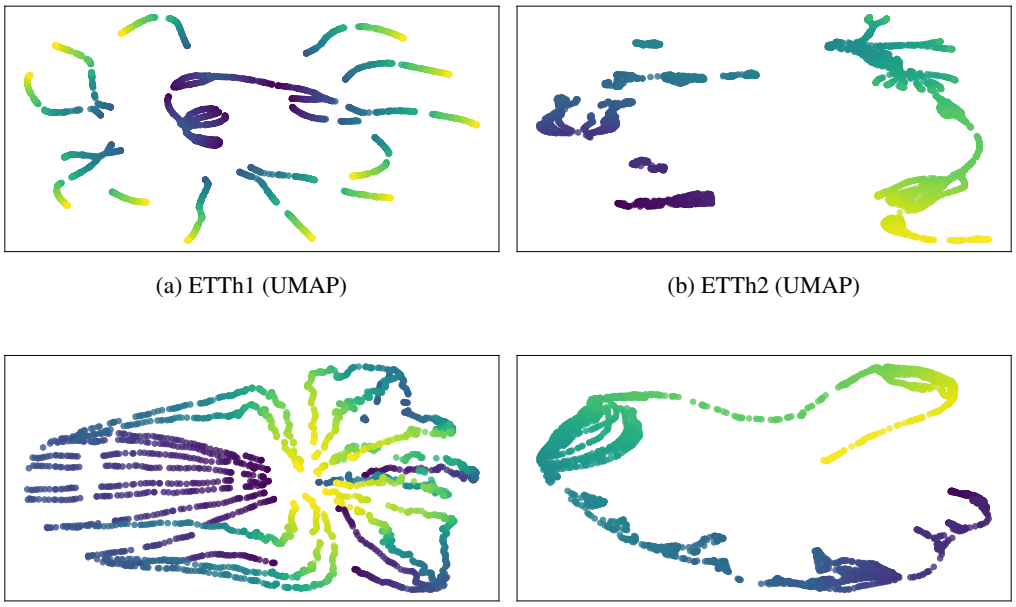
[34] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

A Further latent space visualizations

We visualize representations learned with OliO across three forecasting backbones using UMAP and PHATE. Colors indicate the time indices of historical windows, allowing temporal organization to be inspected within each two-dimensional projection. These qualitative views accompany the forecasting and representation-quality evaluations in Secs. 4 and 4.3.

A.1 Visualizations for the PatchTST backbone

Figs. 5 and 6 show PatchTST representations on ETT datasets. The PHATE projections display connected trajectories with gradual changes in time-index color, while UMAP provides a complementary view of the representation geometry.



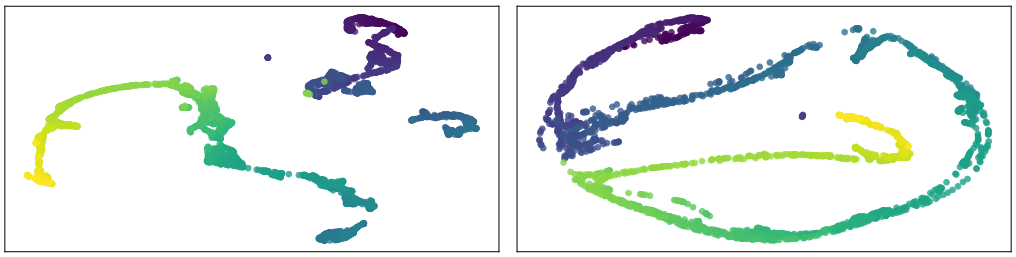
(a) ETTh1 (UMAP)

(b) ETTh2 (UMAP)

(c) ETTh1 (PHATE)

(d) ETTh2 (PHATE)

Figure 5: Latent space visualizations of the PatchTST backbone on ETTh1 and ETTh2 datasets using UMAP (top) and PHATE (bottom). For ETTh2, we utilize the causal pretraining configuration. The continuous trajectories in the PHATE plots demonstrate that OliO effectively preserves the temporal flow.

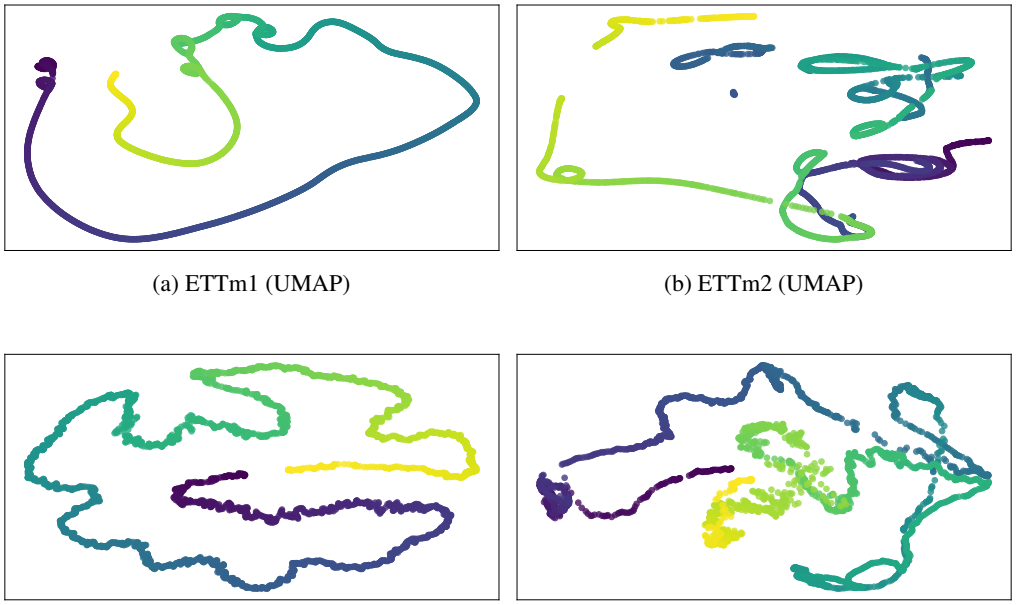


(a) ETTm1 (UMAP) (b) ETTm1 (PHATE)

Figure 6: Latent space visualization of PatchTST on the ETTm1 dataset, comparing UMAP and PHATE embeddings.

A.2 Visualizations for the iTransformer backbone

Fig. 7 shows iTransformer representations on ETTm1 and ETTm2. Both projections exhibit trajectories with temporal color progression, with differences in shape and overlap between the datasets and projection methods.



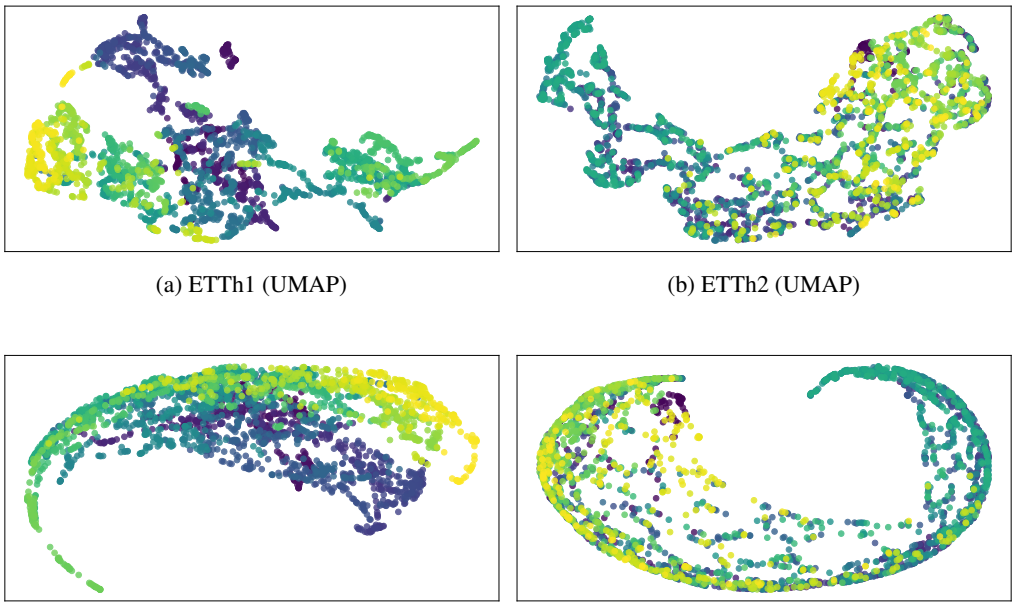
(a) ETTm1 (UMAP) (b) ETTm2 (UMAP)

(c) ETTm1 (PHATE) (d) ETTm2 (PHATE)

Figure 7: Latent space visualizations of iTransformer on ETTm1 and ETTm2 datasets.

A.3 Visualizations for the TimeVLM backbone

Figs. 8–10 show TimeVLM representations on ETT and Weather. Several projections contain overlapping regions and less distinct temporal color progression than the PatchTST and iTransformer examples. These panels document the geometry of the learned multimodal representations; the forecasting results provide the corresponding quantitative evaluation.



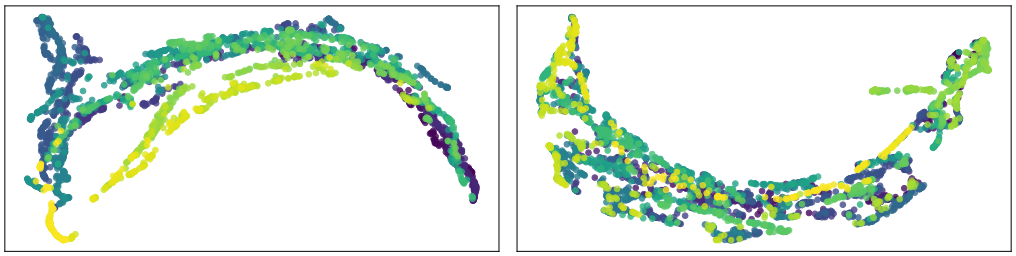
(a) ETTh1 (UMAP)

(b) ETTh2 (UMAP)

(c) ETTh1 (PHATE)

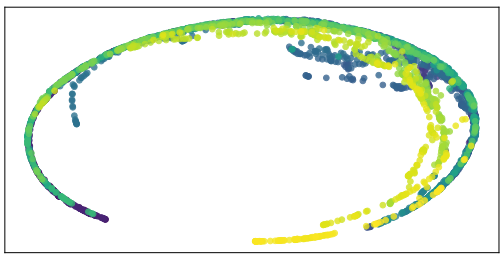
(d) ETTh2 (PHATE)

Figure 8: Latent space visualizations of TimeVLM on ETTh1 and ETTh2 datasets.

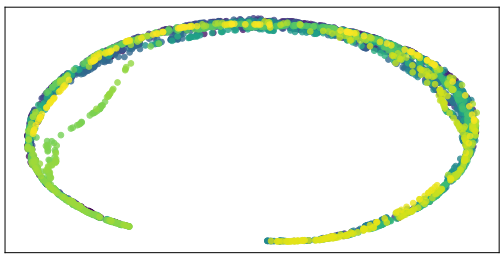


(a) ETTm1 (UMAP)

(b) ETTm2 (UMAP)

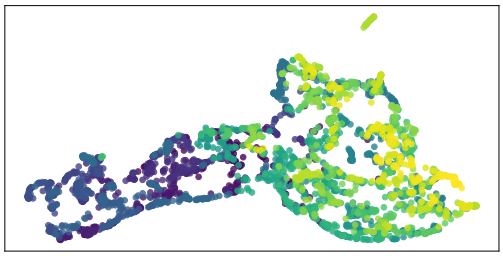


(c) ETTm1 (PHATE)

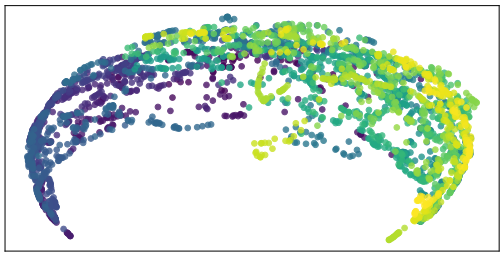


(d) ETTm2 (PHATE)

Figure 9: Latent space visualizations of TimeVLM on ETTm1 and ETTm2 datasets.



(a) Weather (UMAP)



(b) Weather (PHATE)

Figure 10: Latent space visualization of TimeVLM on the Weather dataset, comparing UMAP and PHATE embeddings.

B Mathematical proofs and transition analysis

We provide proofs of the finite-polynomial transition structure and its displacement properties, followed by a comparison with the spectral behavior of a general linear generator.

B.1 Proof of Theorem 1 and polynomial structure

For Theorem 1, let $q \leq d_h$ denote the nilpotency index of K.

Proof. According to the Taylor series expansion of the matrix exponential, we have:

$$e^{\tau K} = I + \tau K + \frac{\tau^2 K^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{(\tau K)^n}{n!} \tag{8}$$

A strictly upper-triangular matrix satisfies $K^{d_h} = 0$. By the definition of q, all powers K^n with $n \geq q$ vanish. Consequently, the infinite series truncates to a finite sum:

$$e^{\tau K} = \sum_{n=0}^{q-1} \frac{\tau^n}{n!} K^n \tag{9}$$

Since $K^0 = I$ and every positive power of K is strictly upper triangular, $e^{\tau K}$ is upper triangular with unit diagonal. Every entry is a polynomial of degree at most $q-1$ in τ. Adding $\tau \boldsymbol{v}$ makes the complete transition a vector-valued polynomial of degree at most $\max\{q-1, 1\}$ for fixed $\boldsymbol{h}$ and $\boldsymbol{v}$. The backward case follows by replacing τ with $-\tau$, including the offset $-\tau\boldsymbol{v}$; the unit-diagonal property and degree bounds are unchanged. This proves both directions of Theorem 1. □

Remark 1 (Implemented signed offset)**.** *In Eqs.* (2) *and* (3)*, OliO directly parameterizes the offset as $\pm\tau\boldsymbol{v}$ using a single shared learnable vector $\boldsymbol{v}$. The offset is linear in lag, so the complete transition retains the finite-polynomial dependence established by Theorem 1.*

Lemma 2 (Squared displacement of a fixed source)**.** *For fixed $\boldsymbol{h}$ and $\boldsymbol{v}$, the squared displacement $\|\mathcal{T}_{\pm\tau}(\boldsymbol{h}) - \boldsymbol{h}\|_2^2$ is a polynomial in τ of degree at most $2\max\{q-1, 1\}$ for either choice of sign.*

Proof. Let $P(\tau) = e^{\tau K} - I$. By Theorem 1, $P(\tau)$ is a polynomial matrix. Including the implemented offset gives

$$\|\mathcal{T}_\tau(\boldsymbol{h}) - \boldsymbol{h}\|_2^2 = \|P(\tau)\boldsymbol{h} + \tau\boldsymbol{v}\|_2^2. \tag{10}$$

The displacement is a vector-valued polynomial of degree at most $\max\{q-1, 1\}$. Squaring and summing its entries gives the stated degree bound. Replacing τ with $-\tau$ gives the backward case with the same degree bound. □

These identities characterize displacement of a fixed source. The training objective aligns independently encoded endpoint representations through cosine similarity, leaving their Euclidean norms unconstrained.

B.2 Local separation and asymptotic growth

For analysis, extend the transition to a real lag $s \in \mathbb{R}$ by $\mathcal{T}_s(\boldsymbol{h}) = e^{sK}\boldsymbol{h} + s\boldsymbol{v}$. This extension agrees with the integer-lag formulation used in pre-training. For fixed $\boldsymbol{h}$ and $\boldsymbol{v}$, define

$$\boldsymbol{d}_h(s) = (e^{sK} - I)\boldsymbol{h} + s\boldsymbol{v}, \qquad D_h(s) = \|\boldsymbol{d}_h(s)\|_2^2. \tag{11}$$

Proposition 3 (Local displacement separation)**.** *If $K\boldsymbol{h} + \boldsymbol{v} \neq 0$, there exists $\epsilon > 0$ such that $D_h(s)$ is strictly increasing for $0 < s < \epsilon$.*

Proof. Expansion at zero gives $\boldsymbol{d}_h(s) = s(K\boldsymbol{h} + \boldsymbol{v}) + O(s^2)$, and hence

$$D_h(s) = s^2\|K\boldsymbol{h} + \boldsymbol{v}\|_2^2 + O(s^3). \tag{12}$$

Since D_h is polynomial, its derivative is $2s\|K\boldsymbol{h} + \boldsymbol{v}\|_2^2 + O(s^2)$, which is positive for sufficiently small positive s. □

The condition holds almost surely for any state distribution assigning zero probability to the affine stationary set $\{\boldsymbol{h} : K\boldsymbol{h} + \boldsymbol{v} = 0\}$.

Proposition 4 (Asymptotic displacement growth)**.** *Suppose $\boldsymbol{d}_h(s)$ has positive degree r and leading vector coefficient $\boldsymbol{c}_r \neq 0$. Then*

$$\|\boldsymbol{d}_h(s)\|_2 = \|\boldsymbol{c}_r\|_2 s^r(1 + o(1)) \quad \text{as } s \to +\infty, \tag{13}$$

and its norm is eventually strictly increasing.

Proof. Factoring out s^r yields $s^{-r}\boldsymbol{d}_h(s) \to \boldsymbol{c}_r$, giving the norm asymptotic. The squared norm has leading term $\|\boldsymbol{c}_r\|_2^2 s^{2r}$, whose derivative is eventually positive. The norm is then positive and strictly increasing as well. □

These results describe local separation and eventual growth. Behavior at intermediate lags depends on the learned coefficients. App. I.6 reports finite-range measurements.

B.3 Algebraic structure and spectral behavior

Corollary 5 (Spectral comparison)**.** *For a general linear generator, each Jordan block contributes an exponential factor $e^{\tau\lambda_i}$ multiplied by a finite polynomial. A nonzero real part of λ_i introduces a mode growing exponentially in one of the two signed temporal directions. For purely imaginary eigenvalues, oscillatory factors may be accompanied by polynomial terms from nontrivial Jordan blocks.*

Proof. Using Jordan normal form over $\mathbb{C}$, write $K = PJP^{-1}$ with blocks $J_i = \lambda_i I + N_i$ of size n_i, where N_i is nilpotent. The block exponential is

$$e^{\tau J_i} = e^{\tau(\lambda_i I + N_i)} = e^{\tau\lambda_i} e^{\tau N_i} = e^{\tau\lambda_i} \left(\sum_{k=0}^{n_i - 1} \frac{\tau^k}{k!} N_i^k \right) \tag{14}$$

Case 1: Positive real part. If $\mathrm{Re}(\lambda_i) > 0$, the factor $|e^{\tau\lambda_i}| = e^{\tau\,\mathrm{Re}(\lambda_i)}$ grows exponentially for positive lag.

Case 2: Negative real part. For $\mathrm{Re}(\lambda_i) < 0$, the block decays in the forward direction because exponential decay dominates its polynomial factor. Writing $\alpha = -\mathrm{Re}(\lambda_i) > 0$ gives

$$\lim_{\tau\to\infty} \frac{\tau^k}{e^{\alpha\tau}} = \lim_{\tau\to\infty} \frac{k\tau^{k-1}}{\alpha e^{\alpha\tau}} = \cdots = \lim_{\tau\to\infty} \frac{k!}{\alpha^k e^{\alpha\tau}} = 0 \tag{15}$$

The same mode grows in the backward direction because reversing the lag reverses the sign of the real part in the exponential factor.

Case 3: Zero real part. For $\lambda_i = i\omega$, the exponential factor has unit magnitude and can be oscillatory. A nontrivial Jordan block adds polynomial factors, so zero real parts alone do not guarantee bounded transport. Nonzero imaginary eigenvalues do not yield the finite-polynomial matrix dependence induced by nilpotency.

Strict upper triangularity supplies nilpotency and exact finite-polynomial transport in both signed directions. The magnitude of the resulting polynomial transients depends on the learned coefficients and the evaluated lag. □

B.4 Computational complexity and polynomial truncation

We analyze the cost of evaluating the transition and the approximation introduced by truncating its polynomial expansion.

Taylor expansion and nilpotency. The continuous transition operator $e^{\tau K}$ is defined by the infinite Taylor series:

$$e^{\tau K} = I + \tau K + \frac{\tau^2 K^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{(\tau K)^n}{n!} \tag{16}$$

By Theorem 1, the nilpotency index satisfies $q \leq d_h$, so the exact series terminates at degree at most $q-1$:

$$e^{\tau K} = \sum_{n=0}^{q-1} \frac{\tau^n}{n!} K^n \tag{17}$$

Polynomial truncation. For $p < q-1$, an approximation retains terms through order p:

$$\hat{S}_p(\tau) = \sum_{n=0}^{p} \frac{(\tau K)^n}{n!} \tag{18}$$

To evaluate this polynomial efficiently, we apply **Horner's method** [10], which reformulates the sum into a nested form. For the matrix exponential, the p-th order approximation can be computed as:

$$\text{Horner}(M, p) = I + M\left(I + \frac{M}{2}\left(I + \frac{M}{3}\cdots\left(I + \frac{M}{p}\right)\cdots\right)\right) \tag{19}$$

where $M = \tau K$ for the forward direction. This nested evaluation uses p matrix–matrix multiplications instead of explicitly forming each power separately [9, 19]. For $p = 0$, the polynomial is simply I. Alg. 1 summarizes the matrix-level computation, with cost $O(pd_h^3)$.

Applying the polynomial directly to $\boldsymbol{h}$ requires only matrix–vector products, giving cost $O(pd_h^2)$ plus $O(d_h)$ for the signed offset, as detailed in Alg. 2. Exact evaluation uses $p = q-1$ and remains cubic in the worst case $q = d_h$; using a fixed lower order reduces cost while approximating the complete transition.

Algorithm 1 Matrix exponential approximation via Horner's method

Input: Matrix $M \in \mathbb{R}^{d_h \times d_h}$, truncation order p
Output: Approximated transition operator $\hat{S}_p \in \mathbb{R}^{d_h \times d_h}$
$\hat{S}_p \leftarrow I_{d_h \times d_h}$ {Initialize with identity matrix}
for $i = p$ **down to** 1 **do**
    $\hat{S}_p \leftarrow I_{d_h \times d_h} + \frac{1}{i} M \cdot \hat{S}_p$ {$\mathcal{O}(d_h^3)$ per iteration}
end for
return $\hat{S}_p$

Algorithm 2 Efficient representation transition via vector-wise Horner's method

Input: Representation $\boldsymbol{h}$, matrix $M = \tau K$, order p, lag $\tau \in \mathbb{N}$, shared offset $\boldsymbol{v}$
Output: Transitioned representation $\hat{\boldsymbol{h}}_\tau \in \mathbb{R}^{d_h}$
$\boldsymbol{y} \leftarrow \boldsymbol{h}$ {Initialize accumulation vector}
for $i = p$ **down to** 1 **do**
    $\boldsymbol{y} \leftarrow \boldsymbol{h} + \frac{1}{i} M\boldsymbol{y}$ {$\mathcal{O}(d_h^2)$ per iteration}
end for
$\hat{\boldsymbol{h}}_\tau \leftarrow \boldsymbol{y} + \tau\boldsymbol{v}$ {Add shared signed offset}
return $\hat{\boldsymbol{h}}_\tau$

Alg. 2 shows the forward evaluation. The backward evaluation uses $M = -\tau K$ and adds $-\tau\boldsymbol{v}$ at the final step; the operation count is unchanged.

Effect of truncation. Truncation omits $E_p(\tau) = \sum_{n=p+1}^{q-1} \tau^n K^n / n!$. The generator can be trained within the truncated parameterization, but optimization does not guarantee recovery of the omitted terms or preservation of forecasting accuracy. The order ablation in Sec. 4.4 evaluates this accuracy–cost trade-off separately from the complete transition.



Figure 11: Speedup of the OliO module's p-th order approximation relative to the full matrix exponential (computed via `torch.matrix_exp`). The benchmark evaluates the computational efficiency of the `TriangularCausalShift` module using 4,096 samples per operation, with values averaged over 5 independent trials on an NVIDIA RTX 3090 GPU. The first-order approximation ($p = 1$) provides an acceleration of approximately $3.5\times$ to $4.5\times$ compared to the exact computation, highlighting the high-throughput capability of OliO in processing high-dimensional temporal representations.

Empirical transition-computation efficiency. Fig. 11 reports a dimension-wise benchmark of the core `OliO` transition module (`TriangularCausalShift`) relative to `torch.matrix_exp`. The benchmark uses 4,096 samples per operation, latent dimensions $\{64, 128, 256, 512\}$, and five independent trials on an NVIDIA RTX 3090 GPU. Each speedup is relative to the full matrix exponential at the same dimension. Speedup varies with representation size and decreases as more polynomial terms are retained. Fig. 4 gives the order-wise summary in the main text. These measurements characterize transition computation during pre-training. Downstream forecasting uses the original backbone after the projection and transition modules have been discarded.

C Datasets

We describe the six benchmark datasets used for full-shot and few-shot forecasting. Table 11 summarizes their characteristics. The eight downstream datasets used in the LightGTS zero-shot evaluation are reported in App. E.2.

Table 11: Statistics of the benchmark datasets used in our forecasting experiments.

Dataset	Channels	Sampling Rate	Time Steps	Train/Val/Test Split	Features
ETTh1, h2	7	1 hour	14,307	8,545 / 2,881 / 2,881	Transformer oil temperature
ETTm1, m2	7	15 min	57,507	34,465 / 11,521 / 11,521	Transformer oil temperature
Weather	21	10 min	52,603	36,792 / 5,271 / 10,540	Meteorological indicators
ECL	321	1 hour	26,211	18,317 / 2,633 / 5,261	Power consumption

C.1 Textual prompts for TimeVLM

For all methods using the TimeVLM backbone, we use the textual prompts supplied by its official implementation without modification. The quotations below document these inherited text inputs, while Table 11 describes the benchmark data used for forecasting.

ETT (ETTh1, ETTh2, ETTm1, ETTm2)

> *"The Electricity Transformer Temperature (ETT) is a crucial indicator in the electric power long-term deployment. This dataset consists of 2 years data from two separated counties in China. To explore the granularity on the Long sequence time-series forecasting (LSTF) problem, different subsets are created, {ETTh1, ETTh2} for 1-hour-level and ETTm1 for 15-minutes-level. Each data point consists of the target value "oil temperature" and 6 power load features. The train/val/test is 12/4/4 months."*

Weather

> *"Weather is recorded every 10 minutes for the 2020 whole year, which contains 21 meteorological indicators, such as air temperature, humidity, etc."*

Electricity (ECL)

> *"Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available. This archive contains 2075259 measurements gathered in a house located in Sceaux (7km of Paris, France) between December 2006 and November 2010 (47 months)."*

The ECL experiments use the 321-channel hourly benchmark listed in Table 11. The Electricity quotation above is retained as part of the inherited TimeVLM prompt configuration.

D Implementation details

In this section, we provide the detailed experimental setup, including the hardware environment, software framework, loss functions, and hyperparameters used for our experiments.

D.1 Experimental setup

All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU with 24GB of memory. We implemented our framework using the PyTorch library. For all downstream forecasting tasks, we employed the L_2 loss for training. The performance was evaluated using two standard metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE), defined as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{20}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{21}$$

where y_i and $\hat{y}_i$ denote the ground truth and the predicted values, respectively, and n is the total number of data points.

Training, validation, and test periods are chronologically separated. Pre-training pairs are drawn from the training period and exclude held-out forecasting targets. We use seeds 2023, 2024, and 2025 for three independent runs; tables report means and sample standard deviations.

D.2 Self-supervised methods

We compared OliO with several state-of-the-art self-supervised learning (SSL) methods, which can be broadly categorized into two groups: Masked and Reconstruction-based methods, and Contrastive Learning methods. Table 12 summarizes the compared methods.

Table 12: Summary of self-supervised learning methods compared in our experiments.

Category	Method	Key Feature
Masked & Reconstruction	TimeSiam [6]	Siamese modeling with temporal masking
	SimMTM [5]	Masked time-series modeling with manifold learning
Contrastive Learning	TimesURL [14]	Universal relational learning for time series
	SoftCLT [12]	Soft contrastive learning for time series
	TS2Vec [32]	Hierarchical contrastive learning on timestamp level

Each SSL objective uses the representation interface available in the selected backbone. For iTransformer [15] and TimeVLM [33], temporal loss components requiring a timestamp axis are omitted when the latent representation has no such axis.

PatchTST retains a temporal latent axis compatible with timestamp-indexed objectives. The other backbones use their available latent interfaces under the same plug-and-play protocol. The evaluated TimeSiam implementation retains temporally separated subseries, Siamese encoding, masking, lineage embeddings, its cross-attention decoder, and past-to-current reconstruction. These configurations define the shared-backbone comparisons reported in Sec. 4.

D.3 Finetuning and hyperparameter search

After pre-training, we fully finetune each backbone for forecasting. For each method–backbone–dataset–horizon setting, we select one configuration by validation MSE from the predefined search space. The selected configuration is fixed before three-seed test evaluation; test performance is excluded from hyperparameter and checkpoint selection. Baseline-specific search spaces follow the protocols documented in the original papers. Specifically, for TimeSiam, we utilized the settings reported in Dong et al. [6, App. A.2, Table 10], which specify a learning rate of 1×10^{-4} and a dropout search space of $\{0, 0.1, 0.2, 0.3\}$. Similarly, for SimMTM, we adopted the configuration from Dong et al. [5, App. A.4, Table 10] with a learning rate of 1×10^{-4} and a dropout rate search space of $\{0, 0.1, 0.2, 0.3\}$. For SoftCLT, we used a learning rate of 1×10^{-3} and searched for the optimal dropout rate within $\{0, 0.1, 0.2, 0.3\}$, as detailed in Lee et al. [12, App. C, Table C.1]. For other SSL methods and backbone models, validation selection uses the following search space:

- Learning rate: $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 1 \times 10^{-5}\}$
- Dropout rate: $\{0, 0.1, 0.2, 0.3\}$

In general, we observed that employing a sequence length for pre-training that is twice that of fine-tuning was more effective for capturing diverse temporal dynamics to model the long-term dependencies. For the fine-tuning stage, a learning rate of 1×10^{-4} and a dropout rate of 0.0 usually yielded the better performance across most benchmarks. Regarding the projection dimension (d_h), as no clear performance trend was observed during our preliminary grid search, we followed common practices in contrastive learning and utilized dimensions within the range of $128 \sim 256$.

For OliO pre-training with the iTransformer backbone, we also performed a grid search to determine the optimal hyperparameters for each dataset. The search space included `projection_dim` $\in \{128, 256\}$, `learning_rate` $\in \{1 \times 10^{-3}, 7 \times 10^{-4}\}$, `train_epochs` $\in \{10, 20\}$, and `max_tau` $\in \{8, 32, 64\}$. The dataset-specific settings were as follows:

- Electricity: `seq_len` $\in \{96, 192\}$, `batch_size` $= 16$.
- ETTh1: `seq_len` $\in \{96, 192\}$, `batch_size` $= 32$.
- ETTh2, ETTm1, ETTm2: `seq_len` $\in \{336, 672\}$, `batch_size` $= 32$.
- Weather: `seq_len` $\in \{336, 672\}$, `batch_size` $= 16$.

For OliO pre-training with the TimeVLM backbone, the search space for `projection_dim`, `learning_rate`, `train_epochs`, and `max_tau` was kept identical to the iTransformer experiments. The dataset-specific configurations were as follows:

- Electricity: `seq_len` $\in \{512, 1024\}$, `batch_size` $= 16$.
- ETTh1, ETTh2, ETTm1, ETTm2, and Weather: `seq_len` $\in \{512, 1024\}$, `batch_size` $= 32$.

For OliO pre-training with the PatchTST backbone, the search space for the common hyperparameters was also maintained as described above. The dataset-specific settings were as follows:

- Electricity: `seq_len` $\in \{336, 672\}$, `batch_size` $= 8$.
- ETTh1, ETTh2, ETTm1, ETTm2, and Weather: `seq_len` $\in \{336, 672\}$, `batch_size` $= 32$.

D.4 Computational cost

Each individual experimental session (including pre-training and forecasting fine-tuning for a single dataset) typically takes approximately one hour on a single NVIDIA RTX 3090 GPU.

D.5 Visualization settings

Dimensionality-reduction tools such as t-SNE [27], UMAP [17], and PHATE [21] provide low-dimensional views of learned representations. We use UMAP and PHATE to obtain complementary two-dimensional projections, with points colored by the time index of each historical window. App. A presents the projections alongside the forecasting and representation-quality evaluations in Secs. 4 and 4.3.

D.6 Baseline model architectures

In our experiments, we evaluated OliO across three representative backbone architectures, each highlighting different structural paradigms in time-series modeling. For all models, we followed the official hyperparameter settings (e.g., embedding dimension, number of layers, and feed-forward network dimension) as specified in their respective repositories.

- **PatchTST** [22]: A representative patch-based, channel-independent model. By segmenting time series into sub-series patches, it captures local semantic information while maintaining computational efficiency. Following the official implementation, the patch size was set to 12.
  - Repository: `https://github.com/yuqinie98/patchtst`
  - Settings: seq_len $= 336$, pred_len $\in \{96, 192, 336, 720\}$.
- **iTransformer** [15]: A channel-dependent architecture that treats each channel as a single token. This approach leverages the multi-head attention mechanism to explicitly capture inter-channel dependencies, making it effective for multivariate forecasting.
  - Repository: `https://github.com/thuml/iTransformer`
  - Settings: seq_len $= 96$, pred_len $\in \{96, 192, 336, 720\}$.

Table 13: Frozen-backbone linear probing results. Each dataset value averages four forecasting horizons. Lower MSE is better.

Backbone	Dataset	TimesURL	OliO
PatchTST	Electricity	0.1846 ± 0.0012	$\mathbf{0.1833 \pm 0.0004}$
	ETTh1	0.4147 ± 0.0021	$\mathbf{0.4107 \pm 0.0009}$
	ETTh2	$\mathbf{0.3409 \pm 0.0006}$	0.3461 ± 0.0011
	ETTm1	$\mathbf{0.3546 \pm 0.0009}$	0.3616 ± 0.0012
	ETTm2	0.2550 ± 0.0006	$\mathbf{0.2528 \pm 0.0005}$
	Weather	0.2469 ± 0.0015	$\mathbf{0.2439 \pm 0.0005}$
	Average	$\mathbf{0.2994 \pm 0.0011}$	0.2997 ± 0.0008
iTransformer	Electricity	0.2266 ± 0.0004	$\mathbf{0.2161 \pm 0.0003}$
	ETTh1	$\mathbf{0.4535 \pm 0.0011}$	0.4536 ± 0.0011
	ETTh2	$\mathbf{0.3796 \pm 0.0013}$	0.3811 ± 0.0012
	ETTm1	0.4176 ± 0.0002	$\mathbf{0.4108 \pm 0.0006}$
	ETTm2	0.2896 ± 0.0005	$\mathbf{0.2862 \pm 0.0007}$
	Weather	0.2805 ± 0.0005	$\mathbf{0.2738 \pm 0.0005}$
	Average	0.3412 ± 0.0007	$\mathbf{0.3369 \pm 0.0007}$
Overall average		0.3203	**0.3183**

- **TimeVLM** [33]: A state-of-the-art multimodal model that integrates vision, text, and time-series data. It aligns temporal features with natural language descriptions to enhance predictive performance.
  - Repository: `https://github.com/CityMind-Lab/ICML25-TimeVLM`
  - Settings: seq_len $= 512$, pred_len $\in \{96, 192, 336, 720\}$.

All backbones use the same prediction horizons. Each backbone retains its standard preprocessing and normalization across the compared SSL methods.

E Additional representation quality and transferability

E.1 Frozen-backbone linear probing

We provide the detailed frozen-backbone linear probing results corresponding to Sec. 4.3. After self-supervised pre-training, all backbone parameters are frozen and only an identical linear forecasting head is trained. We compare OliO with TimesURL using PatchTST and iTransformer across six datasets and four prediction horizons. Each reported dataset value averages the four forecasting horizons. The deviation shown after $\pm$ is the average of their three-seed sample standard deviations, rather than the standard deviation of the horizon-averaged score.

Across the 48 backbone–dataset–horizon configurations, OliO achieves lower MSE than TimesURL in 35 settings. When the four horizons are aggregated at the dataset level, OliO performs better in 8 of 12 backbone–dataset combinations. PatchTST shows comparable overall performance between the two methods, whereas OliO provides a clearer improvement on iTransformer, reducing the average MSE from 0.3412 to 0.3369. Overall, the aggregate MSE decreases from 0.3203 to 0.3183.

E.2 Zero-shot performance with LightGTS-4M

We follow the LightGTS-4M [29] zero-shot protocol, preserving its architecture, forecasting hyperparameters, and pre-training data. A single pretrained checkpoint is evaluated on eight downstream datasets disjoint from pre-training, with four forecasting horizons per dataset. Downstream evaluation uses the pretrained parameters directly, without parameter updates or checkpoint selection. We compare TimesURL, the default LightGTS mask-reconstruction objective, and OliO across the resulting 32 settings.

Table 14: Full zero-shot forecasting results with the LightGTS-4M time-series foundation model [29]. Mask Recon. denotes its default mask-reconstruction pre-training. Best results are in **bold** and the second best are underlined. Ranks are determined from the reported values rounded to four decimal places. If multiple methods share the best reported value, no method is counted as Top-1 and all tied methods are treated as Top-2. Lower MSE is better.

Dataset		LightGTS-4M		
		TimesURL	Mask Recon.	**OliO**
ETTh1	96	.3578	.3490	**.3481**
	192	.3967	.3902	**.3859**
	336	.4213	.4115	**.4061**
	720	.4626	**.4324**	.4345
ETTh2	96	.2814	.2802	**.2758**
	192	.3322	.3353	**.3309**
	336	.3592	.3632	.3592
	720	.4274	.4110	**.4022**
ETTm1	96	.3083	**.2884**	.2905
	192	.3287	.3166	**.3164**
	336	.3526	.3423	**.3417**
	720	.3947	.3866	**.3828**
ETTm2	96	.1863	**.1650**	.1704
	192	.2345	**.2205**	.2206
	336	.2747	.2677	**.2655**
	720	.3309	.3311	**.3277**
Weather	96	.1587	**.1420**	.1450
	192	.2063	**.1818**	.1867
	336	.2587	**.2305**	.2381
	720	.3190	**.2994**	.3028
ECL	96	.1859	.1684	**.1650**
	192	.2264	.2172	**.1994**
	336	.2681	.2698	**.2423**
	720	.3421	.3349	**.3232**
Solar	96	.1560	.1553	**.1453**
	192	.1668	.1684	**.1539**
	336	.1741	.1777	**.1656**
	720	**.1746**	.1835	.1798
Traffic	96	.4561	.4069	**.4037**
	192	.5166	.4897	**.4549**
	336	.5672	.5761	**.5077**
	720	.6760	.6728	**.6377**
Average		.3219	.3114	**.3034**

Table 14 reports horizon-level MSE. The eight-dataset average is 0.3034 for OliO, 0.3114 for mask reconstruction, and 0.3219 for TimesURL. The main text summarizes these same datasets after averaging over the four horizons.

Additional error metric. Across the eight datasets, average MAE is 0.3469 for mask reconstruction and 0.3391 for OliO. Dataset-average MAE also decreases from 0.348 to 0.335 on Electricity, from 0.254 to 0.232 on Solar, and from 0.420 to 0.399 on Traffic. These dataset averages cover the four forecasting horizons and are rounded to three decimal places.

F Few-shot forecasting results

We evaluate forecasting performance using limited labeled training data. Following the sampling protocol of TimeVLM [33], we construct few-shot training sets by randomly selecting 1%, 5%, and 10% of the original training data. These experiments assess the pretrained representations when labeled data for downstream forecasting are scarce.

F.1 Summary of few-shot results

Figs. 12, 13, and 14 summarize the relative MSE across the three few-shot ratios. The relative performance index is the mean MSE of OliO divided by the mean MSE of the corresponding method, each averaged over the same 24 dataset–horizon settings, with OliO fixed at 1.0. Values above 1.0 indicate lower MSE than OliO, while values below 1.0 indicate higher MSE.

The following tables report per-setting means and standard deviations. Ranks use the means rounded to four decimal places, and ties receive the last occupied rank (e.g., $2, 2, 3$ or $1, 3, 3$). Bold and underlined entries denote ranks 1 and 2, respectively; Top-1 and Top-2 counts tally these entries.

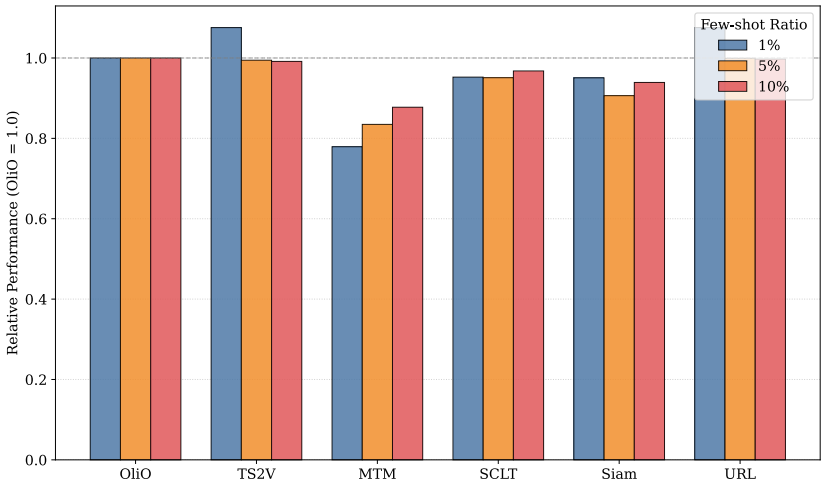


Figure 12: Summary of relative few-shot forecasting performance for **PatchTST** backbone across 1%, 5%, and 10% settings. Performance is normalized relative to OliO ($= 1.0$), where higher values indicate better forecasting quality (lower MSE).

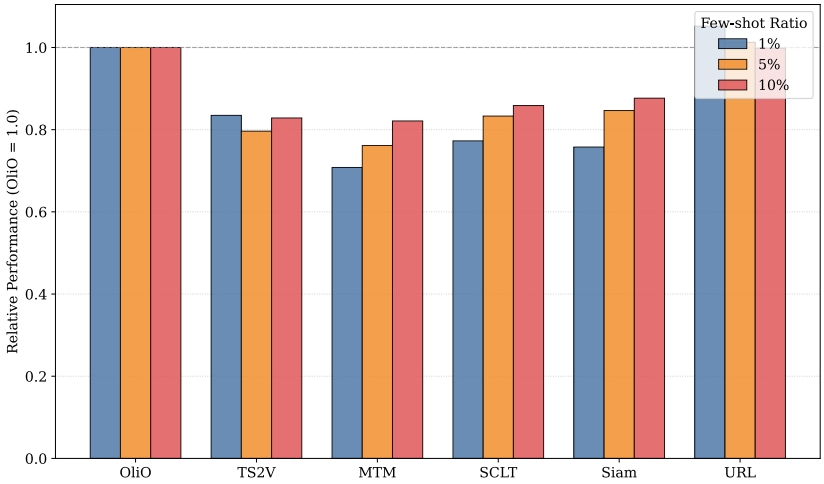


Figure 13: Summary of relative few-shot forecasting performance for **iTransformer** backbone across 1%, 5%, and 10% settings. Performance is normalized relative to OliO (= 1.0), where higher values indicate better forecasting quality (lower MSE).

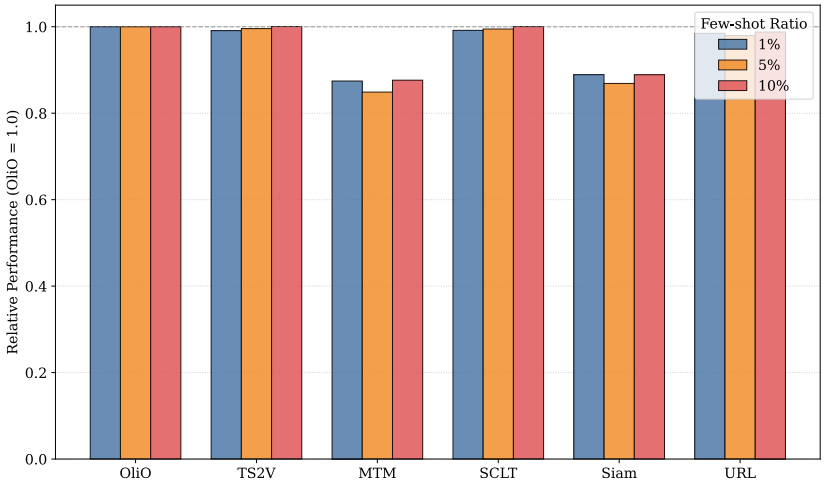


Figure 14: Summary of relative few-shot forecasting performance for **TimeVLM** backbone across 1%, 5%, and 10% settings. Performance is normalized relative to OliO (= 1.0), where higher values indicate better forecasting quality (lower MSE).

F.2 PatchTST few-shot results

Table 15: 1% few-shot forecasting results for PatchTST [22] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 16.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.6167±.0088	.5307±.0033	**.4697±.0042**	**.4594±.0022**	.7179±.0101	.5701±.0036
	192	.6294±.0090	.5401±.0042	**.4979±.0046**	**.4768±.0023**	.7295±.0031	.5781±.0013
	336	.6262±.0079	.5453±.0038	**.5049±.0029**	**.4829±.0014**	.7251±.0019	.5833±.0004
	720	.6198±.0052	.5631±.0017	**.5134±.0072**	**.5058±.0035**	.7210±.0030	.5999±.0016
ETTh2	96	.3471±.0010	.3923±.0026	**.3290±.0037**	**.3788±.0005**	.3931±.0003	.4246±.0005
	192	.3888±.0080	.4205±.0038	**.3790±.0092**	**.4135±.0054**	.4114±.0006	.4371±.0004
	336	.3919±.0006	.4288±.0009	**.3857±.0010**	**.4225±.0006**	.4029±.0011	.4356±.0010
	720	.4312±.0029	.4557±.0015	**.4198±.0001**	**.4472±.0004**	.4399±.0014	.4602±.0009
ETTm1	96	.3393±.0061	.3793±.0024	.3330±.0047	.3760±.0024	.6701±.0001	.5387±.0006
	192	.3679±.0050	.3954±.0029	.3682±.0002	.3950±.0004	.6785±.0002	.5461±.0008
	336	.3958±.0043	.4111±.0025	.3971±.0003	.4116±.0017	.6933±.0003	.5526±.0003
	720	.4503±.0030	.4463±.0027	**.4471±.0003**	**.4424±.0028**	.7117±.0004	.5656±.0005
ETTm2	96	**.1774±.0021**	**.2731±.0015**	.1809±.0032	.2745±.0026	.2716±.0005	.3416±.0004
	192	.2381±.0051	**.3107±.0041**	**.2379±.0027**	.3113±.0041	.3041±.0012	.3606±.0008
	336	**.2880±.0012**	**.3404±.0012**	.2932±.0043	.3462±.0028	.3403±.0012	.3797±.0008
	720	.3797±.0017	.3974±.0007	.3825±.0004	.3981±.0020	.4222±.0009	.4236±.0006
Weather	96	.2264±.0013	.2843±.0010	.1924±.0063	.2474±.0064	.2571±.0002	.3122±.0004
	192	.2550±.0029	.3010±.0032	.2245±.0037	.2705±.0031	.2846±.0006	.3285±.0006
	336	.2954±.0010	.3286±.0012	.2750±.0024	.3059±.0018	.3174±.0003	.3478±.0003
	720	.3510±.0016	.3643±.0023	.3380±.0005	.3490±.0004	.3681±.0001	.3796±.0002
ECL	96	.1655±.0007	.2701±.0007	.1755±.0007	.2819±.0007	.2661±.0009	.3671±.0006
	192	.1792±.0005	.2813±.0004	.1901±.0004	.2942±.0006	.2783±.0014	.3762±.0130
	336	.1957±.0006	.2959±.0006	.2068±.0004	.3085±.0003	.2937±.0013	.3875±.0170
	720	.2337±.0004	.3258±.0005	.2442±.0010	.3372±.0005	.3258±.0022	.4085±.0024
Top1 count		2	3	10	9	0	0
Top2 count		9	3	4	5	0	0

Table 16: 1% few-shot forecasting results for PatchTST [22] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 15.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.6120±.0147	.5269±.0061	.7105±.0029	.5661±.0015	.4884±.0035	.4722±.0021
	192	.6315±.0087	.5385±.0037	.7178±.0004	.5739±.0009	.5192±.0034	.4902±.0017
	336	.6290±.0054	.5434±.0029	.7116±.0040	.5776±.0019	.5264±.0029	.4962±.0013
	720	.6291±.0074	.5617±.0027	.7095±.0018	.5939±.0011	.5328±.0088	.5170±.0044
ETTh2	96	.3785±.0010	.4157±.0008	.3868±.0030	.4207±.0015	.3354±.0031	.3838±.0013
	192	.4061±.0042	.4345±.0023	.4121±.0019	.4372±.0015	.3883±.0105	.4164±.0064
	336	.4036±.0007	.4372±.0003	.4051±.0020	.4370±.0010	.3982±.0033	.4290±.0010
	720	.4392±.0006	.4606±.0005	.4420±.0024	.4615±.0013	.4276±.0007	.4524±.0011
ETTm1	96	.3835±.0071	.4098±.0011	.3587±.0050	.3953±.0031	**.3250±.0039**	**.3668±.0024**
	192	.4085±.0002	.4245±.0007	.3876±.0022	.4109±.0114	**.3650±.0036**	**.3892±.0025**
	336	.4324±.0043	.4366±.0018	.4195±.0022	.4270±.0011	**.3941±.0020**	**.4055±.0007**
	720	.4801±.0073	.4657±.0051	.4717±.0042	.4559±.0027	.4562±.0029	.4444±.0026
ETTm2	96	.2215±.0398	.3082±.0266	.1912±.0055	.2835±.0058	.1813±.0024	.2739±.0024
	192	.2413±.0035	.3156±.0052	.2444±.0002	.3165±.0023	.2395±.0040	.3150±.0040
	336	.3127±.0222	.3615±.0145	.2911±.0014	.3453±.0013	.2929±.0051	.3488±.0038
	720	.3864±.0037	.4018±.0034	**.3793±.0057**	**.3931±.0039**	.3879±.0023	.4022±.0011
Weather	96	.2281±.0198	.2826±.0197	.1863±.0064	.2411±.0066	**.1752±.0043**	**.2277±.0043**
	192	.2527±.0230	.2977±.0218	.2262±.0050	.2724±.0048	**.2165±.0031**	**.2609±.0024**
	336	.2965±.0144	.3262±.0149	.2703±.0120	.3030±.0011	**.2645±.0010**	**.2955±.0009**
	720	.3425±.0013	.3526±.0012	.3395±.0004	.3503±.0002	**.3341±.0004**	**.3442±.0003**
ECL	96	.1974±.0013	.3038±.0012	.1649±.0009	.2684±.0010	**.1554±.0008**	**.2576±.0010**
	192	.2118±.0006	.3160±.0006	.1796±.0003	.2809±.0004	**.1698±.0004**	**.2698±.0006**
	336	.2294±.0100	.3305±.0010	.1961±.0006	.2958±.0005	**.1864±.0005**	**.2848±.0005**
	720	.2646±.0007	.3558±.0004	.2339±.0008	.3244±.0007	**.2252±.0008**	**.3165±.0004**
Top1 count		0	0	1	1	**11**	**11**
Top2 count		0	0	4	7	7	9

Table 17: 5% few-shot forecasting results for PatchTST [22] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 18.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4168±.0021	.4311±.0009	**.4128±.0017**	**.4258±.0009**	.6222±.0138	.5332±.0055
	192	.4463±.0002	.4482±.0001	**.4446±.0021**	**.4456±.0016**	.6465±.0055	.5461±.0024
	336	.4623±.0020	.4593±.0003	**.4595±.0028**	**.4565±.0010**	.6518±.0005	.5535±.0006
	720	.4641±.0009	.4793±.0002	**.4620±.0011**	**.4775±.0005**	.6685±.0042	.5770±.0015
ETTh2	96	.3077±.0101	.3636±.0061	**.3011±.0054**	**.3580±.0037**	.3894±.0008	.4228±.0009
	192	.3650±.0078	.3996±.0047	.3575±.0017	**.3942±.0018**	.4156±.0010	.4396±.0007
	336	.3855±.0021	.4211±.0026	.3804±.0019	.4171±.0013	.4146±.0023	.4434±.0015
	720	.4090±.0116	.4419±.0009	.4083±.0015	.4410±.0014	.4431±.0028	.4625±.0018
ETTm1	96	**.2952±.0018**	**.3480±.0025**	.2984±.0020	.3487±.0020	.3273±.0006	.3726±.0005
	192	**.3341±.0011**	.3742±.0130	.3363±.0011	**.3730±.0012**	.3606±.0013	.3919±.0006
	336	**.3651±.0024**	.3924±.0130	.3679±.0007	**.3920±.0009**	.3915±.0006	.4090±.0002
	720	**.4147±.0029**	**.4230±.0020**	.4218±.0022	.4250±.0004	.4444±.0010	.4382±.0008
ETTm2	96	**.1678±.0006**	**.2595±.0130**	.1701±.0018	.2622±.0009	.1817±.0028	.2738±.0016
	192	**.2238±.0008**	**.2972±.0160**	.2284±.0015	.3006±.0009	.2380±.0021	.3116±.0036
	336	**.2781±.0010**	**.3325±.0007**	.2836±.0013	.3388±.0009	.2909±.0024	.3455±.0015
	720	**.3676±.0018**	**.3848±.0130**	.3710±.0013	.3899±.0005	.3773±.0024	.3943±.0021
Weather	96	.1618±.0001	.2122±.0007	.1639±.0019	.2137±.0021	.1828±.0015	.2382±.0016
	192	.2071±.0022	.2518±.0025	.2081±.0011	.2524±.0012	.2222±.0018	.2684±.0019
	336	.2559±.0014	.2895±.0150	.2576±.0007	.2898±.0003	.2686±.0017	.3014±.0017
	720	.3267±.0003	.3390±.0004	.3290±.0027	.3395±.0022	.3376±.0010	.3484±.0009
ECL	96	.1429±.0005	.2396±.0005	.1510±.0003	.2522±.0004	.2044±.0013	.3108±.0008
	192	.1570±.0001	.2517±.0002	.1652±.0004	.2639±.0004	.2161±.0017	.3198±.0120
	336	.1739±.0004	.2683±.0004	.1818±.0005	.2796±.0006	.2313±.0009	.3316±.0008
	720	.2129±.0001	.3009±.0002	.2214±.0003	.3123±.0004	.2676±.0009	.3579±.0008
Top1 count		8	6	5	8	0	0
Top2 count		11	8	6	3	0	0

Table 18: 5% few-shot forecasting results for PatchTST [22] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 17.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4564±.0002	.4543±.0002	.5686±.0011	.5062±.0007	.4138±.0033	.4275±.0020
	192	.4851±.0011	.4704±.0007	.5866±.0015	.5181±.0006	.4480±.0033	.4475±.0027
	336	.4991±.0037	.4806±.0016	.5877±.0024	.5240±.0013	.4656±.0021	.4601±.0001
	720	.5038±.0033	.5011±.0016	.6054±.0005	.5503±.0003	.4653±.0026	.4791±.0014
ETTh2	96	.3331±.0047	.3822±.0021	.3580±.0014	.4015±.0017	.3024±.0065	.3614±.0043
	192	.3844±.0040	.4165±.0031	.3958±.0061	.4248±.0040	**.3560±.0011**	.3957±.0011
	336	.3970±.0030	.4308±.0021	.3967±.0049	.4303±.0019	**.3775±.0047**	**.4147±.0022**
	720	.4267±.0008	.4534±.0007	.4312±.0020	.4548±.0019	**.4052±.0002**	**.4390±.0006**
ETTm1	96	.3086±.0027	.3569±.0026	.3090±.0019	.3580±.0012	.2962±.0027	.3483±.0022
	192	.3463±.0014	.3790±.0009	.3435±.0044	.3764±.0025	.3396±.0035	.3772±.0022
	336	.3772±.0010	.3981±.0011	.3783±.0008	.3972±.0009	.3784±.0114	.3985±.0056
	720	.4261±.0017	.4294±.0013	.4365±.0056	.4294±.0027	.4235±.0190	.4274±.0008
ETTm2	96	.1696±.0012	.2609±.0024	.1746±.0007	.2659±.0016	.1706±.0010	.2622±.0007
	192	.2280±.0013	.3015±.0023	.2290±.0067	.3024±.0054	.2263±.0013	.3001±.0025
	336	.2852±.0022	.3401±.0021	.2820±.0011	.3358±.0009	.2833±.0028	.3400±.0023
	720	.3697±.0007	.3908±.0007	.3696±.0028	.3891±.0018	.3711±.0020	.3913±.0140
Weather	96	.1705±.0007	.2217±.0008	.1629±.0012	.2126±.0013	**.1597±.0013**	**.2091±.0016**
	192	.2151±.0016	.2581±.0015	.2072±.0004	.2509±.0004	**.2034±.0003**	**.2483±.0007**
	336	.2642±.0018	.2940±.0014	.2560±.0011	.2879±.0011	**.2522±.0011**	**.2851±.0012**
	720	.3353±.0014	.3439±.0013	.3277±.0011	.3382±.0012	**.3256±.0009**	**.3377±.0008**
ECL	96	.1555±.0004	.2580±.0008	.1441±.0001	.2420±.0001	**.1413±.0005**	**.2378±.0005**
	192	.1700±.0007	.2702±.0009	.1595±.0006	.2555±.0007	**.1561±.0002**	**.2509±.0002**
	336	.1871±.0007	.2863±.0006	.1761±.0002	.2719±.0002	**.1731±.0003**	**.2677±.0002**
	720	.2261±.0006	.3179±.0008	.2148±.0003	.3038±.0002	**.2121±.0002**	**.3005±.0002**
Top1 count		0	0	0	0	**11**	**10**
Top2 count		1	1	2	5	4	7

Table 19: 10% few-shot forecasting results for PatchTST [22] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 20.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4063±.0053	.4233±.0037	.4087±.0105	.4236±.0069	.5325±.0088	.4949±.0044
	192	.4328±.0037	.4380±.0030	.4320±.0072	.4362±.0053	.5579±.0048	.5083±.0025
	336	.4484±.0035	**.4484±.0019**	**.4477±.0029**	.4490±.0018	.5623±.0036	.5147±.0021
	720	**.4497±.0032**	**.4694±.0024**	.4505±.0046	.4701±.0032	.5729±.0024	.5371±.0140
ETTh2	96	.2856±.0031	.3484±.0021	**.2843±.0020**	**.3467±.0012**	.3643±.0018	.4051±.0012
	192	.3473±.0020	.3904±.0024	.3456±.0017	**.3857±.0021**	.4000±.0014	.4282±.0009
	336	.3695±.0002	.4100±.0014	**.3693±.0007**	**.4089±.0016**	.4031±.0008	.4352±.0005
	720	.4042±.0005	.4393±.0006	.4007±.0008	**.4365±.0005**	.4376±.0004	.4595±.0000
ETTm1	96	**.2922±.0015**	.3456±.0018	.2932±.0023	**.3445±.0024**	.3116±.0011	.3592±.0004
	192	**.3307±.0023**	.3726±.0017	.3326±.0013	.3702±.0013	.3468±.0023	.3815±.0016
	336	**.3627±.0007**	**.3907±.0008**	.3772±.0010	.3982±.0007	.3764±.0012	.3987±.0001
	720	**.4089±.0011**	**.4193±.0011**	.4152±.0035	.4216±.0034	.4300±.0013	.4288±.0010
ETTm2	96	**.1649±.0015**	**.2550±.0014**	.1685±.0012	.2578±.0021	.1736±.0016	.2651±.0013
	192	**.2203±.0003**	**.2947±.0006**	.2238±.0001	.2965±.0008	.2301±.0004	.3041±.0120
	336	**.2750±.0002**	**.3299±.0003**	.2781±.0002	.3334±.0003	.2845±.0020	.3399±.0012
	720	**.3650±.0010**	**.3826±.0009**	.3716±.0018	.3899±.0023	.3714±.0034	.3905±.0025
Weather	96	.1596±.0019	.2099±.0020	.1599±.0007	.2095±.0008	.1735±.0006	.2265±.0011
	192	.2024±.0007	.2477±.0008	.2035±.0008	.2492±.0003	.2161±.0120	.2609±.0012
	336	.2524±.0002	.2858±.0005	.2530±.0006	.2859±.0006	.2640±.0006	.2957±.0007
	720	.3229±.0006	.3353±.0009	.3246±.0018	.3359±.0013	.3326±.0004	.3430±.0004
ECL	96	.1390±.0001	.2344±.0001	.1422±.0003	.2398±.0005	.1830±.0006	.2909±.0005
	192	.1542±.0002	.2477±.0002	.1573±.0003	.2526±.0005	.1953±.0006	.3004±.0005
	336	.1710±.0002	.2646±.0001	.1742±.0001	.2696±.0001	.2101±.0004	.3130±.0003
	720	.2097±.0001	.2970±.0001	.2224±.0006	.3168±.0140	.2465±.0003	.3404±.0003
Top1 count		9	8	3	5	0	0
Top2 count		9	9	7	6	0	0

Table 20: 10% few-shot forecasting results for PatchTST [22] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 19.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4246±.0021	.4348±.0010	.4869±.0007	.4693±.0004	**.4039±.0158**	**.4202±.0109**
	192	.4513±.0009	.4497±.0008	.5070±.0009	.4810±.0003	**.4256±.0010**	**.4311±.0007**
	336	.4670±.0018	.4612±.0006	.5094±.0021	.4859±.0007	.4537±.0022	.4535±.0021
	720	.4688±.0024	.4819±.0014	.5172±.0009	.5084±.0009	.4541±.0064	.4717±.0040
ETTh2	96	.3026±.0023	.3607±.0140	.3283±.0007	.3794±.0007	.2851±.0021	.3476±.0013
	192	.3610±.0015	.3993±.0011	.3828±.0038	.4132±.0035	**.3450±.0010**	.3867±.0017
	336	.3799±.0006	.4188±.0005	.3928±.0028	.4255±.0190	.3706±.0022	.4103±.0005
	720	.4140±.0002	.4457±.0003	.4197±.0013	.4472±.0018	.4007±.0009	.4366±.0007
ETTm1	96	.2998±.0027	.3487±.0114	.3003±.0035	.3516±.0014	.2957±.0024	.3470±.0007
	192	.3400±.0017	.3756±.0010	.3366±.0019	.3735±.0012	.3319±.0013	**.3701±.0010**
	336	.3703±.0006	.3948±.0009	.3710±.0023	.3940±.0007	.3707±.0039	.3960±.0027
	720	.4198±.0016	.4236±.0005	.4238±.0140	.4250±.0005	.4183±.0240	.4246±.0014
ETTm2	96	.1674±.0015	.2584±.0019	.1708±.0016	.2602±.0015	.1674±.0014	.2572±.0014
	192	.2256±.0012	.2998±.0029	.2291±.0022	.3011±.0017	.2233±.0004	.2979±.0003
	336	.2820±.0280	.3373±.0031	.2806±.0025	.3330±.0018	.2799±.0008	.3354±.0009
	720	.3715±.0220	.3910±.0003	.3684±.0034	.3873±.0025	.3723±.0023	.3918±.0019
Weather	96	.1652±.0006	.2146±.0010	.1579±.0100	.2071±.0013	**.1556±.0008**	**.2052±.0010**
	192	.2105±.0130	.2534±.0013	.2012±.0002	.2456±.0003	**.1994±.0005**	**.2454±.0005**
	336	.2607±.0007	.2906±.0005	.2540±.0030	.2861±.0023	**.2487±.0006**	**.2831±.0009**
	720	.3308±.0008	.3395±.0007	.3251±.0024	.3360±.0160	**.3216±.0007**	**.3346±.0002**
ECL	96	.1455±.0002	.2441±.0003	.1393±.0001	.2355±.0002	**.1381±.0003**	**.2339±.0003**
	192	.1606±.0004	.2575±.0006	.1549±.0004	.2493±.0003	**.1535±.0002**	**.2472±.0003**
	336	.1777±.0002	.2743±.0003	.1718±.0002	.2664±.0003	**.1703±.0001**	**.2644±.0001**
	720	.2167±.0002	.3064±.0002	.2106±.0003	.2989±.0005	**.2088±.0001**	**.2967±.0000**
Top1 count		0	0	0	0	**11**	**11**
Top2 count		1	0	3	5	4	4

F.3 iTransformer few-shot results

Table 21: 1% few-shot forecasting results for iTransformer [15] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 22.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.6050±.0112	.5178±.0026	.6387±.0086	.5302±.0014	.8185±.0045	.6034±.0006
	192	.6829±.0123	.5564±.0047	.7039±.0110	.5629±.0039	.8576±.0042	.6233±.0019
	336	.7136±.0027	.5753±.0002	.7272±.0016	.5793±.0001	.8667±.0035	.6338±.0020
	720	.7371±.0196	.5995±.0079	.7356±.0118	.5978±.0042	.8759±.0025	.6503±.0013
ETTh2	96	.3539±.0082	.3878±.0052	.3526±.0167	.3898±.0098	.3918±.0006	.4114±.0004
	192	.4331±.0057	.4322±.0039	.4277±.0009	.4311±.0014	.4768±.0006	.4575±.0008
	336	.4642±.0085	.4591±.0072	**.4621±.0034**	.4596±.0015	.5007±.0018	.4801±.0016
	720	**.4681±.0060**	**.4706±.0038**	.4710±.0021	.4729±.0013	.4972±.0009	.4855±.0007
ETTm1	96	.4326±.0063	.4246±.0033	.5080±.0002	.4632±.0008	.7998±.0017	.5800±.0004
	192	.4747±.0050	.4451±.0030	.5526±.0052	.4845±.0027	.8141±.0107	.5895±.0029
	336	.5128±.0027	.4685±.0016	.5815±.0070	.5019±.0032	.8289±.0082	.5987±.0023
	720	.5718±.0046	.5001±.0032	.6197±.0039	.5236±.0020	.8588±.0058	.6162±.0018
ETTm2	96	.2026±.0030	.2853±.0037	.2135±.0001	.2941±.0011	.2493±.0003	.3221±.0003
	192	**.2668±.0014**	**.3228±.0011**	.2745±.0030	.3284±.0021	.3028±.0005	.3514±.0002
	336	.3278±.0007	.3617±.0012	.3336±.0011	.3657±.0013	.3586±.0004	.3836±.0003
	720	.4297±.0050	.4160±.0041	**.4295±.0015**	.4160±.0015	.4543±.0005	.4331±.0004
Weather	96	**.2033±.0035**	**.2455±.0030**	.2161±.0005	.2665±.0007	.2225±.0023	.2690±.0110
	192	**.2486±.0025**	**.2812±.0024**	.2592±.0002	.2970±.0017	.2683±.0004	.3023±.0100
	336	**.2994±.0011**	**.3147±.0002**	.3099±.0005	.3299±.0008	.3129±.0008	.3309±.0003
	720	**.3695±.0012**	**.3596±.0010**	.3797±.0007	.3730±.0007	.3812±.0007	.3724±.0007
ECL	96	.2859±.0042	.3669±.0035	.6869±.0091	.6538±.0072	.7011±.0027	.6706±.0016
	192	.2848±.0050	.3684±.0041	.6840±.0049	.6484±.0020	.7474±.0196	.6944±.0111
	336	.3002±.0050	.3815±.0040	.6947±.0026	.6544±.0009	.7545±.0181	.6958±.0100
	720	**.3458±.0030**	**.4126±.0024**	.7337±.0053	.6724±.0027	.8048±.0074	.7173±.0032
Top1 count		7	7	2	0	0	0
Top2 count		14	15	4	3	0	0

Table 22: 1% few-shot forecasting results for iTransformer [15] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 21.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.7401±.0010	.5718±.0022	.7335±.0084	.5706±.0038	**.5008±.0050**	**.4761±.0014**
	192	.7895±.0059	.5962±.0026	.7744±.0064	.5912±.0015	**.5601±.0062**	**.5068±.0019**
	336	.8151±.0005	.6137±.0010	.7879±.0048	.6045±.0024	**.5975±.0056**	**.5248±.0032**
	720	.8253±.0026	.6322±.0014	.7917±.0044	.6207±.0018	**.6108±.0017**	**.5466±.0009**
ETTh2	96	.3634±.0045	.3956±.0028	.3755±.0031	.4015±.0027	**.3377±.0065**	**.3763±.0051**
	192	.4483±.0025	.4418±.0018	.4578±.0052	.4465±.0043	**.4251±.0048**	**.4264±.0039**
	336	.4803±.0024	.4690±.0018	.4835±.0036	.4702±.0012	.4671±.0130	**.4581±.0062**
	720	.4847±.0017	.4793±.0007	.4811±.0067	.4767±.0027	.4731±.0068	.4713±.0021
ETTm1	96	.6927±.0025	.5410±.0006	.6807±.0065	.5365±.0034	**.4310±.0035**	**.4235±.0018**
	192	.7453±.0196	.5627±.0079	.7284±.0028	.5577±.0010	**.4681±.0067**	**.4412±.0041**
	336	.7799±.0066	.5812±.0033	.7713±.0073	.5784±.0025	**.5078±.0051**	**.4656±.0034**
	720	.8229±.0080	.6046±.0045	.8150±.0030	.6014±.0010	**.5658±.0038**	**.4969±.0026**
ETTm2	96	.2342±.0001	.3108±.0002	.2369±.0017	.3130±.0013	**.2023±.0029**	.2853±.0040
	192	.2912±.0006	.3424±.0003	.2993±.0008	.3492±.0005	.2670±.0015	.3231±.0011
	336	.3459±.0003	.3745±.0004	.3589±.0004	.3836±.0003	**.3272±.0019**	**.3616±.0024**
	720	.4417±.0160	.4245±.0013	.4562±.0010	.4343±.0007	.4321±.0089	.4179±.0063
Weather	96	.2148±.0007	.2677±.0140	.2063±.0006	.2487±.0012	.2111±.0019	.2490±.0024
	192	.2638±.0020	.3033±.0028	.2505±.0014	.2833±.0010	.2588±.0012	.2865±.0013
	336	.3115±.0005	.3328±.0005	.3054±.0016	.3223±.0150	.3067±.0026	.3180±.0020
	720	.3803±.0004	.3743±.0010	.3756±.0019	.3666±.0260	.3752±.0013	.3623±.0008
ECL	96	.5081±.0174	.5367±.0129	.6243±.0080	.6229±.0051	**.2691±.0005**	**.3532±.0009**
	192	.6123±.0220	.6076±.0131	.6844±.0136	.6584±.0077	**.2774±.0021**	**.3626±.0016**
	336	.6563±.0065	.6311±.0066	.7374±.0169	.6868±.0113	**.2995±.0010**	**.3814±.0011**
	720	.7102±.0062	.6573±.0023	.7994±.0189	.7159±.0139	.3482±.0026	.4152±.0023
Top1 count		0	0	0	0	**15**	**15**
Top2 count		0	0	3	2	3	6

Table 23: 5% few-shot forecasting results for iTransformer [15] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 24.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4375±.0008	.4371±.0008	.4876±.0014	.4671±.0008	.6023±.0027	.5206±.0016
	192	.5082±.0034	.4778±.0022	.5481±.0018	.4991±.0003	.6923±.0014	.5611±.0006
	336	.5559±.0015	.5014±.0007	.5878±.0033	.5192±.0024	.7224±.0032	.5796±.0007
	720	.5948±.0107	.5394±.0063	.6097±.0068	.5469±.0038	.7577±.0025	.6082±.0010
ETTh2	96	.3206±.0059	.3649±.0054	.3304±.0070	.3758±.0058	.3609±.0031	.3917±.0020
	192	.4016±.0031	.4120±.0013	.4149±.0022	.4232±.0022	.4505±.0022	.4423±.0015
	336	.4437±.0023	.4444±.0026	.4493±.0021	.4518±.0018	.4801±.0014	.4683±.0018
	720	.4504±.0058	.4594±.0037	.4537±.0037	.4627±.0020	.4805±.0015	.4764±.0010
ETTm1	96	.3645±.0037	**.3858±.0022**	.4578±.0016	.4394±.0003	.5590±.0020	.4920±.0016
	192	**.4019±.0013**	**.4024±.0008**	.4906±.0010	.4550±.0010	.6004±.0058	.5109±.0030
	336	**.4345±.0015**	**.4231±.0001**	.5198±.0017	.4730±.0008	.6362±.0100	.5292±.0045
	720	**.5008±.0027**	**.4574±.0004**	.5708±.0017	.5000±.0011	.6843±.0069	.5536±.0030
ETTm2	96	.1913±.0008	.2769±.0012	.1975±.0009	.2825±.0010	.2261±.0006	.3035±.0004
	192	.2555±.0017	.3158±.0024	.2630±.0035	.3227±.0036	.2843±.0004	.3376±.0100
	336	.3203±.0017	.3573±.0012	.3269±.0011	.3615±.0004	.3418±.0005	.3716±.0004
	720	**.4170±.0022**	**.4084±.0020**	.4215±.0021	.4113±.0019	.4392±.0006	.4231±.0004
Weather	96	**.1931±.0012**	**.2339±.0006**	.2154±.0003	.2675±.0019	.2143±.0040	.2623±.0053
	192	**.2391±.0013**	**.2709±.0009**	.2602±.0018	.2986±.0035	.2632±.0006	.2995±.0009
	336	**.2916±.0007**	**.3084±.0008**	.3085±.0007	.3292±.0008	.3111±.0001	.3296±.0006
	720	**.3649±.0013**	**.3556±.0007**	.3775±.0020	.3715±.0024	.3790±.0003	.3710±.0007
ECL	96	**.2159±.0028**	**.3037±.0030**	.5842±.0666	.5829±.0446	.4146±.0013	.4756±.0014
	192	**.2296±.0017**	**.3192±.0019**	.6637±.0171	.6400±.0106	.5404±.0924	.5613±.0599
	336	**.2563±.0017**	**.3429±.0015**	.6716±.0176	.6428±.0090	.4714±.0229	.5187±.0173
	720	**.3060±.0012**	**.3788±.0008**	.7077±.0128	.6606±.0061	.5061±.0332	.5383±.0238
Top1 count		**12**	**13**	0	0	0	0
Top2 count		12	11	0	0	0	0

Table 24: 5% few-shot forecasting results for iTransformer [15] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 23.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.5757±.0060	.5036±.0047	.5393±.0060	.4894±.0033	**.4161±.0034**	**.4238±.0024**
	192	.6608±.0097	.5456±.0034	.6235±.0048	.5303±.0012	**.4751±.0020**	**.4581±.0014**
	336	.6902±.0052	.5686±.0037	.6576±.0026	.5515±.0018	**.5182±.0027**	**.4803±.0016**
	720	.7183±.0058	.5986±.0042	.6784±.0039	.5779±.0021	**.5554±.0039**	**.5193±.0020**
ETTh2	96	.3266±.0058	.3706±.0048	.3381±.0055	.3773±.0047	**.3134±.0051**	**.3588±.0039**
	192	.4094±.0023	.4187±.0023	.4261±.0035	.4272±.0028	**.3905±.0022**	**.4051±.0016**
	336	.4453±.0037	.4487±.0029	.4614±.0041	.4577±.0005	**.4362±.0062**	**.4393±.0041**
	720	.4549±.0046	.4631±.0027	.4617±.0026	.4660±.0036	**.4439±.0124**	**.4540±.0067**
ETTm1	96	.4618±.0012	.4405±.0008	.4303±.0007	.4240±.0008	**.3625±.0025**	.3860±.0012
	192	.4954±.0004	.4574±.0002	.4867±.0120	.4531±.0007	.4057±.0008	.4071±.0003
	336	.5240±.0010	.4755±.0004	.5279±.0031	.4764±.0012	.4377±.0018	.4275±.0012
	720	.5764±.0028	.5036±.0020	.5976±.0027	.5128±.0010	.5034±.0016	.4631±.0011
ETTm2	96	.2104±.0003	.2932±.0003	.2009±.0008	.2837±.0006	**.1893±.0006**	**.2752±.0008**
	192	.2734±.0012	.3305±.0014	.2665±.0005	.3235±.0007	**.2550±.0140**	**.3153±.0019**
	336	.3331±.0003	.3657±.0002	.3295±.0007	.3621±.0003	**.3194±.0024**	**.3563±.0021**
	720	.4316±.0012	.4180±.0010	.4281±.0013	.4153±.0008	.4174±.0018	.4089±.0011
Weather	96	.2145±.0004	.2683±.0004	.1957±.0140	.2365±.0110	.2016±.0022	.2392±.0015
	192	.2619±.0005	.3018±.0013	.2425±.0005	.2737±.0005	.2455±.0021	.2752±.0020
	336	.3099±.0008	.3319±.0014	.2949±.0008	.3111±.0008	.2964±.0012	.3117±.0005
	720	.3785±.0002	.3734±.0003	.3667±.0007	.3573±.0009	.3701±.0013	.3589±.0012
ECL	96	.3698±.0029	.4273±.0027	.3977±.0019	.4593±.0018	.2261±.0015	.3142±.0015
	192	.3964±.0051	.4551±.0044	.4111±.0026	.4713±.0190	.2393±.0020	.3283±.0017
	336	.4259±.0115	.4776±.0080	.4277±.0008	.4838±.0010	.2621±.0019	.3487±.0017
	720	.4925±.0309	.5200±.0233	.4811±.0048	.5182±.0037	.3084±.0010	.3818±.0010
Top1 count		0	0	0	0	**12**	11
Top2 count		0	0	4	4	8	9

Table 25: 10% few-shot forecasting results for iTransformer [15] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 26.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4119±.0018	.4220±.0008	.4389±.0028	.4404±.0020	.4907±.0014	.4694±.0009
	192	.4647±.0019	.4517±.0012	.5096±.0044	.4807±.0018	.5669±.0036	.5082±.0007
	336	.5113±.0013	.4757±.0006	.5544±.0035	.5028±.0023	.6115±.0009	.5313±.0009
	720	.5475±.0023	.5129±.0014	.5768±.0024	.5305±.0019	.6499±.0041	.5660±.0024
ETTh2	96	.3080±.0022	.3557±.0013	.3198±.0098	.3698±.0072	.3378±.0018	.3763±.0011
	192	**.3892±.0019**	.4055±.0010	.4102±.0023	.4204±.0029	.4324±.0032	.4310±.0024
	336	**.4296±.0049**	.4374±.0036	.4478±.0023	.4514±.0008	.4627±.0011	.4577±.0012
	720	**.4363±.0028**	**.4515±.0008**	.4580±.0021	.4662±.0006	.4660±.0034	.4686±.0019
ETTm1	96	.3501±.0009	.3791±.0005	.3863±.0015	.4016±.0006	.4446±.0004	.4378±.0004
	192	**.3869±.0018**	**.3949±.0014**	.4428±.0050	.4303±.0030	.4764±.0008	.4517±.0007
	336	**.4234±.0010**	**.4186±.0002**	.4784±.0063	.4523±.0032	.5116±.0051	.4722±.0032
	720	**.4887±.0004**	**.4550±.0003**	.5357±.0035	.4829±.0020	.5675±.0040	.5009±.0022
ETTm2	96	.1869±.0014	.2730±.0020	.1934±.0006	.2787±.0013	.2080±.0003	.2894±.0003
	192	.2514±.0012	.3118±.0013	.2576±.0012	.3175±.0019	.2700±.0012	.3266±.0011
	336	**.3137±.0010**	**.3515±.0010**	.3234±.0013	.3582±.0013	.3285±.0006	.3620±.0003
	720	.4153±.0009	.4079±.0012	.4205±.0002	.4106±.0008	.4287±.0007	.4164±.0004
Weather	96	**.1858±.0003**	**.2265±.0004**	.2125±.0010	.2666±.0008	.2086±.0006	.2529±.0160
	192	**.2345±.0007**	**.2670±.0007**	.2599±.0013	.2982±.0009	.2610±.0039	.2966±.0046
	336	**.2875±.0006**	**.3041±.0005**	.3083±.0017	.3287±.0032	.3106±.0005	.3292±.0010
	720	**.3625±.0011**	**.3529±.0015**	.3775±.0007	.3709±.0020	.3790±.0001	.3708±.0004
ECL	96	**.1939±.0009**	**.2799±.0009**	.4395±.0725	.4724±.0502	.3904±.0023	.4500±.0019
	192	**.2029±.0007**	**.2903±.0008**	.4856±.0448	.5137±.0320	.4195±.0110	.4738±.0112
	336	**.2220±.0001**	**.3097±.0001**	.5576±.0734	.5593±.0527	.4080±.0023	.4663±.0022
	720	**.2694±.0008**	**.3472±.0009**	.5912±.0165	.5857±.0096	.4437±.0112	.4888±.0012
Top1 count		**15**	**13**	0	0	0	0
Top2 count		9	11	0	0	0	0

Table 26: 10% few-shot forecasting results for iTransformer [15] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 25.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4820±.0011	.4640±.0007	.4607±.0022	.4512±.0010	**.4058±.0015**	**.4169±.0005**
	192	.5458±.0117	.4960±.0046	.5321±.0014	.4893±.0007	**.4555±.0022**	**.4457±.0013**
	336	.5867±.0091	.5176±.0048	.5808±.0011	.5153±.0010	**.4993±.0032**	**.4678±.0018**
	720	.6299±.0101	.5571±.0051	.6090±.0027	.5472±.0017	**.5287±.0024**	**.5033±.0018**
ETTh2	96	.3092±.0018	.3588±.0008	.3186±.0015	.3629±.0016	**.3046±.0020**	**.3533±.0017**
	192	.3973±.0030	.4111±.0034	.4103±.0020	.4171±.0025	.3896±.0033	**.4053±.0017**
	336	.4332±.0033	.4416±.0019	.4493±.0097	.4497±.0061	.4331±.0016	**.4372±.0012**
	720	.4497±.0009	.4614±.0006	.4507±.0029	.4599±.0019	.4402±.0047	.4526±.0020
ETTm1	96	.4176±.0011	.4176±.0008	.3781±.0012	.3941±.0005	**.3481±.0023**	**.3784±.0019**
	192	.4648±.0041	.4414±.0025	.4304±.0004	.4213±.0003	.3888±.0037	.3973±.0024
	336	.5037±.0021	.4657±.0011	.4704±.0018	.4453±.0010	.4274±.0008	.4225±.0007
	720	.5623±.0055	.4968±.0032	.5402±.0012	.4838±.0008	.4925±.0015	.4581±.0009
ETTm2	96	.1951±.0010	.2811±.0014	.1900±.0004	.2739±.0003	**.1847±.0011**	**.2698±.0017**
	192	.2595±.0011	.3199±.0013	.2549±.0002	.3140±.0009	.2514±.0010	.3118±.0013
	336	.3217±.0004	.3585±.0009	.3174±.0007	.3531±.0004	.3146±.0017	.3519±.0014
	720	.4235±.0011	.4128±.0006	**.4141±.0015**	**.4064±.0010**	.4176±.0027	.4087±.0008
Weather	96	.2157±.0011	.2673±.0011	.1895±.0006	.2302±.0008	.1936±.0014	.2326±.0012
	192	.2620±.0009	.2999±.0004	.2378±.0012	.2695±.0006	.2401±.0010	.2707±.0014
	336	.3103±.0010	.3305±.0027	.2911±.0005	.3087±.0006	.2921±.0005	.3079±.0004
	720	.3784±.0007	.3726±.0018	.3642±.0003	.3560±.0006	.3684±.0005	.3578±.0004
ECL	96	.3419±.0002	.4052±.0004	.3621±.0045	.4281±.0034	.1959±.0004	.2864±.0004
	192	.3560±.0071	.4185±.0060	.3714±.0025	.4381±.0021	.2082±.0012	.2991±.0011
	336	.3759±.0031	.4359±.0028	.3835±.0028	.4474±.0015	.2296±.0005	.3191±.0004
	720	.4129±.0077	.4592±.0062	.4310±.0034	.4795±.0024	.2776±.0018	.3567±.0017
Top1 count		0	0	1	1	7	9
Top2 count		0	0	4	3	12	11

F.4 TimeVLM few-shot results

Table 27: 1% few-shot forecasting results for TimeVLM [33] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 28.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**.6168±.0133**	**.5361±.0072**	.6407±.0239	.5452±.0107	.6561±.0172	.5599±.0088
	192	**.6362±.0132**	**.5501±.0072**	.6533±.0079	.5542±.0059	.6821±.0118	.5702±.0080
	336	**.6101±.0128**	**.5442±.0098**	.6314±.0085	.5613±.0051	.6482±.0102	.5744±.0065
	720	**.6476±.0060**	**.5796±.0026**	.6671±.0075	.5886±.0039	.7278±.0080	.6241±.0051
ETTh2	96	**.3746±.0068**	**.4174±.0049**	.3753±.0087	.4182±.0059	.4117±.0117	.4416±.0074
	192	.3879±.0027	.4260±.0019	**.3876±.0029**	**.4256±.0021**	.4536±.0157	.4720±.0115
	336	.3952±.0023	.4316±.0019	**.3948±.0024**	**.4315±.0016**	.4701±.0149	.4903±.0109
	720	**.4335±.0044**	**.4607±.0042**	.4336±.0045	.4608±.0043	.5124±.0146	.5104±.0104
ETTm1	96	.3685±.0037	**.4015±.0030**	**.3672±.0028**	.4019±.0021	.4311±.0098	.4364±.0064
	192	.3916±.0039	.4143±.0023	**.3908±.0036**	**.4141±.0022**	.4614±.0117	.4541±.0074
	336	.4118±.0029	.4235±.0025	**.4110±.0037**	**.4234±.0022**	.4841±.0109	.4720±.0080
	720	.4530±.0019	.4430±.0017	**.4518±.0032**	**.4422±.0020**	.5254±.0128	.5041±.0086
ETTm2	96	.2042±.0034	**.2879±.0027**	**.2025±.0032**	.2885±.0031	.2384±.0079	.3064±.0049
	192	.2505±.0034	.3187±.0027	**.2488±.0024**	**.3176±.0013**	.2941±.0102	.3452±.0064
	336	.2953±.0021	.3422±.0006	**.2951±.0023**	**.3420±.0008**	.3364±.0098	.3776±.0080
	720	.3803±.0022	.3969±.0023	**.3795±.0037**	**.3962±.0035**	.4114±.0141	.4234±.0102
Weather	96	.1949±.0038	.2539±.0037	.1916±.0036	.2506±.0038	.2444±.0022	.2973±.0016
	192	.2334±.0013	.2828±.0013	.2331±.0017	.2825±.0013	.2751±.0014	.3174±.0013
	336	.2715±.0008	.3083±.0007	.2714±.0011	.3083±.0009	.3040±.0003	.3343±.0002
	720	**.3357±.0012**	**.3536±.0012**	.3364±.0017	.3542±.0015	.3549±.0003	.3692±.0001
ECL	96	**.2443±.0058**	**.3417±.0060**	.2458±.0043	.3436±.0045	.2941±.0128	.3841±.0096
	192	**.2487±.0044**	**.3444±.0041**	.2513±.0032	.3475±.0033	.3141±.0114	.3952±.0102
	336	**.2606±.0006**	**.3536±.0007**	.2644±.0004	.3577±.0006	.3286±.0102	.4114±.0006
	720	.2880±.0019	.3711±.0019	.2906±.0014	.3736±.0014	.3564±.0114	.4286±.0098
Top1 count		**10**	**12**	**10**	8	0	0
Top2 count		10	9	5	10	0	0

Table 28: 1% few-shot forecasting results for TimeVLM [33] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 27.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.6278±.0114	.5524±.0097	.6714±.0141	.5924±.0072	.6714±.0141	.5924±.0072
	192	.6511±.0076	.5540±.0037	.6923±.0114	.5858±.0061	.6498±.0096	.5533±.0055
	336	.6199±.0068	.5522±.0038	.6703±.0102	.5952±.0086	.6270±.0085	.5573±.0066
	720	.6584±.0047	.5895±.0017	.7321±.0080	.6244±.0055	.6633±.0093	.5911±.0044
ETTh2	96	.3776±.0058	.4196±.0034	.4201±.0069	.4498±.0049	.3773±.0054	.4194±.0036
	192	.3889±.0037	.4278±.0027	.4498±.0086	.4758±.0063	.3880±.0031	.4268±.0023
	336	.3969±.0023	.4344±.0014	.4703±.0064	.4923±.0051	.3966±.0028	.4339±.0023
	720	.4358±.0037	.4644±.0034	.5124±.0066	.5108±.0049	.4366±.0034	.4634±.0029
ETTm1	96	.3699±.0031	.4034±.0017	.4301±.0064	.4423±.0051	.3694±.0034	.4034±.0019
	192	.3918±.0044	.4152±.0034	.4601±.0102	.4612±.0064	.3914±.0041	.4152±.0029
	336	.4139±.0034	.4256±.0027	.4823±.0086	.4823±.0061	.4134±.0037	.4254±.0031
	720	.4539±.0031	.4442±.0031	.5234±.0080	.5104±.0053	.4532±.0029	.4440±.0033
ETTm2	96	.2052±.0039	.2892±.0027	.2401±.0066	.3102±.0051	.2048±.0031	.2890±.0021
	192	.2518±.0037	.3203±.0023	.2944±.0102	.3498±.0061	.2514±.0034	.3201±.0025
	336	.2966±.0031	.3434±.0017	.3376±.0086	.3812±.0066	.2961±.0033	.3431±.0027
	720	.3811±.0033	.3986±.0041	.4124±.0080	.4256±.0086	.3811±.0034	.3985±.0043
Weather	96	.1918±.0034	.2508±.0035	.2453±.0004	.2977±.0005	**.1915±.0036**	.2506±.0038
	192	.2329±.0018	.2823±.0013	.2730±.0029	.3162±.0029	**.2327±.0014**	**.2822±.0010**
	336	.2714±.0010	.3083±.0009	.3050±.0009	.3351±.0007	**.2712±.0011**	**.3080±.0009**
	720	.3360±.0016	.3539±.0014	.3535±.0007	.3670±.0007	.3359±.0016	.3538±.0015
ECL	96	.2466±.0047	.3451±.0041	.2586±.0080	.3588±.0053	.2501±.0044	.3488±.0054
	192	.2513±.0047	.3488±.0041	.2686±.0086	.3698±.0066	.2536±.0034	.3498±.0043
	336	.2664±.0051	.3596±.0049	.2812±.0064	.3814±.0066	.2688±.0034	.3611±.0043
	720	.2914±.0053	.3752±.0049	**.2640±.0015**	**.3529±.0010**	.2966±.0037	.3796±.0055
Top1 count		0	0	1	1	3	2
Top2 count		4	2	0	0	3	3

Table 29: 5% few-shot forecasting results for TimeVLM [33] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 30.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**.4274±.0069**	**.4399±.0035**	.4343±.0076	.4444±.0041	.6452±.0063	.5503±.0028
	192	**.4433±.0041**	**.4511±.0025**	.4539±.0040	.4576±.0024	.6613±.0057	.5649±.0039
	336	**.4517±.0051**	**.4602±.0038**	.4611±.0049	.4653±.0035	.6546±.0068	.5667±.0063
	720	**.4869±.0051**	**.4939±.0033**	.4976±.0049	.4989±.0032	.6860±.0022	.6009±.0019
ETTh2	96	.3161±.0019	**.3685±.0005**	.3165±.0012	.3690±.0002	.3653±.0024	.4136±.0023
	192	.3626±.0030	.3998±.0005	**.3623±.0030**	.3987±.0006	.3829±.0014	.4244±.0015
	336	.3780±.0028	.4170±.0022	.3780±.0028	.4170±.0022	.3875±.0004	.4302±.0006
	720	**.4083±.0130**	**.4421±.0027**	.4084±.0016	.4422±.0029	.4315±.0027	.4611±.0024
ETTm1	96	.3170±.0044	.3609±.0027	.3156±.0044	.3600±.0027	.3491±.0017	.3890±.0013
	192	.3479±.0034	**.3780±.0031**	.3475±.0034	.3783±.0031	.3747±.0023	.4031±.0014
	336	.3767±.0010	.3998±.0016	.3747±.0009	.3989±.0016	.4005±.0020	.4158±.0010
	720	.4239±.0023	.4242±.0015	.4230±.0023	.4237±.0015	.4454±.0011	.4385±.0007
ETTm2	96	**.1703±.0008**	**.2609±.0007**	.1707±.0008	.2611±.0007	.1921±.0050	.2797±.0040
	192	**.2251±.0027**	**.2983±.0023**	.2252±.0027	.2993±.0023	.2402±.0017	.3105±.0012
	336	.2798±.0021	.3329±.0015	.2790±.0020	.3322±.0015	.2894±.0009	.3407±.0007
	720	.3682±.0007	.3681±.0006	.3681±.0007	.3681±.0006	.3785±.0026	.3962±.0021
Weather	96	**.1665±.0011**	**.2195±.0005**	.1678±.0010	.2210±.0004	.1837±.0009	.2427±.0009
	192	**.2096±.0024**	**.2566±.0018**	.2106±.0026	.2579±.0020	.2268±.0016	.2768±.0012
	336	**.2566±.0015**	**.2917±.0013**	.2567±.0008	.2921±.0009	.2683±.0014	.3054±.0013
	720	.3241±.0010	.3405±.0009	**.3239±.0012**	.3403±.0011	.3327±.0010	.3504±.0009
ECL	96	.1606±.0003	.2650±.0003	.1605±.0005	.2646±.0005	.2121±.0020	.3131±.0016
	192	.1753±.0001	.2767±.0002	.1745±.0002	.2761±.0001	.2200±.0013	.3195±.0009
	336	.1905±.0008	.2913±.0010	.1897±.0007	.2909±.0009	.2325±.0013	.3299±.0010
	720	.2275±.0003	.3201±.0003	.2265±.0002	.3187±.0002	.2670±.0005	.3547±.0004
Top1 count		**10**	**12**	2	0	0	0
Top2 count		1	0	9	8	0	0

Table 30: 5% few-shot forecasting results for TimeVLM [33] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 29.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.4357±.0069	.4453±.0035	.6357±.0014	.5453±.0010	.4319±.0055	.4435±.0032
	192	.4541±.0041	.4576±.0025	.6581±.0054	.5618±.0035	.4524±.0036	.4566±.0022
	336	.4614±.0051	.4656±.0038	.6566±.0019	.5692±.0020	.4604±.0042	.4649±.0032
	720	.4984±.0051	.4993±.0033	.6821±.0046	.5996±.0026	.4978±.0045	.4990±.0030
ETTh2	96	**.3160±.0019**	.3686±.0005	.3623±.0021	.4112±.0130	.3165±.0018	.3689±.0006
	192	.3626±.0030	.3987±.0005	.3845±.0009	.4264±.0130	.3625±.0027	**.3986±.0004**
	336	.3781±.0028	.4171±.0022	.3904±.0011	.4334±.0130	**.3779±.0028**	.4170±.0022
	720	.4086±.0130	.4422±.0027	.4346±.0024	.4643±.0023	.4095±.0025	.4433±.0037
ETTm1	96	.3156±.0044	.3600±.0027	.3499±.0029	.3895±.0024	.3156±.0044	.3600±.0027
	192	.3475±.0034	.3783±.0031	.3743±.0020	.4028±.0014	.3475±.0034	.3783±.0031
	336	.3747±.0009	.3989±.0016	.4005±.0020	.4160±.0010	.3765±.0029	.4006±.0032
	720	.4230±.0023	.4237±.0015	.4454±.0011	.4385±.0004	.4230±.0023	.4237±.0015
ETTm2	96	.1707±.0008	.2611±.0007	.1917±.0023	.2788±.0019	.1707±.0008	.2611±.0007
	192	.2252±.0027	.2993±.0023	.2425±.0026	.3127±.0021	.2253±.0024	.2986±.0020
	336	.2813±.0006	.3352±.0100	.2893±.0013	.3408±.0011	.2790±.0020	.3322±.0015
	720	.3681±.0007	.3681±.0006	.3780±.0016	.3964±.0017	.3681±.0007	.3681±.0006
Weather	96	.1678±.0011	.2210±.0005	.1863±.0011	.2454±.0012	.1677±.0011	.2209±.0005
	192	.2105±.0024	.2578±.0018	.2257±.0013	.2757±.0011	.2104±.0026	.2577±.0020
	336	.2570±.0014	.2931±.0013	.2693±.0015	.3063±.0013	.2569±.0015	.2931±.0013
	720	.3242±.0010	.3405±.0009	.3329±.0007	.3503±.0004	.3240±.0011	.3403±.0010
ECL	96	.1606±.0003	.2646±.0003	**.1583±.0002**	**.2623±.0002**	.1954±.0037	.3010±.0025
	192	.1747±.0001	.2763±.0002	**.1714±.0002**	**.2732±.0004**	.2073±.0010	.3103±.0011
	336	.1901±.0008	.2912±.0010	**.1864±.0002**	**.2872±.0002**	.2209±.0031	.3216±.0021
	720	.2268±.0003	.3191±.0003	**.2184±.0004**	**.3111±.0008**	.2548±.0021	.3470±.0016
Top1 count		1	0	4	4	1	1
Top2 count		1	2	0	0	8	8

Table 31: 10% few-shot forecasting results for TimeVLM [33] backbone (part 1). TS2Vec [32] is denoted as TS2V, and SimMTM [5] as MTM. Best results are in **bold** and the second best are underlined. (SCLT, Siam, and URL results are in Table 32.)

Dataset		**OliO (Ours)**		TS2V [32]		MTM [5]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3912±.0023	**.4169±.0014**	**.3901±.0031**	.4176±.0027	.5677±.0081	.5124±.0039
	192	.4160±.0023	.4343±.0020	.4157±.0023	.4338±.0014	.5819±.0053	.5244±.0037
	336	.4296±.0029	.4449±.0023	.4289±.0026	.4444±.0019	.5699±.0042	.5262±.0034
	720	.4795±.0051	.4939±.0033	.4802±.0091	.4896±.0060	.5939±.0027	.5557±.0014
ETTh2	96	**.2877±.0010**	**.3484±.0015**	.2880±.0002	.3487±.0013	.3507±.0014	.4007±.0013
	192	.3454±.0008	.3848±.0006	.3455±.0010	.3848±.0005	.3770±.0004	.4181±.0006
	336	**.3649±.0020**	.4060±.0023	.3650±.0022	.4060±.0020	.3825±.0007	.4261±.0008
	720	**.4009±.0022**	.4382±.0025	.4010±.0031	.4382±.0031	.4210±.0016	.4536±.0015
ETTm1	96	.3011±.0016	.3520±.0014	.3008±.0018	.3527±.0017	.3248±.0012	.3686±.0011
	192	.3350±.0018	.3738±.0014	.3350±.0018	.3732±.0012	.3541±.0014	.3851±.0010
	336	.3641±.0025	.3920±.0018	.3630±.0022	.3910±.0016	.3823±.0001	.4006±.0004
	720	**.4132±.0005**	**.4209±.0011**	.4139±.0006	.4219±.0140	.4321±.0015	.4313±.0011
ETTm2	96	**.1687±.0006**	**.2584±.0006**	.1689±.0006	.2585±.0005	.1761±.0011	.2660±.0006
	192	.2235±.0004	.2966±.0008	.2221±.0001	.2950±.0007	.2280±.0009	.3002±.0010
	336	**.2753±.0013**	**.3298±.0005**	.2758±.0008	.3300±.0004	.2826±.0029	.3353±.0019
	720	.3666±.0015	.3874±.0013	.3663±.0009	.3872±.0010	.3703±.0011	.3902±.0016
Weather	96	**.1527±.0009**	**.2029±.0006**	.1531±.0010	.2037±.0008	.1734±.0002	.2294±.0004
	192	**.1967±.0004**	**.2439±.0002**	.1973±.0004	.2443±.0001	.2179±.0009	.2659±.0007
	336	.2453±.0005	**.2819±.0004**	.2453±.0006	.2823±.0006	.2626±.0009	.2984±.0008
	720	.3161±.0006	.3332±.0004	.3164±.0006	.3337±.0006	.3277±.0008	.3445±.0008
ECL	96	.1428±.0005	.2420±.0004	.1426±.0006	.2420±.0005	.1732±.0014	.2781±.0014
	192	.1572±.0005	.2540±.0004	**.1567±.0002**	.2540±.0004	.1844±.0005	.2872±.0004
	336	.1732±.0001	.2700±.0004	**.1723±.0001**	.2700±.0004	.1985±.0004	.2999±.0004
	720	.2134±.0005	.3036±.0003	**.2120±.0004**	**.3035±.0004**	.2370±.0004	.3293±.0004
Top1 count		**8**	**8**	4	1	0	0
Top2 count		2	5	4	4	0	0

Table 32: 10% few-shot forecasting results for TimeVLM [33] backbone (part 2). SoftCLT [12] is denoted as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (OliO, TS2V, and MTM results are in Table 31.)

Dataset		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3910±.0028	.4183±.0025	.5611±.0033	.5108±.0021	.3902±.0028	.4177±.0025
	192	.4153±.0024	.4335±.0016	.5787±.0075	.5218±.0043	**.4151±.0022**	**.4334±.0014**
	336	.4289±.0025	.4443±.0018	.5746±.0030	.5296±.0022	**.4286±.0025**	**.4441±.0018**
	720	**.4780±.0076**	**.4882±.0049**	.5856±.0057	.5509±.0034	.4787±.0090	.4888±.0061
ETTh2	96	.2880±.0002	.3487±.0012	.3475±.0018	.3986±.0014	.2885±.0005	.3491±.0015
	192	.3454±.0003	.3848±.0004	.3766±.0004	.4175±.0007	.3457±.0012	.3849±.0005
	336	.3650±.0022	.4060±.0020	.3833±.0002	.4265±.0004	.3654±.0030	.4065±.0026
	720	.4011±.0031	.4383±.0033	.4229±.0028	.4555±.0031	.4014±.0031	.4388±.0032
ETTm1	96	.3008±.0018	.3527±.0017	.3254±.0009	.3697±.0005	.3039±.0013	**.3511±.0009**
	192	.3350±.0018	.3732±.0012	.3532±.0013	.3856±.0010	.3350±.0018	.3732±.0012
	336	.3630±.0022	.3910±.0016	.3819±.0006	.4029±.0006	**.3629±.0022**	**.3908±.0018**
	720	.4139±.0006	.4219±.0014	.4311±.0006	.4284±.0005	.4139±.0006	.4219±.0014
ETTm2	96	.1689±.0006	.2585±.0005	.1766±.0012	.2665±.0017	.1689±.0006	.2585±.0005
	192	.2221±.0001	.2950±.0007	.2283±.0008	.3014±.0008	.2221±.0001	.2950±.0007
	336	.2769±.0020	.3308±.0012	.2809±.0031	.3350±.0033	.2758±.0008	.3300±.0004
	720	.3663±.0009	.3872±.0010	.3725±.0043	.3922±.0035	.3663±.0009	.3872±.0010
Weather	96	.1534±.0010	.2039±.0008	.1749±.0007	.2312±.0006	.1532±.0008	.2037±.0006
	192	.1973±.0003	.2443±.0002	.2187±.0007	.2666±.0006	.1972±.0003	.2442±.0002
	336	.2453±.0006	.2821±.0006	.2629±.0010	.2987±.0009	.2458±.0005	.2823±.0007
	720	**.3159±.0005**	.3332±.0005	.3281±.0004	.3449±.0003	.3168±.0007	.3340±.0005
ECL	96	.1426±.0006	.2420±.0006	.1447±.0002	.2451±.0002	.1668±.0016	.2736±.0015
	192	.1568±.0002	.2542±.0005	.1585±.0002	.2566±.0002	.1809±.0007	.2853±.0007
	336	.1730±.0001	.2708±.0005	.1742±.0000	.2723±.0002	.1951±.0008	.2978±.0007
	720	.2132±.0006	.3048±.0007	.2123±.0007	.3036±.0006	.2297±.0017	.3236±.0136
Top1 count		2	1	0	0	3	4
Top2 count		6	4	1	0	3	2

G Detailed Full-shot Forecasting Results

In this section, we provide the full-shot forecasting results for all backbone architectures, including both MSE and MAE metrics. Note that while the MSE results are consistent with the main text, the MAE results are provided for a comprehensive comparison across all benchmarks.

G.1 Effect size and statistical comparisons

We compare OliO with TimesURL using unrounded three-seed means. Across 72 backbone–dataset–horizon configurations, OliO wins 58, with mean, median, and maximum relative MSE gains of 0.88%, 0.55%, and 5.75%. Counts use unrounded values and can therefore differ from ranks inferred from displayed rounded means.

Our primary analysis averages the four horizons within each of the 18 dataset–backbone groups: OliO wins 17 groups, and the two-sided exact Wilcoxon signed-rank test gives $W = 3$, $p = 3.81 \times 10^{-5}$. Averaging further over backbones gives six dataset-level pairs, all favoring OliO, with $W = 0$, $p = 0.03125$. The supplementary 72-configuration analysis gives $W = 347$, $p = 5.87 \times 10^{-9}$; its horizon-level entries need not be independent, so the grouped comparisons receive greater weight.

For a descriptive comparison with seed variation, let $\Delta = \bar{m}_{\mathrm{URL}} - \bar{m}_{\mathrm{OliO}}$ and let s_{URL} and s_{OliO} be the corresponding three-seed sample standard deviations. In 38 settings, $\Delta > \max(s_{\mathrm{URL}}, s_{\mathrm{OliO}})$; the symmetric criterion favors TimesURL in five settings. This compares effect magnitude with observed seed variation and is not a per-setting significance test. The tables below report the original per-setting means and sample standard deviations; the main-text summary averages these quantities separately and does not estimate uncertainty of the aggregated mean.

G.2 PatchTST Results

Table 33: Full-shot forecasting results for PatchTST [22] backbone (part 1). TS2Vec [32] is denoted as TS2V. Best results are in **bold** and the second best are underlined. (MTM, SCLT, Siam, and URL results are in Table 34.)

Dataset		Rand.		**OliO (Ours)**		TS2V [32]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3735 ± .0047	.3973 ± .0030	.3716 ± .0003	.3975 ± .0001	.3708 ± .0006	**.3953 ± .0003**
	192	.4094 ± .0041	.4191 ± .0023	.4082 ± .0009	**.4175 ± .0008**	.4080 ± .0008	.4190 ± .0006
	336	.4343 ± .0030	.4367 ± .0024	**.4252 ± .0000**	**.4315 ± .0002**	.4311 ± .0026	.4369 ± .0021
	720	.4582 ± .0089	.4711 ± .0047	**.4356 ± .0049**	**.4596 ± .0036**	.4483 ± .0016	.4691 ± .0015
ETTh2	96	.2833 ± .0093	.3418 ± .0030	**.2769 ± .0014**	.3400 ± .0004	.2778 ± .0017	**.3383 ± .0011**
	192	.3638 ± .0071	.3908 ± .0020	**.3392 ± .0002**	.3822 ± .0003	.3406 ± .0001	.3817 ± .0005
	336	.3918 ± .0014	.4131 ± .0011	**.3599 ± .0004**	.4020 ± .0004	.3606 ± .0003	.4017 ± .0001
	720	.3986 ± .0024	.4357 ± .0015	.3934 ± .0009	.4326 ± .0010	**.3910 ± .0020**	.4295 ± .0012
ETTm1	96	.2859 ± .0007	.3387 ± .0006	**.2830 ± .0015**	**.3368 ± .0010**	.2860 ± .0006	.3384 ± .0004
	192	.3266 ± .0005	.3653 ± .0003	**.3242 ± .0007**	.3660 ± .0011	.3279 ± .0002	.3680 ± .0003
	336	.3642 ± .0028	.3886 ± .0015	.3605 ± .0009	**.3877 ± .0005**	.3668 ± .0012	.3904 ± .0014
	720	.4140 ± .0015	.4204 ± .0010	**.4060 ± .0008**	**.4174 ± .0006**	.4162 ± .0009	.4221 ± .0013
ETTm2	96	.1660 ± .0006	.2561 ± .0004	**.1636 ± .0002**	**.2535 ± .0001**	.1669 ± .0001	.2574 ± .0001
	192	.2243 ± .0029	.2951 ± .0008	**.2189 ± .0003**	**.2920 ± .0002**	.2220 ± .0003	.2961 ± .0004
	336	.2770 ± .0013	.3299 ± .0006	**.2727 ± .0003**	**.3268 ± .0002**	.2760 ± .0003	.3310 ± .0004
	720	.3670 ± .0017	.3836 ± .0006	**.3644 ± .0008**	**.3820 ± .0005**	.3684 ± .0002	.3869 ± .0002
Weather	96	.1529 ± .0006	.2020 ± .0005	**.1507 ± .0003**	**.1997 ± .0011**	.1615 ± .0005	.2111 ± .0002
	192	.1959 ± .0007	.2412 ± .0002	**.1948 ± .0004**	**.2403 ± .0005**	.2036 ± .0003	.2485 ± .0003
	336	.2478 ± .0008	.2821 ± .0009	**.2453 ± .0003**	**.2785 ± .0007**	.2528 ± .0005	.2859 ± .0003
	720	.3221 ± .0011	.3347 ± .0013	**.3193 ± .0009**	**.3318 ± .0006**	.3242 ± .0004	.3360 ± .0004
ECL	96	.1356 ± .0000	.2302 ± .0000	.1297 ± .0002	.2228 ± .0003	.1314 ± .0002	.2249 ± .0002
	192	.1515 ± .0001	.2444 ± .0001	.1476 ± .0001	.2401 ± .0001	.1481 ± .0004	.2401 ± .0002
	336	.1682 ± .0002	.2614 ± .0001	**.1633 ± .0001**	.2572 ± .0001	.1653 ± .0004	.2583 ± .0001
	720	.2064 ± .0002	.2937 ± .0001	**.2001 ± .0005**	**.2896 ± .0005**	.2013 ± .0006	.2899 ± .0001
Top1 count		0	0	**18**	**15**	1	2
Top2 count		1	4	3	2	6	3

Table 34: Full-shot forecasting results for PatchTST [22] backbone (part 2). SimMTM [5] is denoted as MTM, SoftCLT [12] as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (Rand., OliO, and TS2V results are in Table 33.)

Dataset		MTM [5]		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3985 ± .0014	.4173 ± .0013	.3754 ± .0007	.3984 ± .0005	.3871 ± .0004	.4099 ± .0005	.3708 ± .0007	.3957 ± .0005
	192	.4299 ± .0011	.4355 ± .0011	.4095 ± .0010	.4179 ± .0008	.4192 ± .0008	.4286 ± .0007	**.4062 ± .0006**	.4187 ± .0011
	336	.4473 ± .0004	.4480 ± .0005	.4357 ± .0020	.4374 ± .0019	.4380 ± .0011	.4426 ± .0008	.4329 ± .0030	.4352 ± .0029
	720	.4560 ± .0021	.4744 ± .016	.4589 ± .0001	.4783 ± .0008	.4547 ± .0005	.4719 ± .0003	.4520 ± .0015	.4707 ± .0013
ETTh2	96	.2783 ± .0007	.3403 ± .0005	.2802 ± .0010	.3434 ± .0006	.2808 ± .0003	.3410 ± .0005	.2773 ± .0002	.3394 ± .0004
	192	.3431 ± .0005	.3827 ± .0007	.3441 ± .0009	.3844 ± .0005	.3440 ± .0005	**.3794 ± .0002**	.3410 ± .0004	.3812 ± .0004
	336	.3642 ± .0011	.4049 ± .0008	.3643 ± .0033	.4024 ± .0003	.3676 ± .0010	**.3989 ± .0004**	.3601 ± .0002	.4007 ± .0005
	720	.3949 ± .0004	.4312 ± .0004	.3960 ± .0031	.4342 ± .0021	.3925 ± .0006	**.4268 ± .0004**	.3926 ± .0010	.4298 ± .0010
ETTm1	96	.2839 ± .0008	.3412 ± .0004	.2917 ± .0017	.3463 ± .0015	.2948 ± .0009	.3450 ± .0002	.2837 ± .0012	.3381 ± .0005
	192	.3255 ± .0019	.3692 ± .0015	.3331 ± .0023	.3730 ± .0009	.3363 ± .0007	.3719 ± .0006	.3252 ± .0003	**.3650 ± .0003**
	336	**.3592 ± .0005**	.3895 ± .0008	.3692 ± .0019	.3947 ± .0010	.3672 ± .0014	.3927 ± .0006	.3627 ± .0005	.3890 ± .0005
	720	.4120 ± .0008	.4219 ± .0006	.4154 ± .0005	.4238 ± .0013	.4165 ± .0020	.4242 ± .0014	.4126 ± .0002	.4203 ± .0002
ETTm2	96	.1725 ± .0006	.2638 ± .0008	.1694 ± .0012	.2590 ± .0013	.1692 ± .0007	.2574 ± .0004	.1658 ± .0003	.2554 ± .0005
	192	.2278 ± .0008	.3036 ± .0004	.2249 ± .0026	.3009 ± .0031	.2279 ± .0003	.3005 ± .0008	.2210 ± .0004	.2947 ± .0005
	336	.2835 ± .0022	.3393 ± .0019	.2803 ± .0019	.3356 ± .0022	.2802 ± .0006	.3340 ± .0007	.2752 ± .0004	.3308 ± .0007
	720	.3706 ± .0010	.3902 ± .0010	.3722 ± .0010	.3913 ± .014	.3672 ± .0005	.3866 ± .0016	.3701 ± .0012	.3894 ± .0007
Weather	96	.1646 ± .0011	.2148 ± .0011	.1540 ± .0001	.2031 ± .0001	.1511 ± .0007	.2003 ± .0013	.1528 ± .0009	.2015 ± .0014
	192	.2043 ± .0014	.2489 ± .0010	.1975 ± .0005	.2434 ± .0009	.1949 ± .0002	.2406 ± .0002	.1976 ± .0003	.2428 ± .0000
	336	.2549 ± .0003	.2872 ± .0004	.2466 ± .0003	.2817 ± .0001	.2467 ± .0004	.2805 ± .0005	.2467 ± .0001	.2806 ± .0004
	720	.3253 ± .0005	.3368 ± .0004	.3201 ± .0003	.3339 ± .0005	.3207 ± .0013	.3326 ± .0011	.3216 ± .0003	.3334 ± .0003
ECL	96	.1387 ± .0000	.2352 ± .0001	.1325 ± .0003	.2265 ± .0003	.1300 ± .0001	.2234 ± .0001	**.1287 ± .0002**	**.2212 ± .0001**
	192	.1541 ± .0000	.2486 ± .0001	.1493 ± .0002	.2419 ± .0003	.1481 ± .0000	.2399 ± .0000	**.1471 ± .0001**	**.2394 ± .0001**
	336	.1707 ± .0001	.2657 ± .0002	.1656 ± .0002	.2589 ± .0003	.1672 ± .0003	.2585 ± .0001	.1638 ± .0007	.2572 ± .0009
	720	.2026 ± .0002	.2914 ± .0002	.2017 ± .0010	.2907 ± .0007	.2042 ± .0007	.2924 ± .0007	.2123 ± .0018	.3010 ± .0019
Top1 count		1	0	0	0	0	3	3	3
Top2 count		1	0	2	1	3	5	9	10

G.3 iTransformer Results

Table 35: Full-shot forecasting results for iTransformer [15] backbone (part 1). TS2Vec [32] is denoted as TS2V. Best results are in **bold** and the second best are underlined. (MTM, SCLT, Siam, and URL results are in Table 36.)

Dataset		Rand.		**OliO (Ours)**		TS2V [32]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**.3903 ± .0007**	**.4055 ± .0002**	.3969 ± .0002	.4101 ± .0003	.4027 ± .0001	.4155 ± .0001
	192	.4448 ± .0004	.4372 ± .0003	.4442 ± .0017	.4368 ± .0006	.4507 ± .0037	.4445 ± .0021
	336	.4884 ± .0030	.4612 ± .0022	.4865 ± .0037	.4607 ± .0027	.4875 ± .0034	.4638 ± .0020
	720	.5120 ± .0056	.4966 ± .0038	.4919 ± .0162	.4814 ± .0064	.4978 ± .0076	.4848 ± .0049
ETTh2	96	**.2988 ± .0029**	**.3493 ± .0019**	.3002 ± .0021	.3500 ± .0008	.3079 ± .0044	.3571 ± .0026
	192	**.3785 ± .0009**	.3979 ± .0010	.3811 ± .0026	.4000 ± .0009	.3923 ± .0051	.4084 ± .0033
	336	.4250 ± .0023	.4327 ± .0007	.4230 ± .0037	.4336 ± .0020	.4333 ± .0022	.4423 ± .0006
	720	.4357 ± .0050	.4500 ± .0028	**.4271 ± .0029**	.4465 ± .0010	.4373 ± .0018	.4535 ± .0014
ETTm1	96	.3456 ± .0011	.3782 ± .0007	**.3291 ± .0008**	**.3655 ± .0008**	.3367 ± .0016	.3697 ± .0010
	192	.3831 ± .0003	.3946 ± .0005	**.3711 ± .0015**	**.3876 ± .0011**	.3777 ± .0005	.3937 ± .0005
	336	.4205 ± .0011	.4183 ± .0007	**.4064 ± .0023**	**.4109 ± .0009**	.4138 ± .0020	.4159 ± .0007
	720	.4896 ± .0014	.4573 ± .0008	**.4730 ± .0027**	**.4492 ± .0013**	.4844 ± .0008	.4538 ± .0004
ETTm2	96	.1857 ± .0001	.2709 ± .0010	**.1812 ± .0022**	**.2637 ± .0019**	.1901 ± .0002	.2759 ± .0001
	192	.2528 ± .0006	.3128 ± .0008	**.2486 ± .0005**	**.3086 ± .0003**	.2548 ± .0004	.3151 ± .0004
	336	.3146 ± .0022	.3516 ± .0016	.3105 ± .0016	**.3486 ± .0005**	.3181 ± .0007	.3555 ± .0003
	720	.4138 ± .0001	.4074 ± .0008	**.4088 ± .0033**	**.4044 ± .0021**	.4184 ± .0005	.4105 ± .0005
Weather	96	.1776 ± .0005	.2182 ± .0006	**.1756 ± .0008**	**.2161 ± .0016**	.2139 ± .0011	.2630 ± .0020
	192	.2263 ± .0010	.2590 ± .0005	**.2256 ± .0012**	**.2585 ± .0009**	.2591 ± .0018	.2976 ± .2976
	336	.2829 ± .0013	**.2999 ± .0008**	.2825 ± .0007	.3002 ± .0006	.3075 ± .0009	.3270 ± .0024
	720	.3590 ± .0008	.3505 ± .0001	.3587 ± .0010	**.3496 ± .0009**	.3774 ± .0007	.3712 ± .0006
ECL	96	.1503 ± .0004	.2419 ± .0004	**.1497 ± .0003**	**.2416 ± .0004**	.2611 ± .0124	.3462 ± .0086
	192	**.1649 ± .0003**	**.2560 ± .0003**	.1671 ± .0011	.2582 ± .0012	.2646 ± .0084	.3503 ± .0056
	336	.1815 ± .0010	.2739 ± .0007	.1800 ± .0006	**.2733 ± .0007**	.2880 ± .0024	.3503 ± .0018
	720	.2131 ± .0002	.3016 ± .0002	**.2107 ± .0002**	**.3000 ± .0003**	.3216 ± .0110	.3922 ± .0079
Top1 count		4	4	**12**	**14**	0	0
Top2 count		2	5	9	6	0	0

Table 36: Full-shot forecasting results for iTransformer [15] backbone (part 2). SimMTM [5] is denoted as MTM, SoftCLT [12] as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (Rand., OliO, and TS2V results are in Table 35.)

Dataset		MTM [5]		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3993 ± .0006	.4118 ± .0004	.4040 ± .0006	.4152 ± .0002	.3914 ± .0011	.4057 ± .0009	.3924 ± .0009	.4076 ± .0006
	192	.4490 ± .0009	.4417 ± .0014	.4540 ± .0080	.4444 ± .0058	**.4408 ± .0013**	**.4342 ± .0005**	.4448 ± .0013	.4388 ± .0008
	336	.4927 ± .0030	.4634 ± .0026	.4932 ± .0027	.4661 ± .0004	**.4824 ± .0028**	**.4559 ± .0021**	.4868 ± .0007	.4612 ± .0004
	720	.5043 ± .0043	.4846 ± .0019	.5203 ± .0054	.4970 ± .0028	**.4812 ± .0064**	**.4738 ± .0030**	.5016 ± .0112	.4873 ± .0051
ETTh2	96	.3181 ± .0021	.3587 ± .0013	.3016 ± .0044	.3520 ± .0028	.3050 ± .0017	.3515 ± .0011	.3037 ± .0002	.3521 ± .0002
	192	.4040 ± .0047	.4095 ± .0019	.3871 ± .0043	.4058 ± .0040	.3794 ± .0016	**.3971 ± .0007**	.3891 ± .0047	.4057 ± .0037
	336	.4496 ± .0061	.4424 ± .0022	.4271 ± .0026	.4333 ± .0023	**.4180 ± .0024**	**.4296 ± .0013**	.4313 ± .0052	.4339 ± .0020
	720	.4386 ± .0010	.4479 ± .0004	.4412 ± .0071	.4541 ± .0045	.4289 ± .0029	**.4464 ± .0021**	.4398 ± .0034	.4519 ± .0022
ETTm1	96	.3341 ± .0002	.3732 ± .0002	.3368 ± .0008	.3712 ± .0001	.3435 ± .0015	.3764 ± .0007	.3315 ± .0022	.3660 ± .0013
	192	.3746 ± .0015	.3945 ± .0010	.3764 ± .0009	.3919 ± .0006	.3837 ± .0002	.3958 ± .0003	.3732 ± .0003	.3901 ± .0005
	336	.4083 ± .0007	.4157 ± .0003	.4097 ± .0007	.4140 ± .0001	.4175 ± .0003	.4173 ± .0001	.4124 ± .0018	.4146 ± .0011
	720	.4784 ± .0005	.4534 ± .0002	.4805 ± .0007	.4521 ± .0002	.4839 ± .0005	.4543 ± .0004	.4812 ± .0014	.4527 ± .0006
ETTm2	96	.1836 ± .0003	.2696 ± .0003	.1875 ± .0017	.2705 ± .0016	.1826 ± .0005	.2657 ± .0005	.1838 ± .0004	.2689 ± .0009
	192	.2488 ± .0008	.3103 ± .0006	.2539 ± .0021	.3125 ± .0010	.2536 ± .0043	.3113 ± .0031	.2517 ± .0011	.3123 ± .014
	336	**.3095 ± .0010**	.3498 ± .0005	.3173 ± .0008	.3516 ± .0006	.3203 ± .0022	.3530 ± .0012	.3134 ± .0011	.3501 ± .0010
	720	.4131 ± .0032	.4091 ± .0018	.4151 ± .0009	.4078 ± .0007	.4191 ± .0018	.4100 ± .0143	.4129 ± .0020	.4062 ± .0012
Weather	96	.2041 ± .0017	.2478 ± .0007	.2156 ± .0056	.2648 ± .0050	.1778 ± .0008	.2190 ± .0007	.1834 ± .0003	.2236 ± .0004
	192	.2499 ± .0006	.2821 ± .0003	.2570 ± .0016	.2913 ± .0027	.2262 ± .0012	.2604 ± .0008	.2314 ± .0004	.2629 ± .0005
	336	.3008 ± .0006	.3163 ± .0005	.3085 ± .0030	.3255 ± .0032	**.2819 ± .0010**	.3014 ± .0008	.2865 ± .0003	.3023 ± .0002
	720	.3706 ± .0009	.3609 ± .0008	.3798 ± .0007	.3706 ± .0005	**.3574 ± .0010**	.3502 ± .0004	.3623 ± .0005	.3518 ± .0006
ECL	96	.3656 ± .0010	.4260 ± .0007	.1646 ± .0004	.2567 ± .0004	.2849 ± .0011	.3656 ± .0009	.1508 ± .0003	.2435 ± .0004
	192	.3655 ± .0021	.4279 ± .013	.1789 ± .0005	.2695 ± .0006	.2908 ± .0023	.3718 ± .0111	.1659 ± .0004	.2572 ± .0004
	336	.3800 ± .0025	.4388 ± .0020	.1982 ± .0007	.2894 ± .0007	.3050 ± .0008	.3836 ± .0004	**.1797 ± .0001**	.2739 ± .0001
	720	.4121 ± .0006	.4584 ± .0005	.2394 ± .0005	.3244 ± .0003	.3138 ± .0007	.3855 ± .0006	.2114 ± .0002	.3015 ± .0001
Top1 count		1	0	0	0	6	6	1	0
Top2 count		3	2	0	2	5	3	5	5

G.4 TimeVLM Results

Table 37: Full-shot forecasting results for TimeVLM [33] backbone (part 1). TS2Vec [32] is denoted as TS2V. Best results are in **bold** and the second best are underlined. (MTM, SCLT, Siam, and URL results are in Table 38.)

Dataset		Rand.		**OliO (Ours)**		TS2V [32]	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3704 ± .0033	.4028 ± .0036	.3669 ± .0024	**.3979 ± .0000**	.3666 ± .0012	.3983 ± .0014
	192	.4074 ± .0006	.4301 ± .0015	.4029 ± .0046	.4251 ± .0000	.4040 ± .0023	.4278 ± .0008
	336	.4659 ± .0034	.4661 ± .0032	.4232 ± .0003	**.4399 ± .0000**	.4287 ± .0038	.4442 ± .0035
	720	.4981 ± .0066	.5011 ± .0037	.4587 ± .0069	.4752 ± .0000	.4634 ± .0095	.4824 ± .0063
ETTh2	96	.2818 ± .0004	.3456 ± .0009	.2755 ± .0000	.3407 ± .0000	.2766 ± .0004	.3407 ± .0009
	192	.3497 ± .0010	.3828 ± .0027	**.3359 ± .0002**	**.3811 ± .0000**	.3369 ± .0013	.3814 ± .0009
	336	.3767 ± .0079	.4093 ± .0036	**.3613 ± .0005**	**.4040 ± .0000**	.3621 ± .0016	.4051 ± .0010
	720	.4117 ± .0036	.4463 ± .0047	.4009 ± .0042	.4407 ± .0000	.4069 ± .0048	.4437 ± .0028
ETTm1	96	.2908 ± .0009	.3483 ± .0004	.2895 ± .0009	**.3468 ± .0000**	.2896 ± .0003	.3477 ± .0006
	192	.3354 ± .0015	.3759 ± .0004	**.3271 ± .0011**	**.3713 ± .0000**	.3310 ± .0022	.3751 ± .0011
	336	.3676 ± .0081	.3966 ± .0027	**.3593 ± .0008**	.3932 ± .0000	.3617 ± .0027	.3938 ± .0016
	720	.4209 ± .0012	.4297 ± .0022	.4096 ± .0029	.4256 ± .0000	.4102 ± .0030	.4256 ± .0021
ETTm2	96	.1690 ± .0006	.2603 ± .0005	.1674 ± .0004	.2584 ± .0000	.1673 ± .0011	.2569 ± .0011
	192	.2251 ± .0020	.2992 ± .0028	.2216 ± .0023	.2964 ± .0000	.2233 ± .0025	.2963 ± .0029
	336	.2745 ± .0027	.3313 ± .0009	.2733 ± .0014	.3293 ± .0000	.2702 ± .0017	.3292 ± .0013
	720	.3676 ± .0018	.3887 ± .0017	**.3521 ± .0014**	.3824 ± .0000	.3540 ± .0025	.3829 ± .0004
Weather	96	.1495 ± .0006	.2004 ± .0004	.1490 ± .0010	**.1988 ± .0000**	.1490 ± .0006	.1991 ± .0016
	192	.1937 ± .0005	.2428 ± .0008	.1937 ± .0003	.2431 ± .0000	.1938 ± .0100	**.2409 ± .0005**
	336	.2445 ± .0005	.2828 ± .0002	**.2433 ± .0002**	**.2793 ± .0000**	.2444 ± .0008	.2804 ± .0003
	720	.3180 ± .0005	.3342 ± .0004	.3148 ± .0003	.3313 ± .0000	**.3147 ± .0003**	**.3310 ± .0007**
ECL	96	.1353 ± .0002	.2346 ± .0004	.1319 ± .0002	.2264 ± .0000	**.1315 ± .0000**	.2266 ± .0001
	192	.1513 ± .0001	.2487 ± .0007	.1483 ± .0001	**.2412 ± .0000**	**.1478 ± .0001**	.2418 ± .0001
	336	.1673 ± .0007	.2648 ± .0008	**.1643 ± .0003**	**.2578 ± .0000**	.1670 ± .0005	.2629 ± .0006
	720	.2080 ± .0013	.3000 ± .0026	**.2044 ± .0005**	**.2924 ± .0000**	.2091 ± .0023	.3005 ± .0270
Top1 count		0	0	**8**	**11**	3	2
Top2 count		1	0	8	4	4	5

Table 38: Full-shot forecasting results for TimeVLM [33] backbone (part 2). SimMTM [5] is denoted as MTM, SoftCLT [12] as SCLT, TimeSiam [6] as Siam, and TimesURL [14] as URL. Best results are in **bold** and the second best are underlined. (Rand., OliO, and TS2V results are in Table 37.)

Dataset		MTM [5]		SCLT [12]		Siam [6]		URL [14]	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	.3666 ± .0016	.3984 ± .0017	**.3660 ± .0005**	.3988 ± .0004	.3779 ± .0001	.4085 ± .0000	.3672 ± .0008	.4000 ± .0008
	192	.4038 ± .0017	.4278 ± .0011	**.4017 ± .0015**	**.4223 ± .0013**	.4074 ± .0007	.4263 ± .0010	.4027 ± .0028	.4234 ± .0026
	336	.4341 ± .0073	.4533 ± .0062	.4232 ± .0042	.4401 ± .0042	.4285 ± .0011	.4428 ± .0011	.4250 ± .0067	.4414 ± .0064
	720	.4930 ± .0009	.4964 ± .0006	**.4556 ± .0049**	**.4736 ± .0041**	.4577 ± .0114	.4769 ± .0070	.4596 ± .0119	.4794 ± .0092
ETTh2	96	.2862 ± .0061	.3470 ± .0036	.2762 ± .0010	.3428 ± .0016	.2777 ± .0003	.3448 ± .0004	**.2754 ± .0006**	**.3401 ± .0007**
	192	.3409 ± .0018	.3868 ± .0014	.3377 ± .0004	.3835 ± .0021	.3409 ± .0018	.3865 ± .0017	.3368 ± .0011	.3822 ± .0021
	336	.3664 ± .0004	.4076 ± .0009	.3624 ± .0007	.4059 ± .0003	.3713 ± .0142	.4093 ± .0082	.3625 ± .0019	.4059 ± .0028
	720	.4073 ± .0038	.4438 ± .0022	.4119 ± .0032	.4493 ± .0032	.4218 ± .0024	.4599 ± .0023	**.3977 ± .0006**	**.4372 ± .0007**
ETTm1	96	.2899 ± .0005	.3478 ± .0003	.2903 ± .0008	.3472 ± .0007	.2940 ± .0016	.3515 ± .0008	.2895 ± .0004	.3480 ± .0005
	192	.3348 ± .0036	.3770 ± .0022	.3308 ± .0009	.3735 ± .0003	.3331 ± .0006	.3742 ± .0004	.3309 ± .0020	.3736 ± .0010
	336	.3663 ± .0027	.3967 ± .0012	.3615 ± .0018	.3930 ± .0004	.3619 ± .0004	.3919 ± .0010	.3605 ± .0003	**.3908 ± .0004**
	720	.4104 ± .0017	.4262 ± .0015	.4102 ± .0025	.4244 ± .0008	.4165 ± .0012	**.4236 ± .0010**	**.4081 ± .0012**	.4244 ± .0008
ETTm2	96	.1691 ± .0004	.2603 ± .0007	.1688 ± .0018	.2590 ± .0011	.1678 ± .0018	.2586 ± .0018	.1673 ± .0011	.2569 ± .0011
	192	.2238 ± .0008	.2978 ± .0011	**.2211 ± .0004**	.2951 ± .0007	.2227 ± .012	**.2950 ± .0014**	.2229 ± .0025	.2963 ± .0029
	336	.2728 ± .0040	.3311 ± .0031	**.2699 ± .0014**	.3279 ± .0015	.2716 ± .0015	**.3269 ± .0016**	.2712 ± .0035	.3293 ± .0028
	720	.3581 ± .0082	.3859 ± .0032	.3544 ± .0011	.3825 ± .0008	.3590 ± .0034	**.3823 ± .0025**	.3540 ± .0025	.3829 ± .0004
Weather	96	.1492 ± .0006	.1990 ± .0011	**.1488 ± .0005**	.1991 ± .0006	.1520 ± .0002	.2021 ± .0002	.1491 ± .0009	.1998 ± .0015
	192	.1946 ± .0003	.2431 ± .0006	.1942 ± .0009	.2411 ± .0009	.1956 ± .0001	.2422 ± .0002	.1938 ± .0003	.2412 ± .0009
	336	.2451 ± .0003	.2826 ± .0005	.2453 ± .010	.2808 ± .0007	.2453 ± .0004	.2809 ± .0003	.2445 ± .0003	.2822 ± .0007
	720	.3163 ± .011	.3335 ± .0005	.3170 ± .0008	.3334 ± .0008	.3151 ± .0008	.3321 ± .0005	.3149 ± .0001	.3312 ± .0004
ECL	96	.1318 ± .0003	.2264 ± .0003	.1325 ± .0001	.2276 ± .0001	.1346 ± .0002	.2321 ± .0002	.1363 ± .0004	.2330 ± .0007
	192	.1487 ± .0008	.2426 ± .0012	.1481 ± .0001	.2424 ± .0002	.1533 ± .0002	.2498 ± .0004	.1534 ± .0005	.2493 ± .0120
	336	.1665 ± .0010	.2612 ± .0130	.1651 ± .0001	.2605 ± .0002	.1692 ± .0110	.2647 ± .0011	.1727 ± .0003	.2720 ± .0004
	720	.2080 ± .0026	.2989 ± .0029	.2077 ± .0001	.2971 ± .0001	.2080 ± .0006	.2978 ± .0007	.2061 ± .0001	.2941 ± .0001
Top1 count		0	0	6	2	0	4	3	3
Top2 count		1	2	4	8	1	1	6	3

H Detailed comparative analysis of performance improvements

Figs. 15, 16, and 17 show backbone-specific forecasting comparisons. Each cell averages the performance index $\mathrm{MSE}_{\mathrm{Rand}}/\mathrm{MSE}_{\mathrm{Method}}$ across the six benchmark datasets at a given prediction horizon. Values above 1.0 correspond to lower MSE than random initialization.

Figs. 18, 19, 20, 21, 22, and 23 instead average the index across the three backbones for each dataset. Together, the two views show how the relative benefit of pre-training varies across architectures, datasets, and horizons. App. G provides per-setting MSE and sample standard deviations.

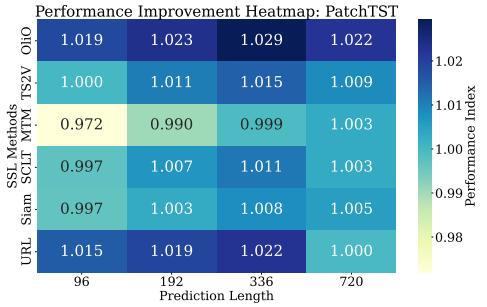


Figure 15: PatchTST backbone.

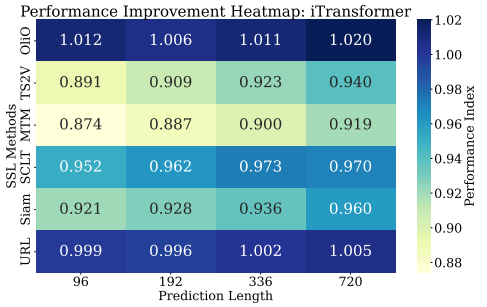


Figure 16: iTransformer backbone.

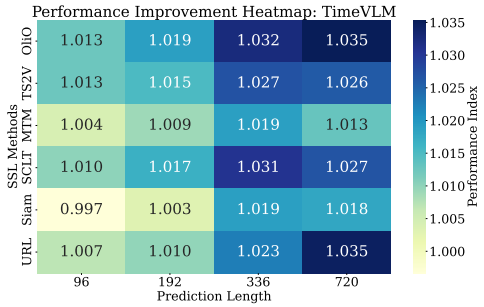


Figure 17: TimeVLM backbone.

Appendix figure: detailed performance improvement heatmaps for each backbone architecture. The values represent the mean performance index (Random MSE / Method MSE) aggregated across six benchmark datasets. Dark blue cells indicate higher performance indices.

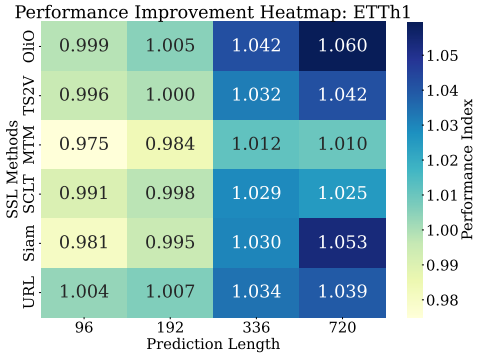


Figure 18: ETTh1 dataset.

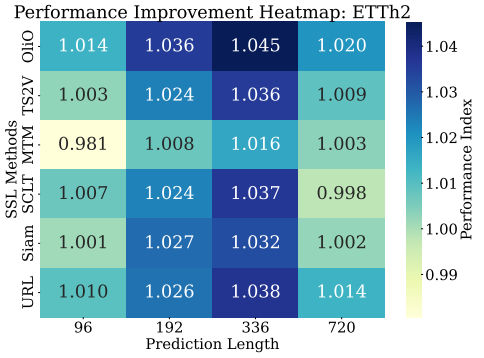


Figure 19: ETTh2 dataset.

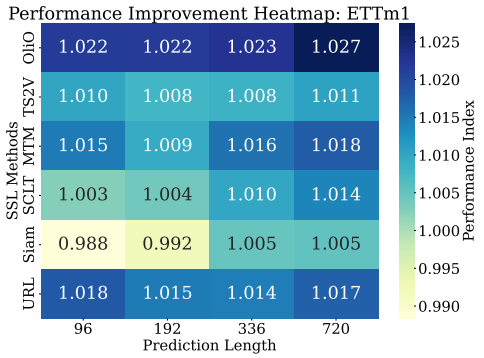


Figure 20: ETTm1 dataset.

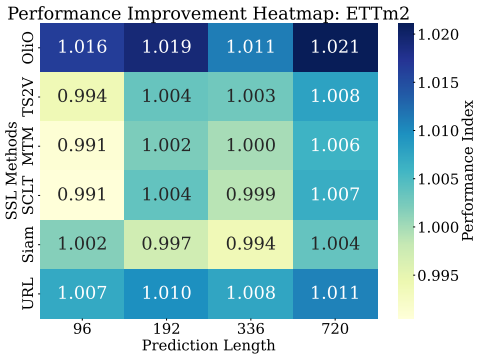


Figure 21: ETTm2 dataset.

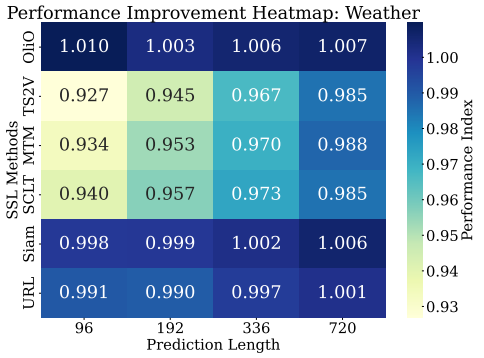


Figure 22: Weather dataset.

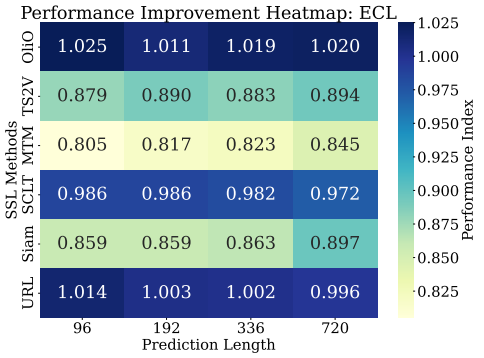


Figure 23: ECL dataset.

Appendix figure: performance improvement heatmaps aggregated over all backbone architectures for each benchmark dataset. The values represent the mean performance index (Random MSE / Method MSE). These visualizations summarize how relative forecasting performance varies across datasets and prediction horizons.

I Detailed ablation study results

In this section, we provide detailed ablation results complementing the analyses in Sec. 4.4. We first report backbone-specific results comparing the bidirectional formulation of OliO with its unidirectional variants in Table 39. We then provide complete results for the transition-matrix constraint and approximation order p in Table 40.

We then report overlap-ratio sweeps, endpoint comparisons, and fixed-pair masking measurements, followed by alternative transition parameterizations. Higher-order transition measurements and finite-range distance ordering complete the analysis.

Table 39: Detailed ablation study results for each backbone architecture. Values represent the average MSE across all prediction horizons $T \in \{96, 192, 336, 720\}$.

Backbone	Method	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL
PatchTST	**OliO (Bi)**	**.4103**	**.3424**	.3434	**.2549**	**.2275**	**.1602**
	Forward Only	.4133	.3431	.3474	.2601	.2137	.2063
	Backward Only	.4126	.3428	**.3432**	.2601	.2303	.1621
iTransformer	**OliO (Bi)**	**.4549**	**.3829**	**.3949**	**.2873**	**.2606**	**.1769**
	Forward Only	.5181	.4015	.4877	.3027	.2867	.3313
	Backward Only	.5117	.3996	.4895	.3047	.2798	.3608
TimeVLM	**OliO (Bi)**	**.4129**	.3434	.3464	.2536	**.2252**	**.1622**
	Forward Only	.4142	**.3430**	.3464	.2515	.2256	.1625
	Backward Only	.4151	**.3430**	**.3463**	**.2503**	.2256	.1625

Table 40: Full computation analysis: evaluation of transition matrix constraints and approximation order $p = 1 \ldots 5$. Values represent the average MSE across all backbones.

Matrix Type	Order (p)	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL
Unconstrained (Full)	Exact	NaN	NaN	NaN	NaN	NaN	NaN
Nilpotent (OliO)	**Exact**	**.4260**	**.3562**	**.3616**	**.2653**	**.2378**	**.1664**
Nilpotent (OliO)	$p = 1$	.4434	.3661	.3752	.2674	.2470	.2137
Nilpotent (OliO)	$p = 2$	.4315	.3624	.3695	.2661	.2480	.2063
Nilpotent (OliO)	$p = 3$	.4315	.3606	.3686	.2678	.2382	.2029
Nilpotent (OliO)	$p = 4$	.4325	.3629	.3720	.2675	.2457	.2051
Nilpotent (OliO)	$p = 5$	.4299	.3646	.3678	.2669	.2464	.1957

I.1 Effect of window overlap

For windows of length l with start separation $0 \leq \tau \leq l$, the overlap ratio is $1 - \tau/l$. We vary τ while keeping the window length, backbone, and remaining training configuration fixed. Table 41 reports forecasting MSE across four overlap ratios.

Table 41: Forecasting MSE under different overlap ratios. Lower is better.

Setting	90%	80%	60%	20%
PatchTST–ETTh1	0.421	0.418	**0.417**	**0.417**
PatchTST–ETTh2	**0.343**	**0.343**	**0.343**	0.344
iTransformer–ETTm1	0.404	0.406	0.403	**0.400**
iTransformer–ETTm2	**0.299**	0.305	0.306	0.306

The lowest reported MSE occurs at different overlap ratios across the four settings. Increasing overlap produces no monotonic improvement in forecasting MSE. Both lag and shared content vary in this sweep; the fixed-pair masking analysis in App. I.3 holds the pair and lag constant.

For Weather, we additionally report an extended sweep spanning both overlap endpoints in Table 42. The preferred ratio depends on the backbone, and increasing overlap does not monotonically improve forecasting MSE.

Table 42: Extended overlap-ratio sweep on Weather. Lower MSE is better.

Backbone	100%	90%	80%	60%	40%	20%	0%
PatchTST	.2299	.2298	.2288	.2288	**.2279**	.2288	.2287
iTransformer	.2871	.2681	.2725	.2818	.2752	**.2668**	**.2668**

I.2 Zero-overlap analysis

Table 43 compares the backbone without OliO with two overlap endpoints. At 100% overlap, $\tau = 0$ and $\mathcal{T}_0(\boldsymbol{h}) = \boldsymbol{h}$, giving a zero-lag control. At 0% overlap, the windows share no observations. These endpoint conditions supplement the standard pre-training range $0 < \tau < l$.

Table 43: Comparison of models without OliO, OliO with full-overlap pairs, and OliO with zero-overlap pairs. Results are averaged over three seeds. Lower MSE is better.

Setting	w/o OliO	100% Overlap	0% Overlap
PatchTST–ETTh1	0.4189	**0.4178**	0.4179
PatchTST–ETTh2	0.3594	**0.3416**	0.3437
PatchTST–Weather	0.2297	0.2299	**0.2287**
iTransformer–ETTm1	0.4097	0.4068	**0.4060**
iTransformer–ETTm2	0.2917	**0.2889**	0.2908
iTransformer–Weather	**0.2615**	0.2871	0.2668

In five of the six settings, the MSE difference between the zero-overlap and full-overlap conditions is within 0.7%. The zero-overlap condition also has lower MSE than the corresponding backbone without OliO in five settings. iTransformer–Weather favors the backbone without OliO, illustrating variation across the evaluated configurations.

I.3 Matched masking control

We further conduct a controlled masking experiment at 50% overlap. The underlying window pair, signed temporal lag, mask ratio, backbone, training schedule, and hyperparameters are held identical. The shared half of the second window is set to zero in the overlap-mask condition; an equal-sized unique half is set to zero in the unique-mask condition. Masked observations retain their original positions, and the input length is unchanged.

Table 44: Matched masking control at 50% overlap. MSE is averaged over ETTh1, ETTh2, ETTm1, ETTm2, Weather, and ECL and horizons $\{96, 192, 336, 720\}$ (24 settings per backbone; 48 overall). "Wins" reports Overlap-Mask / Unique-Mask wins. Lower MSE is better.

Backbone	Overlap Mask	Unique Mask	Gap	Wins
PatchTST	0.29818	**0.29767**	+0.17%	9 / 15
iTransformer	0.35022	**0.34932**	+0.26%	14 / 10
Overall	0.32420	**0.32349**	+0.22%	23 / 25

The aggregate MSE difference is 0.22%. Overlap masking yields lower MSE in 23 settings, and unique masking in 25 settings. The equal-budget conditions therefore produce closely matched aggregate errors and balanced setting-level wins. This diagnostic measures sensitivity to mask location for fixed window pairs and signed lags.

I.4 Alternative transition parameterizations

The parameterization comparison in Sec. 4.4 uses PatchTST and iTransformer across six datasets and four horizons, giving 48 settings. For signed lag s, Bias-only uses $\boldsymbol{h} + s\boldsymbol{v}$, and K-only uses $e^{sK}\boldsymbol{h}$. MLP replaces the complete transition with a generic nonlinear mapping. Spec-K uses a full generator regularized by spectral normalization in place of the strictly upper-triangular generator.

Across the 48 backbone–dataset–horizon settings, OliO achieves an average MSE of 0.3080, compared with 0.3116 for Spec-K, 0.3123 for K-only, 0.3176 for MLP, and 0.3198 for Bias-only. OliO achieves lower MSE than these alternatives in 39/48, 42/48, 45/48, and 47/48 settings, respectively.

Table 45 gives the dataset–backbone averages. Bias-only has the largest aggregate MSE increase among these alternatives. The dataset-level results also show variation: K-only has the lowest MSE on PatchTST–ETTm1, and Spec-K on iTransformer–Weather. Spec-K trains stably in these settings, providing a full-generator comparison alongside the unconstrained generator in Table 40.

Table 45: Dataset-averaged MSE of alternative transition parameterizations. Each value is averaged over four forecasting horizons. Lower MSE is better.

Backbone	Dataset	OliO	Bias-only	K-only	MLP	Spec-K
iTransformer	ECL	**.1769**	.1904	.1786	.1954	.1785
	ETTh1	**.4549**	.4581	.4592	.4600	.4596
	ETTh2	**.3829**	.3865	.3879	.3888	.3873
	ETTm1	**.3949**	.4731	.4005	.4308	.3986
	ETTm2	**.2873**	.2901	.2883	.3035	.2894
	Weather	.2606	.2789	.2840	.2805	**.2600**
PatchTST	ECL	**.1602**	.1634	.1605	.1620	.1605
	ETTh1	**.4102**	.4200	.4132	.4144	.4158
	ETTh2	**.3424**	.3433	.3437	.3433	.3440
	ETTm1	.3434	.3484	**.3431**	.3440	.3458
	ETTm2	**.2549**	.2573	.2577	.2594	.2722
	Weather	**.2275**	.2282	.2309	.2293	.2278

I.5 Higher-order transition analysis

We evaluate the matrix component of the transition at all integer lags from 1 to 64. Table 46 reports eight representative lags, extending the summary in Table 10. The diagnostic compares the exact matrix exponential with its first-order approximation; the additive offset is identical in both affine maps and is excluded from the relative matrix error.

We measure the relative approximation error between the ODE-derived matrix transport and its first-order approximation as

$$r(\tau) = \frac{\left\| e^{\tau K} - (I + \tau K) \right\|_F}{\left\| e^{\tau K} \right\|_F}. \tag{22}$$

Table 46: Relative approximation error between the complete transition and its first-order approximation across temporal lags.

Lag τ	ETTh1	ETTh2
1	0.018	0.000
10	0.617	0.034
20	0.857	0.104
30	0.931	0.182
40	0.965	0.258
50	0.986	0.331
60	1.001	0.402
64	1.004	0.430

For the two evaluated operators, relative error increases monotonically over lags 1–64. The discrepancy is smaller at short lags and grows as higher-order terms contribute more to the matrix transition. The order ablation in Table 40 reports the corresponding forecasting results for separately trained truncation orders.

I.6 Distance ordering over the evaluated lag range

We evaluate the learned checkpoints using the complete affine transition, including its shared lag-dependent offset. For each encoded state $\boldsymbol{h}$, we compute the adjacent-lag increment

$$\Delta_h(k) = \|\mathcal{T}_{+(k+1)}(\boldsymbol{h}) - \boldsymbol{h}\|_2^2 - \|\mathcal{T}_{+k}(\boldsymbol{h}) - \boldsymbol{h}\|_2^2, \qquad k = 0, \ldots, 7. \tag{23}$$

Complete pointwise distance ordering holds for 88%–100% of evaluated representations across checkpoints. At every adjacent lag in every evaluated checkpoint, both the mean and median of $\Delta_h(k)$ are positive. Every evaluated representation also has positive initial transition-velocity norm $\|K\boldsymbol{h} + \boldsymbol{v}\|_2$. These measurements cover integer lags 0 through 8. App. B.2 gives the analytical local and asymptotic properties of fixed-source displacement.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction explicitly detail the main contribution.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations regarding the expressivity of low-order polynomial expansions for highly complex datasets are discussed in Section 4. The impact of static text prompts in specific backbone models (e.g., TimeVLM) on latent visualization is addressed in Section A.3. Computational efficiency and scaling analysis are provided in Section 3.4.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides formal statements and assumptions for all theoretical results, including the polynomial distance relationship. Detailed proofs for these results, along with auxiliary Corollary are provided in Section B.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Detailed setup, including hardware, software, and hyperparameter search space, is provided in Section D.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: An anonymous GitHub repository containing the implementation code and instructions is provided at `https://anonymous.4open.science/r/OliO-46EF`.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Detailed pre-training and fine-tuning configurations, including the hyperparameter search space for each backbone, are provided in Sections C and D.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Experimental results are reported with standard deviations calculated over three independent runs with different random seeds, as shown in the result tables in Section G.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Hardware specifications (single NVIDIA RTX 3090 GPU), compute environment, average execution times, and total compute estimates are detailed in Section D.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in this paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work focuses on foundational representation learning for time series. As such, it does not directly address specific societal applications or impacts.

    Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The proposed framework is a general time series forecasting tool and does not pose significant risks for misuse compared to large-scale generative models.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All benchmark datasets (ETT, Weather, Electricity) are properly credited and cited, and were used in accordance with their research-oriented terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: No new assets (like new datasets or physical artifacts) are introduced in this paper.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This research does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This research does not involve human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were not used as an important, original, or non-standard component of the core methods in this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.