Aurora: Towards Universal Generative Multimodal Time Series Forecasting

Xingjian Wu, Jianxin Jin, Wanghui Qiu, Peng Chen, Yang Shu, Bin Yang, Chenjuan Guo✉
{xjwu,jxjin,onehui,pchen}@stu.ecnu.edu.cn, {yshu,byang,cjguo}@dase.ecnu.edu.cn
East China Normal University

You can subscribe to our Lab:
DecisionIntelligence @ ECNU

My WeChat | Paper | Code

Introduction

Cross-domain generalization is very important in Time Series Forecasting because similar historical information may lead to distinct future trends due to different domain-specific characteristics. As shown in Figure 1, Current research of time series forecasting explores the cross-domain adaptation in two main perspectives:

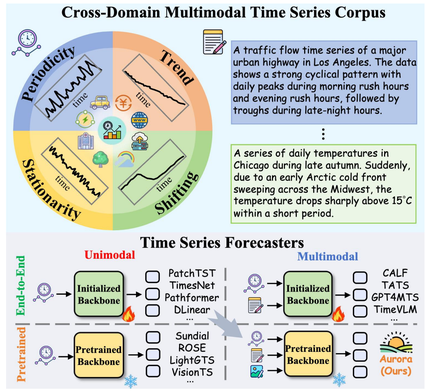


Figure 1 Aurora is pretrained on cross-domain multimodal time series corpus, supporting both text and image information to enhance zero-shot time series forecasting.

1) pre-training on cross-domain time series corpus for unimodal time series foundation models, which partially possess cross-domain generalization capabilities.
2) utilizing cross-modality information in training end-to-end multimodal supervised models, which effectively integrates domain knowledge in forecasting.

For time series foundation models, their capabilites come from single time modality and lack explict domain knowledge guidance, thus hindering the performance; For end-to-end multimodal supervised models, they lack the ability to support zero-shot forecasting in cross-domain scenarios.

Contributions

- We propose a multimodal time series foundation model, called **Aurora**, which is pretrained on cross-domain multimodal time series corpus and supports generative probabilistic forecasting. Through effectively fusing multimodal information during pretraining, Aurora serves as a strong zero-shot forecaster, and can make accurate cross-domain inference.
- We devise a novel cross-modality encoder in Aurora, consisting of token distillation and modality guiding, implemented by merticulously-designed attention structures. It can enhance the temporal representations while effectively fusing representations from texts and images.
- We design a novel flow-matching process in the Aurora Decoder. It obtains multimodal conditions through a Transformer, and obtains future prototypes containing periodic and trend information as the starting points, thus enhancing the ability of flow-matching.
- Experimentally, Aurora achieves state-of-the-art performance on 5 well-recognized benchmarks, including datasets from TimeMMD, TSFM-Bench, ProbTS, TFB, and EPF, covering comprehensive scenarios, i.e., unimodal, multimodal, deterministic, and probabilistic, thus demonstrating a strong out-of-the-box tool of decision intelligence.

Aurora Architecture

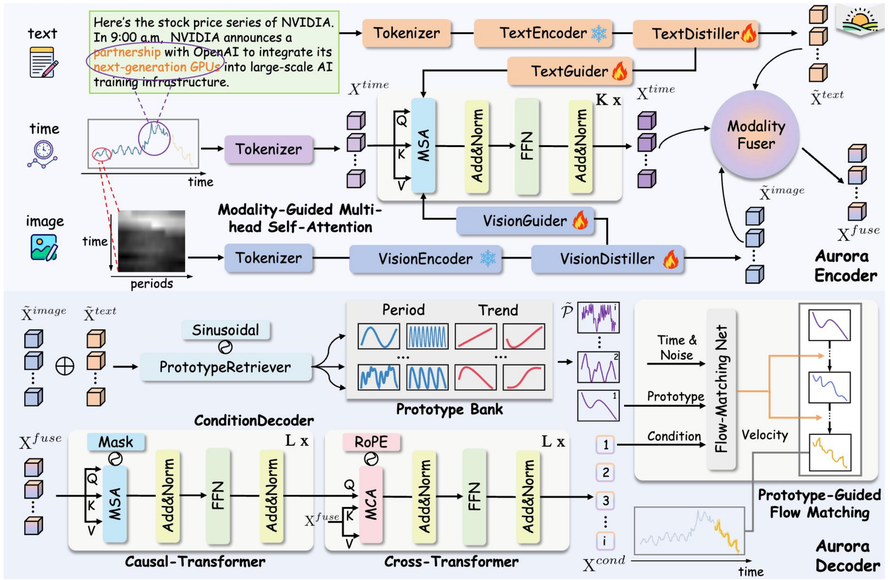


Figure 2 The overview of Aurora. n the Aurora Encoder, the multimodal information is extracted, distilled, and fused. Modality-Guided Multi-head Self-Attention is introduced to inject the domainspecific knowledge into temporal modeling. In the Aurora Decoder, the Prototype-Guided Flow Matching is introduced to support generative probabilistic forecasting.

Modality-Guided Self-Attention Mechanism

$$\mathrm{VAttn} = \mathrm{VisionGuider}(X^{time}, X^{image}) \in \mathbb{R}^{n^{time} \times K^{image}},$$

$$\mathrm{TAttn} = \mathrm{TextGuider}(X^{time}, X^{text}) \in \mathbb{R}^{n^{time} \times K^{text}},$$

$$\mathrm{Corr} = \mathrm{VAttn} \cdot W \cdot \mathrm{TAttn}^T \in \mathbb{R}^{n^{time} \times n^{time}},$$

$$Q = X^{time} \cdot W^Q,\ K = X^{time} \cdot W^K,\ V = X^{time} \cdot W^V$$

$$S = (Q \cdot K^T + \mathrm{Corr})/\sqrt{d^{time}}, O = \mathrm{Softmax}(S) \cdot V,$$

$$O^{norm} = \mathrm{LayerNorm}(X^{time} + O),$$

$$X^{time} = \mathrm{LayerNorm}(\mathrm{FeedForward}(O^{norm}) + O^{norm}),$$

Prototype-Guided Flow Matching

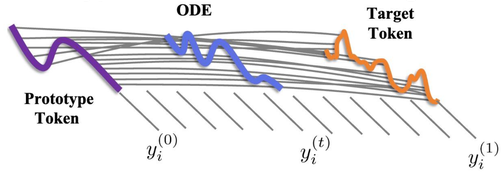


Figure 3 Prototype-Guided Flow Matching. The starting point is set as a prototype instead of a random gaussian noise, which provides an intuitive guidance in generation process.

Algorithm 1 Prototype-Guided Flow Matching

Given condition X_i^{cond}, steps J, and Prototype $\tilde{\mathcal{P}}_i$.
Sample a noise $\epsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
$\Delta t = 1/J, h_i = X_i^{cond}, \hat{y}_i = \tilde{\mathcal{P}}_i + \epsilon_i$
for j **in** $\{0, 1 \ldots, J-1\}$ **do**
  $\hat{y}_i \leftarrow \hat{y}_i + v^{\theta}_{j\Delta t}(\hat{y}_i | h_i)\Delta t$
end for
Return: $\hat{y}_i$

Experiments

Main Results

Table 1: MSE and MAE results on TimeMMD datasets.

Type	Zero-shot Foundation Models										10% few-shot		Full-shot Multimodal End-to-end Supervised Models					
Models	Aurora (Ours)		Sundial (2025)		VisionTS (2025)		ROSE (2025)		MOIRAI (2024)		Aurora (Ours)		GPT4MTS (2025)		TATS (2025)		CALF (2025)		Time-VLM (2025)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	0.272	0.348	0.373	0.392	0.290	0.336	0.345	0.372	0.272	0.403	0.212	0.293	0.225	0.298	0.215	0.301	0.250	0.315	0.237	0.302
Climate	0.865	0.749	1.154	0.881	1.307	0.930	1.475	0.987	1.921	1.095	0.862	0.746	1.182	0.889	1.180	0.887	1.286	0.922	1.195	0.899
Economy	0.033	0.146	0.291	0.432	0.301	0.442	0.289	0.433	0.405	0.512	0.016	0.099	0.017	0.103	0.017	0.104	0.163	0.307	0.024	0.125
Energy	0.255	0.370	0.272	0.367	0.304	0.420	0.386	0.479	0.324	0.417	0.230	0.329	0.262	0.380	0.255	0.368	0.244	0.365	0.260	0.374
Environment	0.276	0.379	0.336	0.416	0.354	0.436	0.392	0.456	0.351	0.403	0.265	0.372	0.323	0.400	0.319	0.396	0.325	0.387	0.319	0.397
Health	1.553	0.850	1.970	0.992	2.436	1.221	2.598	1.201	2.736	1.241	1.343	0.776	1.464	0.799	1.356	0.767	1.491	0.775	1.489	0.834
Security	72.475	4.084	70.441	4.005	79.598	4.597	84.324	4.765	93.245	5.173	70.062	3.988	71.487	4.068	72.406	4.097	76.376	4.300	73.731	4.181
Social Good	0.838	0.516	1.036	0.573	1.126	0.618	1.141	0.581	1.430	0.651	0.814	0.494	0.920	0.450	0.918	0.428	0.906	0.401	0.868	0.444
Traffic	0.161	0.289	0.271	0.405	0.281	0.407	0.341	0.451	0.406	0.468	0.157	0.290	0.203	0.261	0.179	0.238	0.222	0.293	0.216	0.319
1st Count	31	26	4	7	0	4	0	0	1	0	30	23	1	1	4	4	1	8	0	0

Table 2: MSE and MAE results on TSFM-Bench datasets.

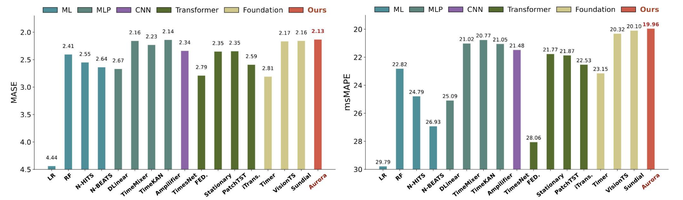
Type	Zero-shot Foundation Models																			
Models	Aurora (Ours)		Sundial (2025)		ROSE (2025)		Timer (2024)		TimesFM (2023)		Chronos (2024)		Time-MoE (2024)		UniTS (2024)		MOIRAI (2024)		TTM (2024)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT (Avg)	0.331	0.376	0.335	0.379	0.393	0.411	0.551	0.478	0.415	0.406	0.442	0.408	0.357	0.390	0.471	0.437	0.382	0.388	0.441	0.430
Weather	0.230	0.267	0.234	0.270	0.265	0.305	0.292	0.313	-	-	0.288	0.309	0.256	0.289	0.275	0.298	0.260	0.275	0.265	0.307
Electricity	0.178	0.275	0.169	0.265	0.234	0.320	0.297	0.375	-	-	-	-	-	-	0.198	0.291	0.188	0.273	0.222	0.317
Traffic	0.524	0.352	-	-	0.588	0.412	0.613	0.407	-	-	0.615	0.421	-	-	-	-	-	-	0.564	0.386
Solar	0.203	0.289	0.221	0.252	0.505	0.549	0.771	0.604	0.500	0.397	0.393	0.319	0.411	0.428	0.845	0.669	0.714	0.704	0.815	0.710
PEMS08	0.563	0.552	-	-	1.369	0.979	0.866	0.695	1.485	0.907	1.707	1.024	-	-	1.253	0.879	-	-	1.730	1.066
Wind	1.151	0.763	1.186	0.772	1.251	0.820	1.201	0.783	1.613	0.870	1.478	0.834	-	-	1.425	0.848	1.299	0.795	1.337	0.829
NYSE	0.528	0.526	0.880	0.642	-	-	0.988	0.704	0.623	0.536	1.129	0.720	-	-	1.220	0.820	-	-	-	-
1st Count	27	21	11	13	3	1	0	0	1	2	0	0	2	2	0	0	0	5	0	0

Table 3: CRPS and NMAE results on ProbTS datasets.

Type	Zero-shot Foundation Models										Full-shot Probabilistic End-to-end Supervised Models					
Models	Aurora (Ours)		Sundial (2025)		Chronos (2024)		MOIRAI (2024)		Lag-Llama (2023)		TSDiff (2023)		CSDI (2022)		TimeGrad (2022)		GRU MAF (2021)	
Metrics	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE	CRPS	NMAE
ETT (Avg)	0.231	0.257	0.231	0.273	0.290	0.316	0.366	0.377	0.273	0.310	0.370	0.465	0.304	0.389	0.493	0.619	0.388	0.475
Weather	0.070	0.076	0.087	0.102	0.142	0.158	0.179	0.143	0.096	0.106	0.132	0.134	0.077	0.093	0.125	0.155	0.133	0.165
Electricity	0.085	0.103	0.081	0.098	-	-	0.247	0.290	-	-	0.407	0.519	/	/	0.102	0.126	0.094	0.122
Traffic	0.220	0.262	-	-	0.269	0.295	-	-	0.330	0.385	0.327	0.392	/	/	0.225	0.264	/	/
Exchange	0.044	0.047	0.045	0.049	0.044	0.047	0.045	0.050	0.057	0.069	0.084	0.111	0.069	0.086	0.082	0.095	0.070	0.083
ILI	0.147	0.166	0.148	0.166	0.170	0.197	0.159	0.197	0.156	0.211	0.248	0.259	0.276	0.290	0.284	0.310	0.262	0.288
1st Count	19	24	8	8	1	1	2	1	0	0	0	0	4	1	1	1	0	0

Table 4: MSE and MAE results on EPF datasets.

Type	Zero-shot Foundation Models										Full-shot End-to-end Supervised Models							
Models	Aurora (Ours)		Sundial (2025)		VisionTS (2025)		ROSE (2025)		MOIRAI (2024)		TimeXer (2024)		iTransformer (2024)		PatchTST (2023)		TimesNet (2023)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	0.288	0.312	0.256	0.277	0.510	0.461	0.666	0.536	0.660	0.538	0.238	0.268	0.265	0.300	0.267	0.284	0.250	0.289
PJM	0.084	0.183	0.088	0.189	0.251	0.366	0.311	0.402	0.330	0.423	0.088	0.188	0.097	0.197	0.106	0.209	0.097	0.195
BE	0.361	0.257	0.371	0.270	0.679	0.457	0.815	0.514	0.837	0.534	0.374	0.241	0.394	0.270	0.403	0.264	0.419	0.288
FR	0.387	0.206	0.392	0.207	0.625	0.393	0.746	0.447	0.751	0.454	0.381	0.211	0.439	0.233	0.411	0.220	0.431	0.234
DE	0.539	0.475	0.541	0.484	0.961	0.687	1.276	0.778	1.251	0.779	0.440	0.418	0.479	0.433	0.461	0.432	0.502	0.446



(a) Mean MASE results. (b) Mean msMAPE results.

Figure 4 MASE and msMAPE results on TFB-univariate datasets.