MOMENT

OPEN-SOURCE foundation models for TIME SERIES analysis

MOMENT: Open-Source Foundation Models for Time Series Analysis

Q: What is the focus of the research paper?
A: MOMENT is a family of open-source foundation models for general-purpose time series analysis. It addresses the challenges of pre-training large models on time series data by compiling a large and diverse collection of public time series called the Time Series Pile and systematically tackling time series-specific challenges. The key contributions of MOMENT include pre-training data, multi-dataset pre-training, and evaluation on diverse tasks and datasets in limited supervision settings

Q: What is MOMENT?
A: MOMENT is a family of open-source foundation models designed for general-purpose time series analysis.

Q: What are the key challenges MOMENT addresses?
A: MOMENT tackles the following challenges in pre-training large time series models:
* Lack of large, cohesive public time series datasets.
* Diverse time series characteristics (sampling rates, channels, lengths, missing values) making multi-dataset training difficult.
* Lack of comprehensive benchmarks to evaluate these models, especially in limited supervision settings.

Q: How does MOMENT address these challenges?
A: MOMENT addresses these challenges by:
* Compiling a large and diverse collection of public time series data called the "Time Series Pile."
* Systematically tackling time series-specific challenges to enable large-scale multi-dataset pre-training.
* Developing a benchmark to evaluate time series foundation models on diverse tasks and datasets in limited supervision settings.

Q: What are the key contributions of MOMENT?
A: MOMENT's key contributions include:
* Pre-training data: The "Time Series Pile" provides a valuable resource for training large time series models.
* Multi-dataset pre-training: MOMENT effectively handles diverse time series data from multiple sources.
* Evaluation: A new benchmark facilitates comprehensive evaluation of time series models in various scenarios.
* Open-source: MOMENT is freely available, promoting accessibility and collaboration in the field.

Q: What datasets were used for pre-training and evaluation?
A:
* Pre-training: The Time Series Pile, comprising data from diverse domains like healthcare, economics, weather, etc.
* Evaluation: M3 and M4 datasets for forecasting; a subset of the UCR Classification Archive for classification; a subset of the UCR Anomaly Archive for anomaly detection.

Q: What are some empirical observations about MOMENT?
A:
* Superior performance: MOMENT excels in anomaly detection and classification, particularly with smaller datasets.
* Effective in limited supervision settings: Performs well in tasks like zero-shot imputation, linear probing for forecasting, and unsupervised representation learning for classification.
* Transparency: MOMENT exhibits high upstream transparency, but model transparency scores are lower due to less understood aspects of time series modeling.

Q: Where can I find MOMENT?
A: Pre-trained models and the Time Series Pile are available on Hugging Face: https://huggingface.co/AutonLab

MOMENT: Open-Source Foundation Models for Time Series Analysis

Q: What types of time series analysis tasks can MOMENT perform?

A: MOMENT can be used for a variety of time series analysis tasks, including:
* Forecasting (short- and long-horizon)
* Classification
* Anomaly detection
* Imputation

Q: How does MOMENT compare to other time series models?
A: MOMENT demonstrates superior performance compared to state-of-the-art deep learning and statistical baselines, especially in tasks like anomaly detection and classification. This is attributed to its pre-training on a large and diverse dataset.

Q: What are the benefits of using a pre-trained time series model like MOMENT?
A: Pre-trained models like MOMENT offer several advantages:
* Faster training: Fine-tuning a pre-trained model requires less time and data than training a model from scratch.
* Improved performance: Pre-training on a large dataset typically leads to better performance on downstream tasks.
* Generalization: Pre-trained models can often generalize well to new, unseen datasets.

Q: What are some potential applications of MOMENT?
A: MOMENT's capabilities can be applied to various domains, including:
* Healthcare: Forecasting patient health, detecting anomalies in medical signals.
* Finance: Predicting stock prices, identifying fraudulent transactions.
* Environmental science: Forecasting weather patterns, detecting anomalies in climate data.
* Manufacturing: Predicting equipment failures, optimizing production processes.

Q: What are the future directions for research on MOMENT and time series foundation models?
A: Future research directions include:
* Exploring new pre-training tasks and datasets.
* Developing more robust evaluation methods for time series models.
* Investigating the interpretability and explainability of time series foundation models.
* Applying MOMENT to real-world problems and evaluating its impact.

Q: How can I contribute to the development of MOMENT?
A: MOMENT is an open-source project, and contributions are welcome! You can contribute by:
* Using MOMENT in your own research or applications.
* Providing feedback on the model's performance and usability.
* Contributing to the codebase or documentation.
* Expanding the Time Series Pile by adding new datasets.

MOMENT: Open-Source Foundation Models for Time Series Analysis

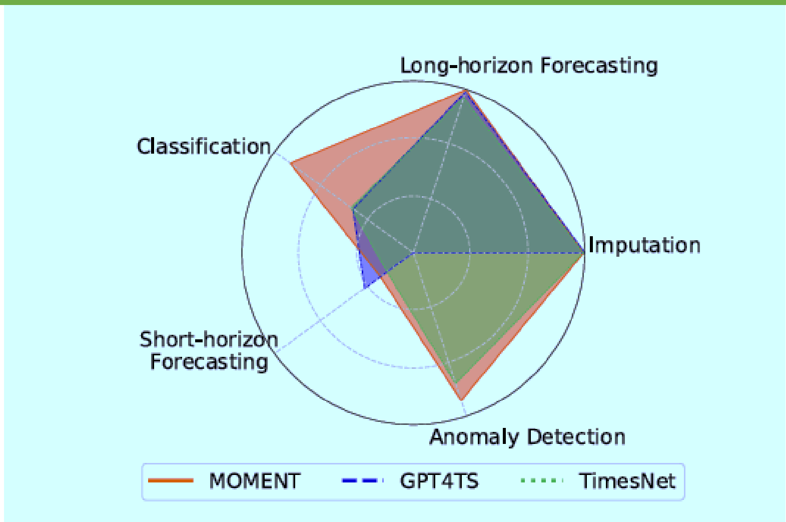


Figure provides a visual summary of the capabilities of their pre-trained time series model.
* **Multiple Time Series Tasks:** The figure showcases MOMENT's performance on various tasks like forecasting, classification, anomaly detection, and imputation.
* **Strong Performance:** The visual demonstrate that MOMENT excels in these tasks, potentially through comparisons with baseline models or visualizations of accurate predictions.

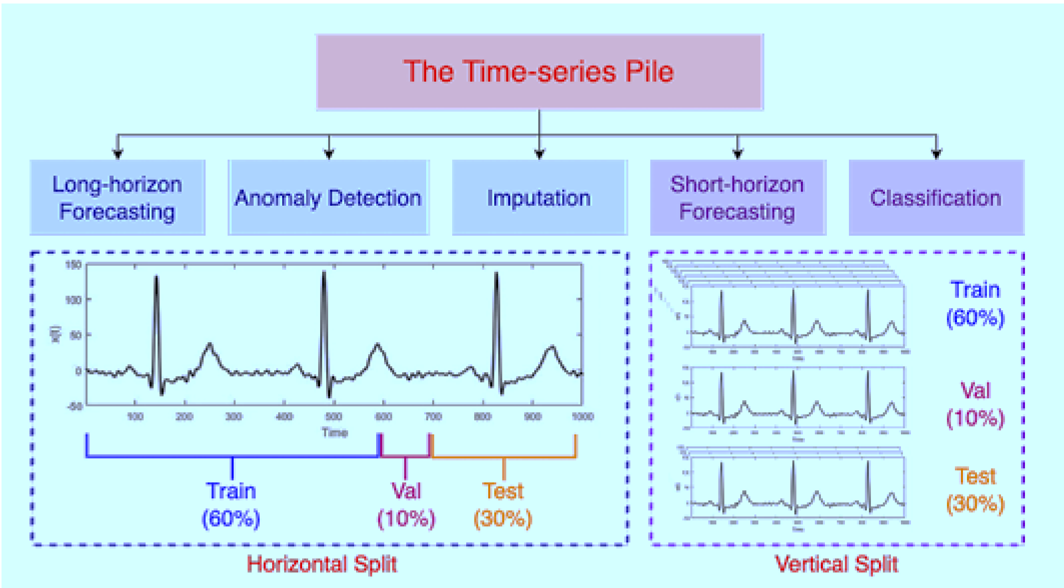
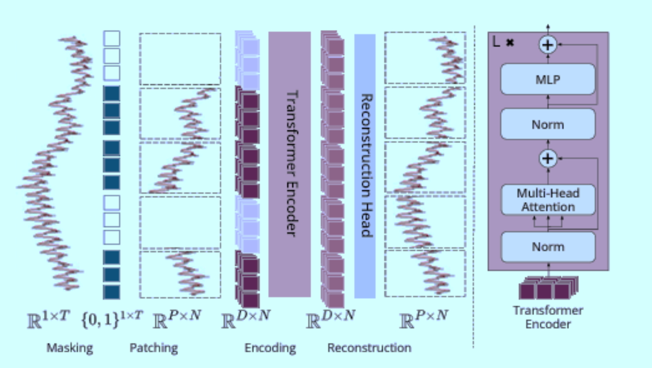


Figure illustrates the "Time Series Pile data splits". Here's a breakdown of what the figure shows and why:
* **Data Splits:** the importance of avoiding data contamination by carefully partitioning datasets into distinct train, validation, and test splits.
* **Visual Representation:** The partitioning for the "Time Series Pile," the large collection of time series data used for training and evaluating MOMENT.
* **Splitting Methodology:** how splits are determined, either using predefined splits provided with datasets or randomly sampling if no predefined splits exist (60% train, 10% validation, 30% test).
* **Training Data:** "We only use the training splits of all datasets for pre-training," which portions of the data are used for pre-training.

MOMENT: Open-Source Foundation Models for Time Series Analysis



Overview of MOMENT

*** Patching:** The figure illustrates how MOMENT handles time series data by breaking it into fixed-length "patches." This reduces computational complexity compared to processing the entire series at once.

*** Masking:** During pre-training, some patches are randomly masked using a special "[MASK]" embedding. This forces the model to learn to predict the missing information.

*** Transformer Encoder:** The masked patches are fed into a transformer encoder, which learns to represent the time series data.

*** Reconstruction Head:** A lightweight prediction head attempts to reconstruct the original time series, including the masked patches, from the encoder's output.

*** Pre-training Objective:** The model is trained to minimize the difference (Mean Squared Error) between the reconstructed time series and the original one.

MOMENT: A Family of Open Time-series Foundation Models

Tasks	Supervision	Datasets	Metrics	Baselines	Experimental Setting
Long-horizon Forecasting	Linear Probing	ETT-h1/h2/m1/m2, Electricity, Traffic, Weather, Exchange, ILI	MSE, MAE	Time-LLM, GPT4TS, TimesNet, PatchTST, FEDFormer, DLinear, N-BEATS, Stationary, LightTS	Look-back window $L = 512$, Forecast horizon $H = \{24, 60\}$ (ILI), $\{96, 720\}$ (rest)
Short-horizon Forecasting	Zero-shot	M3 and M4 competition datasets (subset)	sMAPE[3].	GPT4TS, TimesNet, N-BEATS, AutoARIMA, AutoTheta, AutoETS, Seasonal Naive, Naive, Random Walk	Statistical methods fit on individual time series. Deep learning methods are trained on a source dataset & evaluated on a target dataset of the same temporal resolution.
Classification	Unsupervised representation learning	UCR Classification Archive (subset)	Accuracy	GPT4TS, TimesNet, TS2Vec, T-Loss, TNC, TS-TCC, TST, CNN, Encoder, FCN, MCNN, MLP, ResNet, t-LeNet, TWIESN DTW	All models except MOMENT were trained on each individual dataset. Quality of unsupervised representations measured using the accuracy of a SVM trained on them.
Anomaly Detection	Linear probing, Zero-shot	UCR Anomaly Archive (subset)	Adjusted Best F1 VUS-ROC	GPT4TS, TimesNet, Anomaly Transformer, DGHL, Anomaly Nearest Neighbor	Reconstruction-based anomaly detection with window size $= 512$ MSE between observed and predicted time series is used as the anomaly criterion
Imputation	Linear probing, Zero-shot	ETT-h1/h2/m1/m2, Electricity, Weather	MSE, MAE	GPT4TS, TimesNet, Linear, Naive, Cubic Spline, Nearest Neighbors	Randomly mask contiguous sub-sequences of length 8 Masking ratios: $\{12.5\%, 25\%, 37.5\%, 50\%\}$

Experimental Benchmark

Table outlines the comprehensive benchmark used to evaluate MOMENT's performance across various time series analysis tasks. Here are the key takeaways:

* Five Tasks: MOMENT is tested on five common tasks:

- *** Long-horizon forecasting**
- *** Short-horizon forecasting**
- *** Classification**
- *** Anomaly detection**
- *** Imputation**

*** Limited Supervision: The** emphasis is on evaluating MOMENT in scenarios with limited data and minimal task-specific tuning, highlighting its ability to generalize.

*** Datasets:** Specific datasets used for each task are listed, drawn from established benchmarks like the Informer long-horizon forecasting datasets and the UCR/UEA classification archive.

*** Metrics:** Appropriate evaluation metrics for each task are specified, such as MSE and MAE for forecasting, accuracy for classification, and adjusted best F1 for anomaly detection.

*** Baselines:** MOMENT is compared against various state-of-the-art deep learning and statistical baselines for each task.

*** Experimental Setting:** Key details about the experimental setup are provided, including the look-back window for forecasting, the masking ratio for imputation, and the specific training approach for each baseline.

MOMENT: Open-Source Foundation Models for Time Series Analysis

MOMENT: A Family of Open Time-series Foundation Models

Methods		MOMENT$_{LP}$		Time-LLM		GPT4TS		PatchTST		DLinear		TimesNet		FEDFormer		Stationary		LightTS		N-BEATS	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.154	0.209	-	-	0.162	0.212	0.149	0.198	0.176	0.237	0.172	0.220	0.217	0.296	0.173	0.223	0.182	0.242	0.152	0.210
	192	0.197	0.248	-	-	0.204	0.248	0.194	0.241	0.220	0.282	0.219	0.261	0.276	0.336	0.245	0.285	0.227	0.287	0.199	0.260
	336	0.246	0.285	-	-	0.254	0.286	0.245	0.282	0.265	0.319	0.280	0.306	0.339	0.380	0.321	0.338	0.282	0.334	0.258	0.311
	720	0.315	0.336	-	-	0.326	0.337	0.314	0.334	0.333	0.362	0.365	0.359	0.403	0.428	0.414	0.410	0.352	0.386	0.331	0.359
ETTh1	96	0.387	0.410	0.408	0.429	0.376	0.397	0.370	0.399	0.375	0.399	0.384	0.402	0.376	0.419	0.513	0.491	0.424	0.432	0.399	0.428
	192	0.410	0.426	-	-	0.416	0.418	0.413	0.421	0.405	0.416	0.436	0.429	0.420	0.448	0.534	0.504	0.475	0.462	0.451	0.464
	336	0.422	0.437	-	-	0.442	0.433	0.422	0.436	0.439	0.443	0.491	0.469	0.459	0.465	0.588	0.535	0.518	0.488	0.498	0.500
	720	0.454	0.472	0.523	0.514	0.477	0.456	0.447	0.466	0.472	0.490	0.521	0.500	0.506	0.507	0.643	0.616	0.547	0.533	0.608	0.573
ETTh2	96	0.288	0.345	0.285	0.348	0.285	0.342	0.274	0.336	0.289	0.353	0.340	0.374	0.358	0.397	0.476	0.458	0.397	0.437	0.327	0.387
	192	0.349	0.386	-	-	0.354	0.389	0.339	0.379	0.383	0.418	0.402	0.414	0.429	0.439	0.512	0.493	0.520	0.504	0.400	0.435
	336	0.369	0.408	-	-	0.373	0.407	0.329	0.380	0.448	0.465	0.452	0.452	0.496	0.487	0.552	0.551	0.626	0.559	0.747	0.599
	720	0.403	0.439	0.399	0.435	0.406	0.441	0.379	0.422	0.605	0.551	0.462	0.468	0.463	0.474	0.562	0.560	0.863	0.672	1.454	0.847
ETTm1	96	0.293	0.349	0.384	0.403	0.292	0.346	0.290	0.342	0.299	0.343	0.338	0.375	0.379	0.419	0.386	0.398	0.374	0.400	0.318	0.367
	192	0.326	0.368	-	-	0.332	0.372	0.332	0.369	0.335	0.365	0.374	0.387	0.426	0.441	0.459	0.444	0.400	0.407	0.355	0.391
	336	0.352	0.384	-	-	0.366	0.394	0.366	0.392	0.369	0.386	0.410	0.411	0.445	0.459	0.495	0.464	0.438	0.438	0.401	0.419
	720	0.405	0.416	0.437	0.429	0.417	0.421	0.416	0.420	0.425	0.421	0.478	0.450	0.543	0.490	0.585	0.516	0.527	0.502	0.448	0.448
ETTm2	96	0.170	0.260	0.181	0.269	0.173	0.262	0.165	0.255	0.167	0.269	0.187	0.267	0.203	0.287	0.192	0.274	0.209	0.308	0.197	0.271
	192	0.227	0.297	-	-	0.229	0.301	0.220	0.292	0.224	0.303	0.249	0.309	0.269	0.328	0.280	0.339	0.311	0.382	0.285	0.328
	336	0.275	0.328	-	-	0.286	0.341	0.274	0.329	0.281	0.342	0.321	0.351	0.325	0.366	0.334	0.361	0.442	0.466	0.338	0.366
	720	0.363	0.387	0.366	0.388	0.378	0.401	0.362	0.385	0.397	0.421	0.408	0.403	0.421	0.415	0.417	0.413	0.675	0.587	0.395	0.419
ILI	24	2.728	1.114	3.025	1.195	2.063	0.881	1.319	0.754	2.215	1.081	2.317	0.934	3.228	1.260	2.294	0.945	8.313	2.144	4.539	1.528
	36	2.669	1.092	-	-	1.868	0.892	1.430	0.834	1.963	0.963	1.972	0.920	2.679	1.080	1.825	0.848	6.631	1.902	4.628	1.534
	48	2.728	1.098	-	-	1.790	0.884	1.553	0.815	2.130	1.024	2.238	0.940	2.622	1.078	2.010	0.900	7.299	1.982	4.957	1.585
	60	2.883	1.126	3.245	1.221	1.979	0.957	1.470	0.788	2.368	1.096	2.027	0.928	2.857	1.157	2.178	0.963	7.283	1.985	5.429	1.661
ECL	96	0.136	0.233	-	-	0.139	0.238	0.129	0.222	0.140	0.237	0.168	0.272	0.193	0.308	0.169	0.273	0.207	0.307	0.131	0.228
	192	0.152	0.247	-	-	0.153	0.251	0.157	0.240	0.153	0.249	0.184	0.289	0.201	0.315	0.182	0.286	0.213	0.316	0.153	0.248
	336	0.167	0.264	-	-	0.169	0.266	0.163	0.259	0.169	0.267	0.198	0.300	0.214	0.329	0.200	0.304	0.230	0.333	0.170	0.267
	720	0.205	0.295	-	-	0.206	0.297	0.197	0.290	0.203	0.301	0.220	0.320	0.246	0.355	0.222	0.321	0.265	0.360	0.208	0.298
Traffic	96	0.391	0.282	-	-	0.388	0.282	0.360	0.249	0.410	0.282	0.593	0.321	0.587	0.366	0.612	0.338	0.615	0.391	0.375	0.259
	192	0.404	0.287	-	-	0.407	0.290	0.379	0.256	0.423	0.287	0.617	0.336	0.604	0.373	0.613	0.340	0.601	0.382	0.403	0.274
	336	0.414	0.292	-	-	0.412	0.294	0.392	0.264	0.436	0.296	0.629	0.336	0.621	0.383	0.618	0.328	0.613	0.386	0.426	0.285
	720	0.450	0.310	-	-	0.450	0.312	0.432	0.286	0.466	0.315	0.640	0.350	0.626	0.382	0.653	0.355	0.658	0.407	0.508	0.335

Long-Term Forecasting Performance

* **Focus:** This table presents the results of long-term forecasting experiments, measured by Mean Squared Error (MSE) and Mean Absolute Error (MAE).
* **Datasets:** Results are shown for various datasets, including Weather, ETT variations (hourly and minutely), ILI (Influenza-like Illness), ECL (Exchange rate), and Traffic.
* **Methods:** MOMENT (using linear probing, denoted as MOMENTLP) is compared against several baselines, including Time-LLM, GPT4TS, PatchTST, DLinear, TimesNet, FEDFormer, Stationary, LightTS, and N-BEATS.
* **Horizons:** Forecasting is evaluated at different horizons (96, 192, 336, and 720 time steps into the future), representing long-term predictions.
* **Observations:** The table highlights that PatchTST generally performs the best across most settings. MOMENTLP demonstrates competitive performance, closely trailing PatchTST.

Datasets		MOMENT$_{LP}$ M4	MOMENT$_{LP}$ FR	GPT4TS M4	GPT4TS FR	TimesNet M4	TimesNet FR	N-BEATS M4	N-BEATS FR	ARIMA	Theta	ETS	Seasonal Naive	Naive	Random Walk
M3	Yearly	**16.74**	**16.97**	18.39	**17.40**	27.48	**16.21**	**16.82**	**15.92**	17.90	16.70	16.47	17.54	17.54	16.77
	Quarterly	**10.09**	10.62	**10.18**	10.29	14.41	12.68	11.26	11.30	10.18	9.19	8.99	11.02	11.45	11.72
	Monthly	16.04	16.90	**15.21**	16.37	**15.58**	16.23	**15.63**	16.37	15.95	14.96	14.41	17.74	18.53	19.19
M4	Yearly	-	**14.84**	-	**14.80**	-	**14.40**	-	14.18	16.19	14.04	14.06	16.33	16.33	14.22
	Quarterly	-	12.02	-	11.77	-	13.21	-	12.25	10.86	10.21	10.24	12.55	11.65	11.46
	Monthly	-	15.80	-	15.36	-	15.67	-	15.24	13.68	13.19	13.58	16.00	15.24	15.48

Zero-Shot Short-Horizon Forecasting Performance

* **Focus:** This table presents the results of zero-shot short-horizon forecasting experiments, measured using symmetric Mean Absolute Percentage Error (sMAPE).
* **Datasets:** The experiments are conducted on subsets of the M3 and M4 competition datasets, focusing on yearly, quarterly, and monthly forecasting.
* **Methods:** MOMENT is compared against GPT4TS, TimesNet, N-BEATS, and several statistical methods (ARIMA, Theta, ETS, Seasonal Naive, Naive, Random Walk).
* **Observations:** Statistical methods generally outperform deep learning methods in this zero-shot setting. However, MOMENT, GPT4TS, and N-BEATS achieve lower sMAPE than ARIMA on some datasets (highlighted in bold).

MOMENT: Open-Source Foundation Models for Time Series Analysis

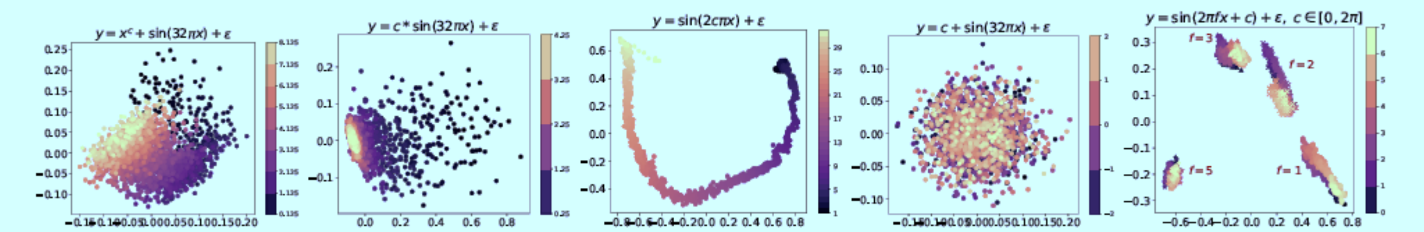
MOMENT: A Family of Open Time-series Foundation Models

	MOMENT$_0$	TimesNet	GPT4TS	TS2Vec	T-Loss	TNC	TS-TCC	TST
Mean	0.794	0.572	0.566	0.851	0.833	0.786	0.793	0.658
Median	0.815	0.565	0.583	0.871	0.849	0.788	0.802	0.720
Std.	0.147	0.238	0.234	0.134	0.136	0.168	0.176	0.220
Mean rank	7.225	13.324	13.318	3.494	5.261	6.937	6.500	11.846
Median rank	7.0	14.0	14.0	3.0	5.0	6.5	6.0	13.0
Wins/Losses	880.5/566.5	325.5/1121.5	326.0/1121.0	1220.0/227.0	1033.0/375.0	885.5/522.5	924.0/484.0	460.0/987.0

	CNN	Encoder	FCN	MCNN	MLP	ResNet	t-LeNet	TWIESN	DTW
Mean	0.751	0.743	0.809	0.702	0.750	0.825	0.348	0.726	0.764
Median	0.773	0.753	0.837	0.718	0.766	0.852	0.333	0.724	0.768
Std.	0.180	0.159	0.188	0.194	0.169	0.177	0.221	0.164	0.152
Mean rank	9.384	8.906	5.565	11.043	9.247	4.406	16.115	10.384	9.071
Median rank	9.0	9.0	4.0	12.0	10.0	3.0	17.0	11.0	9.0
Wins/Losses	684.0/763.0	727.5/719.5	1031.5/415.5	533.0/914.0	696.5/750.5	1137.0/310.0	71.5/1375.5	593.0/854.0	712.5/734.5

Classification Accuracy

* **Focus:** This table compares MOMENT's performance in time series classification against various other methods, including specialized time series classification models.
* **Key Observation:** MOMENT, without any fine-tuning on individual datasets, demonstrates surprisingly good accuracy. It outperforms many specialized methods, indicating its ability to learn generalizable time series representations.
* **Wins/Losses:** The "Wins/Losses" column emphasizes that MOMENT wins more comparisons than it loses against most other methods.
* **Comparison with TimesNet and GPT4TS:** The table highlights that MOMENT surpasses TimesNet and GPT4TS in classification, even though those models are trained on each individual dataset with labels.



(i) Trend *(ii)* Amplitude *(iii)* Frequency *(iv)* Baseline Shift *(v)* Phase

What is MOMENT Learning

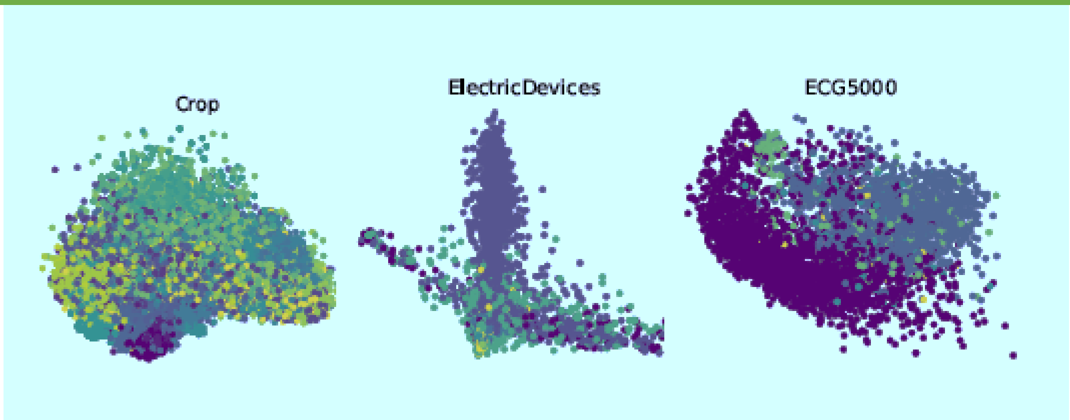
* **Focus:** This figure visually explores the information captured by MOMENT's learned representations. It uses synthetic sine waves with varying characteristics (trend, amplitude, frequency, phase, baseline shift).
* **PCA and t-SNE:** Principal Component Analysis (PCA) and t-SNE are used to visualize the embeddings of these sine waves in a 2-dimensional space.
* **Key Observations:**
    * MOMENT effectively captures changes in trend, amplitude, frequency, and phase.
    * It cannot differentiate between vertically shifted time series due to its normalization process.
* Implication: MOMENT learns representations sensitive to key time series properties, suggesting it's not merely memorizing data but extracting meaningful patterns.

Model	Bit Memory	MNIST	CIFAR-10	IMDb
GPT-2	**1.000**	0.975	**0.711**	0.867
Flan-T5	**1.000**	**0.987**	0.672	0.861
MOMENT	**1.000**	0.982	0.620	**0.872**

Cross-Modal Transfer Experiments

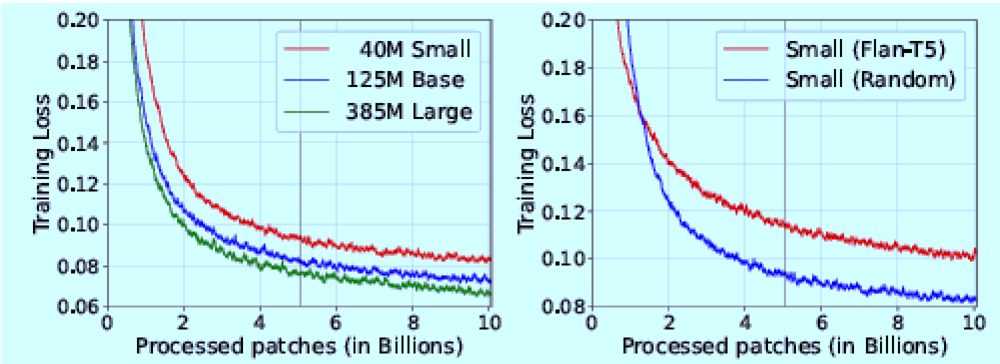
* **Focus:** This table explores MOMENT's ability to transfer its learned representations to other modalities (images, text, binary data) for sequence classification.
* **Frozen Layers:** MOMENT's self-attention and feed-forward layers are frozen, meaning only the final classification layer is trained.
* **Comparison with Language Models:** MOMENT achieves comparable accuracy to GPT-2 and Flan-T5, demonstrating its capacity for cross-modal learning.
* **Implication:** This suggests that MOMENT's pre-training on time series data enables it to learn general sequence modeling principles applicable to other modalities.

MOMENT: Open-Source Foundation Models for Time Series Analysis



PCA and t-SNE Visualizations

* **Focus:** This figure visually demonstrates MOMENT's ability to learn distinct representations for different classes in time series data, even without specific fine-tuning.
* **Datasets:** It uses PCA and t-SNE to visualize the representations learned on the three largest UCR datasets (Crop, Electric Devices, ECG5000).
* **Key Observation:** Different colors, representing different classes, form distinct clusters in the visualization. This indicates that MOMENT captures meaningful class-specific information during pre-training.
* **Implication:** This supports the claim that pre-training enables MOMENT to perform well in classification tasks with minimal task-specific fine-tuning.



Training Losses (MSE)

* **Focus:** This figure explores the impact of model size on training loss.
* **Left Panel:** It shows that larger models achieve lower training loss, consistent with trends observed in language models.
* **Right Panel:** It demonstrates that MOMENT with randomly initialized weights eventually outperforms the same model initialized with Flan-T5 weights.
* **Implication:** This suggests that the Time Series Pile provides sufficient data to train time series foundation models from scratch, without relying on pre-trained language models.

MOMENT: Open-Source Foundation Models for Time Series Analysis

MOMENT: A Family of Open Time-series Foundation Models

Dataset	$\texttt{MOMENT}_0$		$\texttt{MOMENT}_{LP}$		GPT4TS		TimesNet		Naive		Linear		Nearest		Cubic	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	0.082	0.130	0.035	0.075	**0.031**	**0.071**	0.036	0.098	0.119	0.108	0.065	0.067	0.083	0.078	0.601	0.153
ETTh1	0.402	0.403	**0.139**	**0.234**	0.227	0.254	0.175	0.264	1.185	0.658	0.775	0.534	0.900	0.579	2.178	0.916
ETTh2	0.125	0.238	**0.061**	**0.159**	0.109	0.213	0.170	0.286	0.225	0.304	0.135	0.234	0.166	0.252	1.920	0.641
ETTm1	0.202	0.288	**0.074**	**0.168**	0.076	0.146	0.087	0.198	0.455	0.365	0.165	0.229	0.230	0.260	0.858	0.494
ETTm2	0.078	0.184	**0.031**	**0.108**	0.052	0.133	0.112	0.220	0.113	0.191	0.062	0.138	0.079	0.152	0.534	0.356
Electricity	0.250	0.371	0.094	0.211	**0.072**	**0.183**	0.124	0.248	1.474	0.869	0.737	0.592	0.923	0.629	2.257	0.888

Imputation Results

*** Focus:** This table evaluates MOMENT's performance in time series imputation, where the goal is to fill in missing values.

*** Key Observations:**

* MOMENT with linear probing achieves the lowest reconstruction error on all ETT datasets.

* In the zero-shot setting, MOMENT outperforms most statistical interpolation methods, except for linear interpolation.

*** Implication:** MOMENT's pre-training allows it to effectively impute missing values, even with minimal task-specific fine-tuning.

Metric		$\texttt{MOMENT}_0$	$\texttt{MOMENT}_{LP}$	GPT4TS	TimesNet	Anomaly Transformer	DGHL	k-NN
Adj. F_1	Mean	0.585	**0.628**	0.424	0.537	0.492	0.425	0.554
	Median	0.683	**0.778**	0.331	0.541	0.432	0.331	0.595
	Std.	0.377	0.373	0.366	0.389	0.401	0.365	0.393
	Mean rank	3.410	**3.005**	4.862	3.642	4.326	5.071	3.681
	Median rank	**3.00**	**3.00**	5.00	3.50	4.00	5.25	3.75
	Wins/Losses	703.5/472.5	783.0/393.0	419.0/757.0	658.0/518.0	524.0/652.0	378.0/798.0	650.5/525.5
VUS ROC	Mean	0.670	0.684	0.611	0.679	0.661	0.646	**0.706**
	Median	0.677	0.692	0.615	0.692	0.658	0.635	**0.727**
	Std.	0.133	0.146	0.114	0.141	0.147	0.137	0.155
	Mean rank	4.056	3.382	5.193	3.897	3.913	4.403	**3.153**
	Median rank	4.00	**3.00**	6.00	4.00	4.00	4.50	3.00
	Wins/Losses	577.0/599.0	709.0/467.0	354.0/822.0	608.0/568.0	605.0/571.0	509.0/667.0	754.0/422.0

Anomaly Detection Performance

*** Focus:** This table assesses MOMENT's performance in anomaly detection, where the goal is to identify unusual patterns in time series data.

*** Metrics:** It uses Adjusted F1 score and VUS-ROC to evaluate performance.

*** Key Observation:** MOMENTLP (linear probing) achieves near state-of-the-art results in anomaly detection.

*** Comparison with k-NN:** While k-nearest neighbors performs slightly better in terms of VUS-ROC, MOMENTLP has a higher Adjusted F1 score, indicating a better balance between precision and recall.

Constructive comments and feedback are welcomed