Towards Lightweight Time Series Forecasting: a Patch-wise Transformer with Weak Data Enriching

Meng Wang[1], Jintao Yang[1], Bin Yang[2,*], Li Hui[3,4], Tongxin Gong[1], Bo Yang[1], Jiangtao Cui[3]

1. School of Computer Science, Xi'an Polytechnic University, China
2. School of Data Science and Engineering, East China Normal University, China / *Corresponding author
3. School of Computer Science and Technology, Xidian University, China
4. Shanghai Yunxi Technology, China

Xi'an Polytechnic University | East China Normal University | Xidian University

Introduction

Transformer-based time series forecasting faces challenges in balancing resource efficiency and contextual awareness, particularly for edge deployment where computational constraints limit complex architectures. Existing methods suffer from three critical limitations: excessive inference costs in traditional Transformers, neglect of future covariates (e.g., weather forecasts) crucial for abrupt change prediction, and inability to model multimodal weak labels or implicit temporal patterns (e.g., holidays) when explicit covariates are absent. To address these issues, we propose LiPFormer—a lightweight framework combining patch-wise attention to capture global-local patterns without positional encoding, and dual-encoder contrastive learning to unify explicit/implicit covariates through weak label enrichment. This design reduces computation by 30% while maintaining predictive accuracy, achieving efficient contextual integration for dynamic scenarios.

Challenges

➢ Heavyweight Transformers have high computational complexity, limiting deployment in resource-constrained scenarios.

➢ Efficiency-accuracy trade-off requires balancing predictive performance with computational simplicity.

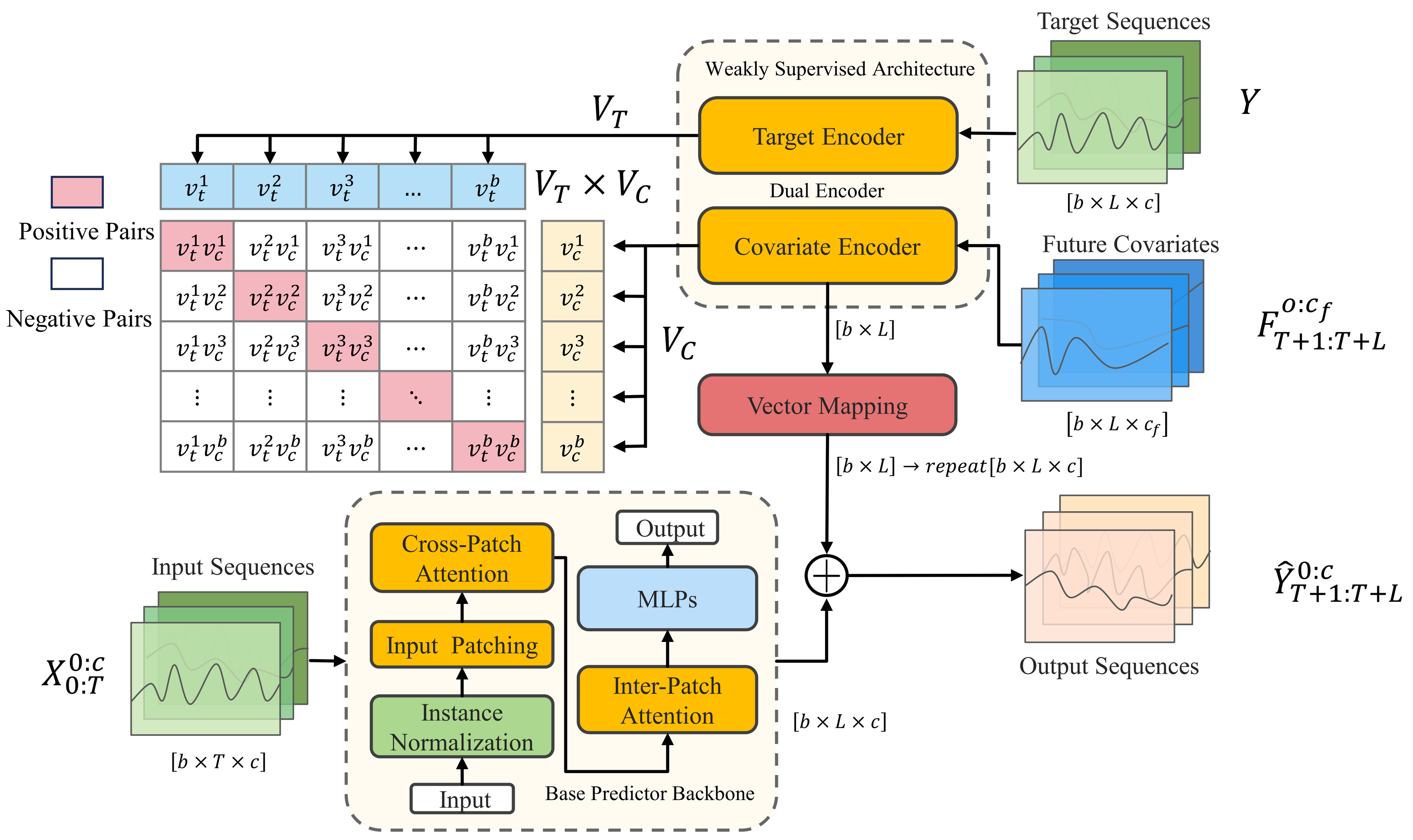
➢ Weak label integration struggles with heterogeneous multimodal data (text/numerical covariates).

Training Methodologies

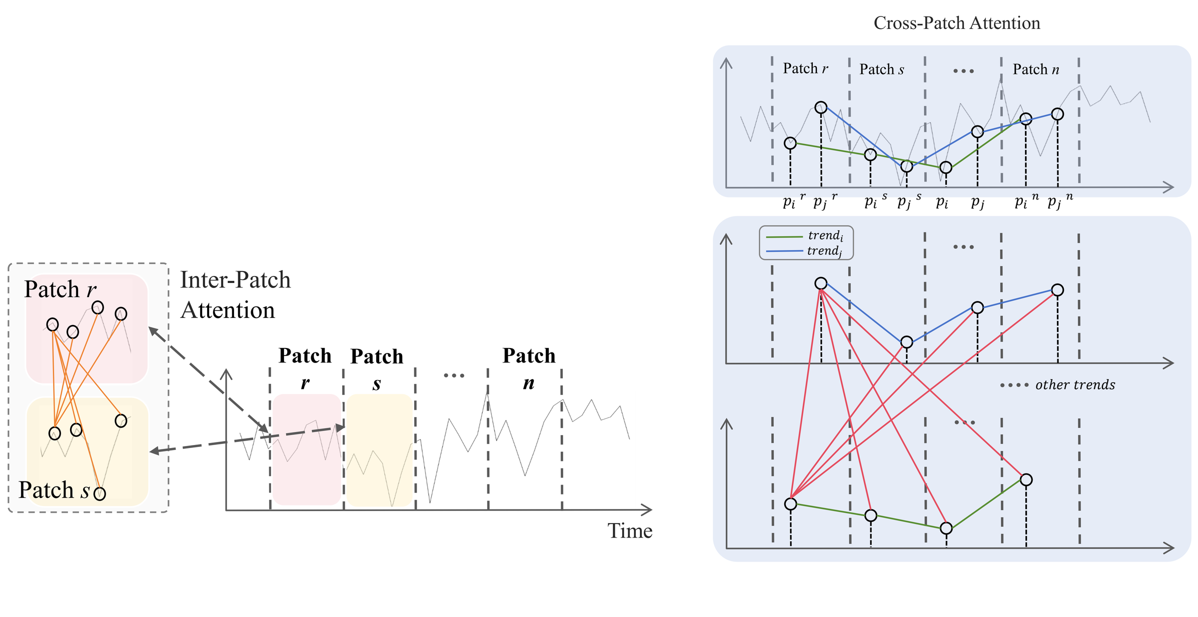
➢ Pre-training with Weak Label Enriching : A contrastive learning-based module uses dual encoders (Covariate Encoder and Target Encoder) to pre-train on "covariate-target" pairs, leveraging weak labels (e.g., temporal attributes, weather) as future covariates. It maximizes cosine similarity between positive pairs (ground truth sequences and covariates) while minimizing negative pairs via symmetric cross-entropy loss.

➢ Prediction-Oriented Training : The Base Predictor processes normalized, patched input sequences through a lightweight backbone. Final predictions are refined using future covariate dual-encoder guidance, optimized with Smooth L1 loss to balance accuracy and robustness during training.

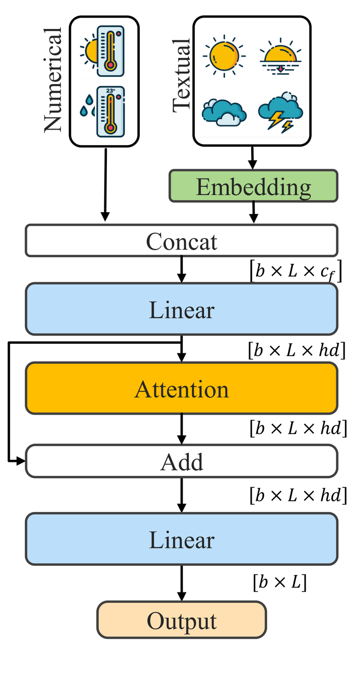
Model Components



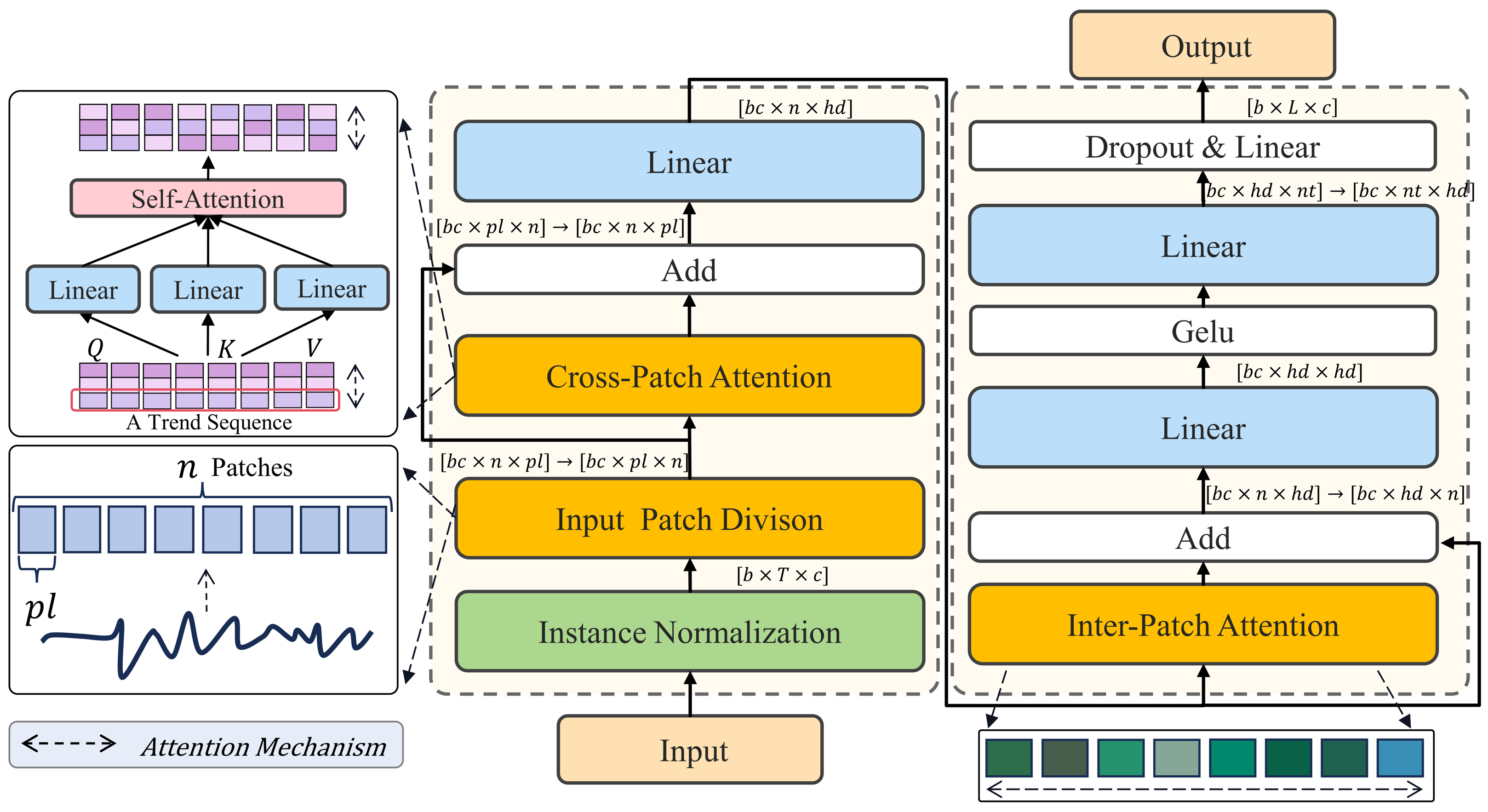
The architecture of LiPFormer.



The construction of Inter Patch and Cross Patch attention.



Covariate Encoder.



Base Predictor block.

Key Components

➢ **Base Predictor**

- **Lightweight Architecture:** Eliminates Layer Normalization (LN), Feed-Forward Networks (FFNs), and Positional Encoding (PE) to simplify the Transformer backbone. Replaces traditional Transformer with linear transformation-based attention for reduced complexity.
- **Patch-wise Attention:** Divides input sequences into patches and introduces two novel patch-wise mechanisms—Inter-Patch Attention (for local coherence preservation) and Cross-Patch Attention (for global temporal dependency capture). This addresses fixed-patch-size limitations by dynamically modeling multi-scale periodic patterns.

➢ **Weakly Supervised Dual Encoder Framework**

- **Explicit Covariate Handling:** For datasets with available future covariates (e.g., weather, time), a co-trained Covariate Encoder maps numerical/textual features into a semantic space aligned with target sequences via Res-attention and linear layers.
- **Implicit Covariate Augmentation:** For scenarios lacking explicit covariates, temporal attributes (e.g., holidays, rush hours) are encoded and embedded into a shared latent space with target sequences. A contrastive learning framework (dual encoder: target sequences vs. covariates) maximizes their correlation, enabling implicit supervision.
- **Seamless Integration:** The weak label enriching module is transplanted into existing forecasting frameworks to enhance predictive capacity while maintaining lightweight design.

Experments

➢ SOTA was achieved on Nine real-world datasets:

- Multivariate long-term time series forecasting results with LiPFormer.

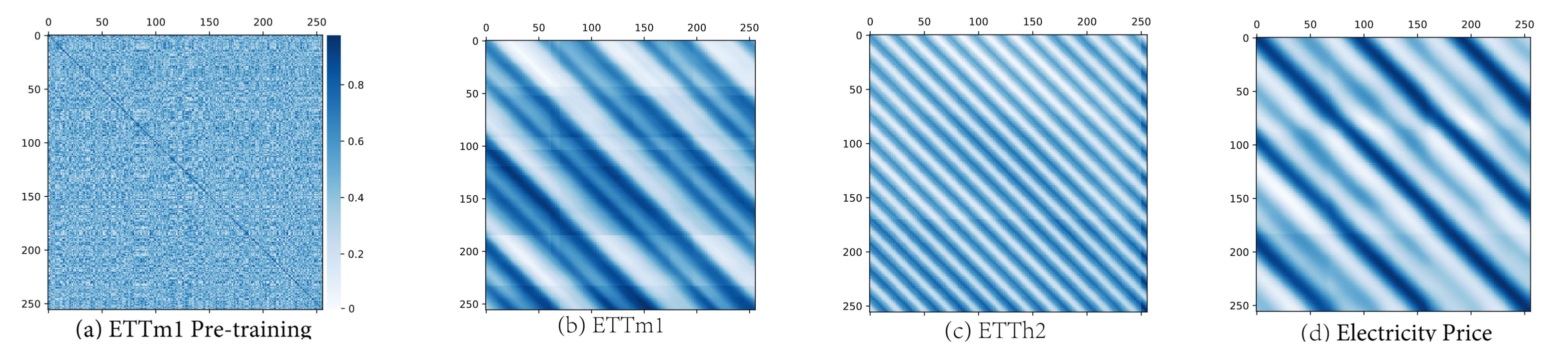
Models		LiPFormer			iTransformer			TimeMixer			FGNN			PatchTST			DLinear			TiDE		
Metric		MSE	MAE	Efficiency	MSE	MAE	Efficiency	MSE	MAE	Efficiency	MSE	MAE	Efficiency	MSE	MAE	Efficiency	MSE	MAE	Efficiency	MSE	MAE	Efficiency
ETTh1	96	0.359	0.379	0.75s	0.392	0.423	1.17s	0.385	0.417	2.87s	0.647	0.561	0.58s	0.375	0.399	2.11s	0.375	0.399	0.28s	0.375	0.398	7.84s
	192	0.404	0.405	0.14s	0.437	0.455	0.19s	0.424	0.443	0.58s	0.688	0.583	0.1s	0.414	0.421	0.32s	0.405	0.436	0.05s	0.412	0.422	0.96s
	336	0.444	0.424	0.27G	0.456	0.469	18.02G	0.456	0.459	1.42T	0.704	0.601	1.05G	0.431	0.436	17.09G	0.439	0.443	4.15M	0.435	0.433	1.13G
	720	0,450	0.453	66K	0.553	0.537	6.4M	0.601	0.559	4.27M	0.772	0.654	592K	0.449	0.466	6.90M	0.472	0.490	18.62K	0.454	0.465	2.53M
ETTh2	96	0.265	0.327	0.80s	0.303	0.364	1.04s	0.296	0.354	2.42s	0.479	0.496	0.6s	0.274	0.336	2.07s	0.289	0.353	0.29s	0.27	0.336	7.83s
	192	0.335	0.374	0.14s	0.409	0.422	0.19s	0.384	0.415	0.59s	0.568	0.540	0.15s	0.339	0.379	0.33s	0.383	0.418	0.04s	0.332	0.38	1.27s
	336	0.364	0.395	0.27G	0.440	0.450	18.02G	0.383	0.423	1.42T	0.692	0.604	1.05G	0.331	0.380	17.09G	0.480	0.465	4.15M	0.36	0.407	1.13G
	720	0.392	0.425	66K	0.439	0.468	6.4M	0.399	0.453	4.27M	1.107	0.774	592K	0.379	0.422	6.90M	0.605	0.551	18.62K	0.419	0.451	2.53M
ETTm1	96	0.296	0.338	3.1s	0.318	0.366	3.56s	0.305	0.358	10.62s	0.403	0.427	4.49s	0.290	0.342	8.55s	0.299	0.343	1.12s	0.306	0.349	32.51s
	192	0.336	0.360	0.55	0.347	0.387	0.71s	0.343	0.379	2.15s	0.426	0.440	0.46s	0.332	0.369	1.54s	0.335	0.365	0.20s	0.335	0.366	4.64s
	336	0.365	0.379	0.27G	0.380	0.405	18.02G	0.371	0.394	1.42T	0.450	0.454	1.05G	0.366	0.453	17.09G	0.369	0.386	4.15M	0.364	0.384	1.13G
	720	0.408	0.413	66K	0.436	0.439	6.4M	0.427	0.423	4.27M	0.498	0.481	592K	0.420	0.533	6.90M	0.425	0.421	18.62K	0.413	0.413	2.53M
ETTm2	96	0.160	0.244	3.24s	0.180	0.273	2.89s	0.181	0.270	9.45s	0.225	0.322	2.74s	0.165	0.255	7.68s	0.167	0.260	1.21s	0.161	0.251	32.01s
	192	0.217	0.285	0.55	0.243	0.315	0.74s	0.239	0.313	2.07s	0.296	0.368	0.56s	0.220	0.292	2.05s	0.224	0.303	0.21s	0.215	0.289	4.21s
	336	0.273	0.322	0.27G	0.299	0.352	18.02G	0.289	0.340	1.42T	0.423	0.460	1.05G	0.278	0.329	17.09G	0.281	0.342	4.15M	0.267	0.326	1.13G
	720	0.348	0.372	66K	0.382	0.405	6.4M	0.452	0.455	4.27M	0.497	0.493	592K	0.367	0.385	6.90M	0.397	0.421	18.62K	0.352	0.383	2.53M
Electricity	96	0.131	0.224	5.12s	0.147	0.249	4.24s	0.134	0.231	-	0.211	0.319	3.63s	0.129	0.222	16.19s	0.140	0.237	2.61s	0.132	0.229	1452.56s
	192	0.147	0.238	1.01s	0.169	0.271	0.97s	0.339	0.414	-	0.226	0.331	0.78s	0.147	0.24	2.20s	0.153	0.249	0.46s	0.147	0.243	255.30s
	336	0.163	0.254	2.94G	0.190	0.292	66.56G	0.280	0.370	2.03T	0.242	0.345	1.51G	0.163	0.259	195.96G	0.169	0.267	47.57M	0.161	0.261	250.86G
	720	0.199	0.285	66K	0.236	0.329	6.4M	0.557	0.933	4.27M	0.274	0.370	592K	0.197	0.29	6.90M	0.203	0.301	18.62K	0.196	0.294	34.10M
Traffic	96	0.382	0.243	3.82s	0.421	0.318	3.14s	0.385	0.276	-	0.721	0.478	-	0.367	0.251	9.84s	0.410	0.282	1.58s	0.336	0.253	2652.14s
	192	0.397	0.255	1.02s	0.455	0.340	0.63s	0.394	0.281		0.777	0.494	-	0.385	0.259	1.69s	0.423	0.287	0.29s	0.346	0.257	399.65s
	336	0.411	0.260	7.91G	0.487	0.359	177G	0.413	0.290		0.813	0.503	4.06G	0.398	0.265	526.22G	0.436	0.296	127.76M	0.355	0.26	1.77T
	720	0.451	0.281	66K	0.555	0.394	6.4M	0.452	0.312		0.856	0.517	592K	0.434	0.287	6.90M	0.466	0.315	18.62K	0.386	0.273	88.49M
Weather	96	0.146	0.186	2.89s	0.159	0.211	2.26s	0.151	0.200	7.17s	0.166	0.236	1.60s	0.152	0.199	4.01s	0.176	0.237	0.49s	0.166	0.222	35.62s
	192	0.189	0.230	0.52s	0.202	0.251	0.50s	0.193	0.242	1.57s	0.208	0.274	0.32s	0.197	0.243	0.67s	0.22	0.282	0.11s	0.209	0.263	7.79s
	336	0.244	0.277	0.78G	0.256	0.291	5.12G	0.242	0.281	532.70G	0.255	0.311	0.39G	0.249	0.283	51.27G	0.265	0.319	12.45M	0.254	0.301	6.17G
	720	0.313	0.326	66K	0.323	0.342	6.4M	0.319	0.334	4.27M	0.314	2.000	592K	0.320	0.335	6.90M	0.323	0.362	18.62K	0.313	0.34	3.93M
Electricity Price	96	0.486	0.424	2.54s	0.677	0.549	1.63s	0.621	0.531	4.52s	0.703	0.552	1.31s	0.635	0.537	2.74s	0.572	0.480	0.38s	0.585	0.480	3.28s
	192	0.528	0.443	0.50s	0.749	0.592	0.31s	0.720	0.586	0.96s	0.729	0.563	0.21s	0.697	0.591	0.50s	0.720	0.447	0.11s	0.618	0.520	0.65s
	336	0.459	0.446	8.31G	0.747	0.591	1.22G	0.704	0.575	12.68G	0.723	0.557	11.79M	0.773	0.615	4.88G	0.651	0.533	1.18M	0.643	0.542	0.22G
	720	0.495	0.467	74.86K	0.797	0.648	6.4M	0.675	0.566	4.27M	0.696	0.545	592K	0.832	0.629	6.90M	0.860	0.520	18.62K	0.632	0.538	2.02M
Cycle	96	0.136	0.221	1.09s	0.182	0.278	1.05s	0.169	0.265	2.79s	0.441	0.464	1.01s	0.160	0.248	1.60s	0.174	0.254	0.22s	0.150	0.236	1.70s
	192	0.145	0.230	0.19s	0.212	0.302	0.18s	0.203	0.303	0.58s	0.469	0.475	0.12s	0.167	0.254	0.29s	0.177	0.254	0.05s	0.158	0.240	0.21s
	336	0.152	0.235	0.84G	0.232	0.318	1.02G	0.146	0.240	6.34G	0.469	0.472	18M	0.179	0.263	2.44G	0.184	0.256	0.59M	0.167	0.246	0.10G
	720	0.159	0.236	74K	0.258	0.337	6.4M	0.175	0.263	4.27M	0.482	0.480	592K	0.214	0.284	6.90M	0.192	0.267	18.62K	0.181	0.256	1.92M
Count		51/12			0/0			0/1			0/0			11/23			0/11			15/29		

- LiPFormer outperforms state-of-the-art models in multivariate time series forecasting, achieving superior accuracy and robustness.

Datasets	Metric	ETTh1 96	ETTh1 192	ETTh1 336	ETTh1 720	ETTh2 96	ETTh2 192	ETTh2 336	ETTh2 720	ETTm1 96	ETTm1 192	ETTm1 336	ETTm1 720	ETTm2 96	ETTm2 192	ETTm2 336	ETTm2 720
Without Cross-Patch attn.	MSE	0.366	**0.401**	**0.441**	0.461	0.270	0.332	0.370	0.394	0.327	0.348	0.386	0.434	0.167	0.225	0.280	0.364
	MAE	0.380	0.424	0.450	0.461	0.329	0.371	0.396	0.427	0.365	0.378	0.398	0.425	0.255	0.295	0.333	0.386
Without Inter-Patch attn.	MSE	0.364	0.413	0.458	0.472	0.271	0.335	0.383	0.586	0.309	0.342	0.370	0.430	0.164	0.220	0.275	0.353
	MAE	0.380	0.408	0.434	0.455	0.329	0.371	0.405	0.477	0.350	0.367	0.384	0.417	0.251	0.289	0.324	0.382
Neither	MSE	0.379	0.417	0.456	0.466	0.273	0.379	0.397	0.422	0.314	0.340	0.377	0.423	0.168	0.224	0.276	0.356
	MAE	0.392	0.412	0.432	0.453	0.332	0.388	0.427	0.441	0.356	0.371	0.393	0.419	0.255	0.293	0.330	0.379
LiPFormer	MSE	**0.359**	0.404	0.444	**0.450**	**0.265**	**0.330**	**0.364**	**0.392**	**0.296**	**0.336**	**0.365**	**0.408**	**0.160**	**0.217**	**0.273**	**0.348**
	MAE	**0.379**	**0.405**	**0.425**	**0.453**	**0.327**	**0.369**	**0.395**	**0.425**	**0.338**	**0.360**	**0.379**	**0.413**	**0.244**	**0.285**	**0.322**	**0.372**

- Cross-Patch outperforms Inter-Patch alone, and the two attention mechanisms are complementary.

- Visualization of the logitsmatrices for Weakly Supervised Architecture.



(a) ETTm1 Pre-training (b) ETTm1 (c) ETTh2 (d) Electricity Price

- Visualization reveals aligned predictions and covariates via diagonal similarity optimization, periodic patterns, and multimodal interactions.