Channel-Aware Low-Rank Adaptation in Time Series Forecasting

Tong **Nie**, Yuewen **Mei**, Guoyang **Qin**, Jian **Sun**, Wei **Ma**

tong.nie@connect.polyu.hk wei.w.ma@polyu.edu.hk

ArXiv GitHub

KEYWORDS:

Long-Term Time Series Forecasting
Low-Rank Adaptation
Channel Independence Channel Mixing

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學
同濟大學 TONGJI UNIVERSITY

Tong Nie[1,2] Yuewen Mei[2] Guoyang Qin[2] Jian Sun[2] Wei Ma[1]

0 CHANNEL-BASED LONG-TERM TIME SERIES FORECASTING ARCHITECTURE

TL/DR; We propose a mixed channel strategy to balance both robustness and capacity based on low-rank adaptation, enhancing performances in long-term series forecasting.

REAL-WORLD ISSUES

- Severe distribution shift
- Channel heterogeneity

EXISTING SOLUTIONS

- Channel Independence
- Channel Dependence

1 " THE DILEMMA OF LTSF " Balance Between Capacity and Robustness in Architectural Design

PARADIGM 1 Channel Dependence

$$\theta^* = \arg\min_{\theta} \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{Y}^{(i)} - f_{\theta}(\mathbf{X}^{(i)})\|_F^2,$$

- Have larger model capacities
- Prone to overfit, less flexible

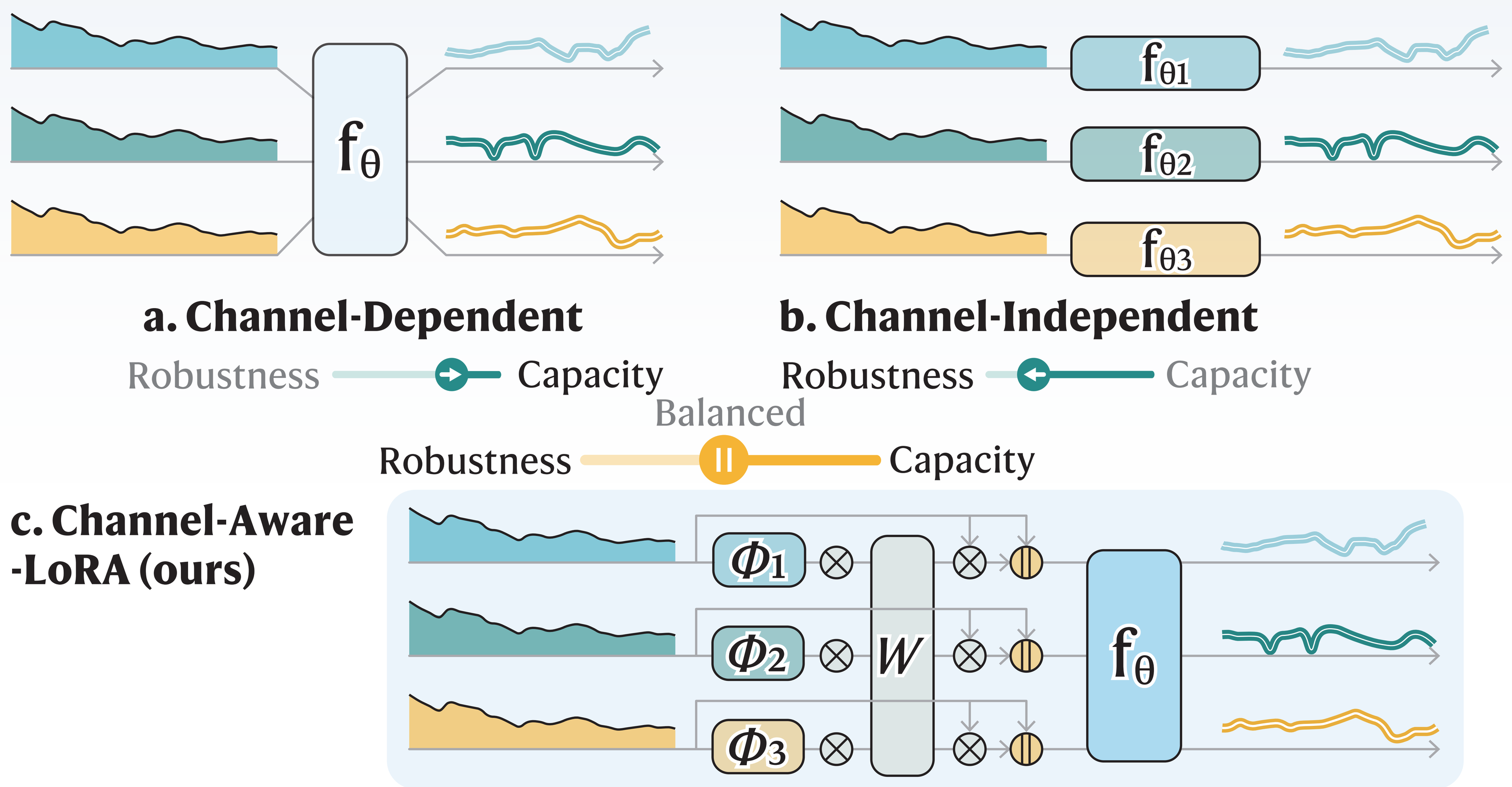
PARADIGM 2 Channel Independence

$$\theta_1^*, \dots, \theta_C^* = \arg\min_{\theta_1,\dots,\theta_C} \frac{1}{NC}\sum_{i=1}^{N}\sum_{c=1}^{C} \|\mathbf{Y}^{(i)}_{:,c} - f_{\theta_c}(\mathbf{X}^{(i)}_{:,c})\|_F^2.$$

- More robust for distribution shift
- Expensive hypothesis spaces

2 BREAKING THE TRADE-OFF Introducing C-LoRA, A Remedy Designed to Combine the Strengths of Both CI and CD in an Elegant Way, based on the Low-Rank Adaptation Method

PARADIGM 1x2
Channel-aware LoRA



How do we design a mixed channel strategy?

GENERAL FORECASTING BACKBONE

- Common components in LTSF architectures
- Popular Transformer- and MLP-based models
- Optional channel mixing operation

General Template for Forecasting Backbones

$$\bar{\mathbf{X}} = \mathtt{Normalization}(\mathbf{X}),$$
$$\mathbf{z}_c^{(0)} = \mathtt{TokenEmbedding}(\bar{\mathbf{X}}_{:,c}), \forall c = 1, \dots, C,$$
$$\mathbf{Z}^{(\ell+1)} = \mathtt{ChannelMixing}(\mathbf{Z}^{(\ell)}), \forall \ell = 0, \dots, L,$$
$$\hat{\mathbf{Y}} = \mathtt{Projection}(\mathbf{Z}^{(L+1)}),$$

PARAMETER-EFFICIENT LOW-RANK ADAPTER

- Low-rank adapter
- Model channel-wise adaptation

CONSIDERING CHANNEL IDENTITY INFORMATION

- Aware of the channel indentity
- Data-adaptive embedding

Channel-aware Low-rank Adaptation

1 $\mathbf{z}_c^{(0)} = \mathtt{MLP}\ (\bar{\mathbf{X}}_{:,c}; \theta_c), \forall c = 1, \dots, C,$

2 $\tilde{\phi}^{(c)} = \mathrm{ReLU}(\phi^{(c),\mathsf{T}}\mathbf{W}) \in \mathbb{R}^{D\times d},$

3 $\mathbf{z}_{c,\phi}^{(0)} = \mathbf{z}_c^{(0),\mathsf{T}}\tilde{\phi}^{(c)} \in \mathbb{R}^{d},$

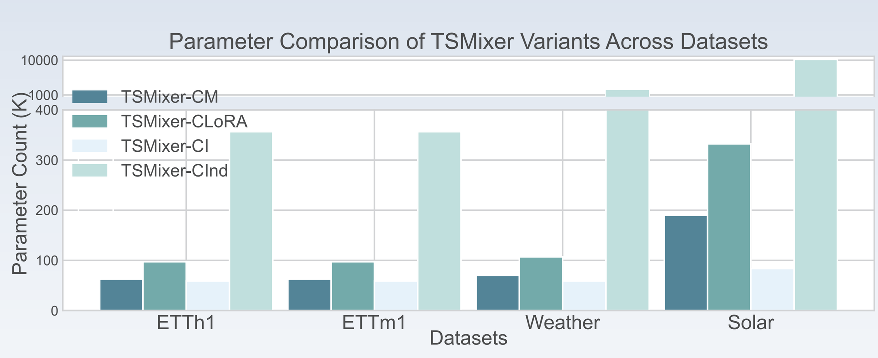
3 A PLUG-IN ENHANCER FOR A WIDE RANGE OF BACKBONES C-LoRA is Seamlessly Adaptable to Various SOTA Time Series Forecasting Backbones, Consistently Improving Their Performances

LTSF Benchmark Results

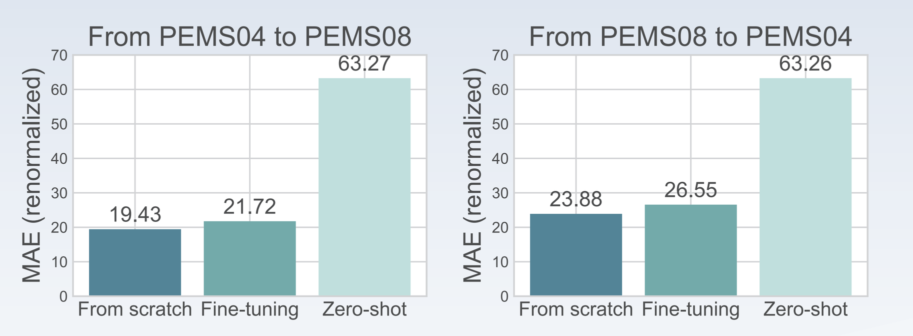
Models		iTransformer		w/ C-LoRA		TSMixer		w/ C-LoRA		RMLP		w/ C-LoRA		FreTS		w/ C-LoRA		FEDformer		w/ C-LoRA		Autoformer		w/ C-LoRA		Informer		w/ C-LoRA		IMP
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	%
ETTm1	96	0.345	0.376	0.331	0.367	0.332	0.370	0.317	0.356	0.337	0.374	0.321	0.360	0.340	0.376	0.330	0.369	0.379	0.419	0.375	0.422	0.505	0.475	0.512	0.480	0.672	0.571	0.577	0.542	3.23
	192	0.384	0.394	0.373	0.390	0.372	0.390	0.358	0.377	0.379	0.391	0.363	0.380	0.383	0.398	0.375	0.398	0.426	0.441	0.411	0.435	0.553	0.496	0.558	0.503	0.795	0.669	0.720	0.634	2.67
	336	0.418	0.415	0.409	0.414	0.405	0.411	0.389	0.400	0.412	0.412	0.395	0.403	0.420	0.425	0.406	0.418	0.445	0.459	0.423	0.447	0.621	0.537	0.618	0.523	1.212	0.871	0.982	0.756	4.51
	720	0.481	0.451	0.479	0.449	0.469	0.447	0.455	0.436	0.478	0.447	0.462	0.440	0.499	0.477	0.479	0.462	0.543	0.490	0.509	0.488	0.671	0.561	0.592	0.520	1.166	0.823	1.121	0.794	3.68
	Avg	0.407	0.409	0.398	0.405	0.395	0.405	0.380	0.392	0.402	0.406	0.385	0.396	0.411	0.419	0.398	0.412	0.448	0.452	0.430	0.448	0.588	0.517	0.570	0.507	0.961	0.734	0.850	0.682	3.59
Electricity	96	0.148	0.240	0.139	0.234	0.177	0.278	0.155	0.256	0.201	0.287	0.168	0.258	0.320	0.403	0.165	0.262	0.195	0.309	0.193	0.307	0.203	0.318	0.193	0.307	0.329	0.407	0.329	0.412	10.55
	192	0.162	0.253	0.160	0.254	0.193	0.293	0.172	0.270	0.209	0.297	0.179	0.268	0.325	0.406	0.176	0.270	0.202	0.315	0.203	0.316	0.225	0.334	0.222	0.330	0.338	0.419	0.347	0.430	8.53
	336	0.178	0.269	0.171	0.266	0.215	0.315	0.191	0.289	0.228	0.316	0.196	0.285	0.370	0.435	0.195	0.289	0.234	0.347	0.230	0.343	0.282	0.377	0.266	0.370	0.364	0.439	0.352	0.434	10.29
	720	0.225	0.317	0.195	0.289	0.260	0.352	0.230	0.321	0.273	0.350	0.238	0.319	0.416	0.474	0.235	0.325	0.261	0.365	0.262	0.365	0.314	0.383	0.299	0.389	0.397	0.460	0.395	0.456	10.24
	Avg	0.178	0.270	0.166	0.261	0.211	0.310	0.187	0.284	0.228	0.313	0.195	0.283	0.358	0.430	0.193	0.287	0.223	0.334	0.222	0.333	0.256	0.353	0.245	0.349	0.357	0.431	0.356	0.433	9.95
Weather	96	0.174	0.214	0.164	0.209	0.181	0.228	0.158	0.206	0.196	0.235	0.163	0.208	0.186	0.241	0.165	0.228	0.220	0.300	0.218	0.299	0.266	0.336	0.234	0.312	0.300	0.384	0.265	0.348	8.35
	192	0.221	0.254	0.209	0.251	0.227	0.263	0.207	0.249	0.240	0.271	0.209	0.249	0.222	0.273	0.210	0.274	0.278	0.344	0.282	0.350	0.307	0.367	0.282	0.344	0.598	0.544	0.381	0.427	8.27
	336	0.278	0.296	0.268	0.294	0.280	0.300	0.266	0.292	0.291	0.307	0.264	0.289	0.272	0.316	0.264	0.317	0.339	0.382	0.349	0.390	0.359	0.395	0.357	0.395	0.578	0.523	0.515	0.511	2.74
	720	0.358	0.349	0.349	0.346	0.353	0.347	0.348	0.345	0.363	0.353	0.342	0.340	0.350	0.381	0.343	0.370	0.409	0.438	0.411	0.420	0.419	0.428	0.415	0.424	1.059	0.741	0.792	0.651	4.47
	Avg	0.258	0.278	0.248	0.275	0.260	0.285	0.245	0.273	0.273	0.292	0.245	0.272	0.258	0.303	0.246	0.297	0.312	0.366	0.315	0.365	0.338	0.382	0.322	0.369	0.634	0.548	0.488	0.484	5.76
Solar	96	0.203	0.237	0.175	0.219	0.222	0.281	0.182	0.272	0.233	0.296	0.213	0.272	0.237	0.300	0.231	0.295	0.242	0.342	0.226	0.319	0.884	0.711	0.603	0.545	0.236	0.279	0.214	0.245	10.96
	192	0.233	0.261	0.211	0.259	0.261	0.301	0.207	0.275	0.260	0.316	0.234	0.292	0.265	0.321	0.261	0.318	0.285	0.380	0.245	0.366	0.834	0.692	0.682	0.563	0.227	0.287	0.241	0.290	7.64
	336	0.248	0.273	0.222	0.261	0.271	0.299	0.212	0.272	0.276	0.323	0.247	0.301	0.283	0.330	0.277	0.325	0.282	0.376	0.246	0.350	0.941	0.723	0.739	0.588	0.262	0.310	0.246	0.307	9.54
	720	0.249	0.275	0.203	0.263	0.267	0.293	0.201	0.262	0.273	0.316	0.244	0.291	0.286	0.326	0.281	0.322	0.357	0.427	0.304	0.410	0.882	0.717	0.801	0.642	0.329	0.355	0.279	0.329	10.05
	Avg	0.233	0.262	0.203	0.251	0.255	0.294	0.201	0.270	0.261	0.313	0.235	0.289	0.268	0.319	0.263	0.315	0.292	0.381	0.255	0.361	0.885	0.711	0.706	0.585	0.264	0.308	0.245	0.293	9.65
PEMS08	96	0.169	0.276	0.109	0.216	0.252	0.355	0.124	0.239	0.284	0.375	0.198	0.300	0.285	0.380	0.220	0.328	0.221	0.325	0.234	0.329	0.613	0.596	0.444	0.508	0.171	0.283	0.165	0.274	19.24
	192	0.188	0.288	0.137	0.239	0.322	0.385	0.151	0.256	0.336	0.409	0.222	0.310	0.335	0.409	0.280	0.371	0.332	0.412	0.340	0.420	1.111	0.829	0.616	0.596	0.202	0.304	0.202	0.298	20.35
	336	0.196	0.289	0.148	0.243	0.326	0.374	0.170	0.265	0.327	0.394	0.266	0.337	0.328	0.397	0.284	0.363	0.247	0.322	0.247	0.321	1.136	0.836	0.716	0.648	0.224	0.309	0.210	0.299	17.27
	720	0.235	0.320	0.186	0.277	0.388	0.422	0.225	0.308	0.372	0.430	0.303	0.372	0.373	0.431	0.327	0.396	0.564	0.556	0.546	0.546	1.409	0.948	0.901	0.733	0.247	0.324	0.238	0.316	16.12
	Avg	0.197	0.293	0.145	0.244	0.322	0.384	0.168	0.267	0.330	0.402	0.247	0.330	0.330	0.404	0.278	0.365	0.341	0.404	0.342	0.404	1.067	0.802	0.669	0.621	0.211	0.305	0.204	0.297	18.30

*Please refer to the paper for full results

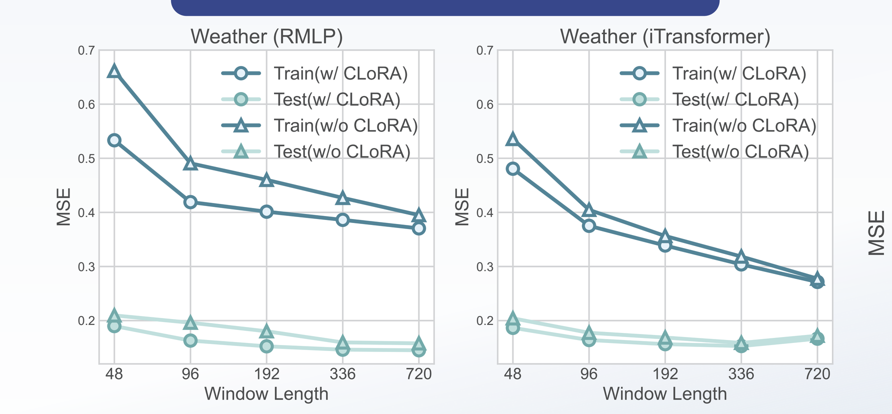
Parameter Efficiency



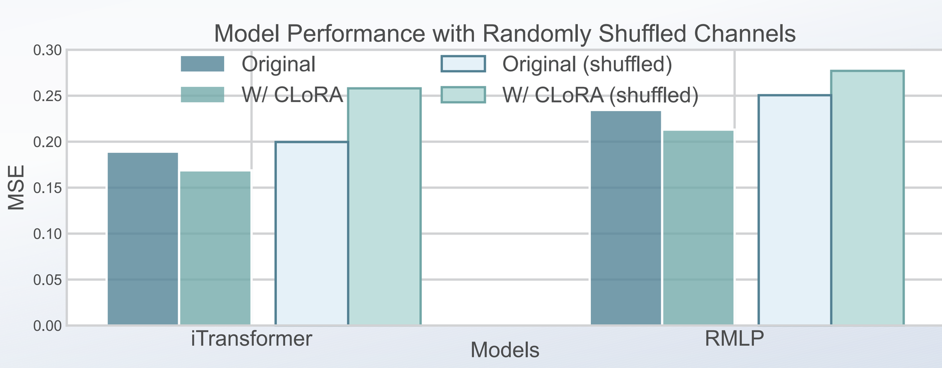
Fine-tune C-LoRA



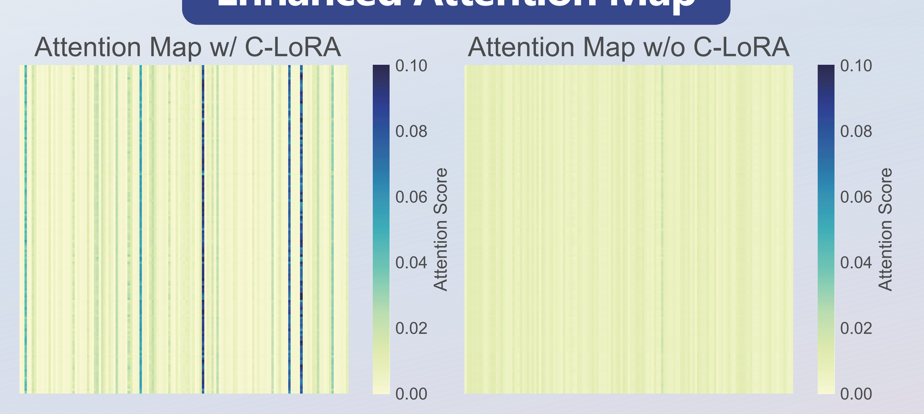
CI versus CD



Impact of Channel Identity



Enhanced Attention Map



OUR FINDINGS

- C-LoRA consistently improves the performances of various popular forecasting models by a large margin.
- It is also efficient, flexible to fine-tune, and beneficial for the exploitation of channel identity.

POTENTIAL DIRECTIONS

- Integrate it into pretrained LLMs
- More tasks such as time series imputation
- Scalability in large-scale spatiotemporal data

TOPS Mobility AI Lab

WE ARE INTERESTED IN:

- Spatiotemporal Data
- Urban Science
- Smart Transportation
- Large Language Models

SCAN TO VIEW MORE OF OUR WORK

CIKM 2024 BOISE OCTOBER 21-25