BRL Seminar

(2024. 11. 12. Tue.)

Channel Identification in Time Series

통합과정 9학기 이승한

Contents

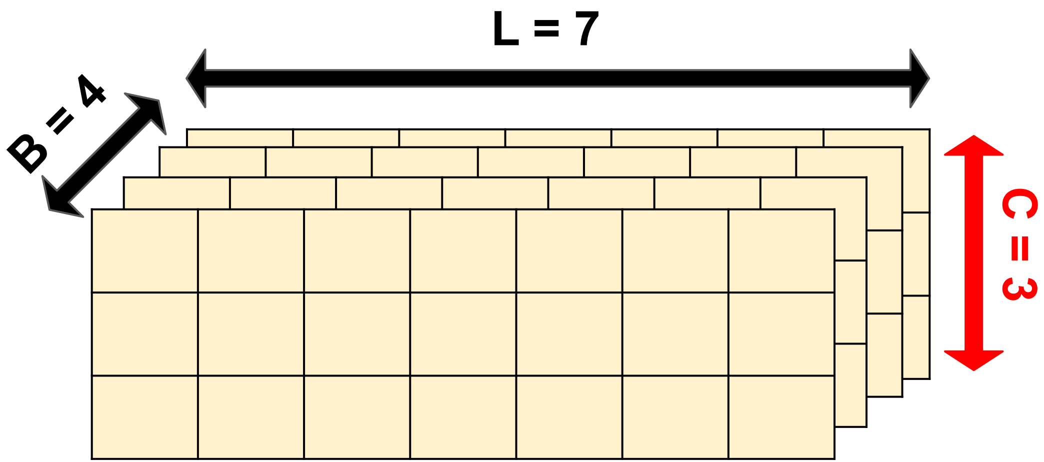
1. Preliminaries

1. Preliminaries

(1) Time Series (TS)

Shape: (B,C,L)

- B: Batch size (# of TS)
- **C: # of Channels**
- L : Length of TS



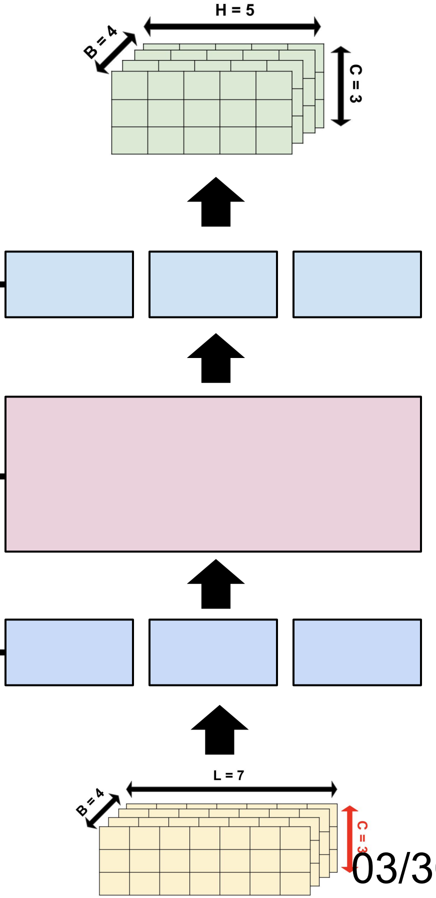
1. Preliminaries

(2) Time Series Forecasting Model

Step 3) Projection layer

Step 2) Encoder

Step 1) Tokenization layer
(= Patch Embedding)



1. Preliminaries

(2) Time Series Forecasting Model

(B,C,**D**) -> (B,C,**H**)
(*H: Forecast horizon)

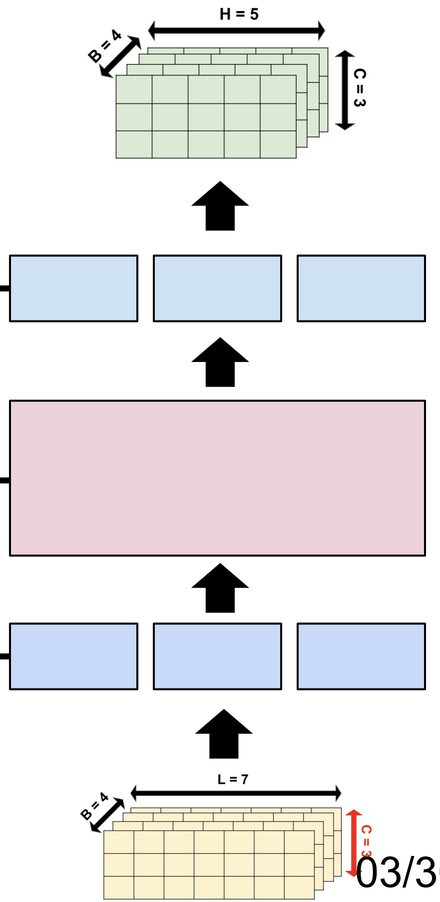
Step 3) Projection layer

(B,C,**D**) -> (B,C,**D**)

Step 2) Encoder

(B,C,**L**) -> (B,C,**D**)

Step 1) Tokenization layer
(= Patch Embedding)



1. Preliminaries

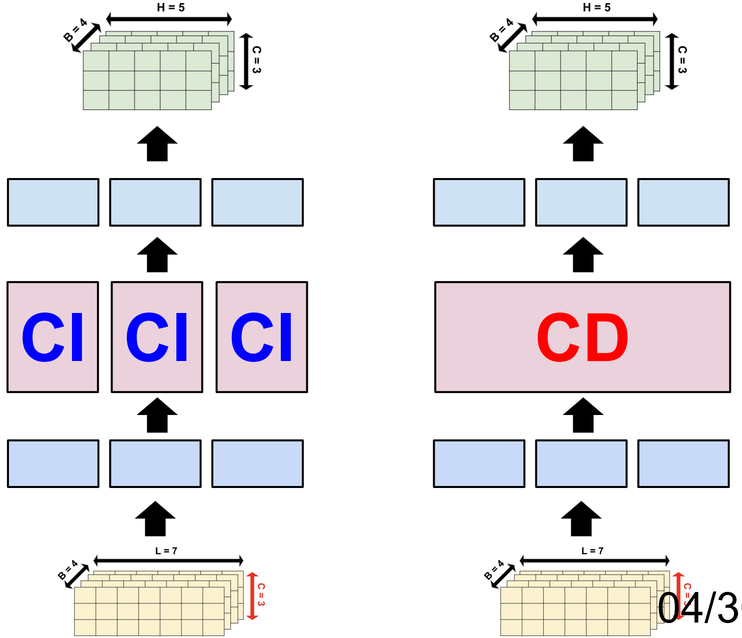
(3) Channel Independence (CI) vs. Channel Dependence (CD)

CI: Ignores dependencies btw channels

- **<u>High Robustness</u>** & Low Capacity

CD: Considers dependencies btw channels

- Low Robustness & **<u>High Capacity</u>**

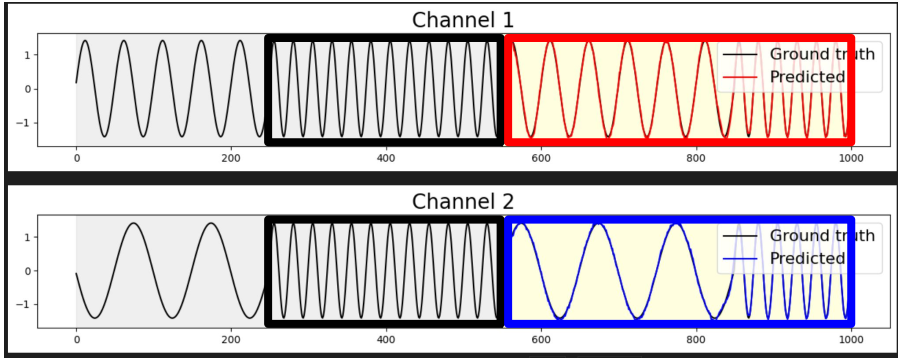


1. Preliminaries

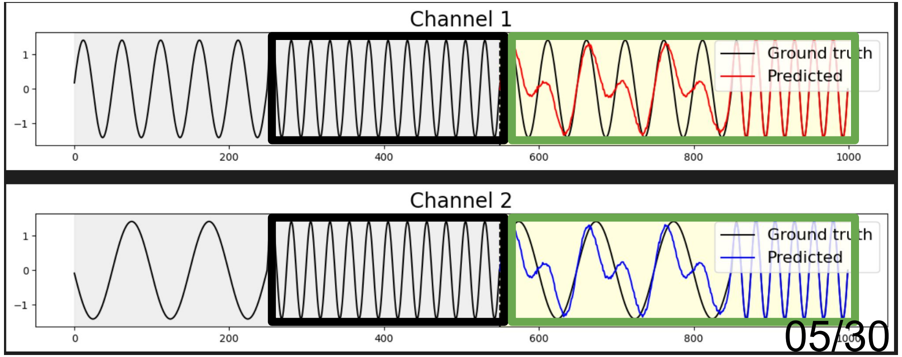
(4) Channel Identification

*Does the model have **<u>ability to distinguish btw channels</u>**?*

No — Same input & **Different** output



Yes — Same input & **Same** output



2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

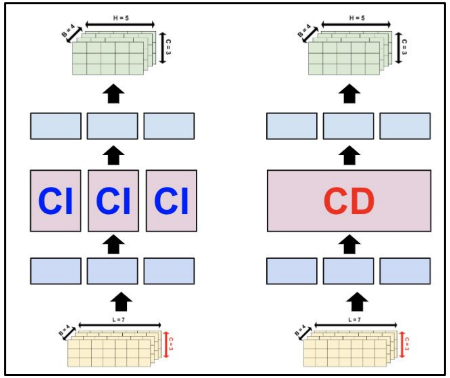
(1) Abstract

- To **balance CI & CD**, propose “C-LoRA”
- **Channel-aware low-rank adaptation (C-LoRA)**
  - Condition CD models on **identity-aware individual components**
  - **Plug-and-play** method
- Improve the performance of **both CI & CD models**

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(1) Abstract

(In 1. Preliminaries)



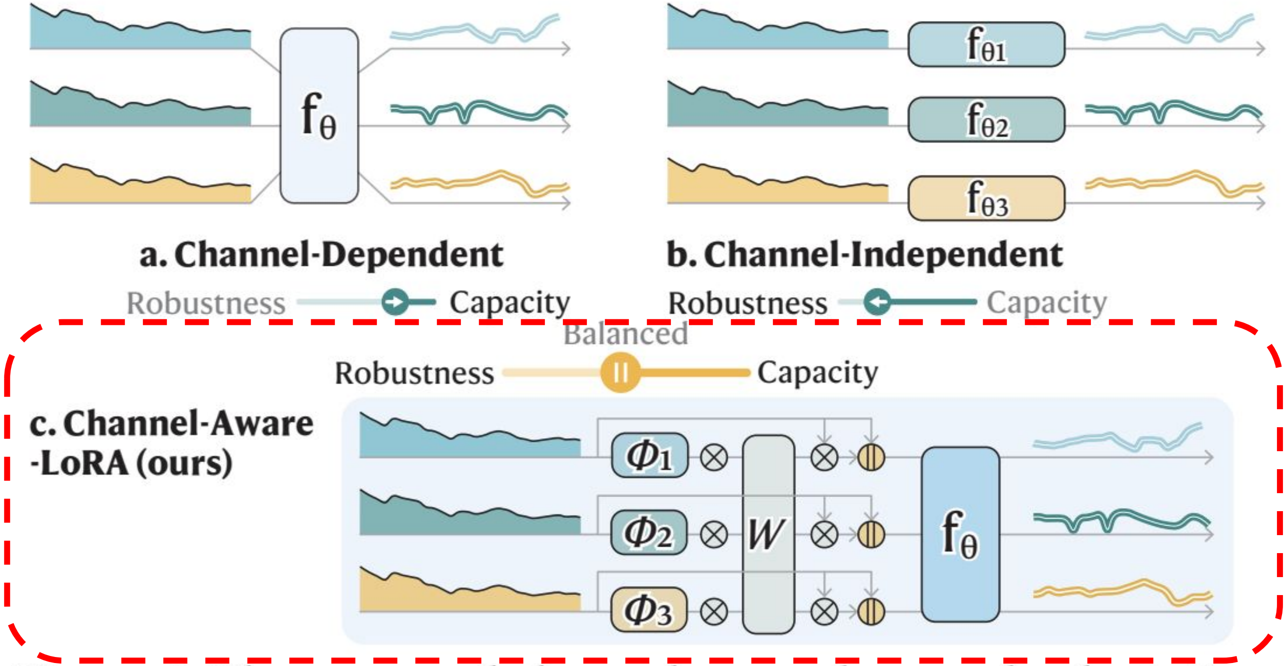


Figure 1: The proposed channel-aware low-rank adaptation.

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(2) Low-rank Adaptation

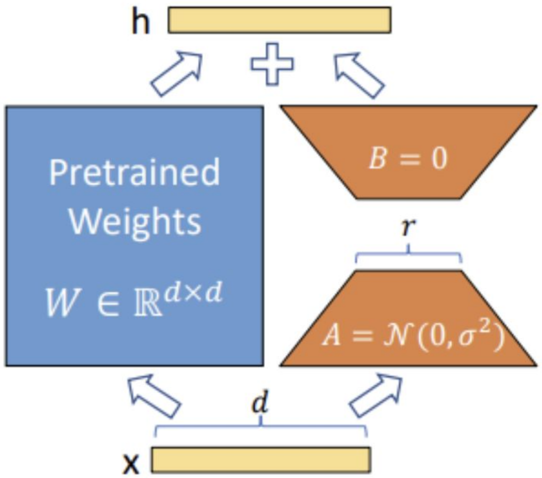


Figure 1: Our reparametrization. We only train A and B.

Pretrained weights

$$W_0 \in \mathbb{R}^{d\times k}$$

Constrain its update with …

$$\underbrace{W_0}_{\text{(Freeze)}} + \Delta W = W_0 + BA$$

$$B \in \mathbb{R}^{d\times r} \quad A \in \mathbb{R}^{r\times k}$$

$$r \ll \min(d, k)$$

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(3) C-LoRA

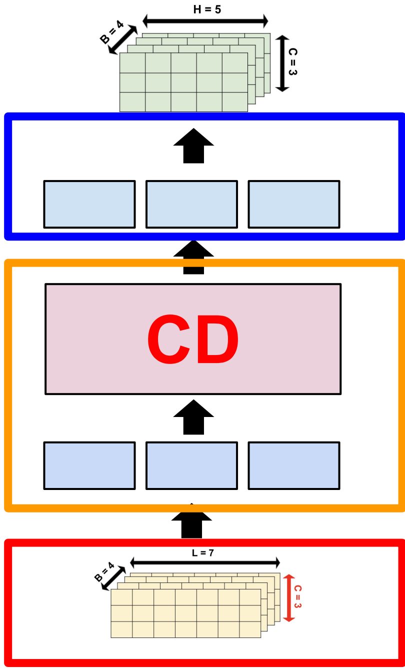
General backbone

$$\bar{\mathrm{X}} = \textsc{Normalization}(\mathrm{X}),$$

$$\mathbf{z}_c^{(0)} = \textsc{TokenEmbedding}(\bar{\mathrm{X}}_{:,c}), \forall c = 1, \ldots, C,$$

$$\text{(Optional): } \mathrm{Z}^{(\ell+1)} = \textsc{ChannelMixing}(\mathrm{Z}^{(\ell)}), \forall \ell = 0, \ldots, L,$$

$$\widehat{\mathrm{Y}} = \textsc{Projection}(\mathrm{Z}^{(L+1)}),$$



2. C ries Forecasting

(3) the nonstationarity of time series, $\textsc{TokenEmbedding} : \mathbb{R}^T \mapsto \mathbb{R}^D$ and $\textsc{Projection} : \mathbb{R}^D \mapsto \mathbb{R}^H$ are usually implemented by MLPs to process temporal features, and $\textsc{ChannelMixing} : \mathbb{R}^{C\times D} \mapsto \mathbb{R}^{C\times D}$ is optional for CD models by Transformer blocks or MLPs.

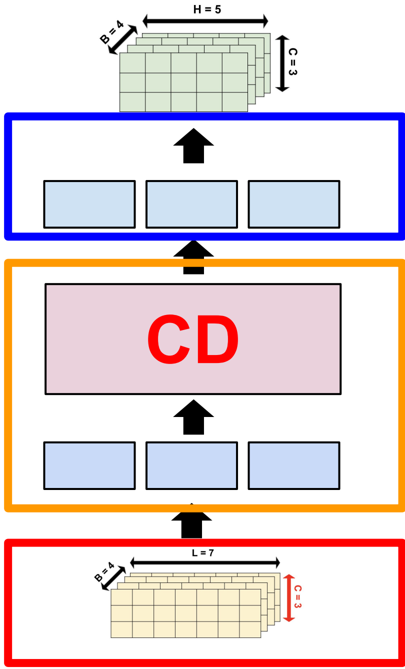
General backbone

$$\bar{\mathbf{X}} = \textsc{Normalization}(\mathbf{X}),$$

$$\mathbf{z}_c^{(0)} = \textsc{TokenEmbedding}(\bar{\mathbf{X}}_{:,c}), \forall c = 1, \ldots, C,$$

$$\text{(Optional): } \mathbf{Z}^{(\ell+1)} = \textsc{ChannelMixing}(\mathbf{Z}^{(\ell)}), \forall \ell = 0, \ldots, L,$$

$$\widehat{\mathbf{Y}} = \textsc{Projection}(\mathbf{Z}^{(L+1)}),$$



2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(3) C-LoRA

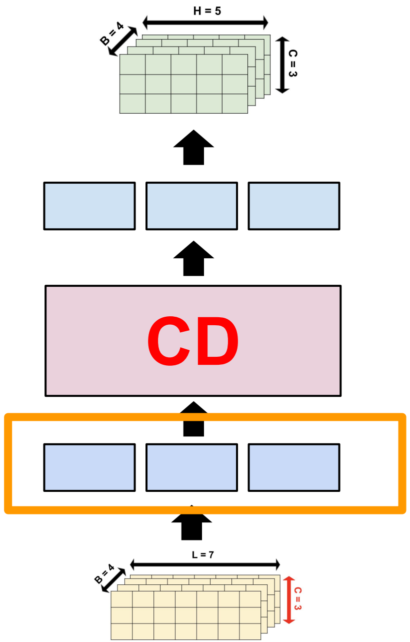
Naive option to reflect channel identity

- **Individual tokenization** for each channel

  => ***Computationally expensive!!***

$$\boxed{\mathbf{z}_c^{(0)} = \mathrm{MLP}_c\left(\overline{\mathbf{X}}_{,c};\theta_c\right), \forall c = 1, \ldots, C}$$

Still, can utilize it as a baseline method for comparison.



2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(3) C-LoRA

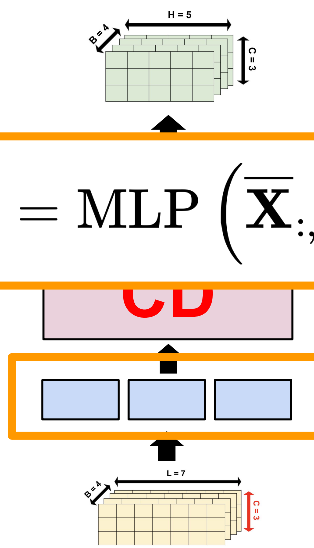
C-LoRA: Channel-wise adaptation in a CD model

- Model the individual channels in a **parameter-efficient way**
- Components
  - (1) **Low-rank adapter** (for each channel): $\phi^{(c)} \in \mathbb{R}^{\boxed{r} \times D}$ intrinsic rank $r \ll D$
  - (2) **Linear layer**: $\mathbf{W} \in \mathbb{R}^{r \times \boxed{d}}$ adaptation dimension
- Result: **Channel-specific parameters** $\boxed{\tilde{\phi}^{(c)} = \mathrm{ReLU}\left(\phi^{(c),\mathrm{T}} \mathbf{W}\right) \in \mathbb{R}^{D \times d}}$

Channel adaptation $\boxed{\mathbf{z}^{(0)}_{c,\phi}} = \boxed{\mathbf{z}^{(0)}_{c}}^{\mathrm{T}} \boxed{\widetilde{\phi}^{(c)}} \in \mathbb{R}^{d}$

Aggregate $\underline{\mathbf{z}^{(0)}_{\phi} = \left\{\boxed{\mathbf{z}^{(0)}_{c,\phi}}\right\}_{c=1}^{C}} \in \mathbb{R}^{C\times d}$

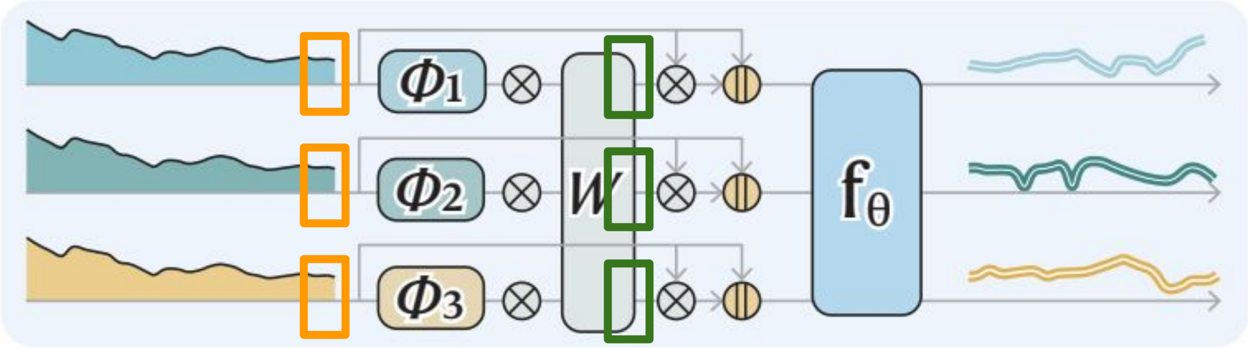
Concat $\mathbf{Z}^{(0)} = \left[\underline{\mathrm{MLP}(\overline{\mathbf{X}};\theta)} \| \underline{\mathbf{Z}^{(0)}_{\phi}}\right] \in \mathbb{R}^{C\times(D+d)}$



$\boxed{\mathbf{z}^{(0)}_{c} = \mathrm{MLP}\left(\overline{\mathbf{X}}_{:,c};\theta\right)}$

- (1) **Low-rank adapter** (for each channel): $\phi^{(c)} \in \mathbb{R}^{\boxed{r}\times D}$ $r \ll D$
- (2) **Linear layer**: $\mathbf{W} \in \mathbb{R}^{r\times \boxed{d}}$ adaptation dimension
- Result: **Channel-specific parameters** $\boxed{\tilde{\phi}^{(c)} = \mathrm{ReLU}\left(\phi^{(c),\mathrm{T}}\mathbf{W}\right) \in \mathbb{R}^{D\times d}}$

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting



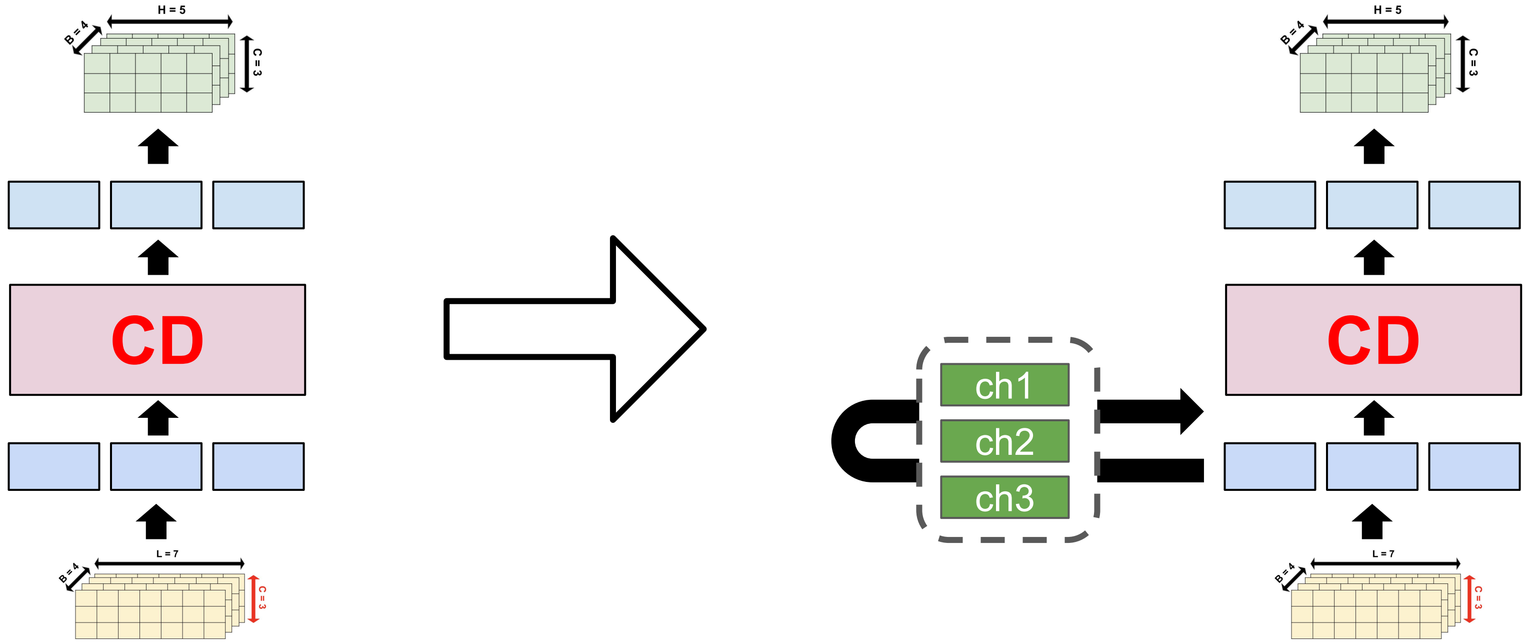
Channel adaptation $\mathbf{z}_{c,\phi}^{(0)} = \mathbf{z}_c^{(0)\top} \widetilde{\phi}^{(c)} \in \mathbb{R}^d$

Aggregate $\mathbf{z}_{\phi}^{(0)} = \left\{ \mathbf{z}_{c,\phi}^{(0)} \right\}_{c=1}^{C} \in \mathbb{R}^{C \times d}$

Concat $\mathbf{Z}^{(0)} = \left[\mathrm{MLP}(\overline{\mathbf{X}}; \theta) \| \mathbf{Z}_{\phi}^{(0)} \right] \in \mathbb{R}^{C \times (D+d)}$

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(3) C-LoRA



2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(4) Experiments

a) TS Forecasting

- Various CI & CD Models
- Plug-and-play method
- Improvement in TS forecasting

Table 1: Results of the LTSF benchmarks. We report the forecast error of different models under different prediction lengths. The input sequence length is set to 96 for all methods. *IMP* shows the average percentage of MSE/MAE improvement of C-LoRA.

Models		iTransformer		w/ C-LoRA		TSMixer		w/ C-LoRA		RMLP		w/ C-LoRA		FreTS		w/ C-LoRA		EDformer		w/ C-LoRA		utoformer		w/ C-LoRA		nformer		w/ C-LoRA		MP
Metric		MSE	MAE	MSE	MAE	SE	MAE	MSE	MAE	SE	MAE	MSE	MAE	SE	MAE	MSE	MAE	SE	MAE	MSE	MAE	SE	MAE	MSE	MAE	SE	MAE	MSE	MAE	%
ETTm1	96	0.345	0.376	0.331	0.367	332	0.370	0.317	0.356	337	0.374	.321	0.360	340	0.376	.330	0.369	379	0.419	0.375	0.422	505	0.475	0.512	0.480	672	0.571	0.577	0.542	.23
	192	0.384	0.394	0.373	0.390	372	0.390	0.358	0.377	379	0.391	.363	0.380	383	0.398	.375	0.398	426	0.441	0.411	0.435	553	0.496	0.558	0.503	795	0.669	0.720	0.634	.67
	336	0.418	0.415	0.409	0.414	405	0.411	0.389	0.400	412	0.412	.395	0.403	420	0.425	.406	0.418	445	0.459	0.423	0.447	621	0.537	0.618	0.523	212	0.871	0.982	0.756	.51
	720	0.481	0.451	0.479	0.449	469	0.447	0.455	0.436	478	0.447	.462	0.440	499	0.477	.479	0.462	543	0.490	0.509	0.488	671	0.561	0.592	0.520	166	0.823	1.121	0.794	.68
	Avg	0.407	0.409	0.398	0.405	395	0.405	0.380	0.392	402	0.406	.385	0.396	411	0.419	.398	0.412	448	0.452	0.430	0.448	588	0.517	0.570	0.507	961	0.734	0.850	0.682	.59
ETTh1	96	0.390	0.406	0.390	0.407	398	0.412	0.396	0.409	405	0.413	.381	0.394	398	0.410	0.402	0.409	376	0.419	0.376	0.417	449	0.459	0.453	0.469	912	0.717	0.874	0.715	.90
	192	0.442	0.436	0.440	0.436	451	0.442	0.451	0.440	460	0.444	.442	0.427	453	0.446	0.460	0.444	420	0.448	0.419	0.443	500	0.482	0.490	0.477	037	0.779	1.044	0.775	.84
	336	0.485	0.459	0.482	0.457	489	0.461	0.493	0.460	505	0.466	.493	0.455	508	0.480	0.509	0.473	459	0.465	0.458	0.466	521	0.496	0.511	0.487	116	0.810	1.146	0.843	.30
	720	0.493	0.485	0.486	0.477	507	0.486	0.498	0.480	514	0.490	.493	0.475	560	0.537	0.547	0.515	506	0.507	0.472	0.486	514	0.512	0.503	0.508	230	0.887	1.170	0.869	.88
	Avg	0.453	0.447	0.450	0.444	461	0.450	0.460	0.447	471	0.453	.452	0.438	480	0.468	0.480	0.460	440	0.460	0.431	0.453	496	0.487	0.489	0.485	074	0.800	1.059	0.801	.28
Electricity	96	0.148	0.240	0.139	0.234	177	0.278	0.155	0.256	201	0.287	.168	0.258	320	0.403	.165	0.262	195	0.309	0.193	0.307	203	0.318	0.193	0.307	329	0.407	0.329	0.412	0.55
	192	0.162	0.253	0.160	0.254	193	0.293	0.172	0.270	209	0.297	.179	0.268	325	0.406	.176	0.270	202	0.315	0.203	0.316	225	0.334	0.222	0.330	338	0.419	0.347	0.430	.53
	336	0.178	0.269	0.171	0.266	215	0.315	0.191	0.289	228	0.316	.196	0.285	370	0.435	.195	0.289	234	0.347	0.230	0.343	282	0.377	0.266	0.370	364	0.439	0.352	0.434	0.29
	720	0.225	0.317	0.195	0.289	260	0.352	0.230	0.321	273	0.350	.238	0.319	416	0.474	.235	0.325	261	0.365	0.262	0.365	314	0.383	0.299	0.389	397	0.460	0.395	0.456	0.24
	Avg	0.178	0.270	0.166	0.261	211	0.310	0.187	0.284	228	0.313	.195	0.283	358	0.430	.193	0.287	223	0.334	0.222	0.333	256	0.353	0.245	0.349	357	0.431	0.356	0.433	.95
Weather	96	0.174	0.214	0.164	0.209	181	0.228	0.158	0.206	196	0.235	.163	0.208	186	0.241	.165	0.228	220	0.300	0.218	0.299	266	0.336	0.234	0.312	300	0.384	0.265	0.348	.35
	192	0.221	0.254	0.209	0.251	227	0.263	0.207	0.249	240	0.271	.209	0.249	222	0.273	.210	0.274	278	0.344	0.282	0.350	307	0.367	0.282	0.344	598	0.544	0.381	0.427	.27
	336	0.278	0.296	0.268	0.294	280	0.300	0.266	0.292	291	0.307	.264	0.289	272	0.316	.264	0.317	339	0.382	0.349	0.390	359	0.395	0.357	0.395	578	0.523	0.515	0.511	.74
	720	0.358	0.349	0.349	0.346	353	0.347	0.348	0.345	363	0.353	.342	0.340	350	0.381	.343	0.370	409	0.438	0.411	0.420	419	0.428	0.415	0.424	059	0.741	0.792	0.651	.47
	Avg	0.258	0.278	0.248	0.275	260	0.285	0.245	0.273	273	0.292	.245	0.272	258	0.303	.246	0.297	312	0.366	0.315	0.365	338	0.382	0.322	0.369	634	0.548	0.488	0.484	.76
Solar	96	0.203	0.237	0.175	0.219	222	0.281	0.182	0.272	233	0.296	.213	0.272	237	0.300	.231	0.295	242	0.342	0.226	0.319	884	0.711	0.603	0.545	236	0.279	0.214	0.245	0.96
	192	0.233	0.261	0.211	0.259	261	0.301	0.207	0.275	260	0.316	.234	0.292	265	0.321	.261	0.318	285	0.380	0.245	0.366	834	0.692	0.682	0.563	227	0.287	0.241	0.290	.64
	336	0.248	0.273	0.222	0.261	271	0.299	0.212	0.272	276	0.323	.247	0.301	283	0.330	.277	0.325	282	0.376	0.246	0.350	941	0.723	0.739	0.588	262	0.310	0.246	0.307	.54
	720	0.249	0.275	0.203	0.263	267	0.293	0.201	0.262	273	0.316	.244	0.291	286	0.326	.281	0.322	357	0.427	0.304	0.410	882	0.717	0.801	0.642	329	0.355	0.279	0.329	0.05
	Avg	0.233	0.262	0.203	0.251	255	0.294	0.201	0.270	261	0.313	.235	0.289	268	0.319	.263	0.315	292	0.381	0.255	0.361	885	0.711	0.706	0.585	264	0.308	0.245	0.293	.65
PEMS04	96	0.159	0.272	0.124	0.237	166	0.285	0.111	0.220	273	0.372	.180	0.292	274	0.372	.211	0.311	220	0.338	0.219	0.336	553	0.583	0.398	0.471	123	0.236	0.125	0.236	6.60
	192	0.182	0.290	0.162	0.266	196	0.166	0.127	0.235	308	0.395	.224	0.320	307	0.392	.248	0.342	313	0.419	0.310	0.418	938	0.772	0.502	0.546	142	0.252	0.136	0.246	2.48
	336	0.186	0.291	0.166	0.268	204	0.312	0.134	0.242	283	0.373	.224	0.322	285	0.375	.247	0.346	231	0.339	0.232	0.339	953	0.772	0.672	0.645	155	0.263	0.148	0.256	3.12
	720	0.242	0.332	0.203	0.298	238	0.345	0.148	0.259	328	0.407	.258	0.352	329	0.406	.286	0.376	643	0.592	0.605	0.577	120	0.823	0.879	0.747	192	0.297	0.162	0.272	4.83
	Avg	0.192	0.296	0.164	0.267	201	0.277	0.130	0.239	298	0.387	.222	0.322	299	0.386	.248	0.344	352	0.422	0.342	0.418	891	0.738	0.613	0.602	153	0.262	0.143	0.253	4.86
PEMS08	96	0.169	0.276	0.109	0.216	252	0.355	0.124	0.239	284	0.375	.198	0.300	285	0.380	.220	0.328	221	0.325	0.234	0.329	613	0.596	0.444	0.508	171	0.283	0.165	0.274	9.24
	192	0.188	0.288	0.137	0.239	322	0.385	0.151	0.256	336	0.409	.222	0.310	335	0.409	.280	0.371	332	0.412	0.340	0.420	111	0.829	0.616	0.596	202	0.304	0.202	0.298	0.35
	336	0.196	0.289	0.148	0.243	326	0.374	0.170	0.265	327	0.394	.266	0.337	328	0.397	.284	0.363	247	0.322	0.247	0.321	136	0.836	0.716	0.648	224	0.309	0.210	0.299	7.27
	720	0.235	0.320	0.186	0.277	388	0.422	0.225	0.308	372	0.430	.303	0.372	373	0.431	.327	0.396	564	0.556	0.546	0.546	409	0.948	0.901	0.733	247	0.324	0.238	0.316	6.12
	Avg	0.197	0.293	0.145	0.244	322	0.384	0.168	0.267	330	0.402	.247	0.330	330	0.404	.278	0.365	341	0.404	0.342	0.404	067	0.802	0.669	0.621	211	0.305	[illegible]	[illegible]	[illegible]

2. Channel-Aware Low-Rank Adaptation in Time Series Forecasting

(4) Experiments

b) Parameter Efficiency

- **CM**: Standard CD
- **CI**: Standard CI
- **CInd**: Standard CI + Projection layer per channel

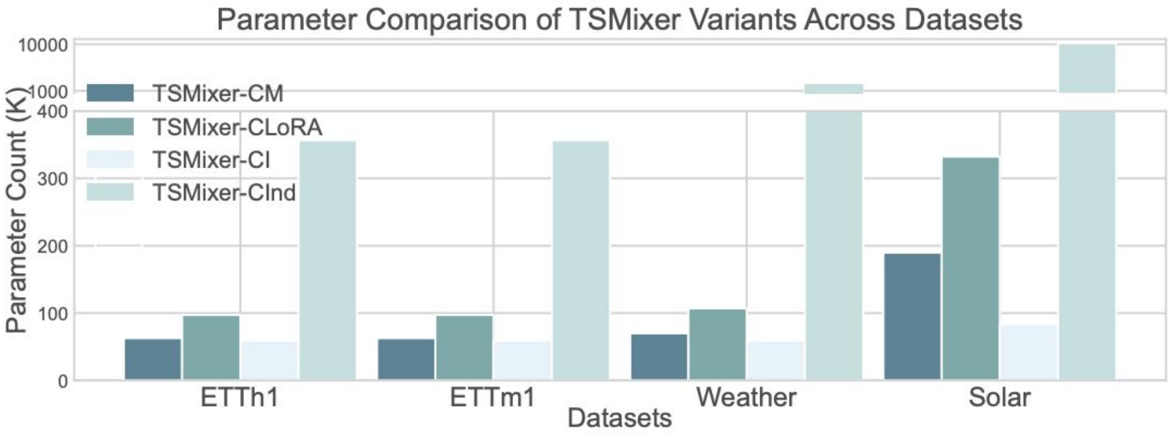


Figure 2: Parameter comparison of different strategies.

2. Channel-Aware Low-Rank Ad… g

Figure 5: randomly shuffling the order of channels (Solar).

w/o C-LoRA **w/ C-LoRA**

(4) Experiments

c) Channel Identity Permutation

- By **randomly shuffling** the order of channels

  => To evaluate the **importance of channel identity**

- Result)
  - (1) Both CI and CD models show a **performance drop** after shuffling
  - (2) Models **with C-LoRA** have **more pronounced error increases**.

  => ***With C-LoRA, can better preserve the channel identity information!***

3. From Similarity to Superiority:

Channel Clustering for Time Series Forecasting

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(1) Abstract

- Lack of channel strategy that balances ..
    - (1) Individual channel treatment
    - (2) Interaction btw channels
- Propose **Channel Clustering Module (CCM)**
    - **Clusters** the channels (based on intrinsic similarity)
    - Leverage **cluster identity** (instead of channel identity)

=> Combining the best of CD and CI worlds

(2) Toy Experiment

Random shuffling the channels

- Compare the performance gap between
  - (1) **w/** shuffling vs. (2) **w/o** shuffling

Table 1. Averaged performance gain from a channel identity information ($\Delta\mathcal{L}(\%)$) and Pearson Correlation Coefficients (PCC) between $\{\Delta\mathcal{L}_{ij}\}_{i,j}$ and $\{\textsc{Sim}(X_i, X_j)\}_{i,j}$. The values are averaged across all test samples.

Base Model Channel Strategy		TSMixer *CD*	DLinear *CI*	PatchTST *CI*	TimesNet *CD*
ETTh1	$\Delta\mathcal{L}(\%)$	2.67	1.10	11.30	18.90
	PCC	- 0.67	- 0.66	- 0.61	- 0.66
ETTm1	$\Delta\mathcal{L}(\%)$	4.41	5.55	6.83	14.98
	PCC	- 0.68	- 0.67	- 0.68	- 0.67
Exchange	$\Delta\mathcal{L}(\%)$	16.43	19.34	27.98	24.57
	PCC	- 0.62	- 0.62	- 0.47	- 0.49

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(2) Toy Experiment

Random shuffling the channels

- Compare the performance gap between
  - (1) **w/** shuffling vs. (2) **w/o** shuffling
- Findings:
  - (1) Shuffling **degrades** the performance => **Importance of channel identity**
  - (2) Correlation btw a) Degradation & b) Channel similarity

    The **Higher similarity** btw channels, the **less degradation similarity**

    => **Importance of channel clustering**

Table 1. Averaged performance gain from a channel identity information ($\Delta\mathcal{L}(\%)$) and Pearson Correlation Coefficients (*PCC*) between $\{\Delta\mathcal{L}_{ij}\}_{i,j}$ and $\{\textsc{Sim}(X_i, X_j)\}_{i,j}$. The values are averaged across all test samples.

Base Model Channel Strategy		TSMixer *CD*	DLinear *CI*	PatchTST *CI*	TimesNet *CD*
ETTh1	$\Delta\mathcal{L}(\%)$	2.67	1.10	11.30	18.90
	PCC	- 0.67	- 0.66	- 0.61	- 0.66
ETTm1	$\Delta\mathcal{L}(\%)$	4.41	5.55	6.83	14.98
	PCC	- 0.68	- 0.67	- 0.68	- 0.67
Exchange	$\Delta\mathcal{L}(\%)$	16.43	19.34	27.98	24.57
	PCC	- 0.62	- 0.62	- 0.47	- 0.49

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(2) Toy Experiment

Random shuffling the channels

- Compare the performance gap between
  - (1) **w/** shuffling vs
- Findings:
  - (1) Shuffling **degr** **nel identity**
  - (2) Correlation btw a) Degradation & b) Channel similarity

    The **Higher similarity** btw channels, the **less degradation similarity**

    => **Importance of channel clustering**

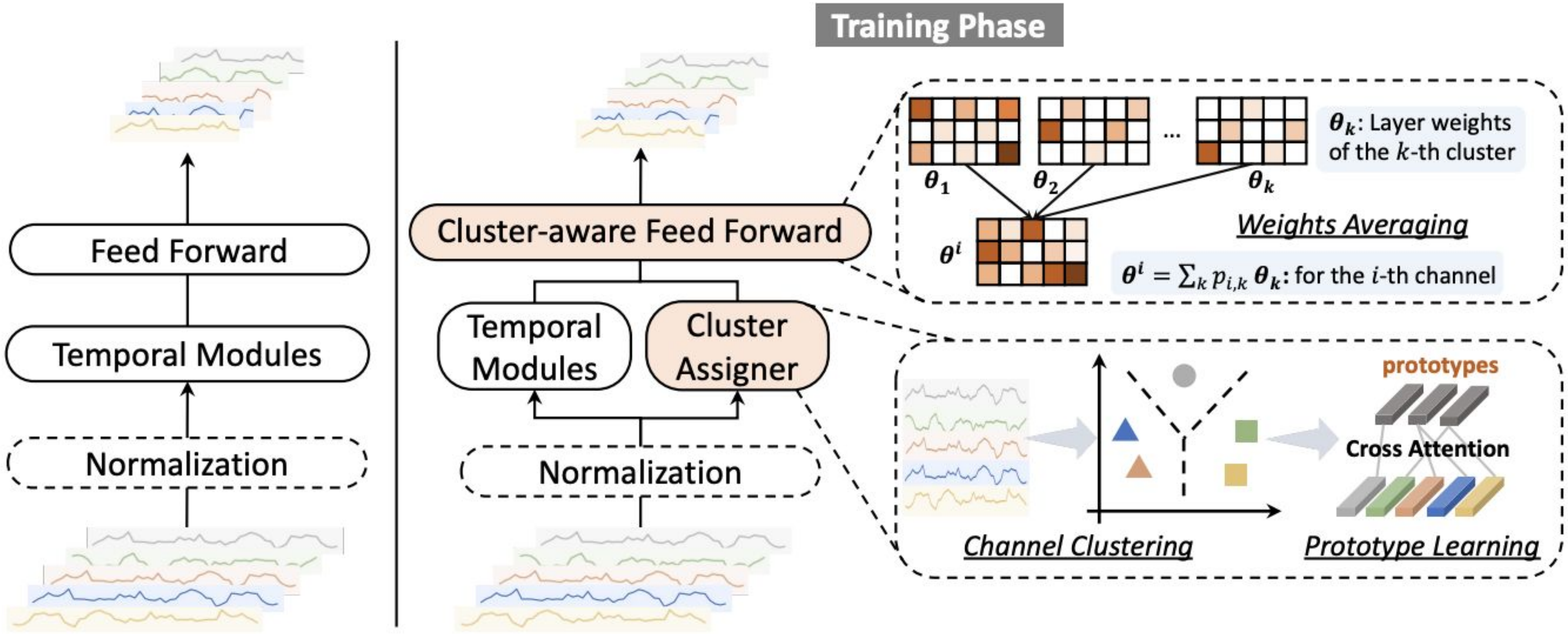
Table 1. Averaged performance gain from a channel identity information ($\Delta\mathcal{L}(\%)$) and Pearson Correlation Coefficients (PCC) between $\{\Delta\mathcal{L}_{ij}\}_{i,j}$ and $\{\textsc{Sim}(X_i, X_j)\}_{i,j}$. The values are averaged across all test samples.

Base Model		TSMixer	DLinear	PatchTST	TimesNet
Channel Strategy		*CD*	*CI*	*CI*	*CD*
ETTh1	$\Delta\mathcal{L}(\%)$	2.67	1.10	11.30	18.90
	PCC	- 0.67	- 0.66	- 0.61	- 0.66
	$\Delta\mathcal{L}(\%)$	4.41	5.55	6.83	14.98
				.68	- 0.67
				.98	24.57
				.47	- 0.49

$$\mathrm{SiM}\left(X_i, X_j\right) = \exp\left(\frac{-\|X_i - X_j\|^2}{2\sigma^2}\right)$$

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(3) Channel Clustering Module (CCM)



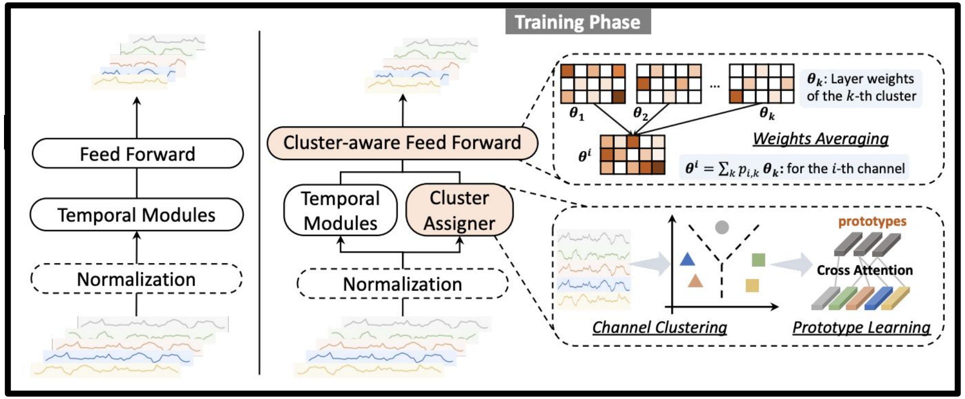
w/o CCM **w/ CCM**

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(3) Channel Clustering Module (CCM)

a) **Cluster Assigner**

- Assigns the cluster for each channel
- Cluster centroid = Prototype



b) **Cluster-Aware FFN**

- Models channels "independently" + Cluster information
- Cluster information= Weighted average of (Linear layer x C)

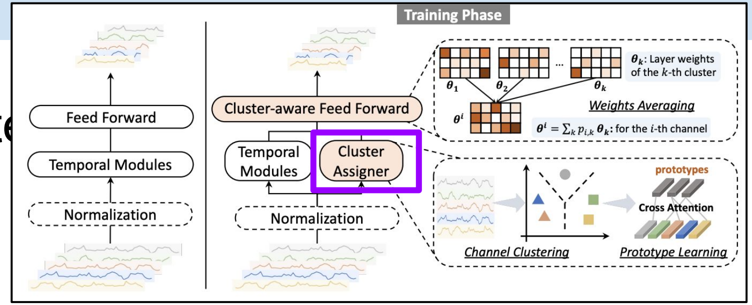
(3) Channel Clustering Module (CCM)

Notation

$X_{[;,i]} \in \mathbb{R}^T$ (X_i for simplicity)

- Input TS: $X = [x_1, \ldots \boldsymbol{x}_T] \in \mathbb{R}^{T \times C}$
- Output TS: $Y = \left[\hat{\boldsymbol{x}}_{T+1}, \ldots, \hat{\boldsymbol{x}}_{T+H}\right] \in \mathbb{R}^{H \times C}$
- Model
    - CI model: $f^{(i)} : \mathbb{R}^T \rightarrow \mathbb{R}^H$ for $i = 1, \cdots, C$,
    - CD model: $f : \mathbb{R}^{T \times C} \rightarrow \mathbb{R}^{H \times C}$

3. From Similarity to Superiority: Channel Cluste

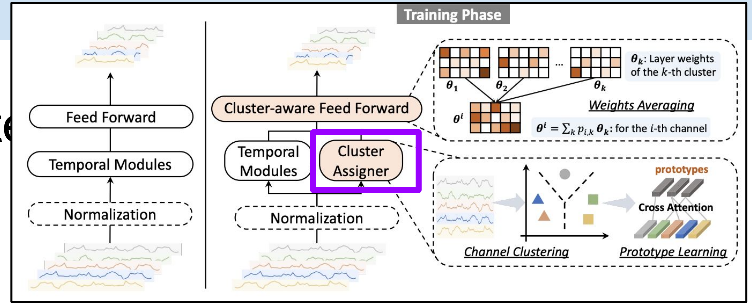


(3) Channel Clustering Module (CCM)

a) **Cluster Assigner**

- (1) TS embedding (for each channel): $X_i \in \mathbb{R}^T \rightarrow \boxed{h_i} \in \mathbb{R}^d$
- (2) Cluster embeddings: $\{\boxed{c_1}, \cdots, c_K\}$, where $c_k \in \mathbb{R}^d$

3. From Similarity to Superiority: Channel Cluste



(3) Channel Clustering Module (CCM)

a) **Cluster Assigner**

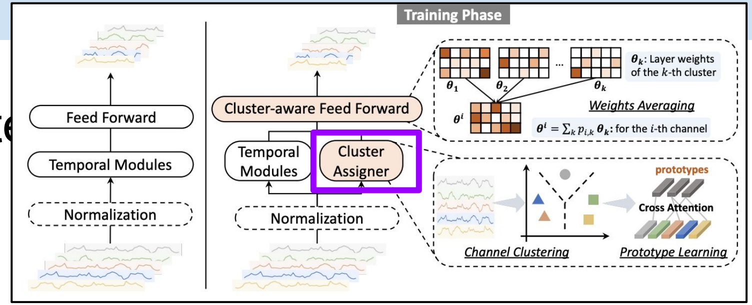
- (1) TS embedding (for each channel): $X_i \in \mathbb{R}^T \to \boxed{h_i} \in \mathbb{R}^d$
- (2) Cluster embeddings: $\{\boxed{c_1}, \cdots, c_K\}$, where $c_k \in \mathbb{R}^d$

Cosine similarity btw (1) & (2):

$$p_{i,k} = \text{Normalize}\left(\frac{c_k}{\|c_k\|}\frac{h_i}{\|h_i\|}\right) \in [0,1]$$

ensures that $\sum_k p_{i,k} = 1$

3. From Similarity to Superiority: Channel Cluste



(3) Channel Clustering Module (CCM)

a) Cluster Assigner

\# of channels \# of clusters

- (3) Clustering membership matrix: $\mathbf{M} \in \mathbb{R}^{\boxed{C} \times \boxed{K}}$

$$\text{where } \mathbf{M}_{ik} \approx \text{Bernoulli}\left(p_{i,k}\right)$$

$\mathbf{P} \in \mathbb{R}^{K\ X\ C}$

c_1	0.1	0.5	0.1	0.4	0.2	0.2
c_2	0.2	0.1	0.2	0.2	0.1	0.1
c_3	0.6	0.2	0.6	0.2	0.3	0.2
c_K	0.1	0.2	0.1	0.2	0.4	0.5

Binary $\longrightarrow$

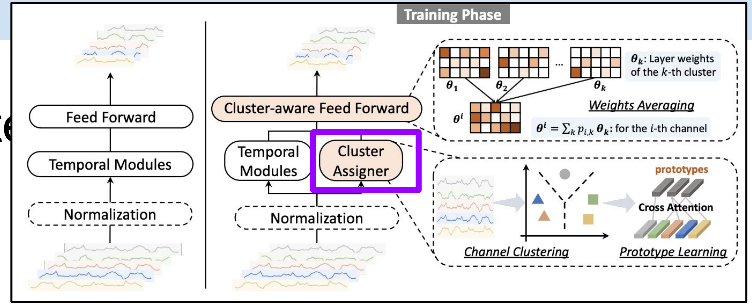
$\mathbf{M} \in \mathbb{R}^{K\ X\ C}$

c_1	0	1	0	1	0	0
c_2	0	0	0	0	0	0
c_3	1	0	1	0	0	0
c_K	0	0	0	0	1	1

채널 5는 Cluster K에 속함

https://www.youtube.com/watch?v=h2PhpY6OFZ8&t=2222s

3. From Similarity to Superiority: Channel Cluste



(3) Channel Clustering Module (CCM)

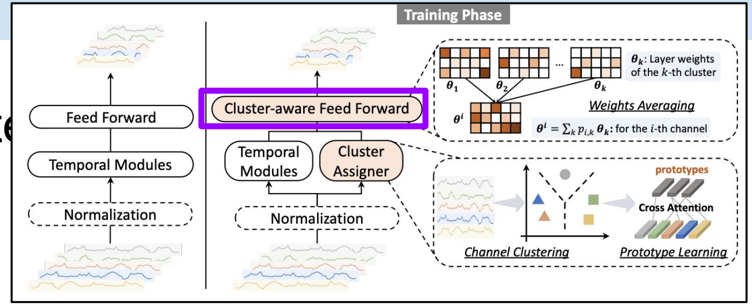
a) **Cluster Assigner**

$$p_{i,k} = \text{Normalize}\left(\frac{c_k^\top h_i}{\|c_k\|\,\|h_i\|}\right) \in [0,1]$$

- (4) Prototype embedding: $\widehat{\mathbf{C}} \in \mathbb{R}^{K\times d}$
  - Updated cluster embedding for subsequent clustering probability
  - Via cross-attention

$$\widehat{\mathbf{C}} = \text{Normalize}\left(\exp\left(\frac{(W_Q\mathbf{C})(W_K\mathbf{H})^\top}{\sqrt{d}}\right) \odot \mathbf{M}^\top\right) W_V\mathbf{H}$$

3. From Similarity to Superiority: Channel Cluste



(3) Channel Clustering Module (CCM)

b) Cluster-aware FFN

Notation

- Linear layer for the k-th cluster: $h_{\theta_k}(\cdot)$
- Embedding of i-th channel before the last layer: Z_i

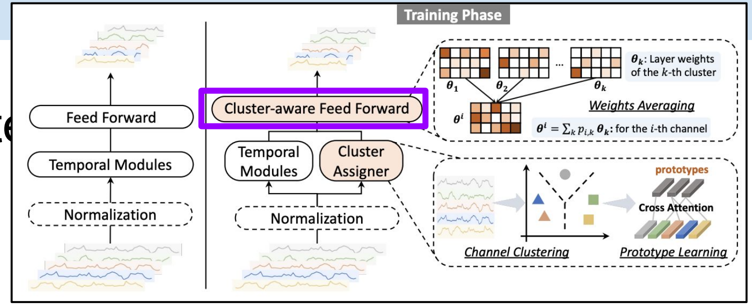
Prediction = **Average across the outputs of all Cluster-aware FFN**

$$\boxed{Y_i = \sum_k p_{i,k} h_{\theta_k}(Z_i)}$$

For computational efficiency, it is equivalent to

$$\boxed{Y_i = h_{\theta^i}(Z_i)} \text{ with avg weights } \theta^i = \sum_k p_{i,k}\theta_k$$

3. From Similarity to Superiority: Channel Cluste



(3) Channel Clustering Module (CCM)

c) Cluster Loss

- Designed for the clustering quality
- Considers both intra-similarity & inter-dissimilarity

$$\mathcal{L}_C = -\boxed{\mathrm{Tr}\left(\mathbf{M}^\top \mathbf{S}\mathbf{M}\right)} + \boxed{\mathrm{Tr}\left(\left(\mathbf{I} - \mathbf{M}\mathbf{M}^\top\right)\mathbf{S}\right)}$$

$$\mathbf{S} \in \mathbb{R}^{C \times C} : \ \text{Channel similarity matrix with } \mathbf{S}_{ij} = \mathrm{Sim}\left(X_i, X_j\right)$$

$$\mathcal{L} = \mathcal{L}_F + \beta\mathcal{L}_C$$

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(4) Experiments

a) TS Forecasting

- Various CI & CD Models
- Plug-and-play method
- Improvement in TS forecasting

Table 4. Long-term forecasting results on 9 real-world datasets in terms of MSE and MAE, the lower the better. The forecasting horizons are $\{96, 192, 336, 720\}$. The better performance in each setting is shown in **bold**. The best results for each row are underlined. The last column shows the average percentage of MSE/MAE improvement of CCM over four base models.

Model		TSMixer		+ CCM		DLinear		+ CCM		PatchTST		+ CCM		TimesNet		+ CCM		IMP(%)
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	**0.361**	**0.392**	0.365	0.393	0.375	0.399	**0.371**	**0.393**	0.375	0.398	**0.371**	**0.396**	0.384	0.402	**0.380**	**0.400**	0.539
	192	0.404	**0.418**	**0.402**	**0.418**	0.405	0.416	**0.404**	**0.415**	0.415	0.425	**0.414**	**0.424**	0.436	0.429	**0.431**	**0.425**	0.442
	336	**0.422**	**0.430**	0.423	**0.430**	0.445	**0.440**	**0.438**	0.443	0.422	0.440	**0.417**	**0.429**	0.491	0.469	**0.485**	**0.461**	0.908
	720	0.463	0.472	**0.462**	**0.470**	0.489	**0.488**	**0.479**	0.497	0.449	**0.468**	**0.447**	0.469	0.521	0.500	**0.520**	**0.493**	0.333
ETTm1	96	0.285	0.339	**0.283**	**0.337**	0.299	**0.343**	**0.298**	**0.343**	0.294	0.351	**0.289**	**0.338**	0.338	0.375	**0.335**	**0.371**	1.123
	192	0.339	**0.365**	**0.336**	0.368	0.335	**0.365**	**0.334**	**0.365**	0.334	0.370	**0.333**	**0.363**	0.374	0.387	**0.373**	**0.383**	0.482
	336	0.361	0.406	**0.359**	**0.393**	0.370	0.386	**0.365**	**0.385**	0.373	0.397	**0.370**	**0.392**	**0.410**	**0.411**	0.412	0.416	0.716
	720	0.445	0.470	**0.424**	**0.421**	0.427	0.423	**0.424**	**0.417**	**0.416**	**0.420**	0.419	0.430	0.478	0.450	**0.477**	**0.448**	1.852
ETTh2	96	0.284	0.343	**0.278**	**0.338**	0.289	0.353	**0.285**	**0.348**	0.278	0.340	**0.274**	**0.336**	0.340	0.374	**0.336**	**0.371**	1.371
	192	0.339	**0.385**	**0.325**	0.393	0.384	0.418	**0.376**	**0.413**	0.341	0.382	**0.339**	**0.355**	0.402	0.414	**0.400**	**0.410**	1.806
	336	**0.361**	0.406	**0.361**	**0.399**	0.442	0.459	**0.438**	**0.455**	0.329	0.384	**0.327**	**0.383**	0.452	0.452	**0.449**	**0.445**	0.823
	720	0.445	0.470	**0.438**	**0.464**	0.601	0.549	**0.499**	**0.496**	0.381	0.424	**0.378**	**0.415**	0.462	0.468	**0.457**	**0.461**	4.370
ETTm2	96	0.171	**0.260**	**0.167**	**0.260**	0.167	0.260	**0.166**	**0.258**	0.174	0.261	**0.168**	**0.256**	**0.187**	**0.267**	0.189	0.270	0.860
	192	0.221	**0.296**	**0.220**	**0.296**	0.284	0.352	**0.243**	**0.323**	0.238	0.307	**0.231**	**0.300**	**0.249**	**0.309**	0.250	0.310	3.453
	336	**0.276**	**0.329**	0.277	0.330	0.369	0.427	**0.295**	**0.358**	0.293	0.346	**0.275**	**0.331**	0.321	0.351	**0.318**	**0.347**	6.012
	720	0.420	0.422	**0.369**	**0.391**	0.554	0.522	**0.451**	**0.456**	**0.373**	0.401	0.374	**0.400**	0.408	0.403	**0.394**	**0.391**	7.139
Exchange	96	0.089	0.209	**0.085**	**0.206**	0.088	0.215	**0.085**	**0.214**	0.094	0.216	**0.088**	**0.208**	0.107	0.234	**0.105**	**0.231**	2.880
	192	0.195	0.315	**0.177**	**0.300**	0.178	0.317	**0.171**	**0.306**	0.191	0.311	**0.185**	**0.309**	0.226	0.344	**0.224**	**0.340**	3.403
	336	0.343	0.421	**0.312**	**0.405**	0.371	0.462	**0.300**	**0.412**	0.343	0.427	**0.342**	**0.423**	0.367	0.448	**0.361**	**0.442**	5.875
	720	0.898	0.710	**0.847**	**0.697**	0.966	0.754	**0.811**	**0.683**	0.888	0.706	**0.813**	**0.673**	0.964	0.746	**0.957**	**0.739**	5.970
ILI	24	**1.914**	0.879	1.938	**0.874**	2.215	1.081	**1.935**	**0.935**	1.593	0.757	**1.561**	**0.750**	2.317	**0.934**	**2.139**	0.936	4.483
	36	1.808	0.858	**1.800**	**0.851**	2.142	0.977	**1.938**	**0.942**	1.768	0.794	**1.706**	**0.780**	1.972	0.920	**1.968**	**0.914**	2.561
	48	1.797	0.873	**1.796**	**0.867**	2.335	1.056	**2.221**	**1.030**	1.799	0.916	**1.774**	**0.892**	2.238	0.940	**2.229**	**0.937**	1.602
	60	1.859	0.895	**1.810**	**0.876**	2.479	**1.088**	**2.382**	1.096	1.850	0.943	**1.735**	**0.880**	**2.027**	**0.928**	2.041	0.930	2.491
Weather	96	0.149	0.198	**0.147**	**0.194**	0.192	0.250	**0.187**	**0.245**	0.149	0.198	**0.147**	**0.197**	0.172	0.220	**0.169**	**0.215**	1.729
	192	0.201	0.248	**0.192**	**0.242**	0.248	0.297	**0.240**	**0.285**	0.194	0.241	**0.191**	**0.238**	0.219	0.261	**0.215**	**0.257**	2.539
	336	0.264	0.291	**0.244**	**0.281**	0.284	0.335	**0.274**	**0.324**	**0.244**	**0.282**	0.245	0.285	0.280	0.306	**0.274**	**0.291**	2.924
	720	0.320	0.336	**0.318**	**0.334**	0.339	0.374	**0.320**	**0.357**	0.320	0.334	**0.316**	**0.333**	**0.365**	**0.359**	0.366	0.362	1.476
Electricity	96	0.142	0.237	**0.139**	**0.235**	0.153	**0.239**	**0.142**	0.247	0.138	0.233	**0.136**	**0.231**	0.168	0.272	**0.158**	**0.259**	2.480
	192	0.154	0.248	**0.147**	**0.246**	0.158	0.251	**0.152**	**0.248**	**0.153**	**0.247**	**0.153**	0.248	0.184	0.289	**0.172**	**0.262**	3.226
	336	0.163	0.264	**0.161**	**0.262**	0.170	0.269	**0.168**	**0.267**	0.170	0.263	**0.168**	**0.262**	0.198	0.300	**0.181**	**0.284**	2.423
	720	0.208	0.300	**0.204**	**0.299**	0.233	0.342	**0.230**	**0.338**	**0.206**	**0.296**	0.210	0.301	0.220	0.320	**0.205**	**0.309**	1.417
Traffic	96	0.376	0.264	**0.375**	**0.262**	**0.411**	0.284	**0.411**	**0.282**	0.360	0.249	**0.357**	**0.246**	0.593	0.321	**0.554**	**0.316**	1.488
	192	0.397	**0.277**	**0.340**	0.279	0.423	0.287	**0.422**	**0.286**	**0.379**	0.256	**0.379**	**0.254**	0.617	0.336	**0.562**	**0.331**	3.175
	336	0.413	0.290	**0.411**	**0.289**	0.438	0.299	**0.436**	**0.297**	0.401	0.270	**0.389**	**0.255**	0.629	**0.336**	**0.579**	0.341	2.120
	720	0.444	0.306	**0.441**	**0.302**	**0.467**	**0.316**	0.471	0.318	0.443	0.294	**0.430**	**0.281**	0.640	**0.350**	**0.587**	0.366	[illegible]

(4) Experiments

b) Comparison with other methods

- Predict Residuals with Regularization (PRReg).
    - Designed to address the non-robustness of the CD strategy.
    - Just another plug-and-play method…

      Not related to channel identification

		CD	CI	+PRReg	+CCM*
ETTh1	Linear	0.402	0.345	**0.342**	**0.342**
	Transformer	0.861	0.655	0.539	**0.518**
ETTm1	Linear	0.404	0.354	0.311	**0.310**
	Transformer	0.458	0.379	0.349	**0.300**
Weather	Linear	0.142	0.169	0.131	**0.130**
	Transformer	0.251	0.168	0.180	**0.164**
ILI	Linear	2.343	2.847	2.299	**2.279**
	Transformer	5.309	4.307	3.254	**3.206**
Electricity	Linear	**0.195**	0.196	0.196	**0.195**
	Transformer	0.250	0.185	0.185	**0.183**

3. From Similarity to Superiority: Channel Clustering for Time Series Forecasting

(4) Experiments

c) Ablation study on # of clusters

Cluster ratio 0.0 => C clusters

Cluster ratio 1.0 => 1 cluster

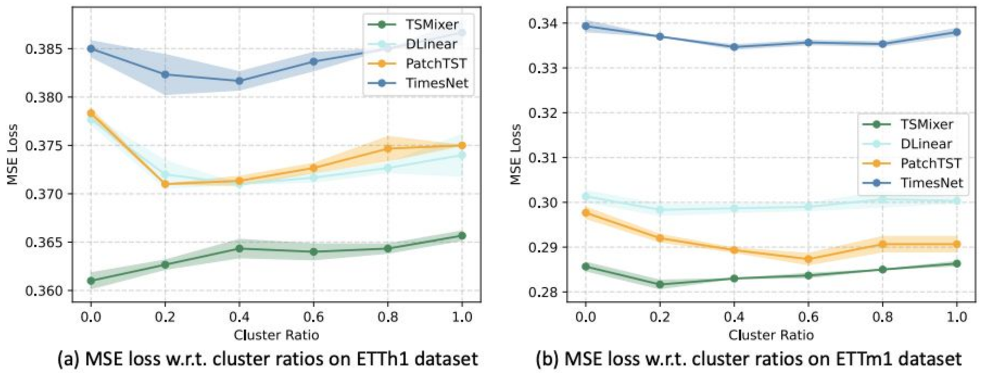


Figure 4. Ablation Study on Cluster Ratios in terms of MSE loss with four base models. The forecasting horizon is 96. (*left*: ETTh1 dataset; *right*: ETTm1 dataset)

4. Conclusion

4. Conclusion

- Capturing channel identity is crucial for TS forecasting
- Desiderata
  - (1) Simple & Effective
  - (2) Plug-and-play method for any TS backbone models
  - (3) Capture/Interpret the relationships btw the channels (i.e., clustering)
- ~ing: Channel identification with ***channel-wise layer normalization***

Thank you!