SDformer: Transformer with Spectral Filter and Dynamic Attention for Multivariate Time Series Long-term Forecasting

Ziyu Zhou, Gengyu Lyu#, Yiming Huang, Zihao Wang, Ziyu Jia and Zhen Yang

Beijing University of Technology

School of Computer Science, Beijing University of Technology, Beijing, China

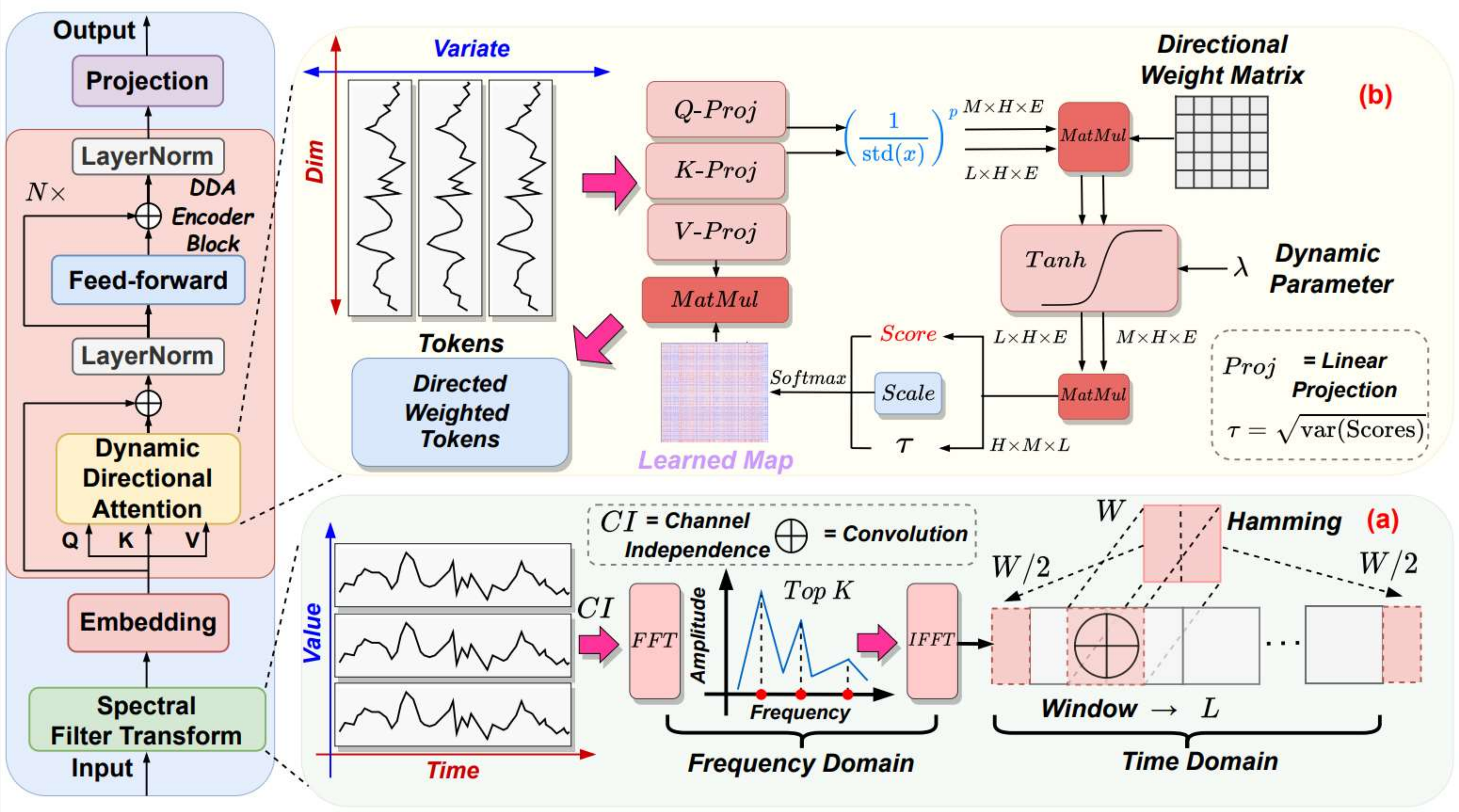
Institute of Automation, Chinese Academy of Sciences, Beijing, China

IJCAI JEJU 2024

Introduction

- Time series are **widely present** in everyday life: Traffic Flow, Weather Variations, Economic Changes, etc.
- **Transformers** have revolutionized time series modeling: iTransformer, Pathformer, Pyraformer, Autoformer, Informer, etc.

Methods



$$\mathbf{X}f_{:,n} = \text{Spectral-Filter-Transform}(\mathbf{X}_{:,n}),$$
$$\mathbf{H}_n^0 = \text{Invert Embedding}(\mathbf{X}f_{:,n}),$$
$$\mathbf{H}^{l+1} = \text{DDAEncoder}(\mathbf{H}^l), \quad l = 0, \dots, L-1,$$
$$\hat{\mathbf{Y}}_{:,n} = \text{Projection}(\mathbf{H}_n^L),$$

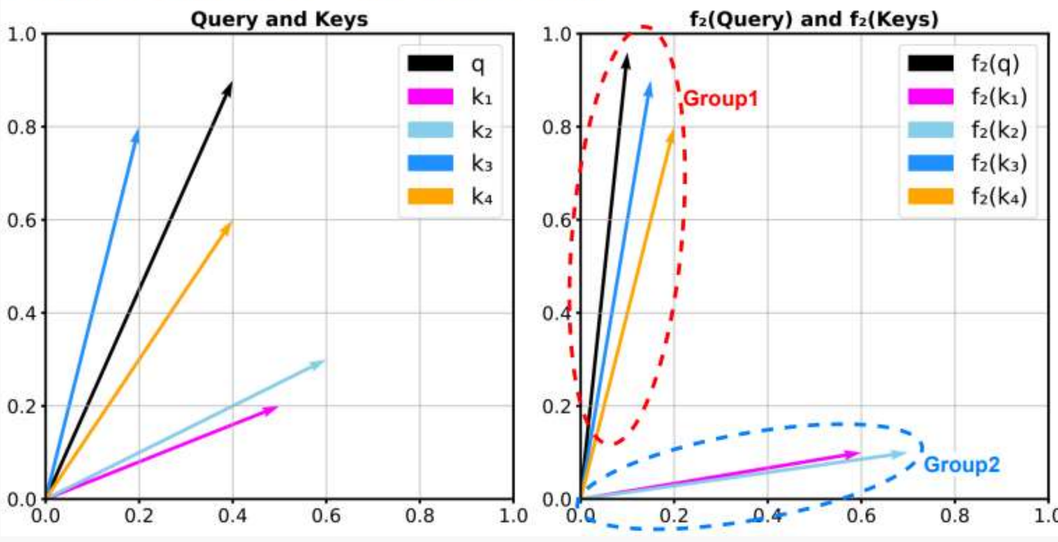
Spectral-Filter-Transform: Frequency Domain Denoising & Time Domain Smoothing

Dynamic-Directional-Attention: Introduce a Kernel Function

Algorithm 1 The Spectral-Filter-Transform module

```
Input: Time series X ∈ R^{T×N}, Length T, Variates N
Output: Denoised and smoothed series X_h ∈ R^{T×N}
Initialization: A Hamming Window w_n sized w; the number of top frequency components k.
for n = 1 to N do
    X_{f_n} = FFT(x_n)                {Fast Fourier Transform}
    X_{f_{k_n}} = TopK(X_{f_n}, k)     {Select k frequencies}
    x_{if_n} = IFFT(X_{f_{k_n}})       {Convert to the time domain}
    x_{p_n} = Reflective Padding(x_{if_n}, w_n)
    x_{h_n} = Applying Window(x_{p_n}, w_n)
end for
X_h = Concat(x_{h_n})   {Concatenate N univariate series}
Return: X_h
```

Analysis on the Kernel Function



Enclose similar Q-K pair, Distance dissimilar pair

Long-term Forecasting Results

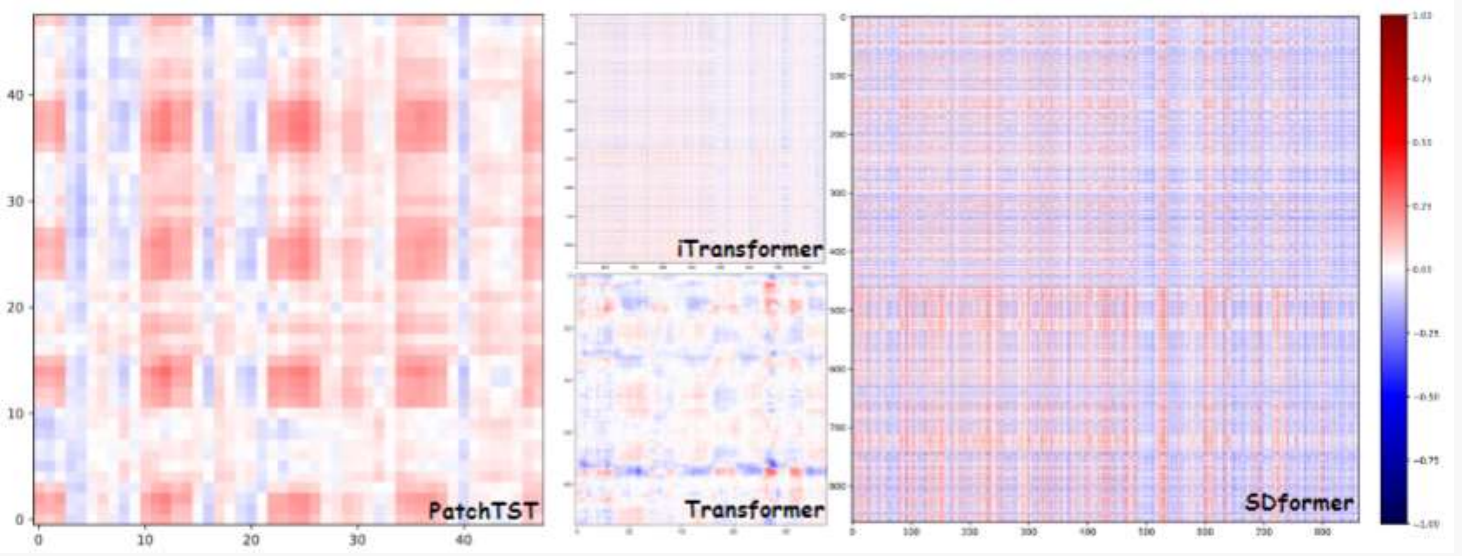
Methods		Ours		iTransformer		DLinear		PatchTST		TimesNet		FEDformer		Autoformer		Stationary		Crossformer		TiDE		LightTS		Informer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.183	0.268	0.184	0.268	0.193	0.292	0.177	0.261	0.187	0.267	0.203	0.287	0.255	0.339	0.192	0.274	0.287	0.366	0.207	0.305	0.308	0.365	0.453	0.596
	192	0.249	0.309	0.253	0.313	0.284	0.362	0.244	0.303	0.249	0.309	0.269	0.328	0.281	0.340	0.280	0.339	0.414	0.492	0.290	0.364	0.311	0.382	0.311	0.382
	336	0.313	0.348	0.312	0.350	0.369	0.427	0.309	0.346	0.321	0.351	0.325	0.366	0.339	0.372	0.334	0.361	0.597	0.542	0.377	0.422	0.442	0.466	1.363	0.887
	720	0.407	0.402	0.412	0.406	0.554	0.522	0.400	0.398	0.408	0.403	0.421	0.415	0.433	0.432	0.417	0.413	1.730	1.042	0.558	0.524	0.675	0.587	3.379	1.338
	Avg	0.288	0.332	0.291	0.334	0.350	0.401	0.255	0.327	0.291	0.333	0.305	0.349	0.327	0.371	0.306	0.347	0.757	0.610	0.358	0.404	0.409	0.436	1.410	0.810
ETTh2	96	0.298	0.345	0.299	0.350	0.333	0.387	0.295	0.344	0.340	0.374	0.358	0.397	0.346	0.388	0.476	0.458	0.745	0.584	0.400	0.440	0.397	0.437	3.755	1.525
	192	0.378	0.394	0.381	0.400	0.477	0.476	0.367	0.391	0.402	0.414	0.429	0.439	0.456	0.452	0.512	0.493	0.877	0.656	0.528	0.509	0.520	0.504	5.602	1.931
	336	0.419	0.427	0.424	0.433	0.549	0.541	0.434	0.443	0.452	0.452	0.496	0.487	0.482	0.486	0.552	0.551	1.041	0.731	0.377	0.422	0.626	0.559	4.721	1.835
	720	0.418	0.437	0.430	0.446	0.831	0.657	0.423	0.445	0.462	0.468	0.463	0.474	0.515	0.511	0.562	0.560	1.104	0.763	0.874	0.679	0.863	0.672	3.647	1.625
	Avg	0.378	0.401	0.384	0.407	0.559	0.515	0.380	0.406	0.414	0.427	0.437	0.449	0.450	0.459	0.526	0.516	0.942	0.684	0.611	0.550	0.543	0.548	4.431	1.729
Weather	96	0.171	0.210	0.176	0.216	0.196	0.255	0.177	0.218	0.172	0.220	0.217	0.296	0.266	0.336	0.173	0.223	0.158	0.230	0.202	0.261	0.182	0.242	0.300	0.384
	192	0.222	0.255	0.225	0.257	0.237	0.296	0.225	0.259	0.219	0.261	0.276	0.336	0.307	0.367	0.245	0.285	0.206	0.277	0.242	0.298	0.227	0.278	0.598	0.544
	336	0.278	0.297	0.281	0.299	0.283	0.335	0.279	0.297	0.280	0.306	0.339	0.380	0.359	0.395	0.321	0.338	0.272	0.335	0.287	0.335	0.282	0.334	0.578	0.523
	720	0.358	0.348	0.360	0.352	0.345	0.381	0.447	0.466	0.365	0.359	0.403	0.428	0.419	0.428	0.414	0.410	0.398	0.418	0.351	0.386	0.352	0.386	1.059	0.741
	Avg	0.258	0.278	0.261	0.281	0.265	0.317	0.354	0.348	0.259	0.287	0.309	0.360	0.338	0.382	0.288	0.314	0.259	0.315	0.271	0.320	0.261	0.312	0.634	0.548
ECL	96	0.150	0.243	0.149	0.240	0.197	0.282	0.181	0.271	0.168	0.272	0.193	0.308	0.201	0.317	0.169	0.273	0.219	0.314	0.237	0.329	0.207	0.307	0.274	0.368
	192	0.164	0.258	0.165	0.257	0.196	0.285	0.187	0.276	0.184	0.289	0.201	0.315	0.222	0.334	0.182	0.286	0.231	0.322	0.236	0.330	0.213	0.316	0.296	0.386
	336	0.180	0.274	0.178	0.271	0.209	0.301	0.203	0.292	0.198	0.300	0.214	0.329	0.231	0.338	0.200	0.304	0.246	0.337	0.249	0.344	0.230	0.333	0.300	0.394
	720	0.211	0.302	0.228	0.312	0.245	0.333	0.245	0.325	0.220	0.320	0.246	0.355	0.254	0.361	0.222	0.321	0.280	0.363	0.284	0.373	0.265	0.360	0.373	0.439
	Avg	0.176	0.269	0.180	0.261	0.212	0.300	0.204	0.291	0.192	0.295	0.214	0.327	0.227	0.338	0.193	0.296	0.244	0.334	0.251	0.344	0.229	0.329	0.311	0.397
Exchange	96	0.087	0.208	0.087	0.207	0.088	0.218	0.089	0.206	0.107	0.234	0.148	0.278	0.197	0.323	0.111	0.237	0.256	0.367	0.094	0.218	0.116	0.262	0.847	0.752
	192	0.177	0.300	0.181	0.304	0.176	0.315	0.186	0.307	0.226	0.344	0.271	0.315	0.300	0.369	0.219	0.335	0.470	0.509	0.184	0.307	0.215	0.359	1.204	0.895
	336	0.331	0.418	0.338	0.422	0.313	0.427	0.310	0.403	0.367	0.448	0.460	0.427	0.509	0.524	0.421	0.476	1.268	0.883	0.349	0.431	0.377	0.466	1.672	1.036
	720	0.829	0.688	0.853	0.696	0.839	0.695	0.864	0.701	0.964	0.746	1.195	0.695	1.447	0.941	1.092	0.769	1.767	1.068	0.852	0.698	0.831	0.699	1.941	1.127
	Avg	0.356	0.404	0.365	0.407	0.354	0.414	0.362	0.404	0.416	0.443	0.519	0.429	0.613	0.539	0.461	0.454	0.940	0.707	0.370	0.413	0.385	0.447	1.550	0.998
Traffic	96	0.377	0.262	0.393	0.268	0.650	0.396	0.460	0.295	0.593	0.321	0.587	0.366	0.613	0.388	0.612	0.338	0.522	0.290	0.805	0.493	0.615	0.391	0.719	0.391
	192	0.396	0.272	0.413	0.277	0.598	0.370	0.464	0.296	0.617	0.336	0.604	0.373	0.616	0.382	0.613	0.340	0.530	0.293	0.756	0.474	0.601	0.382	0.696	0.379
	336	0.413	0.281	0.424	0.283	0.605	0.373	0.480	0.303	0.629	0.336	0.621	0.383	0.622	0.337	0.618	0.328	0.558	0.305	0.762	0.477	0.613	0.386	0.777	0.420
	720	0.447	0.295	0.460	0.301	0.645	0.394	0.514	0.322	0.640	0.350	0.626	0.382	0.660	0.408	0.653	0.355	0.589	0.328	0.719	0.449	0.658	0.407	0.864	0.472
	Avg	0.408	0.278	0.423	0.282	0.625	0.383	0.480	0.304	0.620	0.336	0.610	0.376	0.628	0.379	0.624	0.340	0.550	0.304	0.760	0.473	0.622	0.392	0.767	0.416
ILI	24	2.079	0.900	2.567	0.949	4.830	1.167	1.319	0.754	2.317	0.934	3.228	1.260	3.483	1.287	2.294	0.945	2.527	1.020	8.313	2.144	8.313	2.144	5.764	1.677
	36	2.219	0.933	2.082	0.919	4.454	1.563	1.430	0.834	1.972	0.920	2.679	1.080	3.103	1.148	1.825	0.848	2.615	1.007	6.631	1.902	6.631	1.902	4.755	1.467
	48	1.986	0.881	2.146	0.921	4.099	1.476	1.553	0.815	2.238	0.940	2.622	1.078	2.669	1.085	2.010	0.900	2.359	0.972	7.299	1.982	7.299	1.982	4.763	1.469
	60	2.031	0.915	2.052	0.931	4.207	1.482	1.470	0.788	2.027	0.928	2.857	1.157	2.770	1.125	2.178	0.963	2.487	1.016	7.283	1.985	7.283	1.985	5.264	1.564
	Avg	2.066	0.915	2.212	0.930	4.398	1.422	1.443	0.797	2.139	0.931	2.847	1.144	3.006	1.161	2.077	0.914	2.497	1.004	7.382	2.003	7.382	2.003	5.137	1.544

- 11 baselines for comparison: iTransformer, PatchTST, etc.
- SDformer achieved SOTA performance on 5 datasets
- The number of variates in the MTS ranges from 7 to 862, with error reductions of 15% and 17% in MSE, respectively

Motivation

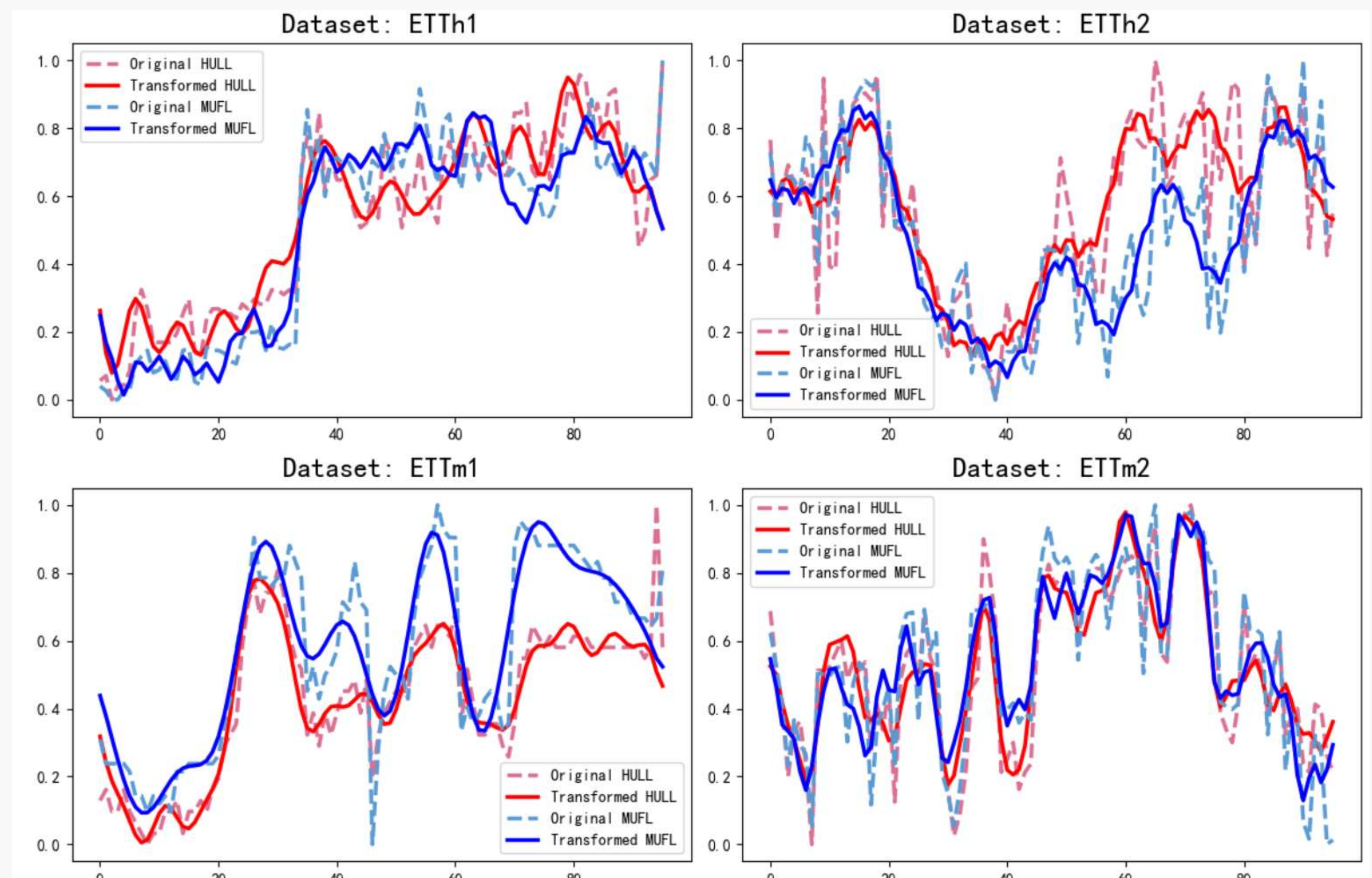
Previous Work failed to simultaneously solve:

- Noise in time series data (Meaningless patterns)
- Smooth Attention Distribution (Row-homogenization in attention maps)



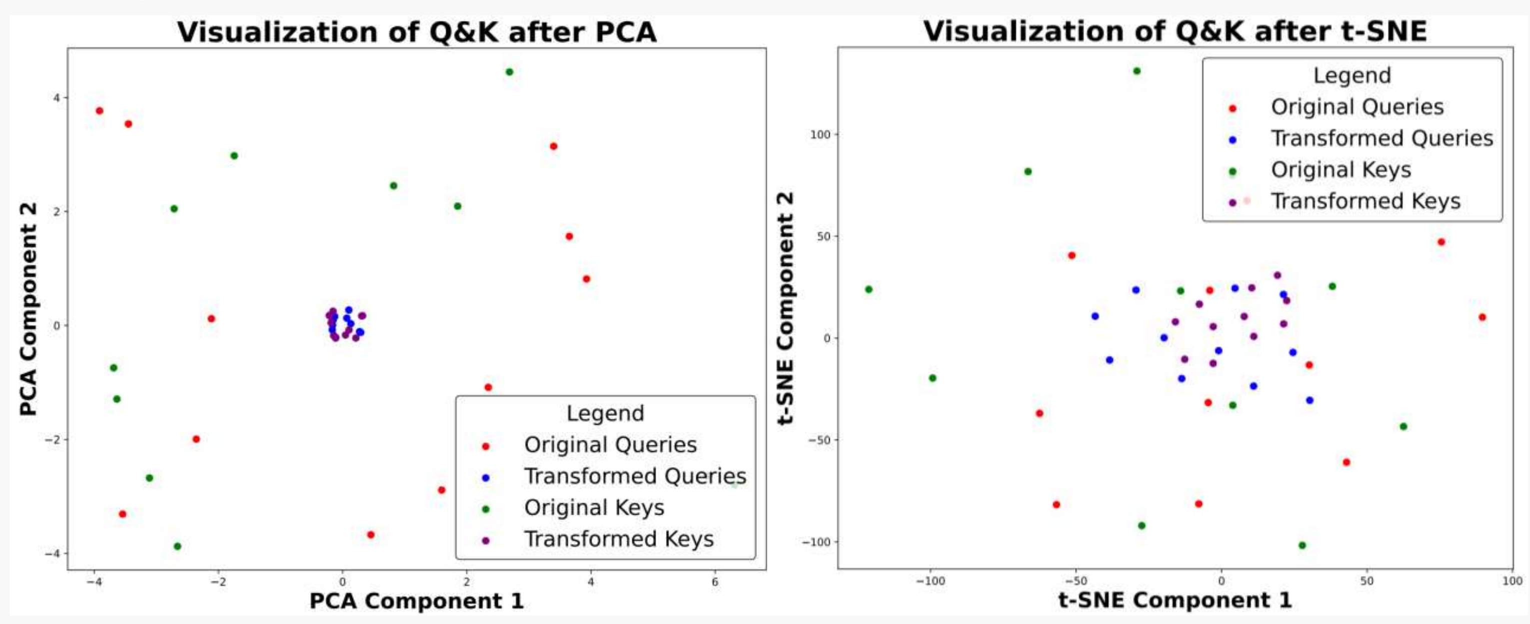
Analysis and Discussion

Spectral Filter Transform

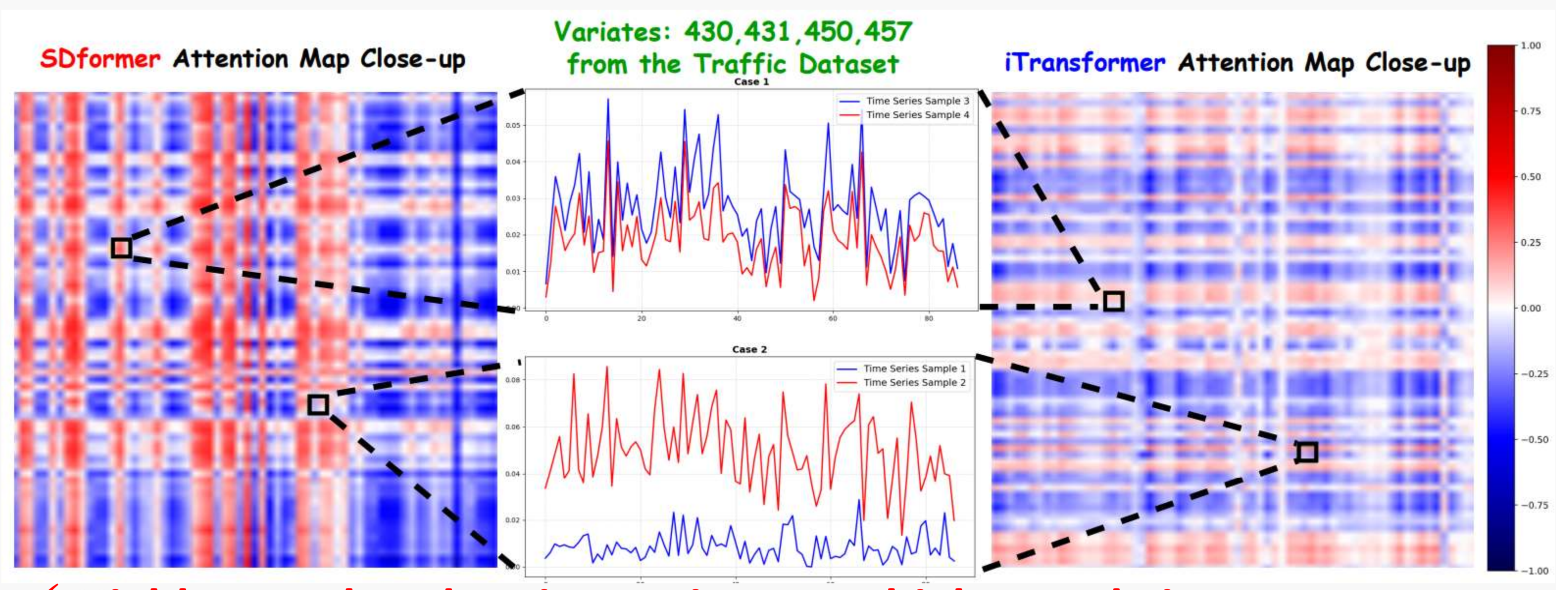


✓ Analogous to a single-layer MLP: Marco Patterns(Trend, Periodicity)

Dynamic Directional Attention

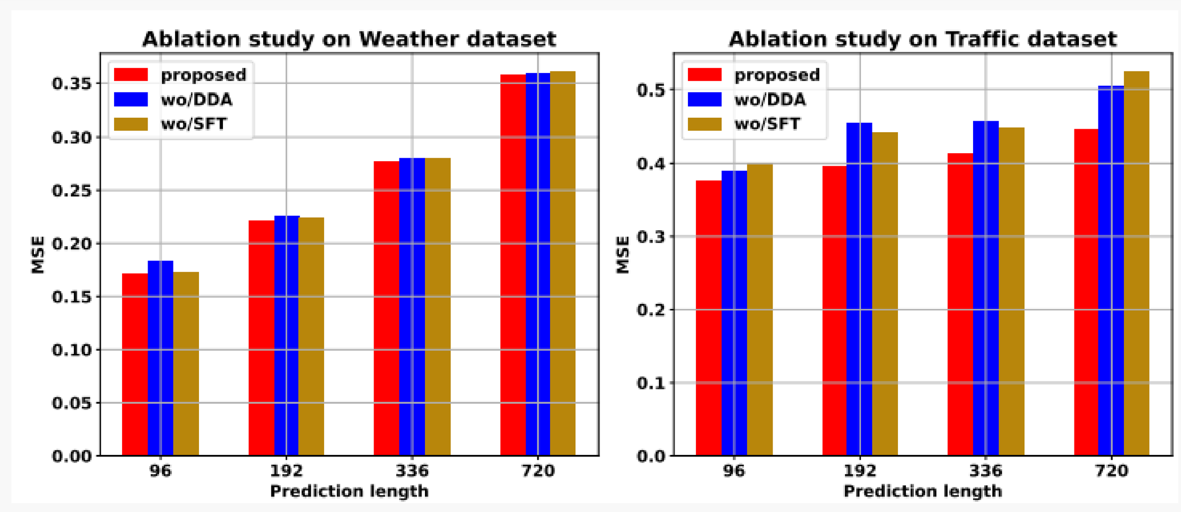


Jointly Analysis for better Interpretability



✓ Highly correlated variate pair scores higher and vice versa

Further Analysis



Method	Parameter	Time	MACs	FLOPs
SDformer	6.41M	165ms	44.18G	88.38G
iTransformer	6.41M	107ms	44.19G	88.38G
PatchTST	6.91M	353ms	526.89G	1053.80G
Crossformer	54.46M	492ms	1597.86G	3195.72G
TimesNet	304.25M	1039ms	3122.55G	6245.02G

Links

Wechat

My Profile

I will join the Time Series Group of the CityMind Lab at HKUST(GZ) in August, focusing on time series self-supervised learning and multi-modality deep learning.