Enhancing Time Series Forecasting through Selective Representation Spaces: A Patch Perspective (NIPS 25 Spotlight)

Reporter: Xingjian Wu
A PHD from ECNU

華東師範大學
EAST CHINA NORMAL UNIVERSITY

SCHOOL OF DATA SCIENCE & ENGINEERING
数据科学与工程学院

Time Series Forecasting

EAST CHINA NORMAL UNIVERSITY

- Time series records temporal data from sensors in real-world cyber-physical systems.

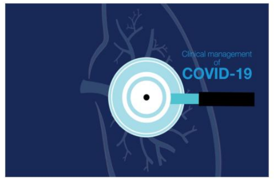
Multivariate Time Series Forecasting aims to predict the next F future timestamps, formulated as $Y = < X_{:,T+1}, \cdots, X_{:,T+F} > \in \mathbb{R}^{N\times F}$ based on the historical time series $X = < X_{:,1}, \cdots, X_{:,T} > \in \mathbb{R}^{N\times T}$ with N channels and T timestamps.

Energy Consumption

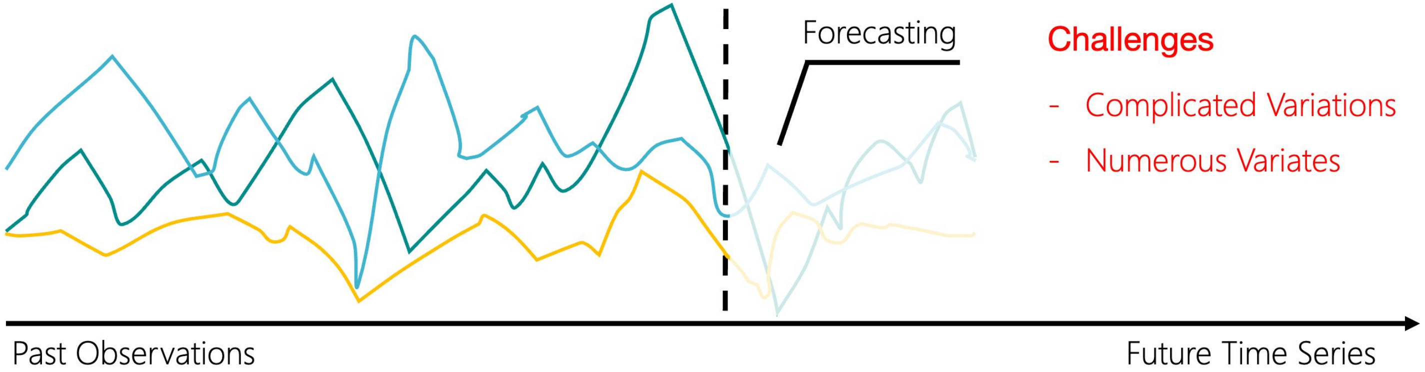
Traffic Flow

Economic Changes

Weather Variations

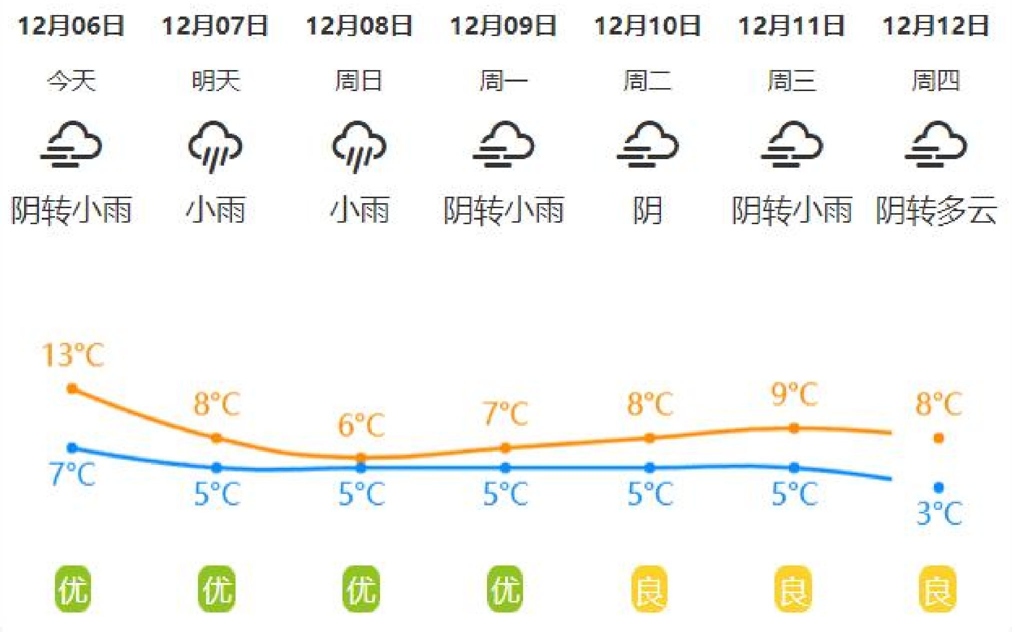


Disease Estimations



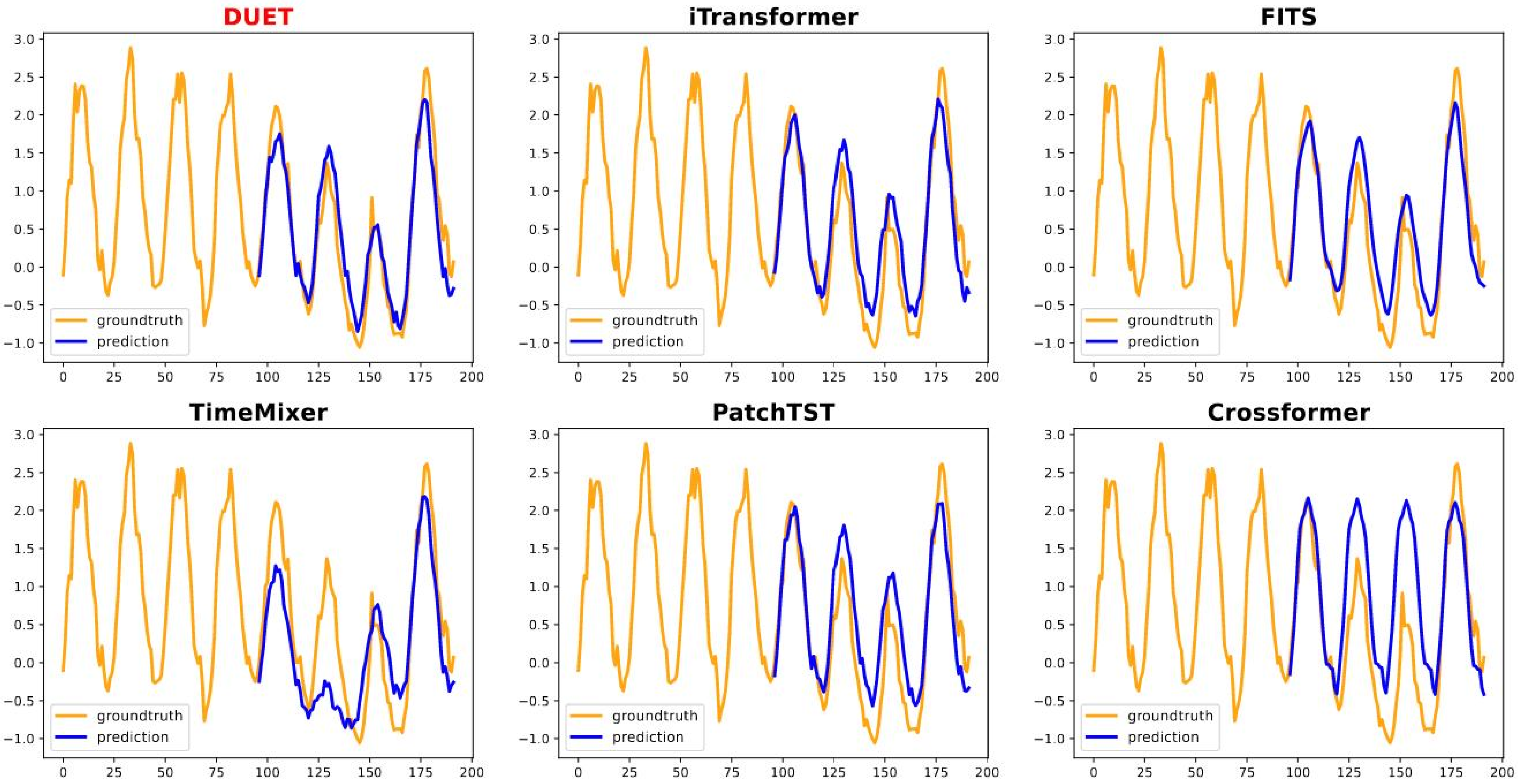
Challenges

- Complicated Variations
- Numerous Variates

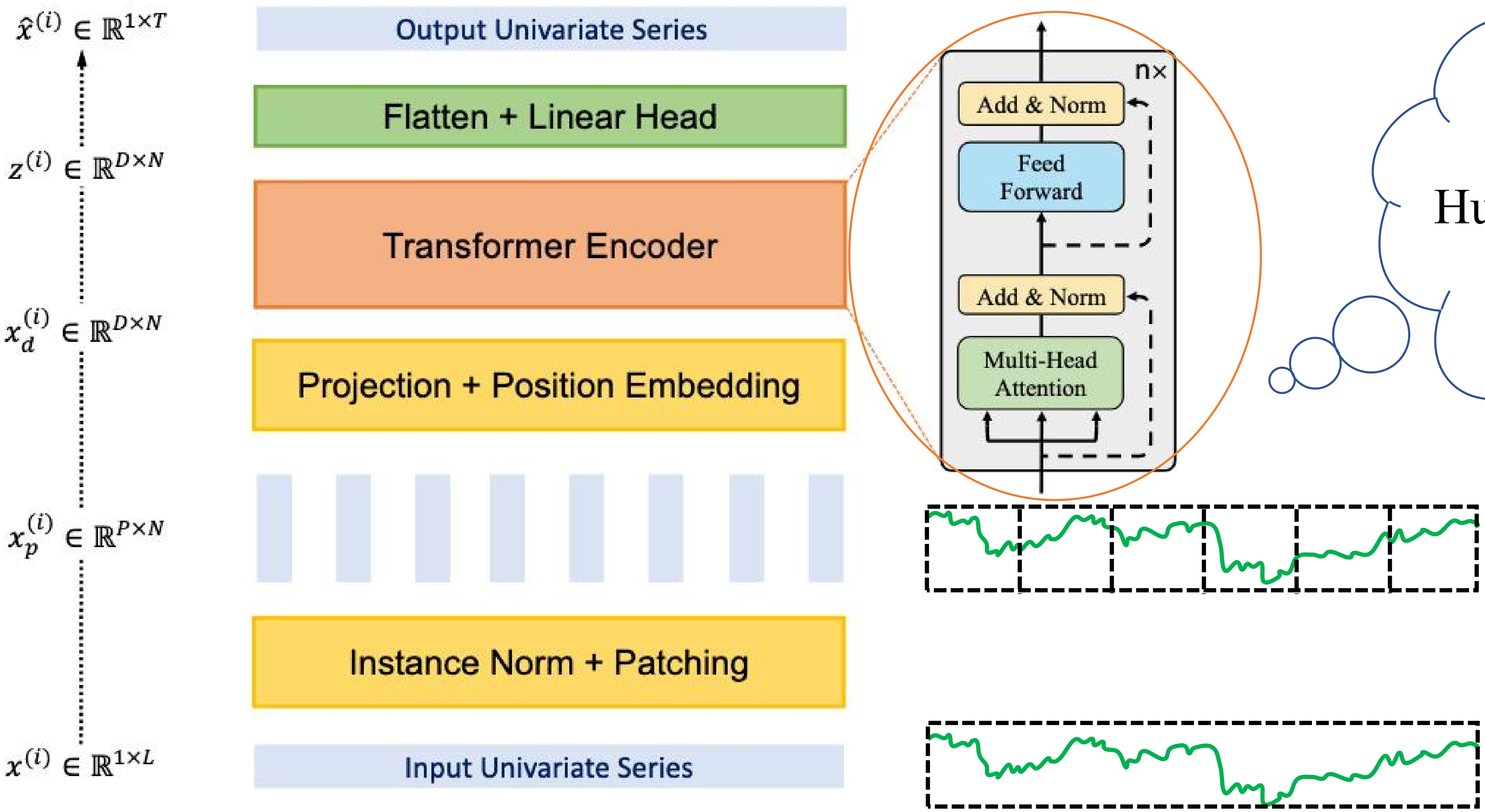


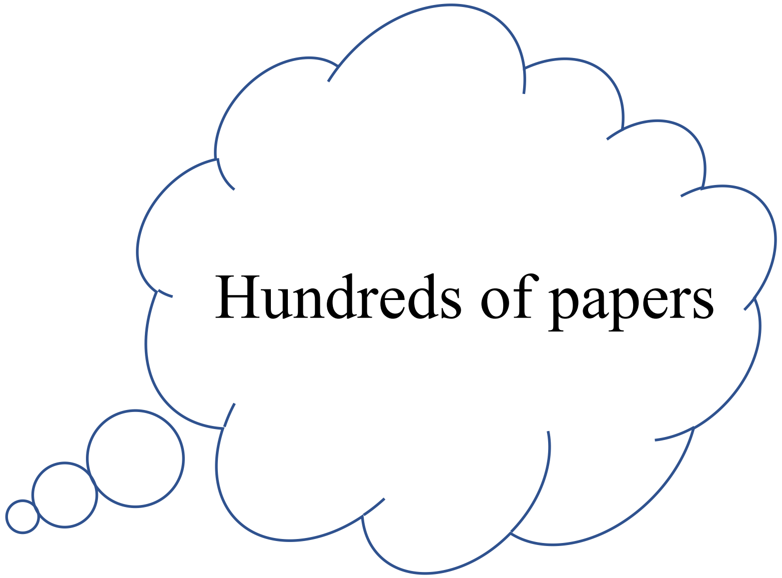
- Differences from classical regression tasks:
  - Somewhat self-supervised (like NLP tasks).
- The underlying principes:
  - Learn to evolve key patterns within past observations.

Forecasting is the evolution of proper patterns

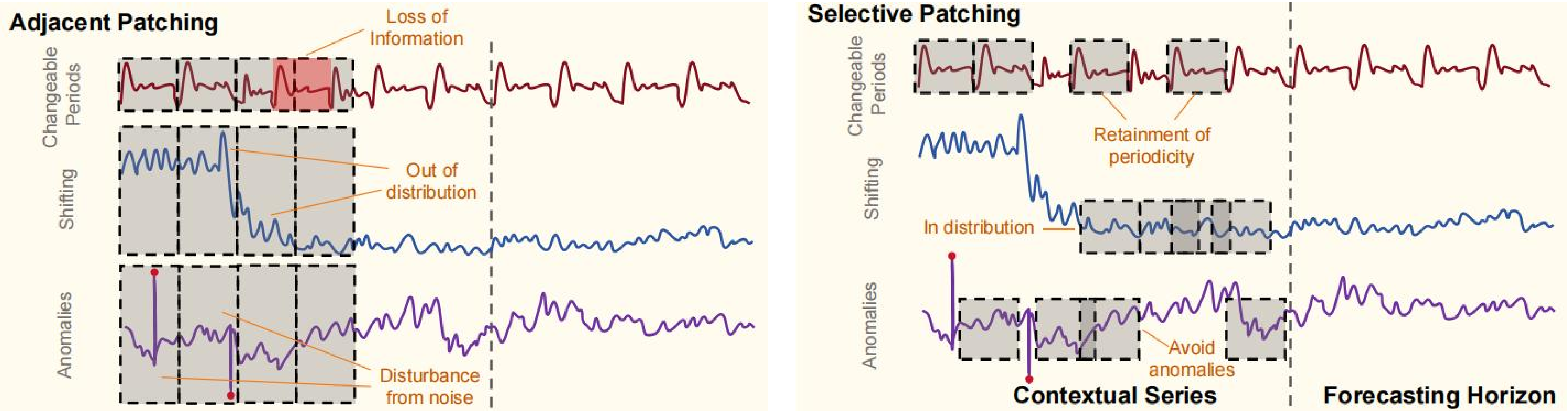


Patch-based Forecasting Models



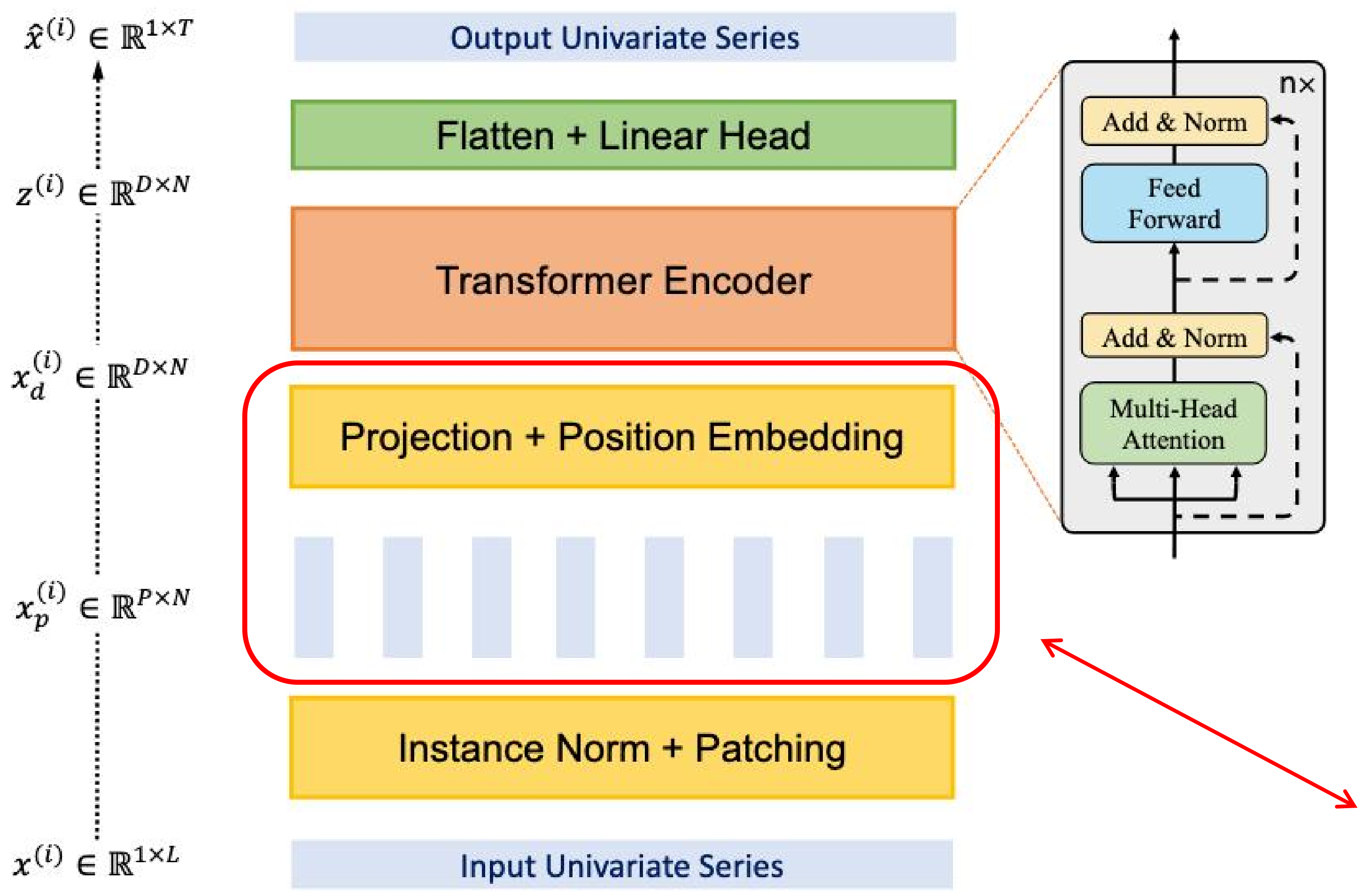


Motivation: Directly Evolve or Select then Evolve?



- Adjacent vs. Selective Patching (both using 4 patches)
  - Adjacent patching partitions time series into adjacent patches, lacking flexibility. Selective patching automatically select most relevant sub-series as patches.
  - The upper part shows examples that conventional adjacent patching may include harmful information, thus hindering the forecasting performance.
  - The lower part shows the selected patches that are more relevant for making the corresponding forecasting, demonstrating the flexibility that selective patches offer.

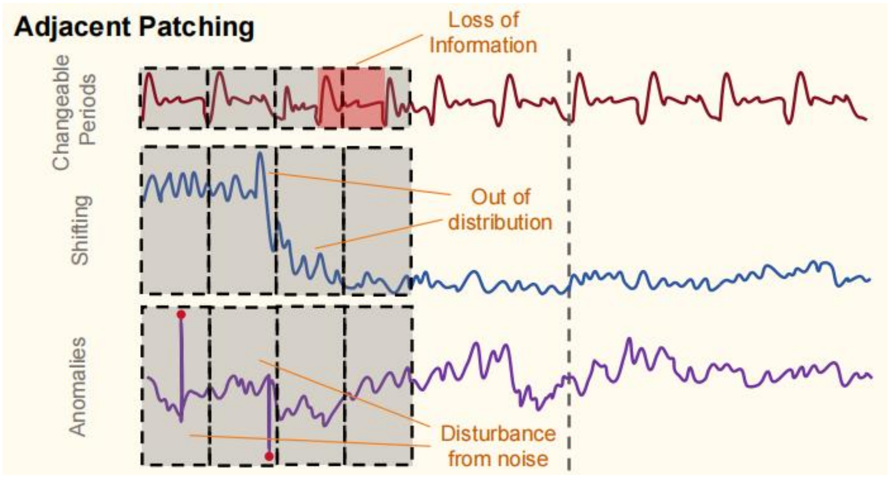
Representation Spaces



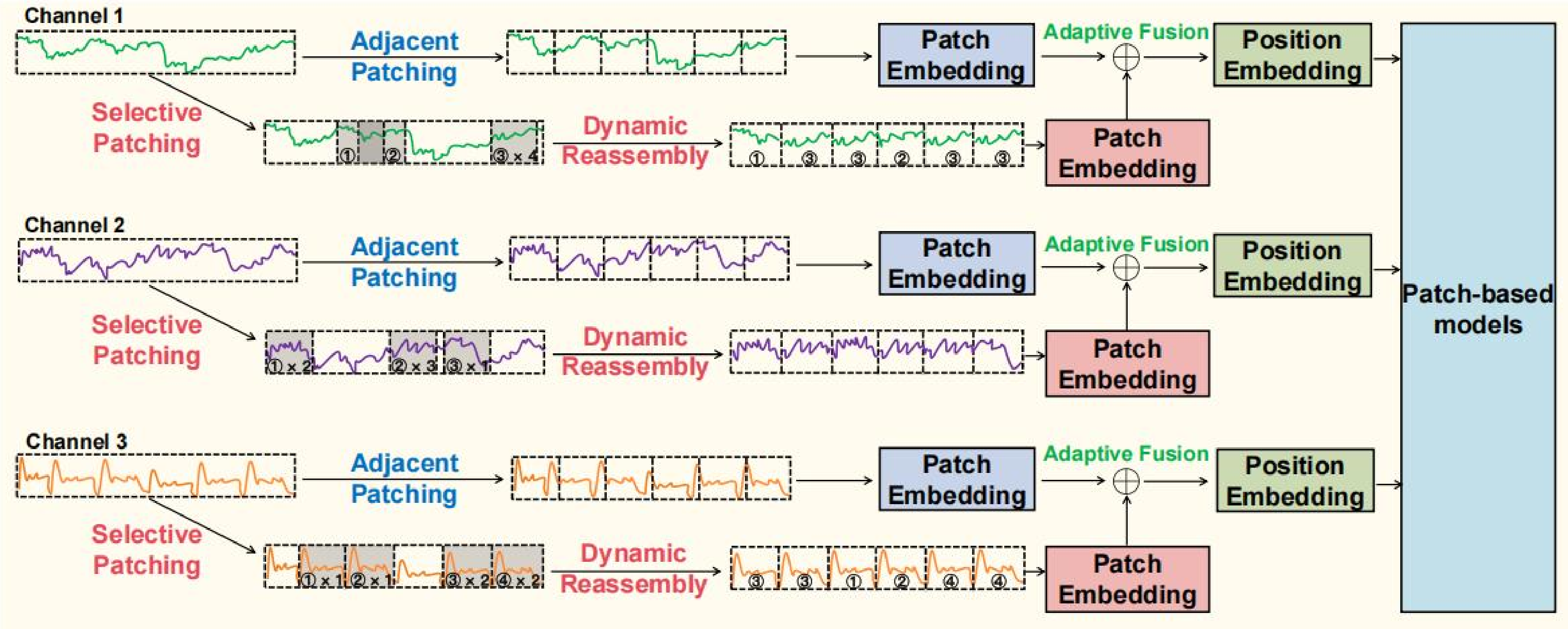
How many representation spaces?

Each configuration of patch-size and stride contributes to a representation space.

What about changeable strides or just random sampling?



Overview: Selective Representation Space (SRS)



- Selective Patching

$$C_{K+n-1}^{n}$$

- Dynamic Reassembly

$$n!$$

- The Total Size of Search Space:

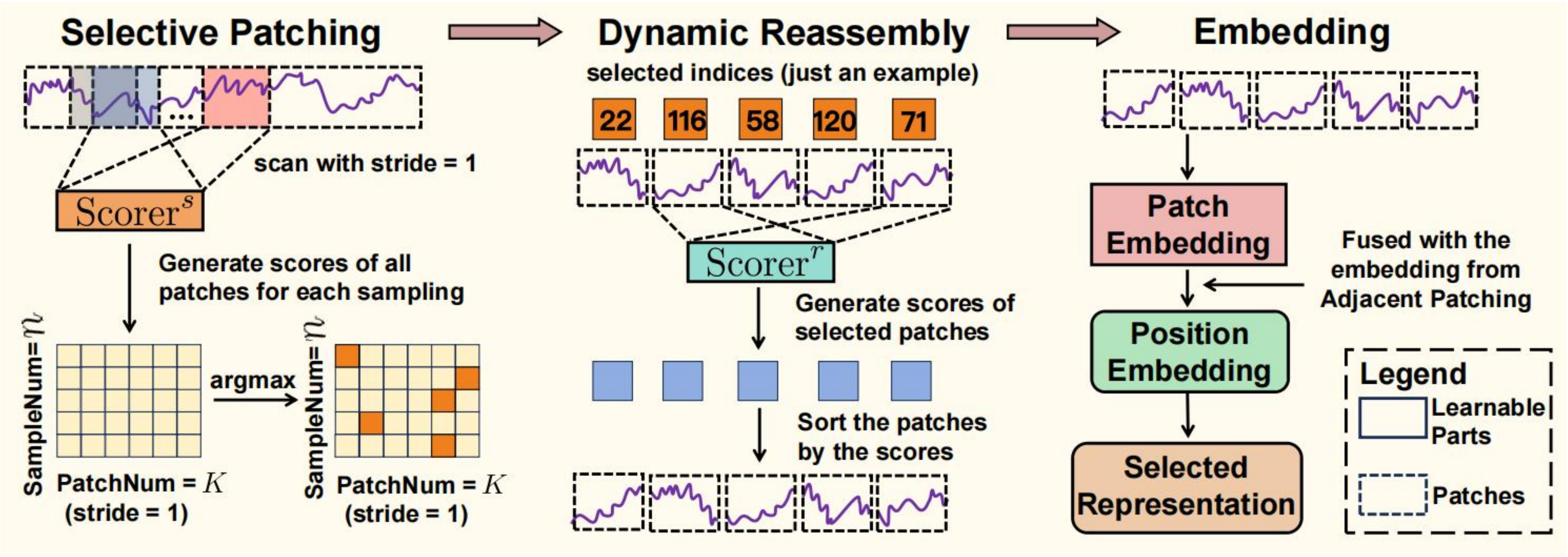
$$C_{K+n-1}^{n} \cdot n!$$

- Given input length T, patch size p, stride s.
- There exists $n = \lceil (T-p)/s \rceil + 1$ patches using adjacent patching.
- There exists $K = (n-1) \cdot s + 1$ potential patches not using adjacent patching.

A very huge search space for representation spaces!

Methodology: “Sort as Search”

华东师范大学
EAST CHINA NORMAL UNIVERSITY



Selective Patching

$$\text{Scorer}^s := \mathbb{R}^{N\times K\times p} \rightarrow \mathbb{R}^{N\times K\times n},$$
$$\mathcal{S}^s = \text{Scorer}^s(\mathcal{P}'), \mathcal{I}^s = \text{Argmax}(\mathcal{S}^s),$$
$$\mathcal{S}^s_{max} = \mathcal{S}^s[\mathcal{I}^s], \mathcal{S}^s_{inv} = \text{detach}(1/\mathcal{S}^s_{max}),$$
$$\mathcal{P}^s_{max} = \mathcal{P}'[\mathcal{I}^s], E^s = \mathcal{S}^s_{max} \odot \mathcal{S}^s_{inv},$$
$$\tilde{\mathcal{P}}^s_{\text{max}} = \mathcal{P}^s_{max} \odot E^s,$$

Dynamic Reassembly

$$\text{Scorer}^r := \mathbb{R}^{N\times n\times p} \rightarrow \mathbb{R}^{N\times n},$$
$$\mathcal{S}^r = \text{Scorer}^r(\tilde{\mathcal{P}}^s_{max}), \mathcal{I}^r = \text{Argsort}(\mathcal{S}^r),$$
$$\mathcal{S}^r_{sort} = \mathcal{S}^r[\mathcal{I}^r], \mathcal{S}^r_{inv} = \text{detach}(1/\mathcal{S}^r_{sort}),$$
$$\mathcal{P}^r_{sort} = \tilde{\mathcal{P}}^s_{max}[\mathcal{I}^r], E^r = \mathcal{S}^r_{sort} \odot \mathcal{S}^r_{inv},$$
$$\tilde{\mathcal{P}} = \mathcal{P}^r_{sort} \odot E^r,$$

C^n_{K+n-1}

- Independently selecting n times from K candidates based on some criterions.
- Reranking n selected patches based on some criterions. $n!$

- Adaptively fusing the embeddings of SRS and adjacent patching.

SRSNet: A simple yet effective baseline

華東師範大學 EAST CHINA NORMAL UNIVERSITY

Datasets	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Solar		Traffic	
Metrics	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae
FEDformer [2022]	0.433	0.454	0.406	0.438	0.567	0.519	0.335	0.380	0.312	0.356	0.219	0.330	0.641	0.628	0.620	0.382
Stationary [2022]	0.667	0.568	0.377	0.419	0.531	0.472	0.384	0.390	0.287	0.310	0.194	0.295	0.390	0.387	0.622	0.340
DLinear [2023]	0.430	0.443	0.470	0.468	0.356	0.378	0.259	0.324	0.242	0.295	0.167	0.264	0.224	0.286	0.418	0.287
TimesNet [2023]	0.468	0.459	0.390	0.417	0.408	0.415	0.292	0.331	0.255	0.282	0.190	0.284	0.211	0.281	0.617	0.327
Crossformer [2023]	0.439	0.461	0.894	0.680	0.464	0.456	0.501	0.505	0.232	0.294	0.171	0.263	0.205	**0.232**	0.522	0.282
PatchTST [2023]	0.419	0.436	0.351	0.395	0.349	0.381	0.256	0.314	0.224	**0.262**	0.171	0.270	0.200	0.284	0.397	0.275
TimeMixer [2024]	0.427	0.441	0.347	0.394	0.356	0.380	0.257	0.318	0.225	0.263	0.185	0.284	0.203	0.261	0.410	0.279
iTransformer [2024]	0.440	0.445	0.359	0.396	0.347	0.378	0.258	0.318	0.232	0.270	0.163	0.258	0.202	0.260	0.397	0.281
Amplifier [2025]	0.421	0.433	0.356	0.402	0.353	0.379	0.256	0.318	**0.223**	0.264	0.163	0.256	0.202	0.256	0.417	0.290
TimeKAN [2025]	0.409	0.427	0.350	0.397	**0.344**	0.380	0.260	0.318	0.226	0.268	0.164	0.258	0.198	0.263	0.420	0.286
SRSNet	**0.404**	**0.424**	**0.334**	**0.385**	0.351	**0.378**	**0.252**	**0.314**	0.226	0.266	**0.161**	**0.254**	**0.183**	0.239	**0.392**	**0.270**

Table 1: Multivariate forecasting average results with forecasting horizons F ∈ {96, 192, 336, 720} for the datasets.

Datasets		ETTh1		Solar	
Metrics		Memory (MB)	Training Time (s)	Memory (MB)	Training Time (s)
Linear	DLinear	828	1.28	815	15.66
	Amplifer	596	1.78	715	14.38
CNN	TimesNet	2,846	14.95	13,141	1812.46
Transformer	FEDformer	8,190	64.83	3,751	227.45
	Stationary	18,386	40.22	18,529	156.27
	Crossformer	3,976	17.13	16,375	205.60
	PatchTST	1,404	2.49	26,777	137.60
	iTransformer	722	4.14	1,015	20.66
MLP	TimeMixer	1,394	7.49	20,602	107.15
	TimeKAN	1,456	5.50	13,109	326.38
	SRSNet	1,012	2.27	6,301	56.149

Table 2: Efficiency comparison between SRSNet and other baselines on ETTh1 and Solar datasets

- SRSNet achieves state-of-the-art peformance compared with recent advanced models.
- SRSNet possesses high efficiency, i.e., fast training speed and low memory occupation.

SRS: A plug-and-play module

Datasets	ETTh1		ETTm2		Solar		Traffic	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MLP	0.430	0.451	0.273	0.336	0.219	0.277	0.413	0.284
+ SRS	0.404	0.424	0.252	0.315	0.183	0.239	0.392	0.270
Improve	**6.05%**	**6.06%**	**7.70%**	**6.31%**	**16.08%**	**13.47%**	**5.26%**	**4.88%**
PatchTST	0.419	0.436	0.256	0.314	0.200	0.284	0.397	0.275
+ SRS	0.404	0.426	0.249	0.307	0.182	0.251	0.386	0.266
Improve	**3.48%**	**2.20%**	**3.00%**	**2.33%**	**8.87%**	**11.13%**	**2.74%**	**3.34%**
Crossformer	0.439	0.461	0.501	0.505	0.205	0.232	0.522	0.282
+ SRS	0.432	0.455	0.454	0.462	0.193	0.225	0.512	0.274
Improve	**1.55%**	**1.45%**	**9.83%**	**8.22%**	**5.66%**	**2.92%**	**1.85%**	**2.84%**
PatchMLP	0.435	0.443	0.261	0.322	0.193	0.250	0.413	0.287
+ SRS	0.422	0.436	0.253	0.315	0.179	0.242	0.402	0.277
Improve	**2.99%**	**1.55%**	**2.94%**	**2.31%**	**7.00%**	**3.11%**	**2.57%**	**3.41%**
xPatch	0.416	0.429	0.253	0.308	0.186	0.210	0.398	0.248
+ SRS	0.406	0.422	0.244	0.303	0.179	0.204	0.389	0.240
Improve	**2.38%**	**1.66%**	**3.59%**	**1.86%**	**4.01%**	**2.61%**	**2.17%**	**3.38%**

Table 3: Five models, i.e., MLP, PatchTST, Crossformer, PatchMLP, xPatch, are considered to show the effectiveness of SRS to work as a plugin

Datasets	Variants	Memory (MB)	Inference (s)	Training (s)	MACs (G)
ETTh1	PatchTST	2,837	5.076	5.131	16.214
	+SRS	2,907	5.722	5.763	16.905
	Overhead	+2.47%	+12.73%	+12.31%	+4.26%
	Crossformer	4,011	27.503	32.613	56.280
	+SRS	4,159	30.311	35.276	56.625
	Overhead	+3.69%	+10.21%	+8.17%	+0.61%
Solar	PatchTST	27,822.08	84.231	88.714	600.261
	+SRS	29,767.68	95.200	101.981	613.790
	Overhead	+6.99%	+13.02%	+14.95%	+2.25%
	Crossformer	17,355	79.031	82.472	61.822
	+SRS	18,978	86.674	90.268	62.174
	Overhead	+9.35%	+9.67%	+9.45%	+0.57%

Table 4: Efficiency analysis of the SRS module.

- About 5% - 15% improvement on classical patch-based models.
- SRS possesses high efficiency, introducing small overload.

Showcases

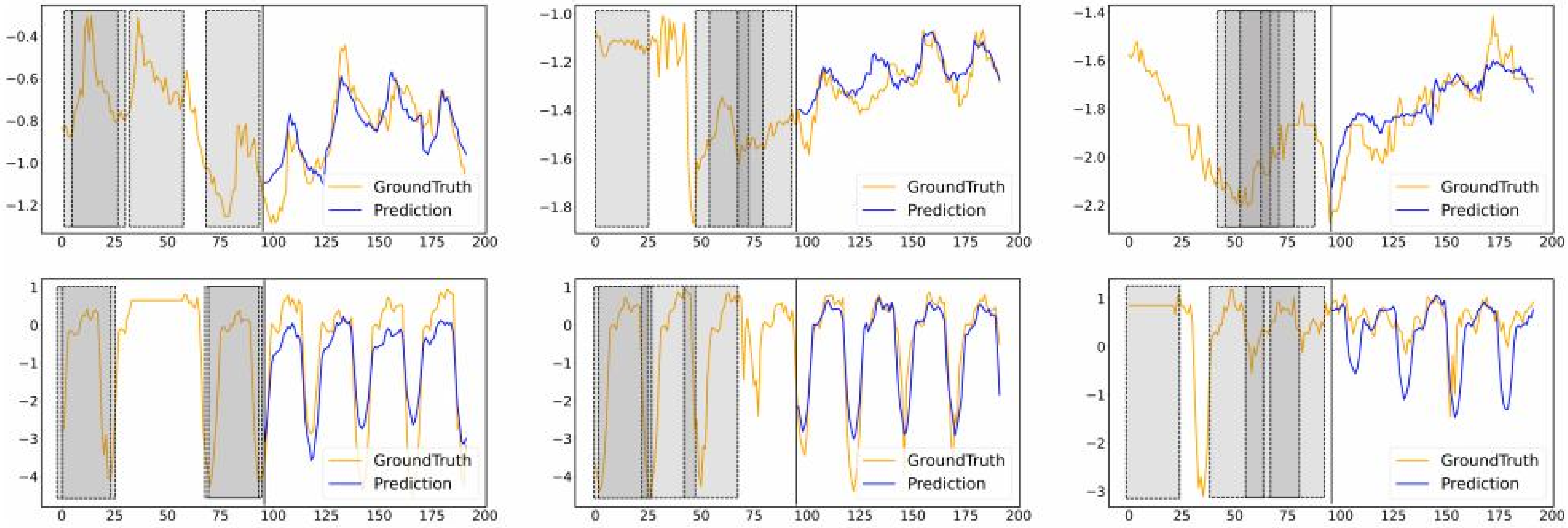


Figure 1: Visualization of input-96-predict-96 results on the ETTh1 dataset. SRSNet effectively processes the special cases with the help of SRS module. The grey rectangles are the selected patches with the size of 24.

Showcases

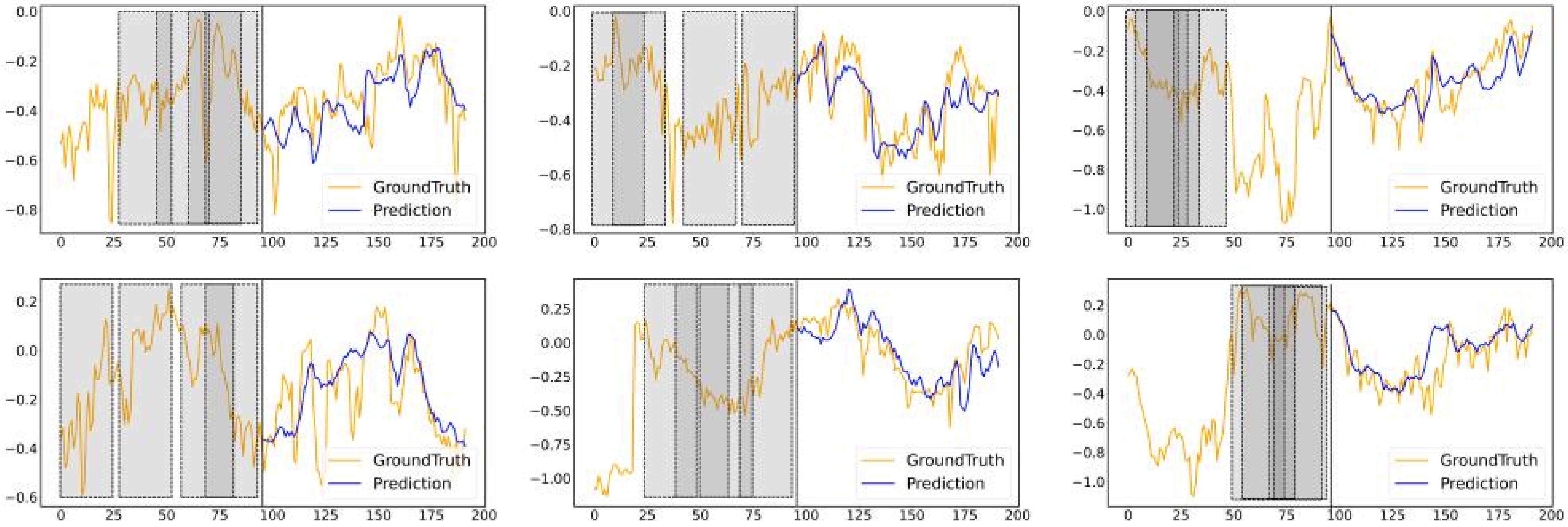


Figure 2: Visualization of input-96-predict-96 results on the ETTm2 dataset. SRSNet effectively processes the special cases with the help of SRS module. The grey rectangles are the selected patches with the size of 24.

Showcases

華東師範大學 EAST CHINA NORMAL UNIVERSITY

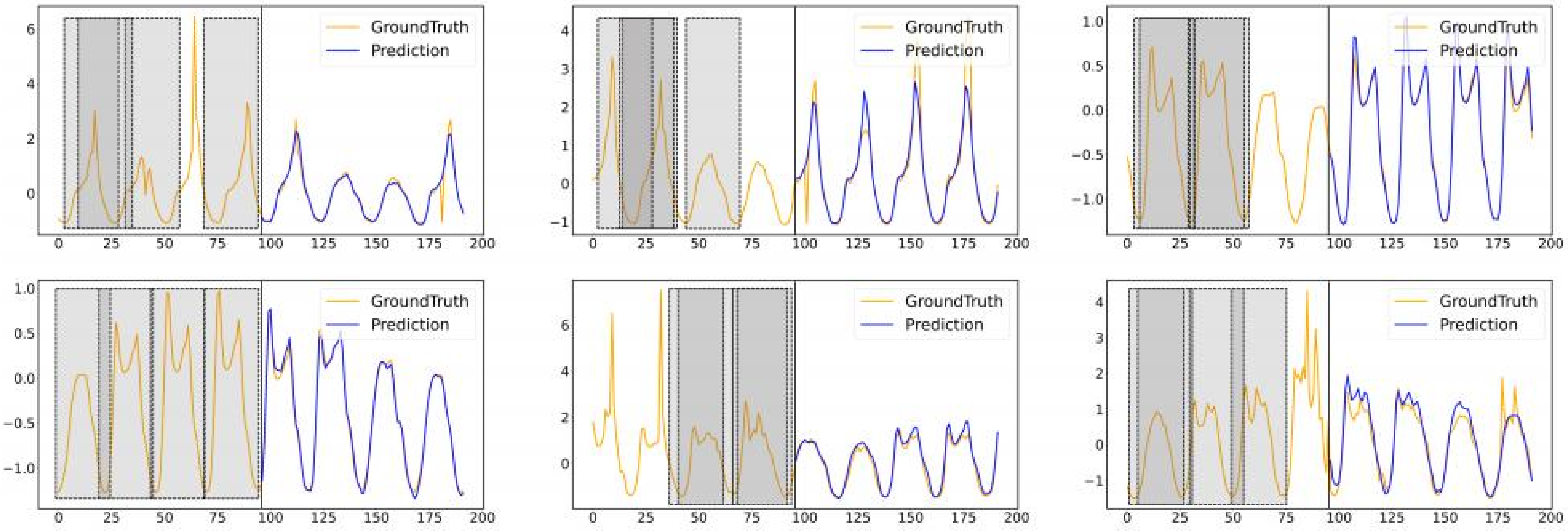


Figure 2: Visualization of input-96-predict-96 results on the Traffic dataset. SRSNet effectively processes the special cases with the help of SRS module. The grey rectangles are the selected patches with the size of 24.

Thank you!