



Temporal 2D-Variation Modeling for General Time Series Analysis

[ICLR 23]

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, Mingsheng Long
School of Software, BNRist, Tsinghua University, Beijing 100084, China

Sunghyun Ahn
skd@yonsei.ac.kr

<2025/03/26>

1 Time Series Anomaly Detection

Time Series Anomaly Detection (TAD)

- 시계열 데이터: 시간 순서대로 나열된 테이블 데이터 (데이터 수(N) x 데이터 측정 시간(T), 변수의 개수에 따라 단변량, 다변량)
- 시계열 이상 탐지: 시계열 데이터에서 이상이 발생한 시간 구간(time intervals)을 찾는 작업
- 다변량 데이터는 여러 변수를 동시에 고려해야 의미있는 분석이 가능할 수 있음 (e.g. 온도는 정상 범위지만 전력 소비량은 증가하면 장비 이상 가능성)



	00:00:00	00:01:00	00:02:00	00:03:00	...
A	102	100	90	130	...
B	0.1	1.5	2.3	0.3	...
C	1002	1200	960	1300	...



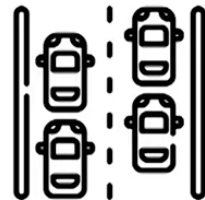
주식 →
이상 금융 탐지



인터넷 속도 →
네트워크 상태 점검



심장 박동 →
건강 상태 확인



시간당 교통량 →
교통 원활 정도 확인

1 Time Series Anomaly Detection

TimesNet

- 다중 주기성(Multi-periodicity)를 고려하여 시계열 분석을 수행한 논문
 - 예를 들어, 기온은 하루마다도 주기성을 가지지만, 크게 보면 일년마다도 주기성을 가지고 있음
 - 크기가 다른 각각의 주기로 시계열을 분할하여 2D 형태의 시계열 데이터를 제작함
- 주기 내 변화(Intraperiod-variation)와 주기 간 변화(Interperiod)를 모두 포착 가능함

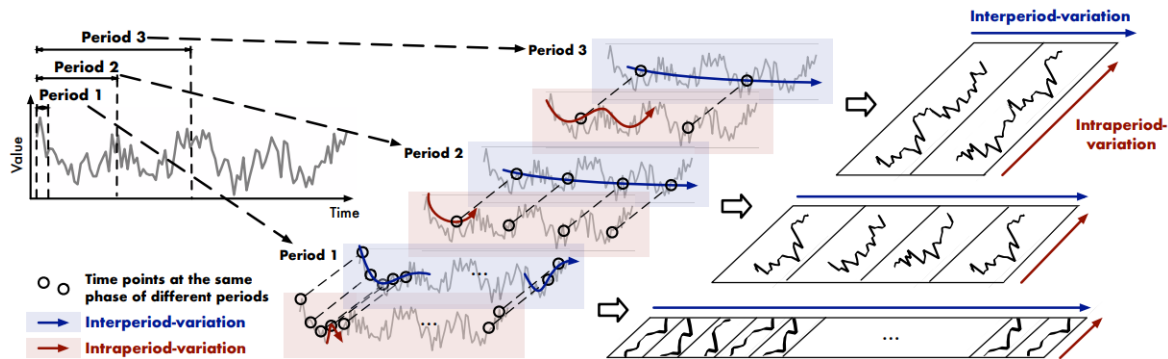


Figure 1: Multi-periodicity and temporal 2D-variation of time series. Each period involves the **intraperiod-variation** and **interperiod-variation**. We transform the original 1D time series into a set of 2D tensors based on multiple periods, which can unify the intraperiod- and interperiod-variations.

1 Time Series Anomaly Detection

TimesNet

Period의 선정 기준은 Fast Fourier Transform(FFT)를 적용하여 산출된 frequency 중 amplitude 값이 가장 높은 Top k개를 이용함

주파수(frequency)는 주기(period)의 역수이므로, 주기를 계산하여 시계열 데이터를 분할할 수 있음

→ 입력 데이터에 존재하는 주기가 큰 신호 (주파수가 작은 신호)와 주기가 작은 신호 (주파수가 큰 신호)를 모두 반영할 수 있도록 함

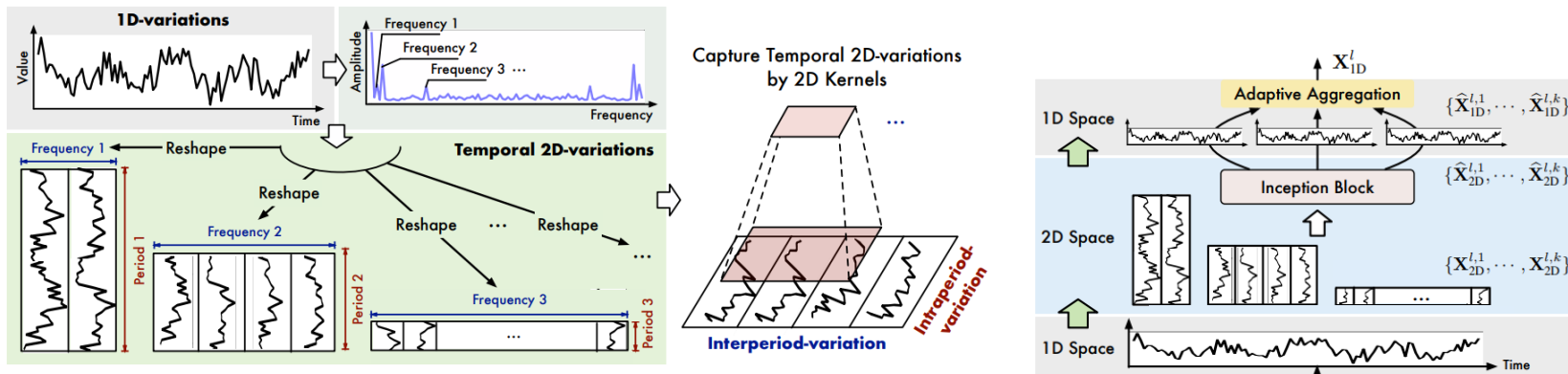


Figure 2: A univariate example to illustrate 2D structure in time series. By discovering the periodicity, we can transform the original 1D time series into structured 2D tensors, which can be processed by 2D kernels conveniently. By conducting the same reshape operation to all variates of time series, we can extend the above process to multivariate time series.

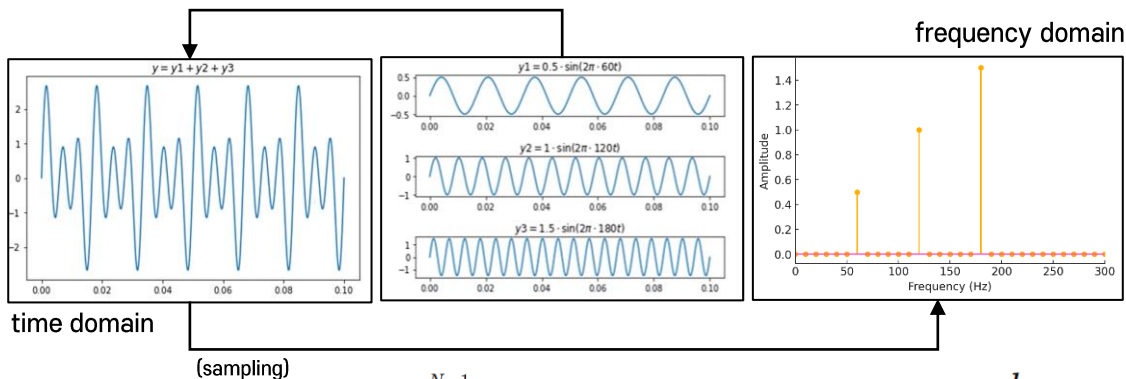
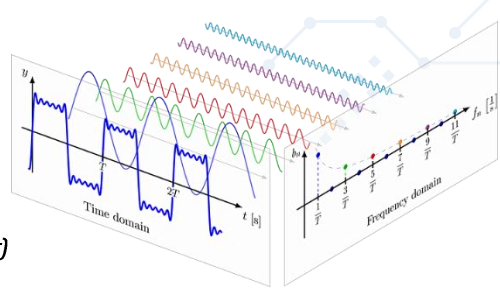
$$\mathbf{A} = \text{Avg} \left(\text{Amp} \left(\text{FFT}(\mathbf{X}_{1D}) \right) \right), \{f_1, \dots, f_k\} = \arg \text{Topk} \left(\mathbf{A} \right), p_i = \left\lceil \frac{T}{f_i} \right\rceil, i \in \{1, \dots, k\}.$$

f_i 는 주파수 인덱스,
 f_i/T 가 실제 주파수
 (샘플링 주파수: 1)

1 Time Series Anomaly Detection

Fourier Transform

- 푸리에 변환: 시간에 대한 함수를 주파수 성분으로 분해하는 변환 (주파수: 1초 동안 파동이 몇 번 반복되는지를 나타낸 값)
- 신호에 어떤 주파수 성분이 존재하는지는 알 수 있지만, 주파수 성분이 어느 시점에 존재하는지는 알 수 없음



$$\omega = e^{-j\frac{2\pi}{N}}$$

$$X = Fx \quad F = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix}$$

주파수 진폭 = 푸리에 행렬 × 신호

낮은 주파수에서 높은 주파수까지 포함된 신호 행렬

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ -2 + 2j \\ -2 \\ -2 - 2j \end{bmatrix}$$

신호와 내적을 통해 해당 주파수가 신호에 얼마나 포함되어 있는지 평가

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, 2, \dots, N-1 \quad f_k = \frac{k}{N} f_s$$

- N: 신호의 길이, k: 주파수 인덱스, j: 복소수 단위
- x[n]: 시간 영역의 입력 신호 (time)
- X[k]: 주파수 영역의 출력 신호 (amplitude)
- f_k: 주파수 (frequency)
- f_s: 샘플링 주파수
- 회전 함수: 복소수 지수함수로, 주파수 성분을 나타냄

1 Time Series Anomaly Detection

TimesNet

- CNN (Inception)을 통해 주기 내 변화와 주기 간 변화를 효과적으로 포착함
- Adaptive Aggregation을 통해 2D tensor의 중요도를 반영함 (FFT과정을 통해 구해진 amplitude 값을 바탕으로 집계)

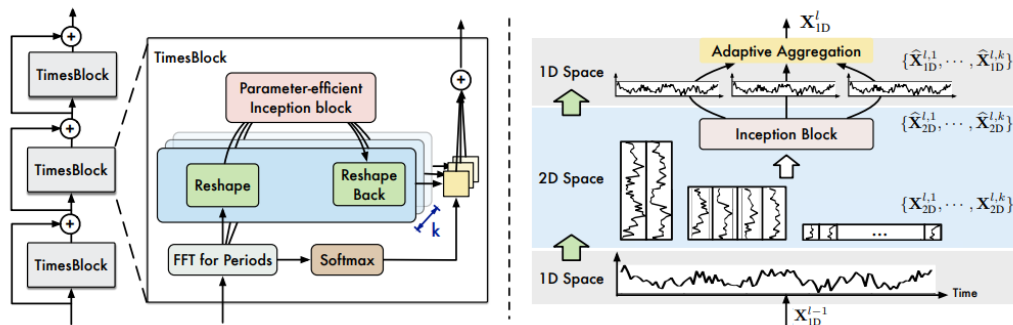


Figure 3: Overall architecture of TimesNet. TimesNet is stacked by TimesBlocks in a residual way. TimesBlocks can capture various temporal 2D-variations from k different reshaped tensors by a parameter-efficient inception block in 2D space and fuse them based on normalized amplitude values.

$$\mathbf{A}^{l-1}, \{f_1, \dots, f_k\}, \{p_1, \dots, p_k\} = \text{Period}(\mathbf{X}_{\text{ID}}^{l-1}),$$

- A: amplitude
- f: frequency
- p: period
- Trunc: truncate

$$\mathbf{X}_{2\text{D}}^{l,i} = \text{Reshape}_{p_i, f_i}(\text{Padding}(\mathbf{X}_{\text{ID}}^{l-1})), i \in \{1, \dots, k\}$$

$$\widehat{\mathbf{X}}_{2\text{D}}^{l,i} = \text{Inception}(\mathbf{X}_{2\text{D}}^{l,i}), i \in \{1, \dots, k\}$$

$$\widehat{\mathbf{X}}_{\text{ID}}^{l,i} = \text{Trunc}(\text{Reshape}_{1, (p_i \times f_i)}(\widehat{\mathbf{X}}_{2\text{D}}^{l,i})), i \in \{1, \dots, k\},$$

$$\widehat{\mathbf{A}}_{f_1}^{l-1}, \dots, \widehat{\mathbf{A}}_{f_k}^{l-1} = \text{Softmax}(\mathbf{A}_{f_1}^{l-1}, \dots, \mathbf{A}_{f_k}^{l-1})$$

$$\mathbf{X}_{\text{ID}}^l = \sum_{i=1}^k \widehat{\mathbf{A}}_{f_i}^{l-1} \times \widehat{\mathbf{X}}_{\text{ID}}^{l,i}$$

1 Time Series Anomaly Detection

TimesNet

- 대부분의 시계열 tasks에서 SOTA 성능 달성함
- Long-term, Short-term forecasting task에서 모두 우수한 성능을 보임

Table 1: Summary of experiment benchmarks.

Tasks	Benchmarks	Metrics	Series Length
Forecasting	Long-term: ETT (4 subsets), Electricity, Traffic, Weather, Exchange, ILI	MSE, MAE	96~720 (ILI: 24~60)
	Short-term: M4 (6 subsets)	SMAPE, MASE, OWA	6~48
Imputation	ETT (4 subsets), Electricity, Weather	MSE, MAE	96
Classification	UEA (10 subsets)	Accuracy	29~1751
Anomaly Detection	SMD, MSL, SMAP, SWaT, PSM	Precision, Recall, F1-Score	100

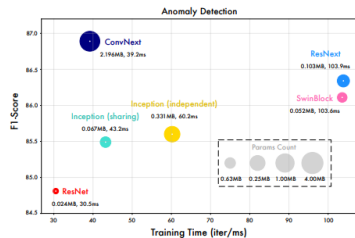
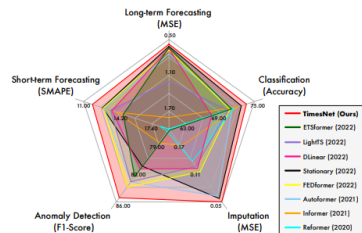


Figure 4: Model performance comparison (left) and generality in different vision backbones (right).

Table 2: Long-term forecasting task. The past sequence length is set as 36 for ILI and 96 for the others. All the results are averaged from 4 different prediction lengths, that is {24, 36, 48, 60} for ILI and {96, 192, 336, 720} for the others. See Table 13 in Appendix for the full results.

Models	TimesNet (Ours)	ETSformer (2022)	LightTS (2022)	DLinear (2023)	FEDformer (2022)	Stationary (2022a)	Autoformer (2021)	Pyraformer (2021a)	Informer (2021)	LogTrans (2019)	Reformer (2020)
Metric	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE
ETTM1	0.400 0.406	0.429 0.425	0.435 0.437	0.403 0.407	0.448 0.452	0.481 0.456	0.588 0.517	0.691 0.607	0.961 0.734	0.929 0.725	0.799 0.671
ETTM2	0.291 0.333	0.293 0.342	0.409 0.436	0.350 0.401	0.305 0.349	0.306 0.347	0.327 0.371	1.498 0.869	1.410 0.810	1.535 0.900	1.479 0.915
ETTh1	0.458 0.450	0.542 0.510	0.491 0.479	0.456 0.452	0.440 0.460	0.570 0.537	0.496 0.487	0.827 0.703	1.040 0.795	1.072 0.837	1.029 0.805
ETTh2	0.414 0.427	0.439 0.452	0.602 0.543	0.559 0.515	0.437 0.449	0.526 0.516	0.450 0.459	0.826 0.703	4.431 1.729	2.686 1.494	6.736 2.191
Electricity	0.192 0.295	0.208 0.323	0.229 0.329	0.212 0.300	0.214 0.327	0.193 0.296	0.227 0.338	0.379 0.445	0.311 0.397	0.272 0.370	0.338 0.422
Traffic	0.620 0.336	0.621 0.396	0.622 0.392	0.625 0.383	0.610 0.376	0.624 0.340	0.628 0.379	0.878 0.469	0.764 0.416	0.705 0.395	0.741 0.422
Weather	0.259 0.287	0.271 0.334	0.261 0.312	0.265 0.317	0.309 0.360	0.288 0.314	0.338 0.382	0.946 0.717	0.634 0.548	0.696 0.602	0.803 0.656
Exchange	0.416 0.443	0.410 0.427 0.385	0.447 0.354 0.414	0.519 0.500	0.461 0.454	0.613 0.539	1.913 1.159	1.550 0.998	1.402 0.968	1.280 0.932	
ILI	2.139 0.931	2.497 1.004	7.382 2.003	2.616 1.090	2.847 1.144	2.077 0.914	3.006 1.161	7.635 2.050	5.137 1.544	4.839 1.485	4.724 1.445

Table 3: Short-term forecasting task on M4. The prediction lengths are in [6, 48] and results are weighted averaged from several datasets under different sample intervals. See Table 14 for full results.

Models	TimesNet (Ours)	N-HITS (2022)	N-BEATS (2019)	ETSformer (2022)	LightTS (2023)	DLinear (2022)	FEDformer (2022a)	Stationary (2021)	Autoformer (2021)	Pyraformer (2021a)	Informer (2019)	LogTrans (2020)	Reformer (2020)
SMAPE	11.829	11.927	11.851	14.718	13.525	13.639	12.840	12.780	12.909	16.987	14.086	16.018	18.200
MASE	1.585	1.613	1.599	2.408	2.111	2.095	1.701	1.756	1.771	3.265	2.718	3.010	4.223
OWA	0.851	0.861	0.855	1.172	1.051	1.051	0.918	0.930	0.939	1.480	1.230	1.378	1.775