11 / 15 / 2024 (Fri)

Time Series Anomaly Detection Using Diffusion Models

Seokho Moon

danny232@korea.ac.kr

School of Industrial and Management Engineering, Korea University

문석호 (Seokho Moon)

- 고려대학교 산업경영공학과 대학원 재학 중
- Data Mining & Quality Analytics Lab (김성범 교수님)
- 석박통합과정 (2019.09 ~)

관심 연구 분야

- Anomaly detection
- Generative models

E-mail

- danny232@korea.ac.kr

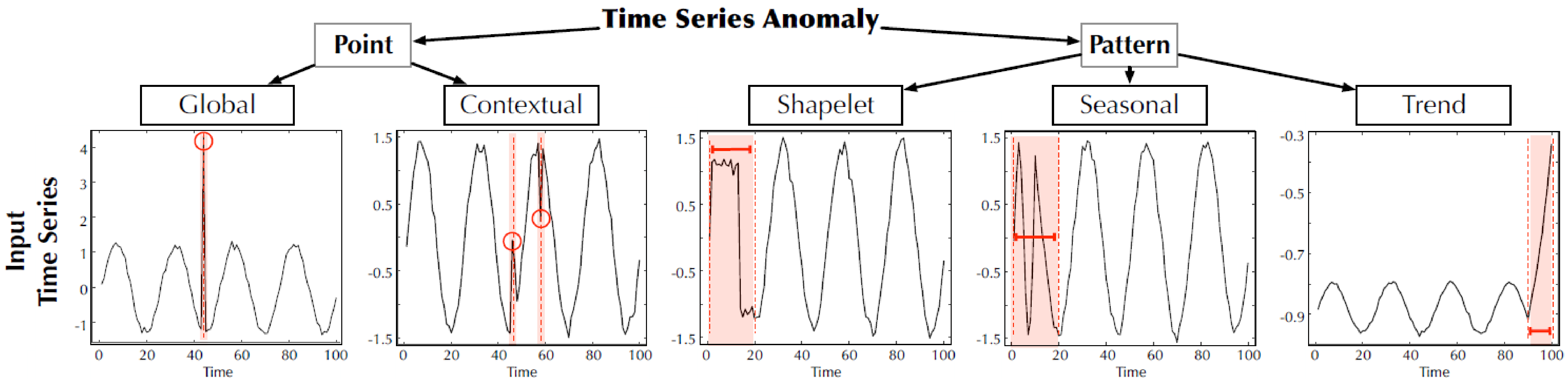
Introduction

- **Time Series Anomaly Detection**

  Multivariate Time series 데이터는 다양한 분야에서 사용되고 있음.

  그 중에서 이상탐지는 매우 중요한 과업 중 하나.

  그러나 실제 time series 의 anomaly 는 매우 다양한 형태로 나타나서 detection 이 어려움.



Xu, J. (2021). Anomaly transformer: Time series anomaly detection with association discrepancy. arXiv preprint arXiv:2110.02642.

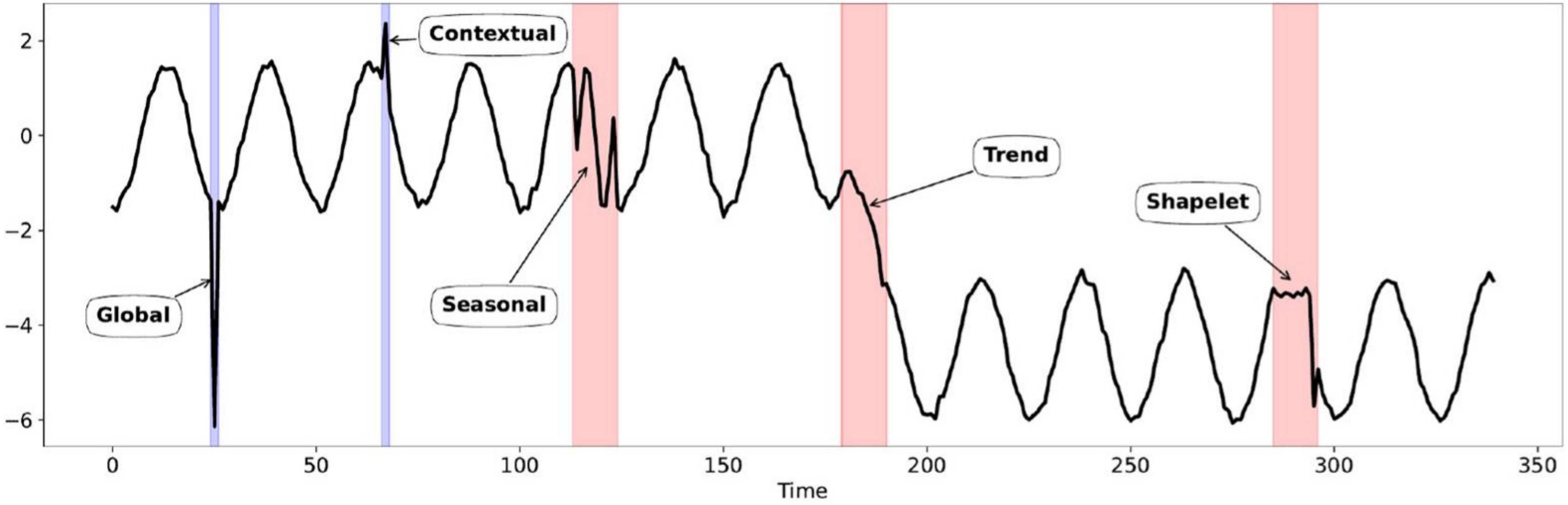
Introduction

- **Time Series Anomaly Detection**

  Multivariate Time series 데이터는 다양한 분야에서 사용되고 있음.

  그 중에서 이상탐지는 매우 중요한 과업 중 하나.

  그러나 실제 time series 의 anomaly 는 매우 다양한 형태로 나타나서 detection 이 어려움.



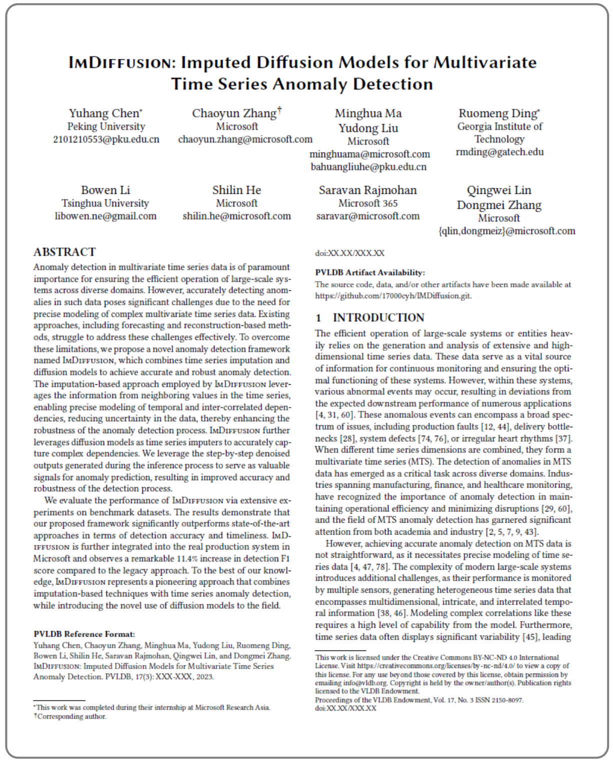
Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. ACM Computing Surveys, 57(1), 1-42.

Introduction

- **(Diffusion-based) Time Series Anomaly Detection**

  [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

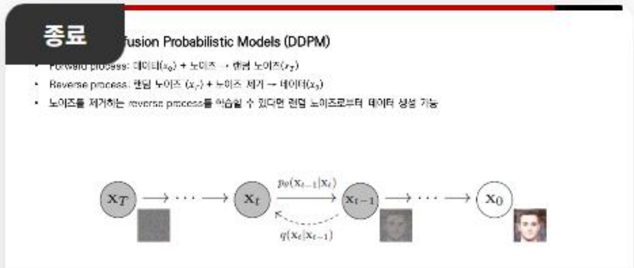
  [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**



Introduction

- **DMQA Seminar**

종료



Score-based Generative Models and Diffu

발표자: 조한샘

2022년 2월 11일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

Conditional Diffusion Models

발표자: 이종현

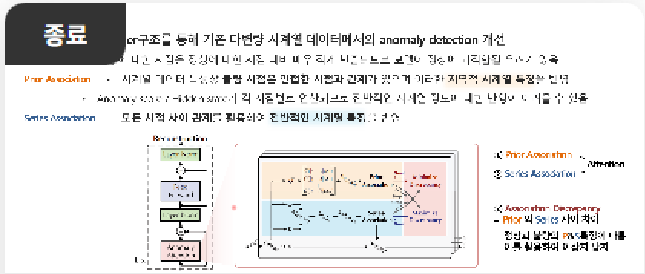
2023년 6월 16일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료



Transformer-based Anomaly Detection in

발표자: 이지윤

2023년 1월 27일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

문제점: 기존 deep learning 방법의 발달에도 불구하고, anomaly concentration 상황에서 성능이 좋지 않음. (prediction, reconstruction-based)

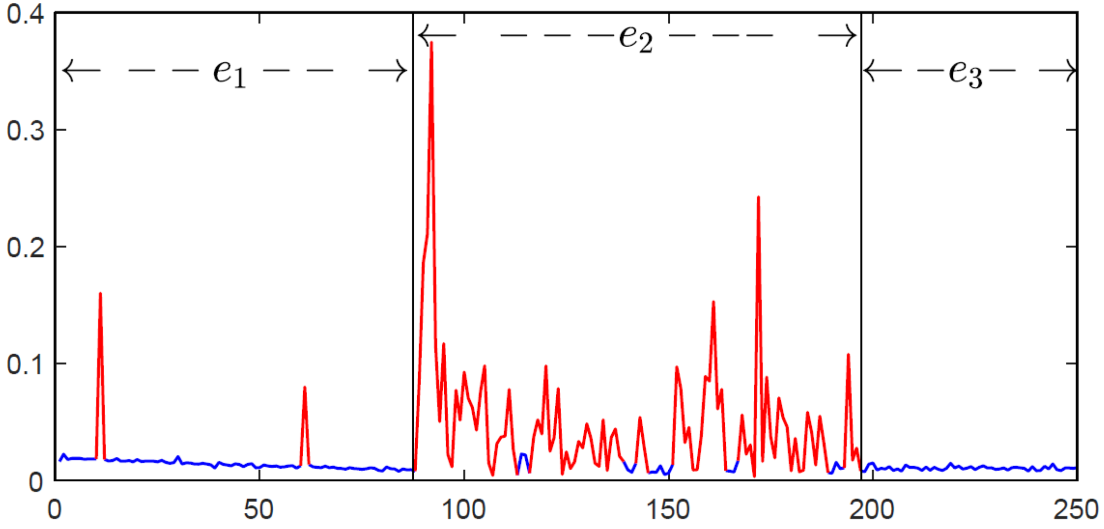


Figure 1: An example of anomaly concentration drawn sampled from the server machine dataset. Blue and red points denote normal and anomalous ones, respectively. Episode e_2 is an anomaly concentration region, and episode e_1 and e_3 are normal data dominant regions.

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

Benchmark dataset 분석: sequence length 가 100 or 1000 이상인 anomaly point 가 70% 혹은 80% 이상인 상황

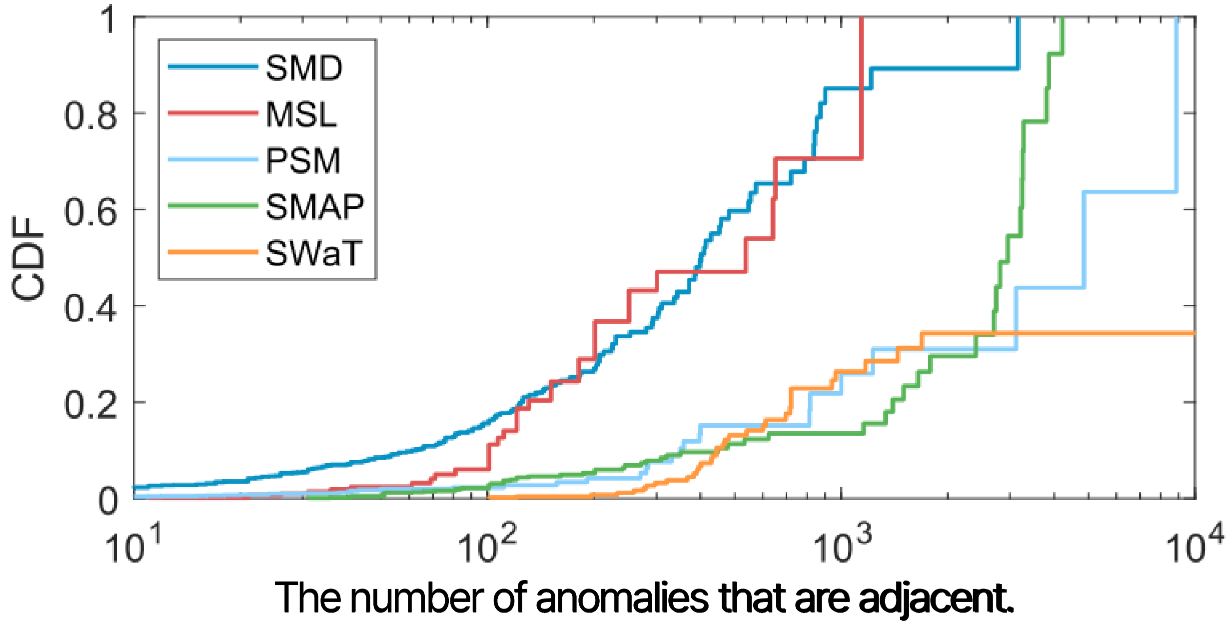
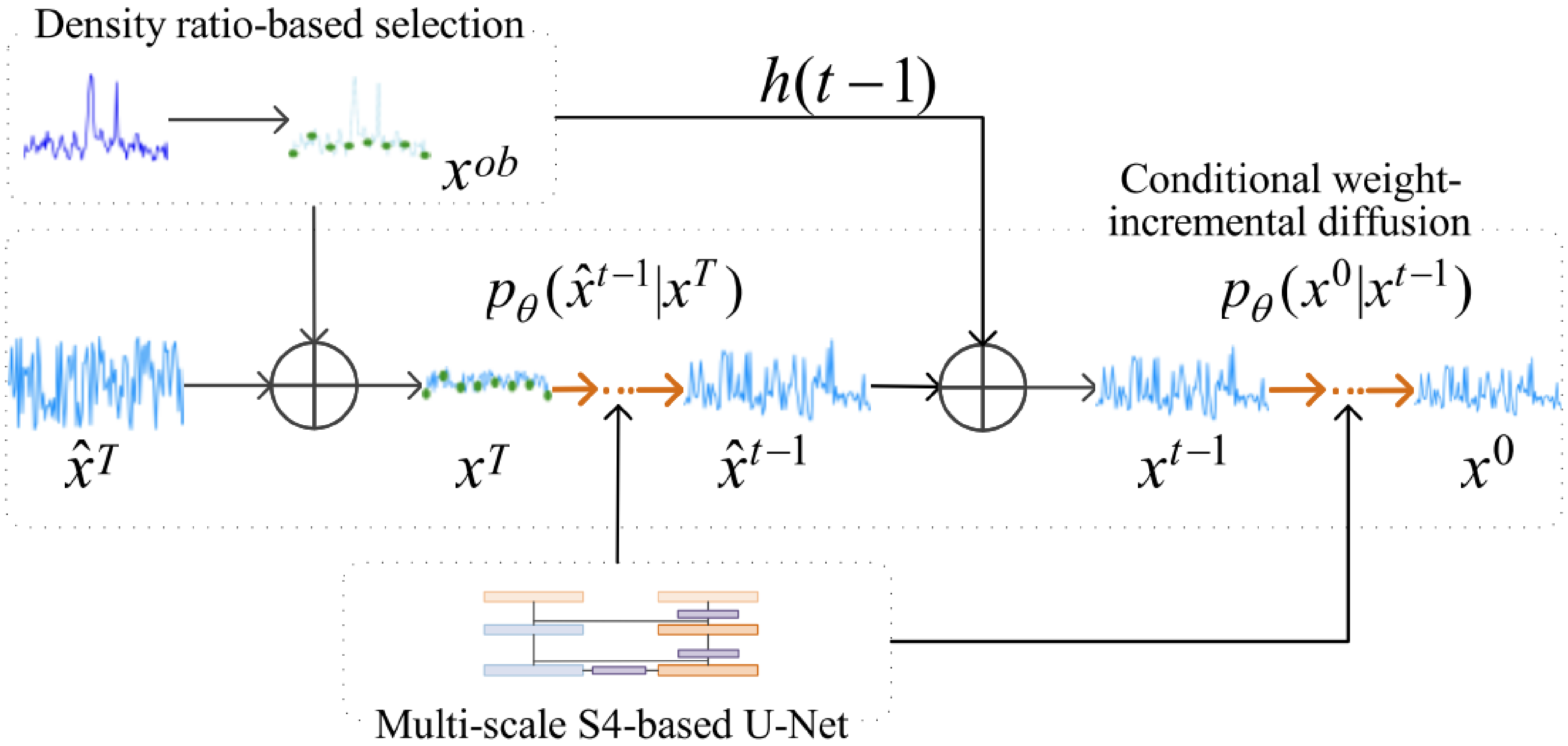


Figure 2: The CDFs of the consecutive anomalous points.

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

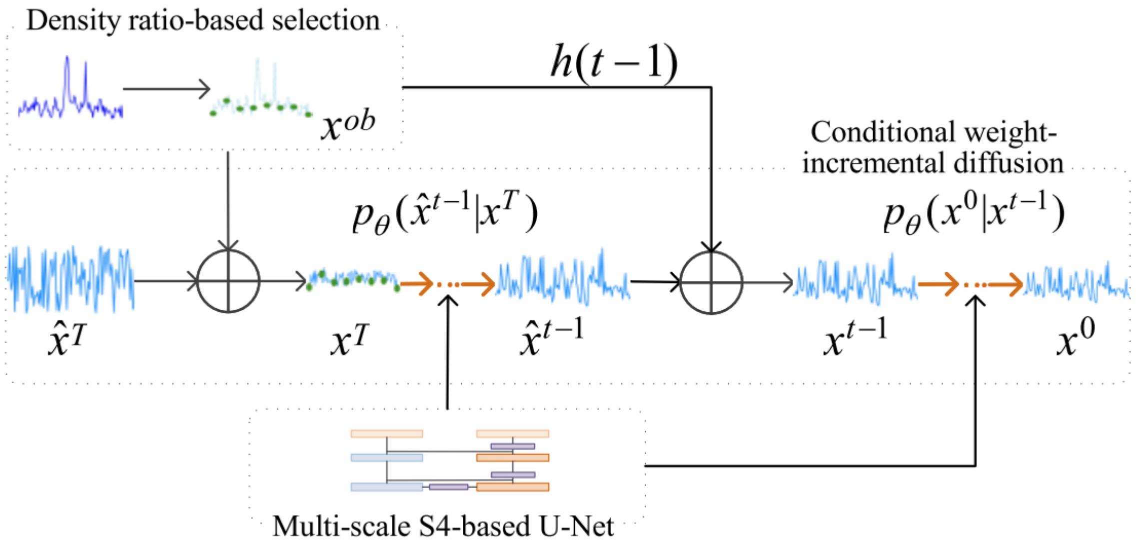
  제안 방법론 DiffAD 구조: 크게 3가지 핵심 요소로 되어 있음



Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

  제안 방법론 DiffAD 구조: 1) Density ratio-based selection



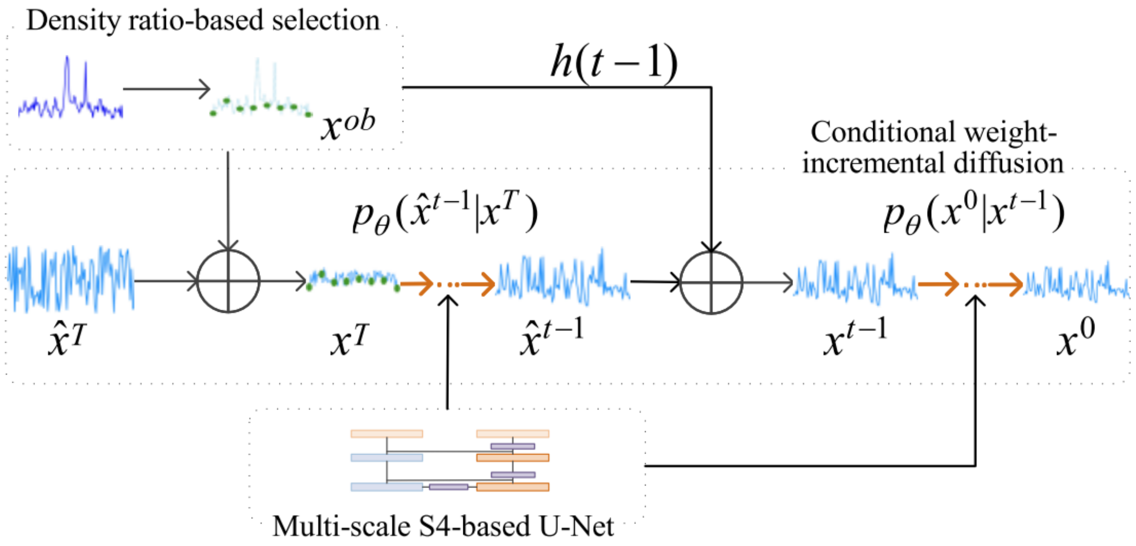
$$g_k(x) = \frac{f_{k-1}(x)}{f_k(x)}, \tag{1}$$

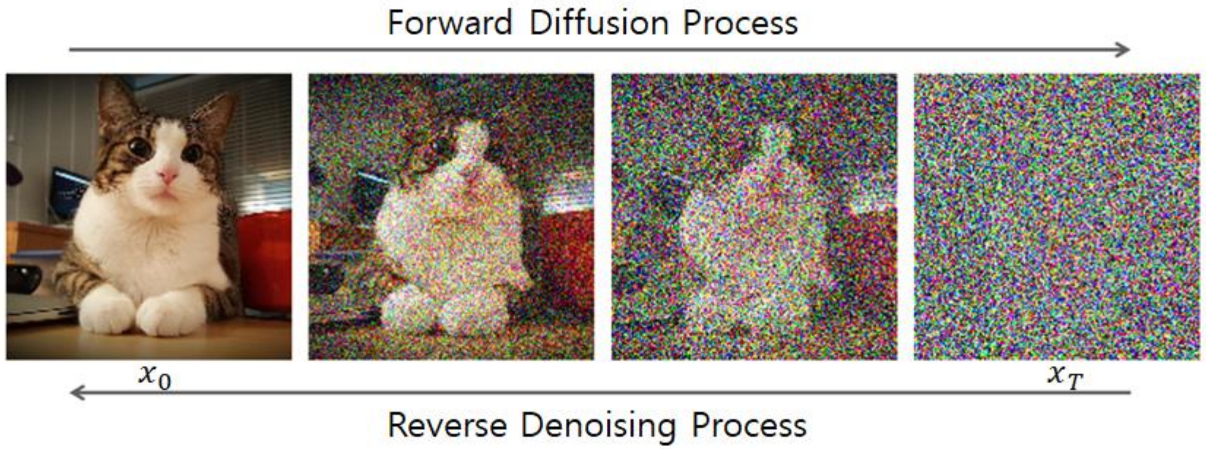
where $f_{k-1}(x)$ and $f_k(x)$ correspond to estimated probability densities of the two consecutive windows respectively. Then, the change score is calculated:

$$\hat{CHG} = \mathrm{Max}\left(0, \quad \frac{1}{2} - \frac{1}{s}\sum_{i=1}^{s} g_k(x^i),\right), \tag{2}$$

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**
  - **제안 방법론 DiffAD 구조:** 2) Conditional weight-incremental diffusion

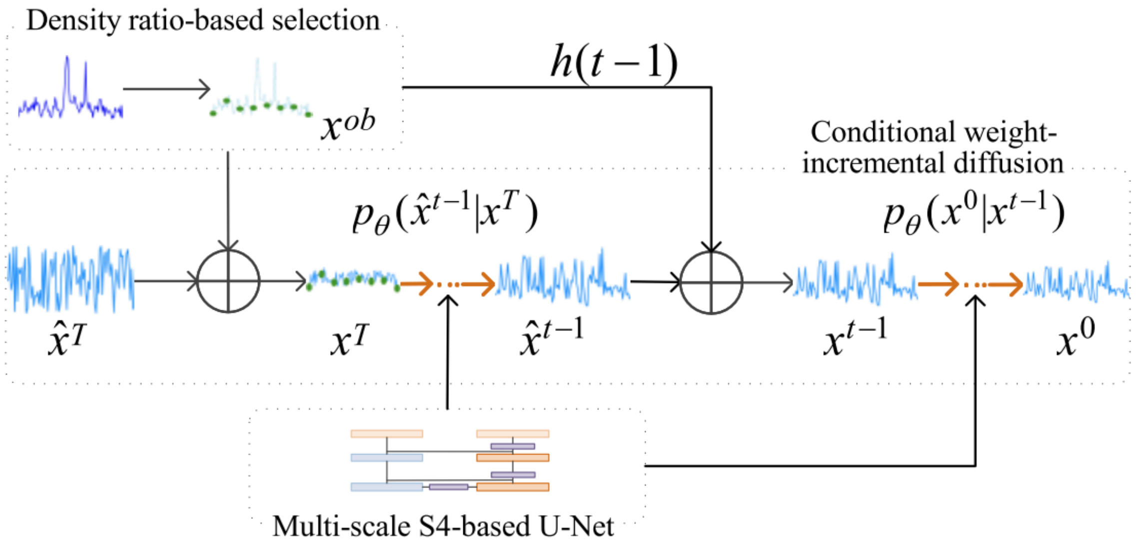




Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

제안 방법론 DiffAD 구조: 2) Conditional weight-incremental diffusion



$$x^T = s \odot x^{ob} + (1-s) \odot \left(g(x^{ob})\gamma + \hat{x}^T(1-\gamma)\right), \tag{5}$$

To parameterize $\mu_\theta(x^t, x^{ob}, t)$, we train a neural denoising model $f_\theta(x^t, x^{ob}, t)$ to predict the noise vector ϵ. The objective is defined as follows:

$$\mathbb{E}_{x^{ob}}\mathbb{E}_{(\epsilon,t)}\left[\|f_\theta(x^t, x^{ob}, t) - \epsilon\|_2^2\right], \tag{7}$$

and $\mu_\theta(x^t, x^{ob}, t)$ can be derived from $f_\theta(x^t, x^{ob}, t)$:

$$\mu_\theta(x^t, x^{ob}, t) = \frac{1}{\sqrt{a_t}}\left(x^t - \frac{1-a_t}{\sqrt{1-\bar{a}_t}}f_\theta(x^t, x^{ob}, t)\right). \tag{8}$$

Consequently, the generative (reverse diffusion) process is:

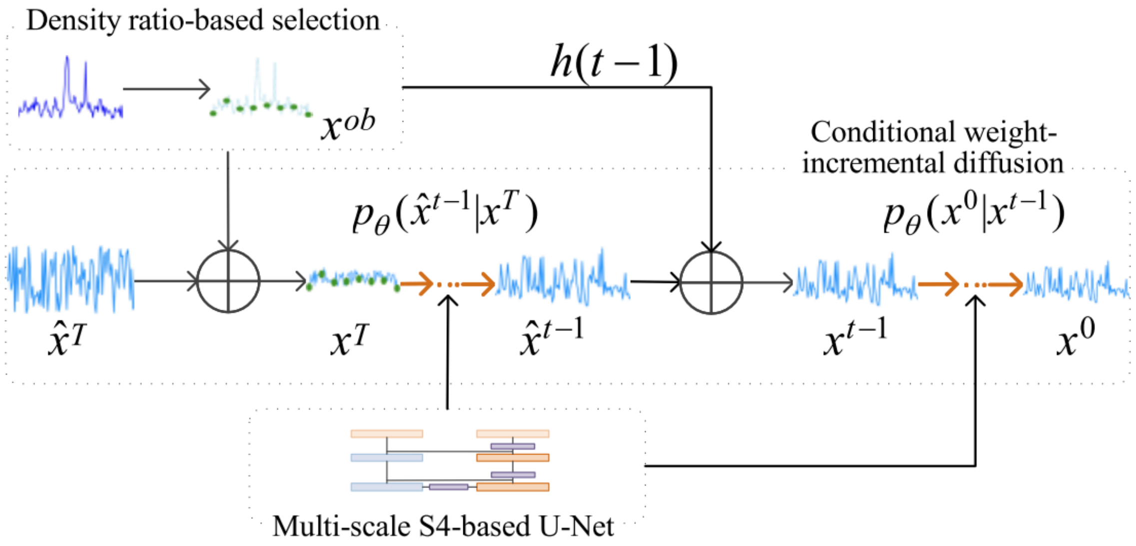
$$\hat{x}^{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x^t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}f_\theta(x^t, x^{ob}, t)\right) + \sqrt{\tilde{\beta}_t}\mathbf{z}, \tag{9}$$

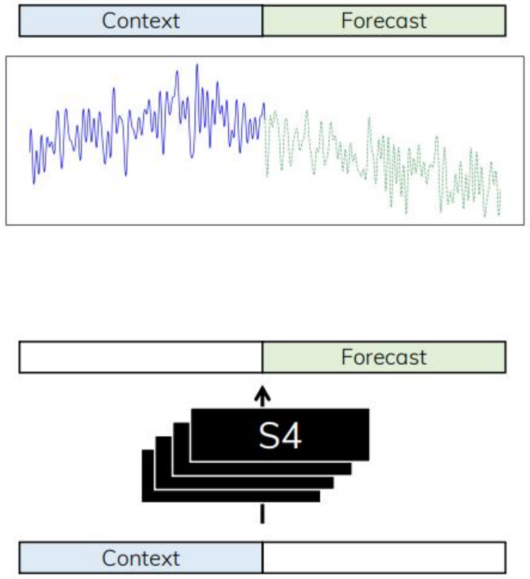
$$x^{t-1} = s \odot \left(\left(1-h(t-1)\right)\hat{x}^{t-1} + h(t-1)x^{ob}\right) + (1-s) \odot \hat{x}^{t-1}, \tag{11}$$

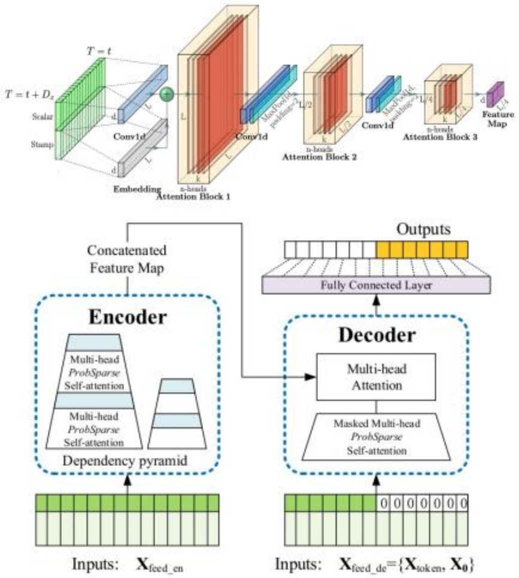
Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

  제안 방법론 DiffAD 구조: 3) Multi-scale S4-based U-Net (structured state-space sequence model)







Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

  Anomaly score 를 설정하는 방법

For anomaly detection, we compare the anomaly score with a given threshold to determine anomalies. For a testing point, its anomaly score is computed based on the estimation error:

$$AS(c_i) = \sum_{k=1}^{d} \|c_i^k - \hat{c}_i^k\|^2, \tag{12}$$

where c_i and $\hat{c}_i$ are the real value and estimated value, respectively, and d refers to the dimension of multivariate time series.

Similar to the previous works [57], we obtain the threshold based on the training data. Given the training data $\mathcal{X} = \{c_1, c_2, ..., c_N\}$, the corresponding decision threshold $\mathcal{T}$ is:

$$\mathcal{T} = \frac{1}{N}\sum_{i=1}^{N} \ell(c_i) + \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left(\ell(c_i) - \ell_{avg}\right)^2}, \tag{13}$$

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

  사용한 데이터셋 설명

Table 1: Descriptive Statistics of Datasets

Datasets	Applications	# Dimension	# point	anomaly(%)
MSL	Space	55	132,046	10.5%
SWaT	Water	51	944,919	12.1%
PSM	Server	25	220,322	27.8%
SMAP	Space	25	562,800	12.8%
SMD	Server	38	1,416,825	4.2%

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

  사용한 데이터셋에 대한 성능

Table 2: Performance comparison between DiffAD and baselines on the five datasets.

Method	MSL			SWaT			PSM			SMAP			SMD		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
DAGMM	89.60	63.93	74.62	89.92	57.84	70.40	93.49	70.03	80.08	86.45	56.73	68.51	67.30	49.89	57.30
MPPCACD	81.42	61.31	69.95	82.52	68.29	74.73	76.26	78.35	77.29	88.61	75.84	81.73	71.20	79.28	75.02
LOF	47.72	85.25	61.18	72.15	65.43	68.62	57.89	90.49	70.61	58.93	56.33	57.60	56.34	39.86	46.68
ITAD	69.44	84.09	76.07	63.13	52.08	57.08	72.80	64.02	68.13	82.42	66.89	73.85	86.22	73.71	79.48
THOC	88.45	90.97	89.69	83.94	86.36	85.13	88.14	90.99	89.54	92.06	89.34	90.68	79.76	90.95	84.99
Deep-SVDD	91.92	76.63	83.58	80.42	84.45	82.39	95.41	86.49	90.73	89.93	56.02	69.04	78.54	79.67	79.10
CSDI	90.46	90.92	90.69	91.66	91.98	91.82	94.30	95.23	94.76	94.23	93.85	94.04	88.32	89.03	88.67
STING	88.25	89.15	88.70	87.28	87.69	87.48	92.35	93.47	92.91	88.97	89.85	89.41	85.14	86.49	85.81
CL-MPPCA	73.71	88.54	80.44	76.78	81.50	79.07	56.02	**99.93**	71.80	86.13	63.16	72.88	82.36	76.07	79.09
LSTM	85.45	82.50	83.95	86.15	83.27	84.69	76.93	89.64	82.80	89.41	78.13	83.39	78.55	85.28	81.78
LSTM-VAE	85.49	79.94	82.62	76.00	89.50	82.20	73.62	89.92	80.96	92.20	67.75	78.10	75.76	90.08	82.30
BeatGAN	89.75	85.42	87.53	64.01	87.46	73.92	90.30	93.84	92.04	92.38	55.85	69.61	72.90	84.09	78.10
OmniAnomaly	89.02	86.37	87.67	81.42	84.30	82.83	88.39	74.46	80.83	92.49	81.99	86.92	83.68	86.82	85.22
InterFusion	81.28	92.70	86.62	80.59	85.58	83.01	83.61	83.45	83.52	89.77	88.52	89.14	87.02	85.43	86.22
ATransformer	92.09	95.15	93.59	91.55	96.73	94.07	96.91	98.90	97.89	94.13	**99.40**	96.69	89.40	95.45	92.33
DiffAD	**92.97**	**95.44**	**94.19**	**98.44**	**96.90**	**97.66**	**97.00**	98.92	**97.95**	**96.52**	97.38	**96.95**	**90.01**	**95.67**	**92.75**

Methods

- [KDD 2023] **(DiffAD) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models**

Ablation study

Table 3: The role of different components

Dataset	Metric	DiffAD-Base	DiffAD-Weight	DiffAD-MS4	DiffAD-CHG	DiffAD (Full)
MSL	P	85.64	87.51	89.84	90.42	**92.97**
	R	86.55	87.46	87.35	92.57	**95.44**
	F1	86.09	87.48	88.58	91.48	**94.19**
SWaT	P	90.12	91.22	94.37	95.55	**98.44**
	R	91.24	91.94	93.65	96.16	**96.90**
	F1	90.68	91.58	94.01	95.85	**97.66**
PSM	P	91.61	92.49	94.05	95.98	**97.00**
	R	93.64	94.54	95.87	96.13	**98.92**
	F1	92.61	93.50	94.95	96.05	**97.95**
SMAP	P	89.45	90.36	92.62	94.28	**96.52**
	R	90.44	91.03	92.99	95.67	**97.38**
	F1	89.94	90.69	92.80	94.97	**96.95**
SMD	P	81.87	83.21	85.64	88.33	**90.01**
	R	85.42	86.89	89.33	92.75	**95.67**
	F1	83.61	85.01	87.45	90.49	**92.75**

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**

  문제상황: 정상 구간에서 forecasting 과 reconstruction 방법은 error 가 크게 나와서 적절한 threshold 를 선정하기 어려움.

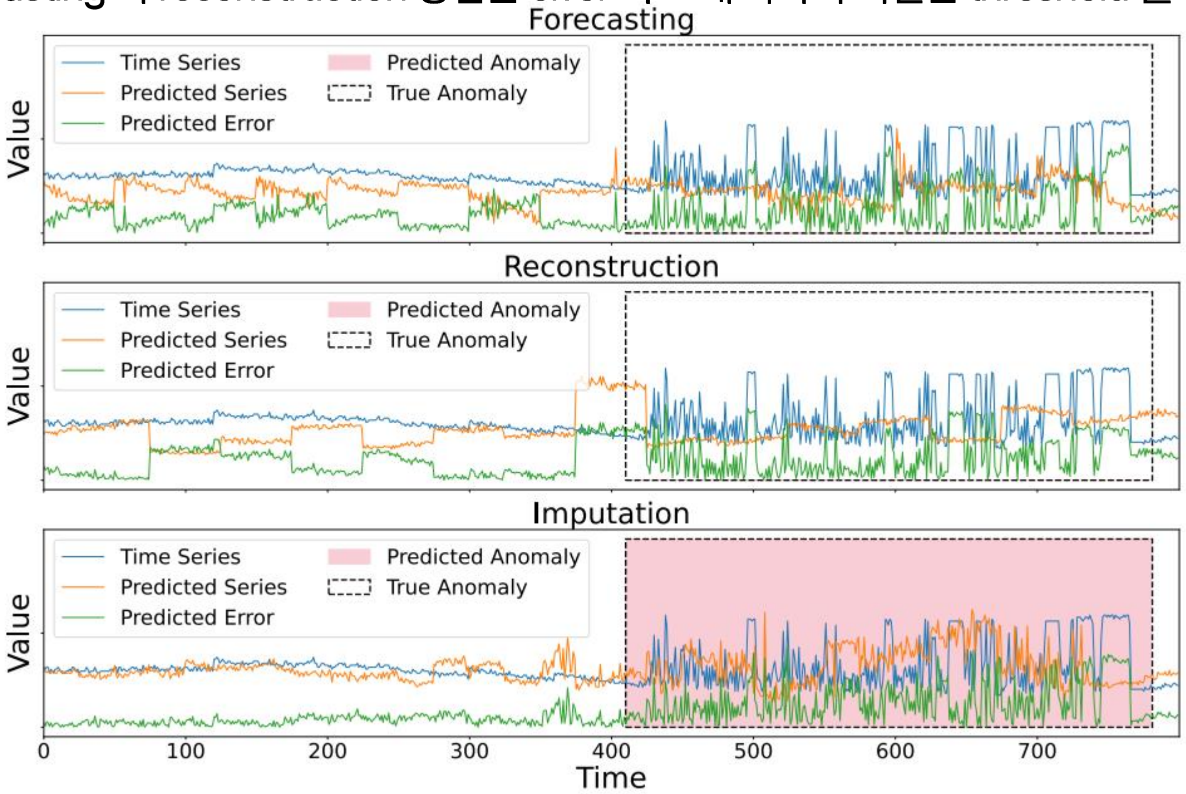


Figure 1: Examples of reconstruction, forecasting and imputation modeling of time series for anomaly detection.

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**

제안 방법론의 전체 framework

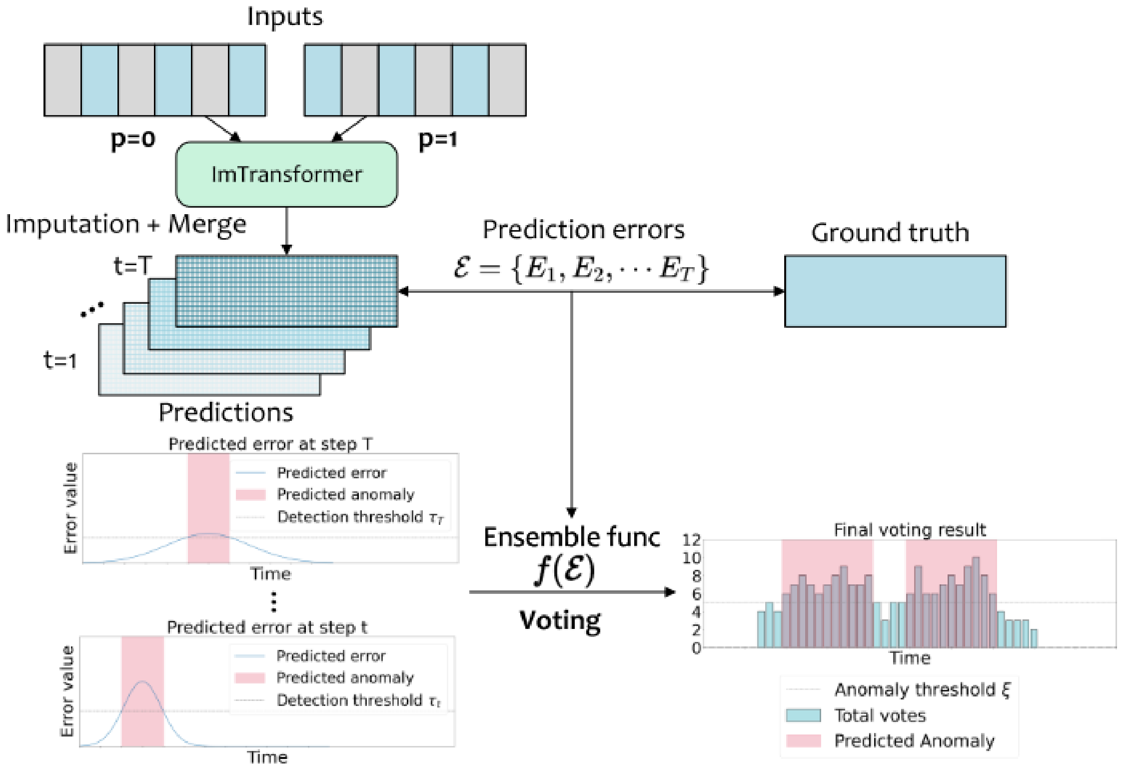


Figure 6: The ensemble anomaly inference of ImDiffusion.

Algorithm 1 The ensemble inference process of ImDiffusion.

Inputs:
Masked data input series $\mathcal{X}_t^{in}$, masking tensors $\mathcal{M}$, a trained denoising model ϵ_Θ, the forward ground truth noise on unmasked region $\epsilon_{1:T}^{\mathcal{M}_1}$, total denosing step T.
Initialise:
Initial noise vector $\mathcal{X}_T$.
for $t = T$ to 1 **do**
Construct two input series $\mathcal{X}_t^{in} = \{\mathcal{X}_t^{\mathcal{M}_0}, \epsilon_t^{\mathcal{M}_1}\}$ with masking $\mathcal{M}$.
Predicting μ_Θ, Σ_Θ using the denosing model ϵ_Θ.
Sampling using equation (9) and obtain predicted $\mathcal{X}_{t-1}$.
Compute prediction error $E_t = ||\mathcal{X} - \mathcal{X}_{t-1}||^2$.
end for
for $t = T$ to 1 **do**
Computing the anomaly prediction label Y_t using Eq. (12).
end for
Aggregating the voted anomaly prediction $\mathcal{V}_l = \sum_{t=1}^{T} y_{t,l}$.
Computing the final anomaly prediction $y_l = \mathbb{1}(\mathcal{V}_l > \xi)$.

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**

제안 방법론 중 ImDiffusion 구조

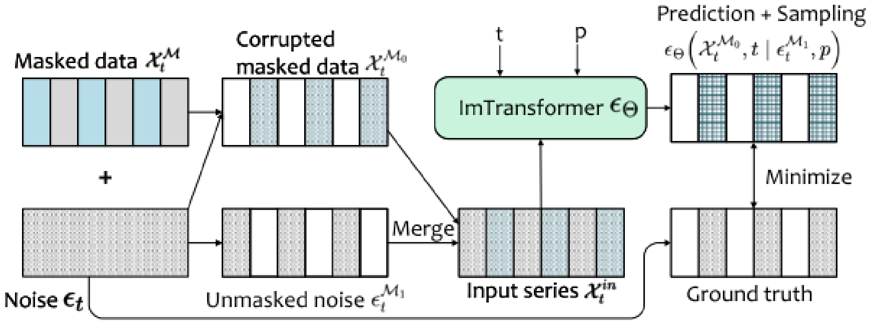


Figure 4: The training process of IMDIFFUSION.

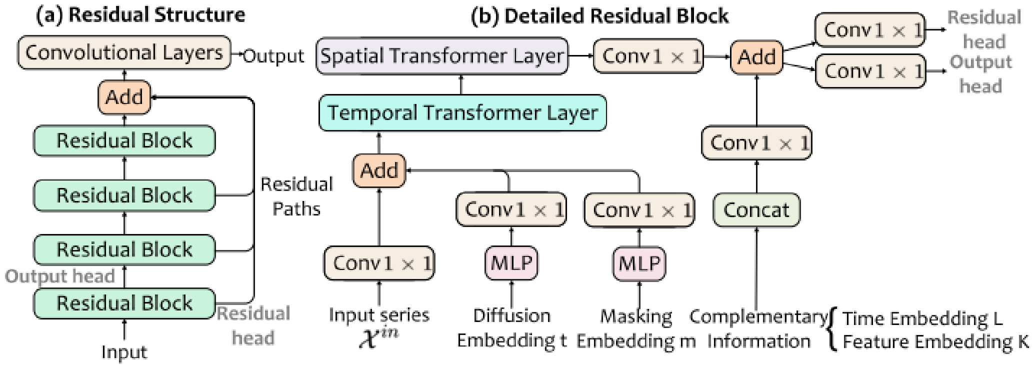


Figure 5: The IMTRANSFORMER architecture, with (a) the residual structure; and (b) the details of a residual block.

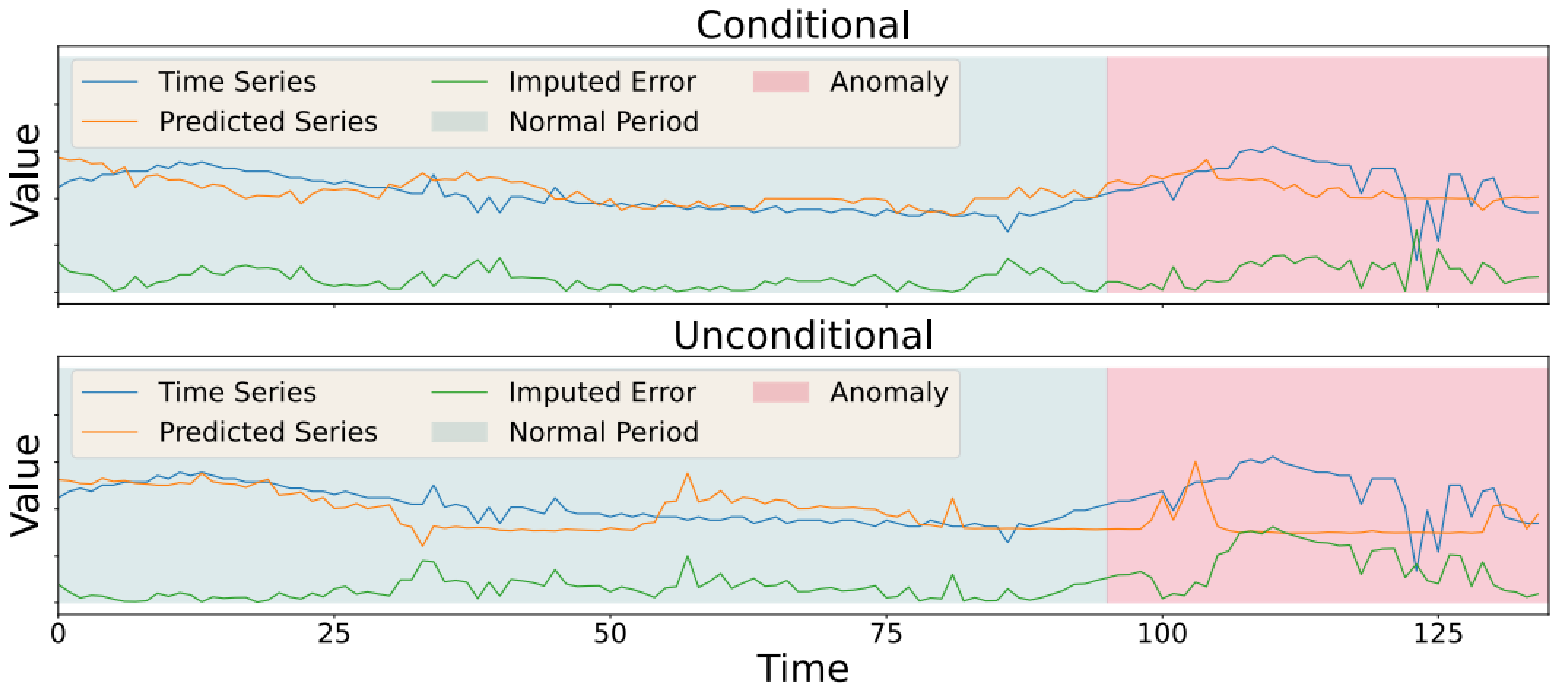
$$\min_{\Theta} \mathcal{L}(\Theta) := \min_{\Theta} \mathbb{E}_{\mathcal{X}_0 \sim q(\mathcal{X}_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t} \|\epsilon - \epsilon_\Theta(\mathcal{X}_t^{\mathcal{M}_0}, t \mid \epsilon_t^{\mathcal{M}_1}, p)\|^2. \tag{11}$$

Once trained, we can utilize the diffusion model to infer the masked values given a random Gaussian noise $\mathcal{X}_T^{\mathcal{M}_0}$, as well as the forward noise sequence added to the unmasked data $\epsilon_{1:T}^{\mathcal{M}_1}$.

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**

제안 방법론 중 Diffusion model 에서의 condition 유무 (noise 를 masked input 에 주는지 여부)



$$p(\mathcal{X}^{\mathcal{M}_0} \mid \mathcal{X}^{\mathcal{M}_1})$$

$$p(\mathcal{X}^{\mathcal{M}_0} \mid \epsilon_{1:T}^{\mathcal{M}_1})$$

Figure 2: Example cases of conditional/unconditional diffusion models for time series anomaly detection.

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection**

Table 2: The Precision (P), Recall (R), F1 and R-AUC-ROC of all anomaly detectors on benchmark datasets. The average values of P, R, F1 and R-AUC-ROC were calculated from 6 individual runs, while F1-std. is the standard deviation across the 6 runs.

Method	SMD					PSM					SWaT				
	P	R	F1	F1-std.	R-AUC-PR	P	R	F1	F1-std.	R-AUC-PR	P	R	F1	F1-std.	R-AUC-PR
IForest	0.2030	0.2130	0.1799	0.0138	0.0257	0.6630	0.4919	0.5641	0.0070	0.2058	0.9764	0.6650	0.7907	0.0020	0.0685
BeatGAN	0.9013	0.8894	0.8797	0.0058	0.3200	0.9204	0.8767	0.8975	0.0178	0.3453	0.9606	0.7020	0.8107	0.0022	0.3215
LSTM-AD	0.3361	0.3229	0.2639	0.0123	0.0399	0.9050	0.7707	0.8313	0.0036	0.2561	**0.9925**	0.6737	0.8026	0.0013	0.3118
InterFusion	0.8815	0.9071	0.8772	0.0226	0.3012	0.9533	0.9128	0.9326	0.0036	0.1896	0.8683	0.8530	0.8600	0.0309	0.1477
OmniAnomaly	0.8751	0.9052	0.8775	0.0083	0.2525	0.9551	0.8859	0.9191	0.0060	0.3718	0.9749	0.7500	0.8470	0.0271	**0.3722**
GDN	0.8460	0.7862	0.7865	0.0109	0.1637	0.8750	0.8385	0.8564	**0.0000**	0.3230	0.1311	0.0585	0.0808	**0.0009**	0.1318
MAD-GAN	0.8851	0.9045	0.8803	0.0384	0.2295	0.8596	0.8838	0.8698	0.0339	0.4416	0.7918	0.5423	0.6385	0.3048	0.2633
MTAD-GAT	0.8836	0.8330	0.8463	0.0316	0.3006	0.8763	0.8725	0.8744	0.0000	0.4116	0.8468	0.8224	0.8344	0.0067	0.3196
MSCRED	0.8567	0.9038	0.8426	**0.0002**	0.2601	0.9555	0.6857	0.7965	0.0102	0.3846	0.4823	0.4065	0.4407	0.3408	0.1668
TranAD	0.8906	0.8982	0.8785	0.0023	0.2941	0.9506	0.8951	0.9220	0.0045	0.3994	0.7025	0.7266	0.6886	0.1089	0.1670
ImDiffusion	**0.9520**	**0.9509**	**0.9488**	0.0039	**0.3821**	**0.9811**	**0.9753**	**0.9781**	0.0072	**0.4711**	0.8988	**0.8465**	**0.8709**	0.0124	0.1939

Method	SMAP					MSL					GCP				
	P	R	F1	F1-std.	R-AUC-PR	P	R	F1	F1-std.	R-AUC-PR	P	R	F1	F1-std.	R-AUC-PR
IForest	0.2886	0.7671	0.4163	0.0026	0.1096	0.6059	0.5328	0.5334	0.0309	0.0942	0.8055	0.7385	0.7370	0.0120	0.1558
BeatGAN	0.8915	0.6781	0.7663	0.0162	**0.1303**	0.7782	0.8512	0.8102	0.0342	0.1421	0.9865	0.9630	0.9717	0.0074	0.2414
LSTM-AD	0.7841	0.5630	0.6533	0.0382	0.1099	0.7330	0.5745	0.6378	0.1473	0.1066	0.9591	0.9575	0.9553	0.0013	0.2610
InterFusion	0.8788	0.7704	0.8204	0.0077	0.1457	0.7688	**0.9464**	0.8442	0.0330	0.1083	0.9361	0.9720	0.9092	0.0005	0.2846
OmniAnomaly	0.8407	**0.9674**	0.8995	0.0078	0.0978	0.8321	0.8125	0.8221	0.0121	0.1290	0.9572	0.9796	0.9668	0.0027	0.2029
GDN	0.9689	0.5401	0.6936	0.0037	0.0961	0.8668	0.8072	0.8360	0.0004	0.1295	0.9648	0.9628	0.9589	0.0011	0.2096
MAD-GAN	0.9547	0.5474	0.6952	0.0013	0.0990	0.7047	0.7841	0.7423	**0.0000**	0.1301	0.9766	0.9558	0.9605	0.0055	0.1867
MTAD-GAT	**0.9718**	0.5259	0.6824	**0.0012**	0.1083	0.7321	0.7616	0.7432	0.0200	0.1278	0.9490	0.9523	0.9461	0.0047	0.2210
MSCRED	0.4107	0.8604	0.2712	0.0625	0.1042	0.5008	0.6088	0.4899	0.0788	0.1090	0.9754	0.9735	0.9712	**0.0006**	0.2068
TranAD	0.8224	0.8502	0.8360	0.0090	0.1077	**0.8951**	0.9297	**0.9115**	0.0051	0.1057	0.9472	0.9812	0.9631	0.0030	0.2026
ImDiffusion	0.8771	0.9618	**0.9175**	0.0095	0.1105	0.8930	0.8638	0.8779	0.0152	**0.2381**	**0.9771**	**0.9825**	**0.9774**	0.0014	**0.3957**

Methods

- [VLDB 2024] **(ImDiffusion) Imputed Diffusion Models for Multivariate Ti**
  Detection

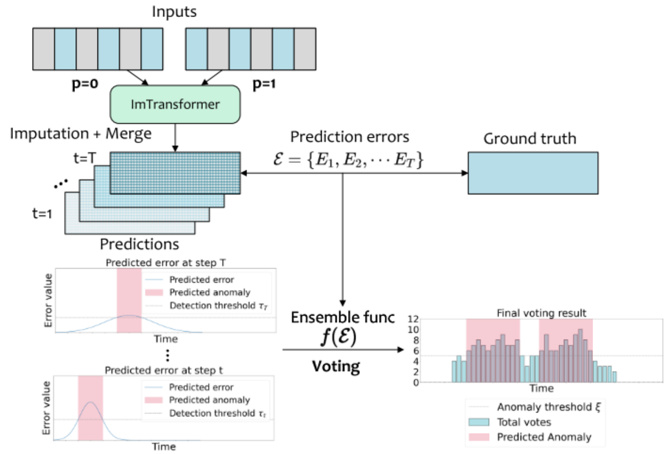
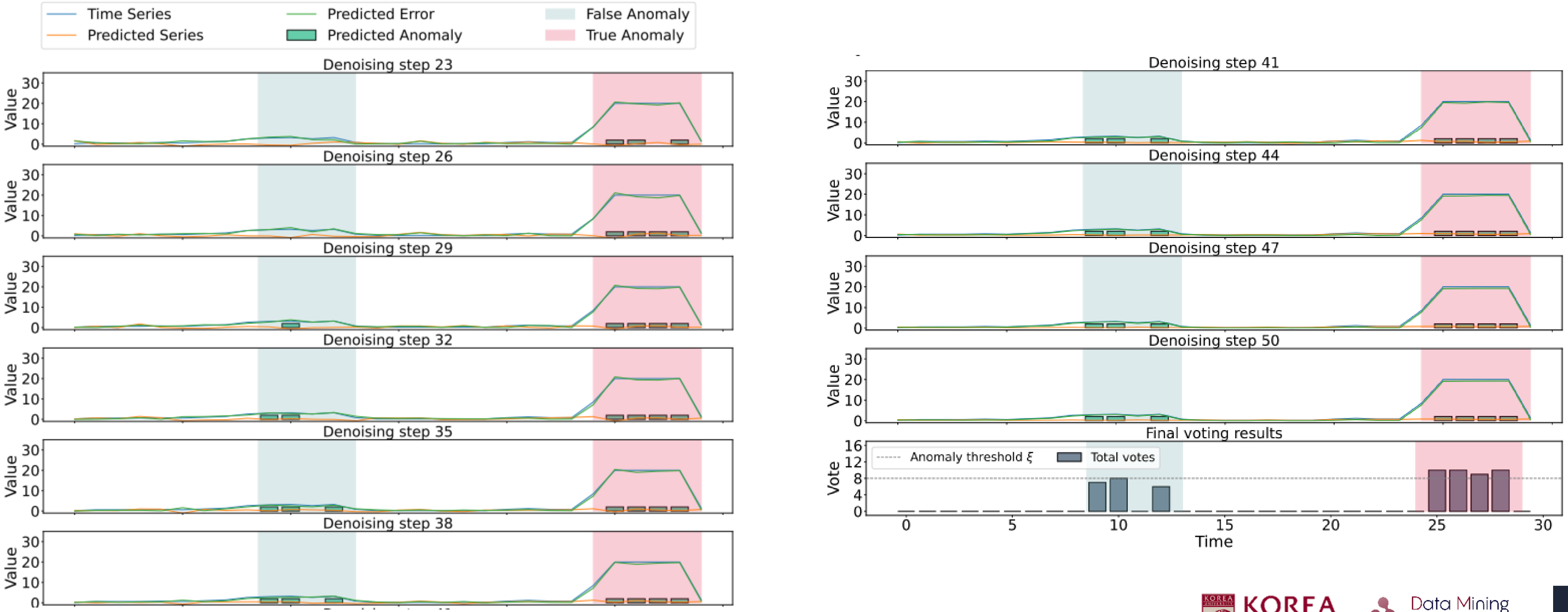


Figure 6: The ensemble anomaly inference of ImDiffusion.



Conclusion

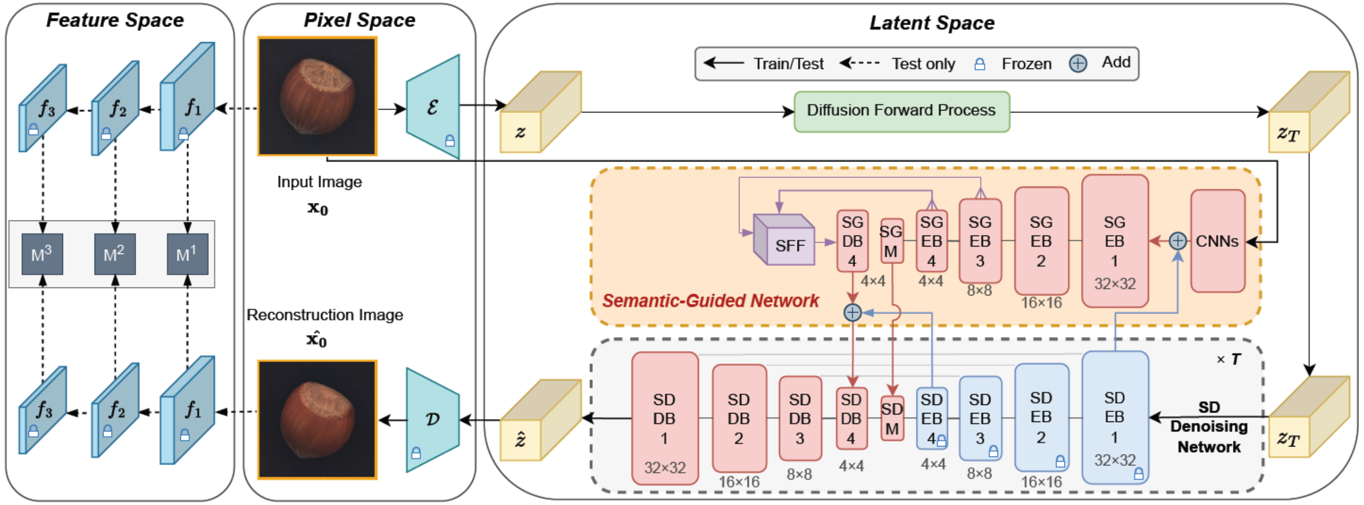
- **Conclusion**

이미지 데이터셋에서 diffusion anomaly detection 과의 차이점

① 모델의 구조가 다름 (U-Net vs Transformer or 4S-based U-Net)

② Input 의 정보를 반영할 때, image 는 대부분 feature map 형태로 condition 을 주는데, time series 는 imputation 을 통해 어느정도 input 의 정보를 raw 단계에서 보존하는 식으로 많이 활용하는 것으로 보임.

③ Diffusion model 이 image 에서는 multi-class anomaly detection 일 때 좋은 성능을 보였고, time series 에서는 imputation 구조 및 모델 설계를 어떻게 하냐에 따라 성능 차이가 나는 것으로 보임. (step 수의 차이도 큼 1000 vs 50, 100)



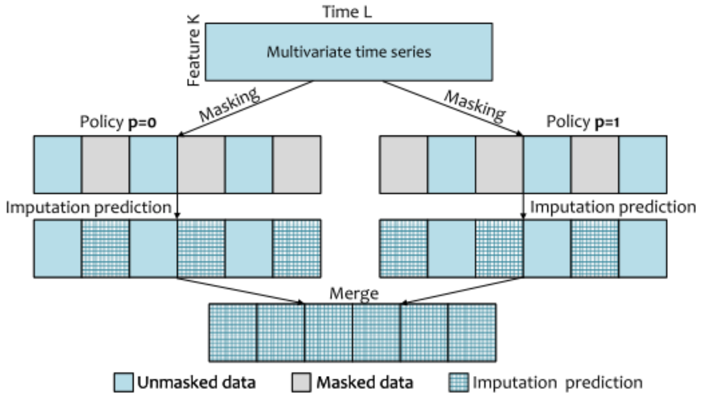


Figure 3: An illustration of the the grating masking and the imputation process under this strategy.

He, H., Zhang, J., Chen, H., Chen, X., Li, Z., Chen, X., ... & Xie, L. (2024, March). A diffusion-based framework for multi-class anomaly detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 8, pp. 8472-8480).

Thank you!

Seokho Moon

danny232@korea.ac.kr

School of Industrial and Management Engineering, Korea University