

USAD : UnSupervised Anomaly Detection on Multivariate Time Series, KDD 2020

Data analysis programming paper review

Kyonggi Univ. Smart I.O.T Lab 이상민 (2021년 12월 09일)

01. Introduction

02. Methods

03. Experiments and Results

01. Introduction

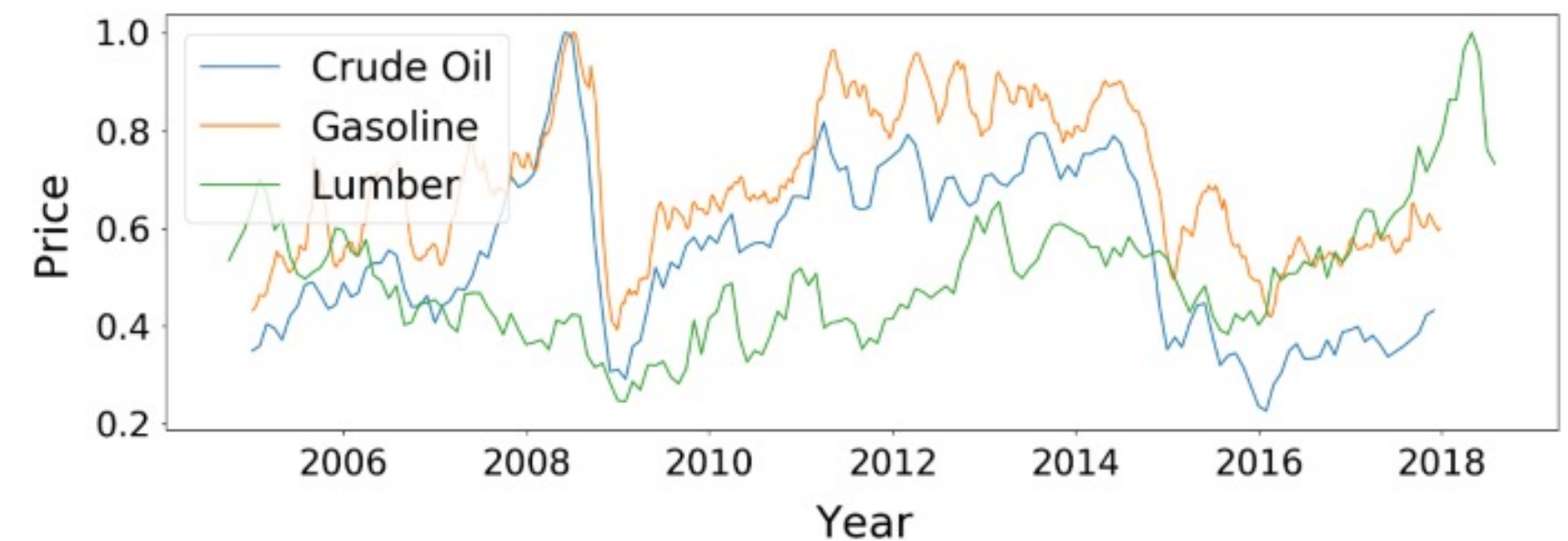
Background

기존 관측과는 상이하여 다른 매커니즘에 의해 생성되었다고 판단할만한 관측값

Unsupervised Anomaly Detection on Multivariate Time Series.

↓
Auto encoder (AE)
Generative adversarial networks (GANs)

↓
각 시간 단위마다 여러 개의 값을 가지는 시계열

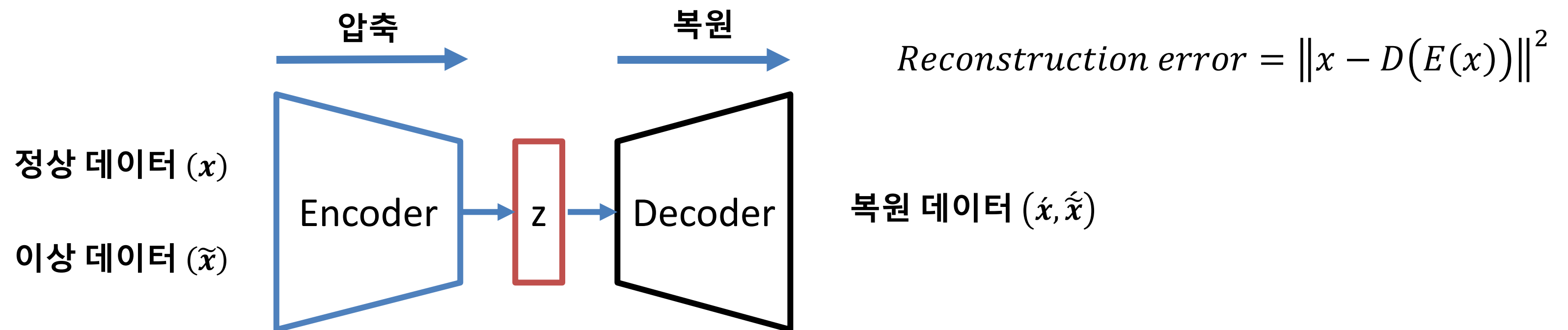


이미지 출처 : <https://link.springer.com/article/10.1007/s10994-019-05815-0?shared-article-renderer>

01. Introduction

Unsupervised anomaly detection example

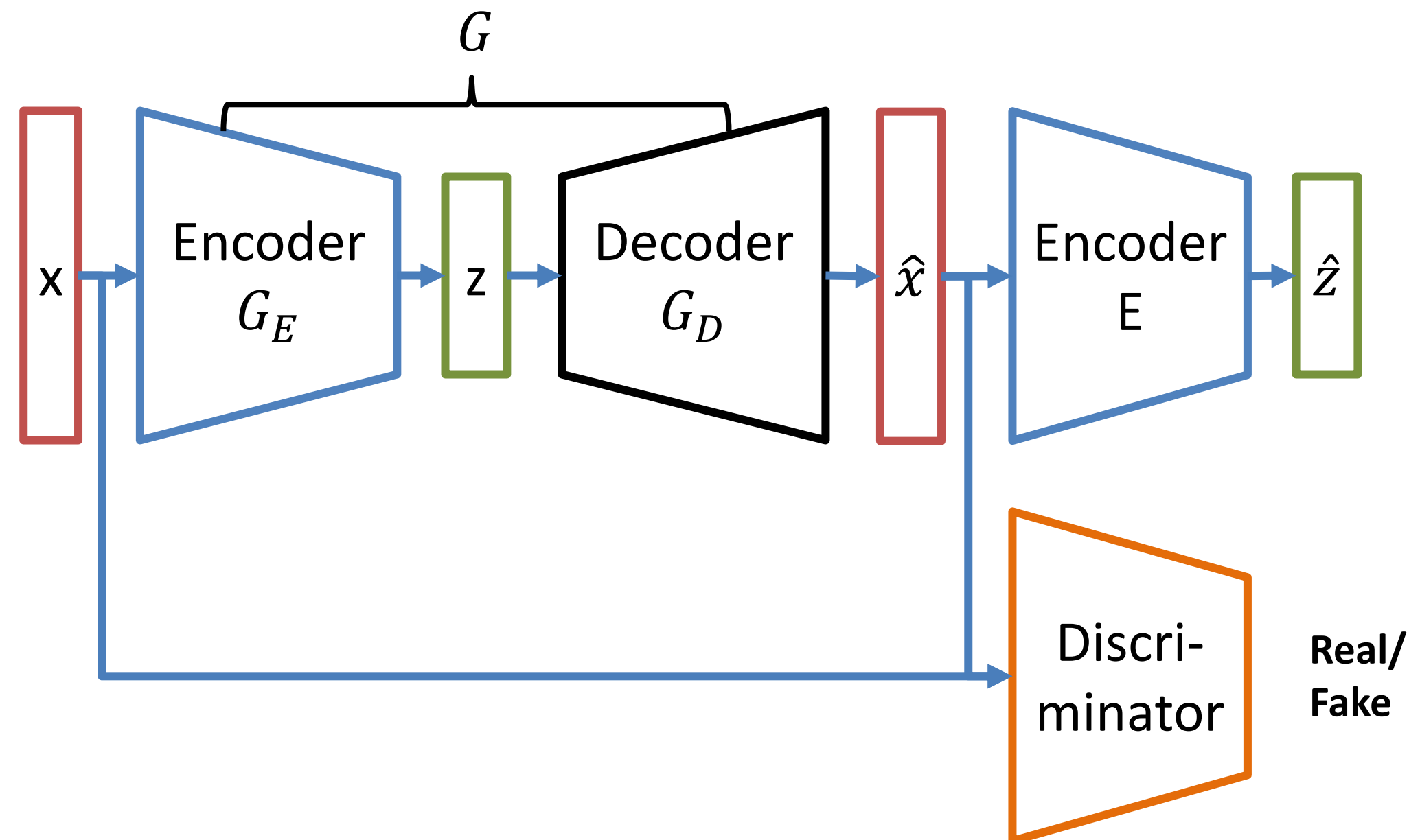
- AE를 활용한 anomaly detection은 데이터를 압축&복원하는 과정에서 발생하는 reconstruction error를 anomaly score로 사용
- 학습에는 정상 데이터만을 사용하기 때문에 이상 데이터가 들어올 경우 큰 reconstruction error가 발생
- 다만 AE는 압축 과정에서 복원에 불필요한 정보를 제거하여 비정상을 탐지할 수 있는 abnormal information이 사라질 가능성이 존재
- 정상 분포와 유사한 비정상 데이터가 들어올 경우 이를 구별할 수 없음 (= 최대한 정상처럼 복원하기 때문)



01. Introduction

Unsupervised anomaly detection example

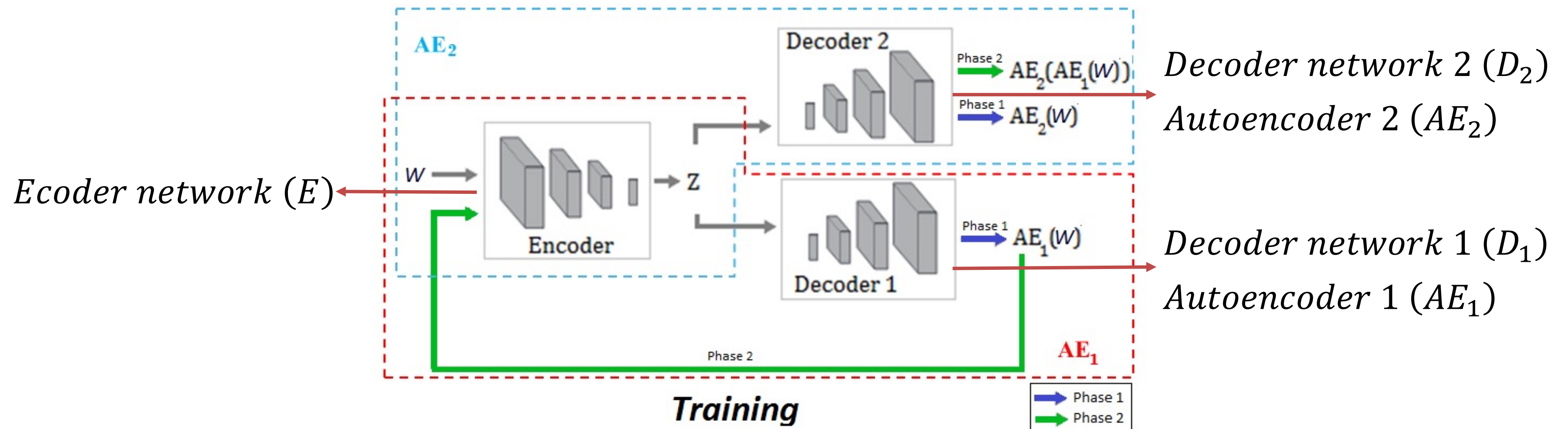
- GANs based 방법의 경우 fake (abnormal)와 real (normal)을 구분하는 discriminator의 도입으로 정상 데이터만을 활용하더라도 보다 상세하게 비정상을 구분할 수 있음(=abnormal information을 포함)
- 판별자를 속이기 위해서는 real 정보 뿐 아니라 fake에 대한 정보를 포함하도록 인코더 및 디코더가 학습하기 때문
- 다만 GANs는 안정적인 학습이 어렵다는 단점이 존재함



02. Methods

USAD architecture

- USAD는 학습이 쉬운 AE의 장점과 abnormal information을 강제할 수 있는 GANs의 장점을 결합한 모델임
- AE를 사용하되 adversarial training을 적용하여 보다 상세한 anomaly detection을 추구하고자 함
- USAD의 architecture는 다음과 같음



$$AE_1 = D_1(E(W)), \quad AE_2 = D_2(E(W))$$

02. Methods

USAD training process

- USAD is trained in two phases

1. Auto encoder training

$$L_{AE_1} = \|W - AE_1(W)\|_2$$

$$L_{AE_2} = \|W - AE_2(W)\|_2$$

2. Adversarial training

- Train AE_2 to distinguish the real data from the data coming from AE_1 , and train AE_1 to fool AE_2
- The objective of AE_1 is to minimize the difference between W and the output of AE_2

$$\min_{AE_1} \|W - AE_2(AE_1(W))\|_2$$

- The objective of AE_2 is to maximize this difference

$$\max_{AE_2} \|W - AE_2(AE_1(W))\|_2$$

02. Methods

USAD training process

실제 데이터의 재구성 오류 가짜 데이터에 대한 AE_2 의 재구성 오류

$$L_{AE_1} = \frac{1}{n} \|W - AE_1(W)\|_2 + \left(1 - \frac{1}{n}\right) \|W - AE_2(AE_1(W))\|_2$$

$$L_{AE_2} = \frac{1}{n} \|W - AE_2(W)\|_2 - \left(1 - \frac{1}{n}\right) \|W - AE_2(AE_1(W))\|_2$$

n : Training epochs

$AE_1(W)$: Fake (abnormal)

- AE_1 의 경우 real에 대한 reconstruction error와 fake에 대한 AE_2 의 reconstruction error 둘 모두 최소일 때 최소
- AE_2 의 경우 real에 대한 reconstruction error가 최소이고 fake에 대한 AE_2 의 reconstruction error가 최대일 때 최소
- 정리하면 AE_1 는 AE_2 가 fake와 real을 구분하지 못하게 만들고, AE_2 는 fake이 들어왔을 때 reconstruction error를 크게 만들도록 학습
- 즉 AE_2 는 정상 데이터와 비정상 데이터의 미세한 차이를 극대화 시키는 역할을 함

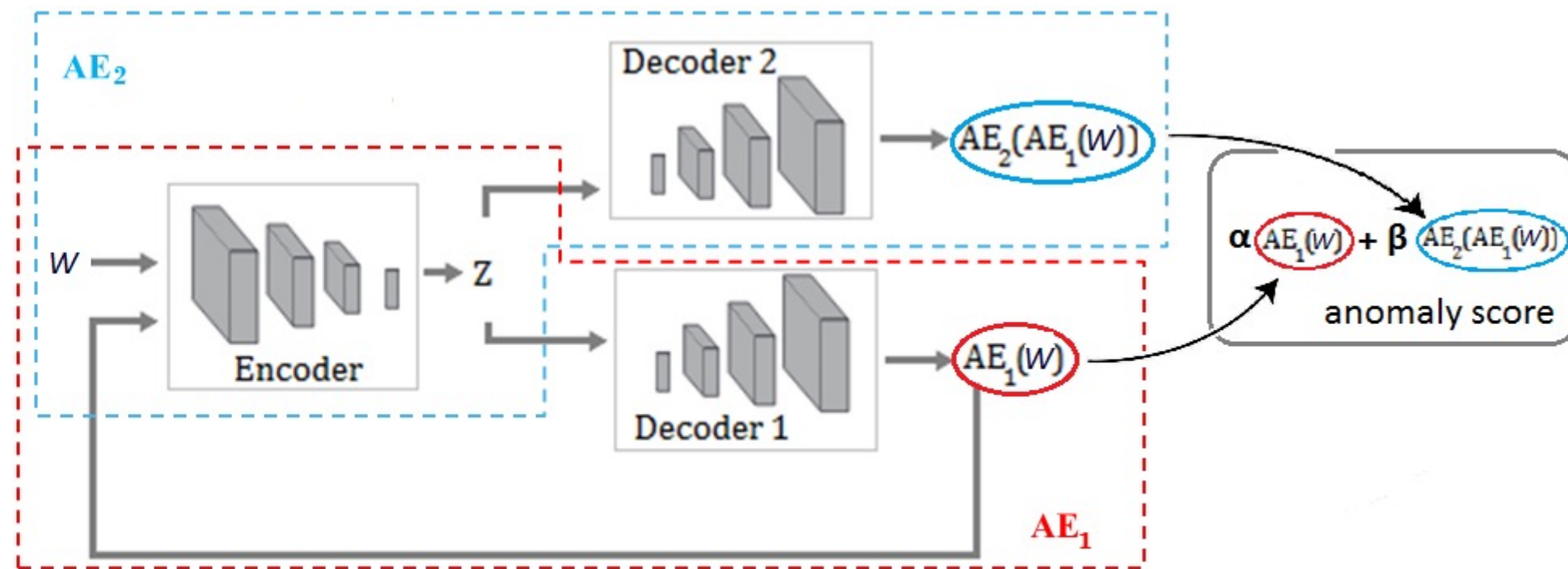
02. Methods

USAD detection process

- 위와 같이 학습된 USAD의 anomaly score 산출 식은 다음과 같음

$$\mathcal{A}(\hat{W}) = \alpha \|\hat{W} - AE_1(\hat{W})\|_2 + \beta \|\hat{W} - AE_2(AE_1(\hat{W}))\|_2$$

- 정상과 매우 유사한 비정상 데이터가 들어왔을 때도 USAD는 이를 탐지할 수 있음



Detection

[USAD detection process]

Parameter setting	# of detection	Detection sensitivity
$\alpha > \beta$	Reduce	Low
$\alpha < \beta$	Increase	High

[Variation by parameter setting]

03. Experiments and Results

Experimental setup

5개의 공공 Dataset 활용

- Secure water treatment (SWaT) - 여과수를 생산하는 실제 산업용 수처리 공장 11일간 연속 운영(정상 7일 비정상 4일)
- Water distribution (WADI) - SWaT의 확장, 물 분배 데이터 세트, 16일간 연속 운영(정상 14일, 비정상 2일)
- Server machine dataset (SMD) - 대규모 인터넷 회사에서 수집한 5주간의 데이터 세트, 28개의 서버 시스템에서 가져온 데이터 크기가 동일한 두 개의 집합으로 나뉘며, 각각 훈련 세트, 테스트 세트.
- Soil moisture active passive (SMAP) - 화성 토양 수분 데이터 세트
- Mars science laboratory (MSL) - 화성 탐사 데이터 세트

Feasibility study : Orange's dataset

웹 사이트에 있는 Orange 광고 네트워크의 기술 및 비즈니스 지표
약 32일에 해당하는 학습 세트와 약 60일에 해당하는 테스트 세트
훈련 세트는 회사에 큰 사고가 없는 연속된 날을 선택.

테스트 세트는 중요한 기간에 해당하는 60일을 선택, 이상 현상은 사고 보고서를 기반으로 전문가가 레이블 지정.

Dataset	Train	Test	Dimensions	Anomalies (%)
SWaT	496800	449919	51	11.98
WADI	1048571	172801	123	5.99
SMD	708405	708420	28*38	4.16
SMAP	135183	427617	55*25	13.13
MSL	58317	73729	27*55	10.72
Orange	2781000	5081000	33	33.72

오랑주 S.A.	
	
원어	Orange S.A.
ISIN	FR0000133308
산업 분야	통신 및 서비스업
창립	1988년
본사 소재지	프랑스 파리 (파리 15구)

03. Experiments and Results

Evaluation Metrics

평가 지표

- Precision (정밀도, P), Recall (재현율, R), F1 score (F1), F1*

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F1 = \frac{P * R}{P + R}, F1^* = 2 * \frac{\bar{P} * \bar{R}}{\bar{P} + \bar{R}}$$

※ \bar{P}, \bar{R} 은 평균 정밀도와 평균 재현율을 의미

TP는 참 긍정, FP는 거짓 긍정, FN은 거짓 부정.

완전성을 기하기 위해 저자는 평균 정밀도와 평균 재현율을 사용하여 F1 점수를 계산하여 측정 값을 보고하며, 이 점수를 $F1^*$ 으로 표시

03. Experiments and Results

Results

- Point-adjust : detect each observation/time-point independently and assigns a label to single time-point (without/with)
: 각 관찰/시점을 독립적으로 감지하고 단일 시점에 레이블을 할당(미포함/포함).

Methods	SWaT						WADI					
	Without			With			Without			With		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
AE	0.9903	0.6295	0.7697	0.9913	0.7040	0.8233	0.9947	0.1310	0.2315	0.3970	0.3220	0.3556
IF	0.9512	0.5884	0.7271	0.9620	0.7315	0.8311	0.2992	0.1583	0.2071	0.6241	0.6155	0.6198
LSTM-VAE	0.9897	0.6377	0.7756	0.7123	0.9258	0.8051	0.9947	0.1282	0.2271	0.4632	0.3220	0.3799
DAGMM	0.4695	0.6659	0.5507	0.8292	0.7674	0.7971	0.0651	0.9131	0.1216	0.2228	0.1976	0.2094
OmniAnomaly	0.9825	0.6497	0.7822	0.7223	0.9832	0.8328	0.9947	0.1298	0.2296	0.2652	0.9799	0.4174
USAD	0.9851	0.6618	0.7917	0.9870	0.7402	0.8460	0.9947	0.1318	0.2328	0.6451	0.3220	0.4296

Methods	SMD				SMAP				MSL			
	P	R	F1	F1*	P	R	F1	F1*	P	R	F1	F1*
AE	0.8825	0.8037	0.8280	0.8413	0.7216	0.9795	0.7776	0.8310	0.8535	0.9748	0.8792	0.9101
IF	0.5938	0.8532	0.5866	0.7003	0.4423	0.5105	0.4671	0.4739	0.5681	0.6740	0.5984	0.6166
LSTM-VAE	0.8698	0.7879	0.8083	0.8268	0.7164	0.9875	0.7555	0.8304	0.8599	0.9756	0.8537	0.9141
DAGMM	0.6730	0.8450	0.7231	0.7493	0.6334	0.9984	0.7124	0.7751	0.7562	0.9803	0.8112	0.8537
OmniAnomaly	0.9809	0.9438	0.9441	0.9620	0.7585	0.9756	0.8054	0.8535	0.9140	0.8891	0.8952	0.9014
USAD	0.9314	0.9617	0.9382	0.9463	0.7697	0.9831	0.8186	0.8634	0.8810	0.9786	0.9109	0.9272

03. Experiments and Results

Results

- 알고리즘 별 전체 datasets의 평균 성능(\pm standard deviation)은 다음과 같음

	P	R	F1	F1*
AE	0.77(0.21)	0.76(0.24)	0.73(0.19)	0.86 (0.04)
IF	0.64(0.17)	0.68(0.11)	0.62(0.12)	0.60 (0.09)
LSTM-VAE	0.72(0.15)	0.80(0.25)	0.75 (0.18)	0.86 (0.04)
DAGMM	0.62(0.21)	0.76(0.29)	0.65(0.22)	0.79 (0.04)
OA	0.73(0.25)	0.95(0.04)	0.78(0.19)	0.91(0.04)
USAD	0.84(0.12)	0.80(0.25)	0.79(0.18)	0.91(0.04)

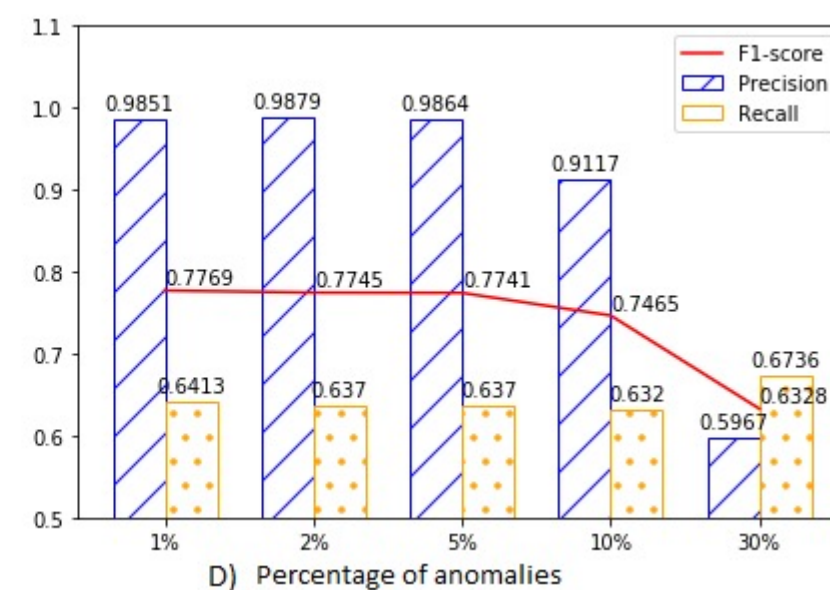
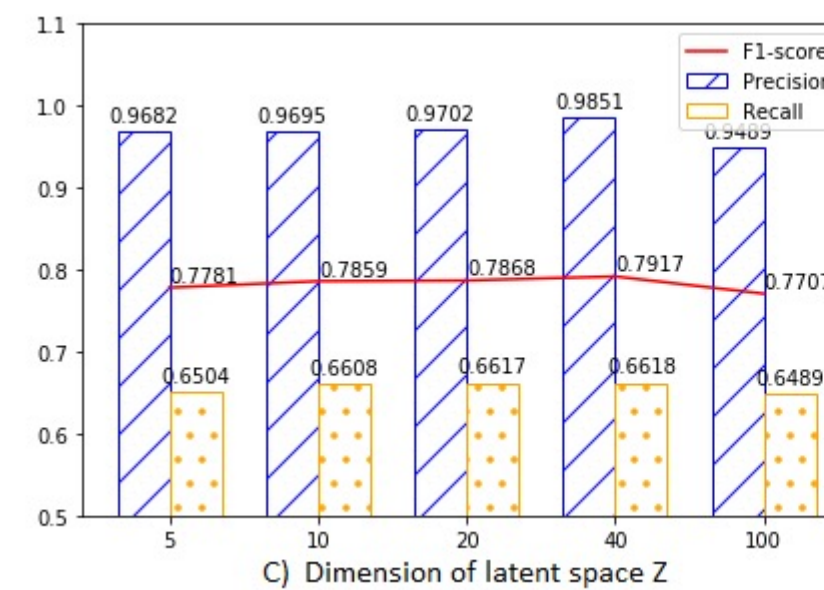
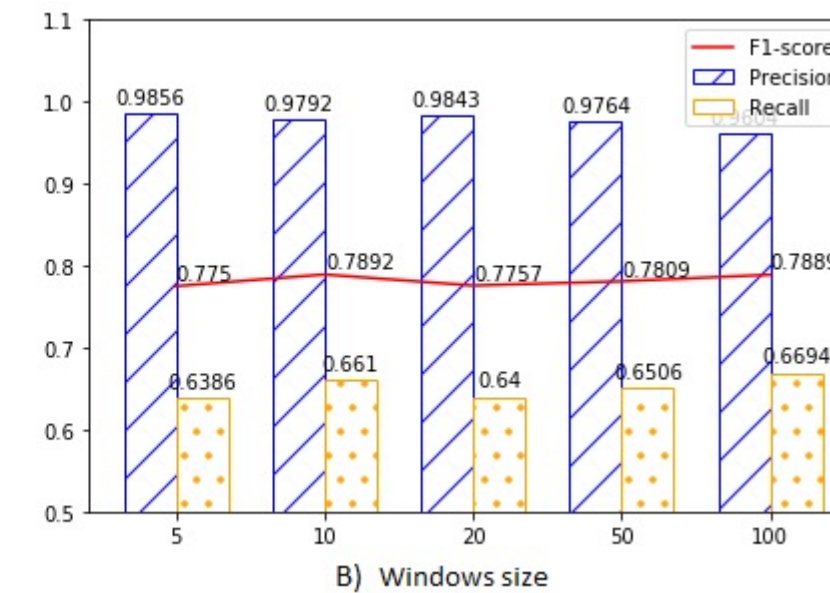
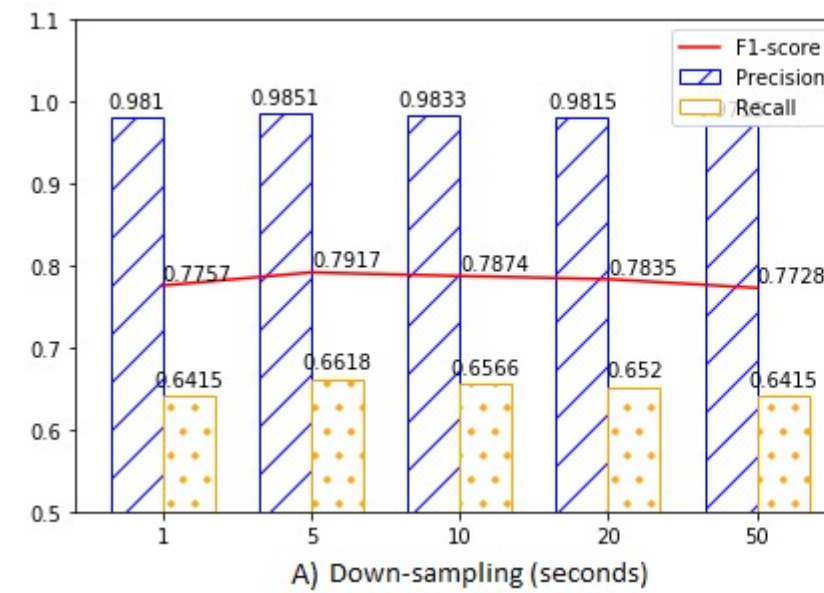
- α 와 β 에 따른 SWaT dataset의 성능은 다음과 같음

α	β	FP	TP	F1
0.0	1.0	604	35,616	0.7875
0.1	0.9	580	35,529	0.7853
0.2	0.8	571	35,285	0.7833
0.5	0.5	548	34,590	0.7741
0.7	0.3	506	34,548	0.7738
0.9	0.1	299	34,028	0.7684

03. Experiments and Results

Results

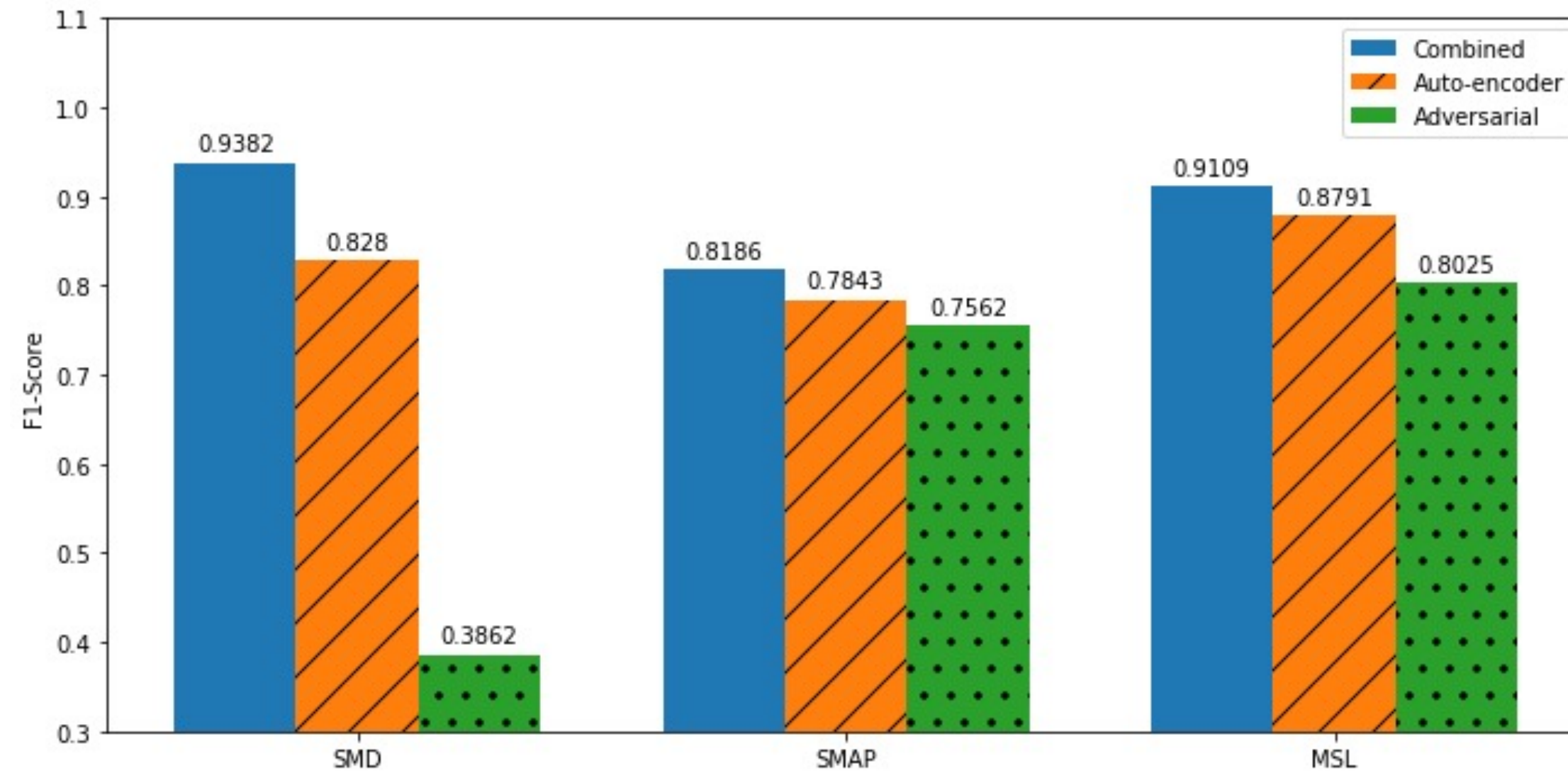
- Down-sampling / windows size / latent space's dimension / percentage of anomalies에 따른 ablation study의 결과



03. Experiments and Results

Results

- With/without adversarial training



감사합니다!