Anomaly detection using imaging of time series data

DMQA Open Seminar

2022.09.23

백민재

- **백민재(Min Jae Baek)**
  - 고려대학교 산업경영공학과 재학 중
  - Data Mining & Quality Analytics Lab(김성범 교수님 연구실)
  - 석사 과정(2022.03 ~)

- **연구 관심 분야**
  - Machine learning & Deep learning Algorithms
  - Multivariate Time Series Data

- **E-mail**
  - E-mail: minjae7306@korea.ac.kr

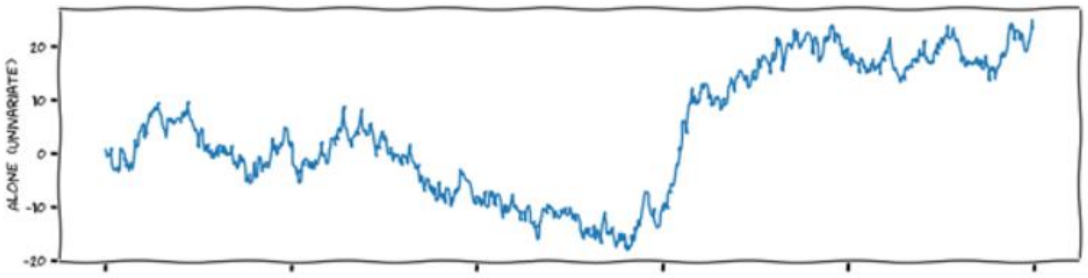
Contents

Introduction

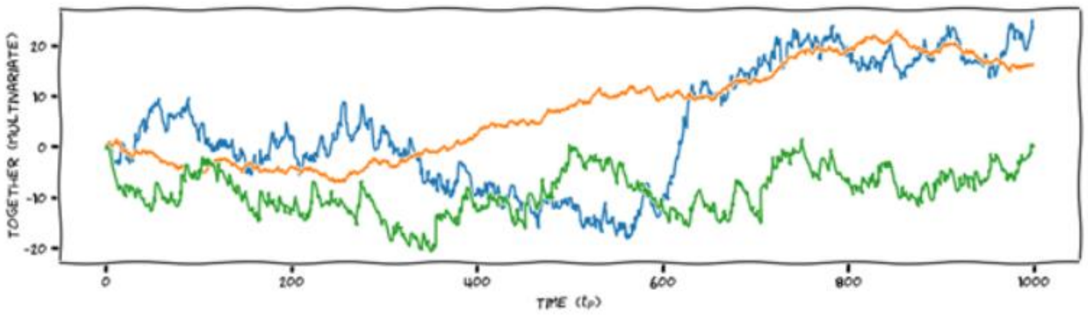
Introduction

❖ 시계열 데이터(Time series data)

- 일정한 시간 동안 수집 된 일련의 순차적으로 정해진 데이터 셋의 집합
  - ➢ 단변량 시계열(Univariate time series): 변수가 하나인 시계열 데이터
  - ➢ 다변량 시계열(Multivariate time series): 변수가 2개 이상인 시계열 데이터
- 발전소, 스마트 팩토리, 헬스케어 시스템 등 다양한 분야에서 실시간 센서 데이터 수집이 증가
- 다변량 시계열(Multivariate time series)을 이용한 이상탐지(Anomaly detection)의 수요가 증가



Univariate time series

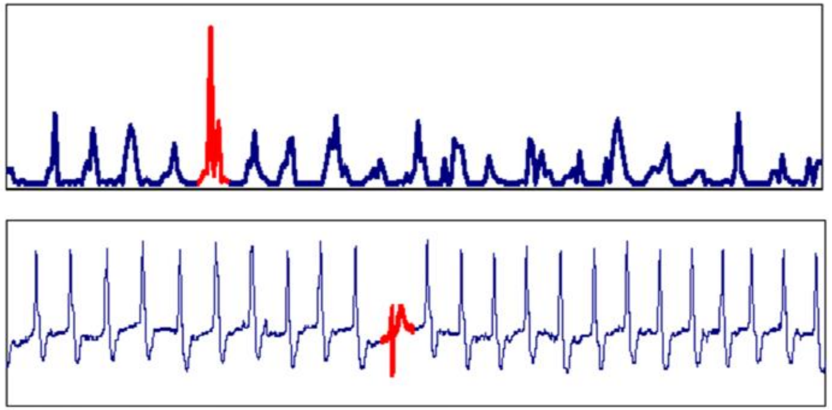


Multivariate time series

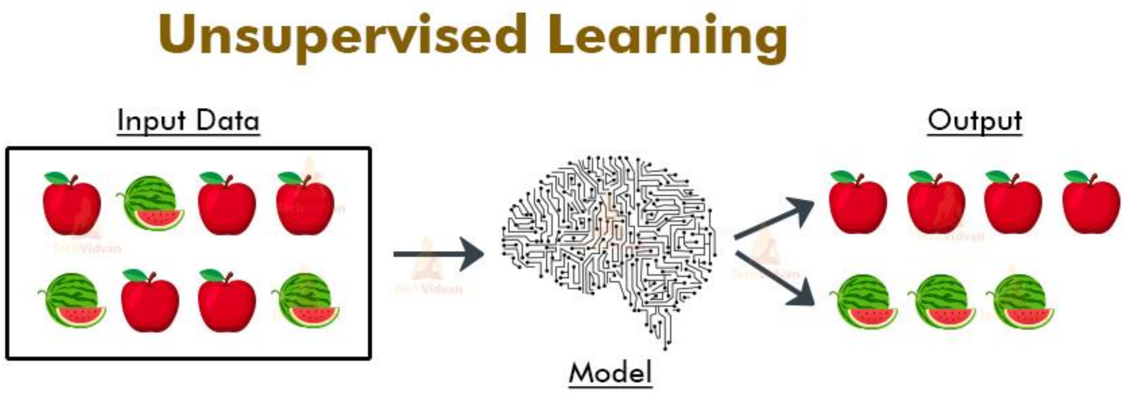
Introduction

❖ **시계열 데이터의 이상탐지(Anomaly detection of time series)**

- 시계열 데이터를 통해 이상 패턴 이나 이상징후 를 찾아내는 것을 의미
- 전문가를 통한 비정상 레이블링(labeling) 필요
- 정상데이터에 비해 비정상데이터의 양이 극히 적은 불균형 데이터로 구성
  - ➢ 정상 데이터만을 활용한 비지도 학습(Unsupervised learning) 방법을 적용



Anomaly detection of time series

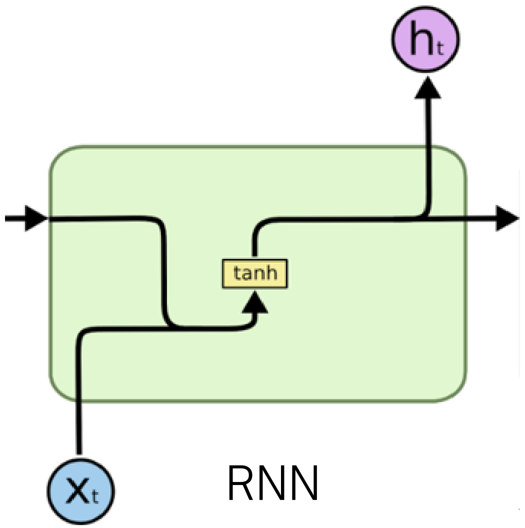


Unsupervised learning

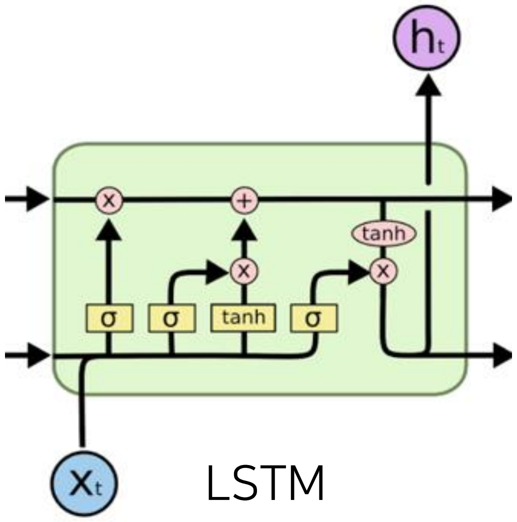
Introduction

❖ **시계열 데이터의 딥러닝(Deep learning)**

- 시간적 흐름을 고려하기 위해 Sequence 데이터를 다루는 RNN, LTSM 구조가 주로 사용
- 기본적으로 Fully connected 구조로 되어 있어 변수들간의 관계성 파악 어려움
- 변수가 증가 할 수록 계산의 복잡성이 증가
- Real world data(산업 데이터 등)에는 Noise가 존재하는데, 모델의 성능이 Noise의 영향을 많이 받음



RNN



LSTM

Anomaly detection using imaging of time series data

Anomaly detection using imaging of time series data

Paper

❖ A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data

- Zhang, Chuxu, et al. "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data." Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019.**(359회 인용)**

A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data

Chuxu Zhang[§]*, Dongjin Song[†]*, Yuncong Chen[†], Xinyang Feng[‡]*, Cristian Lumezanu[†], Wei Cheng[†], Jingchao Ni[†], Bo Zong[†], Haifeng Chen[†], Nitesh V. Chawla[§]

[§]University of Notre Dame, IN 46556, USA

[†]NEC Laboratories America, Inc., NJ 08540, USA

[‡]Columbia University, NY 10027, USA

[§]{czhang11,nchawla}@nd.edu, [†]{dsong,yuncong,lume,weicheng,jni,bzong,haifeng}@nec-labs.com, [‡]xf2143@columbia.edu

Abstract

Nowadays, multivariate time series data are increasingly collected in various real world systems, *e.g.*, power plants, wearable devices, *etc.* Anomaly detection and diagnosis in multivariate time series refer to identifying abnormal status in certain time steps and pinpointing the root causes. Building such a system, however, is challenging since it not only requires to capture the temporal dependency in each time series, but also need encode the inter-correlations between different pairs of time series. In addition, the system should be robust to noise and provide operators with different levels of anomaly scores based upon the severity of different incidents. Despite the fact that a number of unsupervised anomaly detection algorithms have been developed, few of them can jointly address these challenges. In this paper, we propose a Multi-Scale Convolutional Recurrent Encoder-Decoder (MSCRED), to perform anomaly detection and diagnosis in multivariate time series data. Specifically, MSCRED first constructs multi-scale

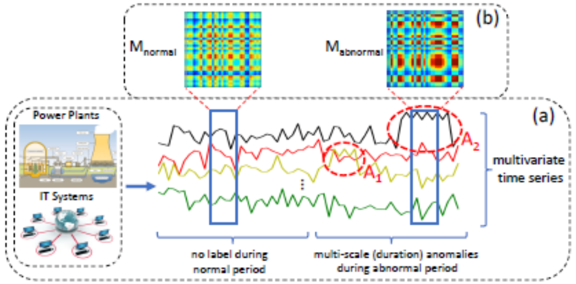


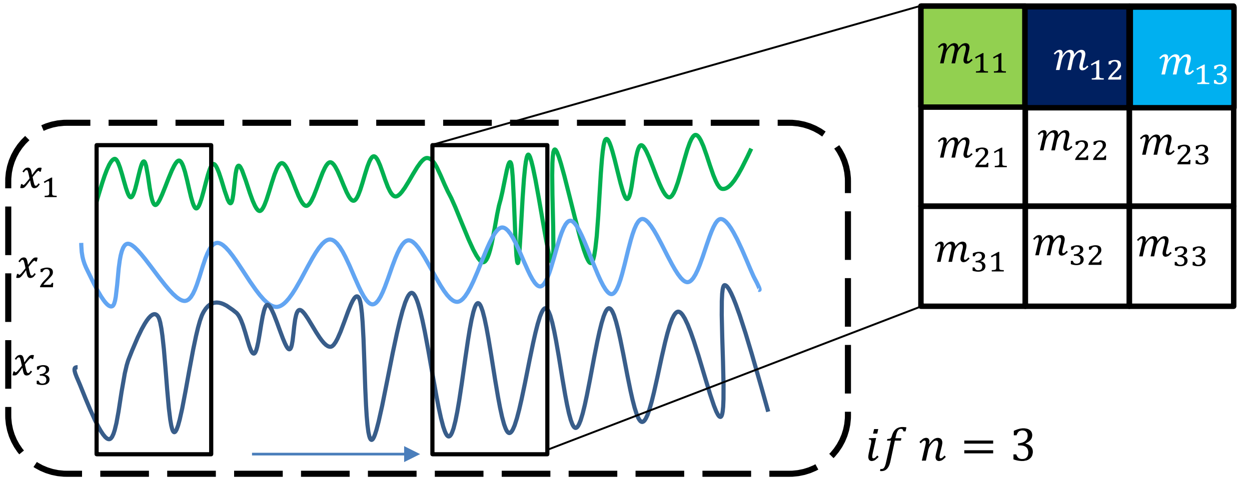
Figure 1: (a) Unsupervised anomaly detection and diagnosis in multivariate time series data. (b) Different system signature matrices between normal and abnormal periods.

A critical task in managing these systems is to detect anomalies in certain time steps such that the operators can take further actions to resolve underlying issues. For instance, an

MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

❖ Signature Matrix

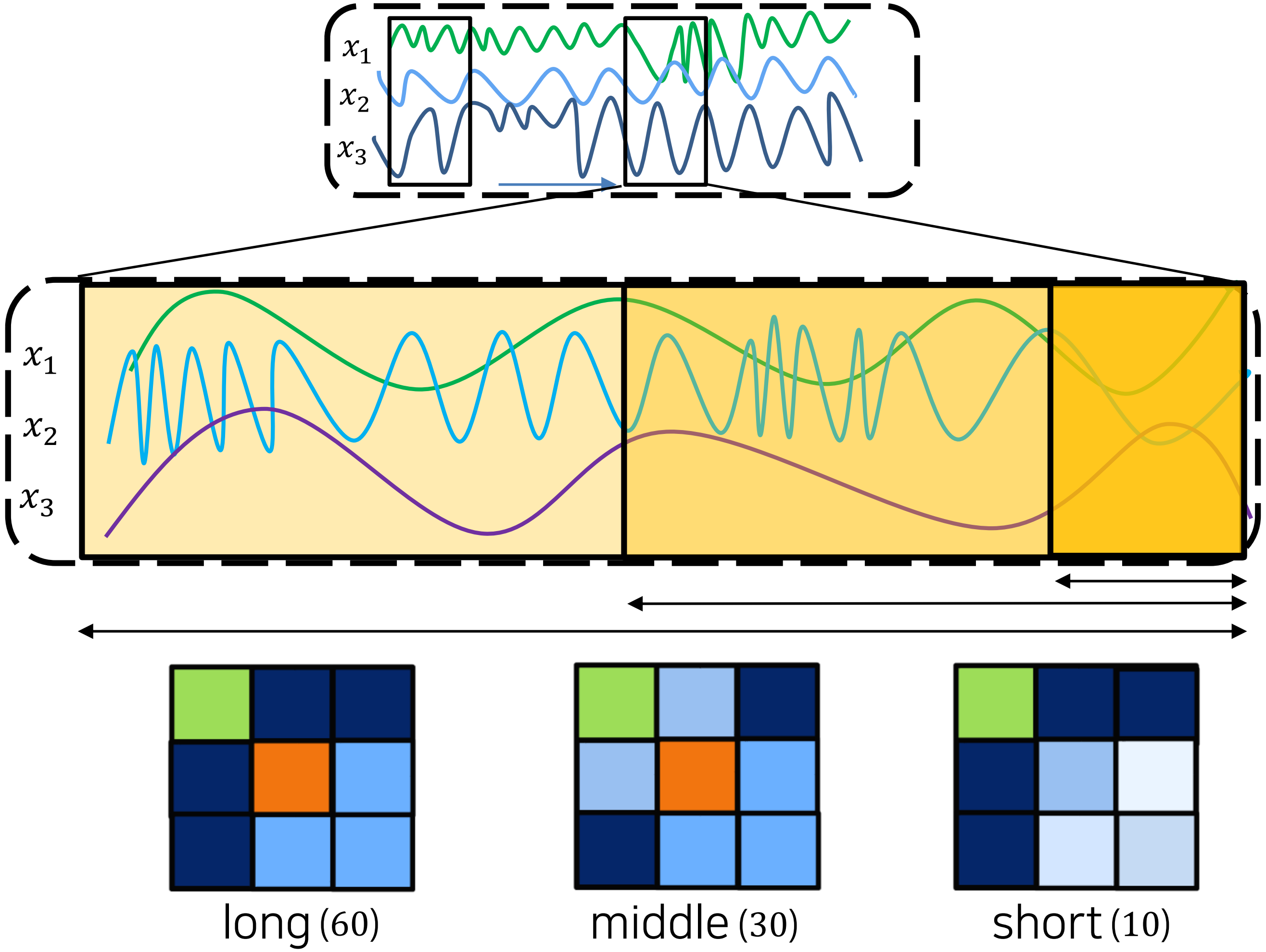
- $T: X = \{x_1, x_2, x_3, \cdots, x_n\}$ (n: $data$ 변수의 개수, T: 전체 Time 길이)
- $m_{ij}^t = \frac{1}{w}\sum_{\delta=0}^{w}(x_i^{t-\delta}x_j^{t-\delta})$ (w :window size, x_i : i th sensor data, $x_i^{t-\delta}$: $t-\delta$ 시점에서의 point)



- Time series data간의 correlation 을 반영
- Noise에 강건함을 가짐

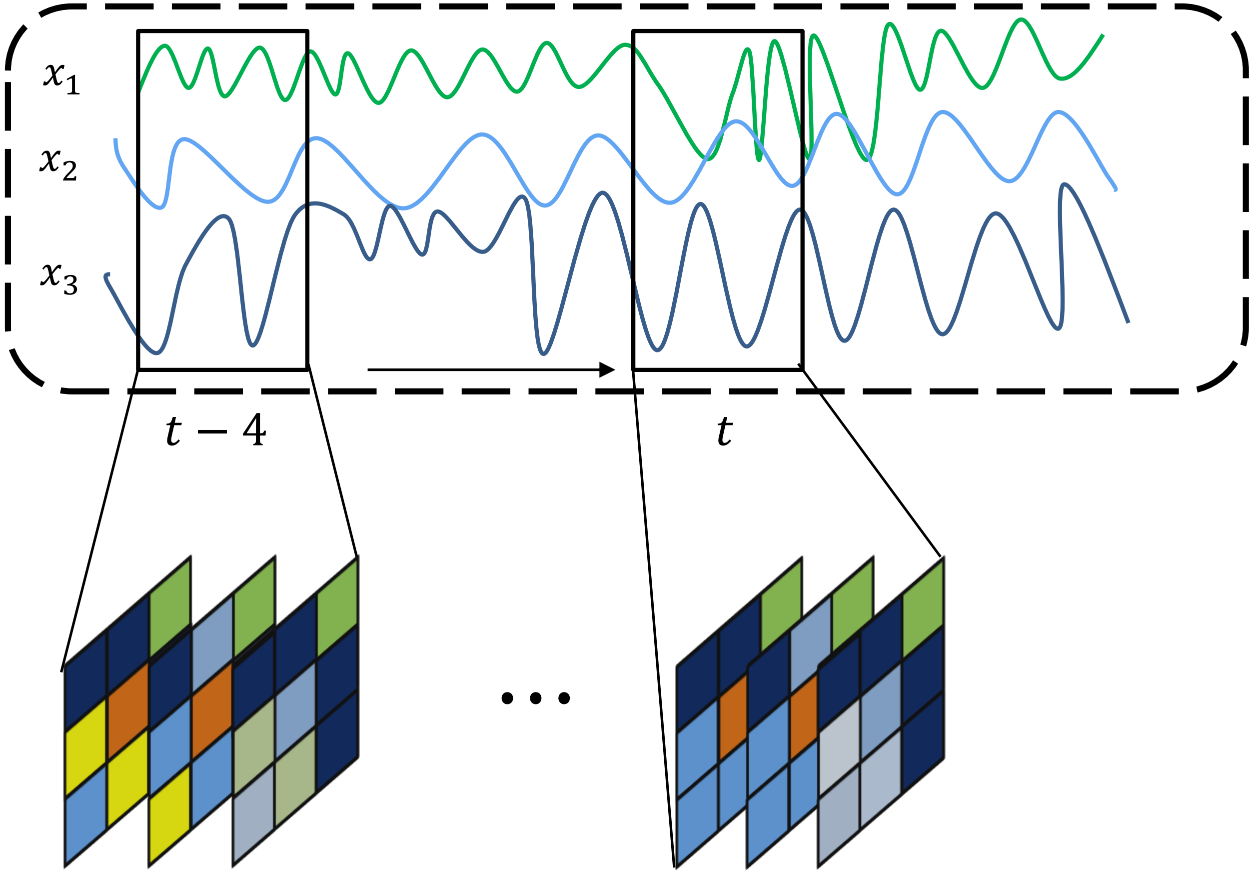
MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

❖ Signature Matrix



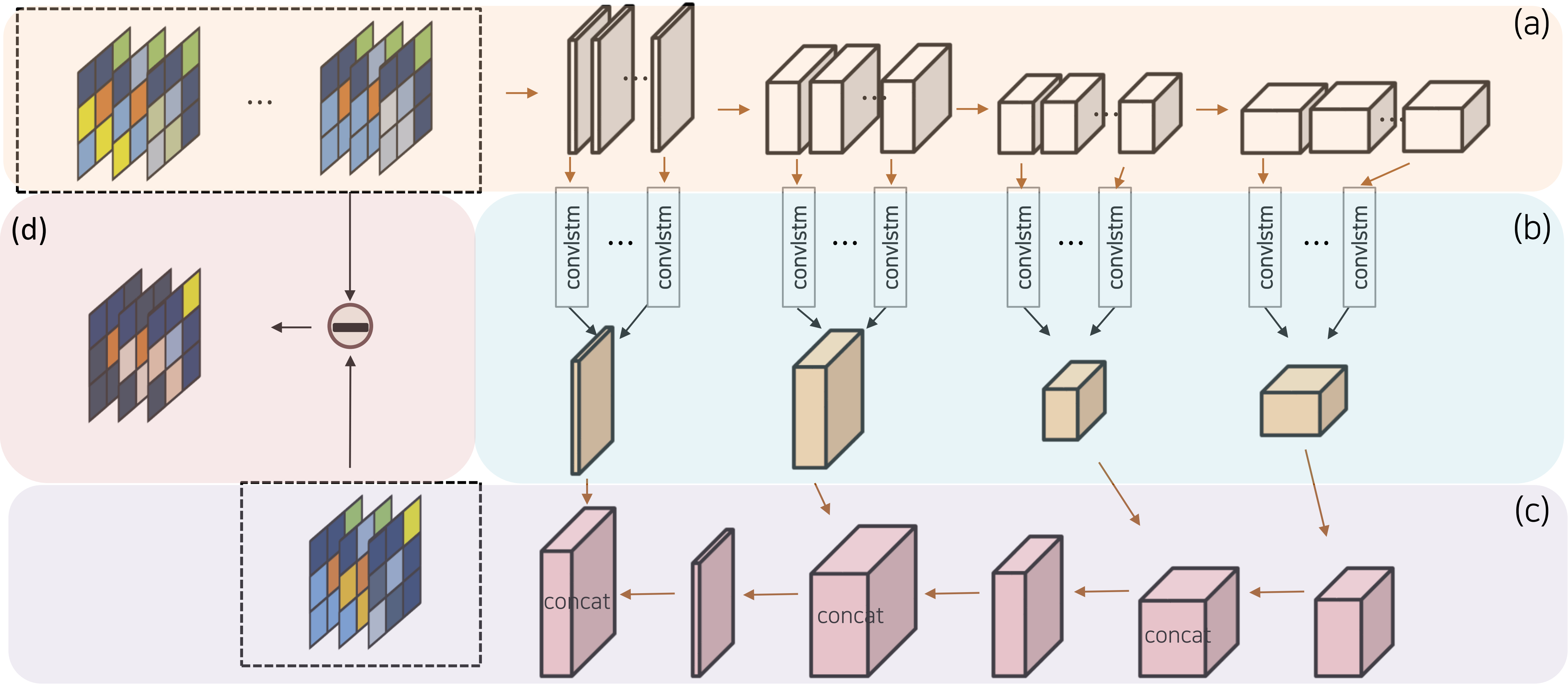
MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

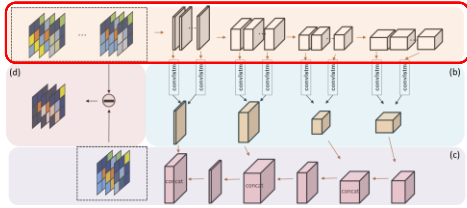
❖ Signature Matrix



MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

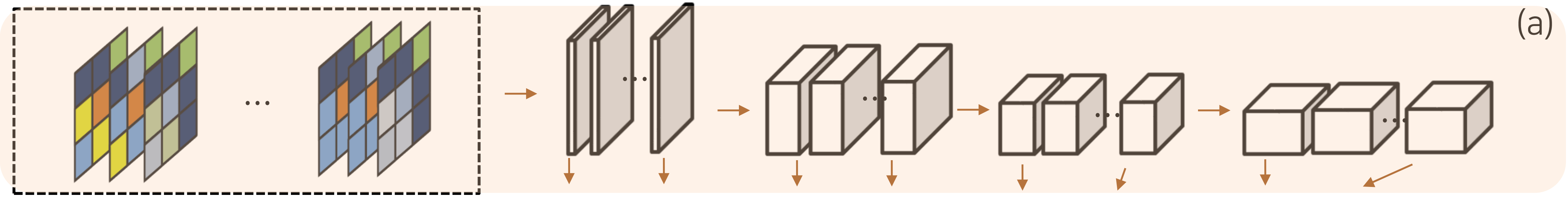
❖ MSCRED Framework





MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

❖ MSCRED Framework - Convolutional encoder



$(30 \times 30 \times 3) \times 5$ $(30 \times 30 \times 32) \times 5$ $(15 \times 15 \times 64) \times 5$ $(8 \times 8 \times 128) \times 5$ $(4 \times 4 \times 256) \times 5$

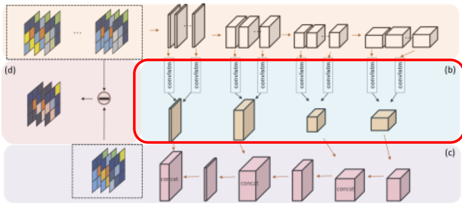
- Signature matrix를 input으로 사용
- Spatial pattern을 encoding 하기 위해서 Fully Convolutional encoder를 사용
- $\chi^{t,l} = f(W^l * \chi^{t,l-1} + b^l)$ $f(\ ): activation\ function,$

  $\chi^{t,l}: output\ feature\ map$

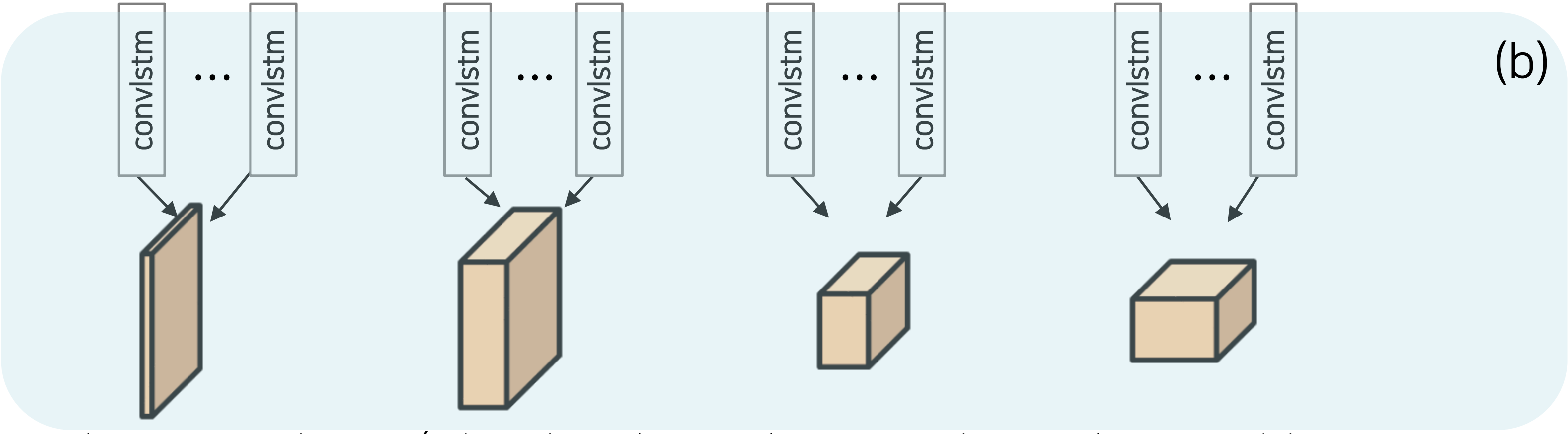
  $W^l: kernel$ $*: convolution$

  $b^l: bias$

MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

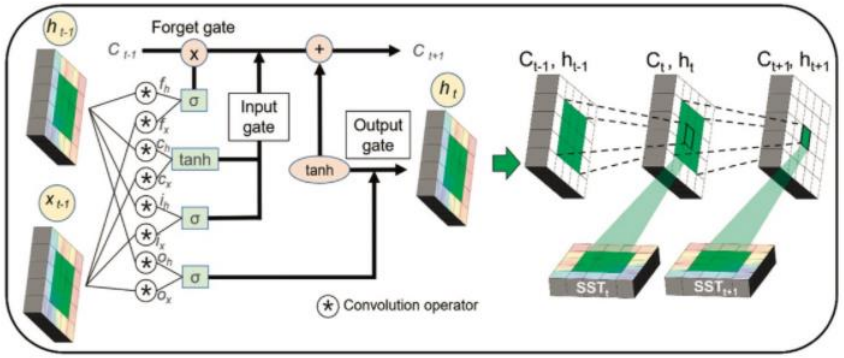


❖ MSCRED Framework – Attention based ConvLSTM



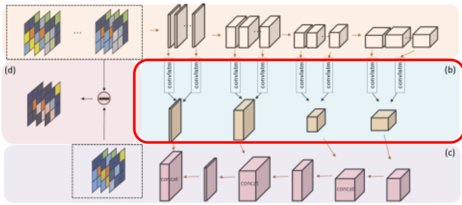
$(30 \times 30 \times 32)$ $(15 \times 15 \times 64)$ $(8 \times 8 \times 128)$ $(4 \times 4 \times 256)$

➢ ConvLSTM



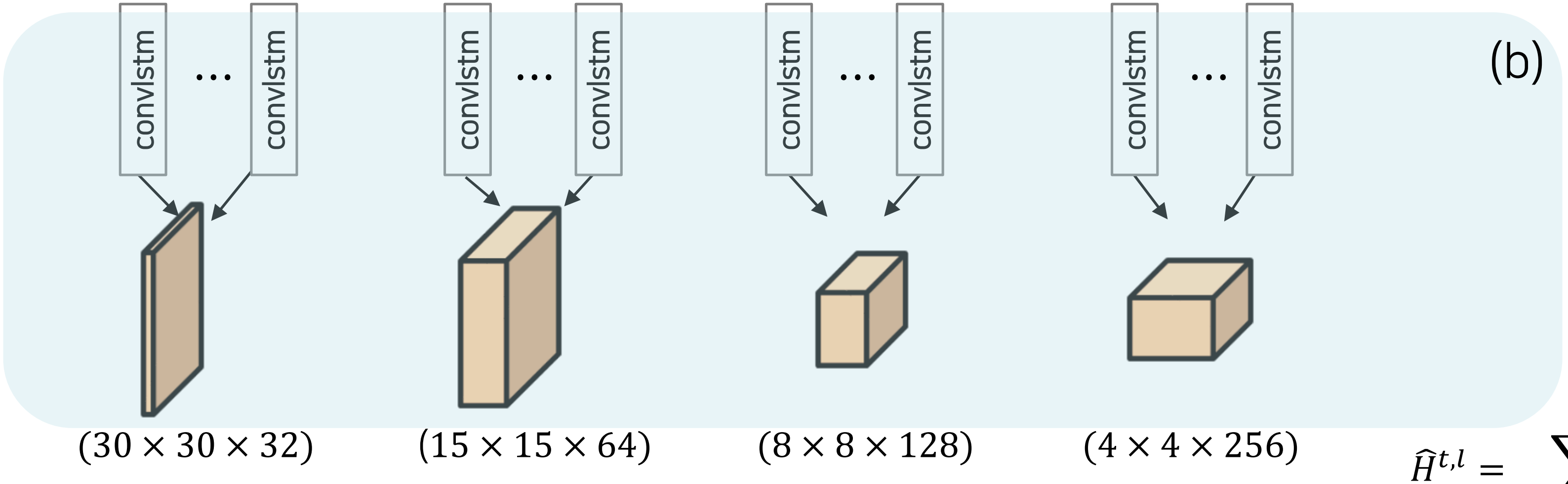
- LSTM의 연산을 Convolution 연산으로 대체
- 시간적, 공간적 정보를 모두 처리 가능

- Convolutional encoder를 통해 encoding된 feature map($h = 5$)를 input으로 사용
- ConvLSTM을 사용하여 feature map 각각의 hidden state를 추출
- 마지막 hidden state의 유사도를 기반으로 Attention weight를 도출Temporal information을 추출
- Attention weight 합을 통해 Temporal information을 가진 최종 feature map 도출



MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

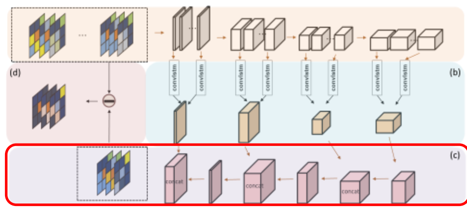
❖ MSCRED Framework – Attention based ConvLSTM



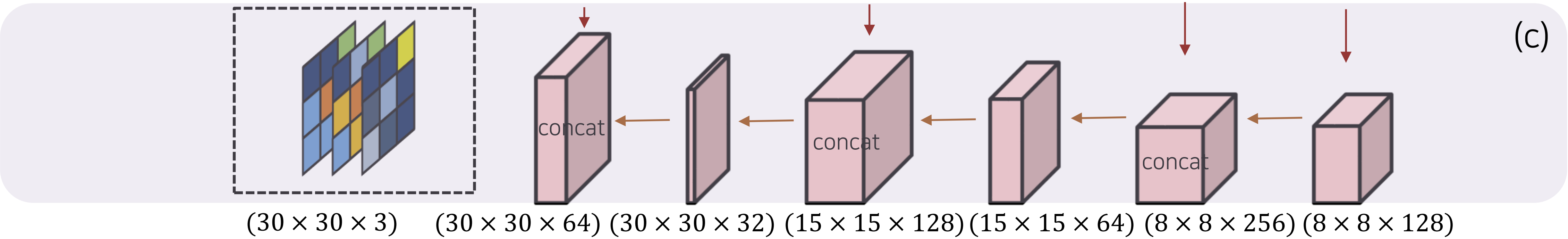
$$\hat{H}^{t,l} = \sum_{i\in(t-h,t)} \alpha^i H^{i,l}, \alpha^i = \frac{exp\left\{\frac{Vec(H^{t,l})^T Vec(H^{i,l})}{\chi}\right\}}{\sum_{i\in(t-h,t)} exp\left\{\frac{Vec(H^{t,l})^T Vec(H^{i,l})}{\chi}\right\}}$$

- Convolutional encoder를 통해 encoding된 feature map($h = 5$)를 input으로 사용
- ConvLSTM을 사용하여 feature map 각각의 hidden state를 추출
- 마지막 hidden state의 유사도를 기반으로 Attention weight를 도출하여 Temporal information을 추출
- Attention weight 합을 통해 Temporal information을 가진 최종 feature map 도출

MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)

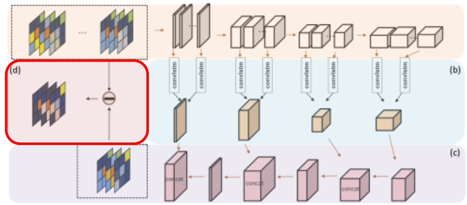


❖ MSCRED Framework – Convolutional decoder

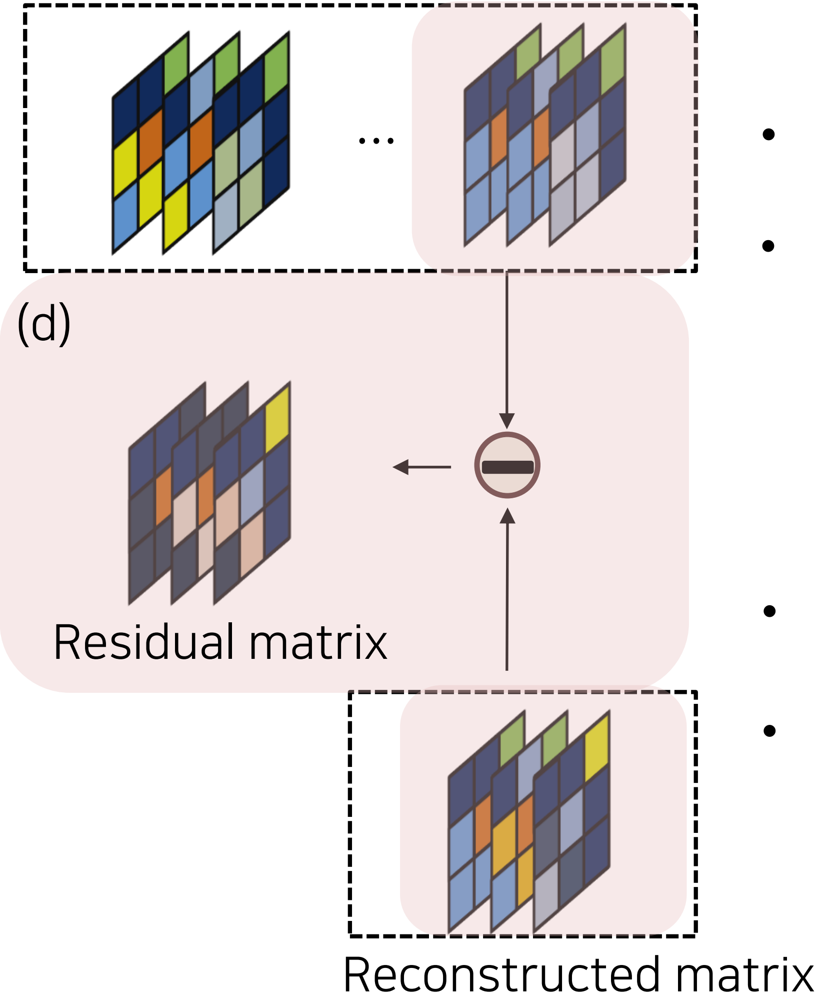


- 4번째 Attention based ConvLSTM으로 추출된 feature map은 Deconvolution 실시
- 나머지 Attention based ConvLSTM으로 추출된 feature map과 Deconvolution layer의 output간의 concat을 실시
- concat된 feature map을 다음 Deconvolution layer의 input으로 사용
- $\hat{\chi}^{t,l-1} = \begin{cases} f(\hat{W}^{t,l} \circledast \hat{H}^{t,l} + \hat{b}^{t,l}), & l = 4 \\ f(\hat{W}^{t,l} \circledast [\hat{H}^{t,l} \oplus \hat{\chi}^{t,l}] + \hat{b}^{t,l}), & l = 3,2,1 \end{cases}$

MSCRED(Multi-Scale Convolutional Recurrent Encoder-Decoder)



❖ MSCRED Framework – Loss function and anomaly score



- Residual matrix: t 시점 Signature matrix와 Reconstructed matrix 간의 차이
- Residual matrix 성분의 제곱의 합이 최소가 되는 방향으로 학습

$$\boldsymbol{loss\ fuction} = \sum_{t}\sum_{c=1}^{s}\left\|\boldsymbol{x}_{:,:,c}^{t,0} - \hat{\boldsymbol{x}}_{:,:,c}^{t,0}\right\|_F^2$$

- Anomaly score: Residual matrix 성분 중 특정 기준치를 넘는 성분의 개수
- Anomaly score가 threshold를 넘으면 이상으로 판단

Experiments

❖ Data

- Synthetic data 1개 , real world data 1개
- 평가 지표 : precision, recall, f1 score

Statistics	Synthetic	Power Plant
# time series	30	36
# points	20,000	23,040
# anomalies	5	5
# root causes	3	3
train period	$0 \sim 8{,}000$	$0 \sim 10{,}080$
valid period	$8{,}001 \sim 10{,}000$	$10{,}081 \sim 18{,}720$
test period	$10{,}001 \sim 20{,}000$	$18{,}721 \sim 23{,}040$

Experiments

❖ Performance Evaluation

- RQ1 : Baseline보다 성능이 높은가?
- RQ2 : Model의 variants의 결과를 비교할 때는 성능이 어떻게 달라지는가?

Method	Synthetic Data			Power Plant Data		
	Pre	Rec	F_1	Pre	Rec	F_1
OC-SVM	0.14	0.44	0.22	0.11	0.28	0.16
DAGMM	0.33	0.20	0.25	0.26	0.20	0.23
HA	0.71	0.52	0.60	0.48	0.52	0.50
ARMA	0.91	0.52	0.66	0.58	0.60	0.59
LSTM-ED	1.00	0.56	0.72	0.75	0.68	0.71
$\text{CNN}^{ED(4)}_{ConvLSTM}$	0.37	0.24	0.29	0.67	0.56	0.61
$\text{CNN}^{ED(3,4)}_{ConvLSTM}$	0.63	0.56	0.59	0.80	0.72	0.76
$\text{CNN}^{ED}_{ConvLSTM}$	0.80	0.76	0.78	0.85	0.72	0.78
MSCRED	**1.00**	**0.80**	**0.89**	**0.85**	**0.80**	**0.82**
Gain (%)	–	30.0	23.8	13.3	19.4	15.5

✓ Baseline보다 성능 높음을 확인

Experiments

❖ Performance Evaluation

- RQ1 : Baseline보다 성능이 높은가?
- RQ2 : Model의 variants의 결과를 비교할 때는 성능이 어떻게 달라지는가?

Method	Synthetic Data			Power Plant Data		
	Pre	Rec	F_1	Pre	Rec	F_1
OC-SVM	0.14	0.44	0.22	0.11	0.28	0.16
DAGMM	0.33	0.20	0.25	0.26	0.20	0.23
HA	0.71	0.52	0.60	0.48	0.52	0.50
ARMA	0.91	0.52	0.66	0.58	0.60	0.59
LSTM-ED	<u>1.00</u>	<u>0.56</u>	<u>0.72</u>	<u>0.75</u>	<u>0.68</u>	<u>0.71</u>
$\mathrm{CNN}_{ConvLSTM}^{ED(4)}$	0.37	0.24	0.29	0.67	0.56	0.61
$\mathrm{CNN}_{ConvLSTM}^{ED(3,4)}$	0.63	0.56	0.59	0.80	0.72	0.76
$\mathrm{CNN}_{ConvLSTM}^{ED}$	0.80	0.76	0.78	0.85	0.72	0.78
MSCRED	**1.00**	**0.80**	**0.89**	**0.85**	**0.80**	**0.82**
Gain (%)	–	30.0	23.8	13.3	19.4	15.5

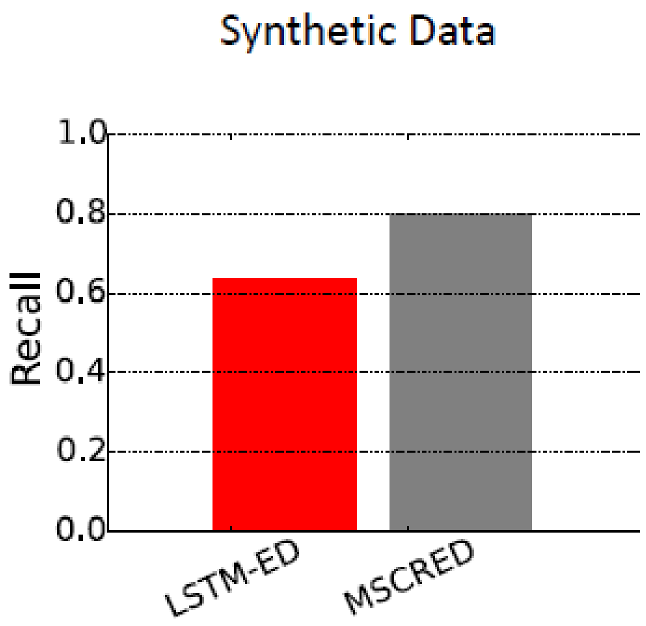
✓ Attention을 도입하였을 때 성능이 상승

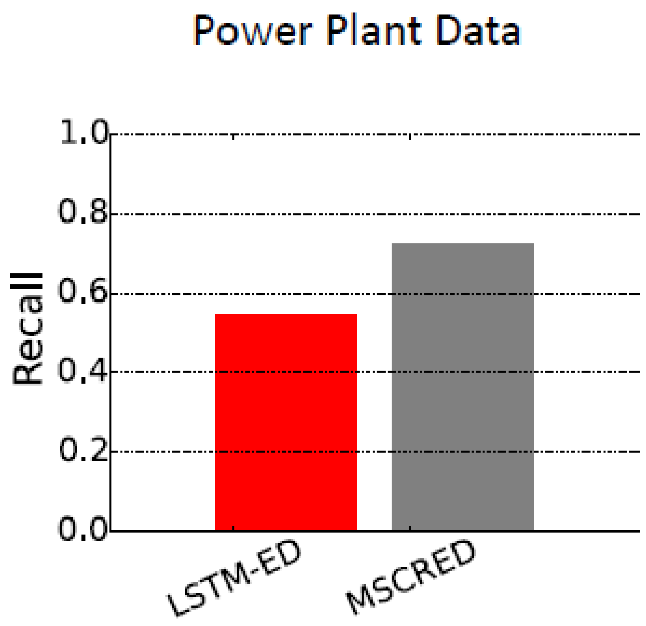
✓ ConvLSTM의 개수가 증가 할 수록 성능이 상승

Experiments

❖ Performance Evaluation

- RQ3 : Alarm이 울린 원인(Root cause)를 잘 찾아 내는가?
  - ✓ Recall 기준으로 MSCRED가 성능이 높음을 확인

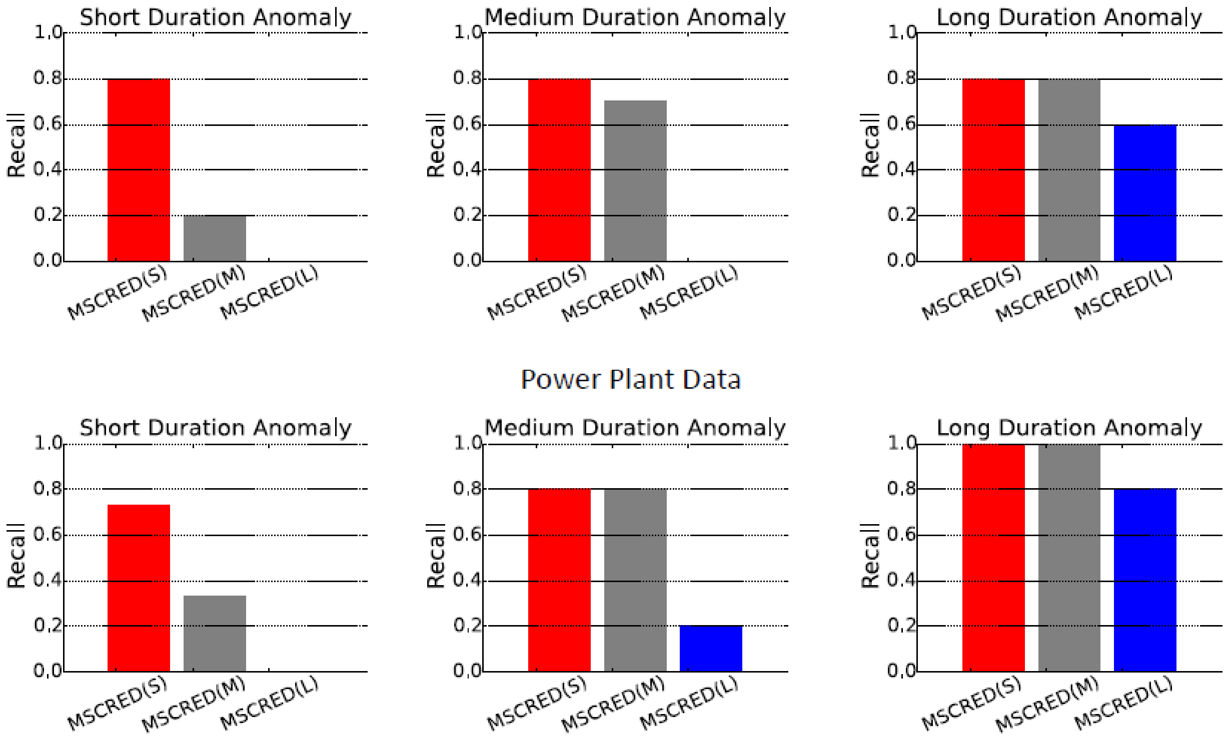


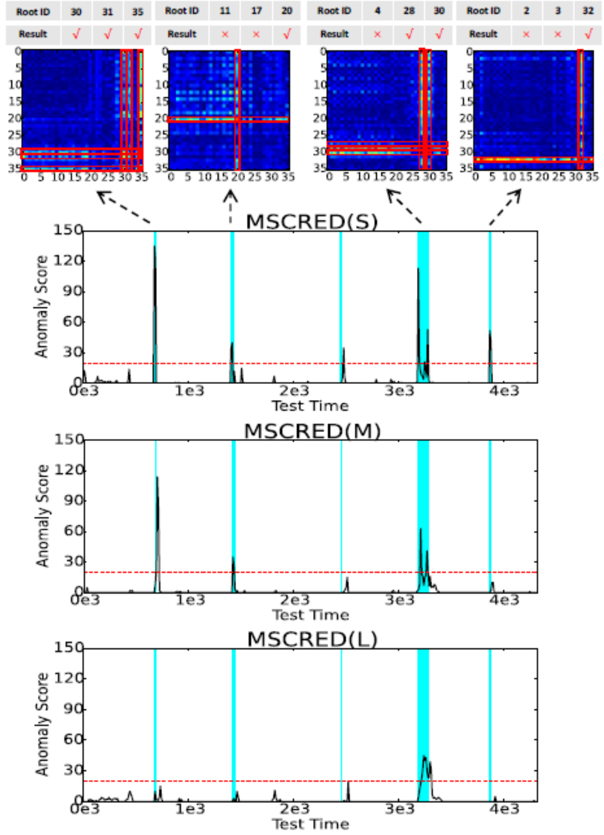


Experiments

❖ Performance Evaluation

- RQ4 : Duration(10,30,60)에 따라 Anomaly 의 차이가 어떻게 나는가?
  - ✓ 각 duration에 따라 탐지 가능한 구간의 차이가 있음을 확인

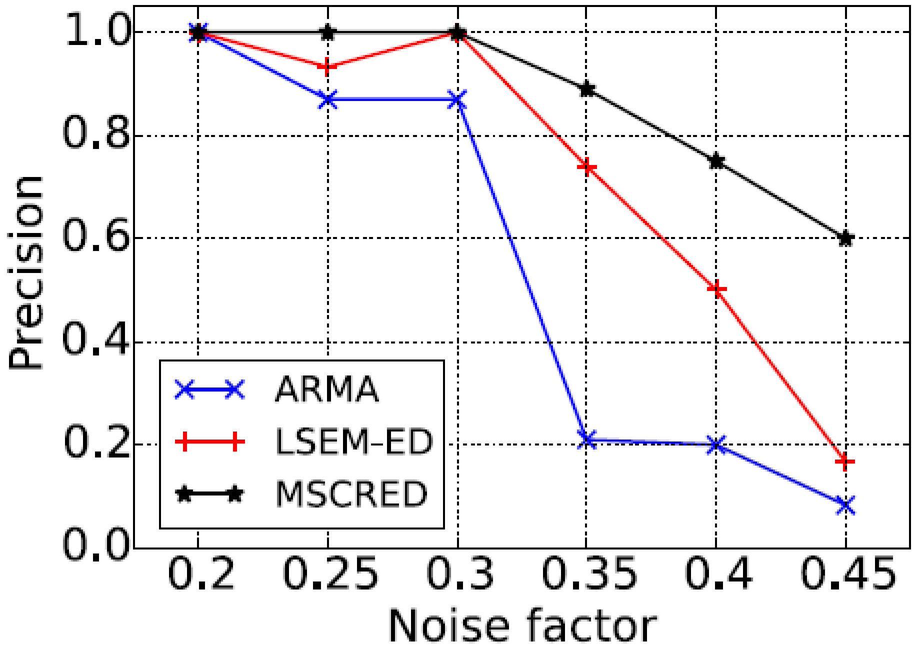


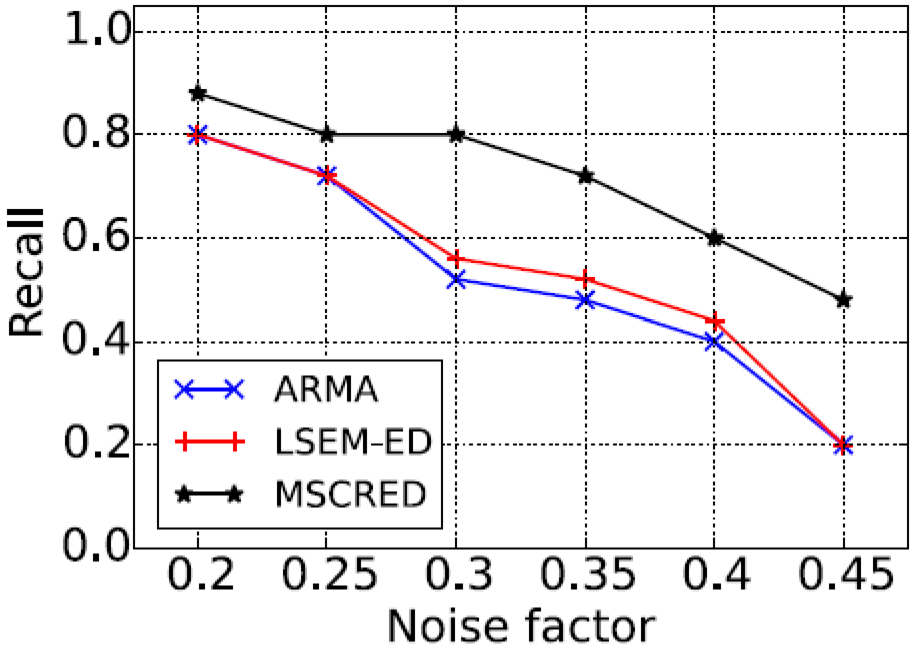


Experiments

❖ Performance Evaluation

- RQ5 : noise에 강건한가?
  - ✓ MSCRED가 noise에 모두 강함을 확인





Anomaly detection using imaging of time series data

Paper

❖ GAN-based Anomaly Detection and Localization of Multivariate Time Series Data for Power Plant

- Choi, Yeji, et al. "Gan-based anomaly detection and localization of multivariate time series data for power plant." *2020 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 2020..**(28회 인용)**

2020 IEEE International Conference on Big Data and Smart Computing (BigComp)

GAN-based Anomaly Detection and Localization of Multivariate Time Series Data for Power Plant

Yeji Choi
Center for Imaging Media Research
Korea Institute of Science and Technology
Seoul, Korea
cyjcyj91@kist.re.kr

Hyunki Lim
Center for Imaging Media Research
Korea Institute of Science and Technology
Seoul, Korea
hyunkilim@kist.re.kr

Heeseung Choi
Center for Imaging Media Research
Korea Institute of Science and Technology
Seoul, Korea
hschoi@kist.re.kr

Ig-Jae Kim
Center for Imaging Media Research
Korea Institute of Science and Technology
Seoul, Korea
drjay@kist.re.kr

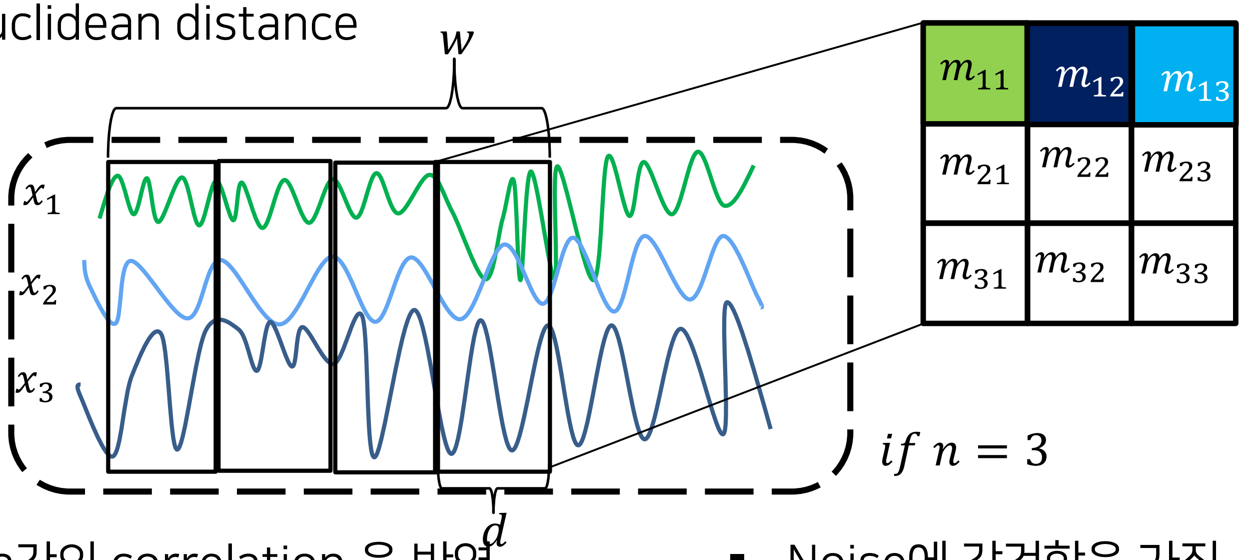
***Abstract*—Recently, as real-time sensor data collection increases in various fields such as power plants, smart factories, and health care systems, anomaly detection for multivariate time series data analysis becomes more important. However, extracting significant features from multivariate time series data is still challenging because it simultaneously takes into account the correlation between the pair of sensors and temporal information of**

recursive connection. However, these methods assume full connectivity between variables, making it difficult to interpret correlation between variables as the number of variables increases [1]. Also, they have a large computational complexity as the length of the time series data increases. Convolutional Neural Networks(CNN), which is widely used in image

Proposed Method

❖ Distance image

- $T: X = \{x_1, x_2, x_3, \cdots, x_n\}$ (n: $data$ 변수의 개수, T: 전체 Time 길이)
- $m_{ij}^t = \frac{1}{d}\sum_{\delta=0}^{d-1}\|x_i^{t-\delta} - x_j^{t-\delta}\|$ (d :duration, x_i : i th sensor data, $x_i^{t-\delta}$: $t-\delta$ 시점에서의 point, w: Timewindow)
  - Distance : Euclidean distance



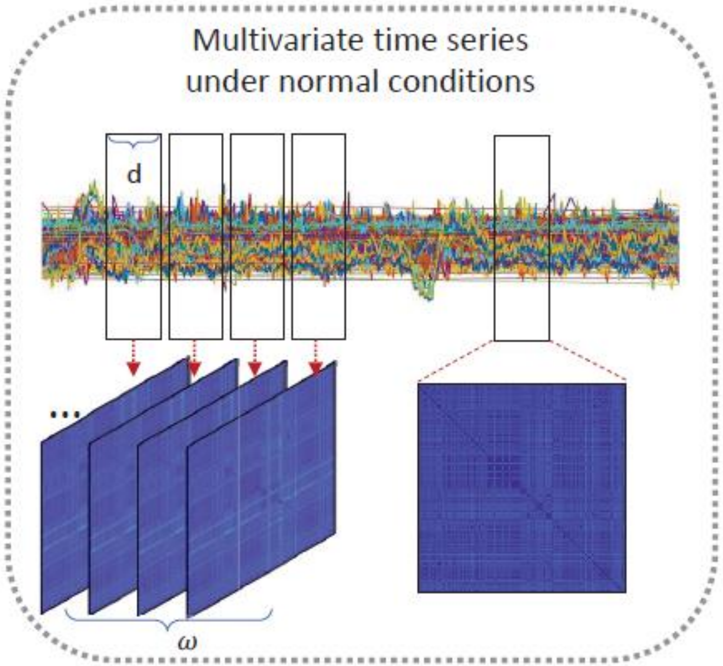
- Time series data간의 correlation 을 반영
- Noise에 강건함을 가짐
- Temporal 정보를 반영

Proposed Method

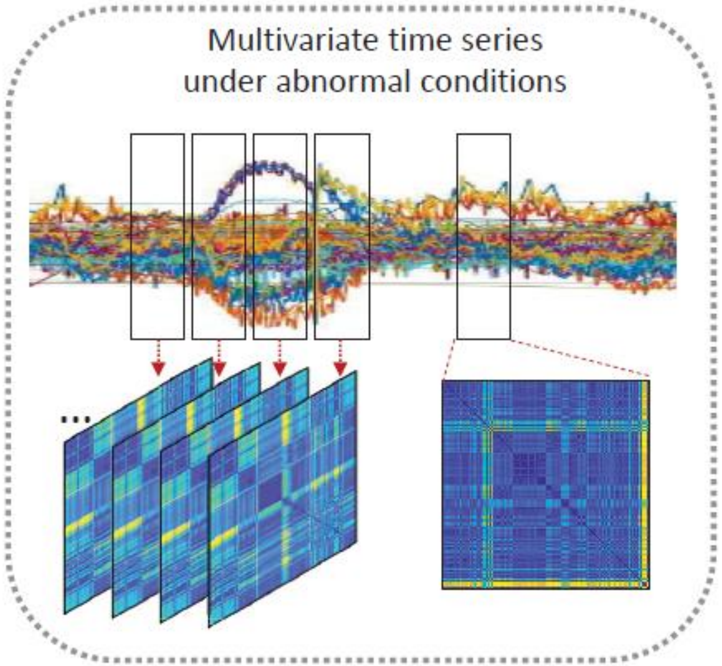
❖ Distance image

- $d = 10, w = 30$

Normal distance image

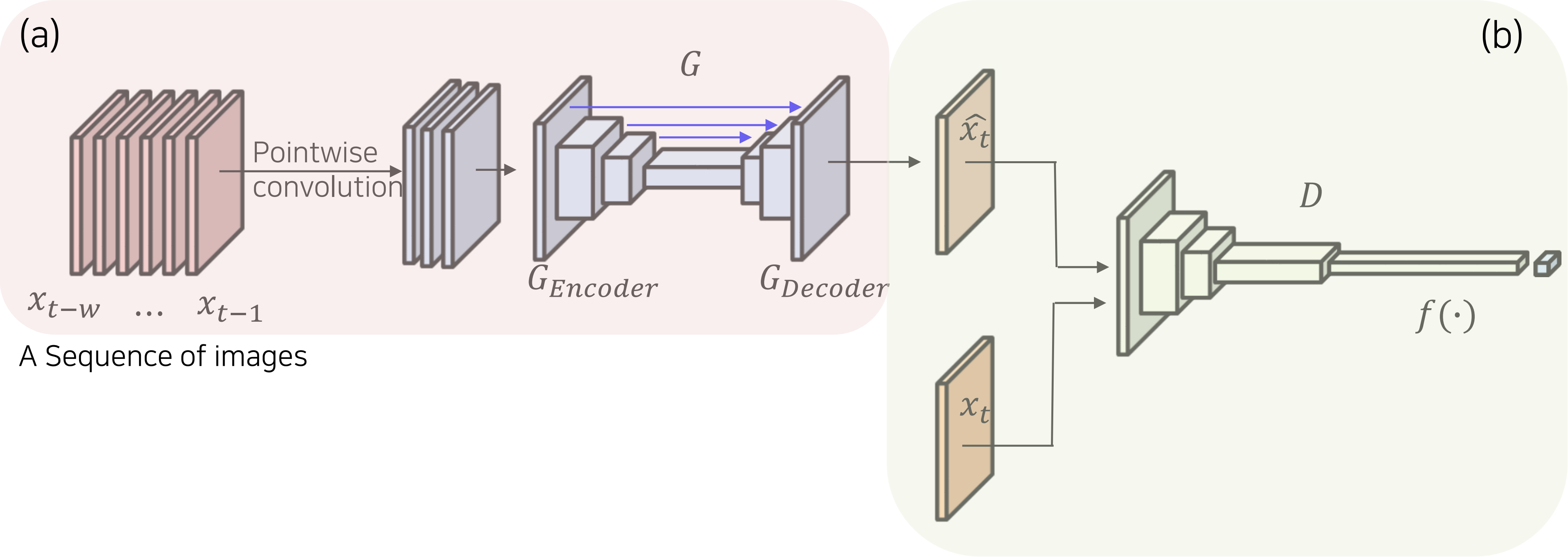


Abnormal distance image



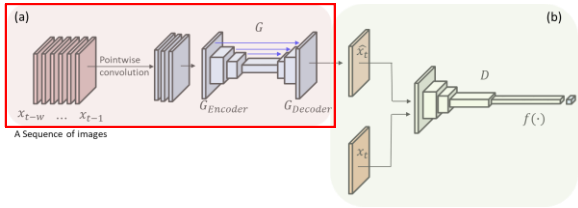
Proposed Method

❖ Model Framework

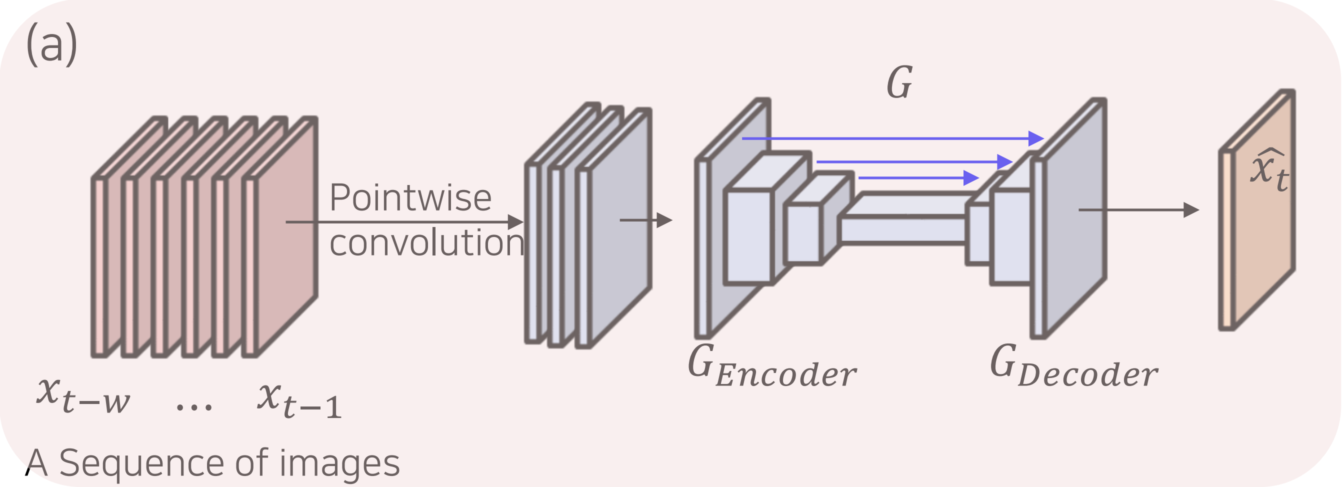


A Sequence of images

Proposed Method

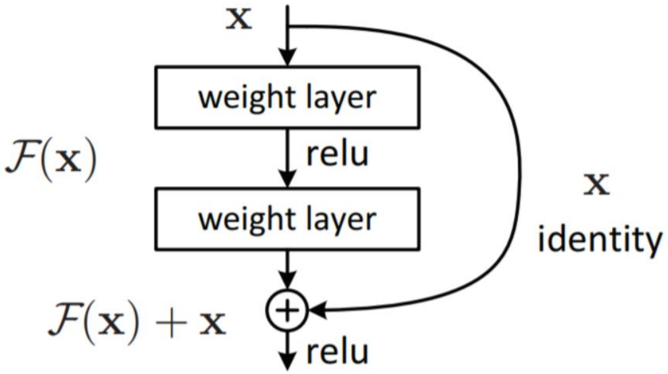


❖ Model Framework-Generator

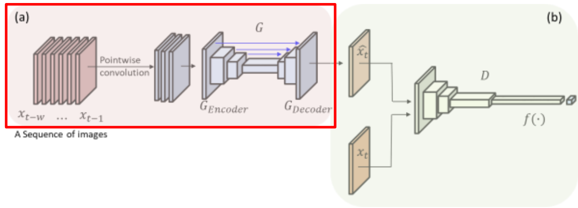


- Encoder, Decoder 구조(skip connection 구조 포함)
- $(t - w \sim t - 1)$ 다음의 distance image를 generate
- Point-wise Convolution을 사용
  - Temporal 정보를 포착
  - 변수들간의 correlation을 represent하는 spatial 정보를 추출

➢ skip connection

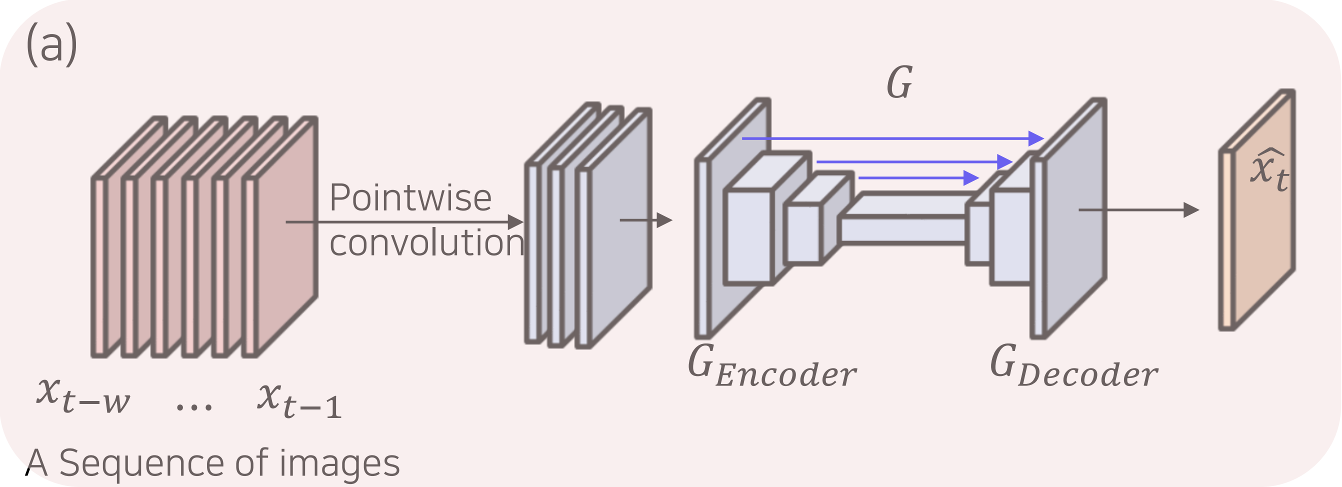


하나의 layer의 output을 몇 개의 layer를 건너뛰고 다음 layer의 input에 추가하는 것



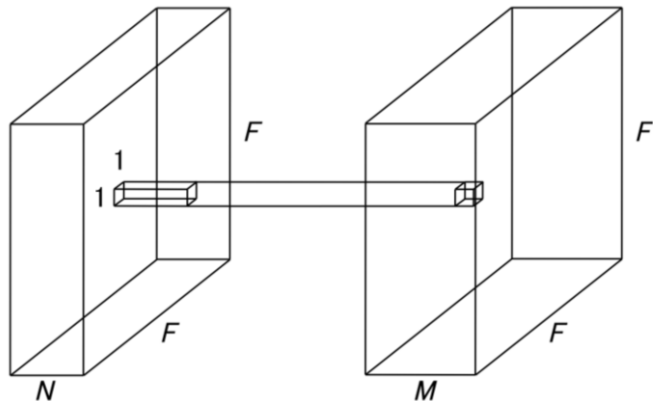
Proposed Method

❖ Model Framework-Generator



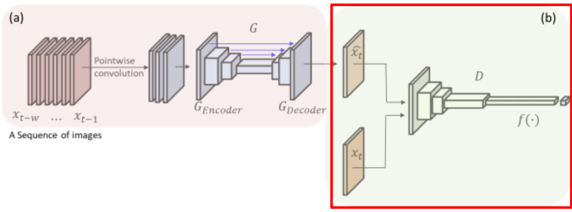
- Encoder, Decoder 구조(skip connection 구조 포함)
- $(t - w \sim t - 1)$ 다음의 distance image를 generate
- Point-wise Convolution을 사용
  - Temporal 정보를 포착
  - 변수들간의 correlation을 represent하는 spatial 정보를 추출

➢ Point-wise Convolution



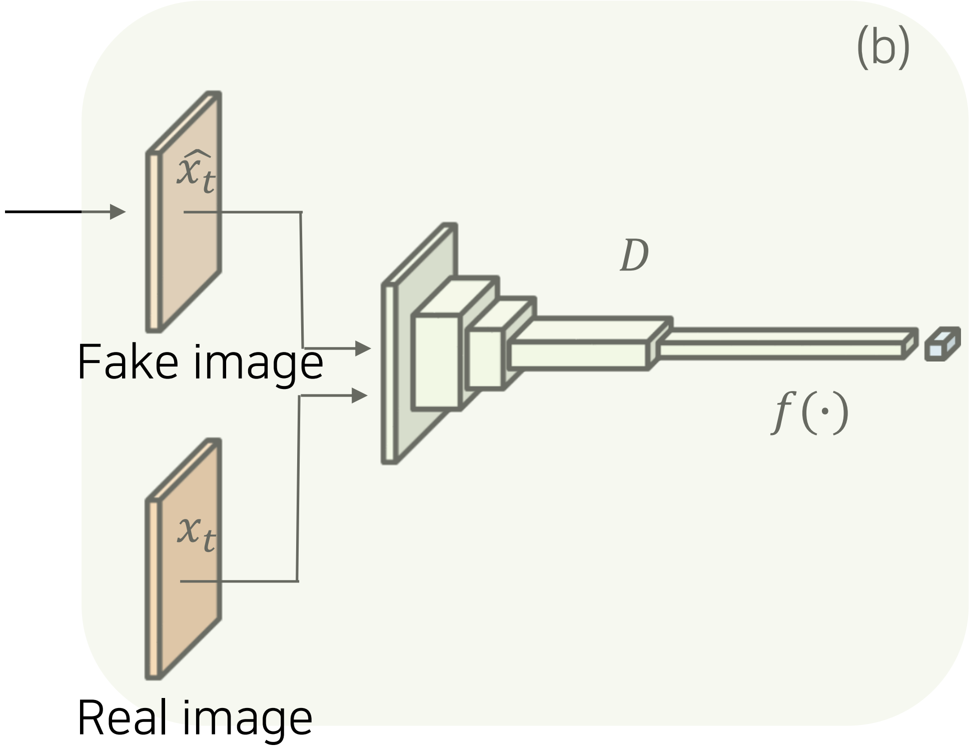
1x1xC filter 사용

출력 채널 수를 줄여 계산량 감소



Proposed Method

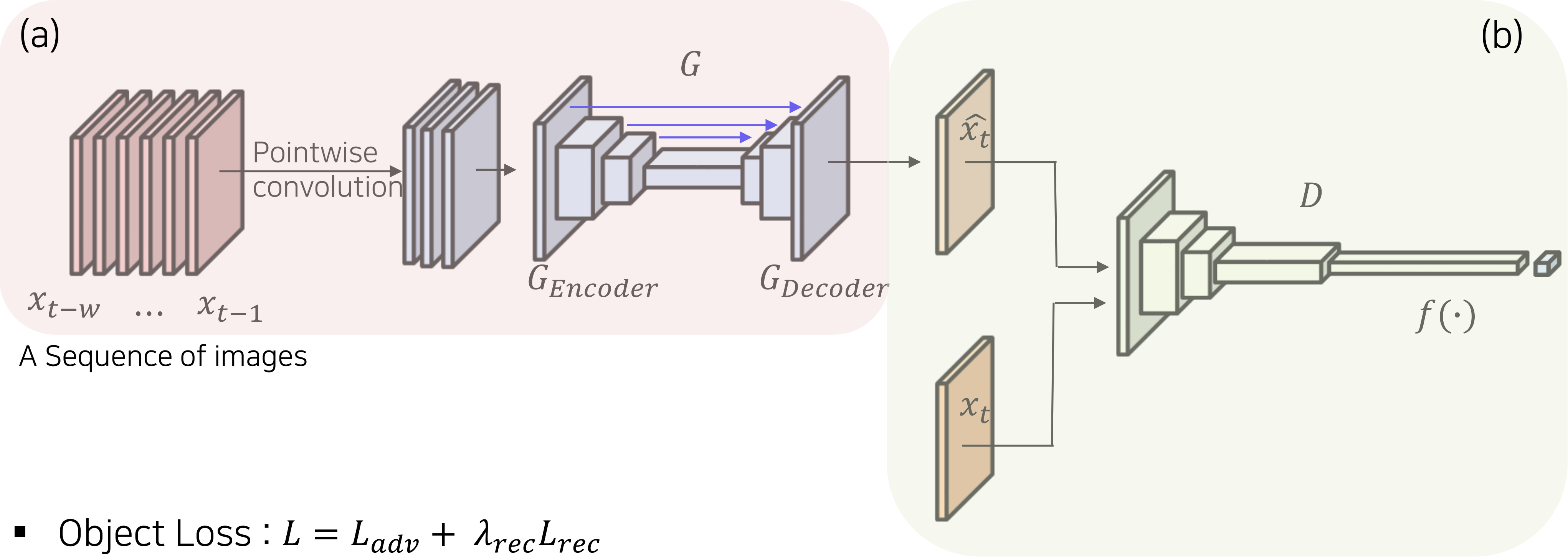
❖ Model Framework - Discriminator



- Real image와 Fake image를 구별

Proposed Method

❖ Object Loss

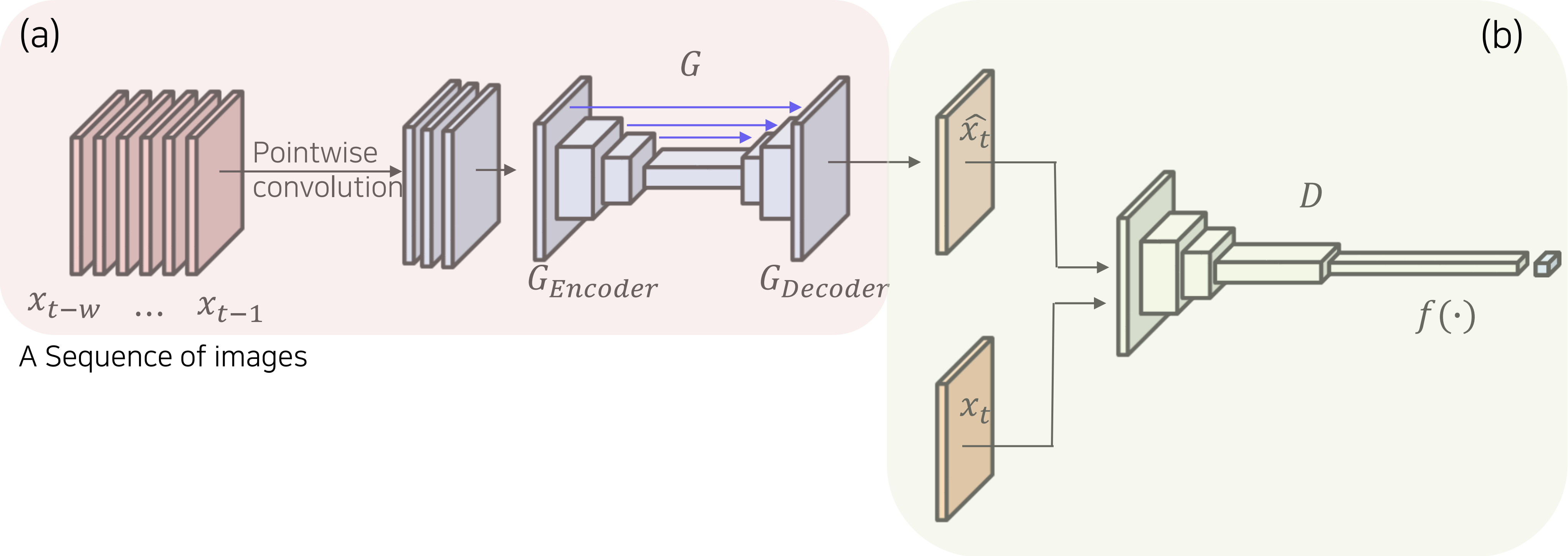


A Sequence of images

- Object Loss : $L = L_{adv} + \lambda_{rec} L_{rec}$
- Reconstructed Loss : $L_{rec} = \mathbb{E}_x \| m^t - G(m^{t-w}, \cdots, m^{t-1}) \|_1$
- Adversarial Loss : $L_{adv} = \mathbb{E}_x[D(m^t)] - \mathbb{E}_x\left[D\left(G(m^{t-w}, \cdots, m^{t-1})\right)\right] - \lambda_{gp} \mathbb{E}_{\hat{m}}[(\|\nabla_{\hat{m}} D(\hat{m})\|_2 - 1)^2]$

Proposed Method

❖ Object Loss

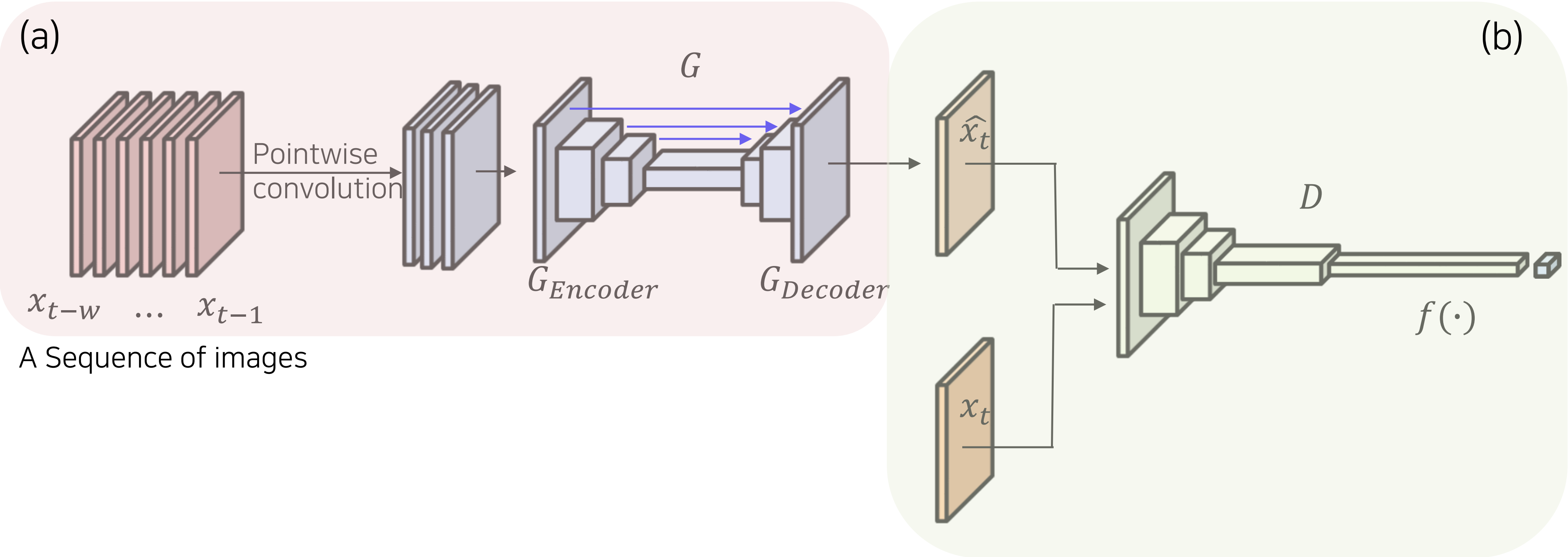


A Sequence of images

- Adversarial Loss : $L_{adv} = \mathbb{E}_x[D(m^t)] - \mathbb{E}_x\big[D\big(G(m^{t-w},\cdots,m^{t-1})\big)\big] - \underbrace{\lambda_{gp}\mathbb{E}_{\hat{m}}[(\|\nabla_{\hat{m}}D(\hat{m})\|_2 - 1)^2]}_{\text{gradient penalty}}$
  - WGAN-GP: Wasserstein GAN objective with gradient penalty
  - WGAN의 weight clipping 제한을 해소를 한 WGAN-GP Loss를 사용하여, Training process를 안정화 시킴

Proposed Method

❖ Object Loss



A Sequence of images

- Reconstructed Loss : $L_{rec} = \mathbb{E}_x \| m^t - G(m^{t-w}, \cdots, m^{t-1}) \|_1$
  - G가 L_{adv} 에 의해서 D를 속이는 방향으로 학습이 잘 되지만, data의 Contextual information을 학습하지 못함
  - $L_1\ distance$를 사용하여 이미지가 흐릿해지는 것을 완화

Proposed Method

❖ Anomaly score function

- Anomaly score function : $\Phi(t) = L_{rec}^{t} + \lambda_f L_f^{t}$
  - ➢ Reconstructed Loss : $L_{rec} = \mathbb{E}_x \| m^t - G(m^{t-w}, \cdots, m^{t-1}) \|_1$
  - ➢ Feature Loss : $L_f = \| f(m^t) - f(G(m^{t-w}, \cdots, m^{t-1})) \|$

    $f(\cdot)$: Discriminator의 마지막 layer로 부터 feature vector
- Normal 분포와 다를 경우 L_{rec} 상승
- L_f에서 input image와 generated image의 dissimilarity 산출
- Anomaly score function을 통해서 abnormality의 정도를 측정
- Anomaly score가 threshold 보다 높으면 이상으로 판단

Experiments

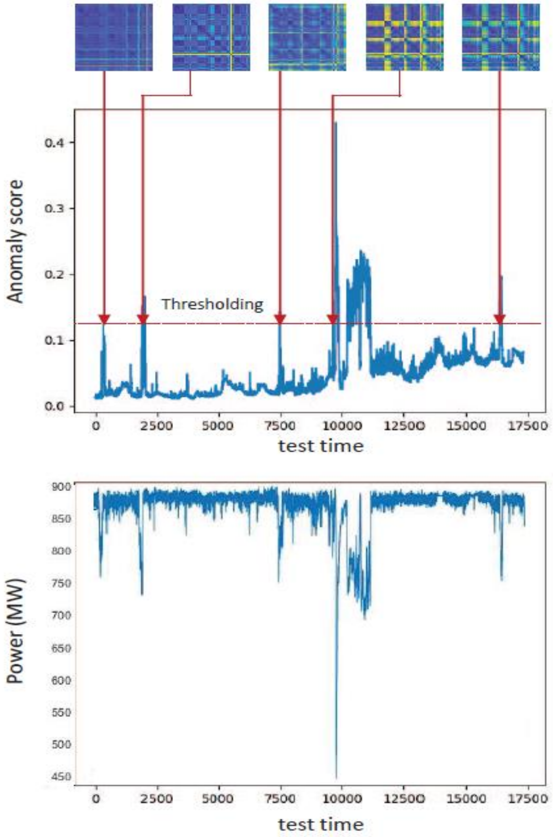
❖ Data

- real world data 1개

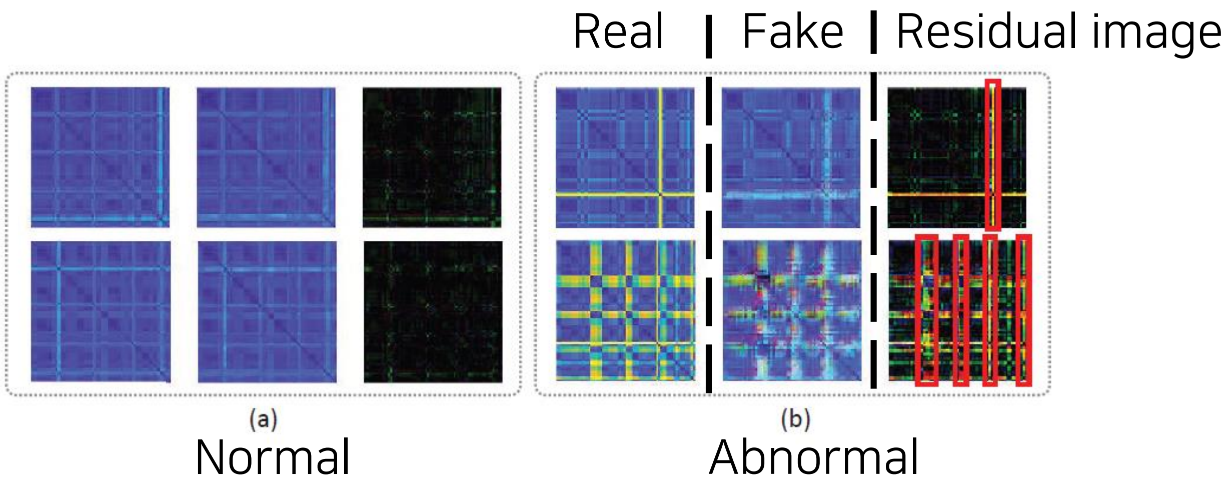
Statistics	Power Plant
# time series	193
# points	691,200
# anomalies	4
train	0 ~ 120,960
valid	120,961 ~ 151,200
test	151,201 ~ 691,200

Experiments

❖ Performance Evaluation



- Power의 값이 변할 때 Anomaly score도 변화가 발생함을 확인



- Residual image를 통하여 Anomaly 부분 확인 가능
  - ➢ Residual image : test image와 generated image 간의 차이
    (Real) (Fake)

Conclusion

❖ 시계열 데이터의 이상탐지는 주로 비지도 학습으로 진행

❖ RNN, LSTM 구조는 계산의 복잡성 및 변수들 간의 관계성을 제대로 반영하지 못한다는 단점이 존재

❖ 내적(inner product) 또는 거리(distance) 기반으로 시계열 데이터 이미지화를 통해 변수들간의 관계성을 반영

❖ Convolutional 연산을 통해 계산 복잡성을 감소

❖ Enocder-Decoder 구조 및 GAN구조를 활용하여 이상 탐지 성능을 증명

감사합니다.