객체 이미지로 사전 학습된 PatchCore 기반 고위험 발전소 시계열 이상 탐지

최용훈 [1], 조유진 [2], 장준원 [3], 김민석*
[1,2,3,*]상명대학교 휴먼지능로봇공학과

[1]choiyonghoon1103@gmail.com, [2]yj889941@gmail.com, [3]pierrot7794@gmail.com,
*minsuk.kim@smu.ac.kr

Time Series Anomaly Detection in Critical Power Plants Using PatchCore Pre-trained on Object Images

Choi Yong Hoon, Jo Yu Jin, Jang Jun Won, Kim Min Suk*

*Dept. of Human Intelligence & Robot Engineering. Sangmyung University.

요 약

최근, 고위험 발전소 내 시계열 데이터 기반의 자동화된 이상 탐지 시스템의 필요성이 대두되고 있다. 따라서 본 논문은 이차계통 발전소에서 수집된 시계열 데이터를 이용한 이상 탐지 시스템을 위한 딥러닝 모델을 제안한다. 제안된 이상 탐지 시스템의 정확성과 안정성의 강화를 위해 시계열 데이터를 이미지로 변환하여 딥러닝 모델 성능을 향상하였다. 이는 MTF, PAA 와 GAF 를 이용하여 시계열 데이터를 이미지 데이터로 변환 후, PatchCore 모델을 통해 이상탐지를 수행할 수 있는 방법이다. 실험 결과, PAA 변환 후 GAF 변환을 한 데이터가 가장 높은 성능을 달성하였으며, 변환 과정에서 패턴 명확성을 높여 PatchCore 모델이 이상 데이터를 높은 성능으로 판별하는 것을 확인하였다. 이를 통해 본 연구는 한정된 데이터를 이용하여 이차계통 발전소 환경에서의 이상 탐지 효율성과 안정성 강화를 기여하는데 목적이 있다.

Ⅰ. 서 론

다양한 산업 분야에서 상태 모니터링은 예방 및 유지보수의 핵심 요소로서 자동화 시스템의 필요성과 중요성이 지속적으로 대두되고 있다. 특히 이차계통 발전소를 비롯한 고위험 발전소의 경우, 이상 징후를 즉각적으로 감지하고 대응하는 것이 매우 중요하다. 이는 발전소 시설의 안정성과 신뢰성 유지에 필수적인 요소로 작용하며 센서 정보 기반의 자동화된 이상 감지 시스템에 필요하다. 이러한 시스템은 운영 중 발생할 수 있는 잠재적 위협을 사전에 탐지하고 안정성을 강화할 수 있으므로 이차계통 발전소 환경에서 발생하는 센서 데이터의 시계열 데이터 분석을 통한 이상 징후 탐지 시스템을 구현할 필요성이 있다.

최근 연구에서는 시계열 데이터의 복잡한 패턴을 보다 효과적으로 추출하고 분석하기 위해 이미지로 변환하여 딥러닝 모델을 학습시키는 방법이 제안되었다. 이러한 접근 방식은 시계열 데이터의 시간적 연속성과 특징을 유지하며 이미지 기반 딥러닝 모델의 강점을 활용하여 높은 성능을 달성하였다.

본 논문에서는 앞서 제시된 방법을 기반으로 이차계통 발전소에서 수집된 시계열 데이터를 이용한 이미지 기반의 딥러닝 모델을 활용한 이상 탐지 시스템을 제안한다. 이 방법은 시계열 데이터를 이미지로 변환한 후, 객체 인식 분야에서 높은 성능을 보인 PatchCore[1]를 사용하여 이상 탐지를 수행하고 이를 비교한다. 이에 대한 전체적인 과정은 Figure 1 에서 확인할 수 있다.

PatchCore 는 기존의 대규모 객체 기반 데이터셋에서 사전 학습된 모델로, 제한된 데이터 상황에서도 효과적인 특성 학습이 가능하도록 한다. 이를 통해 본 논문에서는 이차계통 발전소에서 수집된 시계열 데이터를 이용한 빠르고 정확한 이상 탐지 시스템을 구현하여 발전소 시설의 안정성 강화를 목표로 한다.

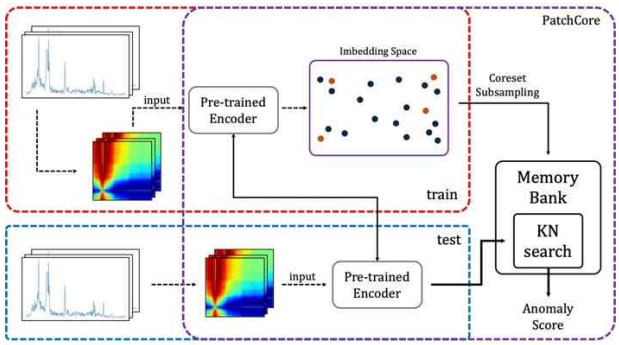

Figure 1. Model Architecture Overview

Ⅱ. 시계열 기반의 이차계통 발전소 이상 탐지

2.1. 데이터셋

본 연구에서는 이차계통 발전소의 배관에 대한 누출 시나리오를 모방하는 시뮬레이션 상황에서 수집한 시계열 데이터를 사용한다. 20kHz ~ 100kHz 범위의 초음파 신호로 누출을 감지하고, 20kHz 이상 신호는 증폭하여 수집되었다. 최종적으로 0.25Hz 간격의 20kHz ~ 100kHz 범위의 320 의 길이를 가지는 평균 스펙트럼 값 으로 구성된다. 각 유체의 입력과 파이프의 구멍 직경에 따라 A, B, C 클래스로 분류되며, 클래스별 데이터 구성은 표 1 과 같다.

Table 1. Summary of Dataset Configurations

Class		Train	Test
A	anomaly	None	1001
	normal	500	501
B	anomaly	None	778
	normal	300	63
C	anomaly	None	1169
	normal	500	515

2.2. 시계열 데이터 이미지 변환

본 연구에서 사용하는 시계열 데이터는 이미지로 변환하여 딥러닝 모델 학습에 사용하므로 MTF, GAF 를 이용한 데이터 변환 과정이 필요하다. 이 과정에서 시계열 데이터의 패턴을 시각적으로 표현하고 딥러닝 모델이 데이터를 효과적으로 학습하도록 한다.

MTF 는 시계열 데이터를 공간적 연속성으로 변환하는 기법으로 데이터를 구간화 하여 각 구간에서 나타나는 상태의 전환 확률을 계산한 후, 이를 이미지 형태로 변환한다. GAF 는 시계열 데이터를 정규화 한 후, 각 데이터 점들을 폴라 좌표계에서 각도로 변환하여 코사인 및 사인함수를 이용하여 GAF 이미지를 생성한다. 추가로 GAF 를 수행하기에 앞서 PAA 기법을 적용하여 시계열 데이터를 압축한다. PAA 는 시계열 데이터를 구간별로 평균화하여 차원을 축소하는 기법으로 데이터의 주요 특징을 유지하면서 계산 복잡성을 줄인다. 본 연구에서는 PAA 를 이용하여 시계열 데이터를 단순화하고 데이터의 주요 특성을 강조하도록 유도하는데 사용하였다.

이러한 데이터 변환 과정을 통해 생성된 이미지는 Figure 2 에 나타난 바와 같이, 시계열 데이터의 시각적 패턴을 드러내며 딥러닝 모델의 학습에 활용된다.

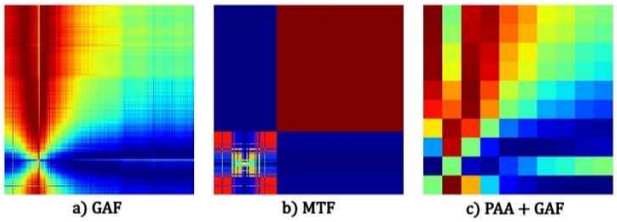

Figure 2. Results of Image Transform

2.3. PatchCore

PatchCore 모델은 오브젝트 기반의 이미지 데이터셋으로 사전 학습된 인코더 $f(\cdot)$를 이용하여 입력 이미지 I의 패치를 추출하고 각 패치 x_i를 고차원 공간에 매핑한다. 이를 기반으로, 서브샘플링을 통해 특징 벡터 $\{z_1, z_2, z_3, z_4, \dots, z_n\}$ 들의 유클리드 거리 $d(z_i, z_j) = \|z_i - z_j\|$를 연산하여 가장 먼 거리의 벡터를 선택 후, 선택된 벡터와 가장 먼 거리의 벡터를 선택하는 과정을 반복하여 메모리 뱅크 M 를 구성한다. 해당 과정은 수식 1 과 같다.

$$z_{next} = argmax_{z \in \{z_1, z_2, \dots, z_n\} \setminus M} min_{m \in M} d(z, m) \quad (1)$$

구성된 메모리 뱅크 $M = \{m_1, m_2, m_3, m_4, \dots, m_n\}$ 를 통해 새로운 데이터가 들어오면 동일한 인코더를 이용해서 특징 벡터 z_{new} 를 추출하고 이를 메모리 뱅크 M 에 저장된 벡터 들과의 최소 거리로 이상 점수를 산출한다. 해당 과정은 수식 2 와 같다.

$$Anomaly\ \ Score\ \ (z_{new}) = min_{m \in M} d(z_{new}, m) \quad (2)$$

본 논문에서는 ImageNet[2]으로 사전 학습된 WideResNet-50[3]을 사용하여 PatchCore 학습을 진행한다.

III. 실험

본 논문에서는 사전 학습된 PatchCore 를 통해 이차계통 발전소 데이터를 학습하고 이에 대한 검증을 진행한다. 이상 탐지에 대한 평가 지표로는 이진 분류에 대한 정확도를 표현하는 ROC(Receiver Operating Characteristic) 곡선의 수치적 성능인 AUC(Area Under the ROC Curve)를 이용한다. 각 class 별 성능은 Table 2 와 같다.

Table 2. AUC score

transform \ class	A	B	C
MTF	89.7	94.1	83.7
GAF	91.6	99.4	96.8
PAA + GAF	99.4	99.8	99.8

실험 결과, PAA 변환 후 GAF 변환을 한 데이터가 99.4% ~ 99.8% 의 AUC 점수로 가장 높은 성능을 달성하였다. 이를 통해, PAA 가 시계열 데이터의 주요 특징을 강조하면서도 데이터의 노이즈를 줄여, GAF 변환 시 중요한 패턴을 더욱 명확하게 했으며, 시각적 패턴의 명료성을 높여 PatchCore 모델이 이를 더 정확하게 인식할 수 있음을 확인할 수 있다. 또한 해당 결과는 이미지로 변환된 시계열 데이터가 객체 기반의 이미지로 사전 학습된 모델에서도 높은 성능을 달성할 수 있음을 시사한다.

IV. 결론

본 논문에서는 한정된 이차계통 발전소 데이터를 통한 이상 탐지를 구현하고 기존의 이미지 기반 모델을 활용한 빠르고 정확한 성능을 달성하는 것을 목표로 한다. 이를 위해 이차계통 발전소에서의 배관 누출 상황을 모방한 데이터 셋을 통해 이상 탐지를 수행하였다. 또한, MTF, GAF, PAA 를 통해 시계열 데이터를 이미지로 변환하고 이를 객체 이미지 기반으로 사전 학습된 PatchCore 로 학습하고 검증하였다. 이를 통해 기존의 이미지 기반의 모델을 활용하여 시계열 데이터를 빠르고 정확하게 처리할 수 있음을 확인하였다. 따라서 본 연구는 이차계통 발전소 환경에서의 이상 탐지 시스템의 효율성과 정확성을 강화하며, 안정성의 확보에 기여할 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government. (Ministry of Science and ICT, MIST) (No. RS-2022-00144000 & 2022R1G1A1010564)

참 고 문 헌

[1] Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P. "Towards total recall in industrial anomaly detection," IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 14318-14328. 2022.

[2] Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L. "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248-255. 2009.

[3] Zagoruyko, S., Komodakis, N. "Wide residual networks," arXiv preprint arXiv:1605.07146. 2016.