

# 약한 지도학습을 이용한 오토인코더 기반 오디오 데이터 이상 탐지 알고리즘의 성능 향상

문용근\*, 권민성\*, 이병주\*\*, 노정훈<sup>o</sup>

## Performance Enhancement of Autoencoder-Based Audio Data Anomaly Detection via Weakly Supervised Learning

Yong-Geun Moon\*, Min-Seong Kwon\*, Byungju Lee\*\*, Jung-hoon Noh<sup>o</sup>

### 요약

최근 산업시설의 고장 및 진단을 위해 딥 러닝을 이용한 이상 탐지 기술이 큰 관심을 얻고 있다. 일반적으로 고장 진단 및 예측 분야는 다른 분야와는 달리 이상 데이터가 정상 데이터 수에 비해 현저히 적다는 특징이 있다. 따라서, 대부분의 데이터가 정상이라고 가정하고 데이터의 레이블 없이 학습하는 비지도 학습을 적용하는 것이 일반적이다. 하지만, 대부분의 정상 데이터로만 학습된 모델에서 높은 정확도의 정상과 이상 데이터 분류성능을 기대하기는 어렵다. 본 논문에서는 비지도 학습의 대표적인 모델인 오토인코더의 성능을 개선하는 방법으로 소량의 이상 데이터를 학습에 사용하는 약한 지도학습 기반 오토인코더를 제안한다. 제안하는 기법이 소량의 이상 데이터만으로도 낮은 복잡도를 유지하면서 기존의 여러 이상탐지 모델 대비 성능 향상에 효과적임을 실험을 통해 보인다.

**키워드** : 이상 탐지, 인공신경망, 오토인코더, 약한 지도학습

**Key Words** : Anomaly detection, Artificial neural network, Autoencoder, Weakly supervised learning

### ABSTRACT

Recently, deep-learning based anomaly detection technology for a failure diagnosis of industrial facilities is receiving great attention. In general, the field of fault diagnosis and prediction has a characteristic that the number of abnormal data is significantly smaller than the number of normal data. Therefore, it is common to apply unsupervised learning assuming that most data are normal and train a model without labels. However, it is difficult to expect high accuracy in unsupervised learning owing to the lack of knowledge of anomaly features. In this paper, we propose a weakly supervised learning-based autoencoder that uses a small amount of abnormal data for learning features of anomalies, thus improve the performance of autoencoder. Experiments show that the proposed weakly supervised autoencoder is effective in improving the detection performance compared to the various conventional anomaly detection models while having low complexity.

※ 이 연구는 금오공과대학교 학술연구비로 지원되었음 (2022년)

\* First Author : Kumoh National Institute of Technology, Department of Aeronautics, Mechanical and Electronic Convergence Engineer, masuriji@kumoh.ac.kr, 학생회원

<sup>o</sup> Corresponding Author : Kumoh National Institute of Technology, Department of Aeronautics, Mechanical and Electronic Convergence Engineer, jhnoh@kumoh.ac.kr, 정회원

\* Kumoh National Institute of Technology, 20170058@kumoh.ac.kr, 학생회원

\*\* Incheon National University, bjlee@inu.ac.kr, 종신회원

논문번호 : 202210-232-C-RE, Received September 29, 2022; Revised December 31, 2022; Accepted January 2, 2023

## I. 서론

산업 전반에 걸쳐 4차 산업혁명이 빠르게 진행되고 있는 가운데 생산설비의 이상은 단순히 설비 하나의 이상 뿐만 아니라 연결된 다른 설비, 이미 생산된 제품 등 많은 부분에서 손실을 야기한다. 최근 IoT 기술의 발달로 생산 환경에서 많은 데이터를 수집할 수 있게 되었고 수집한 데이터를 활용하여 설비의 이상을 판단하는 이상 탐지 알고리즘에 대한 연구가 활발히 진행되고 있다<sup>1</sup>. 또한 제조업 뿐만 아니라 사이버 보안, 금융, 의료 분야 등 다양한 분야에서 이상 탐지 알고리즘이 응용되고 있다<sup>2,3</sup>. 이처럼 다양한 분야에서의 이상 탐지를 위한 다양한 기법이 제안되었지만<sup>4-6</sup>, 기존의 기법은 높은 차원과 비선형적 특성을 가지는 데이터는 구분하는데 한계를 보인다. 최근에는 이러한 한계를 극복하기 위해 딥러닝 기반의 이상 탐지 기법에 대한 폭넓은 연구가 이루어지고 있으며, 기존의 기법에 비해 높은 이상 탐지 성능을 보이는 것으로 또한 확인되었다<sup>7,8</sup>.

딥러닝 기반 이상 탐지 알고리즘의 경우 다수의 레이블 데이터 확보가 어렵고 이상 데이터가 매우 적게 발생하는 환경에 따라 정상 데이터만을 이용하는 비지도 학습(Unsupervised learning)이 주로 사용되고 있다<sup>9</sup>. 하지만 비지도 학습의 경우 레이블 없이 학습을 진행하는 만큼 레이블이 있는 데이터를 사용하는 지도학습에 비해 성능이 낮다는 단점이 있다. 최근에는 비지도 학습의 단점인 낮은 성능을 보완하는 기법으로 매우 적은 양의 이상 데이터를 활용하여 학습하는 약한 지도학습(Weakly supervised learning) 기법이 제시되고 있다<sup>10,11</sup>.

약한 지도학습은 비지도 학습과 다르게 이상 데이터를 사용하면서 지도학습과는 다르게 매우 적은 양의 이상 데이터만을 학습에 사용하는 기법이다. 이는 지도학습의 단점인 많은 수의 데이터를 확보해야 하는 비실용성과, 정상 데이터만을 사용하여 학습하는 비지도 학습의 단점인 낮은 정확도를 모두 해결할 수 있는 기법이다. 비지도 학습 이상 탐지 알고리즘의 경우 학습을 통해 전체 데이터의 군집을 형성하는 것을 목표로 하는 데 비해, 약한 지도학습을 활용한 이상 탐지 알고리즘은 정상 데이터는 비지도 학습과 같이 군집을 형성하도록 학습하고 이상 데이터는 군집에서 멀어지는 방향인 역방향의 학습을 진행한다. 따라서, 매우 적은 양의 이상 데이터라도 이를 활용할 경우 비지도 학습에 비해 전체 알고리즘에서의 성능 향상을 기대할 수 있다.

본 논문에서는 이상 탐지 알고리즘에서 주로 사용되는 비지도 학습 기법인 오토인코더(Autoencoder) 모델에서, 정상 데이터만을 사용하는 비지도 학습이 아닌 소량의 이상 데이터를 활용하는 약한 지도학습 기반의 오토인코더를 제안한다. 제안되는 이상탐지 알고리즘은 특별히 산업현장에서의 고장 및 진단에 관한 적용을 고려한다. 산업현장에서의 이상 및 고장 탐지를 위해 사용되는 데이터는 대부분 진동이나 음향의 형태를 띠게 된다. 이에 산업현장의 오디오 데이터셋을 통하여 제안된 오토인코더의 성능 향상을 확인한다. 하지만, 제안되는 약한 지도학습 기반 오토인코더는 특정 분야에 제한되지 않고 다른 형태의 데이터셋에도 적용 가능하다.

## II. 관련 연구

### 2.1 오토인코더 기반 이상탐지

비지도 학습의 대표적 모델인 오토인코더는 학습 과정에서 입력과 출력을 같게 하도록 학습한다<sup>11,12</sup>. 학습 과정에서 오토인코더의 손실 함수는 입력 데이터와 모델이 복원한 출력 데이터와의 차를 손실 함수로 정의한다. 입력  $\mathbf{x}$ 와 오토인코더를 통해 복원한 출력 데이터  $AE_{\theta}(\mathbf{x})$ 에 대한 오토인코더의 손실함수  $L(\mathbf{x}; \theta)$ 는 다음과 같다.

$$L(\mathbf{x}; \theta) = \|\mathbf{x} - AE_{\theta}(\mathbf{x})\|^2 \quad (1)$$

여기서  $AE_{\theta}(\cdot)$ 는 오토인코더의 함수이며  $\theta$ 는 오토인코더의 모델 파라미터이다. 결국 다수의 학습 데이터에 대해 손실함수  $L(\mathbf{x}; \theta)$ 가 최소화하도록 오토인코더가 학습되고 오토인코더는 정상 데이터의 특징을 인지하게 된다. 이렇게 학습된 오토인코더는 정상 데이터에 대해서는 학습된 특징을 바탕으로 입력과 큰 차이 없이 복원하지만, 이상 데이터에 대해서는 학습되지 않는 특징을 복원하지 못하게 되어 높은 손실 값을 출력한다.

오토인코더의 높은 실용성과 활용성으로 인해 이를 응용한 많은 연구들이 진행되었다. AnoGAN<sup>13</sup>는 잡음으로부터 모조 데이터를 생성하는 생성모델과, 생성 모델이 생성한 모조 데이터와 실제 데이터를 분류하는 분류모델로 이루어져 있다. 서로 적대적으로 학습을 마친 분류모델은 생성모델이 생성한 데이터 뿐만 아니라 실제 불량 데이터 또한 높은 이상점수를 산출할 수 있다고 가정한다.

USAD<sup>14</sup>는 오토인코더에 적대적 학습을 접목한

네트워크로, 인코더를 공유하는 두 개의 오토인코더로 구성되어 두 단계로 학습한다. 오토인코더2는 학습과정에서 주어진 데이터가 오토인코더1이 생성한 데이터인지, 원본 데이터인지를 구분하도록 학습된다.

DAGMM<sup>[15]</sup>은 오토인코더를 사용하여 얻은 저차원 표현을 통해, 가우시안 혼합 모델 프레임워크를 바탕으로 우도를 계산한다. 언급한 논문 모두 오토인코더를 기반으로 하고 실제로 다양한 이상탐지 분야에서 성능을 보였지만, 소수의 비정상 데이터를 고려하지 않는 비지도학습 기반 이상탐지이기에 성능 한계를 보인다. 또한 DAGMM의 경우 음향데이터에 대해서 적용된 사례<sup>[16]</sup>에서 보여지듯이 하이퍼파라미터에 따라 성능이 크게 변화하여 이에 대한 최적화가 중요하게 작용한다.

## 2.2 Deep-SVDD

Deep SVDD<sup>[10]</sup>는 기존의 전통적 이상탐지기법인 SVDD (Support vector data description)에서 커널 함수를 통한 데이터의 매핑을 딥러닝 신경망을 통한 매핑으로 대체하고, SVDD와 같이 Hypersphere를 통해 데이터를 매핑함으로써 이상 탐지 문제를 단일 분류 문제로 대체하여 해결하는 기법이다. Deep SVDD는 정상 데이터의 특징이 신경망을 통해 Hypersphere의 중앙으로 매핑 되도록 학습한다. 정상 데이터와 이상 데이터의 구분은 매핑 된 데이터와 Hypersphere의 중앙과의 거리를 통해 결정된다.

## 2.3 지도학습

지도학습(Supervised learning)은 딥 러닝의 학습과정에서 레이블이 있는 데이터를 이용하여 학습하는 기계학습 기법을 말한다. 이상탐지 어플리케이션에서는 정상과 이상의 명확한 레이블이 있는 데이터로 학습을 진행하는 만큼 손실함수의 방향이 명확하게 제시되어 전체 성능이 높다는 장점이 있다. 하지만, 이상탐지 분야의 특성 상 이상 데이터가 현저히 적을 뿐만 아니라 이상 패턴 또한 다양하기 때문에 비정상 샘플을 확보하는 데 시간과 비용이 많이 든다는 단점이 있다. 또한 학습 시 정상 데이터와 이상 데이터의 비율이 크게 차이가 나기 때문에 클래스 불균형 문제를 해결해야 한다.

한편 정상 샘플의 클래스가 단일 클래스가 아닌 여러 개의 클래스를 가질 때, 지도학습 기반으로 분류 모델을 학습시키는 방식도 사용될 수 있다. 미국 전자전기공학회(IEEE) 주최 Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop

<sup>[16]</sup>에서도 상위권 분류 모델의 상당수가 지도학습인 MobileNetV2<sup>[17]</sup> 또는 MobileFaceNetV2<sup>[18]</sup> 기반의 알고리즘을 채용하고 있다. 하지만 모델의 복잡도가 비지도학습에 비해 높으며, 단일 모델의 이상탐지에는 적용이 불가능 하다.

## 2.4 DevNet

DevNet<sup>[9]</sup>은 기존 이상 탐지 알고리즘의 입력 데이터의 축소에 따른 특징 추출을 사용하지 않는다. 대신에 입력 데이터 모두를 신경망에 연결하여 학습에 사용하는 중단 간 학습을 통해 이상 점수를 생성하고 생성한 이상점수의 분포를 이용하여 이상 탐지를 수행하는 딥러닝 알고리즘이다. 이는 차원축소에 따른 특징 추출과정에서 최적이지 아니거나 이상과 관련이 없는 표현의 발생을 제거하여 학습 과정 중 데이터의 효율을 높이는 알고리즘이다.

또한 DevNet은 소량의 이상 데이터를 이용하여 정상 데이터와 이상 데이터의 이상 점수의 분포가 분리 되도록 하는 약한 지도학습을 적용한다. 하지만, DevNet은 활용되는 소량의 이상 데이터에 대해 과 적합 되는 경향이 있다. 따라서, 훈련에 사용되지 않은 이상 데이터에 대해서는 좋은 성능을 기대하기 어렵다.

# III. 약한 지도학습 기반 오토인코더

## 3.1 오디오 데이터 전처리

딥 러닝을 통하여 오디오 데이터를 분석하고 활용하기 위해서는 그 특성을 잘 표현하는 특징을 추출하여야 한다. 고전적인 오디오 분석에서는 수행 목적에 따라 시간 또는 주파수 도메인에서 특징을 추출하고 분석과 분류를 하는 것이 일반적이다. 하지만, 오디오 데이터는 시계열 특성을 보유하고 있을 뿐만 아니라, 주파수 기반 전처리를 통해 영상, 이미지의 특성을 가질 수도 있다.

최근, 음악 및 음성 학습 분야에서 오디오의 시계열 데이터를 주파수 변환을 통해 스펙트로그램(spectrogram) 이미지를 생성하고 이를 기존의 고도화된 시각적 이미지 분석 모델을 통해 분석하고 해석하는 시도가 많이 이루어지고 있다. 본 연구는 오디오 데이터의 딥 러닝 학습에 가장 많이 활용되는 멜-스펙트로그램(mel-spectrogram)을 이용하여 오디오 데이터의 특징을 추출한다. 파이썬의 librosa 패키지를 통하여 추출된 멜-스펙트로그램 이미지는 그림 1과 같다. 본 연구에서 사용한 FFT의 길이와 window의 길이는 2048이고 홉 길이는 512이다. 128개의 mel-filter

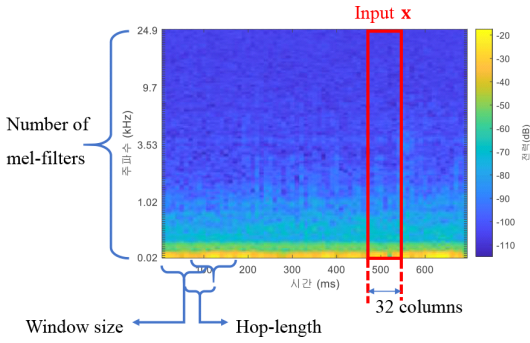


그림 1. 멜-스펙트로그램 변환 예시  
Fig. 1. The example of mel-spectrogram conversion.

를 적용하고, 데이터를 시간에 대해 32개의 열단위로 분할하여 딥러닝 모델에 사용할 데이터를 생성한다.

### 3.2 오토인코더를 이용한 비지도학습 과정

제안되는 기법은 그림 2와 같이 크게 두단계의 과정으로 이루어져 있다. 두 단계 모두 같은 구조의 오토인코더로 이루어져 있지만, 서로 다른 목적을 가지고 있다. 본 연구에서 공통적으로 사용된 오토인코더의 구조는 4개의 128차원의 완전 연결층 (FCN)을 이용해 8차원 잠재공간으로 축소하는 인코더와 8차원의 잠재공간을 4개의 128차원 완전 연결층을 이용해 입력과 같은 크기의 데이터로 복구하는 디코더로 구성되어 있다. 활성화 함수로 정류 선형 유닛(ReLU)이 디코더의 출력층을 제외한 각 완전 연결층 다음에 사용된다.

첫번째 훈련은 레이블이 없는 데이터를 통해 식 (1)

의  $L(\mathbf{x}; \theta)$ 를 최소화하도록 훈련된다. 훈련된 모델을 통해서 모든 데이터의 이상점수  $L(\mathbf{x}; \theta)$ 의 분포를 구하여 정상과 이상 데이터 판별 기준점인  $\alpha$ 를 구한다. 이상점수  $L(\mathbf{x}; \theta)$ 는 입력과 오토인코더를 통해 복원된 출력의 유클리디안 거리의 제곱이기 때문에, 전체 데이터의 이상점수는 감마 분포  $\Gamma(k, \theta)$ 를 따른다고 가정한다. 여기서  $k$ 는 감마 분포의 형상 모수이고  $\theta$ 는 척도모수이다.  $k$ 와  $\theta$ 는 모든 훈련데이터의 이상점수 히스토그램 분포를 토대로 추정한다. 추정된 감마 분포  $\Gamma(k, \theta)$ 와 사전에 정의된 허용 가능한 오경보율 (False alarm rate)  $\rho$ 를 통해 다음의 조건을 만족하는 기준점  $\alpha$ 를 구한다.

$$P_{\Gamma(k, \theta)}(L(\mathbf{x}; \theta) < \alpha) = \rho \quad (2)$$

$\rho$ 는 하이퍼파라미터로 일반적으로 0.8~1사이의 값으로 세팅 된다.

### 3.3 약한 지도학습 과정

두번째 훈련은 약한 지도학습으로서 첫번째 훈련에서 얻어진 모델을 바탕으로 소량의 이상 데이터가 훈련에 추가된다. 소량의 이상 데이터의 이상점수가 기준점  $\alpha$ 보다 위에 분포되도록 모델이 훈련된다. 약한 지도학습 모델의 손실함수  $L_w(\mathbf{x}; \theta)$ 는 Contrastive loss<sup>[9]</sup>에 기반하여 다음과 같이 표현된다.

$$L_w(\mathbf{x}; \theta) = yL(\mathbf{x}; \theta) + (1 - y) \max(\alpha - L(\mathbf{x}; \theta), 0) \quad (3)$$

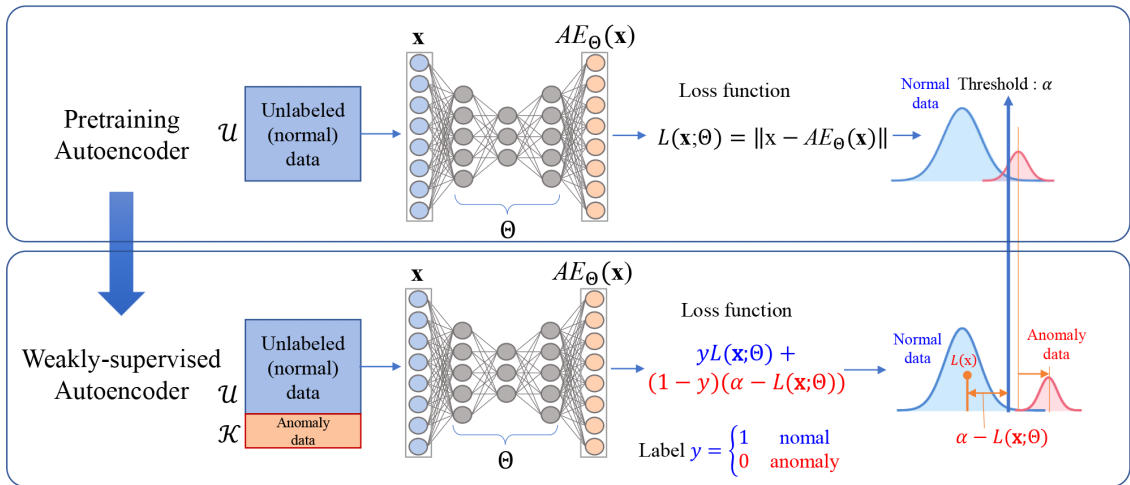


그림 2. 약한 지도 학습 기반 오토인코더 블록도  
Fig. 2. The block diagram of weakly supervised learning based autoencoder.

식 (3)에서  $y$ 는 데이터의 레이블로 정상 데이터는 1, 이상 데이터는 0이다. 약한 지도학습은 그림 2와 같이 정상 데이터에 대해 오토인코더의 입, 출력 오차 최소화를 목표로 학습하고 이상 데이터는 오차가 이상 데이터 판별 기준점  $\alpha$ 보다 크게 분포하도록 학습한다. 약한 지도학습 과정을 통해 딥러닝 모델의 복잡도를 유지하고 이상 데이터가  $\alpha$ 보다 작게 분포하는 것을 억제하여 이상 데이터에 대한 판별 정확도를 높일 수 있다.

한편 정상 데이터가 이상 데이터 보다 큰 비율을

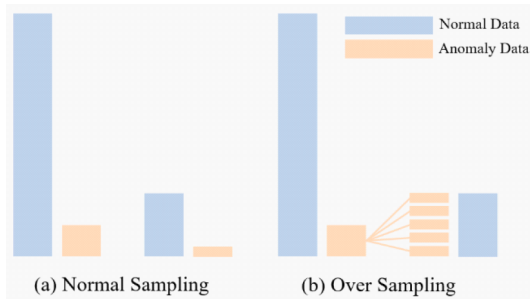


그림 3. 훈련데이터 미니배치 오버샘플링 개념도  
Fig. 3. The concept for oversampling of mini-batch generation.

**Algorithm 1** Training weakly supervised autoencoder

```

Pre-training
Input: Unlabeled dataset  $\mathcal{U}$ 
Output:  $\Theta$ ,  $\alpha$  s.t.  $P_{\Gamma}(L(\mathbf{x}; \Theta) \leq \alpha) = \rho$ 
1: Randomly initialize  $\Theta$ 
2: for  $i = 1$  to  $\text{pre\_n\_epoch}$  do
3:   for  $j = 1$  to  $\text{n\_batch}$  do
4:      $\mathcal{B} \leftarrow$  Randomly sample  $b$  from  $\mathcal{U}$ 
5:      $\text{loss} \leftarrow \frac{1}{b} \sum_{\mathbf{x} \in \mathcal{B}} L(\mathbf{x}; \Theta)$ 
6:     Perform Adam optimizer w.r.t.  $\Theta$ 
7:   End for
8: End for

Main training
Input:  $\Theta$ ,  $\alpha$ , unlabeled dataset  $\mathcal{U}$ , labeled anomaly dataset  $\mathcal{K}$ ,  $y$ 
Output:  $\Theta$ 
1: for  $i = 1$  to  $\text{n\_epoch}$  do
2:   for  $j = 1$  to  $\text{n\_batch}$  do
3:      $\mathcal{B} \leftarrow$  Randomly sample  $b$  with a half from  $\mathcal{U}$ 
       and an another half from  $\mathcal{K}$ 
4:     if  $y == 0$  then
5:        $\text{loss} \leftarrow \frac{1}{b} \sum_{\mathbf{x} \in \mathcal{B}} L(\mathbf{x}; \Theta)$ 
6:     else if  $y == 1$  then
7:        $\text{loss} \leftarrow \frac{1}{b} \sum_{\mathbf{x} \in \mathcal{B}} \max(\alpha - L(\mathbf{x}; \Theta), 0)$ 
8:     end if
9:     Perform Adam optimizer w.r.t.  $\Theta$ 
10:   End for
11: End for
    
```

알고리즘 1. 약한 지도학습 기반 오토인코더 의사코드  
Algorithm 1. The pseudo-code of weakly supervised learning based autoencoder.

차지하는 환경에서는 정상 데이터가 전체 학습을 주도하므로 이상 데이터에 대한 학습을 무력하게 만들고, 이는 전체 학습성능에 영향을 준다. 따라서 배치 단계에서 데이터의 비율을 조정하는 기법을 이용해 배치 내에서 정상, 이상 데이터 비율을 조정하는 기법을 사용하였다.

그림 3은 데이터 배치에서 오버 샘플링의 개념도이다. (a)는 기존의 배치 샘플링 기법인 정상 샘플링 기법이다. 정상 샘플링 기법의 경우 전체 정상, 이상 데이터의 비율이 크게 차이 남에 따라 배치에서의 데이터 비율 또한 전체 데이터 비율과 같이 크게 차이가 난다. (b)의 오버 샘플링 기법은 전체 데이터 비율이 크게 차이 남에도 불구하고 이상 데이터를 추가로 샘플링 함으로써 배치의 데이터 비율이 같게 한다. 데이터 배치에서 오버 샘플링을 통해 매우 낮은 비율을 차지하는 이상 데이터에 대해 정상 데이터와 같은 비율의 학습을 수행함을 목표로 한다. 제안되는 약한 지도학습 기반 오토인코더의 전체 과정을 알고리즘1에 나타내었다.

**IV. 실험 환경 및 결과**

본 논문에서 DCASE 2020의 Task2 데이터 셋<sup>[20]</sup>과 휴대전화 단말기 특정 모델의 진동모터 진동음 데이터를 사용하였다. DCASE2020의 Task2는 기계의 동작음을 분석하여 해당 기계의 정상, 불량을 판별하는 과제이다. 해당 Task는 Toy-car, Toy-Conveyor, Pump, Slider, Fan, Valve 6가지 기계의 오디오 데이터로 구성되어 있다. 휴대전화 단말기 진동모터 진동음 데이터는 실제 휴대전화 생산 과정에서 불량을 검수하기 위해 사용된 데이터로 총 2개 기종의 데이터를 사용하였다. 각각의 데이터 셋의 구성은 표 1과 같다. 데이터 셋의 오디오 데이터를 3.1장에서 언급한 멜-스펙트로그램 변환 과정을 거쳐 특징 데이터를 추

표 1. 데이터셋의 구성  
Table 1. Dataset configuration

	정상데이터	이상데이터
ToyCar	5400	1059
ToyConveyor	5399	1110
Fan	4075	1475
Pump	3749	456
Slider	3504	890
휴대전화 모델 A	2924	326
휴대전화 모델 B	2486	318

출하였다.

모델을 평가하는 지표로는 ROC-AUC(Area under the ROC curve)를 사용하였다. ROC는 적중률(True positive rate)대 오경보율(False positive rate)의 그래프이다. ROC-AUC(이하 AUC)는 ROC 곡선 아래 영역의 넓이를 의미하며, 높은 AUC는 클래스를 구별하는 모델의 성능이 좋다는 의미를 가지고 있다<sup>21)</sup>.

정상 데이터만을 사용한 비지도 사전 학습과 약한 지도 학습 모두 epoch을 2000회로 학습하였다. 최적화 알고리즘으로는 ADAM을 사용하고 학습률을 0.001로 설정하였다. 하이퍼파라미터  $\rho$ 는 0.99으로 설정하였고, 이는 Fan데이터셋에 대하여 실험적으로 얻어진 대략적인 최적의 값이다. 실험을 위한 데이터는 테스트를 위한 정상, 이상데이터 각각 200개를 제외하고는 모두 학습에 사용되었다. 표 2에 나타난 하이퍼파라미터의 변화에 따른 성능을 보면,  $\rho$ 가 0.98에서 가장 높은 성능을 보이지만 0.95, 0.99에서도 큰 차이를 보이지는 않아, 다른 데이터 셋에서도 제안된 알고리즘의 전반적인 성능을 분석하기에는  $\rho$ 를 0.99로 적용하기에 큰 무리가 없다고 판단된다.

먼저, DCASE Task2의 Fan 데이터셋을 활용하여 훈련 과정에서 사용한 이상 데이터 수에 따른 AUC의 성능 차이를 확인한다. 이상 데이터의 수를  $n \in$

표 2.  $\rho$ 의 변화에 따른  $\alpha$ 와 AUC성능 비교  
Table 2. Comparison of  $\alpha$  and AUC performance according to  $\rho$

$\rho$	0.8	0.95	0.99	0.999
$\alpha$	13.11	16.01	19.00	22.96
AUC	0.897	0.899	0.923	0.917

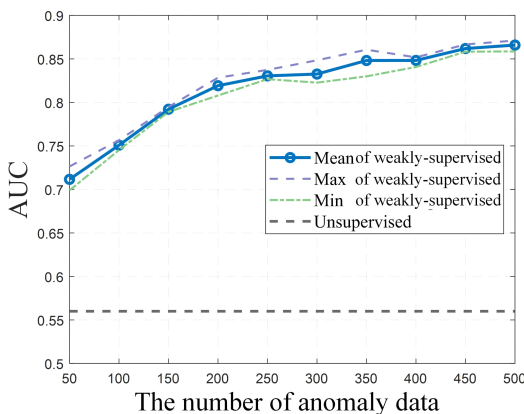


그림 4. 훈련에 사용된 이상 데이터 개수에 따른 성능 비교  
Fig. 4. Performance comparison according to the number of labeled anomaly data used in training

{50, 100, 150, 200, 250, 300, 350, 400, 450, 500}으로 설정하고, 평가 데이터는 정상 200개, 이상 200개로 설정하였으며, 실험은 각  $n$ 마다 4회를 반복하였다. 학습에 사용한 이상 데이터의 개수에 따른 AUC변화는 그림 4와 같다. 기존의 비지도 학습 오토인코더 모델의 경우 AUC가 0.56로 나타난 반면, 제안한 약한 지도 학습 알고리즘을 통한 학습으로 이상 데이터를 50개만 추가한 경우라도 평균 AUC가 0.71로 증가함을 알 수 있다. 또한 학습에 사용된 이상 데이터의 수가 많아질수록 전체 AUC는 상승됨을 보였다.

테스트를 위한 200개의 이상데이터를 제외하모든 이상 데이터를 사용하였을 때 오토인코더와 약한 지도 학습이 적용된 오토인코더, USAD, 지도학습 기반의 MobileFaceNetV2<sup>18)</sup>의 이상 탐지 성능 비교는 표 3과 같다. 오토인코더는 비교를 위한 가장 기본적인 비지도학습 모델이며, USAD는 앞서 언급한 바와 같이 오토인코더와 적대적 훈련이 동시에 적용된 모델이다. MobileFaceNetV2<sup>18)</sup>는 DCASE2020에서 전체 데이터셋에 대하여 6번째로 높은 성능을 보인 모델이다. 특히, Fan데이터에 대하여 가장 높은 성능을 보았다.

표 3에서 보여지듯이 USAD의 경우 기본 모델인 오토인코더에 비해 비슷하거나 약간의 성능 향상만을 보인다. 적대적 학습이 적용되어도 비지도학습의 한계로 성능향상이 제한적임을 알 수 있다. DCASE2021에서 USAD와 유사한 GANomaly기법이 적용된 결과가 [23]에서 보고되었는데, 오토인코더 대비 유의미한 성능차이는 보이지 않는다. 이에 비해 약한 지도 학습이 적용될 경우 사전 훈련 결과인 기본 오토인코더에 비해 성능이 크게 향상 됨을 확인 할 수 있다.

표 3. 각 데이터셋에 대한 AUC 성능 비교  
Table 3. AUC Performance comparison for each dataset

	MobileFaceNetV2	Autoencoder (AE)	USAD	Weakly-supervised AE
ToyCar	0.93	0.78	0.81	0.97
ToyConveyor	0.64	0.72	0.70	0.96
Fan	0.88	0.65	0.66	0.92
Pump	0.91	0.72	0.72	0.92
Slider	0.99	0.84	0.81	0.98
Valve	0.92	0.66	0.67	0.82
휴대전화 모델 A	0.55	0.96	0.95	0.99
휴대전화 모델 B	0.55	0.91	0.91	0.96
평균	0.80	0.78	0.78	0.94

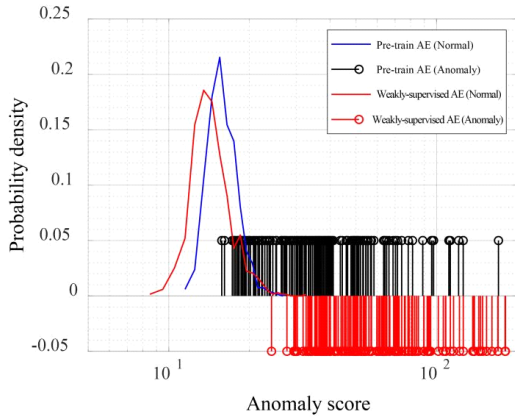


그림 5. 약한 지도학습기반 오토인코더의 적용에 따른 휴대전화 모델 B의 이상점수 분포도 변화  
 Fig. 5. Variation of anomaly score distribution due to the application of weakly supervised learning-based autoencoder

결국, 모델의 복잡도를 높이지 않고도 MobileFaceNetV2보다 높은 성능을 얻을 수 있다. 휴대전화 진동모터 진동음 데이터의 경우 MobileFaceNetV2보다 일반적인 오토인코더가 비교적 우수한 성능을 보였고, 약한 지도학습의 적용시 AUC성능이 1에 가깝게 높이 향상됨을 확인할 수 있다.

휴대폰 모델 B 진동모터의 진동음 데이터에서 약한 지도학습 적용에 따른 이상 점수 분포의 변화를 그림 5에 나타내었다. 정상 데이터는 정규화된 히스토그램으로, 이상 데이터는 stem 형식으로 모든 이상 데이터의 이상점수를 표시하였다. 그림 5에서 보여지듯이 사전학습의 결과와 비교하여 약한 지도학습의 정상과 이상 데이터가 더 크게 분리되어 분포되는 것을 확인할 수 있다.

### V. 결 론

본 논문에서는 이상 탐지 분야에서 비지도 학습의 대표적인 모델인 오토인코더의 성능을 개선하는 방법으로 소량의 이상 데이터를 학습에 사용하는 약한 지도학습 기반의 오토인코더를 제안하였다. 제안하는 기법의 성능을 평가하기 위해 기계음 이상탐지 및 분석에 관련된 두 종의 데이터 셋을 사용하였다. 실험을 통하여 제안된 기법이 소량의 이상 데이터만으로도 기존 비지도학습 기반 딥러닝 모델의 낮은 복잡도를 유지하면서 기존의 지도학습, 비지도학습 대비 이상 탐지 성능 향상에 효과적임을 확인하였다.

### References

- [1] M.-H. Kim and K.-H. Jin, "Development of a deep learning algorithm for anomaly detection of manufacturing facility," *J. KIICE*, vol. 26, no. 2, pp. 199-206, 2022. (<https://doi.org/10.6109/jkiice.2022.26.2.199>)
- [2] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers & Secur.*, vol. 28, no. 1-2, pp. 18-28, 2009. (<https://doi.org/10.1016/j.cose.2008.08.003>)
- [3] M. Ahmed, A. N. Mahmood, and M. R. Islam, "A survey of anomaly detection techniques in financial domain," *Future Gener. Comput. Syst.*, vol. 55, pp. 278-288, 2016. (<https://doi.org/10.1016/j.future.2015.01.001>)
- [4] B. Scholkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Adv. Neural. Inf. Process. Syst.*, vol. 12, pp. 582-588, 1999. (<https://dl.acm.org/doi/10.5555/3009657.3009740>)
- [5] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, pp. 93-104, 2000. (<https://dl.acm.org/doi/10.1145/335191.335388>)
- [6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data.*, vol. 6, no. 1, pp. 1-39, 2012. (<https://dl.acm.org/doi/10.1145/2133360.2133363>)
- [7] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier detection: Methods, models, and classification," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1-37, 2020.
- [8] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1-38, 2021. (<https://dl.acm.org/doi/fullHtml/10.1145/33810>)

- 28)
- [9] G. Pang, et al., “Deep anomaly detection with deviation networks,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. & Data Min.*, pp. 353-362, 2019.  
(<https://dl.acm.org/doi/10.1145/3292500.3330871>)
- [10] L. Ruff, R. A. Vandermeulen, N. Gornitz, A. Binder, E. Muller, K.-R. Muller, and M. Kloft, “Deep semi-supervised anomaly detection,” in *Proc. ICLR*, 2019.  
(<https://doi.org/10.48550/arXiv.1906.02694>)
- [11] C. Zhou and R. C. Paffenroth, “Anomaly detection with robust deep autoencoders,” in *Proc. 23rd ACM SIGKDD Intl. Conf. Knowl. Discov. & Data Min.*, pp. 665-674, 2017.  
(<https://dl.acm.org/doi/10.1145/3097983.3098052>)
- [12] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga, “Outlier detection with autoencoder ensembles,” in *Proc. SIAM Int. Conf. Data Min.*, pp. 90-98, 2017.  
(<https://doi.org/10.1137/1.9781611974973.11>)
- [13] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” in *Proc. Int. Conf. Inf. Process. Med. Imag.*, pp. 146-157, 2017.  
([https://doi.org/10.1007/978-3-319-59050-9\\_12](https://doi.org/10.1007/978-3-319-59050-9_12))
- [14] J. Audibert, P. Michiardi, et al., “USAD: UnSupervised anomaly detection on multivariate time series,” in *Proc. 26th ACM SIGKDD Intl. Conf. Knowl. Discov. & Data Min.*, pp. 3395-3404, 2020.  
(<https://dl.acm.org/doi/10.1145/3394486.3403392>)
- [15] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, “Deep autoencoding Gaussian mixture model for unsupervised anomaly detection,” in *Proc. Int. Conf. Learn. Represent.*, pp. 1-19, 2018.  
(<https://openreview.net/forum?id=BJJLHbb0->)
- [16] H. Purohit, R. Tanabe, T. Endo, K. Suefusa, Y. Nikaido, and Y. Kawaguchi, “Deep autoencoding GMM-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization,” in *Proc. 5th Wkshp. DCASE2020*, Nov. 2020.  
(<https://doi.org/10.48550/arXiv.2009.12042>)
- [17] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, “Description and discussion on DCASE 2021 Challenge Task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions,” *arXiv preprint arXiv:2106.04492*, 2021.  
(<https://doi.org/10.48550/arXiv.2106.04492>)
- [18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510-4520, 2018.  
(<https://doi.org/10.48550/arXiv.1801.04381>)
- [19] Q. Zhou, “ArcFace based sound MobileNets for DCASE2020 task2,” DCASE2020 Challenge, Tech. Rep., 2020.  
([https://dcase.community/documents/challenge2020/technical\\_reports/DCASE2020\\_Zhou\\_22\\_t2.pdf](https://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Zhou_22_t2.pdf))
- [20] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1735-1742, 2006.  
(<https://doi.org/10.1109/CVPR.2006.100>)
- [21] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, “MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” in *Proc. DCASE*, pp. 209-213, 2019.  
(<https://doi.org/10.48550/arXiv.1909.09347>)
- [22] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” in *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 233-240, 2006.  
(<https://dl.acm.org/doi/10.1145/1143844.1143874>)



[23] W. Zhu, J. Ou, Y. Zeng, Y. Zhou, and H. Liu, "Unsupervised anomalous sound detection via semi-supervised ganomaly adversarial training," DCASE2021 Challenge, Tech. Rep., 2021. ([https://dcase.community/documents/challenge2021/technical\\_reports/DCASE2021\\_Liu\\_13\\_t2.pdf](https://dcase.community/documents/challenge2021/technical_reports/DCASE2021_Liu_13_t2.pdf))

**문 용 근 (Yong-Geun Moon)**



2021년 2월 : 금오공과대학교 전자공학부 졸업  
2023년 2월 : 금오공과대학교 전자공학부 석사  
<관심분야> 무선통신, 이상탐지, 임베디드 AI  
[ORCID:0000-0003-0297-8099]

**권 민 성 (Min-Seong Kwon)**



2023년 2월 : 금오공과대학교 전자공학부 졸업  
2023년 3월~현재 : 울산과학기술원 전자공학과 석사과정  
<관심분야> 무선통신, 이상탐지, 머신러닝  
[ORCID:0000-0003-4825-5267]

**이 병 주 (Byungju Lee)**



2008년 2월 : 고려대학교 전파통신공학 학사  
2014년 8월 : 고려대학교 전파통신공학 박사  
2014년 9월~2015년 8월 : 서울대학교 뉴미디어통신공동연구소 선임연구원  
2015년 9월~2017년 5월 : Purdue 대학교 박사후연구원  
2017년 6월~2020년 8월 : 삼성전자 삼성리서치 차세대통신연구센터 책임연구원  
2020년 9월~2022년 2월 : 금오공과대학교 전자공학부 조교수  
2023년 3월~현재 : 인천대학교 정보기술대학 정보통신공학과 조교수  
<관심분야> 차세대 이동통신 시스템, 신호처리  
[ORCID:0000-0002-1710-2324]

**노 정 훈 (Jung-hoon Noh)**



2009년 2월 : 고려대학교 전파통신공학과 학사 졸업  
2014년 2월 : 고려대학교 컴퓨터전파통신공학과 공학박사  
2014년 2월~2021년 2월 : 국방과학연구소 선임연구원  
2021년 3월~현재 : 금오공과대학교 전자공학부 조교수  
<관심분야> 자유공간 광통신, 차세대 이동통신 시스템, 딥러닝 기반 이상탐지, 임베디드 AI  
[ORCID:0000-0002-4599-0584]