

롤러 베어링 내륜 데이터에서 강인한 이상 탐지를 위한 다중 작업 학습 기반 인코더 미세조정 기법

박준휘*¹, 이재현**¹, 김남중*², 박창준*³, 황준***, 나승우****, 곽정환**²

Multi-Task Learning-based Encoder Fine-Tuning Method for Robust Anomaly Detection in Roller Bearing Inner Ring Data

Junhwi Park*¹, Jaehyun Lee**¹, Namjung Kim*², Changjoon Park*³, Joon Hwang***,
Seungwoo Ra****, and Jeonghwan Gwak**²

This research was supported by the Regional Innovation System & Education (RISE) program through the Chungbuk Regional Innovation System & Education Center funded by the Ministry of Education (MOE) and Chungcheongbuk-do, Republic of Korea (2025-RISE-11-004-03), the National Research Foundation of Korea (NRF) grant funded by the Korean government (MOE) (No. RS-2023-00248444), and the Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea under Grant 2022H1D3A2A02060097.

요약

롤러 베어링 내륜에 결함이 발생하면 회전 불균형과 비정상적인 동적 하중이 유발되어 인명 및 재산에 큰 피해가 발생할 수 있다. 본 논문에서는 롤러 베어링 내륜 데이터에서 강인한 이상 탐지를 수행하기 위해 두 가지 기법을 제안한다. 먼저, 복잡한 배경을 가지는 롤러 베어링 내륜 데이터에서 구성 요소별 관심 영역을 정밀하게 추출하는 컴퓨터 비전 기반 전처리 기법을 제안한다. 또한 사전학습 인코더의 일반화 성능에 과도하게 의존하는 기존 이상 탐지 기법의 한계를 극복하기 위해 재구성 작업과 패치 분류 작업을 결합한 다중 작업 학습 기반 인코더 미세조정 기법을 제안한다. 이를 통해 본 연구에서는 제안 기법이 복잡한 배경을 갖는 롤러 베어링 내륜 데이터에서 효과적이고 강인한 이상 탐지 성능을 제공함을 입증하였으며 실제 산업 현장에서의 적용 가능성을 제시한다.

Abstract

Defects in the inner ring of roller bearings can cause rotational imbalance and abnormal dynamic loads, potentially leading to severe damage to human safety and property. This paper proposes two methods to achieve robust anomaly detection on roller bearing inner ring data. First, we introduce a computer vision-based preprocessing method that precisely extracts component-wise regions of interest from roller bearing inner ring data with complex backgrounds. Second, to overcome the limitations of existing anomaly detection methods that overly rely on the generalization capability of pre-trained encoders, we propose a multi-task learning-based encoder fine-tuning approach that combines reconstruction and patch classification tasks. Through these approach, we demonstrate that the proposed method provides effective and robust anomaly detection performance on roller bearing inner ring data with complex backgrounds, and we further present its applicability in real-world industrial environments.

Keywords

roller bearing inner ring, visual anomaly detection, multi-task learning, reconstruction, patch classification

* 국립한국교통대학교 교통·에너지융합학과 박사과정

- ORCID¹: <https://orcid.org/0009-0000-8168-2557>

- ORCID²: <https://orcid.org/0009-0001-4225-6632>

- ORCID³: <https://orcid.org/0009-0000-7305-6097>

** 국립한국교통대학교 소프트웨어학과(**² 교신저자)

- ORCID¹: <https://orcid.org/0009-0006-0120-1340>

- ORCID²: <https://orcid.org/0000-0002-6237-0141>

*** 국립한국교통대학교 항공·기계설계전공 교수

- ORCID: <https://orcid.org/0009-0001-1722-8752>

**** 국립한국교통대학교 이차전지공학과 교수

- ORCID: <https://orcid.org/0009-0007-2726-040X>

· Received: Feb. 04, 2026, Revised: Mar. 16, 2026, Accepted: Mar. 19, 2026

· Corresponding Author: Jeonghwan Gwak

Dept. of Computer Science, Korea National University of Transportation,
South Korea

Tel.: +82-43-841-5852, Email: jgwak@ut.ac.kr

1. 서 론

롤러 베어링(Roller bearing)은 내륜(Inner ring), 외륜(Outer ring), 롤링 요소(Rolling elements) 등으로 구성되어 높은 하중과 충격에 대한 우수한 내구성이 요구되는 중장비, 자동차, 산업용 로봇 등의 분야에서 널리 활용된다[1][2]. 롤러 베어링의 결함은 설치된 장비의 오작동의 주된 원인이 되며 이는 장비 사용자의 생명을 위협할 수 있는 위험 요소로 작용한다[3][4]. 롤러 베어링의 내륜은 결합된 축과 함께 회전하는 역할을 하므로 마모, 균열 등과 같은 결함이 발생할 경우 회전 불균형과 비정상적인 동적 하중이 유발될 수 있다[5]. 따라서 생산 단계에서 신속한 이상 탐지를 수행하는 것은 불량 제품의 유통을 방지하고, 장비의 신뢰성과 안전성을 확보하는 데 필수적이다.

그림 1은 롤러 베어링 내륜 생산 과정에서 각 구성 요소 대단면(LC, Large Cross-Section), 소단면(SC, Small Cross-Section), 외경(OD, Outer Diameter)에 대한 품질 검사를 위해 촬영한 이미지 데이터의 예시이다. 초록색 박스와 같이 각 구성 요소에서 발생하는 이상의 경우 대단면, 소단면, 외경에서 모두 국소적으로 발생한다.

이와 같은 국소 이상은 정상 데이터와 비정상 데이터의 분포 차이를 감소시키기에 이상 탐지 모델의 성능을 제한하는 원인이 된다. 또한 전체 이미지

에서 롤러 베어링 각 구성 요소에 대해 이상 탐지가 수행되어야 하는 관심 영역은 그림 1의 빨간색 마스크 영역과 같이 이미지의 일부에 국한된다. 관심 영역을 제외한 영역은 이상 탐지 모델의 판단과정에서 노이즈로 작용하여 모델이 결함 관련 특징에 집중하는 데 어려움을 초래한다.

산업 도메인 이상 탐지를 위한 딥러닝 기법은 MVTEC AD(MVTEC Anomaly Detection)[6], VisA (Visual Anomaly)[7], Real-IAD(Real-world Industrial Anomaly Detection)[8]과 같은 벤치마크 데이터 세트를 중심으로 다양한 시도[9]-[11]가 이루어져 왔다. 최근에는 사전 학습된 인코더(EN_{pre} , Pre-trained Encoder)의 일반화 성능을 활용한 접근법이 벤치마크 데이터 세트에서 높은 성능을 달성하며 주목받는 현황이다. 하지만 EN_{pre} 의 일반화 성능에 과도하게 의존하는 경우 산업 도메인 데이터가 갖는 분포와 결함 특성을 충분히 반영하지 못해 성능 저하가 발생할 수 있다.

이를 해결하기 위해 본 연구에서는 롤러 베어링 내륜 데이터에서 효과적인 이상 탐지를 위한 다중 작업 학습 기반 인코더 미세조정 기법을 제안한다. 제안 기법은 재구성(Reconstruction)을 통해 정상 데이터의 픽셀 수준 특성을 학습하고, 패치 분류(Patch classification)를 통해 국소적인 이상 특징을 명시적으로 반영함으로써 이상 탐지 성능을 향상시킨다.

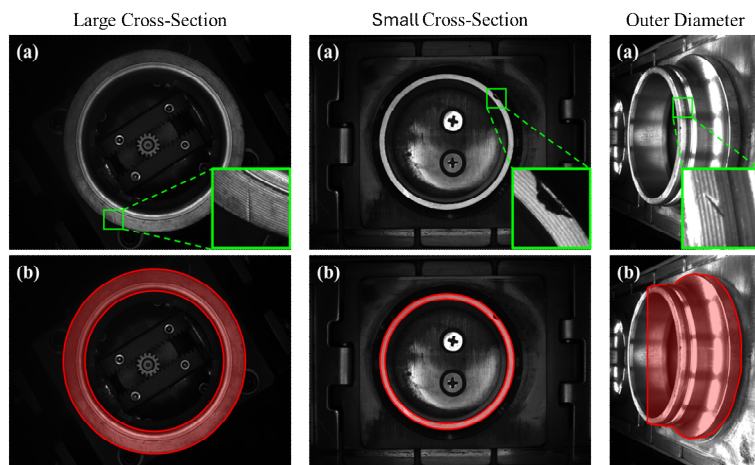


그림 1. 롤러 베어링 내륜 데이터 예시 (a) 원본 이미지 및 이상 영역(초록색 박스), (b) 이상 탐지 관심 영역(빨간색 마스크)

Fig. 1. Example of roller bearing inner ring data

(a) Original image and anomalous region(Green box), (b) Region of interest for anomaly detection.(Red mask)

본 연구의 주요 내용 및 기여는 다음과 같다.

1) 롤러 베어링 내륜 데이터의 구조적 특성을 분석하여 구성 요소(대단면, 소단면, 외경)에 대해 이상 탐지가 요구되는 관심 영역을 정의하고 이에 적합한 전처리 기법을 제안한다.

2) 재구성 및 패치 분류 작업을 수행하는 다중 작업 학습 기반 인코더 미세조정 기법을 통해 EN_{pre} 의 일반화 성능을 유지하면서도 정상 데이터의 특징을 보다 효과적으로 학습할 수 있도록 한다.

3) 실제 산업 데이터를 활용한 연구로써 제안 기법의 기존 기법 대비 강인한 성능을 확인하고 실제 산업 현장에서의 적용 가능성을 제시한다.

II. 관련 연구

본 절에서는 산업 도메인 이상 탐지를 위해 EN_{pre} 의 일반화 성능을 활용한 기법들을 기반 모델 구조에 따라 분류하여 소개한다. 구체적으로 합성곱 신경망(CNN, Convolutional Neural Network) 기반 이상 탐지 기법과 트랜스포머(Transformer) 기반 이상 탐지 기법으로 구분하고, 각 접근 방식의 특징과 한계를 분석한다. 이를 통해 기존 기법들의 공통적인 제약점을 도출하고, 본 연구의 접근 방법이 갖는 차별성을 명확히 한다.

2.1 합성곱 신경망 기반 이상 탐지 기법

이상 탐지 수행을 위해 합성곱 신경망 기반의 EN_{pre} 를 활용하는 모델은 대표적으로 PaDiM(Patch Distribution Modeling)[12]과 PatchCore[13]가 있다. 두 모델 모두 ImageNet[14]과 같은 대규모 데이터 세트로 학습된 ResNet[15] 계열 모델인 Wide ResNet-50-2[16]를 활용하며 각 모델의 특징은 다음과 같다.

PaDiM은 사전학습된 합성곱 신경망의 중간 특징 맵으로부터 위치별 임베딩 분포를 모델링하는 방법으로 각 공간 위치에 대해 다변량 가우시안 분포를 추정하여 이상 점수를 계산한다. PatchCore의 경우 합성곱 신경망 기반 특징을 패치 단위로 추출한 후 정상 데이터로부터 구성된 메모리 बैं크를 기반으로

최근접 이웃 탐색을 통해 이상 탐지를 수행한다. 이러한 합성곱 신경망 기반 이상 탐지 기법들은 사전 학습된 모델의 국소적 특징 표현을 효과적으로 활용함으로써 높은 이상 탐지 성능을 달성하였다.

그러나 합성곱 신경망의 구조적 특성에 따라 전역적인 문맥 정보를 충분히 반영하는 데에는 한계가 있다. 또한, EN_{pre} 의 일반화 성능에 대한 의존도가 높기에 사전학습 데이터와 산업 도메인 데이터의 분포가 큰 차이를 보이는 경우 특징 추출 능력이 저하될 수 있으며, 이로 인해 이상 탐지 성능의 감소가 발생할 가능성을 가진다.

2.2 트랜스포머 기반 이상 탐지 기법

트랜스포머 기반의 EN_{pre} 를 활용하여 이상 탐지를 수행하는 모델은 대표적으로 ViTAD[17]와 Dinomaly[18]가 있으며 MVTec AD, VisA, Real-IAD와 같은 산업 이상 탐지 벤치마크에서 합성곱 신경망 기반 접근법 대비 우수한 성능을 달성하였다. 두 모델 모두 DINO(Self-distillation with no labels)[19] 기반 자기지도 학습을 통해 사전학습된 비전 트랜스포머(Vision transformer)[20] 인코더를 활용하며, 인코더의 특징맵을 입력으로 하는 디코더(DE)를 통해 이상 탐지를 수행한다. 디코더는 정상 데이터에 대해 인코더의 특징맵을 복원하도록 학습되며, 테스트 단계에서는 비정상 데이터 입력 시 인코더 특징과 디코더의 특징 간의 코사인 유사도 기반 거리를 이상 점수로 활용한다.

ViTAD의 경우, 인코더의 특정 레이어에서 추출된 특징 맵에 대응하여 디코더의 개별 레이어가 각각 복원을 수행하는 구조를 가지며 Dinomaly는 인코더의 여러 레이어에서 추출된 특징 맵을 평균하여 하나의 통합된 표현을 구성하고, 이에 대응되는 디코더 특징 맵 또한 평균하여 이상 탐지를 수행한다. 그러나 이러한 트랜스포머 기반 이상 탐지 기법들 또한 EN_{pre} 의 표현 능력에 크게 의존한다는 특징을 가진다. 더 나아가 트랜스포머 기반 모델의 구조적 특성상 전역적 문맥 정보를 효과적인 모델링에 강점을 가지지만 국소적이고 미세한 이상 패턴에 대한 표현 능력이 제한될 수 있다.

2.3 제안 기법의 차별성

본 연구에서는 기존 이상 탐지 기법들의 한계를 극복하기 위해, 다중 작업 학습 기반 EN_{pre} 미세조정 방법론을 제안한다는 점에서 차별성을 가진다. 기존 기법들은 EN_{pre} 의 표현을 고정하거나 제한적으로 활용함으로써 일반화 성능을 유지하고자 하였으나, 이로 인해 산업 도메인 특유의 세밀한 정상 패턴을 충분히 반영하는 데에는 한계가 존재하였다.

반면, 정상 데이터의 세밀한 특징 학습을 위해 인코더를 직접 미세조정하는 경우, 데이터가 제한적인 산업 환경에서는 사전학습을 통해 확보된 일반화 성능이 저하되는 문제가 발생할 수 있다. 본 연구에서는 이러한 문제를 완화하기 위해, 다중 작업 학습 구조를 통해 정상 데이터의 세밀한 표현 학습과 EN_{pre} 의 일반화 성능 유지를 동시에 달성하고자 한다. 이를 통해 기존 단일 작업 기반 이상 탐지 기법 대비, 제한된 데이터 환경에서도 안정적이고 강한 이상 탐지를 가능하게 한다.

III. 제안 기법

3.1 컴퓨터 비전 기반 전처리 파이프라인

본 절에서는 롤러 베어링 내륜의 대단면, 소단면, 외경 영상으로부터 관심 영역을 특정하고 이외의 영역을 제거하기 위한 컴퓨터 비전 기반 전처리 파이프라인을 제안한다. 그림 2의 제안된 파이프라인

은 영상의 각기 다른 구조적 특성을 반영하여 관심 영역 검출의 정밀도를 향상시키는 CGMP (Contrast-Guided Morphological Preprocessing) 모듈, 원형 관심 영역을 탐지하고 크롭 및 마스킹을 수행하는 CGCM(Circle-Guided Cropping & Masking) 모듈, 타원형 관심 영역을 기반으로 크롭을 진행하는 EGGCC(Ellipse-Guided Global Center Cropping) 모듈로 구성된다.

CGMP 모듈의 경우 모든 클래스의 전처리 단계에서 공통적으로 적용되는 과정으로 관심 영역의 경계가 다른 영역 대비 상대적으로 높은 대비를 갖는다는 구조적 특성을 활용하여 CLAHE(Contrast Limited Adaptive Histogram Equalization)를 적용하여 대비를 강화한다. 이후 Gaussian Blur를 통해 노이즈를 억제하고 Otsu 기준에 따른 이진화(Binarization)를 적용하여 전경과 배경을 분리한다. 이후 형태학적 닫힘(Morphological closing)을 적용하여 이진화 과정에서 발생하는 경계 소실 및 미세 공극을 보완한다.

CGCM 모듈은 대단면 및 소단면 영상의 관심 영역이 원형 구조를 갖는다는 특성에 기인하여 그림 3에서 확인할 수 있는 CGMP 모듈의 결과를 입력으로 기하학적 원 후보를 생성한다. 생성된 후보군에 대해 최소 면적 기준을 적용하여 노이즈를 제거한 후 전체 대단면, 소단면 데이터 세트로부터 추정된 원의 평균 반지름 r_{mean} 을 중심으로 허용 오차 δ 를 갖는 구간 $[r_{mean} - \delta_r, r_{mean} + \delta_r]$ 을 만족하는 경우 유효 후보로 간주한다.

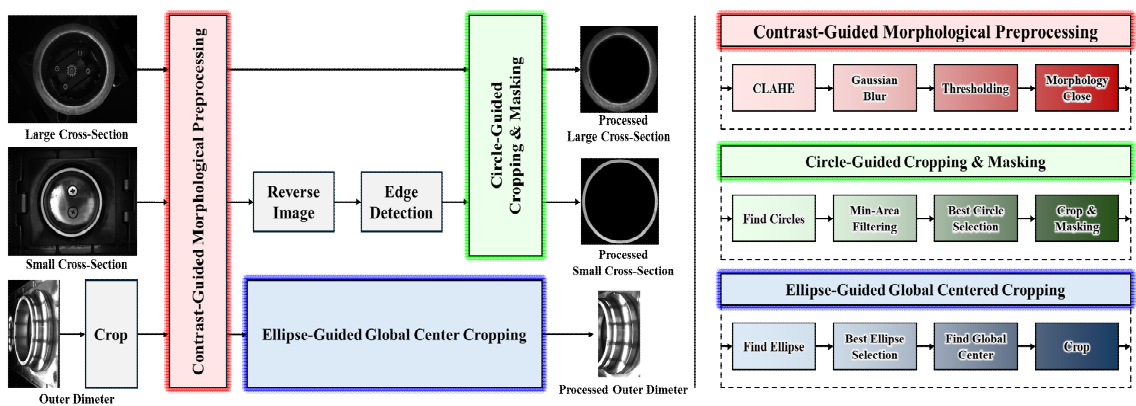


그림 2. 컴퓨터 비전 기반 전처리 파이프라인
Fig. 2. Computer vision-based preprocessing pipeline

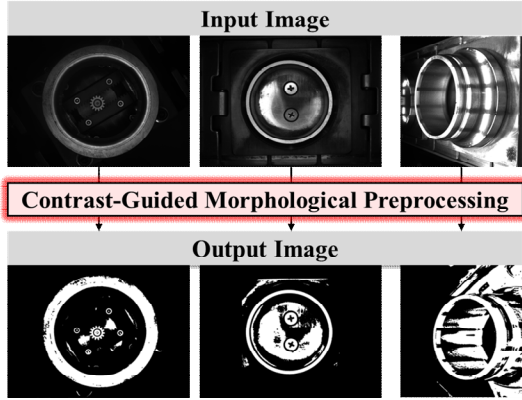


그림 3. CGMP 모듈 처리 결과 시각화
Fig. 3. Visualization of the CGMP module results

이후 크기 기준으로 후보 순서를 내림차순 정렬하고 상위 k 개의 후보군 중 입력 영상의 중심 픽셀과 가장 근접한 후보군을 최종 관심 영역으로 채택한다. 이후 선택된 원의 중심 (c_x, c_y) 과 반지름 c_r 을 기준으로 내부 경계 c_{inner} 와 외부 경계 c_{outer} 를 다음과 같이 정의한다.

$$c_{inner_r} = c_r - \Delta c_r \quad (1)$$

$$c_{outer_r} = c_r + \Delta c_r \quad (2)$$

식 (1)과 식 (2)를 통해 이중 원 구조를 생성하고 c_{inner} 및 c_{outer} 의 내·외부 영역을 마스킹한다. 최종적으로 관심 영역은 원의 중심 (c_x, c_y) 에서 상하좌우 방향으로 c_{inner} 픽셀만큼 확장된 사각 영역이며 해당 픽셀 인덱스 범위를 슬라이딩하여 그림 4와 같이 최종 관심 영역을 추출한다.

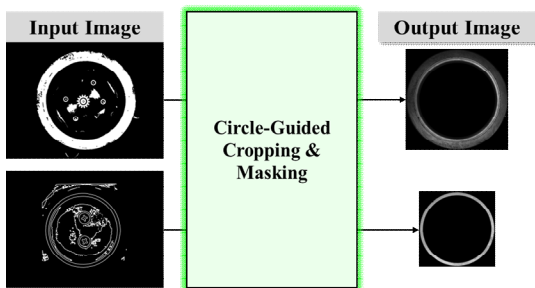


그림 4. CGCM 모듈 결과 시각화
Fig. 4. Visualization of the CGCM module results

외경 영상의 관심 영역 추출을 위한 EGGCC 모듈은 외경 데이터가 세로 방향으로 길게 분포한 타원형 구조를 갖는다는 특성을 반영하여 CGMP 모듈의 결과에 대해 타원 검출을 수행한다. 검출된 타원 후보들은 면적 기준 내림차순 정렬되며 각 후보에 대해 타원 경계를 균일한 각도 간격으로 샘플링하여 생성된 모든 경계점이 영상의 너비 W 와 높이 H 로 정의된 범위 내에 포함되는 경우 해당 타원을 관심 영역으로 간주한다. 위 절차는 전체 외경 이미지 집합에 동일하게 적용되며 각 영상에서 추정된 타원 중심 좌표의 평균을 계산하여 기준 중심으로 사용한다. 이후 모든 영상을 수동 검증하여 기준 중심을 기준으로 상하좌우 방향의 크롭 오프셋 파라미터를 설정하고 해당 파라미터를 전체 데이터 세트에 동일하게 적용하여 그림 5와 같이 최종 크롭 영상을 생성한다.

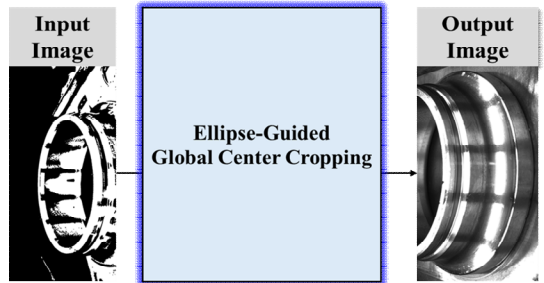


그림 5. EGGCC 모듈 결과 시각화
Fig. 5. Visualization of the EGGCC module results

롤러 베어링 내륜 데이터 세트는 그림 6과 같이 세 가지 밝기 조건으로 구성되어있다. 모든 구성 요소의 세 가지 밝기 조건 중 관심 영역과 배경 간의 명도 대비가 가장 뚜렷하게 나타나는 0번째 밝기 (Brightness Level 0) 영상을 선택하여 전처리를 진행하였다. 소단면 데이터의 경우 CGMP 모듈 적용 후에 영상 반전과 원형 경계 성분을 명확히 강조하기 위한 Canny 에지 검출을 추가로 진행한 후 CGCM 모듈을 적용하였다. 또한 외경 데이터는 동일한 밝기 조건 하에서도 샘플 간 히스토그램 분포의 편차가 크기에 기하학적 구조 검출의 성능 저하를 고려하여 좌우 및 상하 영역에 대해 사전에 정의된 크기대로 크롭을 수행한 뒤 전처리를 적용하여 관심 영역 검출의 신뢰도와 공간적 일관성을 확보한다.

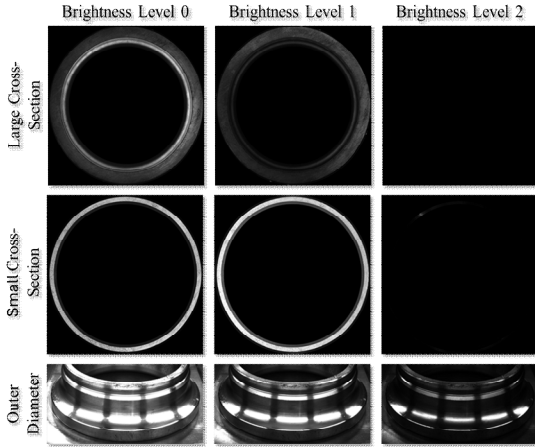


그림 6. 밝기별 이미지 데이터 예시
Fig. 6. Examples of images under different brightness levels

이후 모델 학습에 사용되는 전처리된 이미지는 서로 다른 밝기의 이미지를 결합(Concatenation)하여 모델 학습 및 추론 시 입력으로 활용한다. 이러한 밝기 기반 입력 구성은 동일한 이상이라 하더라도 조명 조건에 따라 표현되는 시각적 특성이 달라질 수 있다는 점을 반영하기 위한 것이다. 그림 6의 예시와 같이 대단면 및 소단면 데이터의 경우 밝기 2

이미지는 촬영된 객체를 거의 확인할 수 없기에 밝기 0과 밝기 1 이미지 간의 중간 밝기 이미지를 생성하여 두 밝기 이미지와 함께 결합하였다. 반면, 외경 데이터의 경우 세 가지 밝기 조건에서 충분한 시각적 특징을 확인할 수 있어, 각 밝기 조건의 이미지를 모두 결합하여 입력으로 활용하였다.

3.2 다중 작업 학습 기반 인코더 미세조정 기법

그림 7과 같이 다중 작업 학습 기반 인코더는 재구성 작업과 패치 분류 작업을 통해 EN_{pre} 미세조정을 진행하였다. 재구성 작업의 경우 정상 데이터를 입력으로 학습 가능한 사전학습 인코더 EN_{pre}^{train} 를 통해 특징맵을 추출하고 다시 입력 데이터와 동일하게 출력하는 과정을 통해 재구성 데이터의 픽셀 수준 특성을 학습할 수 있다. 패치 분류 작업의 경우 비정상 데이터를 입력으로 비전 트랜스포머 모델의 특성에 따라 각 패치 단위로 정상과 비정상 패치 분류를 진행한다. 이 과정을 통해 EN_{pre}^{train} 은 정상 패치와 비정상 패치의 의미적 특징을 특징맵 단위에서 구분할 수 있는 능력을 학습할 수 있다.

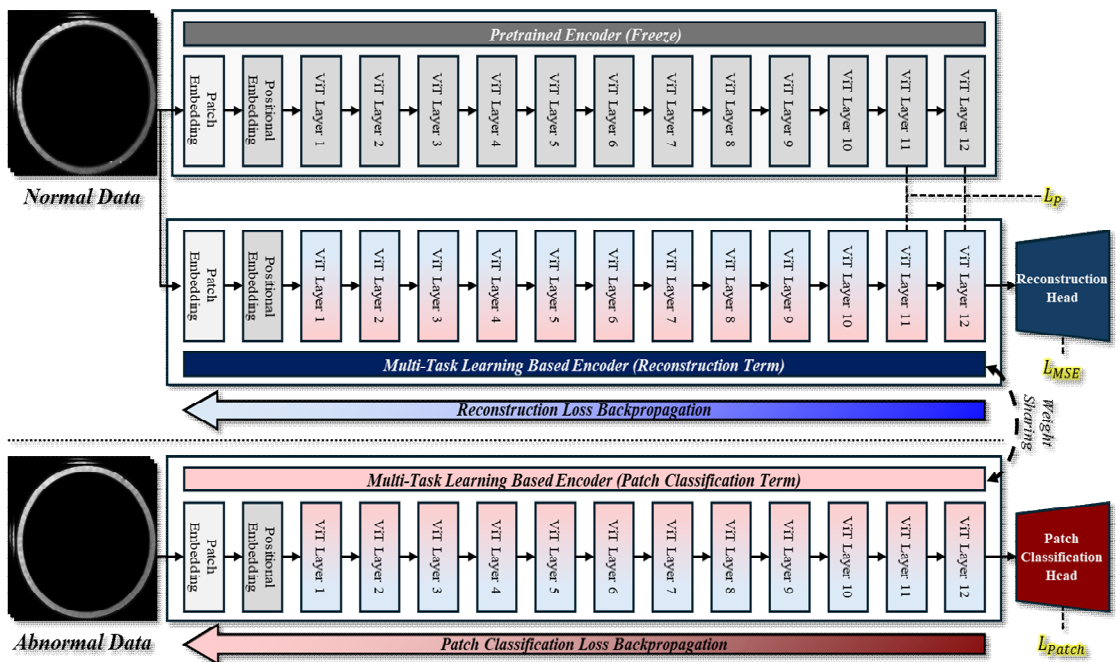


그림 7. 다중 작업 학습 기반 인코더 미세조정 기법 도식
Fig. 7. Architecture of multi-task learning-based encoder fine-tuning method

다중 작업 학습의 근본적인 목적은 한 개 이상의 작업을 동시에 학습하여 각 작업이 상호적으로 작용하는 표현을 인코더가 학습하는 데 있다. 즉, 재구성 작업을 통해 정상 데이터의 전역적이고 세밀한 픽셀 수준 특성을 학습하는 동시에, 패치 분류 작업을 통해 국소적인 이상 패턴의 의미적 특징을 명시적으로 반영함으로써 보다 강한 정상 데이터 특징 표현 획득을 목표로 한다. 그러나 두 가지 이상의 작업을 동일한 인코더에서 동시에 학습하는 경우, 각 작업의 최적화 방향이 상충할 수 있으며 이로 인해 각 작업의 전역 최소점 탐색이 어려울 수 있다[21]. 특히 재구성 작업은 정상 데이터의 픽셀 단위에서 입력 특징을 동일하게 재구성하는 방향으로 학습되는 반면, 패치 분류 작업은 패치 단위에서 정상 패치와 비정상 패치 간의 판별 경계를 강조하는 방향으로 학습되기에 인코더 파라미터 업데이트 과정에서 불안정성이 증가할 수 있다. 이를 해결하기 위해 본 연구에서는 작업 간 간섭을 완화하고 안정적인 인코더 미세조정을 달성하기 위한 학습 방식을 적용하였다.

다중 작업 학습 기반 인코더 미세조정 기법의 목적은 정상 데이터의 세밀한 표현 학습과 EN_{pre}^{train} 의 일반화 성능 유지에 있다. 이 중 정상 데이터의 분포와 구조적 특성을 안정적으로 학습하는 측면에서 입력 데이터를 복원하는 재구성 작업이 패치 분류 작업보다 적합한 학습 신호를 제공한다. 이에 본 연구에서는 재구성 작업을 다중 작업 학습의 주된 학습 목표로 설정하고, 패치 분류 작업은 국소적 이상 특징을 보완적으로 강화하기 위한 보조 작업으로 활용하였다. 세부적인 다중 작업 학습 전략은 다음과 같다.

먼저 재구성 작업의 안정적인 성능 확보를 위해 정상 데이터 x_{normal} 만을 이용한 재구성 학습을 진행한다. 초기 20 에폭(Epoch) 동안 EN_{pre}^{train} 과 재구성 헤드 H_{recon} 은 $SSIM$ (Structural Similarity Index Measure) 기반 시각 손실(L_P , Perceptual Loss)과 입력과 출력 간 픽셀 차이를 측정하는 최소 제곱 오차(MSE , Mean Squared Error) 기반 손실 L_{MSE} 의 합으로 구성된 손실 함수 L_{RC} 를 최적화하도록 학습된다. 여기서, L_P 는 EN_{pre} 와 EN_{pre}^{train} 에서 추출된

특징맵 간의 구조적 유사도를 계산함으로써 정의된다. 이를 통해 입력 데이터의 구조적 특성을 반영함과 동시에 EN_{pre} 의 표현이 EN_{pre}^{train} 의 특징맵과 정렬되도록 유도한다. L_P 는 다음과 같이 정의된다.

$$L_P = \sum_{l \in L} \left(1 - SSIM(F_l^{train}(x), F_l^{pre}(x)) \right) \quad (3)$$

$F_l(x)$ 는 인코더의 l 번째 레이어에서 추출된 특징맵을 의미하며 L 은 L_P 계산에 활용한 레이어 집합이다. L_P 의 목적은 재구성 결과가 입력 이미지와 특징 표현 관점에서 구조적으로 정렬되도록 유도함으로써 재구성 품질을 향상시키는데 있다. 비전 트랜스포머는 패치 간 상관관계 학습을 진행하면서 레이어가 깊어질수록 각 패치의 지역 정보가 통합되어 전역적인 구조 및 문맥 정보를 내포한 고수준 특징 표현을 형성한다. 이와 같은 특성은 재구성 결과와 입력 이미지 간의 전역적 구조 및 문맥 정렬을 목표로 하는 L_P 의 목적과 자연스럽게 부합하기에 본 연구에서는 $L = \{11, 12\}$ 로 설정하였다. L_{MSE} 의 경우 EN_{pre}^{train} 이 추출한 특징맵을 입력으로 하여 재구성 헤드 H_{RC} 가 출력한 재구성 결과 \hat{x} 와 입력 x_{normal} 간의 픽셀 단위 차이를 측정하며 다음과 같이 정의된다.

$$L_{MSE} = \|x_{normal} - \hat{x}\|_2^2 \quad (4)$$

최종적으로 L_{RC} 는 L_P 와 L_{MSE} 의 합으로 정의된다.

$$L_{RC} = L_P + L_{MSE} \quad (5)$$

이와 같은 재구성 기반 학습은 정상 데이터의 구조적 특성과 사전학습 인코더의 일반화 표현을 동시에 반영하는 특징 공간을 형성하는 데 기여한다. 그러나 재구성 작업만으로는 이상 영역에 대한 명시적인 판별 기준을 학습하는 데 한계가 존재한다. 특히 국소적으로 발생하는 미세한 이상 패턴의 경우 정상 데이터 중심의 재구성 학습만으로는 정상과 비정상 간의 경계를 충분히 분리하기 어렵다.

이를 보완하기 위해 재구성 학습을 통해 안정화

된 인코더를 기반으로 패치 단위의 이상 판별을 수행하는 패치 분류 작업을 단계적으로 도입한다. 패치 분류 작업은 비정상 데이터 $x_{abnormal}$ 을 입력으로 각 패치 단위에서 정상 및 비정상 여부를 예측하도록 학습되며 이를 통해 인코더가 국소적인 이상 특징을 보다 명시적으로 반영할 수 있도록 유도한다. 구체적으로 입력 데이터 $x \in \mathbb{R}^{3 \times H \times W}$ 는 비전 트랜스포머의 특성에 따라 패치 단위로 분할되며 패치 크기를 $P \times P$ 라고 할 때 총 $N = HW/P^2$ 개의 패치로 구성된다. 각 패치는 EN_{pre}^{train} 을 통해 $\mathbb{R}^{N \times hidden}$ 크기의 특징맵으로 변환되며 패치 분류 헤드 H_{patch} 는 각 패치에 대해 정상 또는 비정상 여부를 예측한다. 패치 분류 작업에서는 각 패치 i 에 대해 예측된 확률 \hat{y}_i 과 대응되는 정답 y_i^{patch} 간의 크로스 엔트로피(CE, Cross Entropy) 손실을 계산하며 전체 패치에 대한 평균 손실을 패치 분류 손실로 다음과 같이 정의한다.

$$L_{patch} = \frac{1}{N} \sum_{i=1}^N CE(\hat{y}_i, y_i^{patch}) \quad (6)$$

이와 같은 패치 단위의 분류 학습을 통해 인코더는 이미지 전체의 전역적인 정상 패턴뿐만 아니라, 국소적으로 발생하는 미세한 이상 영역에 대해서도 민감하게 반응하는 특징 표현을 학습할 수 있다.

이후 재구성 작업과 패치 분류 작업을 5 에폭 단위로 교대로 수행하며, 각 단계에서 동일한 인코더 파라미터를 공유하여 업데이트한다. 이러한 교대 학습 전략을 통해 인코더는 정상 데이터의 세밀한 구조적 특성을 보존하는 동시에, 정상 패치와 비정상 패치를 효과적으로 구분할 수 있는 판별적 표현을 동시에 학습한다.

IV. 실험 결과 및 분석

본 절에서는 다중 작업 학습 기반 인코더 미세조정 기법의 효과성을 검증하기 위해 사용한 데이터 세트 구성과 실험 설정을 설명한다. 또한 ViTAD와 Dinomaly에서 활용되는 기존 사전학습 인코더 EN_{pre} 를 제안 기법을 통해 학습한 EN_{pre}^{train} 으로 대

체하여 정상 데이터 표현 학습 및 이상 탐지 성능 측면에서의 개선 효과를 분석한다.

4.1 실험 환경 설정

모든 실험은 PyTorch 2.9.1(CUDA 12.8) 환경에서 수행되었으며, NVIDIA RTX A6000 GPU 4개로 구성된 서버에서 진행하였다. 학습 및 테스트 결과의 재현성을 위하여 모든 실험 결과는 동일한 랜덤 시드에서 수행하였다.

표 1. 데이터 세트 분포
Table 1. Dataset distribution

Type	Class	Counting unit	Normal	Abnormal	Total
Train	LC	Sample	30	6	36
		Image	90	18	108
	SC	Sample	94	7	101
		Image	282	21	303
	OD	Sample	38	12	50
		Image	114	36	150
Test	LC	Sample	7	23	30
		Image	21	69	90
	SC	Sample	24	29	53
		Image	72	87	159
	OD	Sample	9	46	55
		Image	27	138	165
Total	Sample	202	123	325	
	Image	606	369	975	

실험에 활용한 데이터 세트의 분포는 표 1에서 제시한다. 해당 데이터는 실제 롤러 베어링 이상 탐지 산업 공정에서 획득한 이미지 데이터로 LC는 롤러 베어링 상부를, SC는 롤러 베어링 하부를, OD는 베어링 측면을 60도 간격으로 균일하게 촬영된 영상으로 구성된다. 각 클래스의 이미지는 서로 다른 세 가지 밝기 조건에서 촬영되었으며 본 연구에서는 세 장의 밝기 이미지를 하나의 샘플로 정의하였다. 위 과정을 통해 구축된 학습 데이터에는 부위별 이상 발생 빈도 차이로 인해 정상 및 비정상 데이터 수에 불균형이 나타난다. 정상 및 비정상 데이터의 분류는 각 영상 내 발생하는 국소적 표면 찍힘의 존재 여부에 따라 정상, 비정상으로 분류하였다. 학습 데이터는 제안한 인코더 미세조정 학습과 ViTAD 및 Dinomaly 모델 학습에 사용하였으며, 테

스트 데이터는 제안 기법의 효과성을 검증하기 위한 모델 성능 평가에 활용하였다. 구체적으로, 학습 데이터 중 정상 데이터는 인코더 미세조정 과정에서의 재구성 작업 학습과 ViTAD 및 Dinomaly의 학습에 사용하였으며, 비정상 데이터는 다중 작업 학습 중 패치 분류 작업 학습에 활용하였다.

각 모델 학습을 위한 하이퍼파라미터 설정은 다음과 같다. ViTAD와 Dinomaly 모델 학습에 사용한 하이퍼파라미터는 이미지 해상도를 제외하고 동일하게 설정하였다. 두 모델 모두 최적화기(Optimizer)는 StableAdamW를 사용하였으며 학습률은 2×10^{-3} , β 값은 (0.9, 0.999), weight decay는 1×10^{-4} , ϵ 는 1×10^{-10} 으로 설정하였다. 학습률 스케줄러는 Warm-up이 포함된 Cosine Annealing 기반의 WarmCosineScheduler를 적용하였으며, 초기 학습률 2×10^{-3} , 최종 학습률 2×10^{-4} , warm-up iteration은 100으로 설정하였다. 전체 학습 iteration 수는 두 모델 모두 1,750으로 동일하게 적용하였다.

입력 이미지 해상도는 Dinomaly의 경우 448×448 , ViTAD의 경우 256×256 이며 배치 크기는 모두 16으로 설정하였다. 데이터 증강은 학습 과정에서 대단면 및 소단면 데이터에 대해 최대 ± 15 도의 랜덤 회전 증강을 배치 단위로 무작위 적용하였다.

이미지 단위(I-)와 픽셀 단위(P-)에서 제안 기법의 공정한 정량적 평가를 위해 본 연구에서 설정한 평가지표는 다음과 같다. 이미지와 픽셀 단위에서 공통적으로 AUROC(Area Under the Receiver Operating Characteristic Curve), AP(Average Precision), F1-score로 선정하였으며 픽셀 단위에서는 AUPRO(Area Under the Per-Region Overlap)를 추가로 선정하였다.

AUROC는 다음 식 (7)을 통해 계산된다.

$$AUROC = \int_0^1 TPR(FPR)d(FPR) \quad (7)$$

식 (7)과 같이 AUROC는 전체 범위에 걸친 참 양성 비율(TPR)과 거짓 양성 비율(FPR)의 관계를 반영하므로, 모델의 전반적인 이미지 단위 및 픽셀 단위 분류 성능을 신뢰성 있게 평가할 수 있는 지표이다.

AP의 경우 식 (8)과 같이 정밀도(P, Precision)와 재현율(R, Recall) 곡선(ROC Curve)의 면적으로 계산된다.

$$AP = \sum_{n=1}^N (R_n - R_{n-1})P_n \quad (8)$$

식 (8)에서 R_n 과 P_n 은 ROC Curve에서 재현율과 정밀도의 지점을 의미한다. 이를 통해 정상과 비정상 데이터가 불균형한 상황에서도 모델의 이상 탐지 성능을 신뢰성 있게 평가할 수 있다.

F1-score의 경우 정밀도와 재현율의 조화평균을 통해 계산되며 본 연구에서는 모델이 예측한 이상 점수를 기반으로 가장 높은 F1-score를 가지는 임계값을 통해 계산하였다.

AUPRO의 경우 정상 영역에서의 과도한 오탐지를 억제하면서 분할 성능을 측정하는 지표이다. PRO(Per-Region Overlap)는 각 이상 영역 단위에서 예측된 분할 결과와 실제 이상 영역 간의 겹침 정도를 측정하는 지표이다. AUPRO는 식 (9)와 같이 거짓 양성 비율에 따른 PRO 값을 적분하여 얻은 면적으로 정의된다.

$$AUPRO = \int_0^{FPR(\tau)} PRO(\tau)d(FPR(\tau)) \quad (9)$$

이와 같은 평가지표를 통해 본 연구에서는 제안 기법의 이상 탐지 성능을 이미지 단위 및 픽셀 단위에서 공정하게 평가한다. 이러한 평가지표 구성은 단일 임계값에 의존하지 않고 다양한 관점에서 모델의 성능을 종합적으로 분석할 수 있도록 한다. 특히 이미지 단위와 픽셀 단위 평가를 함께 고려함으로써 분류 및 분할 성능을 동시에 비교할 수 있다. 또한 모델의 예측 결과를 기반으로 정성적 분석을 수행하여 제안 기법의 성능을 기존 방법들과 종합적으로 비교한다.

4.2 실험 결과 및 분석

본 절에서는 ViTAD와 Dinomaly에서 사용되는 기존 사전학습 인코더 기반 구조와 제안한 다중 작업 학습을 통해 미세조정된 인코더를 적용한 구조 간

의 성능 차이를 분석한다. 구체적으로 기존의 $DE\{EN_{pre}(x)\}$ 구조를 다중 작업 학습 기반 인코더 EN_{pre} 로 대체한 $DE\{EN_{pre}^{train}(x)\}$ 구조를 적용하여 롤러 베어링 내륜 데이터 이상 탐지에서의 효과성을 확인한다. 정량적인 실험 결과는 표 2에서 확인할 수 있다.

실험 결과를 통해 Dinomaly의 경우 제안한 다중 작업 학습 기반 인코더 EN_{pre}^{train} 을 적용했을 때 평균적으로 모든 성능 지표가 향상된 것을 확인할 수 있다. 특히 이미지 단위 및 픽셀 단위 평가 지표 전반에서 AUROC, AP, F1-score가 모두 증가하여, 인코더 미세조정이 정상 및 이상 패턴의 분포 학습과 국소 이상 위치 추정 모두에 긍정적인 영향을 미쳤음을 확인하였다. 반면 ViTAD의 경우 제안 기법 적용 시 I-AUROC 및 I-AP 등 이미지 단위 지표에서는 전반적인 성능 향상이 나타났으나, I-F1과 P-AUROC 지표에서는 상대적으로 제한적인 개선 또는 소폭의 성능 저하가 관찰되었다.

이는 ViTAD 구조가 학습에 활용하는 인코더의 특징맵이 Dinomaly 보다 제한적이기에 인코더 미세조정에 따른 품질 높은 특징맵의 기여가 상대적으로 부족해서 발행한 것으로 분석된다. 하지만 ViTAD와 Dinomaly 모두에서 평균 성능 기준으로는

제안한 인코더 미세조정 기법이 이상 탐지 성능을 전반적으로 향상시키는 경향을 보였으며 특히 두 모델 모두 외경 클래스에서 이상 탐지 성능이 크게 개선된 것을 확인하였다.

각 클래스별 성능 지표를 확인해보면 대단면과 소단면에서는 EN_{pre} 기반 모델과 EN_{pre}^{train} 기반 모델의 성능 차이가 미미했다. 하지만 외경 클래스의 경우 제안 기법을 적용하였을 때 성능이 크게 개선된 것을 확인할 수 있다. 이는 학습에 활용된 대단면 및 소단면 데이터가 외경 데이터에 비해 배경 정보가 상대적으로 적고, 균일한 원형 구조를 가지는 등 분포가 비교적 단순하기 때문으로 판단된다. 반면 외경 클래스의 경우 제안 기법을 적용하였을 때 성능이 크게 개선된 것을 확인할 수 있다. 외경 데이터는 배경 영역이 넓고 구조적 복잡성이 높아 다양한 노이즈 요소가 포함되기 쉬운 특성을 가지며, 이로 인해 단일 작업기반 인코더에서는 정상 패턴과 이상 패턴 간의 구분이 어려운 경우가 발생한다. 제안한 다중 작업 학습 기반 인코더 미세조정 기법은 이러한 복잡한 분포에서도 정상 데이터의 구조적 특성을 보다 효과적으로 학습함으로써, 외경 클래스에서 이상 탐지 성능을 유의미하게 향상시키는 결과를 보였다.

표 2. ViTAD와 Dinomaly의 정량적 실험 결과
Table 2. Quantitative Experimental Results of ViTAD and Dinomaly

Model	Encoder type	Class	I-AUROC	I-AP	I-F1	P-AUROC	P-AP	P-F1	P-AUPRO
ViTAD	EN_{pre}	LC	0.9441	0.9839	0.9388	0.8661	0.0095	0.0488	0.4212
		SC	0.7514	0.8009	0.8060	0.9589	0.0154	0.0562	0.8005
		OD	0.6256	0.8969	0.9109	0.8837	0.0099	0.0443	0.5685
		Mean	0.7737	0.8939	0.8852	0.9029	0.0116	0.0498	0.5967
ViTAD	EN_{pre}^{train}	LC	0.8944	0.9687	0.9200	0.8616	0.0120	0.0644	0.4124
		SC	0.7069	0.7614	0.8060	0.9471	0.0141	0.0632	0.7638
		OD	0.8237	0.9623	0.9149	0.8782	0.0370	0.1102	0.6488
		Mean	0.8083	0.8975	0.8803	0.8956	0.0211	0.0792	0.6084
Dinomaly	EN_{pre}	LC	0.9503	0.9880	0.9565	0.9589	0.0273	0.0885	0.7868
		SC	0.8764	0.9111	0.8254	0.9897	0.0692	0.1670	0.9449
		OD	0.7126	0.9219	0.9109	0.9769	0.1023	0.1751	0.8639
		Mean	0.8464	0.9403	0.8976	0.9752	0.0662	0.1436	0.8652
Dinomaly	EN_{pre}^{train}	LC	0.9627	0.9899	0.9545	0.9566	0.0286	0.0870	0.7746
		SC	0.8750	0.9210	0.8235	0.9880	0.0689	0.1746	0.9343
		OD	0.8527	0.9700	0.9167	0.9833	0.1736	0.2567	0.8969
		Mean	0.8968	0.9603	0.8982	0.9760	0.0904	0.1727	0.8686

그림 8에서는 Dinomaly 모델을 대상으로 EN_{pre} 와 EN_{pre}^{train} 기반 모델이 출력한 이상 맵을 통해 정성적 결과를 확인할 수 있다. 모든 클래스에서 EN_{pre}^{train} 기반 모델이 이상 영역에 대해 높은 확신을 보이는 것을 확인할 수 있다. 그러나 이러한 반응은 실제 이상 영역 외의 정상 또는 배경 영역에서도 오검출(False positive)을 동반하며, 이로 인해 픽셀 단위 성능 지표가 상대적으로 낮게 나타나는 원인으로 작용한다. 특히 이상 패턴의 시각적 특징이 미약하면 모델이 정상 패턴과 이상 패턴 간의 경계를 충분히 학습하지 못해 불안정한 이상 맵을 생성하는 경향을 확인하였다.

그림 9는 재구성 작업과 패치 분류 작업을 통해 EN_{pre}^{train} 학습 과정에서 도출된 평가 지표 및 손실 함수 수렴 곡선이다. 학습 초기에는 각 작업이 번갈아가며 학습됨에 따라 평가 지표 및 손실 함수 수렴이 안정적이지 않은 모습을 확인할 수 있다. 하지만 학습 중반 이후부터는 모든 지표가 안정적으로 수렴하는 양상을 보인다. 이는 다중 작업 학습을 통해 인코더가 재구성 및 패치 분류 작업을 동시에 고려한 공통 표현을 효과적으로 학습하였음을 의미한다.

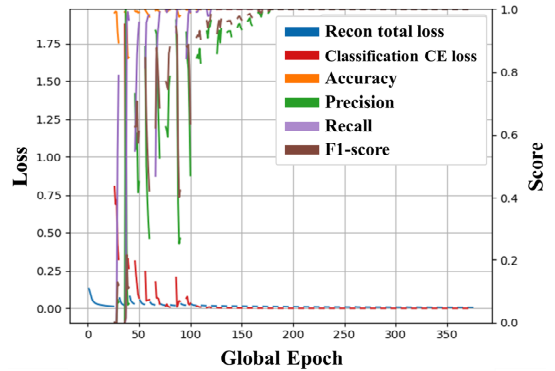


그림 9. 다중 작업 학습 기반 인코더 미세조정 평가 지표 및 손실 함수 수렴 그래프

Fig. 9. Convergence of evaluation metrics and loss in multi-task learning - based encoder fine-tuning

그림 10은 인코더 학습 마지막 에폭에서의 재구성 및 패치 분류 결과를 시각적으로 나타낸 것이다. 제안한 다중 작업 학습 기반 인코더는 재구성과 패치 분류 성능 간의 균형을 효과적으로 유지하며 특정 작업에 편향 없이 모든 작업에서 높은 성능을 보이는 것을 확인할 수 있다. 이는 다중 작업 학습을 통해 인코더가 각 작업에 필요한 특징을 상보적으로 학습하였음을 의미한다.

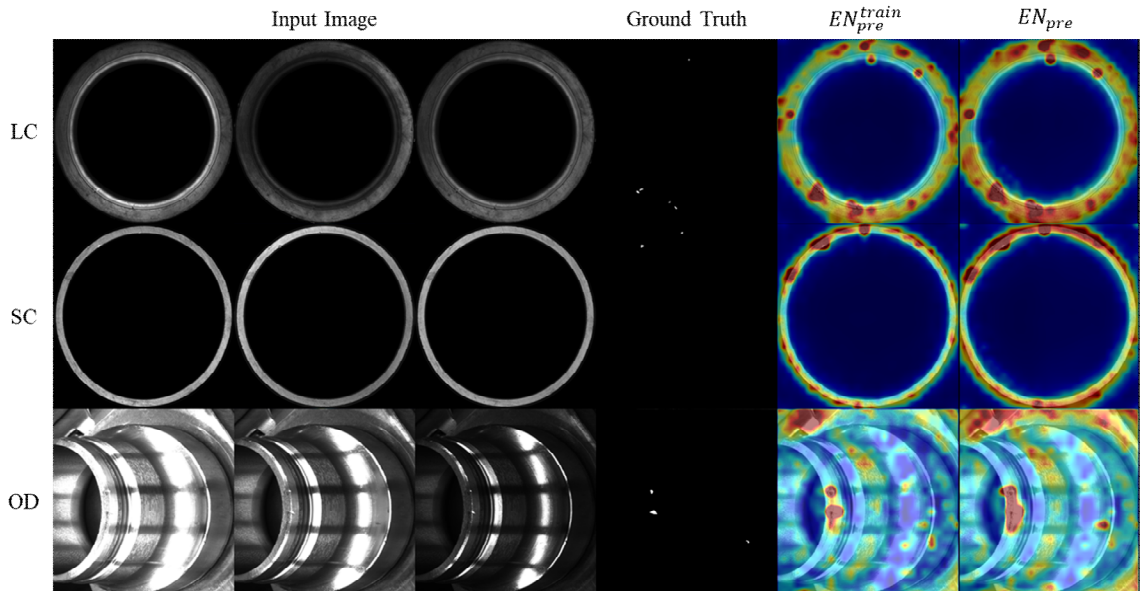


그림 8. Dinomaly의 정성적 실험 결과
Fig. 8. Qualitative experimental results of dinomaly

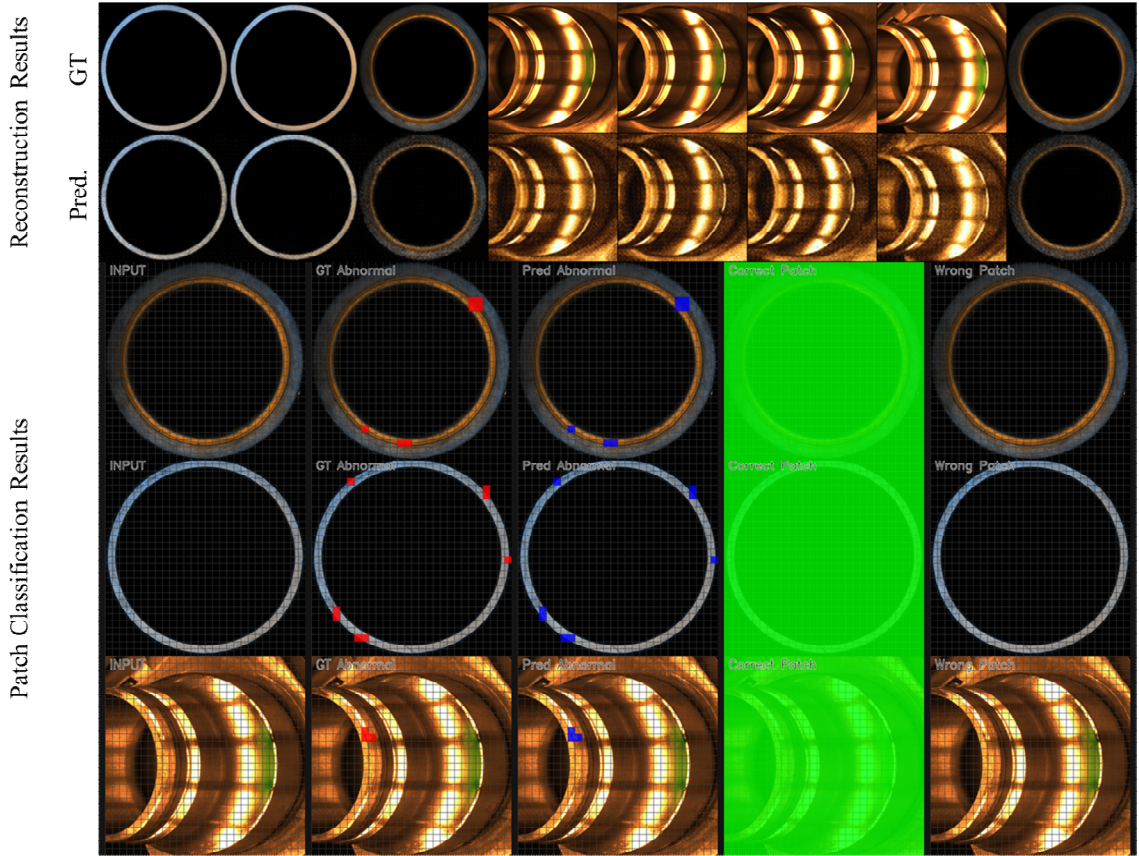


그림 10. 다중 작업 학습 기반 인코더 재구성 및 패치 분류 결과
 Fig. 10. Reconstruction and patch classification results of the multi-task learning - based encoder

이는 인코더 미세조정 과정에서 한 작업의 성능 향상이 다른 작업의 성능 저하로 이어지지 않았음을 의미하며, 제안한 학습 전략이 안정적인 공동 최적화를 달성했음을 보여준다.

제안 기법은 정상 데이터의 분포를 효과적으로 모델링함과 동시에 비정상 데이터에 해당하는 패치의 특징을 명확하게 구분하여 학습할 수 있음을 확인하였다. 이를 통해 인코더는 데이터 분포 변화나 국소적인 이상 패턴에도 강한 표현을 학습할 수 있으며, 전반적인 이상 탐지 성능 향상에 기여할 수 있음을 시사한다.

V. 결론 및 향후 연구

본 연구에서는 기존 이상 탐지 모델이 가지는 EN_{pre} 의 일반화 능력에 과도하게 의존하는 한계를

극복하고자 재구성 작업과 패치 분류 작업을 결합한 다중 작업 학습 기반 인코더 미세조정 기법을 제안하였다. 제안 기법은 정상 데이터의 전역적 구조 특성과 국소적인 이상 패턴을 동시에 학습할 수 있도록 설계하였다. 실험 결과, 제안 기법을 적용한 모델은 ViTAD 및 Dinomaly 대비 전반적으로 이상 탐지 성능 향상을 보였으며 특히 배경 복잡성과 분포 변동성이 큰 외경 데이터에서 이상 탐지 성능이 유의미하게 개선되는 것을 확인하였다. 이는 다중 작업 학습을 통한 인코더 미세조정이 기존 재구성 기반 이상 탐지 모델의 일반화 한계를 효과적으로 완화할 수 있음을 확인하였다. 또한 실제 산업 데이터를 기반으로 수행된 연구로써 제안 기법의 성능 개선을 통해 산업 현장에서의 적용 가능성을 제시하였다.

향후 연구에서는 롤러 베어링 내륜 결합 데이터

의 수량을 확장하여 보다 다양한 조건에서 제안한 방법의 일반화 성능을 검증할 예정이다. 또한 인코더의 표현력을 추가로 향상시키기 위해 트랜스포머 헤드 기반 부스팅 전략을 도입하는 방안도 함께 고려할 계획이다.

References

- [1] B. Sun, Z. Sheng, P. Song, H. Sun, F. Wang, X. Sun, and J. Liu, "State-of-the-Art Detection and Diagnosis Methods for Rolling Bearing Defects: A Comprehensive Review", *Applied Sciences*, Vol. 15, No. 2, Jan. 2025. <https://doi.org/10.3390/app15021001>.
- [2] L. Wen, G. Yang, L. Hu, C. Yang, and K. Feng, "A new unsupervised health index estimation method for bearings early fault detection based on Gaussian mixture model", *Engineering Applications of Artificial Intelligence*, Vol. 128, No. 107562, Article No. 107562, Feb. 2024. <https://doi.org/10.1016/j.engappai.2023.107562>.
- [3] M. E. Haque, M. N. Absur, F. A. Farid, M. K. Siam, J. Uddin, and H. A. Karim, "BearFusionNet: A Multi-Stream Attention-Based Deep Learning Framework with Explainable AI for Accurate Detection of Bearing Casting Defects", *Computers, Materials & Continua*, Vol. 86, No. 3, Jan. 2026. <https://doi.org/10.32604/cmc.2025.071771>.
- [4] L. Jia, T. W. S. Chow, and Y. Yuan, "GTFE-Net: A Gramian Time Frequency Enhancement CNN for bearing fault diagnosis", *Engineering Applications of Artificial Intelligence*, Vol. 119, No. 105794, Article No. 105794, Mar. 2023. <https://doi.org/10.1016/j.engappai.2022.105794>.
- [5] G. Wu, T. Yan, G. Yang, H. Chai, and C. Cao, "A Review on Rolling Bearing Fault Signal Detection Methods Based on Different Sensors", *Sensors*, Vol. 22, No. 21, Article No. 8330, Oct. 2022. <https://doi.org/10.3390/s22218330>.
- [6] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD -- A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", *Proc. IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit. (CVPR)*, Long Beach, CA, USA, pp. 9584-9592, Jun. 2019. <https://doi.org/10.1109/CVPR.2019.00982>.
- [7] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation", *Proc. Eur. Conf. on Comput. Vis. (ECCV)*, Tel Aviv, Israel, pp. 392-408, Oct. 2022. https://doi.org/10.1007/978-3-031-20056-4_23.
- [8] C. Wang, W. Zhu, B.-B. Gao, Z. Gan, J. Zhang, Z. Gu, S. Qian, M. Chen, and L. Ma, "Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection", *Proc. IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit. (CVPR)*, Seattle, WA, USA, pp. 22883-22892, Jun. 2024. <https://doi.org/10.1109/CVPR52733.2024.02159>.
- [9] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, "Deep Industrial Image Anomaly Detection: A Survey", *Machine Intelligence Research*, Vol. 21, No. 1, pp. 104-135, Jan. 2024. <https://doi.org/10.1007/s11633-023-1459-z>.
- [10] V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM - A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection", *Proc. IEEE/CVF Int. Conf. on Comput. Vis. (ICCV)*, Montreal, QC, Canada, pp. 8330-8339, Oct. 2021. <https://doi.org/10.1109/ICCV48922.2021.00822>.
- [11] Z. You, L. Cui, Y. Shen, K. Yang, X. Lu, Y. Zheng, and X. Le, "A Unified Model for Multi-class Anomaly Detection", *Proc. Adv. in Neural Inf. Process. Syst. (NeurIPS)*, New Orleans, LA, USA, Vol. 35, pp. 4571-4584, Oct. 2022. <https://doi.org/10.48550/arXiv.2206.03687>.
- [12] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A Patch Distribution Modeling

- Framework for Anomaly Detection and Localization", Proc. ICPR Int. Workshops and Challenges, Online, Vol. 12664, pp. 475-489, Mar. 2021. https://doi.org/10.1007/978-3-030-68799-1_35.
- [13] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection", Proc. IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 14318-14328, Jun. 2022. <https://doi.org/10.1109/CVPR52688.2022.01392>.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database", Proc. IEEE Conf. on Comput. Vis. and Pattern Recognit. (CVPR), Miami, FL, USA, pp. 248-255, Jun. 2009. <https://doi.org/10.1109/CVPR.2009.5206848>.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", Proc. IEEE Conf. on Comput. Vis. and Pattern Recognit. (CVPR), Las Vegas, NV, USA, pp. 770-778, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.90>.
- [16] S. Zagoruyko and N. Komodakis, "Wide Residual Networks", arXiv preprint arXiv:1605.07146, Jun. 2017. <https://doi.org/10.48550/arXiv.1605.07146>.
- [17] J. Zhang, X. Chen, Y. Wang, C. Wang, Y. Liu, X. Li, M.-H. Yang, and D. Tao, "Exploring plain ViT features for multi-class unsupervised visual anomaly detection", Computer Vision and Image Understanding, Vol. 253, No. 104308, Article No. 104308, Mar. 2025. <https://doi.org/10.1016/j.cviu.2025.104308>.
- [18] J. Guo, S. Lu, W. Zhang, F. Chen, H. Li, and H. Liao, "Dinomaly: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection", Proc. IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 20405-20415, Jun. 2025. <https://doi.org/10.1109/CVPR52734.2025.01900>.
- [19] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging Properties in Self-Supervised Vision Transformers", Proc. IEEE/CVF Int. Conf. on Comput. Vis. (ICCV), Montreal, QC, Canada, pp. 9650-9660, Oct. 2021. <https://doi.org/10.1109/ICCV48922.2021.00951>.
- [20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", Proc. Int. Conf. on Learn. Representations (ICLR), Online, Jan. 2021.
- [21] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient Surgery for Multi-Task Learning", Proc. Advances in Neural Inf. Process. Syst. (NeurIPS), Online, Vol. 33, pp. 5824-5836, Dec. 2020.

저자소개

박 준 휘 (Junhui Park)



2024년 2월 : 국립한국교통대학교
AI로봇공학과(공학사)
2026년 2월 : 국립한국교통대학교
교통·에너지융합학과(공학석사)
2026년 3월 ~ 현재 :
국립한국교통대학교
교통·에너지융합학과 박사과정

관심분야 : 산업 이상탐지, 인공지능, 멀티 태스크 러닝

이 재 현 (Jaehyun Lee)



2025년 2월 : 국립한국교통대학교
컴퓨터공학과(공학사)
2025년 3월 ~ 현재 :
국립한국교통대학교
소프트웨어학과 석사과정
관심분야 : 거대 언어 모델,
경량화, 양자화, 모빌리티

김 남 중 (Namjung Kim)



2023년 2월 : 국립한국교통대학교
소프트웨어학전공(공학사)
2025년 2월 : 국립한국교통대학교
소프트웨어학과(공학석사)
2025년 3월 ~ 현재 :
국립한국교통대학교
교통·에너지융합학과 박사과정

관심분야 : 산업 이상탐지, 생의학 LLM, 인공지능

곽 정 환 (Jeonghwan Gwak)



2014년 8월 : 광주과학기술원
인공지능/기계학습(공학박사)
2014년 9월 ~ 2016년 4월 :
광주과학기술원 박사후 연구원
2016년 5월 ~ 2017년 5월 :
광주과학기술원
전기컴퓨터공학부 연구교수

2017년 6월 ~ 2019년 3월 : 서울대학교병원 연구교수

2019년 4월 ~ 2021년 9월 : 국립한국교통대학교

컴퓨터소프트웨어학과 조교수

2021년 10월 ~ 2024년 9월 : 국립한국교통대학교

컴퓨터소프트웨어학과 부교수

2024년 9월 ~ 현재 : 국립한국교통대학교

컴퓨터소프트웨어학과 교수

관심분야 : 딥러닝, 컴퓨터비전, 신호 및 영상처리,

인공지능, 기계학습

박 창 준 (Changjoon Park)



2023년 2월 : 국립한국교통대학교
AI로봇공학과(공학사)
2025년 2월 : 국립한국교통대학교
교통·에너지융합학과(공학석사)
2025년 3월 ~ 현재 :
국립한국교통대학교
교통·에너지융합학과 박사과정

관심분야 : 시계열 데이터 분석, 인공지능, 기계학습

황 준 (Joon Hwang)



1994년 8월 : 충남대학교 대학원
기계공학과(공학박사)
1996년 8월 ~ 현재 :
국립한국교통대학교
항공·기계설계전공 교수
2008년 6월 ~ 2020년 2월 : 산업부
지역주력산업 육성사업단장

2021년 6월 ~ 현재 : 교육부 대학중점연구소
연구소장/사업단장

2025년 11월 ~ 현재 : 교육부 AI첨단제조공동연구센터
센터장

관심분야 : AI첨단제조, 지능형회격설계, 시스템모니터링

나 승 우 (Seungwoo Ra)



1992년 2월 : 조선대학교
기계설계공학과(공학사)
1994년 8월 : 조선대학교
산업기계공학과(공학석사)
2002년 8월 : 조선대학교
기계설계공학과(공학박사)
2021년 2월 ~ 현재 :

국립한국교통대학교 이차전지공학과 교수

관심분야 : 기계설계, 스마트 팩토리, 산업 이상탐지,
인공지능