

안정적 재구성을 통한 비지도 멀티모달 이상 탐지 방법

Unsupervised Multimodal Anomaly Detection via Stable Reconstruction

정 세 영¹, 이 상 준^{1,*}(Seyoung Jeong¹ and Sang Jun Lee^{1,*})¹Department of Electronics and Information Engineering, Jeonbuk National University

Abstract: Anomaly detection is essential for maintaining product quality and operational safety in industrial environments. However, conventional RGB-based unsupervised methods struggle to capture depth variations and geometric defects, which limits their ability to detect structural anomalies. To address this, we propose a multimodal unsupervised anomaly detection framework that integrates RGB images and 3D information through a stable reconstruction module designed to suppress unstable reconstruction errors in normal regions and enhance the reconstruction gap between normal and anomalous samples. Extensive experiments on the MVTec 3D-AD and Eyecandies datasets demonstrate that the proposed multimodal fusion approach consistently outperforms single-modality baselines in terms of AUROC and AUPRO.

Keywords: Anomaly detection, unsupervised learning, multimodal anomaly detection, deep learning, computer vision

1. 서론

산업 제조 환경에서 발생하는 이상(anomaly) 탐지는 품질 관리와 설비 안정성을 유지하기 위한 핵심 기술로 활용되고 있다. 생산 라인에서 발생하는 결함은 제품 신뢰도 저하뿐 아니라 경제적 손실로 직결되어, 다양한 산업 분야에서는 자동화된 이상 탐지 시스템 수요가 지속적으로 증가하고 있다 [1,2]. 그러나 실제 제조 현장에서는 정상 샘플은 대량으로 수집 가능한 반면, 이상 데이터는 발생 빈도가 매우 낮고 그 형태 또한 일정하지 않아 레이블링이 어렵다. 따라서 정상 데이터만을 활용해 정상 분포를 학습하고, 이러한 분포를 기반으로 이상 여부를 판별하는 비지도 기반 이상 탐지 연구가 활발히 이루어지고 있다 [3].

기존 연구의 대부분은 RGB 영상 기반의 이상 탐지를 중심으로 발전해왔다. RGB 영상은 표면의 색상, 질감과 같은 표면 정보를 포착하는 데에는 효과적이지만, 기하학적 변형, 깊이 변화, 구조적 결함을 인지하는 데에는 근본적인 한계를 지닌다. 특히 조명 변화, 반사, 재질 특성 등에 따라 RGB 기반 방법은 탐지 성능이 크게 흔들리며, 하나의 모달리티만으로 안정적인 이상 탐지를 수행하기 어렵다.

RGB 기반 접근의 한계로 인해 최근에는 멀티모달 이상 탐지가 새로운 방향으로 주목받고 있다 [4]. 이에 대한 선행 연구인 M3DM [5]은 포인트 클라우드를 활용하여 RGB 영상이 포착하기 어려운 구조적 변화와 기하학적 왜곡을 효과적으로 보완함으로써 멀티모달 기반 이상 탐지의 가능성을 제시하였다. 구체적으로, 사전학습된 특징 추출기를 사용하여

RGB 이미지와 3D 포인트 클라우드로부터 각각 특징 임베딩을 추출한 뒤, 픽셀 단위의 cosine similarity를 기반으로 두 모달리티 간 특징을 정렬 및 융합한다. 이후 융합된 특징을 포함한 세 개의 메모리 बैं크를 구성하고, 테스트 단계에서는 입력 샘플의 임베딩과 메모리 बैं크에 저장된 정상 임베딩 간의 거리를 계산함으로써 이상 여부를 판단한다. 그러나 이러한 메모리 बैं크 기반 멀티모달 이상 탐지 방법은 각 모달리티 및 융합된 특징을 모두 저장해야 하므로, 멀티모달 환경에서 메모리 사용량이 급격히 증가하는 한계를 가진다.

이러한 한계를 극복하기 위해 최근에는 모달리티 간의 관계를 학습하는 재구성 기반 이상 탐지 방법이 제안되고 있다. CFM [6]은 각 모달리티의 특징 임베딩을 추출한 후, 한 모달리티의 특징으로부터 다른 모달리티의 특징을 예측하는 cross-modal 임베딩 예측 구조를 학습한다. 정상 데이터에 대해서는 안정적인 재구성이 가능하지만, 이상 데이터의 경우 모달리티 간 일관성이 깨지면서 재구성 오차가 증가한다는 점을 이용해 이상을 탐지한다. 이와 같은 재구성 기반 접근은 메모리 사용량을 줄이고, 멀티모달 정보 간의 상호보완적 특성을 효과적으로 활용할 수 있다는 장점을 지닌다. 그러나 기존의 재구성 기반 이상 탐지 방법은 정상 데이터에 대한 안정적인 재구성 가정을 충분히 확립하지 못하여, 이상 샘플이 입력되었을 때 정상과 이상 간의 재구성 차이를 명확하게 구분하지 못하는 한계를 지닌다. 본 논문에서는 RGB 이미지와 3D 포인트 클라우드를 재구성하는 과정에 Stable reconstruction module (SRM)을 통합함으로써, 정상 데이터에

* Corresponding Author

Manuscript received December 4, 2025; revised January 31, 2026; accepted February 15, 2026

정세영: 전북대학교 전자정보공학부 대학원생(wjtdpdud@jbnu.ac.kr, ORCID[®] 0009-0007-3992-5632)이상준: 전북대학교 부교수(sj.lee@jbnu.ac.kr, ORCID[®] 0000-0002-9312-6299)

※ 상기 논문은 2025년도 ICROS-KROS 호남제주 그랜드 합동 학술대회에서 초안이 발표되었음.

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임 (IITP-2025-RS-2024-00439292).

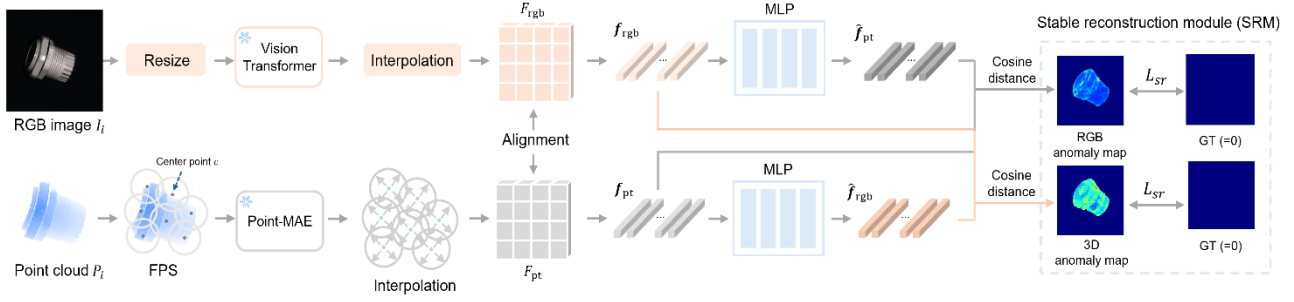


그림 1. 제안하는 멀티모달 이상탐지 네트워크의 파이프라인.

Fig. 1. The pipeline of our proposed network for multimodal anomaly detection.

대한 재구성성을 안정화하고 정상과 이상 간의 재구성 갭을 보다 뚜렷하게 하고자 한다. 이를 통해 재구성 기반 멀티모달 이상 탐지에서 발생하는 오탐을 줄이고, 이상에 대한 민감도를 효과적으로 향상시키는 것을 목표로 한다. 제안하는 방법의 효과를 검증하기 위해, MVTEC3D-AD [7] 및 Eyecandies [8] 데이터셋에서 기존 멀티모달 이상 탐지 기법들과의 정량적·정성적 성능 비교 및 분석을 수행한다.

II. 본론

1. 전체적인 파이프라인

본 논문에서는 재구성 품질의 불안정성을 완화하고 정상 데이터와 이상 데이터 간의 재구성 격차를 보다 명확히 구분하기 위한 새로운 멀티모달 이상 탐지 네트워크를 제안한다. 그림 1은 제안하는 모델의 파이프라인을 나타낸다. 사전 학습된 특징 추출기를 사용하여 특징 임베딩을 추출한 뒤, 이를 기반으로 목표 모달리티 방향으로 재구성을 수행하며, 이 과정에서 정상 데이터의 완전한 재구성을 촉진하고 불필요한 재구성 노이즈를 억제하기 위해 SRM을 통합한다. 테스트 단계에서는 각 모달리티에 대해 재구성 오차 기반의 이상 맵을 생성하고, 픽셀 단위 곱셈을 통해 이를 결합함으로써 최종 이상 맵을 도출한다.

2. 특징 추출 및 정렬

입력 RGB 이미지 $I_i \in \mathbb{R}^{224 \times 224 \times 3}$ 에 대해서는 ImageNet으로 사전 학습된 DINO 기반 ViT-B/8 모델[13]을 사용하여 최종 레이어에서 $28 \times 28 \times 768$ 크기의 특징 맵을 추출한 뒤, bilinear interpolation을 통해 $224 \times 224 \times 768$ 크기의 RGB 특징 맵 F_{rgb} 를 얻는다. $N \times 3$ 크기의 포인트 클라우드 P_i 는 ShapeNet으로 사전 학습된 Point-MAE [14]를 사용하며, farthest point sampling (FPS)을 통해 1,024개의 포인트 그룹으로 분할한 후, 각 그룹에 포함된 32개의 포인트를 입력으로 하여 대표 중심점에 대해 1,152차원의 특징 벡터를 추출한다. 이후 거리 기반 보간을 적용해 모든 원본 포인트에 대한 1,152차원 특징 벡터 P_{pt} 를 복원하고, 각 3D 포인트에 대응되는 RGB 픽셀 좌표 정보를 이용해 이를 2D 이미지 평면으로 투영함으로써 $224 \times 224 \times 1152$ 크기의 포인트 기반 특징 맵 F_{pt} 를 구성한다. 이와 같이 RGB 이미지와 포인트 클라우드로부터 픽셀 단위로 정렬된 특징 맵 F_{rgb} 와 F_{pt} 를 획득하며, 각 픽셀에 대응되는 특징 벡터 f_{rgb} 와 f_{pt} 는 이후 네트워크의 입력으로 사용된다.

3. Stable reconstruction module

RGB 이미지와 포인트 클라우드로부터 픽셀 단위로 정렬된 특징 맵 F_{rgb} 와 F_{pt} 를 획득한 이후, 두 모달리티 간의 의미적 대응 관계를 학습하기 위해 양방향 특징 매핑 함수를 도입한다. 구체적으로, 두 개의 MLP를 사용하여 각각 RGB 특징을 포인트 기반 특징 공간으로, 그리고 포인트 기반 특징을 RGB 특징 공간으로 투영한다.

제안하는 양방향 특징 매핑 구조를 효과적으로 학습하기 위해, 본 연구에서는 모달리티 간 의미적 정렬을 위한 cross-modal projection loss와 정상 데이터에서의 재구성 오류를 안정화하기 위한 stable reconstruction loss를 공동으로 사용한다. 먼저, 서로 다른 모달리티 간의 의미적 일관성을 확보하기 위해 코사인 유사도 기반의 투영 손실을 정의한다. 구체적으로, RGB 및 포인트 클라우드 모달리티에서 추출된 패치 임베딩 $f_{g,rgb}$ 와 $f_{g,pt}$ 를 각각 반대 모달리티의 특징 공간으로 투영하고, 예측된 특징과 대응되는 정답 특징 간의 방향 유사도를 최대화한다. 픽셀 위치 i 에 대한 투영 손실은 다음과 같이 정의된다.

$$L_{proj}^i = \left(1 - \frac{f_{rgb}^i \cdot \hat{f}_{pt}^i}{\|f_{rgb}^i\| \|\hat{f}_{pt}^i\|}\right) + \left(1 - \frac{f_{pt}^i \cdot \hat{f}_{rgb}^i}{\|f_{pt}^i\| \|\hat{f}_{rgb}^i\|}\right). \quad (1)$$

여기서 $\hat{f}_{rgb}^i$ 와 $\hat{f}_{pt}^i$ 는 각각 반대 모달리티의 패치 임베딩을 대응되는 MLP를 통해 투영하여 얻은 픽셀 i 에서의 예측 특징을 의미한다. 해당 손실은 모든 픽셀 위치에 대해 독립적으로 계산된다. 그러나 코사인 유사도 기반 정렬은 특징 벡터의 방향 정보만을 고려하므로, 크기차이로 인해 발생하는 재구성 불안정성을 충분히 반영하지 못한다. 이를 보완하기 위해, 본 연구에서는 예측 임베딩과 정답 임베딩 간의 정규화된 차이를 직접 최소화하는 stable reconstruction loss를 추가로 정의한다. 모달리티 $m \in \{rgb, pt\}$ 에 대해, 픽셀 i 에서의 정규화된 재구성 오차는 다음과 같이 정의된다.

$$d_m^i = \left\| \frac{f_m^i}{\|f_m^i\|} - \frac{\hat{f}_m^i}{\|\hat{f}_m^i\|} \right\|_2. \quad (2)$$

이를 바탕으로, 두 모달리티의 재구성 오차를 모두 고려한 stable reconstruction loss는 다음과 같이 계산된다.

$$L_{sr} = \frac{1}{N} \sum_{i=1}^N ((d_{rgb}^i)^2 + (d_{pt}^i)^2). \quad (3)$$

표 1. MVTEC3D-AD 데이터 세트에 대한 정량적 결과 비교.

Table 1. Comparison of quantitative results on the MVTEC3D-AD dataset.

	Method	Bagel	Cable gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Average
I-AUROC	DepthAE [7]	0.648	0.502	0.650	0.488	0.805	0.522	0.712	0.529	0.540	0.552	0.595
	VoxelAE [7]	0.510	0.540	0.384	0.693	0.446	0.632	0.550	0.494	0.721	0.413	0.538
	AST [9]	0.983	0.873	0.976	0.971	0.932	0.885	0.974	0.981	1.000	0.797	0.937
	BTF [10]	0.938	0.765	0.972	0.888	0.960	0.664	0.904	0.929	0.982	0.726	0.873
	EasyNet [11]	0.991	0.998	0.918	0.968	0.945	0.945	0.905	0.807	0.994	0.793	0.926
	M3DM [5]	0.994	0.909	0.972	0.976	0.960	0.942	0.973	0.899	0.972	0.850	0.945
	CFM [6]	0.994	0.888	0.984	0.993	0.980	0.888	0.941	0.943	0.980	0.953	0.954
	ITNM [12]	0.992	0.951	0.988	0.950	0.999	0.876	0.919	0.965	0.991	0.850	0.948
	Ours	0.998	0.931	0.985	0.999	1.000	0.893	0.991	0.925	0.989	0.963	0.967
P-AUROC	AST [9]	-	-	-	-	-	-	-	-	-	-	0.976
	BTF [10]	0.996	0.991	0.997	0.995	0.995	0.972	0.996	0.998	0.995	0.994	0.993
	M3DM [5]	0.995	0.993	0.997	0.985	0.985	0.984	0.996	0.994	0.997	0.996	0.992
	CFM [6]	-	-	-	-	-	-	-	-	-	-	0.993
	ITNM [12]	-	-	-	-	-	-	-	-	-	-	0.995
	Ours	0.997	0.996	0.998	0.980	0.993	0.995	0.999	0.999	0.999	0.998	0.995
AUPRO	DepthAE [7]	0.432	0.158	0.808	0.491	0.841	0.406	0.262	0.216	0.716	0.478	0.481
	VoxelAE [7]	0.467	0.750	0.808	0.550	0.765	0.473	0.721	0.918	0.019	0.170	0.564
	AST [9]	0.970	0.947	0.981	0.939	0.913	0.906	0.979	0.982	0.889	0.940	0.944
	BTF [10]	0.976	0.967	0.979	0.974	0.971	0.884	0.976	0.981	0.959	0.971	0.964
	EasyNet [11]	0.839	0.864	0.951	0.618	0.828	0.836	0.942	0.889	0.911	0.528	0.821
	M3DM [5]	0.970	0.971	0.979	0.950	0.941	0.932	0.977	0.971	0.971	0.975	0.964
	CFM [6]	0.979	0.972	0.982	0.945	0.950	0.968	0.980	0.982	0.975	0.981	0.971
	ITNM [12]	0.980	0.973	0.982	0.947	0.981	0.967	0.981	0.983	0.982	0.974	0.975
	Ours	0.980	0.979	0.982	0.952	0.969	0.967	0.982	0.982	0.980	0.981	0.976

최종적으로, 전체 학습 목적 함수는 cross-modal projection loss와 stable reconstruction loss의 가중합으로 구성된다:

$$L = L_{\text{proj}}^{\text{rgb}} + L_{\text{proj}}^{\text{pt}} + \alpha L_{\text{sr}}. \quad (4)$$

여기서 α 는 stable reconstruction loss의 기여도를 조절하는 가중치로, 본 연구에서는 0.2로 설정하였다. 이와 같이 projection loss와 reconstruction loss를 공동으로 최적화함으로써, 제안하는 손실 함수는 정상 데이터에 대한 재구성 품질을 안정화하고, 임베딩의 불안정성으로 인해 발생하는 오차를 효과적으로 감소시킨다.

III. 실험

1. 데이터셋

본 연구에서는 제안한 방법의 성능을 평가하기 위해 멀티모달 이상 탐지를 위한 표준 벤치마크인 MVTEC 3D-AD와 Eyecandies 데이터셋을 사용하였다. MVTEC 3D-AD 데이터셋은 총 10개의 산업 카테고리로 구성되어 있으며, 이상 위치를 정밀하게 평가할 수 있는 픽셀 단위 주석을 제공한다. 학습 세트는 2,656개의 정상 샘플로 이루어져 있고, 테스트 세트는 정상 및 이상 샘플을 포함한 총 1,197개로 구성되어 있다.

다. Eyecandies 데이터셋은 산업용 컨베이어 벨트 환경을 모사한 합성 데이터셋으로, 10개의 식품 카테고리에 대해 10,000개의 학습 샘플과 4,000개의 테스트 샘플을 포함한다. 특히, 실제 산업 환경을 반영하기 위해 다양한 조명 조건 및 조명 변화가 시뮬레이션 되어, 조명 변화에 따른 외관 변동성이 존재하는 상황에서도 이상 탐지 성능을 평가할 수 있도록 구성되어 있다. 두 데이터셋 모두 각 샘플에 대해 RGB 이미지와 포인트 클라우드를 함께 제공하며, 이를 통해 각 픽셀 위치에서 색상 정보와 대응되는 3차원 좌표를 동시에 활용할 수 있다.

2. 평가지표 및 실험환경

제안한 방법의 성능은 비지도 이상 탐지 분야에서 널리 사용되는 지표인 Area Under the Receiver Operating Characteristic Curve (AUROC)와 Area Under the Per-Region Overlap curve (AUPRO)를 사용하여 정량적으로 평가하였다. Image-level AUROC (I-AUROC)는 정상 이미지와 이상 이미지 간의 판별 능력을 평가하며, Pixel-level AUROC (P-AUROC)는 픽셀 단위 이상 위치 탐지 정확도를 측정한다. 또한 AUPRO는 예측된 이상 영역과 실제 이상 영역 간의 중첩 정도를 기반으로 분할 성능을 평가하며, 본 실험에서는 AUPRO 곡선의 임계값을 0.3으로 설정하였다. 모든 실험은 NVIDIA H100 GPU 환경

표 2. Eyecandies 데이터 세트에 대한 정량적 결과 비교.

Table 2. Comparison of quantitative results on the Eyecandies dataset.

	Method	Candycane	Chocolate cookie	Chocolate praline	Confetto	Gummy bear	Hazelnut truffle	Licorice sandwich	Lollipop	Marshmallow	Peppermint candy	Average
I-AUROC	AST [9]	-	-	-	-	-	-	-	-	-	-	0.780
	M3DM [5]	0.624	0.958	0.958	1.000	0.886	0.758	0.949	0.836	1.000	1.000	0.897
	EasyNet [11]	0.737	0.934	0.866	0.966	0.717	0.822	0.847	0.863	0.977	0.960	0.869
	CFM [6]	0.680	0.931	0.952	0.880	0.865	0.782	0.917	0.840	0.998	0.962	0.881
	Ours	0.710	0.962	0.958	0.971	0.946	0.830	0.936	0.886	1.000	0.966	0.917
P-AUROC	AST [9]	-	-	-	-	-	-	-	-	-	-	0.902
	M3DM [5]	0.974	0.987	0.962	0.998	0.966	0.941	0.973	0.984	0.996	0.985	0.977
	CFM [6]	0.983	0.982	0.964	0.989	0.949	0.946	0.969	0.980	0.995	0.987	0.974
	Ours	0.984	0.986	0.974	0.994	0.969	0.962	0.968	0.982	0.994	0.994	0.981
	AST [9]	-	-	-	-	-	-	-	-	-	-	0.744
AUPRO	M3DM [5]	0.906	0.923	0.803	0.983	0.855	0.688	0.880	0.906	0.966	0.955	0.882
	CFM [6]	0.942	0.902	0.831	0.965	0.875	0.762	0.791	0.913	0.939	0.949	0.887
	Ours	0.938	0.897	0.820	0.963	0.864	0.808	0.792	0.897	0.941	0.956	0.888

에서 PyTorch 2.2.2와 CUDA 12.1을 사용하여 수행되었으며, 제안한 모델은 Adam 옵티마이저를 사용해 학습률 0.001로 총 250 epoch 동안 학습되었다.

3. MVTEC 3D-AD 데이터셋에서의 비교 실험 결과

표 1은 I-AUROC, P-AUROC, AUPRO를 기준으로 한 정량적 성능 비교 결과를 보여준다. 제안하는 방법은 모든 평가 지표에서 가장 우수한 평균 성능을 달성하며, 기존 접근법 대비 뚜렷한 성능 향상을 보였다. 특히 Tire와 Cookie와 같이 기하학적 변형이 큰 카테고리에서 성능 향상이 두드러지게 나타났는데, 이는 기존 재구성 기반 방법들이 이러한 경우 불안정한 이상 점수를 생성하는 경향이 있기 때문이다. 또한 제안한 방법은 평균 P-AUROC에서도 가장 높은 성능을 기록하여, 픽셀 단위 이상 위치 탐지에서도 우수한 성능을 보였다. 이는 SRM이 정상 영역에서 발생하는 불필요한 이상 반응을 효과적으로 억제함으로써, 정상 영역과 이상 영역 간의 픽셀 수준 판별을 보다 안정적으로 수행함을 의미한다.

제안한 방법의 정성적 성능을 평가하기 위해, MVTEC 3D-AD 데이터셋에서 생성된 최종 이상 맵을 대표적인 기존 방법들과 비교하였다. 그림 2는 기존 방법과의 정성적 비교

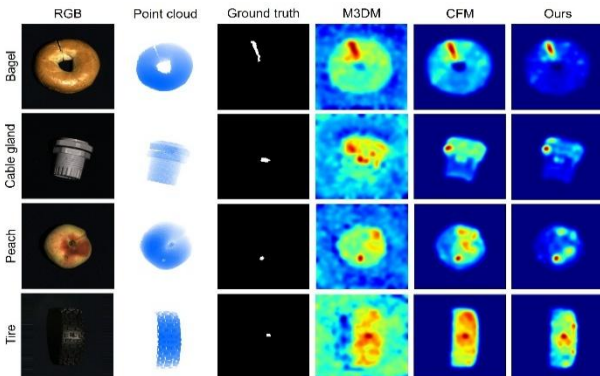


그림 2. MVTEC 3D-AD 데이터 세트에 대한 정성적 결과 비교.

Fig. 2. Qualitative comparison of anomaly localization results on the MVTEC 3D-AD dataset.

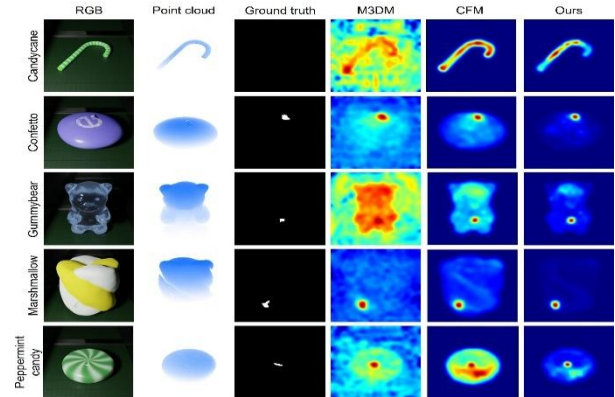


그림 3. Eyecandies 데이터 세트에 대한 정성적 결과 비교.

Fig. 3. Qualitative comparison of anomaly localization results on the Eyecandies dataset.

결과를 보여준다. 기존 방법들은 Cable gland, Peach, Tire와 같이 표면이 불규칙하거나 곡률이 큰 카테고리에서 불안정한 이상 맵을 생성한다. 반면, 제안하는 방법은 정상 영역에 낮은 이상 점수를 부여함으로써 정상 영역과 이상 영역을 보다 명확하게 구분한다. 특히 Cable gland 카테고리는 조명 변화에 민감한 복잡한 기하 구조를 가지는데, 기존 방법들은 조명 영향이 큰 영역에 높은 이상 점수를 부여하는 반면, 제안한 방법은 실제 이상 영역만을 보다 정확하게 탐지하고 분할한다. 또한 Peach 및 Tire 카테고리의 경우, 기존 방법들은 객체 표면 전반에 걸쳐 높은 이상 점수를 부여하는 경향이 있으나, 제안한 방법은 안정적 재구성 손실을 통해 이러한 잡음 이상 반응을 효과적으로 완화하여 보다 정제되고 안정된 이상 맵을 생성한다.

4. Eyecandies 데이터셋에서의 비교 실험 결과

본 절에서는 Eyecandies 데이터셋을 대상으로 기존 방법들과의 정량적 및 정성적 성능 비교 결과를 제시한다. 표 2는 I-AUROC, P-AUROC 그리고 AUPRO 지표를 기준으로 한 정량적 성능 비교 결과를 보여준다. 제안한 방법은 대표적인

최신기법들과 비교했을 때, Eyecandies 데이터셋에 존재하는 다양한 조명 변화에도 불구하고 모든 평가 지표에서 우수한 성능을 달성하였다. 이러한 결과는 보다 도전적인 환경에서도 제안한 방법이 강한 일반화 성능을 보임을 입증한다. 특히 이미지 수준 평가에서 제안한 방법은 Chocolate cookie, Chocolate praline, Gummybear, Peppermint candy 등 여러 객체 카테고리에서 뚜렷한 성능 향상을 보였다. 해당 카테고리에는 미세한 표면 결함과 복잡한 텍스처 변화를 특징으로 하며, 기존 방법에서는 이러한 특성으로 인해 이상 맵의 품질이 저하되었다. 반면, 제안하는 방법은 특징 매핑 과정에서 stable reconstruction loss를 도입함으로써 안정적인 이상 판별 성능을 유지한다. 또한 제안한 방법은 가장 높은 AUPRO 점수를 달성하여, 낮은 오탐률 제약 조건하에서도 강건한 분할 성능을 보임을 확인하였다. 전반적으로 제안한 방법은 다양한 객체 카테고리과 환경 조건 전반에 걸쳐 뛰어난 일반화 성능을 보이며, 미세한 결함과 복잡한 표면 기하를 포함하는 정밀 산업 이상 탐지 과제에 매우 효과적임을 보여준다.

Eyecandies 데이터셋의 정성적으로 비교 결과는 그림 3에 제시되어 있다. 그림에서 확인할 수 있듯이, M3DM은 정상 영역의 상당 부분에 대해 높은 이상 점수를 부여하는 경향이 있는 반면, CFM은 크로스모달 특징 매핑을 통해 이러한 잡음을 부분적으로 완화한다. 그러나 CFM 역시 일부 정상 영역에서 과도한 이상 반응을 보이며, 이상 영역의 경계가 상대적으로 흐릿하게 나타나는 한계를 가진다. 이에 반해 제안한 방법은 정상 영역에서의 잡음을 효과적으로 억제하고, 실제 정답 마스크와 가까운 이상 맵을 생성한다. 특히 Gummybear 카테고리의 경우, 제안한 방법은 정상적인 텍스처 영역에서 발생하는 오탐지를 방지하면서 결함 영역을 정확하게 분할한다. 이러한 정성적 결과는 SRM을 통합함으로써 재구성 과정을 안정화한 제안 방법의 효과를 명확히 보여준다.

VI. 결론

본 논문에서는 RGB 이미지와 포인트 클라우드 데이터를 함께 활용하는 멀티모달 이상 탐지 프레임워크를 제안하였다. 기존 재구성 기반 이상 탐지 방법들이 정상 영역에서도 높은 재구성 오류를 발생시키는 한계를 해결하기 위해, 본 연구에서는 재구성 안정성을 향상시키는 Stable reconstruction module을 통합하였다. 제안한 방법은 정상 샘플에 내재된 변동성을 보다 정확하게 모델링하고, 실제 임베딩과 예측 임베딩 간의 차이를 최소화함으로써 재구성 과정에서 발생하는 잡음을 효과적으로 억제한다. 정성적 실험 결과를 통해 정상 영역에서의 불필요한 재구성 오류가 감소하고, 정상과 이상 샘플 간의 재구성 격차가 보다 뚜렷해짐을 확인하였다. 또한 제안한 방법은 다양한 객체 카테고리과 복잡한 산업 환경 전반에 걸쳐 이상을 효과적으로 탐지하며, MVTec 3D-AD와 Eyecandies 데이터셋에서 이미지 수준 및 픽셀 수준 평가 지표 모두에서 기존 방법 대비 우수한 성능을 달성하였다.

REFERENCES

[1] S. U. Jung and S. G. Kim, "Real-time anomaly detection in CNC machining using multi-head attention VAE with SPC-based thresholding," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 31, no. 12, pp. 1474–1479, Dec.

2025.

[2] S. Lee and J. Kim, "Anomaly detection in encrypted inverted pendulum systems: Simulation-based parameter analysis," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 31, no. 12, pp. 1353–1360, Dec. 2025.

[3] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2022.

[4] S. Jeong and S. J. Lee, "An Analytical study on the effectiveness of multimodal fusion for unsupervised anomaly detection," *Proc. of 2024 ICROS-KROS Honam Jeju Regional Conference (in Korean)*, 2025.

[5] Y. Wang, J. Peng, J. Zhang, R. Yi, Y. Wang, and C. Wang, "Multimodal industrial anomaly detection via hybrid fusion," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8032–8041, 2023.

[6] A. Costanzino, P. Z. Ramirez, G. Lisanti, and L. Di Stefano, "Multimodal industrial anomaly detection by crossmodal feature mapping," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17234–17243, 2024.

[7] P. Bergmann, X. Jin, D. Sattlegger, and C. Steger, "The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization," *arXiv preprint arXiv:2112.09045*, 2021.

[8] L. Bonfiglioli, M. Toschi, D. Silvestri, N. Fioraio, and D. De Gregorio, "The Eyecandies dataset for unsupervised multimodal anomaly detection and localization," *Proceedings of the Asian Conference on Computer Vision*, pp. 3586–3602, 2022.

[9] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Asymmetric student-teacher networks for industrial anomaly detection," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2592–2602, 2023.

[10] E. Horwitz and Y. Hoshen, "Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2968–2977, 2023.

[11] R. Chen, G. Xie, J. Liu, J. Wang, Z. Luo, J. Wang, and F. Zheng, "EasyNet: An easy network for 3D industrial anomaly detection," *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 7038–7046, 2023.

[12] J. Wang, X. Wang, R. Hao, H. Yin, B. Huang, X. Xu, and J. Liu, "Incremental template neighborhood matching for 3D anomaly detection," *Neurocomputing*, vol. 581, p. 127483, 2024.

[13] 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. Comput. Vis.*, pp. 9650–9660, 2021.

[14] Y. Pang, E. H. F. Tay, L. Yuan, and Z. Chen, "Masked autoencoders for 3D point cloud self-supervised learning," *World Scientific Annual Review of Artificial Intelligence*, vol. 1, p. 2440001, 2023.



정 세 영

2025년 전북대학교 전자공학부 학사.
2025년~ 현재 전북대학교 대학원 전자정보공학부 석사과정 재학 중. 관심분야는 결함 검출 및 이상 탐지.

**이 상 준**

2011년 포항공과대학교 전기전자공학과
학사. 2018년 포항공과대학교 전기전자 공
학과 석박사 통합과정. 2020년 9월 ~ 현재
전북대학교 전자공학부 교수. 관심 분야는
컴퓨터비전, 인공지능 및 로보틱스.