

소량 불량 데이터 기반 크롬 도금 손톱깍이 표면검사를 위한 비전-언어 캐시 어댑터 프레임워크

김 경 민 *

A Vision-Language Cache Adapter Framework for Chrome-Plated Nail Clipper Surface Inspection under Few-Shot Defect Data

Gyeongmin Kim*

요 약

크롬 도금 손톱깍이 표면검사는 고광택 곡면, 랙 가림, 공정 변화, 불량 표본 부족으로 일반 지도학습만으로 안정화하기 어렵다. 본 논문은 공정 조건·랙 이력·전후 공정 영상을 연결하고, 소량 불량 이미지와 텍스트 설명을 갱신형 캐시로 통합하는 비전-언어 검사 프레임워크를 제안한다. 예비 검증에서 특징 어댑터 기준선은 1/7(14.29%), 22장 통합 평가는 13/22(59.09%, 95% 신뢰구간 36.4~79.3%)의 이상탐지 정확도를 기록하였다. 백본별 신뢰구간은 서로 겹쳐 우열을 확정할 수 없었고, 학습에 재사용된 13장 파일럿은 독립 근거에서 제외하였다. 촬영 표준화, 독립 평가셋, 클래스 균형, 랙 단위 추적을 포함한 검증 절차를 제시하여 희소 불량 환경에서 모델 갱신성과 공정 추적성의 중요성을 보인다.

Abstract

Surface inspection of chrome-plated nail clippers is hard because reflections, rack occlusion, process drift, and rare defects weaken supervised models. We link process signals, rack identifiers, and paired images with a few-shot vision-language cache. The framework combines object detection, anomaly detection, feature adaptation, and a trainable multimodal cache. A feature-adapter baseline yielded 1/7 accuracy; a separate 22-image evaluation yielded 13/22 anomaly accuracy (59.09%; exact 95% confidence interval, 36.4-79.3%). Four backbone intervals overlapped on 17 defects, so no ranking was conclusive. We present a traceable framework and an independent-validation protocol. These preliminary findings require larger rack-disjoint validation.

Keywords

vision-language model, chrome plating, nail clipper, few-shot learning, anomaly detection

* 삼성데이터서비스(주) 차장, 대한민국산업현장교수
- ORCID: <https://orcid.org/0009-0008-7075-1455>

• Received: Jun. 25, 2026, Revised: Jul. 22, 2026, Accepted: Jul. 25, 2026
• Corresponding Author: Gyeongmin Kim
Samjung Data Service Co., Ltd.
Tel.: +82-2-2029-0156, Email: formin97@gmail.com

I. 서 론

손톱깍이와 같은 소형 금속 소비재의 크롬 도금 품질은 표면 광택, 색상 균일성, 스크래치, 핀홀(Pin-hole), 버(Burr), 전류 조건 변화에 따른 변색과 직접 연결된다. 대상 생산 라인에서는 공정 단위 불량 발생률이 약 30%에 이르고 재작업이 불가능한 완전 폐기율도 약 3% 수준으로 보고되었으며, 숙련 검사자의 육안 검사에 의존하는 현행 방식은 검사 편차와 숙련 노하우 단절의 위험을 안고 있다.

공정 불량률이 높다는 사실이 학습에 사용할 수 있는 불량 이미지가 충분하다는 것을 의미하지는 않는다. 첫째, 재작업 대상 불량은 대부분 라인에서 즉시 재처리되므로 촬영과 라벨링을 거쳐 데이터셋으로 축적되는 표본은 소수에 그친다. 둘째, 불량은 저전류 변색, 니켈 도금 불량, 찌힘, 고전류 버닝, 복합 결함 등 세부 유형으로 나뉘며 유형별 발생 빈도가 극도로 불균형하여 일부 유형은 관측 자체가 드물다. 셋째, 공정 조건 개선에 따라 불량 분포가 계속 변하므로 신규 유형은 항상 소표본 상태에서 검사 체계에 편입되어야 한다. 따라서 본 연구는 유형별 라벨 이미지가 수 장 수준인 소량학습(Few-shot) 조건을 현실적 제약으로 설정한다.

이러한 조건에서 본 연구가 다루는 질문은 두 가지이다. 첫째, 사전학습된 CLIP(Contrastive Language-Image Pre-training) 계열 비전-언어 모델을 동결한 채 소량의 불량 이미지와 텍스트 설명만으로 어느 수준의 이상탐지와 유형 분류가 가능한가. 둘째, 이러한 모델을 공정 조건 랙(Rack) 이력과 연계하여 운영하려면 어떤 시스템 구조와 검증 절차가 필요한가. 이를 위해 YOLO(You Only Look Once) 기반 개체 분리, PatchCore 기반 이상 점수, 갱신형 이미지-텍스트 캐시 어댑터(T-IDEA, Trainable Image Description Enhanced CLIP-Adapter)를 랙 단위 검사 레코드로 통합하는 프레임워크를 설계하고, 예비 실험으로 적용 가능성과 검증 한계를 평가한다.

본 논문의 기여는 세 가지이다. 첫째, 공정 데이터와 검사 영상을 랙 단위로 연결하는 검사 프레임워크를 제시한다. 둘째, 캐시 어댑터의 구성 요소와 판정 절차를 수식으로 정의하고 CLIP-Adapter 및

IDEA(Image Description Enhanced CLIP-Adapter)와의 차이를 명확히 한다. 셋째, 내부 검증값과 독립 평가를 구분하고 신뢰구간과 실패 사례를 함께 보고하여 산업 적용 전 검증 절차를 명시한다. 이하 II장에서 관련 연구를 정리하고, III장에서 대상 공정과 데이터 구성을, IV장에서 제안 프레임워크를, V장에서 실험과 결과를 제시하며, VI장의 논의와 VII장의 결론으로 마무리한다.

II. 관련 연구

산업 표면검사는 결함 표본이 부족하고 정상 표본이 상대적으로 풍부한 조건을 자주 다룬다. 지도 학습 기반 CNN(Convolutional Neural Network) 분류는 유형별로 수백 장 이상의 라벨 이미지를 요구하므로 신규 라인이나 희소 결함에는 적용이 어렵다. PatchCore[1]는 정상 패치 특징을 대표 메모리로 저장하고 최근접 검색으로 이상 점수를 산출하여 정상 표본만으로 콜드스타트 검사를 구성할 수 있으나, 고광택 금속 표면처럼 조명과 자세 변화가 큰 경우에는 정상 메모리의 촬영 조건 안정성이 성능에 직접적인 영향을 준다.

CLIP[2]은 대규모 이미지-텍스트 쌍의 대조학습으로 이미지와 텍스트를 같은 임베딩 공간에 정렬하여, 불량 유형을 단순 라벨이 아니라 자연어 설명으로 정의할 수 있게 한다. SigLIP(Sigmoid Loss for Language-Image Pre-training)[3]은 소프트맥스 대조학습을 쌍 단위 시그모이드 손실로 대체하여 배치 의존성을 줄였고, SigLIP2[4]는 다국어 사전학습과 밀집 특징 개선을 더해 표현 품질을 높였다. 언어 인코더를 교체한 AltCLIP(Altering the Language Encoder in CLIP)[5], 대규모 학습 기법을 개선한 EVA-CLIP[6] 등의 백본 변형도 제안되었다. 본 연구는 이들 백본을 동결한 채 비교 사용한다.

소량 적용 방법으로, CLIP-Adapter[7]는 동결 백본 특징에 잔차 연결된 경량 어댑터를 학습하여 소량 표본에 적응하며 본 연구의 기준선으로 사용하였다. 본 논문의 T-IDEA 캐시 어댑터는 IDEA[8]의 캐시 구조를 확장한 것으로, 이미지-텍스트 쌍 캐시를 이용해 백본 재학습 없이 경량 적응만으로 소량 분류를 수행한다. 텍스트 설명의 생성과 정제에는

BART(Bidirectional and Auto-Regressive Transformers) [9]와 같은 생성 모델을 활용할 수 있다. 캐시 기반 적응은 클래스 추가와 갱신이 메모리 갱신으로 처리되므로, 불량을 대량으로 수집하기 어려운 도금 공정에서 모델 갱신 비용을 낮출 수 있다.

비전-언어 기반 산업 이상탐지 연구로는 WinCLIP(Window-based CLIP)[10]과 AnomalyCLIP[11]이 대표적이다. 전자는 프롬프트 양상블과 윈도우 단위 집계로 제로샷·소수샷 이상 분류와 분할을 수행하고, 후자는 객체 비의존 프롬프트 학습으로 제로샷 이상탐지를 일반화한다. 다만 두 방법은 공개 벤치마크 중심으로 설계되어 도금 공정 조건이나 랙 이력 같은 현장 메타데이터와의 결합과 운영 중 캐시 갱신은 다루지 않는다. 본 연구는 갱신 가능한 이미지-텍스트 캐시를 공정 조건·랙 추적 레코드와 연결하는 배치 구조를 제시한다는 점에서 차별화된다.

III. 대상 공정 및 데이터 구성

대상 라인의 핵심 공정은 전처리, 전해 산세, 초음파 탈지, 전해탈지, 니켈 스트라이크, 광택 니켈, 크롬 도금의 7단계이다. 각 단계에는 온도와 수소이온지수(pH) 센서, 정류기 전류·전압 계측이 결합되고, 랙 상단의 QR(Quick Response) 코드 또는 바코드로 랙 단위 이력이 추적된다. 제품은 핵심 공정 전후 스테이션에서 카메라로 촬영되며, 판정 결과는 품질 점수와 공정 조건이 결합된 데이터베이스로 저장된다.

검사 스테이션은 공정 위치별 관찰 목적에 따라 전해탈지 후, 광택 니켈 후, 크롬 도금 후의 세 지점으로 구성된다. 전해탈지 후 스테이션은 스크래치와 오염 잔류를 조기에 확인하여 후속 공정으로의 불량 전파를 줄이고, 광택 니켈 후 스테이션은 헤어라인과 얼룩처럼 광택막 형성 이후 두드러지는 결함을 확인하며, 크롬 도금 후 최종 스테이션은 편흔, 고전류 버닝, 국부 박리 등 최종 외관 판정 항목을 담당한다. 각 스테이션에서 랙 식별자(ID, Identifier), 촬영 시점, 카메라 ID, 조명 조건, 공정 센서 값이 함께 기록된다.

실험 데이터는 대상 생산 라인에서 수집된 랙 적

재 영상과 개별 제품 영상으로 구성되며, 실험별 데이터 구성은 표 1과 같다. 개체 분리 실험은 학습 16장과 검증 6장(개체 334개)의 랙 영상을, 어댑터 기준선 실험은 스크래치·편흔·정상 3클래스의 학습 9장과 검증 7장을 사용하였다. T-IDEA 파일럿은 저전류 변색 5샷, 니켈 도금 불량 3샷, 찍힘 3샷의 이미지-텍스트 쌍 11개를 지원 집합으로 하고 13장(참불량 11장)으로 평가하였으며, 백본 비교는 불량 17장을 포함한 22장을 대상으로 하였다. 정상 캐시 점검에는 정상 랙 영상 2건(개체 36개)이 사용되었다. 이처럼 유형별 가용 표본이 수 장 수준인 구성은 I 장에서 정의한 소량학습 제약을 그대로 반영한다.

문제는 두 층위로 정의된다. 첫째, 랙 전체 영상에서 개별 손톱깍이 영역(ROI, Region of Interest)을 안정적으로 분리하고 고광택 반사와 가림을 줄여야 한다. 둘째, 정상과 불량을 판별한 뒤 저전류 변색, 니켈 도금 불량, 찍힘, 고전류 버닝, 복합 결함과 같은 세부 유형을 적은 라벨로 구분해야 한다. 따라서 개체 탐지, 이상탐지, 비전-언어 기반 불량 유형 분류를 분리하되 랙 ID와 공정 조건으로 다시 통합하는 구조가 필요하다.

표 1. 실험별 데이터 구성

Table 1. Data composition per experiment

Exp.	Data	Eval.	Metric
YOLO	16 img	6 img	mAP
CLIP-A	9 img	7 img	Acc.
T-IDEA	11 sets	13 img	Acc.*
Backb.	cache	22 img	Rec./CI
PCore	2 racks	36 obj	Score

Note. Table 1: Exp.=experiment; img=images; CLIP-A=CLIP-Adapter; Acc.=accuracy; Rec.=recall; Data/Eval.=training-or-support/evaluation size; sets=image-text support pairs; *=internal validation reused during training; Backb.=backbone comparison; cache=shared support cache; PCore=PatchCore; obj=objects; mAP=mean average precision (precision and recall are reported in the text); CI=confidence interval.

공정-검사 통합 데이터 구조는 표 2와 같다. 전류·전압·온도·수소이온지수는 공정 상태를 나타내며, 랙 식별자·공정 단계·시각은 검사 이력을 연결한다. 카메라·조명·편광 조건과 검사 점수·신뢰도를 함께 기록하여 반사·가림의 영향과 불량 판정 근거를 추적한다.

표 2. 공정-검사 통합 데이터 구조

Table 2. Process-Inspection data

Part	Items	Role
Process	I, V, T, pH	Variables
Rack	QR, ID, stage, time	Link
Vision	Optics	Glare/occ.
Output	Score/conf.	Basis

Note. Table 2: I/V/T=current/voltage/temperature;
ID=identifier; QR=Quick Response; conf.=confidence;
Optics=camera/light/polarizer; Glare/occ.=glare/occlusion.

IV. 제안 프레임워크

그림 1과 같이 제안 구조는 데이터 계층, 광학 계층, 모델 계층, 운영 계층으로 구성된다. 데이터 계층은 랙 ID, 공정 단계, 시간, 전류, 전압, 온도, pH, 카메라 ID를 하나의 검사 레코드로 묶고, 광학 계층은 상부 카메라와 좌우 측면 카메라, 돔 조명, 암시야 바 조명, 편광 필터, PLC(Programmable Logic Controller) 트리거를 결합해 반사와 가림을 완화한다. 모델 계층은 YOLOv8[12]로 개별 제품 ROI를 분리하고, PatchCore 정상 메모리로 이상 점수를 산출하며, T-IDEA 캐시 어댑터로 불량 유형을 분류한다. 각 판정 결과는 검사 레코드에 결합되어 운영 계층의 권고와 제어 판단에 사용된다.

비전-언어 백본을 채택하는 이유는 두 가지이다. 첫째, 불량 유형은 “저전류로 인한 변색”과 같이 언어로 정의되는 현장 지식이므로, 텍스트 설명을 클래스 정의의 앵커로 사용하면 소수 이미지만으로도 유형 경계를 임베딩 공간에 지정할 수 있다. 둘째, 신규 불량 유형이 발생했을 때 텍스트 설명과 소수 이미지를 캐시에 추가하는 것만으로 분류 체계를 확장할 수 있어 백본 재학습 없이 검사 체계를 갱신할 수 있다. 텍스트 키는 이미지 키와 상보적으로 작동하여, 촬영 조건 변화로 이미지 유사도가 흔들릴 때 유형 정의를 안정화하는 역할을 한다.

T-IDEA 캐시 어댑터는 동결된 이미지 인코더와 텍스트 인코더 위에 구성된다. 질의 이미지의 정규화 임베딩을 q 라 하고, 유형별 소량 지원 집합에서 이미지 키 행렬 K_I , 텍스트 키 행렬 K_T , 원-핫 라벨 행렬 Y 를 구성한다(파일럿 기준 특징 차원 768, 지원 11쌍, 4클래스). 질의와 캐시의 멀티모달 유사도 $S(q)$ 는 식 (1)과 같이 이미지 유사도와 텍스트 유사도의 가중합으로 정의하며, α 는 두 모달리티의 혼합 비율이다.

$$S(q) = \alpha q K_I^T + (1 - \alpha) q K_T^T \quad (1)$$

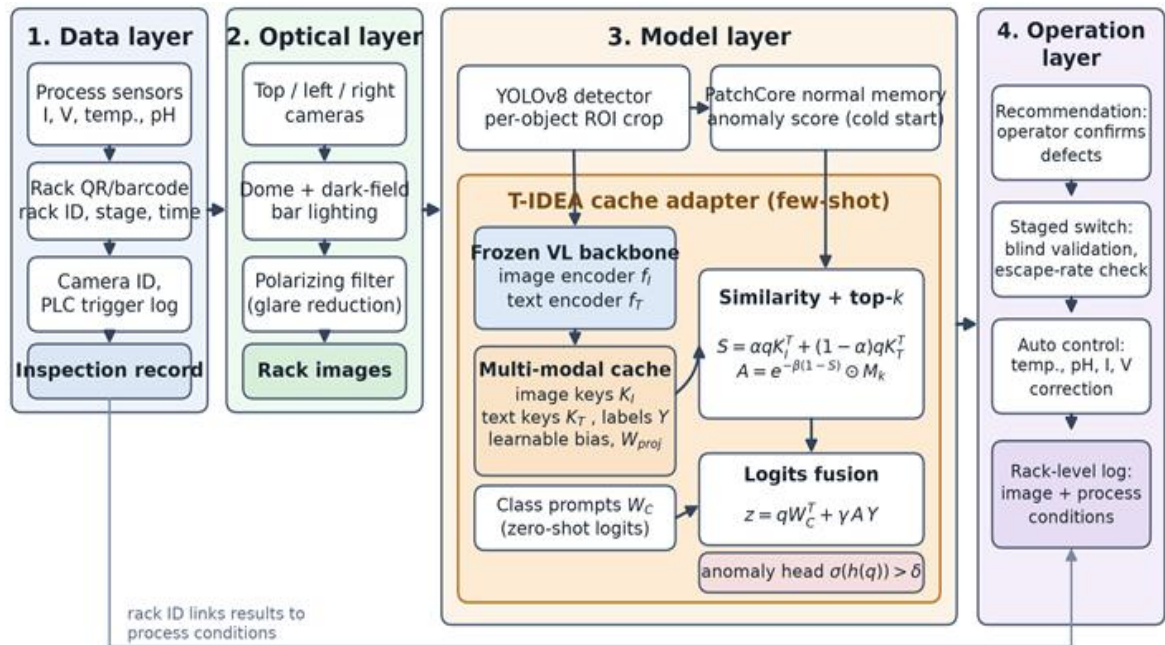


그림 1. 제안하는 비전-언어 캐시 어댑터 프레임워크 구조

Fig. 1. Architecture of the proposed vision-language cache adapter framework

유사도는 식 (2)와 같이 지수 활성화로 캐시 친화도 $A(q)$ 로 변환되며, 상위 k 개 항목만 남기는 마스킹 M_k 로 소표본 캐시의 잡음을 억제한다. β 는 친화도의 예리함을 조절하는 스케일 파라미터이다.

$$A(q) = M_k(\exp(-\beta(1 - S(q)))) \quad (2)$$

최종 유형 로짓 $z(q)$ 는 식 (3)과 같이 캐시 친화도를 라벨 행렬로 사상한 캐시 로짓과 클래스 프롬프트 임베딩 W_c 의 제로샷 로짓을 가중치 γ 로 융합하여 얻는다.

$$z(q) = qW_c^T + \gamma A(q)Y \quad (3)$$

지원 키에는 학습형 바이어스 항이 더해지고 선행 투영이 질의와 키 임베딩을 정렬하며, 이들 파라미터는 지원 집합에 대한 교차 엔트로피 손실로 소량 적응된다. 이상 판정은 유형 분류와 분리된 학습형 이상 헤드 h 로 수행하여, 이상 확률 $\sigma(h(q))$ 가 임계값 δ 를 넘으면 불량으로 판정하고 식 (3)의 최댓값 클래스로 유형을 결정한다.

제안 구조가 CLIP-Adapter[7] 및 IDEA[8]와 다른 점은 세 가지이다. 첫째, CLIP-Adapter는 클래스 집합이 바뀌면 어댑터 가중치를 재학습해야 하지만, 제안 구조는 지원 캐시 갱신만으로 유형을 추가한다. 둘째, 단일 분류 문제를 다루는 IDEA와 달리 이상탐지(이진 판정)와 유형 분류를 분리한 2단 판정을 사용하여, 미지 유형이 특정 클래스로 강제 분류되는 위험을 줄인다. 셋째, 학습형 바이어스, 선행 투영, 상위 k 개 마스킹을 결합해 수 샷 수준 캐시의 과적합과 잡음을 완화하고, 판정 결과를 랭크 단위 공정 레코드와 결합하는 운영 절차까지 포함한다.

운영 계층은 권고 모드와 자동 제어 모드로 구분한다. 권고 모드에서는 모델이 불량 유형과 관련 공정 조건 후보를 제시하고 작업자가 확정하며, 충분한 표본과 재현성 검증이 축적된 뒤에 온도, pH, 전류, 전압 보정을 포함하는 자동 제어 모드로 확장한다.

V. 예비 실험 및 결과

실험은 NVIDIA Tesla T4 GPU(Graphics Processing

Unit) 환경에서 PyTorch 기반으로 수행되었다. 개체 분리는 YOLOv8n(파라미터 약 3.0M, 입력 640×640 , 시드 0)을 최대 1,000 epoch에 조기 종료(patience 100) 조건으로 학습하였고, 어댑터 기준선은 CLIP 백본에서 5 epoch, T-IDEA는 CLIP ViT-L/14(Vision Transformer-Large, 패치 크기 14) 백본(특징 차원 768, 시드 42)에서 60 epoch 적응 학습하였다. 이상 판정 임계값 δ 는 0.55를 사용하였다. 본 장의 모든 수치는 실험 당시 저장된 실행 기록과 결과 자료에서 산출한 값이다.

CLIP-Adapter 기준선은 스크래치, 핀홀, 정상 3 클래스(학습 9장, 검증 7장)에서 5 epoch 동안 검증 정확도 14.29%(1/7)에 머물렀다. 학습 손실은 1.0396에서 1.0389로 사실상 정체되었고, 단일 이미지 추론의 최고 신뢰도도 0.3382로 3클래스 균등분포 수준이었다. 즉 이 기준선은 수 장 수준의 표본으로는 잔차 어댑터가 유의미한 결정 경계를 학습하지 못함을 보여준다.

T-IDEA 파일럿은 저전류 변색, 니켈 도금 불량, 찍힘의 지원 11쌍으로 60 epoch 적응한 뒤 13장 평가에서 이상탐지 13/13, 참 불량 11장의 유형 분류 11/11을 기록하였다. 그러나 이 수치를 기준선 14.29%와 직접 비교하여 텍스트 설명 추가만으로 정확도가 100%에 도달했다고 해석해서는 안 된다. 두 실험은 클래스 집합(스크래치·핀홀·정상 대 저전류 변색·니켈 불량·찍힘)과 적응 방식(잔차 어댑터 5 epoch 대 캐시 적응 60 epoch)이 서로 다르고, 무엇보다 13장 평가 표본이 60 epoch 학습 중 검증에 반복 사용되어 독립 테스트가 아닌 내부 검증값이기 때문이다. 정확 이항 신뢰구간으로도 13/13과 11/11의 하한은 각각 75.3%와 71.5%에 그친다. 실제로 동일 가중치에서 학습 도메인의 정상 표본은 이상 확률 0.3751로 정상 판정되었지만, 촬영 조건이 다른 신규 정상 표본은 이상 확률 0.9047로 오탐되어 촬영 조건 변화에 대한 일반화 한계가 확인되었다.

동일한 지원 캐시로 네 백본을 비교한 실험에서, 재구성 가능한 불량 17장의 유형 분류 정확도(마이크로 평균 재현율)는 표 4와 같이 AltCLIP 2/17(11.8%), CLIP 5/17(29.4%), SigLIP2 6/17(35.3%), EVA-CLIP[6] 8/17(47.1%)이었으나, Clopper-Pearson 95% 신뢰구간이 서로 크게 겹쳐 현재 표본 규모로

는 백본 간 우열을 통계적으로 확정할 수 없다. 클래스 불균형을 보정한 클래스 평균(매크로) 재현율은 AltCLIP 6.7%, CLIP 18.9%, EVA-CLIP 26.7%, SigLIP2 27.2%로 마이크로 평균과 순위가 일부 달라지며, 고전류 버닝(1장)과 전처리 복합(1장) 클래스는 모든 백본에서 0%를 기록하였다. 이는 클래스 불균형 조건에서 단일 정확도 수치보다 클래스별 재현율과 표본 보강이 중요함을 보여준다.

개체 분리와 정상 메모리 계층도 실행 기록으로 정량화하였다. YOLOv8n은 검증 6장(개체 334개)에서 정밀도 0.955, 재현율 0.865, mAP50 0.928, mAP50-95 0.671, 이미지당 추론 4.6 ms를 기록하여 락 영상에서의 ROI 분리가 실용 수준임을 확인하였다. PatchCore 정상 메모리는 정상 락 영상 2건의 개체 36개에 이상 점수 11.54~14.94를 부여하고 모두 정상으로 판정하였으나, 정상 영상 2건만으로는 점수 대역의 재현성이나 결함 검출력을 입증할 수 없다.

불량 17장과 정상 5장을 합한 22장 통합 평가의 이상탐지 정확도는 59.09%(13/22, 95% 신뢰구간 36.4~79.3%)로, 제조 현장이 요구하는 수준에는 미치지 못한다. 보존된 실행 기록에는 22장 평가의 이미지별 예측과 점수가 남아 있지 않아 정밀도, 재현율, F1-점수, AUROC(Area Under the Receiver Operating Characteristic Curve)를 산출할 수 없었으며, 이는 예비 실험 기록 관리의 한계이다. 이에 따라 VII장의 향후 검증 절차에 혼동행렬 기반 지표와 임계값 독립 지표(AUROC)를 포함한 지표 세트를 명시하였다. 또한 13장 내부 검증과 22장 평가의 표본 중복 여부와 락 단위 분리 근거가 기록에 없어 두 수치를 직접 비교하지 않고 탐색적 결과로만 해석한다.

표 3. 예비 실험 결과 요약과 해석

Table 3. Summary of preliminary results

Exp.	Scale	Result	Note
CLIP-A	9/7	14.29%	Weak
T-IDEA	13/11	13/13; 11/11	Reuse
Backb.	17def	2-8/17	Overlap
Rare	1-2	0% all	Balance

Note. Table 3: CLIP-A=CLIP-Adapter; 9/7=train/val images; 13/11=13 tests/11 defects; Backb.=backbone comparison; 2-8/17=range across four backbones; Reuse=test reused in training; Overlap=95% CIs overlap; 0% all=zero recall for all backbones.

표 4. 백본별 불량 분류 정확도와 95% 신뢰구간

Table 4. Backbone accuracy with 95% CI

Backbone	k/n	Acc.	95% CI
AltCLIP	2/17	11.8%	1.5-36.4
CLIP	5/17	29.4%	10.3-56.0
SigLIP2	6/17	35.3%	14.2-61.7
EVA-CLIP	8/17	47.1%	23.0-72.2

Note. Table 4: k/n=correct/total on the defective subset (17 images) with the T-IDEA cache adapter; CI=exact Clopper-Pearson 95% interval (%).

VI. 논 의

도금 공정은 약품 농도, 온도, 표면 상태가 시간에 따라 변하므로 한 번 학습된 뒤 고정되는 모델은 장기 안정 운전에는 불리하다. 캐시 기반 어댑터는 신규 결함 표본과 텍스트 설명을 추가해 임베딩 메모리를 갱신할 수 있으나, 갱신은 명확한 절차 아래에서만 수행되어야 한다. 신규 표본을 지원 집합에 추가할 때에는 별도의 검증 표본으로 분리 평가를 수행하고 임계값과 가중치를 재조정된 뒤 권고 정확도와 불량 유출률을 함께 점검해야, 잘못 라벨링된 표본이 누적되는 것을 막을 수 있다.

고광택 표면 검사의 신뢰성은 촬영 조건의 재현성에 크게 의존한다. V장의 신규 정상 표본 오탐 사례가 보여주듯 조명과 자세 변화는 정상품 오탐의 직접 원인이 되므로, 돔 조명·암시야 바 조명·편광 필터 조합의 고정과 카메라 노출·화이트 밸런스의 주기적 교정이 선행되어야 한다. 모델 출력은 우선 공정 엔지니어가 확인하는 권고 모드로 사용하고, 자동 제어 전환은 충분한 기간의 블라인드 검증과 불량 유출률 분석이 누적된 후에 단계적으로 결정하는 것이 타당하다.

VII. 결 론

본 논문은 크롬 도금 손톱깍이 표면검사를 대상으로 공정 조건 데이터와 비전-언어 기반 소량학습 검사를 결합하는 프레임워크를 제시하였다. 7단계 공정 센싱과 락 이력 추적, 다중 카메라와 조명 구성, YOLO/PatchCore/CLIP-Adapter 기준선, SigLIP 계열 T-IDEA 캐시 어댑터를 하나의 통합 검사 구조

로 정리하고, 캐시 어댑터의 구성 요소와 판정 절차를 수식으로 정의하였다.

본 연구의 한계는 다음과 같다. 첫째, 유형별 표본이 수 장 수준이어서 신뢰구간이 넓고 백본 간 우열을 확정할 수 없다. 둘째, T-IDEA 파일럿 수치는 학습 중 검증에 재사용된 표본에 대한 내부 검증값이며 독립 테스트 성능이 아니다. 셋째, 이상탐지 정확도 59.09%는 양산 요구 수준에 미달하며, 이미지별 기록 부재로 정밀도·재현율·F1-점수·AUROC를 보고하지 못하였다. 넷째, 촬영 조건 변화에 따른 정상품 오탐이 확인되었다.

향후 연구는 첫째, 이미지·랙 단위로 분리한 독립 평가셋과 결합 유형별 균형 표본을 구축하고, 둘째, 혼동행렬 기반 정밀도·재현율·F1-점수와 AUROC를 포함한 지표 세트를 상시 산출하며, 셋째, 정상 시점별 베이스라인을 수백 장 규모로 확보하고 현장 조명 반복성 실험으로 측정 재현성을 정량화한다. 넷째, 복합 결함을 위한 다중 라벨 분류와 적응형 임계값을 도입하고 텍스트 키, 상위 k개 마스킹, 로짓 융합 등 구성 요소별 기여를 분리 검증하는 절제(Ablation) 실험을 수행하며, 다섯째, 공정 조건-결함 원인 상관 분석과 장기 권고 모드 운영 로그 분석을 수행한다. 이러한 검증이 축적되면 제안 구조는 희소 불량 환경에서 갱신 가능한 검사 기준 모델로 활용될 수 있을 것이다.

Acknowledgment

본 논문은 Zenodo에 공개된 사전 공개 초안[13]의 확장 및 개정본이다.

References

- [1] K. Roth, et al., "Towards Total Recall in Industrial Anomaly Detection", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., New Orleans, LA, USA, pp. 14318-14328, Jun. 2022. <https://doi.org/10.1109/CVPR52688.2022.01392>.
- [2] A. Radford, et al., "Learning Transferable Visual Models From Natural Language Supervision", Proc. 38th Int. Conf. Machine Learning, PMLR, Online, Vol. 139, pp. 8748-8763, Jul. 2021. <https://doi.org/10.48550/arXiv.2103.00020>.
- [3] X. Zhai, B. Mustafa, A. Kolesnikov, and L. Beyer, "Sigmoid Loss for Language Image Pre-Training", Proc. IEEE/CVF Int. Conf. Comput. Vis., Paris, France, pp. 11975-11986, Oct. 2023. <https://doi.org/10.1109/ICCV51070.2023.01100>.
- [4] M. Tschannen, et al., "SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features", arXiv preprint, arXiv:2502.14786, Feb. 2025. <https://doi.org/10.48550/arXiv.2502.14786>.
- [5] Z. Chen, G. Liu, B.-W. Zhang, Q. Yang, and L. Wu, "AltCLIP: Altering the Language Encoder in CLIP for Extended Language Capabilities", Findings ACL, Toronto, Canada, pp. 8666-8682, Jul. 2023. <https://doi.org/10.18653/v1/2023.findings-acl.552>.
- [6] Q. Sun, Y. Fang, L. Wu, X. Wang, and Y. Cao, "EVA-CLIP: Improved Training Techniques for CLIP at Scale", arXiv preprint, arXiv:2303.15389, Mar. 2023. <https://doi.org/10.48550/arXiv.2303.15389>.
- [7] P. Gao, et al., "CLIP-Adapter: Better Vision-Language Models with Feature Adapters", Int. J. Comput. Vis., Vol. 132, No. 2, pp. 581-595, Feb. 2024. <https://doi.org/10.1007/s11263-023-01891-x>.
- [8] Z. Ye, F. Jiang, Q. Wang, K. Huang, and J. Huang, "IDEA: Image Description Enhanced CLIP-Adapter", arXiv preprint, arXiv:2501.08816, Jan. 2025. <https://doi.org/10.48550/arXiv.2501.08816>.
- [9] M. Lewis, et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension", Proc. 58th ACL, Online, pp. 7871-7880, Jul. 2020. <https://doi.org/10.18653/v1/2020.acl-main.703>.
- [10] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, and O. Dabeer, "WinCLIP: Zero-/Few-Shot Anomaly Classification and

- Segmentation", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Vancouver, BC, Canada, pp. 19606-19616, Jun. 2023. <https://doi.org/10.1109/CVPR52729.2023.01878>.
- [11] Q. Zhou, G. Pang, Y. Tian, S. He, and J. Chen, "AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection", Proc. 12th Int. Conf. Learn. Represent. (ICLR), Vienna, Austria, May 2024. <https://doi.org/10.48550/arXiv.2310.18961>.
- [12] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8", software, version 8.0.0, Jan. 2023. <https://github.com/ultralytics/ultralytics>. [accessed: Jun. 25, 2026].
- [13] G. Kim, "Few-Shot Vision-Language Inspection for Chrome-Plated Nail Clippers Using CLIP-Adapter and SigLIP2/T-IDEA Cache Adaptation", Zenodo, preprint, Jun. 2026. <https://doi.org/10.5281/zenodo.20770610>.

저자소개

김 경 민 (Gyeongmin Kim)



2017년 9월 ~ 2019년 8월 :
한국공학대학교
소프트웨어융합공학과(공학석사)
2025년 1월 ~ 현재 :
대한민국산업현장교수
2007년 9월 ~ 현재 :
삼정데이터서비스(주) 차장

관심분야 : 비전-언어 모델, 표면검사 자동화, 소량학습,
산업 이상탐지, 명세 기반 개발(SDD,
Specification-Driven Development), 제조 공정 데이터
모델링