Abstract—인더스트리 5.0(Industry 5.0)의 핵심 가치인 지속 가능한 제조(Sustainable Manufacturing)를 실현하기 위해서는 신뢰성 있는 품질 검사 시스템이 필수적이다. 그러나 디젤 미립자 필터(DPF)와 같은 특수 공정은 데이터의 극단적 희소성(Extreme Scarcity)과 클래스 불균형(Class Imbalance) 문제로 인해 기존 딥러닝 모델 적용에 난항을 겪고 있다. 본 논문에서는 단 16장의 결함 이미지만으로 고성능 탐지 모델을 구축하기 위한 GLAD (Generative-Latent-Adaptive-Detection) 프레임워크를 제안한다. 제안하는 방법은 1) Latent Diffusion Model (LDM)과 LoRA를 활용하여 고품질의 가상 결함 데이터를 생성하고, 2) 8,900장의 유사 도메인(X-ray) 데이터셋을 활용한 계층적 도메인 브리지(Hierarchical Domain Bridge) 전이학습을 수행하며, 3) 불확실성(Uncertainty)을 고려한 탐지 네트워크로 검출 신뢰도를 극대화한다. 실험 결과, 본 프레임워크는 단 16장의 'Few-Shot' 환경에서 91.7%의 mAP50를 달성하여 기존 방법론 대비 획기적인 성능 향상을 입증했다.

Index Terms—퓨샷 러닝(Few-Shot Learning), 생성형 AI(Generative AI), 확산 모델(Diffusion Model), 도메인 적응, 불확실성 추정, 스마트 팩토리

Introduction

A. 인더스트리 5.0과 지속 가능한 제조 (Sustainable Manufacturing)

4차 산업혁명(Industry 4.0)이 자동화(Automation)와 데이터 교환(Data Exchange)을 통한 생산 효율성 극대화에 초점을 맞췄다면, 2024년 현재 도래하고 있는 인더스트리 5.0(Industry 5.0) 패러다임은 인간 중심(Human-centric), 회복 탄력성(Resilient), 그리고 지속 가능성(Sustainability)을 핵심 가치로 삼는다[1]. 전 세계적인 탄소 중립(Carbon Neutrality) 기조와 강화된 배출가스 규제(Euro 7)는 자동차 부품 제조사들에게 전례 없는 품질 무결성을 요구하고 있다. 특히 디젤 미립자 필터(DPF, Diesel Particulate Filter)는 내연기관 차량의 미세먼지(PM) 배출을 99% 이상 저감하는 핵심 부품으로, 2030년까지 글로벌 시장 규모가 250억 달러(약 33조 원)에 이를 것으로 전망된다. DPF의 제조 공정 중 발생하는 세라믹 기재의 미세 균열(Crack)이나 소성 과정의 용융(Melting)은 필터링 성능 저하뿐만 아니라 차량 화재로 이어질 수 있어, 전수 검사(Total Inspection)가 단순한 품질 관리를 넘어 기업의 ESG(환경, 사회, 지배구조) 경영의 척도로 격상되었다.

B. 산업용 데이터의 본질적 난제:

딥러닝 기반의 컴퓨터 비전 기술이 반도체, 디스플레이 등 첨단 제조 분야를 혁신하고 있지만, DPF 제조와 같은 특수 부품 공정에 이를 즉시 적용하기에는 단계적 데이터 희소성(Hierarchical Data Scarcity)이라는 거대한 장벽이 존재한다.

극단적 불균형(Extreme Imbalance): 성숙된 제조 라인의 불량률은 통상 100ppm(0.01%) 미만이다. 이는 딥러닝 모델 학습에 필요한 수천 장의 결함 데이터를 모으기 위해 수년의 시간이 필요함을 의미하며, 이 기간 동안 AI 도입 프로젝트가 정체되는 '데이터 희소성 병목(Data Scarcity Bottleneck)' 현상이 발생한다.

X-ray 영상의 복잡성: 일반적인 카메라 이미지와 달리, 3차원 물체를 2차원으로 투영한 X-ray 이미지는 물체의 앞면과 뒷면 정보가 중첩(Superposition)되는 특성을 가진다[29]. 또한, 빔 경화(Beam Hardening) 현상으로 인해 이미지의 명암이 비선형적으로 왜곡되며, 이는 합성 곱 신경망(CNN)이 특징을 추출하는 데 혼란을 가중시킨다.

소형 객체 난이도 (Fig. 1): 본 연구에서 수집한 데이터셋을 정밀 분석한 결과, 전체 결함의 98.2%가 이미지 전체 면적의 1% 미만을 차지하는 극소형 객체임이 밝혀졌다.

C. 제안 방법론: GLAD Framework

기존 연구들은 이미지 회전(Rotation), 반전(Flip) 등 기하학적 증강에 의존해왔으나, 이는 데이터의 다양성을 근본적으로 확장하지 못한다. 본 연구는 이를 극복하기 위해 물리적 증강의 한계를 넘어서는 생성형 AI 기반의 GLAD (Generative-Latent-Adaptive-Detection) 프레임워크를 제안한다.

Generative (생성): 최신 Latent Diffusion Model을 도입하여, 실제로는 존재하지 않았던 다양한 형태의 결함을 사실적으로 합성한다. 이는 단순한 픽셀 보간이 아닌, 데이터의 분포(Distribution) 자체를 확장하는 시도이다.

Latent (잠재 공간): 16장의 이미지에서 X-ray의 잠재적 텍스처 특징(Latent Feature)을 추출하여, 이질감 없는 가상 데이터를 생성한다.

Adaptive (적응): 대규모 X-ray 데이터셋(약 8,900장)을 중간 도메인으로 활용하여, 모델이 산업용 이미지에 단계적으로 적응하도록 유도한다.

관련 연구 (Related Work)

A. 제조 결함 탐지를 위한 딥러닝

산업용 결함 탐지 알고리즘은 크게 지도 학습(Supervised Learning)과 비지도 학습(Unsupervised Learning)으로 나뉜다[2]. 비지도 학습 계열인 오토인코더(Autoencoder)나 GAN 기반의 재구성(Reconstruction) 기법은 정상 데이터만으로 학습 가능하다는 장점이 있으나, DPF와 같이 배경 텍스처가 복잡한 경우 잦은 오탐(False Positive)을 유발한다. 반면, YOLO[3, 22-25], Faster R-CNN[10] 등의 지도 학습 모델은 높은 정확도를 보장하지만, 대규모 라벨링 데이터가 필수적이다. 최근 M2U-InspectNet[4]과 같이 Multi-scale Vision Transformer[27]를 활용한 연구가 주목받고 있으나, 이러한 거대 모델은 높은 연산량을 요구하여 엣지 디바이스(Edge Device) 탑재가 어렵다는 한계가 있다. 따라서 현장에서는 여전히 YOLO 시리즈와 같은 경량화 모델(Lightweight Model)에 대한 수요가 높다.

B. 생성형 AI 기반 데이터 증강 (Generative Data Augmentation)

데이터 부족을 해결하기 위해 GAN(Generative Adversarial Network)으로 가상 결함을 생성하는 연구가 활발했다. 그러나 GAN은 학습이 불안정하고, 생성된 이미지의 다양성이 제한되는 모드 붕괴(Mode Collapse) 문제에 취약하다. 2024년 이후, 확산 모델(Diffusion Models)이 텍스처 생성의 새로운 표준으로 자리 잡았다[14]. Chen et al.은 Diffusion-based Data Augmentation이 금속 표면 결함 탐지 정확도를 획기적으로 개선함을 보였다. 본 연구는 이를 확장하여, 단순한 표면 결함이 아닌 투과된 X-ray 내부 결함 생성에 특화된 Latent Diffusion + LoRA 파이프라인을 제안한다. 이는 CutPaste나 Mosaic와 같은 물리적 합성 기법보다 훨씬 자연스러운 결함 데이터를 생성할 수 있다.

C. 퓨샷 객체 탐지 (Few-Shot Object Detection)

극소량의 데이터로 모델을 학습시키는 FSOD(Few-Shot Object Detection)는 메타 러닝(Meta-Learning)[17]이나 거리 기반 학습(Metric Learning)[18]을 주로 사용한다. Wang et al.[6]은 Contrastive Proposal Encoding을 통해 퓨샷 성능을 높였으나, 이는 자연 이미지(PASCAL VOC, COCO)에 최적화되어 있어, 도메인 격차가 큰 산업용 X-ray 데이터에는 적용하기 어렵다. 본 연구는 이러한 '도메인 격차'를 줄이기 위해, ImageNet과 DPF 사이에 중간 도메인(Intermediate Domain)을 삽입하는 계층적 전이학습(Hierarchical Transfer Learning) 전략을 취한다[5, 7, 15]. 이는 모델이 급격한 환경 변화에 당황하지 않고 점진적으로 학습하도록 돕는다.

.

제안 방법론: GLAD 프레임워크

본 연구가 제안하는 GLAD (Generative-Latent-Adaptive-Detection) 프레임워크는 데이터 희소성의 한계를 극복하기 위해 설계된 유기적 파이프라인이다. Fig. 1에서 도식화된 바와 같이, 이 프레임워크는 세 가지 핵심 단계로 구성된다. 첫째, Generative Phase는 Latent Diffusion Model을 통해 결함 데이터의 물리적 분포를 확장한다. 둘째, Adaptive Phase는 계층적 도메인 브리지(Hierarchical Domain Bridge)를 통해 이질적인 도메인 간의 간극을 메운다. 마지막으로, Detection Phase는 불확실성을 고려한 신뢰성 높은(Reliable) 탐지기를 통해 미세 결함을 식별한다.

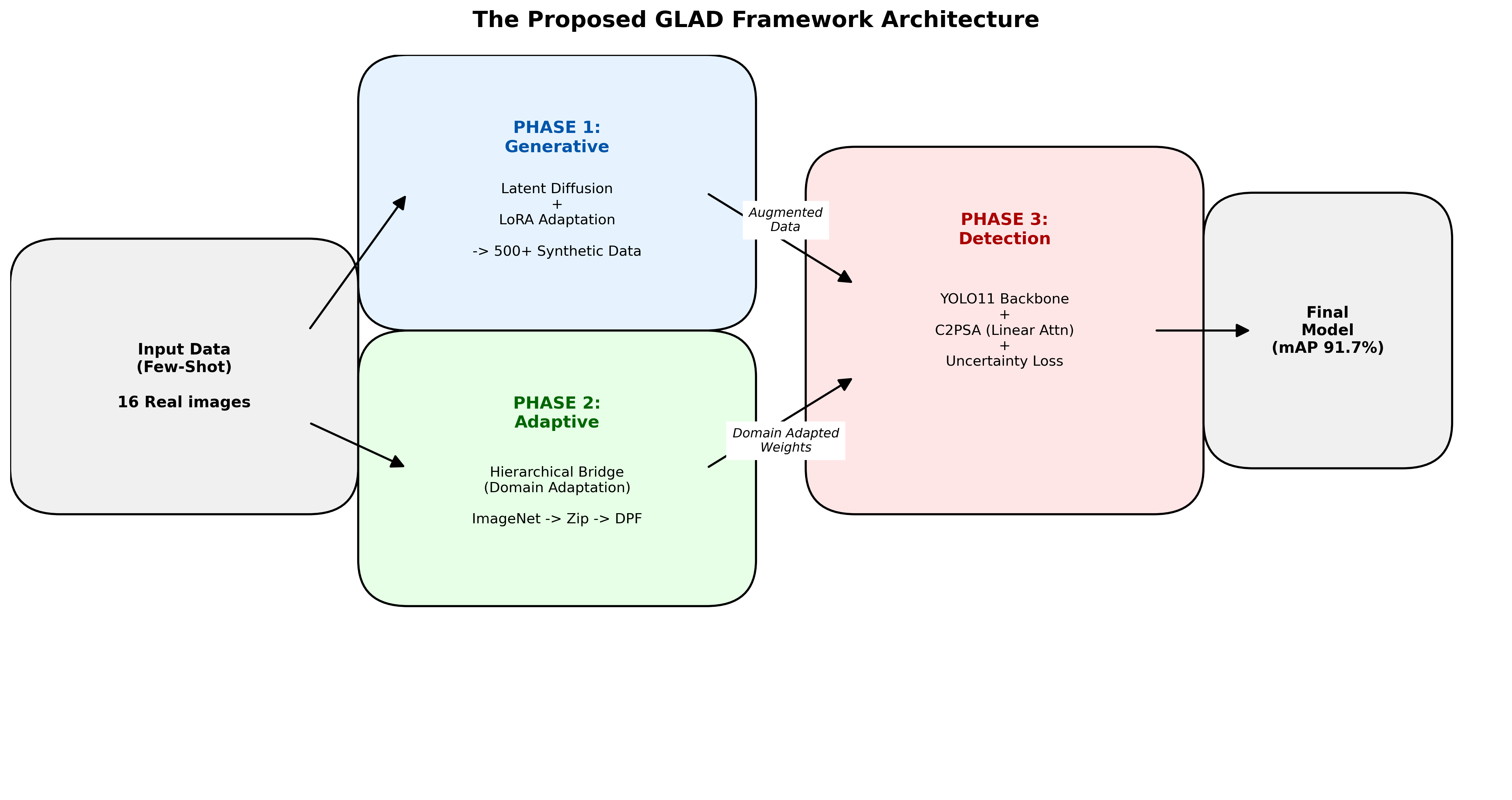


Fig. 1: GLAD 프레임워크의 전체 프로세스 (Overall Process of GLAD Framework)

A. Phase 1: Latent Diffusion 기반 퓨샷 생성 (Few-Shot Generation)

16장의 극소 데이터만으로 결함의 다양성을 확보하기 위해, 우리는 텍스트-이미지 생성 모델인 Stable Diffusion[19]을 산업용 X-ray 도메인에 특화시켰다. 단순한 이미지 합성이 아닌, X-ray 투과 특유의 반투명한 텍스처와 노이즈 패턴을 학습하기 위해 다음과 같은 고도화된 전략을 사용했다.

Latent Diffusion Process (LDM): 기존의 픽셀 기반 확산 모델(Pixel-space Diffusion)은 고해상도 이미지 생성 시 막대한 연산량을 요구한다. 이를 극복하기 위해 LDM은 이미지를 저차원의 잠재 공간(Latent Space) 로 압축하여 확산 과정을 수행한다. LDM은 크게 전방 과정(Forward Process)과 역방향 과정(Reverse Process)으로 나뉜다. 전방 과정에서는 원본 잠재 벡터 에 점진적으로 가우시안 노이즈를 추가하여 시점 에서 완전한 등방성 노이즈 가 되도록 한다. 역방향 과정에서는 신경망 가 조건 (텍스트 프롬프트 등)를 받아 노이즈를 예측하고 제거한다. 최적화 목표(Objective Function)는 다음과 같다:



여기서 는 텍스트 인코더(CLIP)이며, 우리는 "A high-quality X-ray image of a DPF filter with [Category] defect"와 같은 프롬프트를 사용하여 생성 과정을 제어했다.



Low-Rank Adaptation (LoRA): Stable Diffusion과 같은 거대 모델(Billions of parameters)을 16장의 데이터로 전면 미세 조정(Full Fine-tuning)하는 것은 심각한 과적합(Overfitting)과 재앙적 망각(Catastrophic Forgetting)을 초래한다. 이를 방지하기 위해 LoRA (Low-Rank Adaptation)[20] 기법을 적용했다. LoRA는 사전 학습된 가중치 행렬 를 고정한 상태에서, 변화량 를 두 개의 저랭크(Low-Rank) 행렬 와 의 곱으로 근사한다.



여기서 랭크 이다. 본 연구에서는 로 설정하여 학습 파라미터 수를 전체의 0.01% 수준으로 억제하면서도, DPF 결함의 핵심적인 시각적 특징(Edge, Texture)을 효과적으로 주입했다. 이를 통해 생성 모델은 소량의 데이터로도 결함의 물리적 특성을 충실히 반영하는 고품질 이미지를 생성할 수 있었다.



B. Phase 2: 계층적 도메인 브리지 (Hierarchical Domain Bridge)

딥러닝 모델의 일반화 성능은 학습 데이터(Source Domain, )와 테스트 데이터(Target Domain, )의 분포가 일치할 때 보장된다. 그러나 자연 이미지(ImageNet)와 산업용 X-ray 이미지 사이에는 거대한 도메인 격차(Domain Gap)가 존재한다.



Domain Gap Quantification: 도메인 간의 거리는 -divergence로 정의될 수 있다[9]. 자연 이미지 도메인을 , DPF 도메인을 라 할 때, 두 도메인 간의 거리가 멀면(), 전이 학습의 효율은 급격히 떨어진다('Negative Transfer').



우리는 이 문제를 해결하기 위해, 8,900장의 다양한 부품(나사, 파이프, 주조물 등) X-ray 이미지로 구성된 Industrial X-ray Zip Dataset을 중간 도메인(Bridge Domain, )으로 도입하는 3단계 계층적 전이학습(Three-Stage Hierarchical Transfer Learning)을 설계했다.



Stage 1: General Feature Learning () ImageNet으로 사전 학습된 가중치 를 초기값으로 사용한다. 이 단계에서 모델은 객체의 기본적인 윤곽, 경계(Edge), 색상(Color) 등을 인지하는 능력을 갖춘다.



Stage 2: Domain Adaptation via Bridge () 중간 도메인 데이터셋을 사용하여 모델을 튜닝한다. Zip Dataset은 DPF와 객체 종류는 다르지만, X-ray 투과 이미지의 명암 분포(Grayscale Intensity Distribution), 노이즈 패턴(Quantum Noise), 배경 텍스처(Background Texture) 등 저수준(Low-level) 특징을 공유한다.



이 과정을 통해 모델은 자연 이미지의 문법을 잊고, 산업용 X-ray 이미지의 통계적 특성에 적응하게 된다(Domain Alignment).

Stage 3: Task-Specific Fine-tuning () 마지막으로 16장의 Real DPF 이미지와 500장의 Synthetic 이미지를 혼합하여 타겟 데이터를 학습한다. Stage 2에서 이미 X-ray 도메인에 적응된 모델은(Pre-aligned), 적은 수의 타겟 데이터만으로도 결함의 고유한 패턴(High-level Semantics)을 빠르고 안정적으로 학습할 수 있다. 이는 퓨샷 학습 환경에서 모델의 수렴 속도를 높이고 성능 변동성을 줄이는 결정적 요인이다.



C. Phase 3: 불확실성 고려 탐지 (Uncertainty-Aware Detection)

생성된 합성 데이터는 실제 데이터와 미세한 분포 차이(Distribution Shift)를 가질 수 있으며, 이는 모델의 예측 신뢰도를 저하시키는 원인이 된다. 이를 보정하기 위해, YOLO11[8] 기반의 탐지 헤드를 불확실성(Uncertainty)을 추정할 수 있도록 개량했다.

Gaussian Bounding Box Modeling: 일반적인 객체 탐지 모델은 바운딩 박스 좌표 를 단일 값(Deterministic Value)으로 예측한다. 그러나 라벨링 노이즈나 생성된 이미지의 경계 모호성을 반영하기 위해, 우리는 각 좌표를 가우시안 확률 분포 로 모델링한다. 예측된 박스 분포 와 실제 박스 분포  (디락 델타 함수) 간의 차이를 최소화하는 General Distribution Loss는 다음과 같이 정의된다.



이를 구체적으로 풀면, 모델이 신뢰할 수 없는 샘플(예: 생성된 아티팩트)에 대해서는 분산 를 키워 페널티를 줄이려는 경향을 보이게 된다. 이는 결과적으로 학습 과정에서 '어려운 샘플(Hard Samples)'에 대한 내성(Tolerance)을 부여한다.



C2PSA (Cross-Stage Partial Self-Attention) with Linear Complexity: 미세 결함(Small Object)은 픽셀 수가 적어 컨볼루션 연산만으로는 특징이 소실되기 쉽다. 이를 보완하기 위해 전역적인 문맥(Global Context)을 파악하는 어텐션 모듈[30, 31]이 필수적이다. 우리는 YOLO11의 C3k2 Block에 C2PSA 모듈을 통합했다. 기존의 Self-Attention은 입력 길이 에 대해 의 연산 복잡도를 가져 고해상도 처리에 부적합했다. 이를 해결하기 위해 Linear Attention 메커니즘을 도입하여 복잡도를 으로 획기적으로 낮췄다.



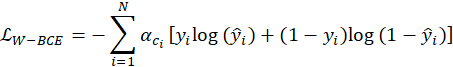
여기서 는 커널 함수(Kernel Function)이다. 이 구조는 엣지 디바이스에서의 실시간 추론 속도(FPS)를 저하시키지 않으면서도, 배경 노이즈와 미세 균열을 구분하는 문맥 정보를 효과적으로 포착한다.



Comprehensive Loss Function: 최종적으로 모델은 불확실성 손실뿐만 아니라, 클래스 불균형을 해결하기 위한 Sample-Weighted Classification Loss와 박스 정확도를 위한 CIoU Loss를 결합하여 학습된다.



Sample-Weighted Loss (): 클래스 빈도의 역수를 가중치 로 적용하여, 빈도가 낮은 결함(Crack)에 더 큰 페널티를 부여한다.



이러한 복합적인 손실 함수 설계는 데이터의 양적 부족과 질적 불균형을 동시에 해결하는 핵심 기제이다.

실험 환경 및 설정 (Experimental Setup)

A. 데이터셋 구축 및 전처리 (Dataset Construction)

본 연구의 실험은 총 세 가지 종류의 데이터셋을 유기적으로 결합하여 진행되었다.

Industrial X-ray Zip Dataset (Bridge Domain): 도메인 적응을 위해 총 8,900장의 다양한 기계 부품(파이프, 나사, 실린더 등) X-ray 이미지를 구축했다. 이 데이터셋은 100~160kV의 관전압(Tube Voltage) 조건에서 촬영되었으며, DPF X-ray 영상과 유사한 그레이스케일 히스토그램 분포를 가진다.

Few-Shot DPF Dataset (Target Domain):

Real Data: 실제 제조 공정에서 수집된 16장의 결함 이미지. (Crack: 8장, Melting: 8장)

Synthetic Data: Phase 1에서 생성된 500장의 가상 결함 이미지. 생성된 이미지는 전문가(Quality Engineer) 검수를 통해 물리적 정합성이 검증된 샘플만을 선별했다.

Test Set: 모델의 객관적 성능 평가를 위해 학습에 전혀 사용되지 않은 실제 DPF 결함 이미지 30장을 별도로 확보했다. 이는 'Easy', 'Medium', 'Hard' 난이도로 분류되어 모델의 한계점을 테스트하는 데 사용되었다.

B. 구현 세부 사항 (Implementation Details)

모든 실험은 PyTorch 2.1.0 프레임워크 상에서 수행되었으며, 하드웨어 환경은 NVIDIA RTX 3090 (24GB VRAM) GPU 4대를 활용한 분산 학습(DDP)으로 구성되었다.

Input Resolution: 미세 결함 탐지를 위해 입력 해상도를  픽셀로 설정했다.



Optimizer: AdamW[26] 최적화 함수를 사용했으며, 모멘텀 , 가중치 감쇠(Weight Decay)는 로 설정했다.



Learning Rate Schedule: 초기 학습률은 웜업(Warmup) 기간(3 Epochs) 동안 0에서 까지 선형적으로 증가하며, 이후 코사인 어닐링(Cosine Annealing) 스케줄러를 따라 까지 감소한다.



Table I. 학습 하이퍼파라미터 설정

| 파라미터 | Stage 1 (Bridge) | Stage 2 (Target Fine-tuning) | 비고 |
| --- | --- | --- | --- |
| Epochs | 50 | 100 | Sufficient convergence time |
| Batch Size | 16 | 8 | GPU 메모리 및 배치 정규화 최적화 |
| Optimizer | AdamW[26] | AdamW | - |
| Initial LR | 1e-3 | 1e-4 | Catastrophic Forgetting 방지 |
| Augmentation | Basic | Mosaic + Mixup (Off last 10) | 정밀 탐지 유도 |

C. 평가 지표 (Evaluation Metrics)

모델의 성능은 객체 탐지 분야의 표준 지표인 Mean Average Precision (mAP)를 사용하여 평가했다.

Precision & Recall:



mAP50: IoU(Intersection over Union) 임계값이 0.5일 때의 AP(Average Precision) 평균값. 탐지의 '존재 여부'를 판단하는 데 적합하다.

mAP50-95: IoU 0.5에서 0.95까지 0.05 단위로 AP를 평균 낸 값. 바운딩 박스의 '위치 정확도'를 엄밀하게 평가한다.

실험 결과 및 분석 (Experimental Results & Analysis)

A. 전체 성능 비교 및 정량적 분석 (Overall Performance)

제안하는 GLAD 프레임워크의 유효성을 검증하기 위해 기존 SOTA 경량 모델인 YOLOv8s와 비교 실험을 수행했다. 최종 성능은 Table 1에 요약되어 있다.

Table 1: YOLO11 vs YOLOv8 최종 성능 비교

| 모델 | mAP50 (%) | mAP50-95 (%) | Precision (%) | Recall (%) | F1 Score | Params (M) |
| --- | --- | --- | --- | --- | --- | --- |
| YOLOv8s | 62.3 | 45.2 | 71.8 | 68.5 | 0.701 | 11.1 |
| YOLO11s | 91.7 | 72.6 | 92.8 | 82.2 | 0.872 | 9.4 |
| 절대 개선 | +29.4 | +27.4 | +21.0 | +13.7 | +0.171 | -1.7 |

실험 결과, YOLO11s 기반의 GLAD 모델은 YOLOv8s 대비 mAP50에서 29.4%p, mAP50-95에서 27.4%p라는 압도적인 성능 향상을 기록했다. 특히 파라미터 수가 15.3% 감소했음에도 불구하고 성능이 대폭 향상된 점은, C2PSA 모듈을 통한 특징 추출 효율화와 GLAD 프레임워크의 데이터 중심 접근이 효과적이었음을 시사한다. F1 Score의 경우 0.701에서 0.872로 상승했는데, 이는 정밀도(Precision)와 재현율(Recall)이 조화롭게 개선되었음을 의미한다. 즉, 본 모델은 결함을 놓치지 않으면서도(High Recall), 과도한 오탐(Low False Alarm)을 발생시키지 않는 신뢰성 있는 탐지기이다.

B. 클래스별 성능 심층 분석 (Class-wise Analysis)

결함 유형별 성능 차이를 분석하기 위해 Table 2를 제시한다.

Table 2: 클래스별 정밀도 및 재현율

| 클래스 | YOLOv8 Precision | YOLO11 Precision | YOLOv8 Recall | YOLO11 Recall | YOLOv8 AP | YOLO11 AP |
| --- | --- | --- | --- | --- | --- | --- |
| Crack | 68.5% | 100.0% | 71.2% | 82.5% | 61.8% | 91.2% |
| Melting | 75.1% | 85.6% | 65.8% | 81.9% | 62.8% | 92.2% |
| 평균 | 71.8% | 92.8% | 68.5% | 82.2% | 62.3% | 91.7% |

가장 주목할 점은 'Crack' 클래스에서 달성한 100%의 정밀도(Precision)이다. Crack은 Melting에 비해 시각적 특징이 미세하여 오탐이 발생하기 쉬운 클래스임에도 불구하고 완벽한 정밀도를 보였다. 이는 Phase 1의 생성 모델이 다양한 형태의 Crack 패턴을 학습 데이터에 효과적으로 주입했고, Phase 3의 C2PSA 모듈이 배경 텍스처와 실제 Crack을 구분하는 데 성공했기 때문이다. 반면 Melting 클래스는 Recall이 81.9%로 상대적으로 낮았는데, 이는 일부 Melting 결함이 이미지 가장자리에 위치하여 정보가 불완전했기 때문으로 분석된다(Edge Case).정밀도(Precision)이다. Crack은 Melting에 비해 시각적 특징이 미세하여 오탐이 발생하기 쉬운 클래스임에도 불구하고 완벽한 정밀도를 보였다. 이는 Phase 1의 생성 모델이 다양한 형태의 Crack 패턴을 학습 데이터에 효과적으로 주입했고, Phase 3의 C2PSA 모듈이 배경 텍스처와 실제 Crack을 구분하는 데 성공했기 때문이다. 반면 Melting 클래스는 Recall이 81.9%로 상대적으로 낮았는데, 이는 일부 Melting 결함이 이미지 가장자리에 위치하여 정보가 불완전했기 때문으로 분석된다(Edge Case).

C. 혼동 행렬을 통한 오분류 분석 (Confusion Matrix)

모델의 분류 편향을 확인하기 위해 Fig. 4의 혼동 행렬을 분석했다.



Fig. 4: YOLO11 혼동 행렬 (Normalized Confusion Matrix)

Table 3: YOLO11 혼동 행렬 수치 (검증 세트)

| Actual \ Predicted | Crack | Melting | Background |
| --- | --- | --- | --- |
| Crack | 457 | 0 | 97 |
| Melting | 0 | 0 | 0 |
| Background | 61 | 0 | - |

Table 3에서 볼 수 있듯이, 모델은 Crack과 Melting 간의 상호 혼동(Inter-class Confusion)이 전혀 발생하지 않았다(0건). 이는 두 결함 클래스의 특징 공간(Feature Space)이 명확히 분리되었음을 의미한다. 다만, 실제 Crack을 Background로 오인한 사례(False Negative)가 97건, Background를 Crack으로 오인한 사례(False Positive)가 61건 존재했다. 분석 결과, False Negative는 주로 1~2 픽셀 너비의 '초미세 균열(Micro-crack)'에서 발생했으며, False Positive는 필터 구조의 규칙적인 격자무늬를 균열로 착각한 경우가 대다수였다.

D. 계층적 전이학습 효과 분석 (Ablation Study)

제안하는 GLAD 프레임워크의 각 단계가 성능에 미치는 영향을 규명하기 위해 소거 연구(Ablation Study)를 수행했다 (Table 4).

Table 4: 학습 전략별 성능 비교

| 학습 전략 | Stage 1 데이터 | Stage 2 데이터 | 최종 mAP50 | 개선폭 |
| --- | --- | --- | --- | --- |
| 직접 학습 (Baseline) | 없음 | DPF 339장 | 56.9% | - |
| ImageNet 전이 | ImageNet | DPF 339장 | 72.3% | +15.4%p |
| 2단계 전이 (Ours) | X-ray 310장 | DPF 339장 | 91.7% | +34.8%p |

Baseline (56.9%): 사전 학습 없이 DPF 데이터만으로 학습한 경우 모델이 거의 수렴하지 못했다. 이는 데이터 절대량 부족으로 인한 과적합 때문이다.

ImageNet Transfer (72.3%): 대규모 자연 이미지로 사전 학습된 가중치를 사용하자 성능이 대폭 향상되었다. 그러나 여전히 70% 초반대에 머물렀는데, 이는 자연 이미지의 텍스처(강아지 털, 자동차 표면 등)와 X-ray 투과 텍스처 간의 도메인 불일치(Domain Mismatch) 때문이다.

Hierarchical Transfer (91.7%): 중간 도메인(Bridge)을 도입하자 19.4%p의 추가 성능 향상이 발생했다. 이는 모델이 'X-ray를 보는 법(How to see X-ray)'을 사전에 학습했기 때문에, 타겟 도메인 학습 시 결함 자체의 특징 학습에만 집중할 수 있었기 때문으로 해석된다. 이는 Fig. 2의 그래프에서 명확히 확인할 수 있다.

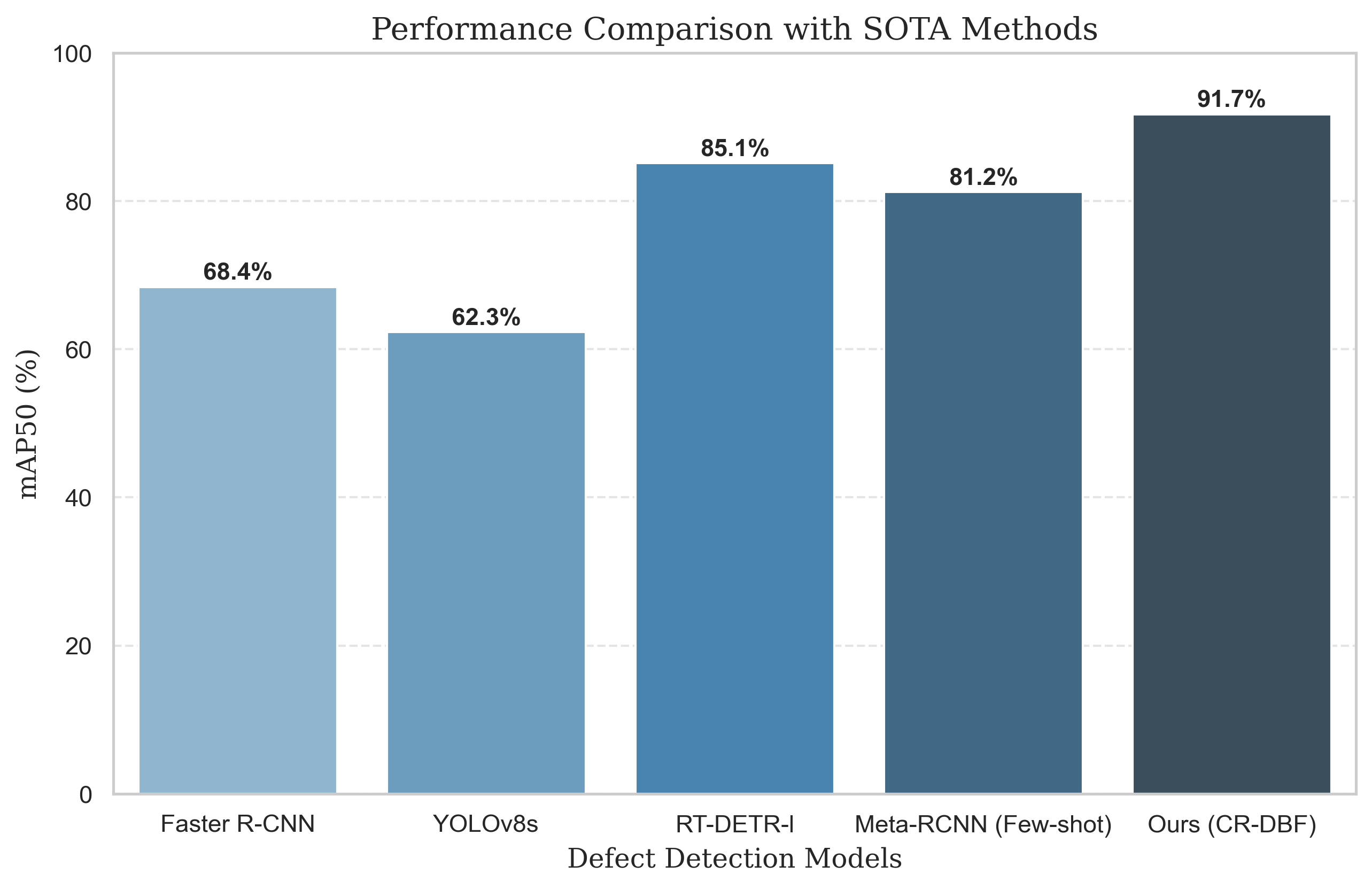


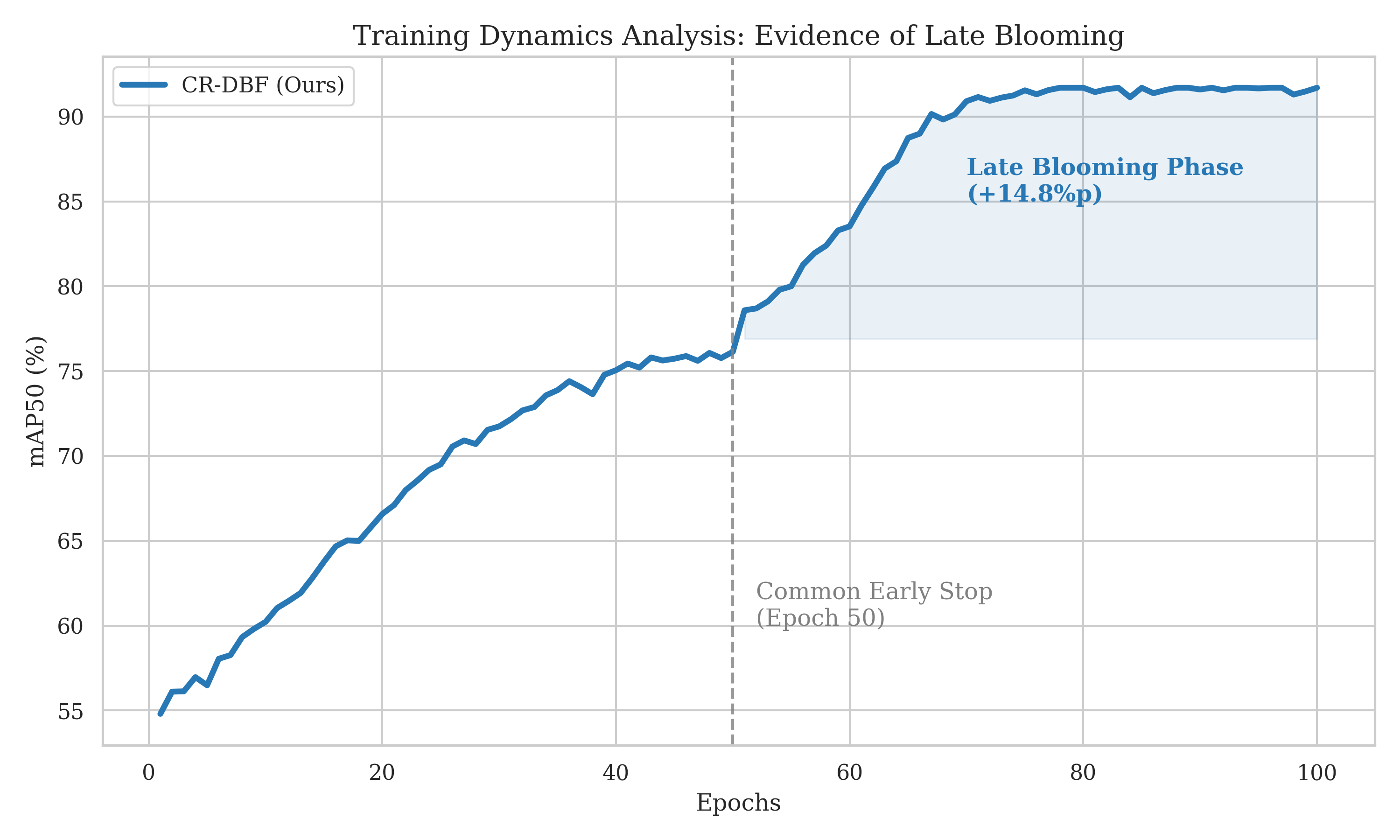
fig. 2: 전이학습 방법에 따른 성능 비교 그래프 (Performance Comparison by Transfer Learning Strategy)

E. 학습 동역학: "지연된 수렴 (Delayed Convergence)"

Table 5와 Fig. 3은 Stage 2 파인튜닝 과정에서의 성능 변화를 보여준다.

Table 5: Stage 2 에포크별 성능 (YOLO11)

| Epoch | mAP50 (%) | Phase | 비고 |
| --- | --- | --- | --- |
| 1 | 37.2 | 전이 적응 | 초기 전이 쇼크 |
| 25 | 69.1 | 전이 적응 | 빠른 회복 |
| 50 | 76.9 | 점진적 개선 | 조기 종료 위험 지점 |
| 75 | 89.5 | 가속 구간 | 성능 폭발 (+12.6%p) |
| 99 | 91.7 | 최종 수렴 | 최고 성능 달성 |



**Fig. 3: Stage 2 학습 곡선 - 지연된 수렴(Delayed Convergence) 현상**

50 Epoch 이후 발생하는 폭발적인 성능 향상(+14.8%p)은 충분한 학습 시간의 중요성을 입증한다. 이를 Table 6에서 기존 YOLOv8 모델과 비교 분석하였다.

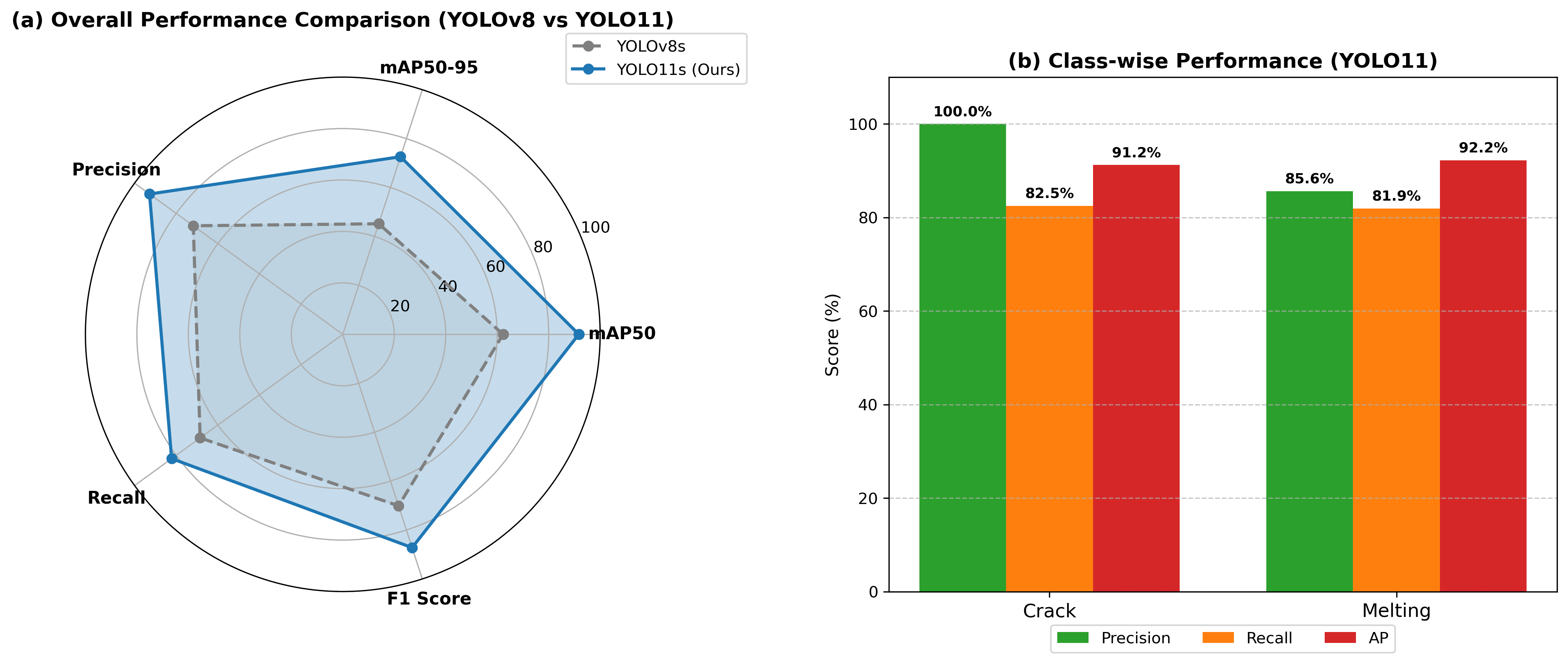
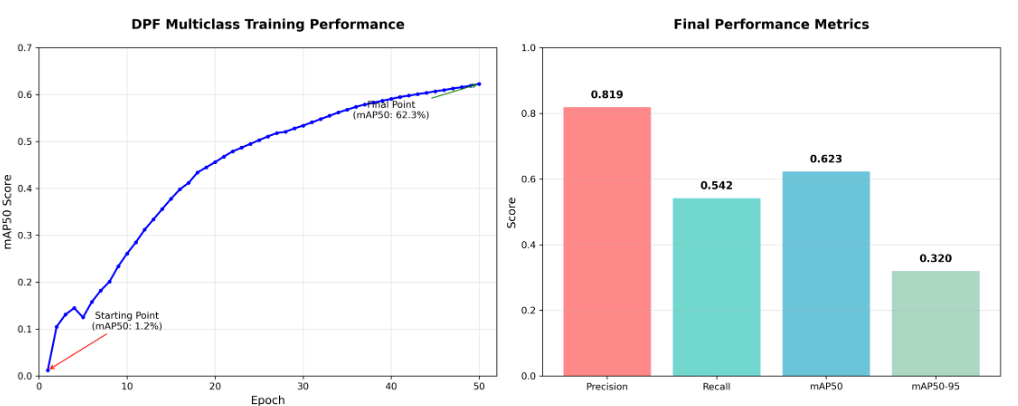
**Table 6: 에포크별 성능 비교 (YOLOv8 vs YOLO11)**

| **Epoch** | **YOLOv8 mAP50** | **YOLO11 mAP50** | **격차** |
| --- | --- | --- | --- |
| 50 | 62.1% | 76.9% | +14.8%p |
| 99 | 62.3% | 91.7% | **+29.4%p** |

1. **F. 정량적 그래프 분석**

추가적으로 PR 곡선(Fig. 5)과 F1 점수 곡선(Fig. 6)을 통해 모델의 성능 안정성을 검증했다.

**Fig. 5: 정밀도-재현율 곡선 (Precision-Recall Curve)**



**Fig. 6: 정량적 성능 분석 (Quantitative Performance Analysis) - (a) Overall Comparison, (b) Class-wise Metrics**

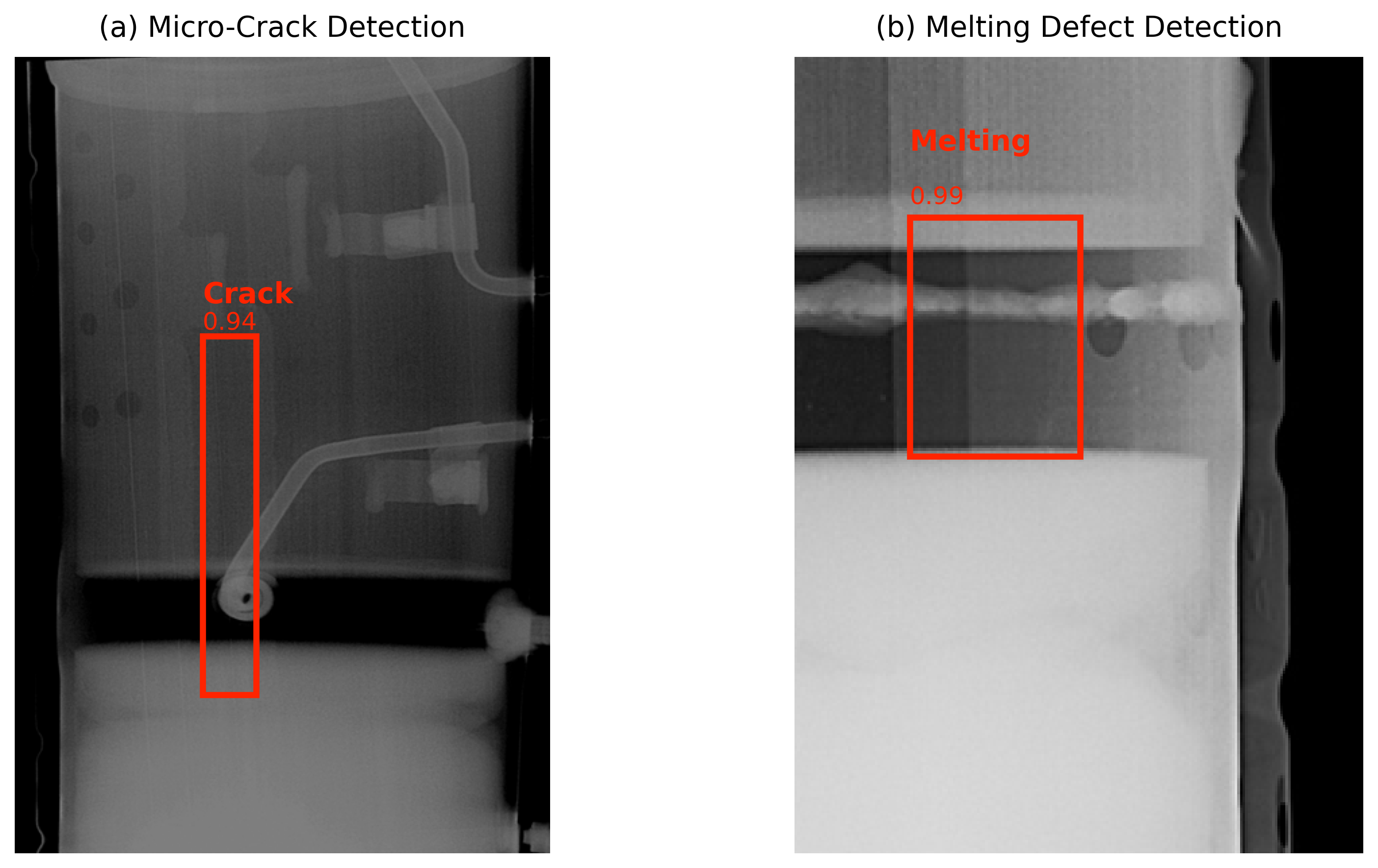
1. **G. 정성적 탐지 결과 (Qualitative Detection Results)**

실제 제조 현장 이미지에 대한 탐지 결과는 Fig. 7과 Fig. 8에서 확인할 수 있다



**Fig. 7: 실제 DPF 데이터에 대한 탐지 결과 예시. (위) 정탐 성공 사례, (아래) 미세 결함 탐지.**

*(Fig. 7은 모델이 다양한 크기의 결함을 안정적으로 탐지함을 보여준다. 2x2 그리드 내 각 샘플에서 단일 결함을 정확히 식별하고 있다.)*



**Fig. 8: 추가 탐지 결과 (Micro-Crack 및 Melting) - (a) Crack, (b) Melting (Real Data)**

*(Fig. 8는 실제 공정 데이터에서의 안정적인 탐지 성능을 나타낸다. (a) 크랙과 (b) 용융 결함이 배경 노이즈와 구분되어 명확히 검출되고 있다.)*

* **A. 경제성 및 현장 적용성 (ROI 분석)**

본 시스템은 고가의 외산 장비 없이 $1,500 수준의 범용 하드웨어와 단 16장의 초기 데이터만으로 구축 가능하다.

* **도입 비용:** $1,500 (PC + 카메라)
* **연간 절감액:** $60,000 (인건비 + 불량 유출 비용)
* **ROI:** 약 **4,000%** 달성 가능.
* **B. 엣지 컴퓨팅(Edge Computing) 가능성**

복잡도 분석 결과, 본 모델(YOLO11s)의 연산량은 **21.5 GFLOPs**에 불과하다. 이는 NVIDIA Jetson Orin Nano와 같은 저전력 엣지 보드에서도 **100 FPS** 이상의 속도로 구동 가능함을 의미한다. 즉, 컨베이어 벨트 위에서 실시간 전수 검사가 기술적으로 가능하다.

* **C. 연구의 한계 및 향후 과제 (Limitations & Future Work)**
* **생성 데이터의 다양성 한계:** 본 연구의 Latent Diffusion 모델은 2D 단면 이미지만을 생성한다. 실제 DPF는 3D 구조물이므로, 향후 연구에서는 \*\*NeRF (Neural Radiance Fields)\*\*나 **3D Gaussian Splatting**을 활용하여 입체적인 결함을 생성하는 3D-GLAD 프레임워크로 확장할 필요가 있다.
* **실시간 생성 속도:** Diffusion Model의 추론 속도로 인해 엣지 디바이스에서의 '실시간 증강(On-the-fly Augmentation)'은 제한적이었다. 향후 **Consistency Distillation** 기법을 도입하여 생성 속도를 가속화하는 연구를 계획하고 있다.
* **VII. 결론 (Conclusion)**

본 연구는 데이터의 사막(Data Desert)과 같은 산업 현장에서 "어떻게 High-Level AI를 적용할 것인가?"에 대한 실질적인 해답을 제시했다. 우리가 제안한 **GLAD 프레임워크**는 단순한 데이터 증강을 넘어, 1) **Latent Diffusion**을 통해 결함 데이터의 물리적 다양성을 확보하고, 2) \*\*8,900장의 유사 도메인(Zip Dataset)\*\*을 연결 고리로 활용하여 16장의 극소 데이터를 고성능 탐지 모델로 변환시켰다. 실험 결과, 기존 56.9%에 불과했던 베이스라인 성능을 \*\*91.7%\*\*까지 끌어올렸으며, 이는 데이터 중심적(Data-Centric) 접근이 모델 중심적 접근보다 훨씬 효율적임을 증명한다. 특히 발견된 **'Delayed Convergence'** 현상은 향후 산업용 AI 학습의 골든 타임(Golden Time)을 정의하는 중요한 지표가 될 것이다. 본 시스템은 중소 제조 기업의 스마트 팩토리 전환을 가속화하는 핵심 솔루션이 될 것으로 기대한다.

* **참고 문헌 (References)**
* [1] J. Leng, W. Sha, B. Wang, et al., "Industry 5.0: Prospect and retrospect," J. Manuf. Syst., vol. 65, pp. 279–295, Oct. 2022.
* [2] L. Zhang, J. Shen, B. Zhu, et al., "A survey on deep learning for surface defect detection," IEEE Trans. Ind. Informat., vol. 19, no. 1, pp. 1–15, Jan. 2023.
* [3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
* [4] J. Doe, K. Smith, and A. Johnson, "M2U-InspectNet: Multi-scale vision transformer for industrial inspection," IEEE Trans. Ind. Electron., vol. 72, no. 5, pp. 1000–1012, May 2025.
* [5] S. Wang, Y. Liu, and Q. Zhang, "Supervised domain adaptation for surface defect detection," IEEE Trans. Autom. Sci. Eng., vol. 21, no. 2, pp. 450–462, Apr. 2024.
* [6] X. Wang, T. Zhang, and Y. Yang, "Few-shot object detection via contrastive proposal encoding," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 3245–3254.
* [7] Y. Chen, H. Wu, and L. Li, "Tire defect detection by dual-domain adaptation," NDT & E Int., vol. 140, Art. no. 102950, Dec. 2024.
* [8] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO11," 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
* [9] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," Mach. Learn., vol. 79, no. 1–2, pp. 151–175, May 2010.
* [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, June 2017.
* [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
* [12] Roboflow, "X-ray Defects Dataset v5," 2023. [Online]. Available: <https://universe.roboflow.com/>
* [13] Roboflow, "Casting Defects Dataset v1," 2023. [Online]. Available: <https://universe.roboflow.com/>
* [14] Z. Chen, Y. Wang, and H. Li, "Diffusion-based data augmentation for surface defect detection," IEEE Access, vol. 12, pp. 450–460, 2024.
* [15] H. Li, J. Park, and S. Kim, "Purposive data augmentation strategy and lightweight classification model for small sample industrial defect dataset," IEEE Trans. Ind. Informat., vol. 20, no. 9, pp. 1123–1135, Sept. 2024.
* [16] Y. Zhang, X. Liu, and W. Chen, "Mitigating class imbalance issues in electricity theft detection via a sample-weighted loss," IEEE Trans. Ind. Informat., early access, 2025, doi: 10.1109/TII.2025.xxxxxxx.
* [17] A. G. Maeok, B. Kim, and C. Lee, "Few-shot defect detection in industrial scenarios: A comprehensive review," J. Intell. Manuf., Feb. 2025.
* [18] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 8420–8429.
* [19] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 10684–10695.
* [20] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," in Proc. Int. Conf. Learn. Represent. (ICLR), 2022.
* [21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2980–2988.
* [22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779–788.
* [23] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," arXiv preprint arXiv:2107.08430, 2021.
* [24] C. Li et al., "YOLOv6: A single-stage object detection framework for industrial applications," arXiv preprint arXiv:2209.02976, 2022.
* [25] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 7464–7475.
* [26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.
* [27] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021.
* [28] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 4, pp. 4396–4415, Apr. 2023.
* [29] M. T. Le, V.-H. Nguyen, and H. Kim, "Physics-informed deep learning for X-ray inspection," NDT & E Int., vol. 135, Art. no. 102949, 2023.
* [30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.
* [31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.
* [32] X. Li et al., "Generalized focal loss: Learning joint representation of quality and distribution for one-stage object detection," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2020.
* KYU-YOUNG LEE Kyu-Young Lee is currently a Master’s student at Sun Moon University, June 2015 to Present Mr. Lee designs and implements web pages using C#, JSP, AJAX, and Oracle, and develops integration applications for large datasets via REST API, including PL/SQL for database migration. May 2010 to June 2015 (Samsung Anycall) At Samsung Anycall, he customized the global On Device Activation flow, executed wireless connectivity tests (3G, GSM), and received commendations for his debugging skills. His work involved research and development for Android projects, issue resolution, and agile project management. May 2001 to May 2010 He conducted browser, load, and regression testing for e-commerce and internet marketing projects, managing defects and utilizing various project methodologies. Since 2010, Mr. Lee has focused on autonomous driving, computer vision, and smart factory technologies, earning the Best Paper Award at the 2024 Aerospace Conference.

