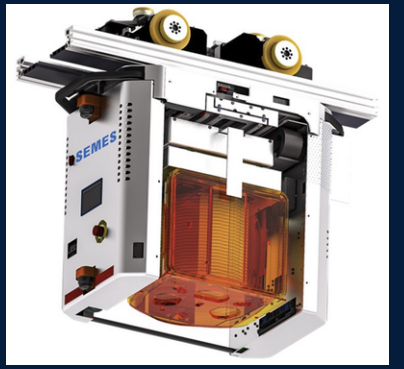


## 2D 비전 기반 OHT 센서 신호 처리용 온디바이스 AI 시스템 구축



제한된 엣지 환경에서의 극단적 추론 가속 및 최적화 파이프라인

### 산업적 배경 & 문제 정의

- 산업적 요구 → 팍 OHT, 24h 가동, 충돌·결함 실시간 감지 필수
- 비용 한계 → 3D LiDAR 고비용 / 클라우드 전송 시 지연·보안
- 해결 방안 → Jetson 단독 추론 + 2D 비전 + 시계열

### 실시간 OHT 인지 파이프라인

- 1단계 (객체 인지): YOLOv8n 기반 경량 객체 탐지 모델을 활용한 전방 객체 실시간 인식
- 2단계 (객체 추적): ByteTrack 알고리즘을 결합하여 연속 프레임 상의 이동 궤적 및 속도 계산
- 3단계 (충돌 예측): Bounding Box 크기 변화율 기반 충돌 예상 시간(TTC) 산출 및 위험 감지

### 레일 이상 탐지

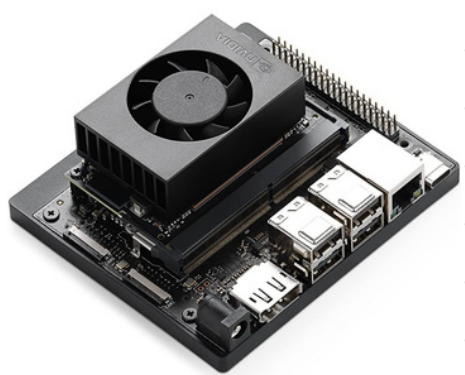
- 비지도 학습 적용: Autoencoder 기반 모델로 정상 레일 이미지 학습 후 실시간 재건 오차 추출
- 오탐지 최소화: 1D-CNN 시계열 필터링을 결합하여 일시적 노이즈를 배제하고 실제 결함(크랙, 이물질 등)만 정밀 타격

### 프로젝트 계획

"캡스톤1은 추론 최적화를 위한 시스템 인프라 구축에 집중하며, 캡스톤2 고도화를 위한 견고한 기반이 된다"

구분	캡스톤 1 (본 과제)	캡스톤 2 (후속)
레이어	시스템·하드웨어 레벨	모델 레벨
목적	인프라 + 베이스라인 + 마이크로 분석	본격 압축 + 도메인 통합
기법	ONNX/TRT 변환, 양자화(FP16/INT8), Calibration, Operator Fusion	Knowledge Distillation, Pruning, QAT
데이터	COCO 사전학습 (통제 변수)	AGV → OHT 도메인 데이터

### 환경설정



- Hardware: Jetson Orin Nano 8GB
- JetPack: 6.1(L4T R36.4, CUDA 12.6)
- PyTorch 2.5.0a0
- Torchvision: 0.20.0
- TensorRT: 10.3.0
- ONNX Runtime: 1.19
- Camera: USB webcam
- Nsight: 2026.2.1

### 최종 기대효과 및 확장성

- 기술적 가치: 무거운 AI 모델을 엣지용으로 압축/가속하는 엔지니어링 파이프라인 확립
- 산업적 가치: 라이다 센서 비용 절감 및 통신 없는 실시간 OHT 주행 안전성 확보
- 미래 확장성: 스마트팩토리, AGV, AMR 등 다양한 실내 물류 로봇 분야로의 기술 전이

### 캡스톤2 로드맵

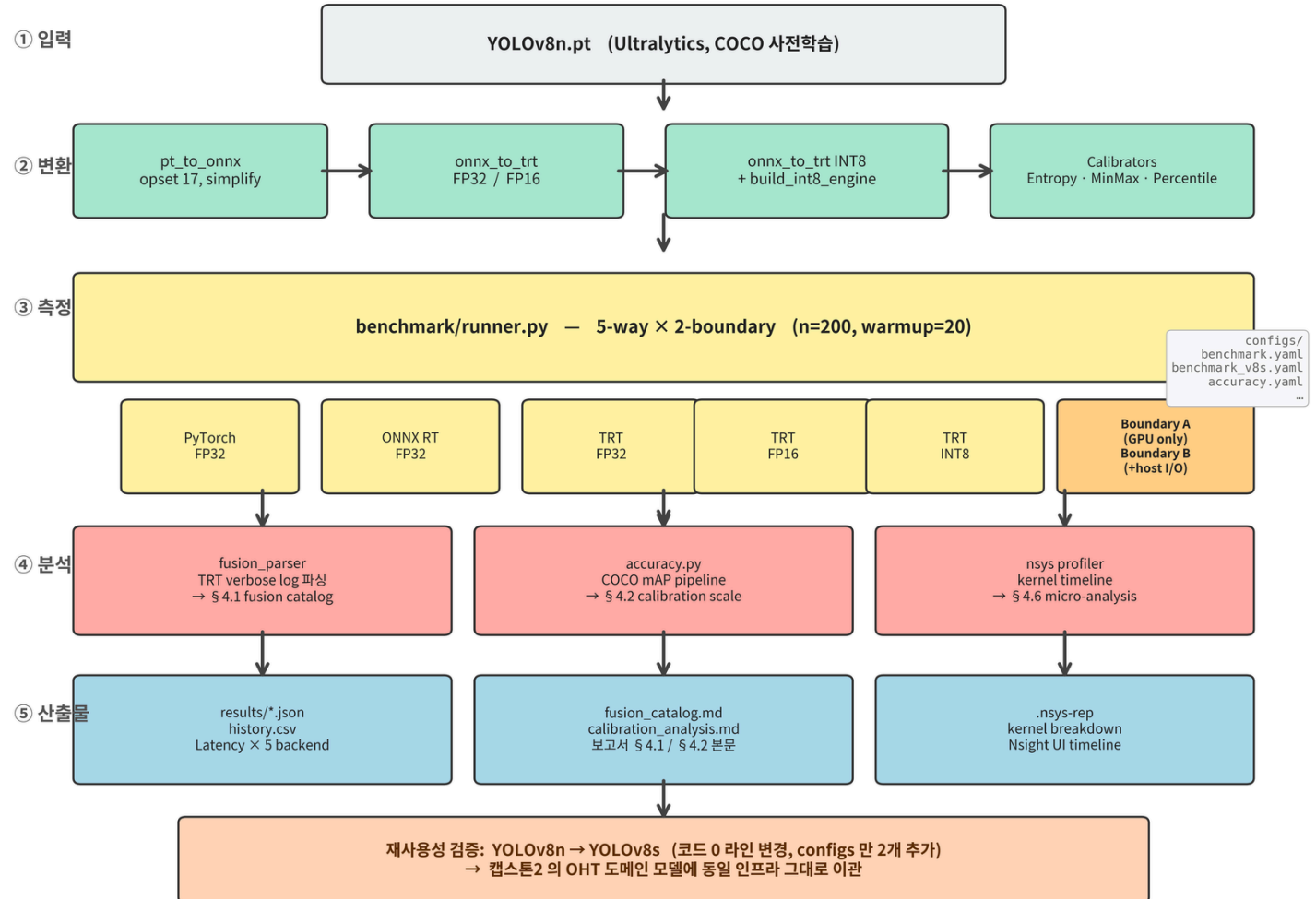
System Layer (COCO pretrained, 본 캡스톤1)  
 → Model Layer (AGV/OHT 도메인 데이터셋)  
 + KD / Pruning / QAT - Autoencoder 기반 이상 감지·다중 작업(검출 + 트래킹 + 이상)

### 시스템 아키텍처

"YOLOv8n.pt 한 모델을 5가지 백엔드에서 측정·분석하는 자동화 파이프라인"

추론 최적화 인프라

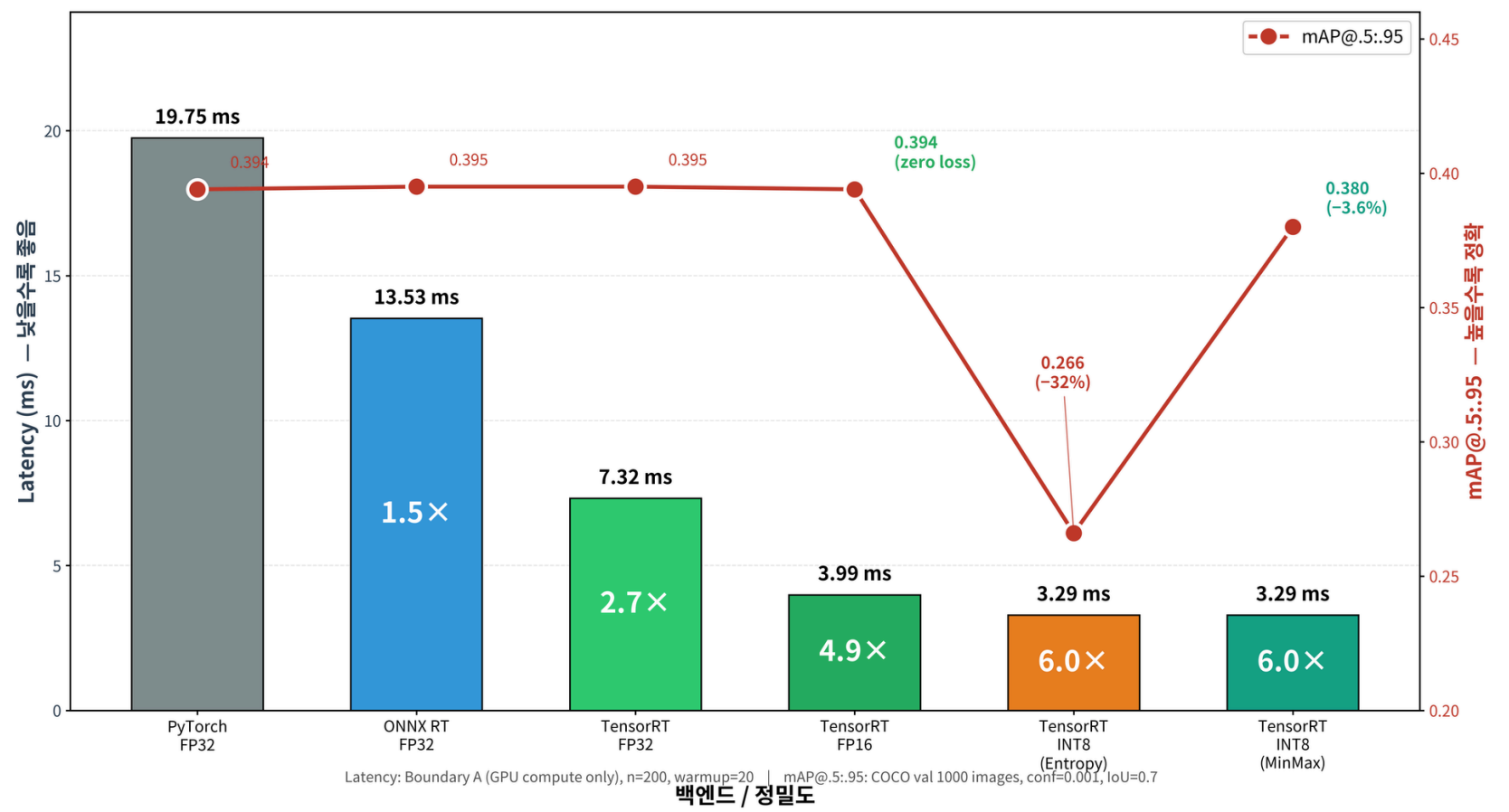
도메인 독립적 측정·분석 파이프라인 (캡스톤1의 핵심 산출물)



### 5-WAY 백엔드 비교 그래프

"본 과제의 핵심 산출물 ①: 5-way x 2-boundary 정량 분석"

YOLOv8n on Jetson Orin Nano 8GB: 5-way 백엔드 비교



PyTorch FP32 대비 TRT INT8에서 6배 가속, 그러나 mAP는 -3.6%만 손실. FP16까지는 정확도 손실 0 (mAP 0.394 → 0.394)

### 시스템 재사용성 검증 그래프

"본 과제의 핵심 산출물 ②: 인프라의 모델 독립성 실증"

인프라 재사용성 검증: YOLOv8n → YOLOv8s (3.5x 큰 모델, 코드 0 라인 변경)

