AIMS Lab | 한국항공대학교 조재훈 교수 연구실

We build vision AI that works in the real world

aimslab.vercel.app



조재훈 교수

Jaehoon Cho

- 연구실 정보
  - 현재 인원 : 석사 과정 2 명 + 학부연구생 3 명
  - 산업체 인력과 협업 가능 (현대자동차 , 삼성전자 , KETI 등)
  - 대학원 진학을 위해서는 학부 인턴 생활을 권장함
  - 홈페이지 : https://aimslab.vercel.app/

Autonomous Driving | VLM / VQA | Knowledge Distillation
Quantization | Monocular Depth | RGB-D Dataset
Drone Perception | Adverse Weather

❖ 목표

- 예측할 수 없는 상황에서도 AI 가 세상을 보고 이해하도록 실제 환경에서 동작하는 Vision AI 를 연구함

❖ 연구주제

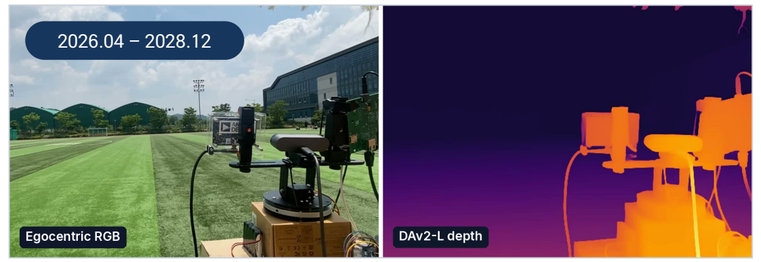
- Physical AI: 드론 (UAV), 자율주행 , Edge device (핸드폰 , 소형기기) 등
- AI 모델 최적화 경량화 : Knowledge Distillation, 양자화 , 경량화 등
- 실제 환경을 고려한 AI 모델 : 악천후 (안개 , 눈 , 비) 와 열화 상황에 강인한 AI 모델

❖ Ongoing Projects



Drone Perception in Adverse Weather

고도 · 자세 정보를 활용한 물리 기반 멀티센서 융합으로 안개 · 비 속에서도 강건한 드론 인지 기술을 개발함



Outdoor Smart Glasses

기능 모듈 재구성을 통해 스마트 글래스의 실외 깊이 추정 성능을 높이는 기술을 개발함

Research Theme

Theme Topic #1 — Physical AI: Robust Perception for Autonomous Driving & UAV



❖ Grounded VLM Reasoning for Driving Decisions (Scene VQA)
❖ LiDAR · Radar · Camera Fusion for 3D Spatial Understanding (BEV)

Autonomous Driving

VLM-based Driving Scene VQA



Front camera · real urban scene

GROUNDED REGIONS
[A] drivable area — lane area to stay in
[B] traffic light — signal state ahead
[C] lead car — vehicle in front

Q. What are you paying attention to?

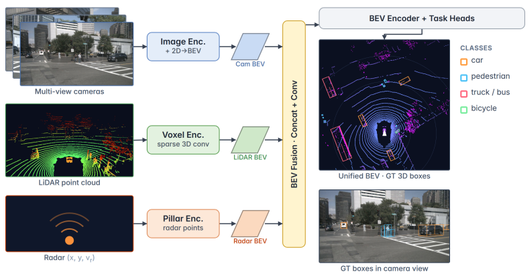
Text Tokenizer

Vision Encoder ViT → Projector MLP → visual tokens
text tokens

Large Language Model
decoder · fine-tuned on driving QA

A. Currently, I'm focusing on the **drivable area** [A] to ensure we stay within the lane. I'm also monitoring the **traffic** [B] in front of our **lead car** [C] to anticipate my next action.

Multi-Sensor Fusion → Unified BEV — nuScenes

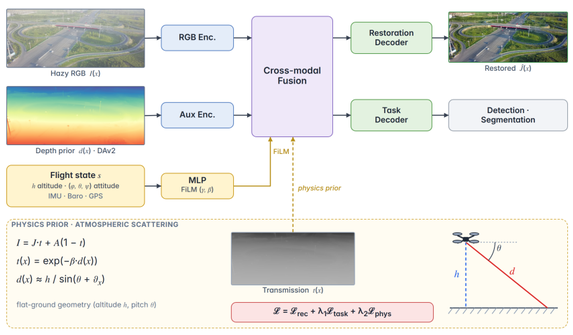


Cameras carry rich semantics but weak geometry; LiDAR and radar give precise 3D structure but little semantics. Lifting each sensor to one **bird's-eye-view grid** fuses both, cell by cell. Real nuScenes sample; boxes are ground-truth annotations.

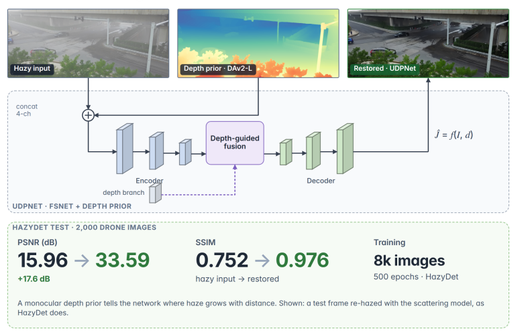
❖ Physics-guided Fusion with Flight State (Altitude · Attitude)
❖ Atmospheric Scattering & Depth Priors for Fog · Rain-Robust Perception

Drone (UAV)

Altitude·Attitude-Aware Physics-Guided Fusion — Ongoing



Depth-Prior Dehazing for UAV — HazyDet



A monocular depth prior tells the network where haze grows with distance. Shown: a test frame re-hazed with the scattering model, as HazyDet does.

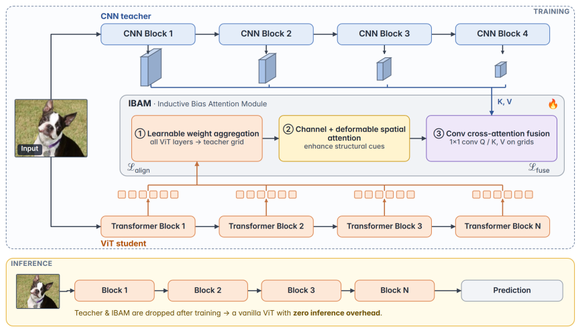
Theme Topic #2 — AI Model Optimization & Lightweight 3D Depth Estimation



❖ Knowledge Distillation: CNN → Lightweight ViT (No Inference Overhead)
❖ INT8 Quantization for Edge NPU: Latency −67% · Model Size −65%

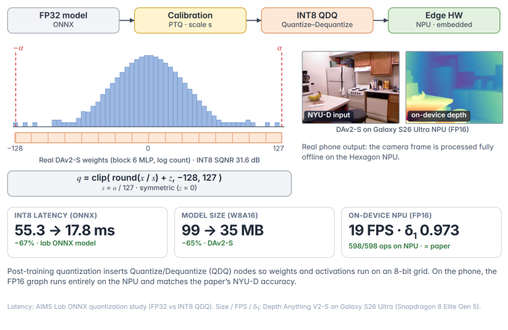
Model Optimization

Grid-Preserving KD: CNN Teacher → ViT Student — arXiv 2026 · iBKD



Teacher & IBAM are dropped after training → a vanilla ViT with **zero inference overhead**.

FP32 → INT8 Quantization — On-device



Real phone output: the camera frame is processed fully offline on the Hexagon NPU.

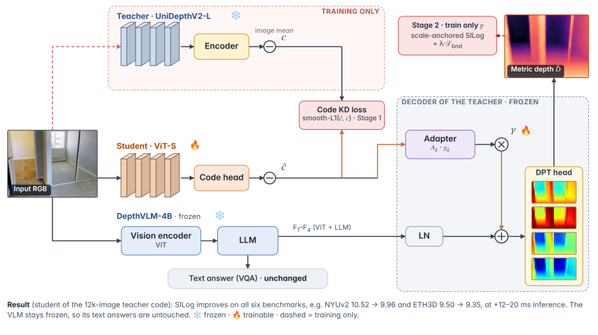
Post-training quantization inserts Quantize/Dequantize (QDQ) nodes so weights and activations run on an 8-bit grid. On the phone, the FP16 graph runs entirely on the NPU and matches the paper's NYU-D accuracy.

Latency: AIMS Lab ONNX quantization study (FP32 vs INT8 QDQ). Size / FPS / δ_1: Depth Anything V2-S on Galaxy S26 Ultra (Snapdragon 8 Elite Gen 5).

❖ Teacher → Student Code Distillation for Efficient Metric Depth in VLMs
❖ Confidence-guided Distillation: Stereo Teacher → Shallow Mono Student

Efficient Depth

VLM Metric Depth · Teacher → Student Code — Ongoing



Result (student of the 12k-image teacher code): SILog improves on all six benchmarks, e.g. NYUv2 10.52 → 9.96 and ETH3D 9.50 → 9.35, at +12–20 ms inference. The VLM stays frozen, so its text answers are untouched. ❄ frozen · 🔥 trainable · dashed = training only.

Stereo Confidence → Monocular Depth

