2024 embedded VISION SUMMIT®

Optimized Vision Language Models for Intelligent Transportation System Applications

Tae-Ho Kim

CTO & Co-Founder

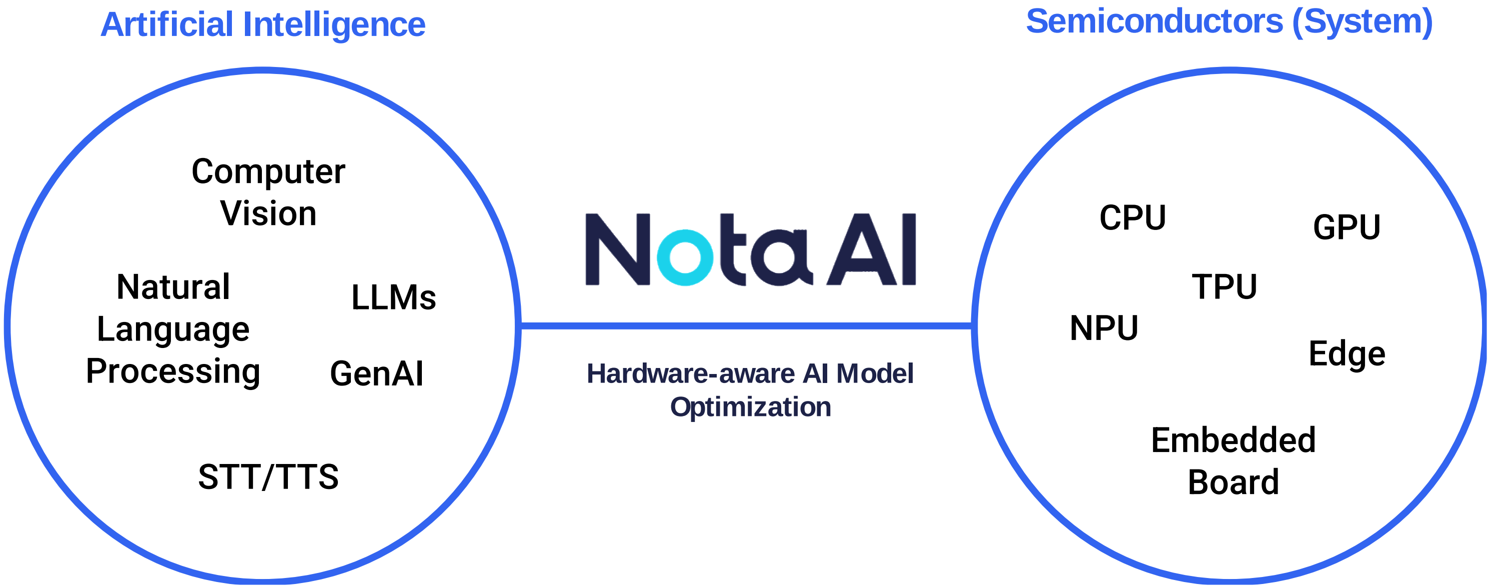
Nota Inc.

Nota AI

Introduction

- In this talk, we explore the **challenges in ITS**.
- How **vision language model (VLM)** can solve these challenges.
- **Future work** will also be addressed.

Identity of NotaAI



Nota AI bridges the gap between AI & semiconductors.

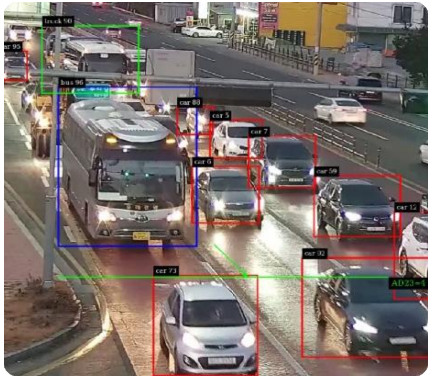
Nota AI's Main Product: NetsPresso

NetsPresso® simplifies AI model optimization for target devices with automated processes.

Platform & Edge Solutions: Elevating Excellence

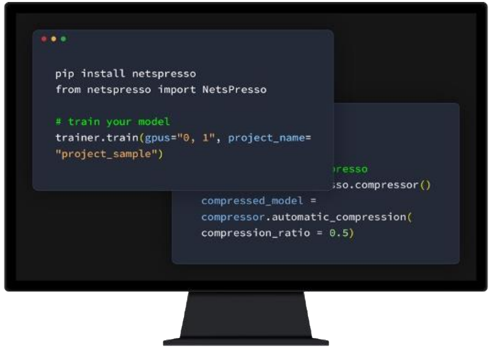
Solution

Nota ITS



Platform

NetsPresso



Solution

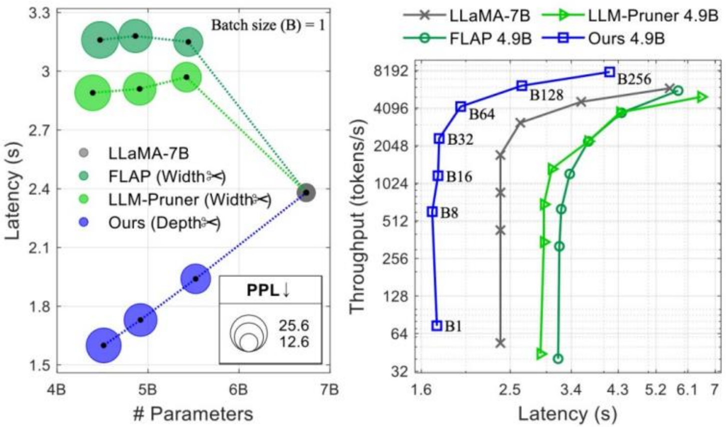
Nota DMS

Using NetsPresso, Nota AI also has created a solution business in ITS

NotaAI's Expertise

embedded VISION SUMMIT

LLM Optimization



Stable Diffusion Optimization

ICML International Conference On Machine Learning 40 Years

ICCV23 PARIS

(a) Efficient General-purpose T2I

(b) Efficient Personalized T2I

(c) Efficient I2I (Different Styles)

Device	SDM-v1	Base [Ours]	Small [Ours]	Tiny [Ours]
AGX Orin 32GB	4.9s [53]	3.4s (-31%)	3.2s (-35%)	2.8s (-43%)
iPhone 14	5.6s [52]	4.0s (-29%)	3.9s (-29%)	3.9s (-29%)

Nota AI also specializes in GenAI compression

Intelligent Transportation System (ITS)

Enhancing Traffic Flow: Real-Time Lightweight ITS Solutions

Smart Intersection System

Real-time incident detection and analysis for safe road condition

VRU Safety Solutions

AI-based real-time hazardous situation screening and control system

Automatic Incident Management System

Real-time incident detection and analysis for safe road condition

AI Smart Parking

Real-time parking occupancy and parking facility utilization analysis

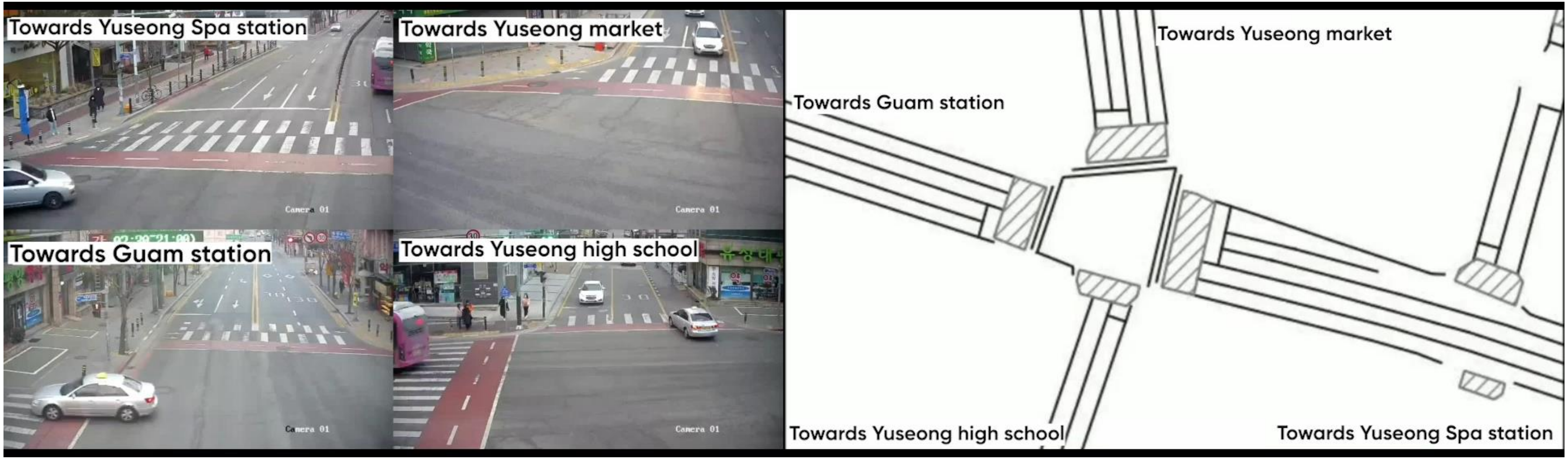
Use Case 1: Smart Intersection System

DAEJEON
DAEJEON METROPOLITAN CITY

- Daejeon metropolitan city ITS construction project
- Intersection CCTV AI video analysis (600 ch)
- 98% accuracy in traffic volume counting in night and rainy conditions

Use Case 2: AI Safe Crossing

Daejeon City | Implementing LDM using vehicle trajectory prediction information



Use Case 3: VRU Safety Solutions

embedded VISION SUMMIT

National Highways | C-ITS: AI video analysis for smart crosswalk safety transmitted through V2X communication

Use Case 4: AI Smart Parking

The City of SAN DIEGO

- USA San Diego outdoor parking management (Caltrans)

Milton Keynes City Council

- UK Milton Keynes stadium outdoor parking management

Challenges in ITS

- Ill-posed problem: How can we define “road debris”?
- Contextual problem: How can we define “accidents” with object detection?
- Rare dataset: How can we obtain dataset?

Ill-posed Problem: Road Debris

Can it be detected by legacy AI models?

Contextual Problem: Accident Detection

Can it be detected by legacy AI models?

Nota AI

Challenges in ITS

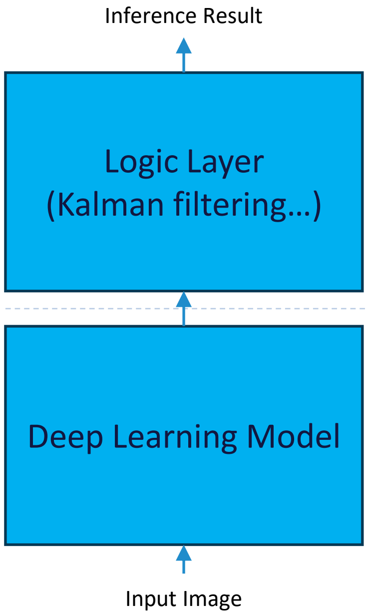
- Ill-posed problem: How can we define "Road debris"?
- Contextual problem: How can we define "Accidents" with Object detection?
- Rare dataset: How can we obtain dataset?

Requires super-generalized model: Foundation Model

Foundation Model on the Edge: Challenges in Industrial AI

embedded VISION SUMMIT

As-is | **Problem**



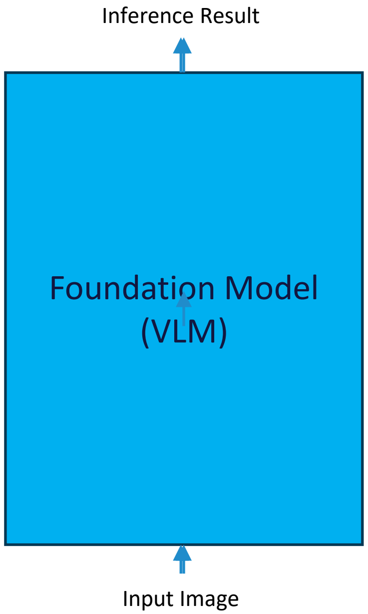
- Rule-based algorithm is fragile.
- Logic added for new requirements.
- Errors on the object detection/tracking propagate to logic layers.
- On-site calibration.

- Data drift.
- Data acquisition on rare events is hard.
- Sophisticated model composition.
- On-site calibration.

Foundation Model on the Edge: Challenges in Industrial AI

embedded VISION SUMMIT

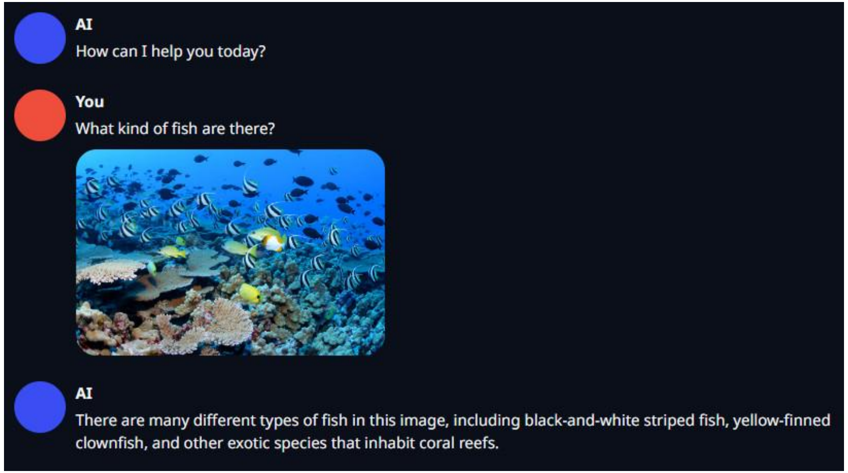
To-be	Features



- VLM is capable of comprehending complex scene.
- VLM already contains various logic.
- VLM is robust on data drift.
- VLM is aware of rare events.
- VLM needs less or no calibration.
- Still, VLM is not understanding video.

Vision Language Model: LLaVA

LLaVA



Live LLaVA

Source: jetson-ai-lab.com

Working Prototype of LLaVA on Accidents Detection

Benchmark on Models

Llava-13B (Jetson AGX Orin)	Quantization	Tokens/sec	Memory
`text-generation-webui`	4-bit (GPTQ)	2.3	9.7 GB
`llava.serve.cli`	FP16 (None)	4.2	27.7 GB
`llama.cpp`	4-bit (Q4_K)	10.1	9.2 GB
`local_llm`	4-bit (MLC)	21.1	8.7 GB

Source: jetson-ai-lab.com

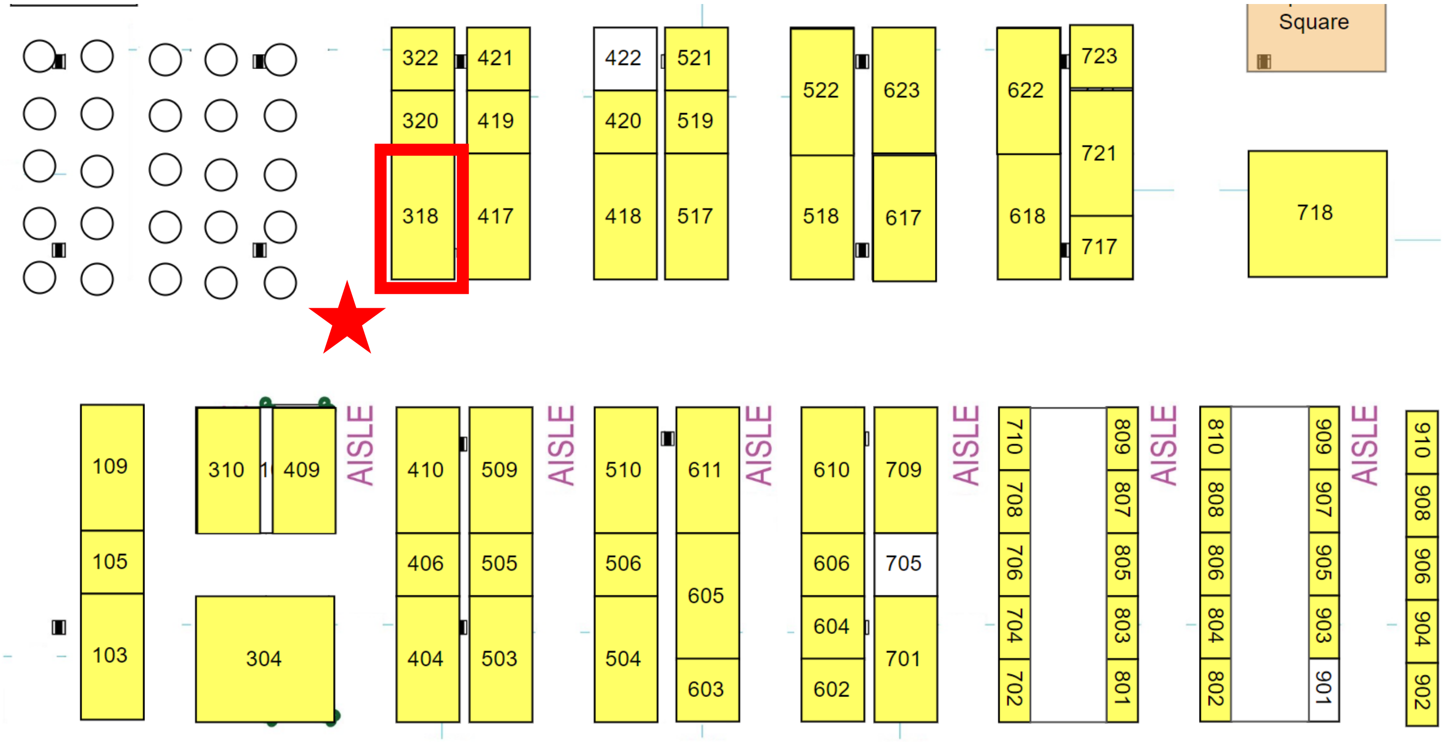
Future Work

- More advanced optimization required
- VLM needs to comprehend temporal consistency
- Domain adaptation might be required for user specific scenario
- Interface for product required
- Prompt engineering is required for higher performance

Conclusion

- Industrial AI is already a widely used technology, but technology is limited when the problem is complex and underdetermined.
- Among GenAI models, a Vision Language Model (VLM) can understand complex scenes, so it could analyze complex queries and events.
- For example, in ITS, road debris and car accidents are severe problems, but this couldn’t be solved by legacy AI models.
- Using VLMs, these problems can be solved as well.
- However, VLMs are still computationally expensive, so lightweight VLMs are required for the next step.

Visit Nota AI @318 !



Nota AI

Thank you for your attention!