Using NVIDIA Isaac Sim for AGV Data Collection & Algorithm Development

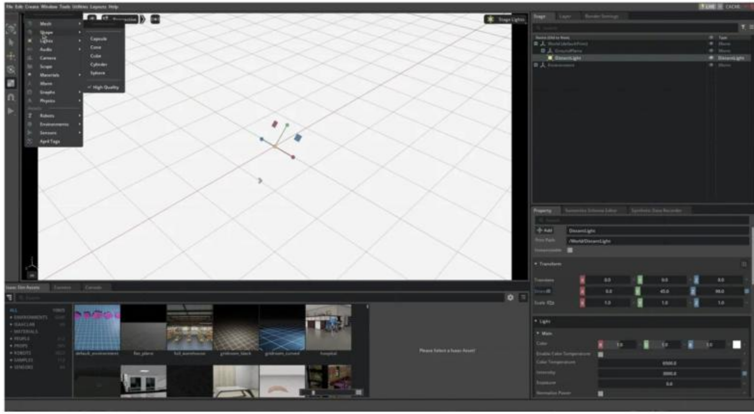
Why We Need a New Data Collection Method

- Access limitations: OPC-only
- Safety risk: cannot stage near-collision events on the real AGV
- Time consuming
- Low data diversity: static routes and environments → poor edge-case coverage

What is NVIDIA Isaac Sim?

- High-fidelity warehouse/factory simulation on Omniverse (RTX)
- Sensors: LiDAR, RGB, RGB-D,

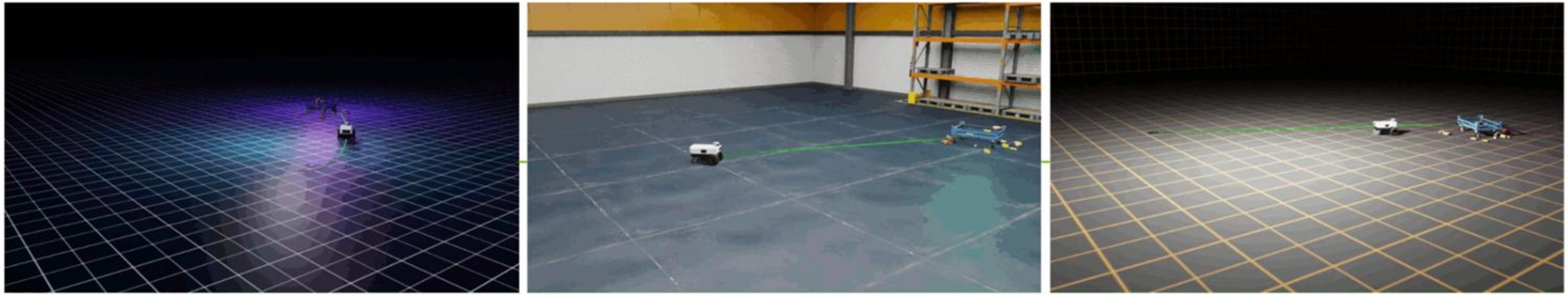
Source	frameId	nodeNamespace	topicName	type
Camera RGB	(device_name)_(data_type)	(device_name)/(data_type)	image_raw	rgb
Camera Depth	(device_name)_(data_type)	(device_name)/(data_type)	image_rect_raw	depth
Lidar	base_scan		scan	laser sca
Lidar	base_scan		point_cloud	point_cl
TF			tf	tf



- Drag-and-drop GUI to start; Python for automation at scale
- Widely used in industry and academia

4) What Data Can Isaac Sim Generate?

- Multiple scenes of collision event



The scenario uses the left and right camera sensors of Nova Carter

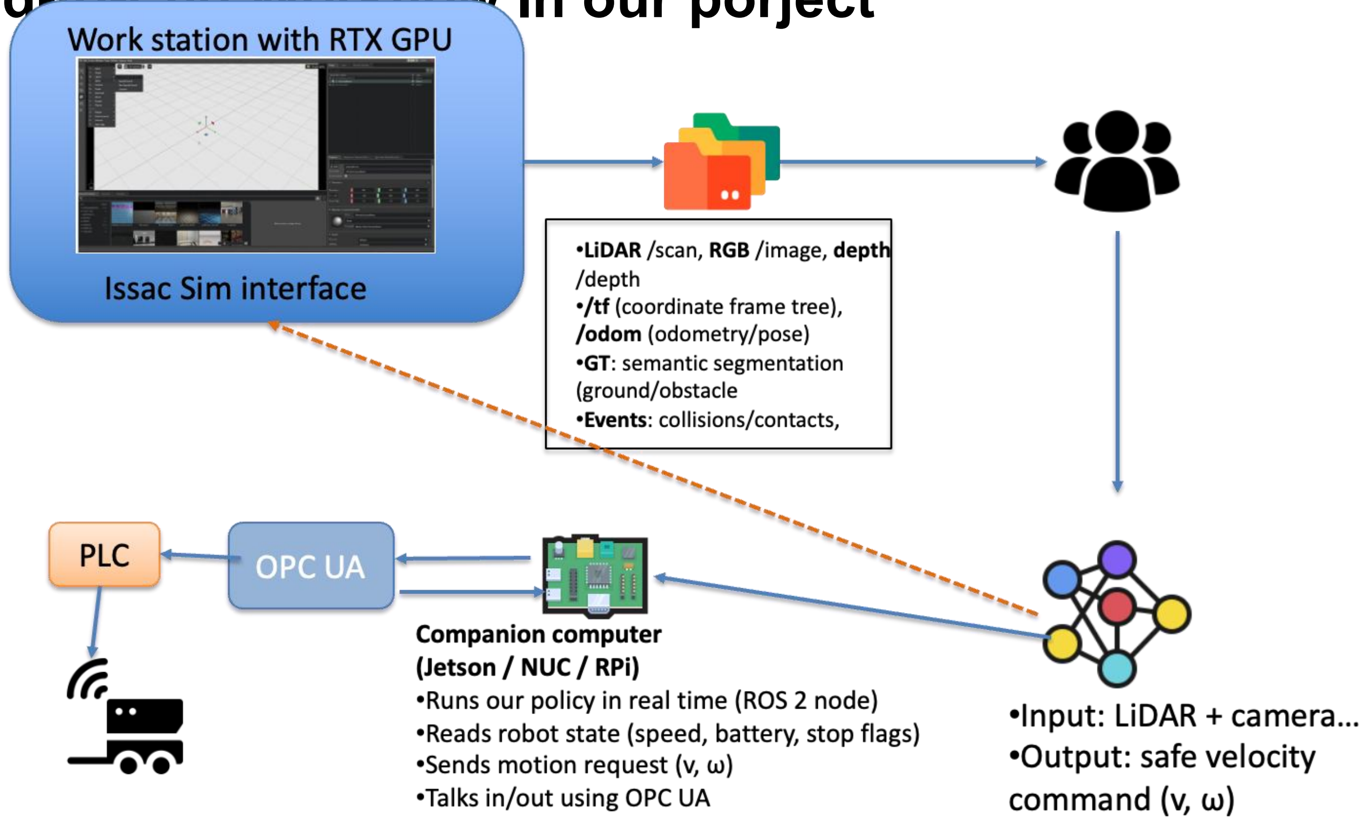
5) Advaced Function:Replicator / SDG (Synthetic Data Generation)

- Automatic labels:

Normals	Per-pixel surface orientation vectors (x,y,z).
Poses	Exact position/orientation for robot & objects.
Distances / contacts	Exact robot-to-obstacle distances; contact/ collision events with who/when.
Trajectories / timestamps	Robot/object paths over time; precise times for every frame.

- **Domain randomization**: lighting, textures/materials, clutter, motion
- Stress hard cases: reflective shelves, dark boxes, glass/transparent surfaces, low light
- Scriptable dynamic actors (pedestrians, forklifts) to generate varied scenarios
- Export COCO/KITTI or custom; configs in YAML/JSON for reproducibility

End-to-End Workflow In our porject



First meeting

Define

1. Research problem :
Camilo-> same as PW, VLA->lightweight (VLA as baseline)
Lynn-> how to enable VLA to incorporate not only camera data but also other modality

Thomas->try baseline first, keep flexible in the future.

2. Potential contribution:

Camilo->defining a best output space for AGV control using VLA

Lynn->lightwieght machine model

3.Required data:
Camilo->Lidar, camera,
Lynn-> Lidar, camera, language (NLP)

Donato-> AGV kinetic data, all potential useful data.

4.To do:
DO suvery on VLA controing AGV!!!

Objective: Collision avoidance, lightweight

Camilo:
LLMs for the recommendation of clinical decisions
->
keywords: LLM, VLM

Donato:
Deep Clustering for Time Series
->
keywords: Time series prediction, explainity AI

PW:
sensor fusion:

keyword

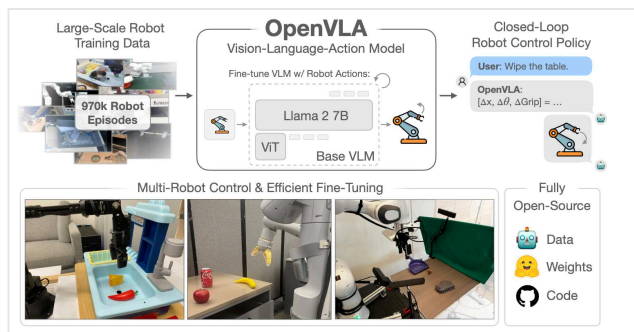
Lyn:
keyworldL generative AI, Federated learning

Tomasz:
Software development, security.

Team3 proposal:
Vision language Action (VLA) model distilled Light-weight semantic RL agent for AGV collision

Motivation:

Reinforcement Learning (RL) has long been a dominant paradigm for decision-making and control in robotics and autonomous systems. However, its reliance on task-specific reward signals and limited semantic understanding often constrains its generalization to new environments. In contrast, recent advances in **Vision–Language–Action (VLA)** models present a promising alternative. By leveraging large-scale multimodal pretraining, VLA models inherently possess a **generalized understanding of real-world event semantics**, allowing them to interpret scenes, infer intent, and adapt actions across diverse contexts.



Research Problem:

While Vision–Language–Action (VLA) models demonstrate impressive generalization and semantic understanding, their **token-based sequential prediction** nature introduces significant computational overhead during inference and training. This high cost limits their deployment in real-time robotic applications, such as dynamic collision avoidance, where low-latency decision-making is critical.

Objective

The challenge lies in **distilling or adapting VLA representations into more efficient control policies**

Research Problem:

Current VLA is majorly train on Robotic control, wheather current openVLA can adapt to AGV control is remain unknown

Objective

Test what current VLA can achieve in AGV avoidance guidence

Research Problem:

Current VLA is majorly is train on RGB image, whether incoperate additional sensor data would be beneficial to automatic guidance?

Objective

Exploring fusion technique to extend VLA of different modality feature