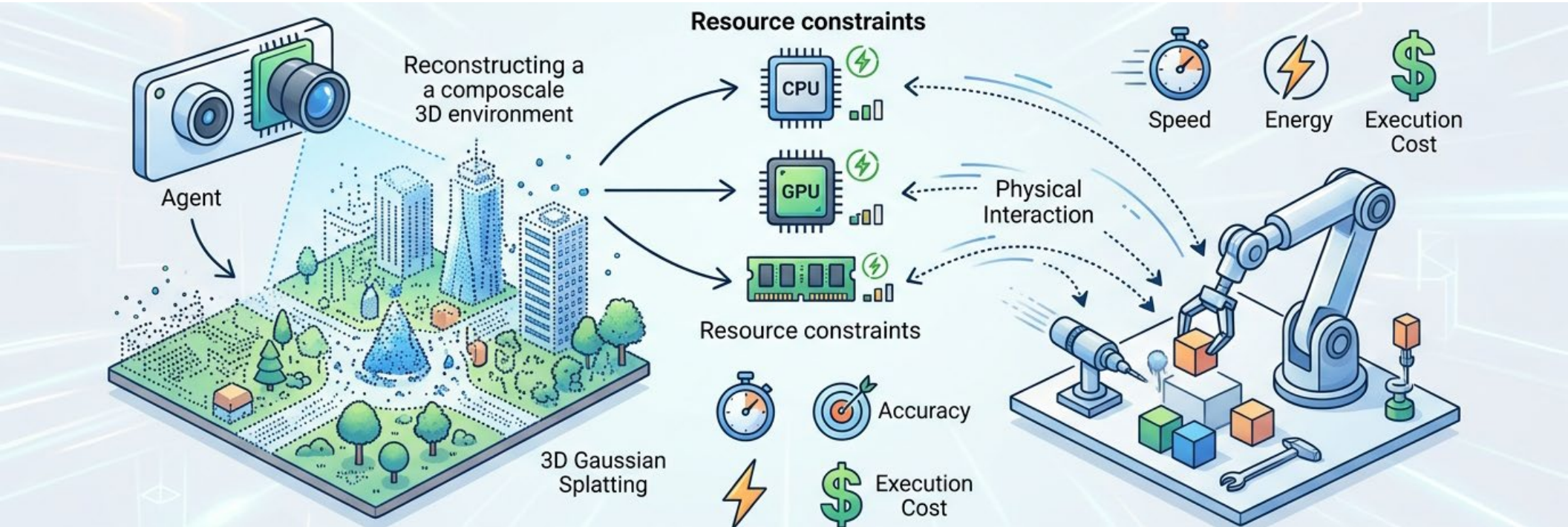
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Efficient Spatial and Embodied Intelligence under Resource Constraints

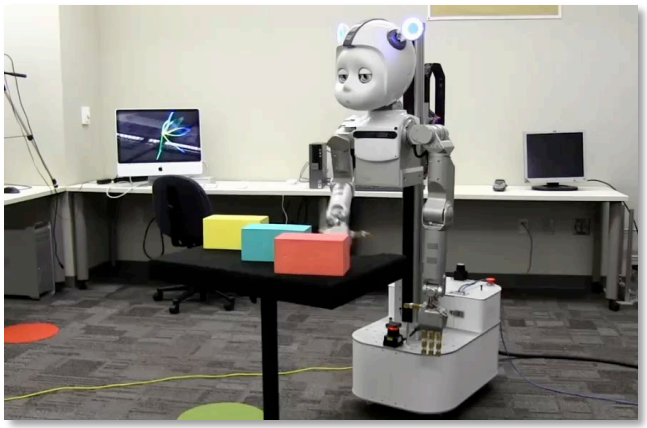
Chaojian Li

Assistant Professor at HKUST

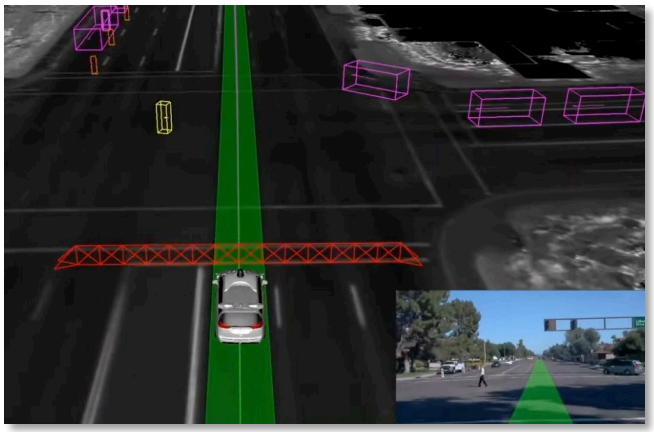


Ubiquitous Spatial & Embodied Intelligence

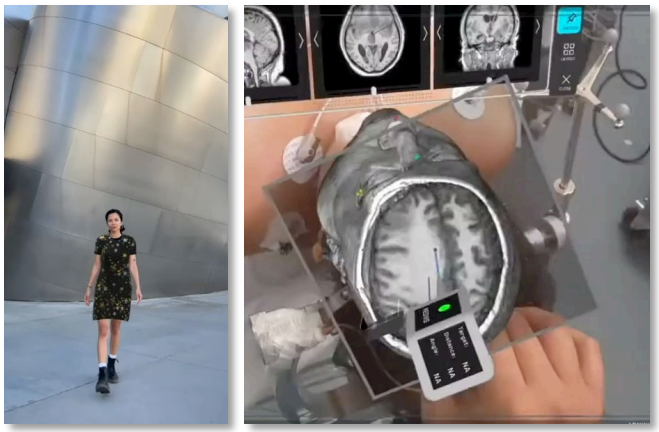
- Every **application**, on every **device**, all at **once**



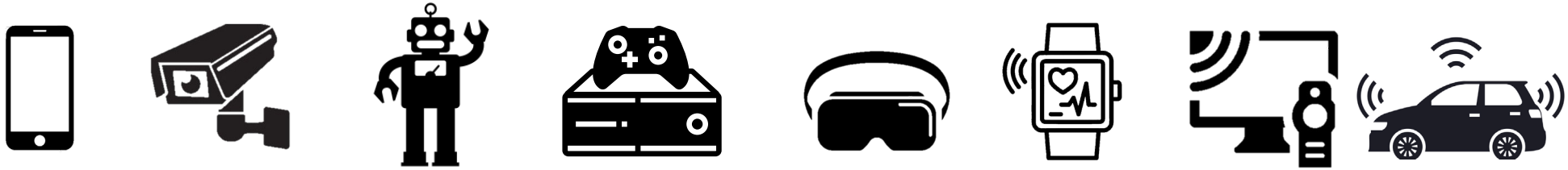
Embodied AI



Autonomous Driving



AR/VR



Efficient Deployment: Not Yet Possible

- **Gap** exists between vision and reality

Embodied AI

Autonomous Driving

AR/VR

Efficient Deployment: Not Yet Possible

- **Gap** exists between vision and reality

Embodied AI

Autonomous Driving

AR/VR

Constrained by Power Consumption

Efficient Deployment: Not Yet Possible

- **Gap** exists between vision and reality

Embodied AI

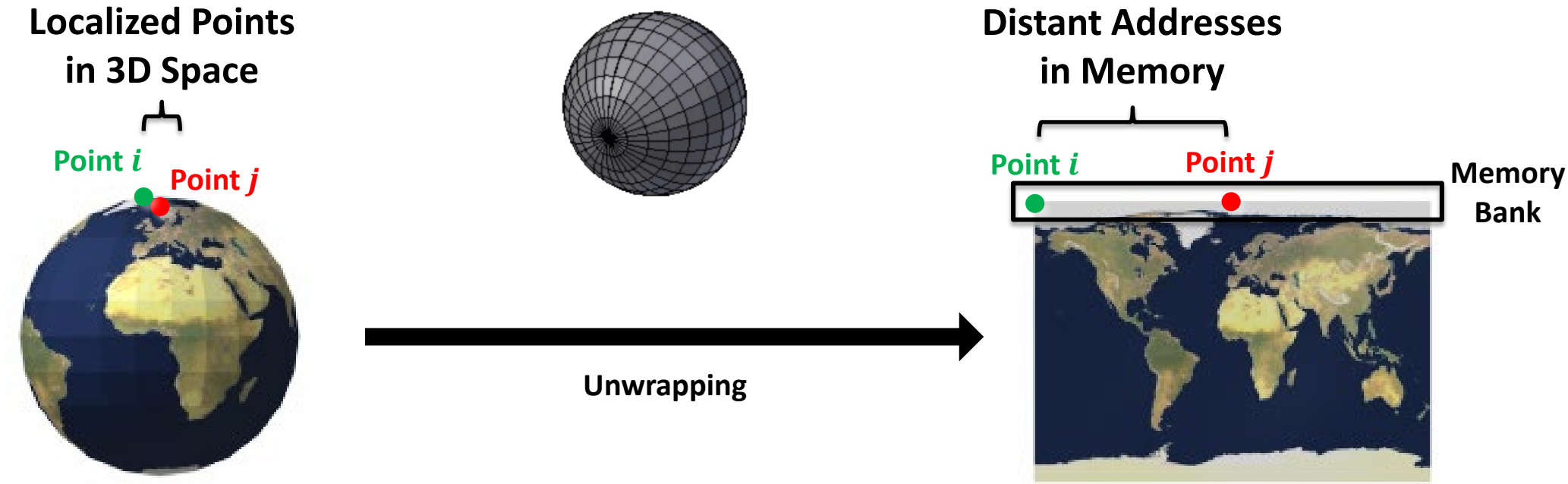
Autonomous Driving

AR/VR

Constrained by
Area/Form Factor Size

Our Work on Tackling Memory Efficiency

- Memory allocation plays a more critical role in spatial & embodied intelligence than in image/text applications
  - **Irregular & Massive**



Example Application: 3D Reconstruction

- We target reconstructing city-scale scenes

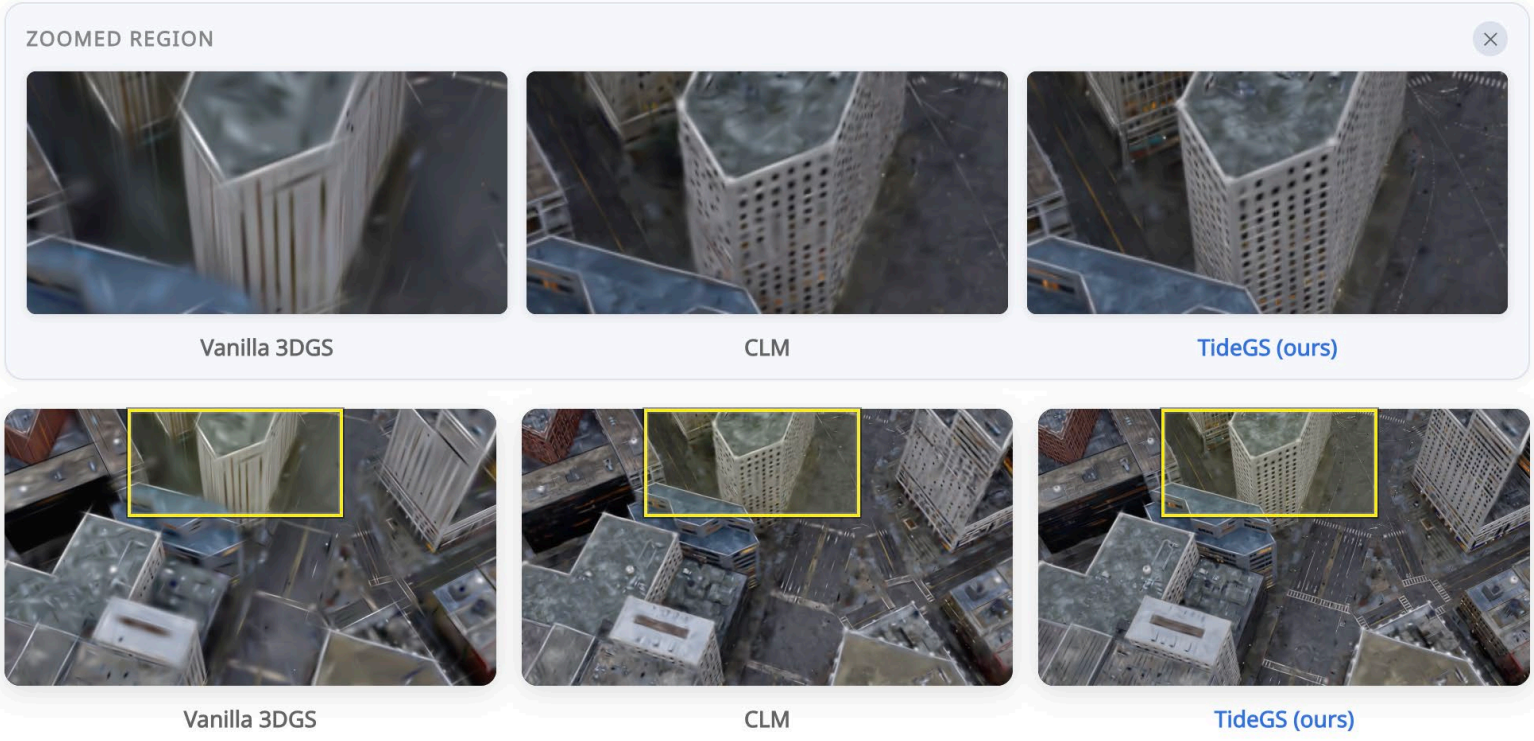
Our Solution: TideGS

- Billion-primitive 3D Gaussian Splatting on a single consumer GPU (RTX A5000 24GB)

TideGS: Results

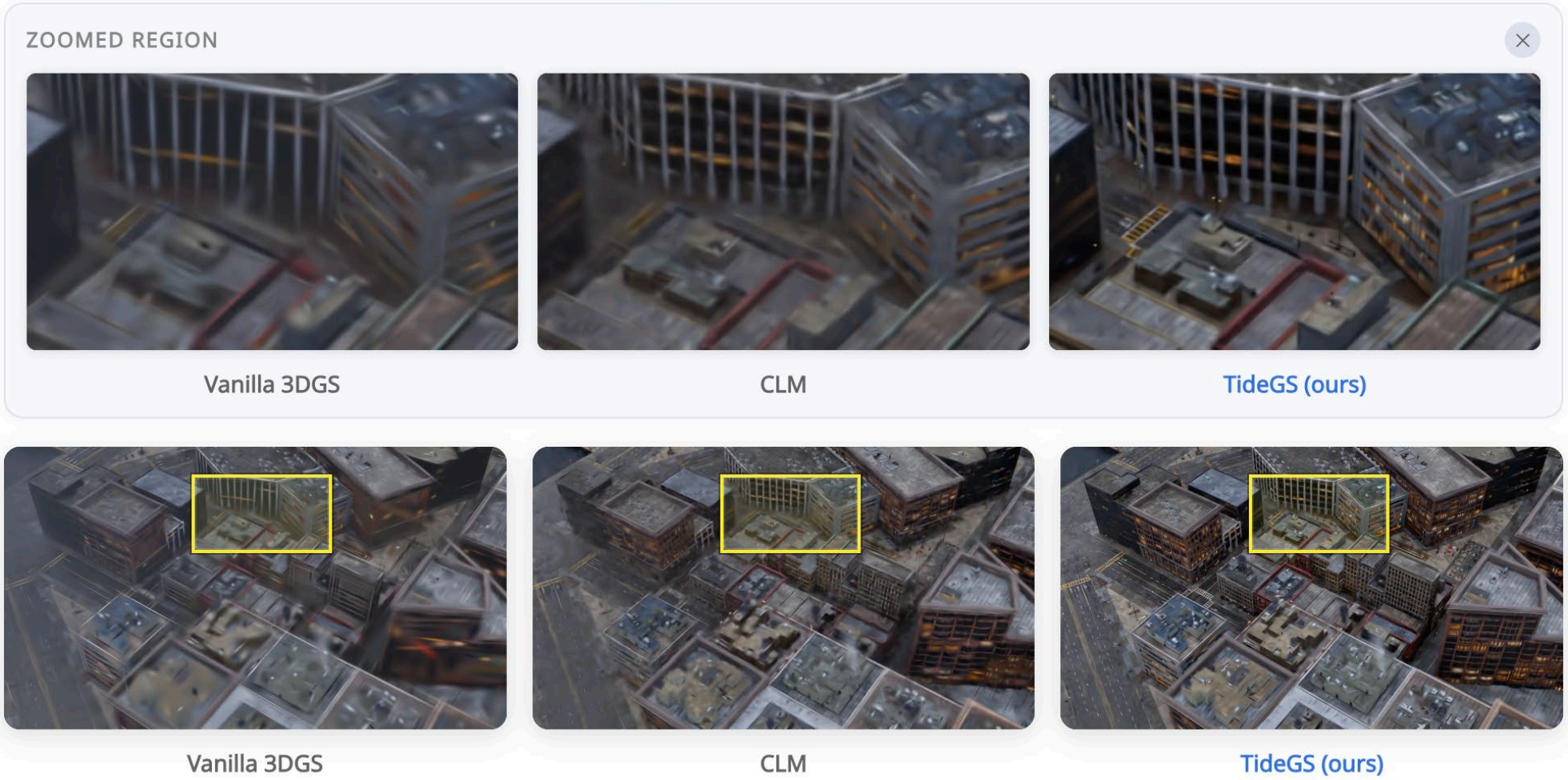
- Higher GPU utilization enables larger scenes, leading to better visual quality

Method	Scale (N)	Backing Store	PCIe (GB/it)↓	GPU Util.(%)↑	Iter (ms)↓
Naive Offload	∼102 M	System RAM	— *Out of Memory (OOM)* —		
CLM (Zhao et al., 2025)	∼102 M	System RAM	0.41	37.0	100.8
TideGS (Ours)	∼102 M	NVMe SSD	**0.10**	43.3	**90.7**
CLM (Zhao et al., 2025)	∼1.1 B	System RAM	— *Out of Memory (OOM)* —		
TideGS (Ours)	∼1.1 B	NVMe SSD	0.97	**49.5**	525.6



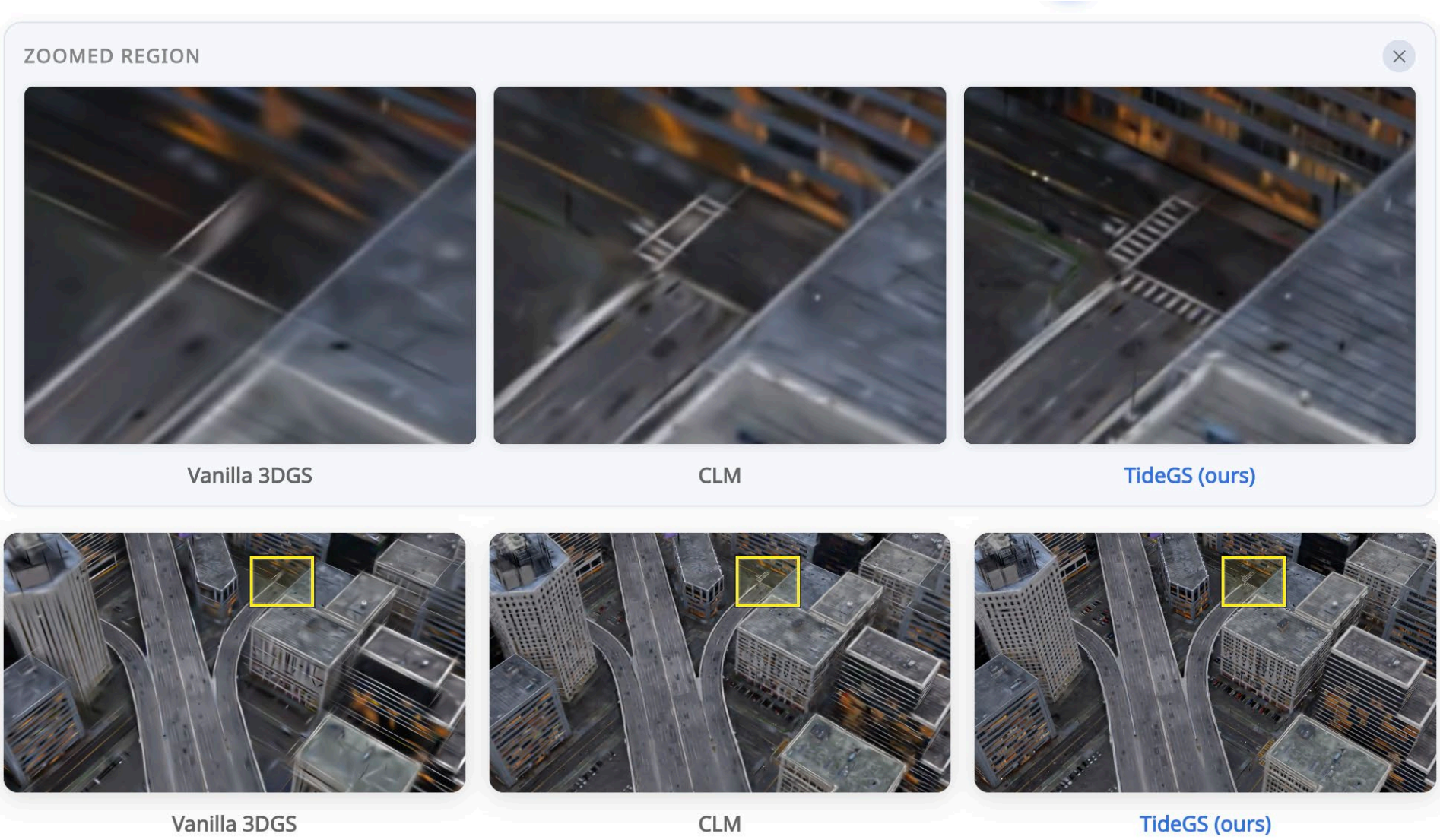
TideGS: More Visualization

- Sharper window details on buildings



TideGS: More Visualization

- Clearer road view



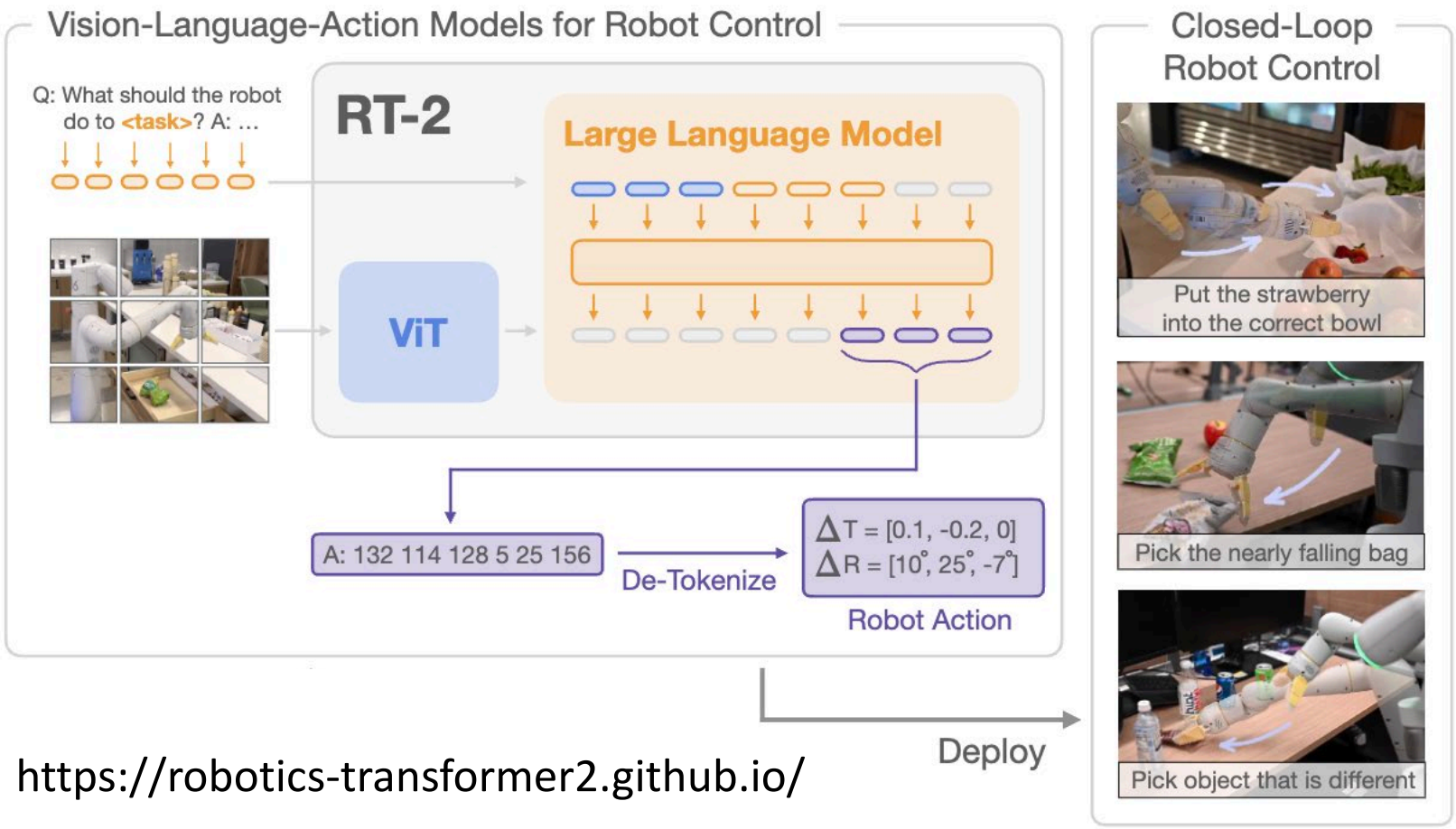
Our Work on Tackling Energy Efficiency

- Energy efficiency is one of the key aspects for scalable and sustainable embodied intelligence in elder care

Our Target

Example Application: VLA Model

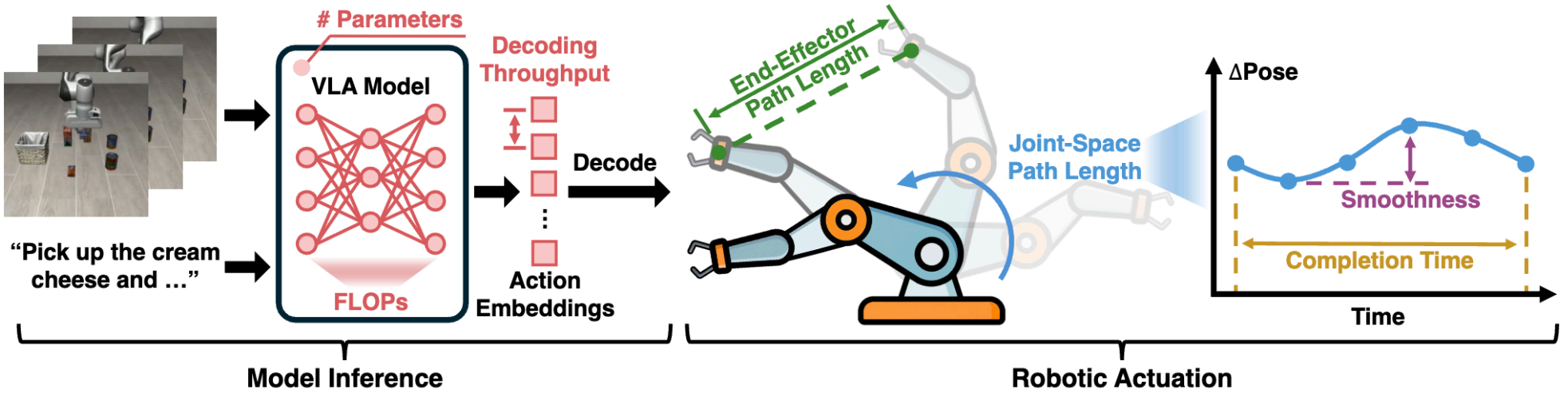
- Vision-Language-Action (VLA) models represent robot actions as another language



https://robotics-transformer2.github.io/

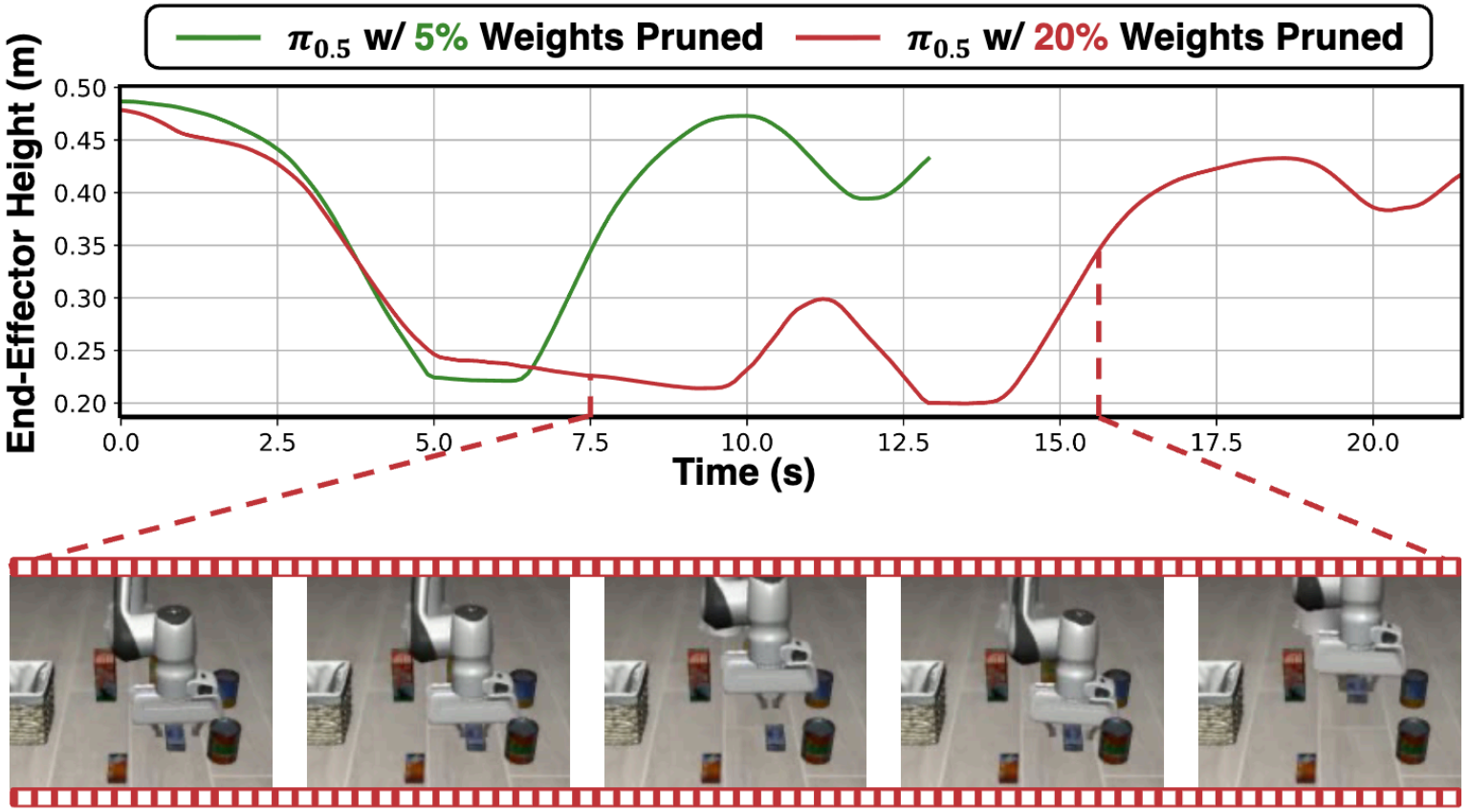
Our Solution: Revisiting Efficiency Metrics

- Model Inference Efficiency:
  - FLOPs, # Params, Decoding throughput, …
- Robotic Actuation Efficiency:
  - End-effector path length, smoothness, …



Our New Observations

- Lighter model inference → lower actuation efficiency



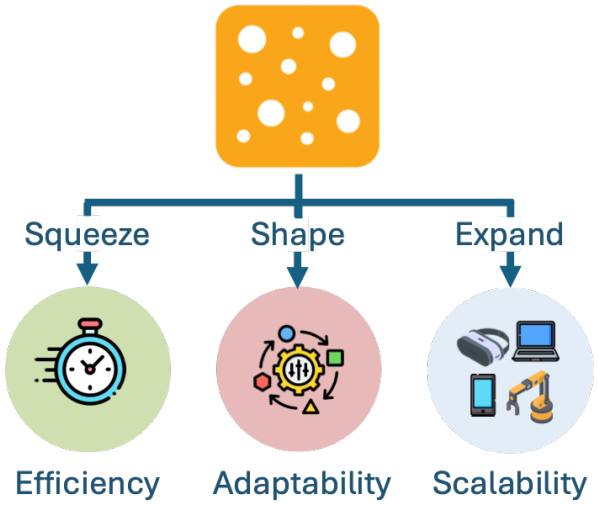
VLA Model	$\pi_{0.5}$ w/ 5% Weights Pruned	$\pi_{0.5}$ w/ 20% Weights Pruned
Success Rate	100%	100%
# Parameters	0.95×	0.80×
Completion Time	13 s	21 s

Summary

- Both memory and energy consumption are key factors in spatial and embodied intelligence applications

- Optimization should be considered from a system- and architecture-level perspective

Lab Website



chaojian@ust.hk

Thanks! We welcome collaboration!