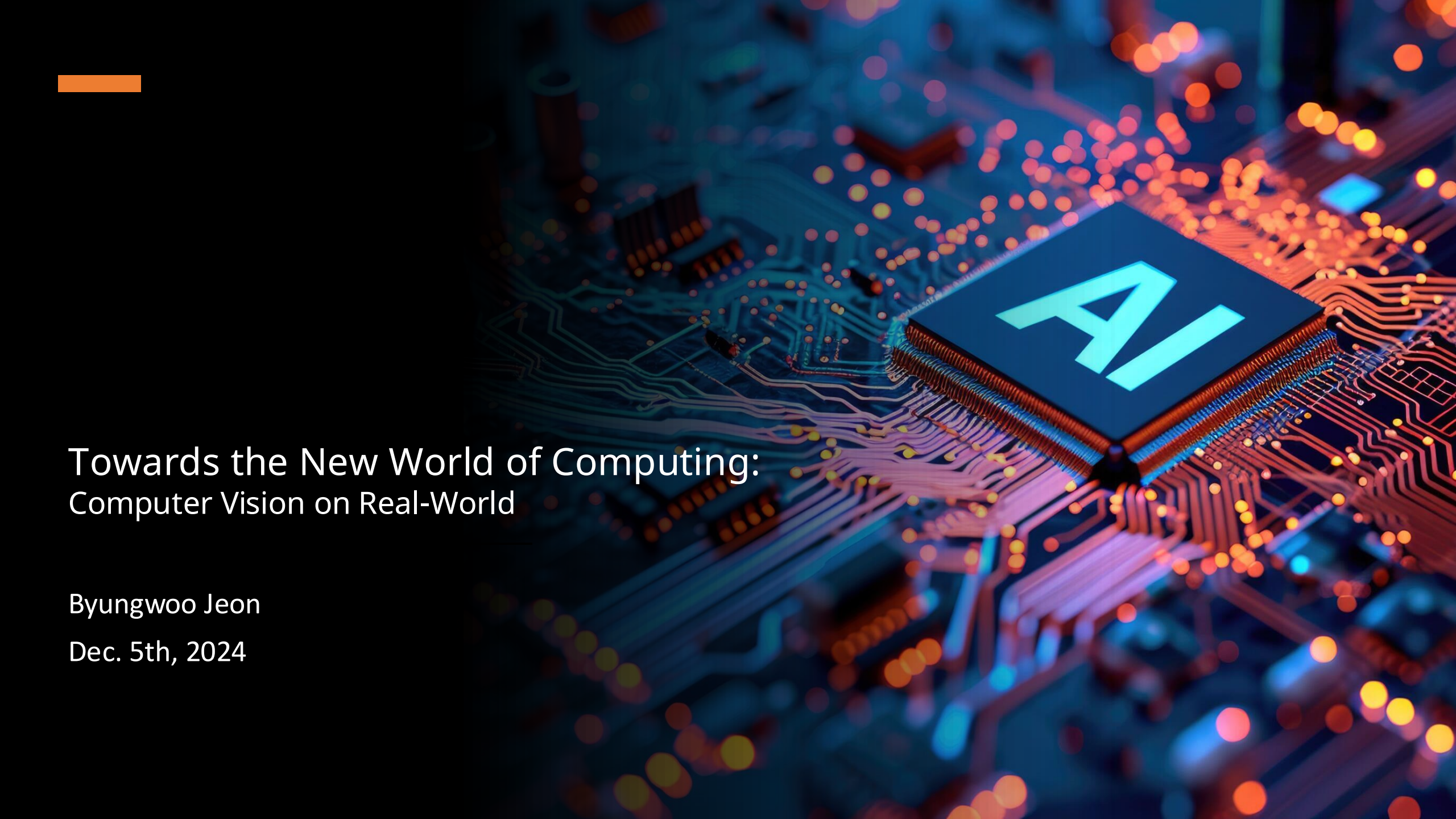


Towards the New World of Computing:
Computer Vision on Real-World

Byungwoo Jeon

Dec. 5th, 2024

Hello, World 👋

Byungwoo Jeon

- B.S. in CSE, **Korea Univ.**
- M.S/Ph.D student in AI, **KAIST**
  - ALIN-LAB advised by Dr. Jinwoo Shin
- **GDG KU** Founder & 1st Lead (22.08 ~ 23.07)

Summary of Today's Talk

Agenda

- Introduction
  - What is computer vision?
  - Historical background of computer vision
- Applications on real-world scenarios
  - 3D Vision
  - Robotics
  - Autonomous Driving
- Discussions & Summary

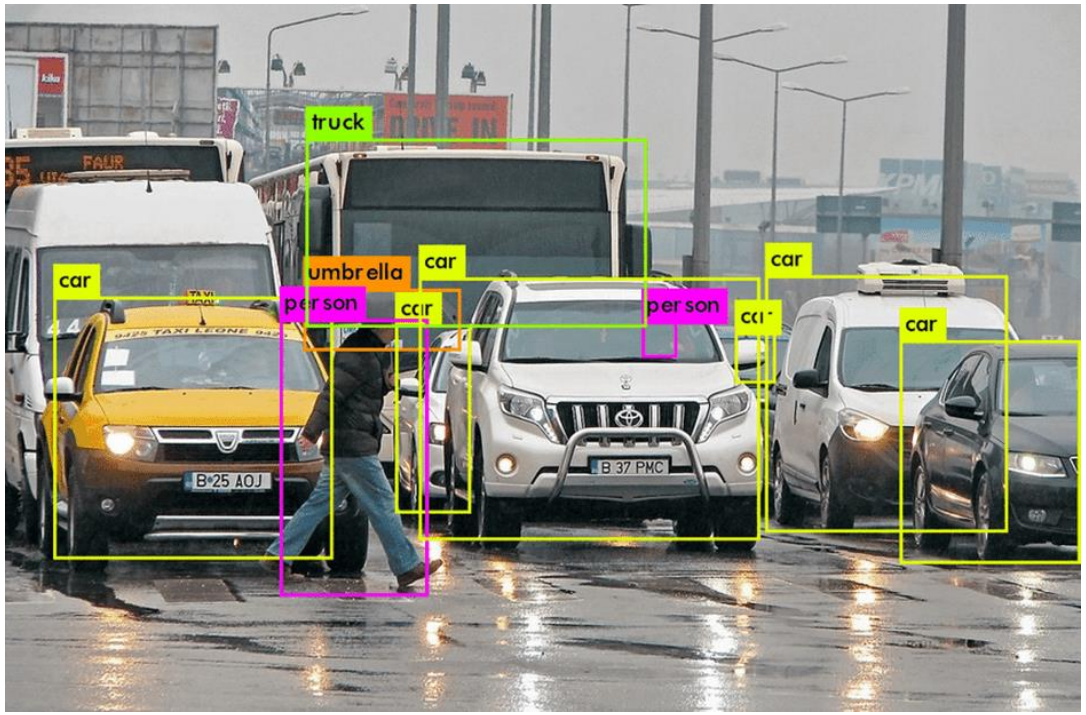
Introduction

🤔 What is Computer Vision?

Enables machines to **interpret** and **understand** visual data (*e.g.*, images, videos) from the world

Introduction

Examples of Computer Vision



Object Detection

Segmentation

Introduction

Examples of Computer Vision

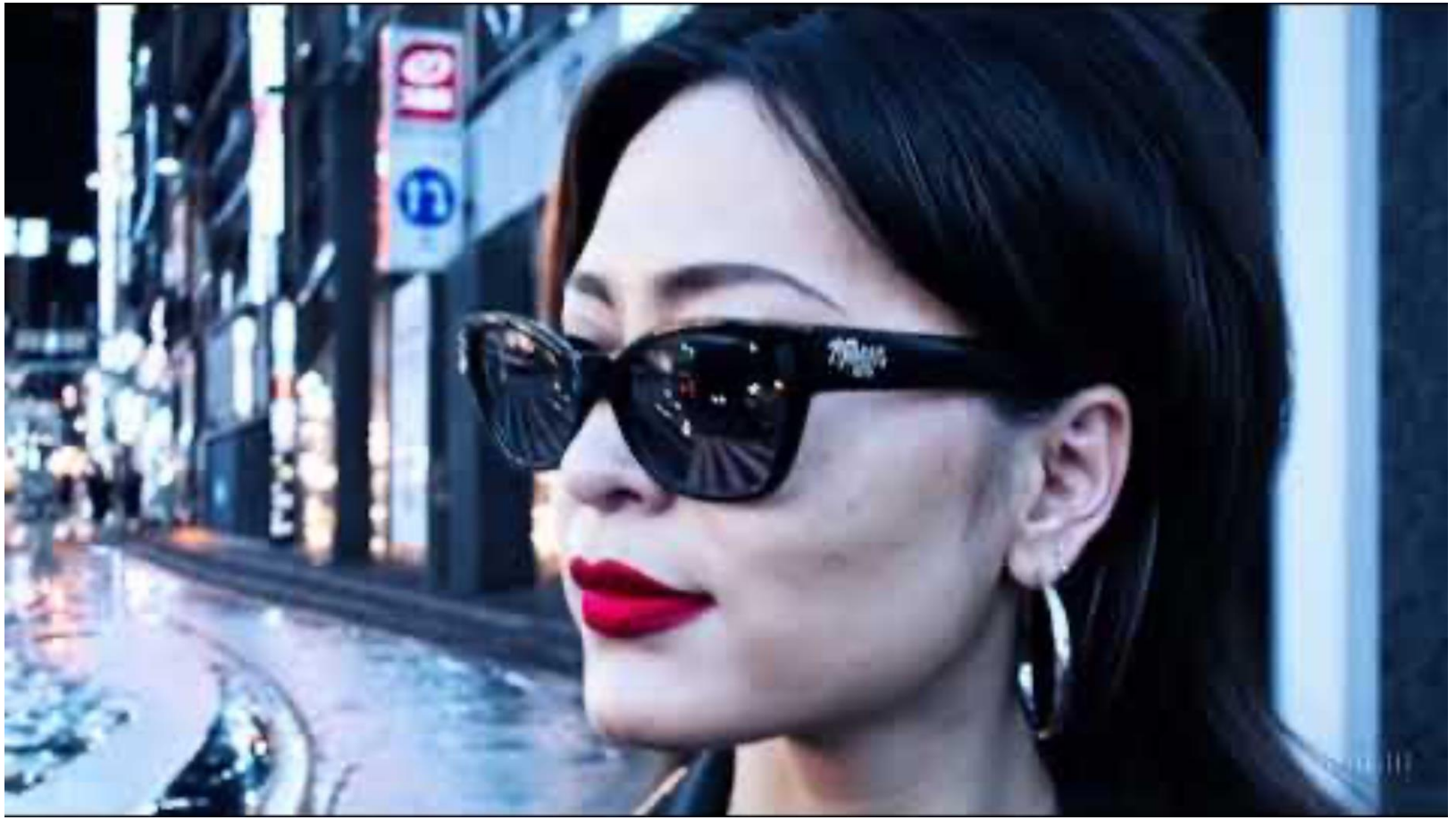
Image Matching

Prompt: ~*~aesthetic~*~ #boho #fashion, full-body 30-something woman laying on microfloral grass, candid pose, overlay reads Stable Diffusion 3.5, cheerful cursive typography font.

Image/Video Generation

Introduction

Examples of Computer Vision



Video Generation by OpenAI, SORA

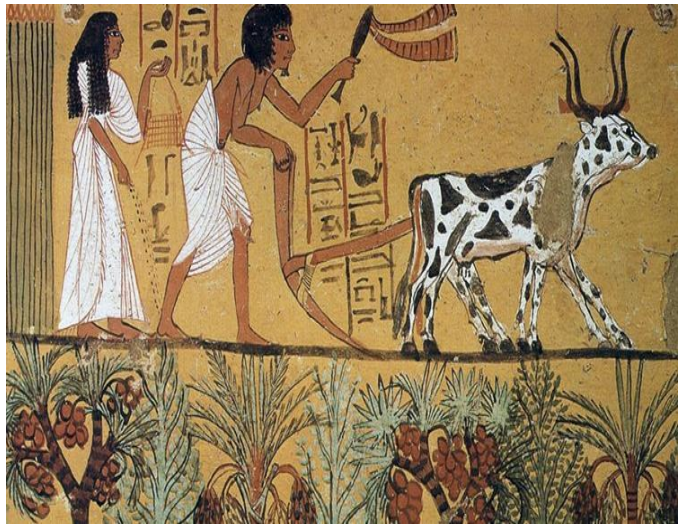
Introduction

Fundamental Question

🤔 Nothing special other than replacing human labor?

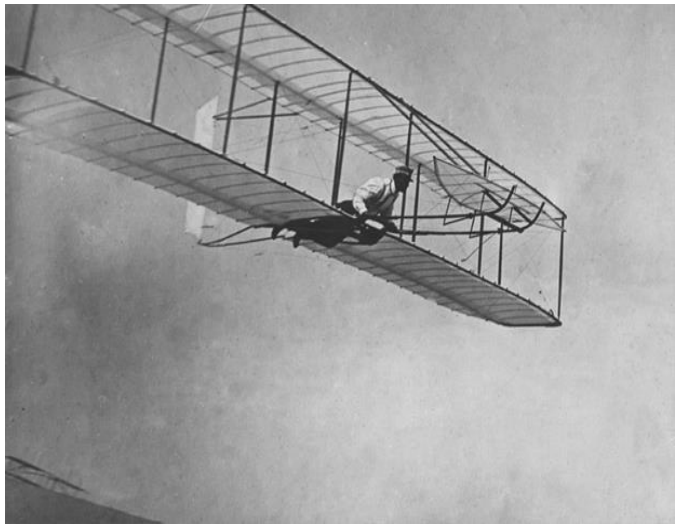
Introduction

History of Human



Land (B.C. 12C ~)

Sea (14C ~)



Air (1903 ~)

Introduction

History of Human

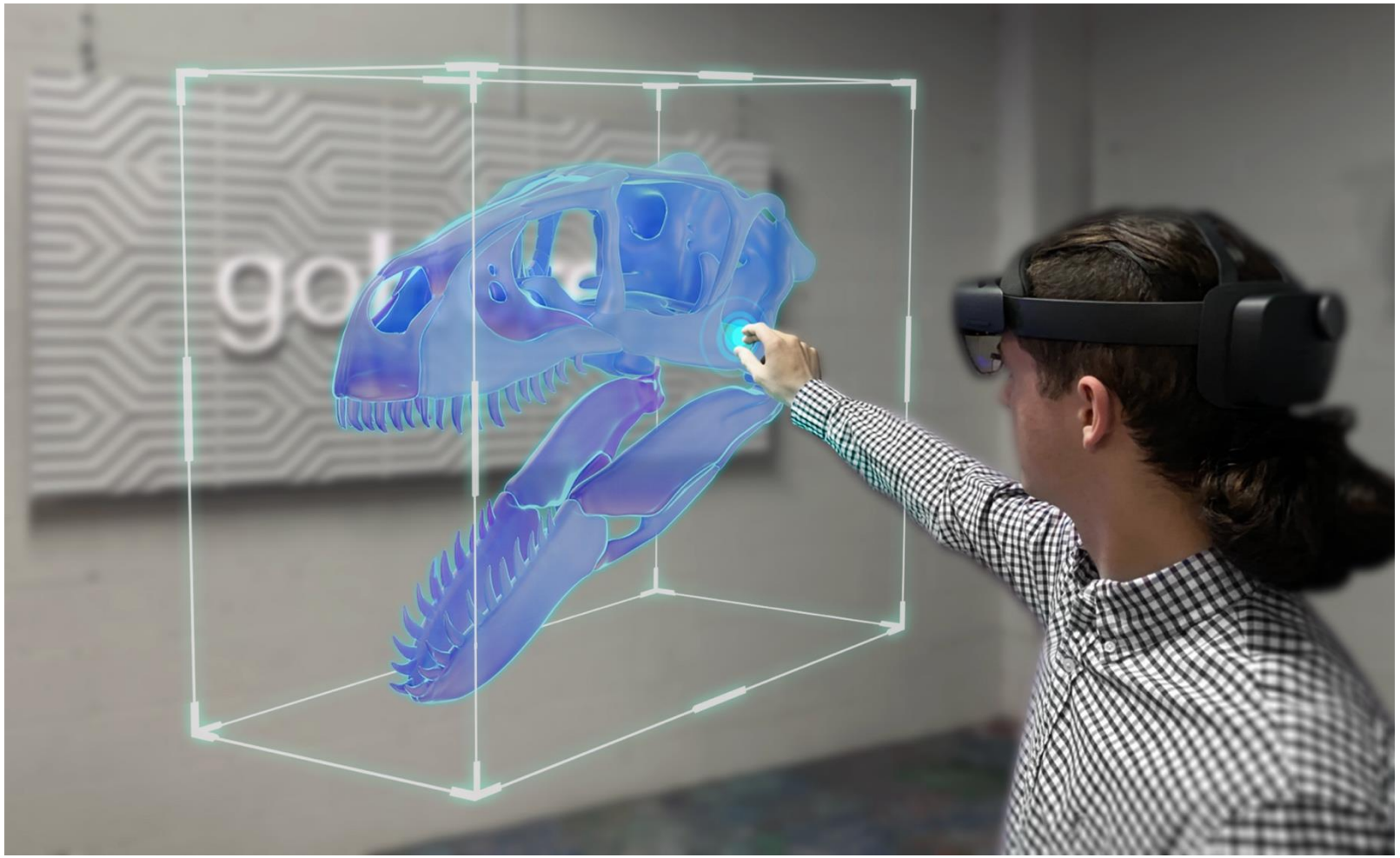
Space (1961~)

Digital (2004 ~)

What's next?

Introduction

History of Human

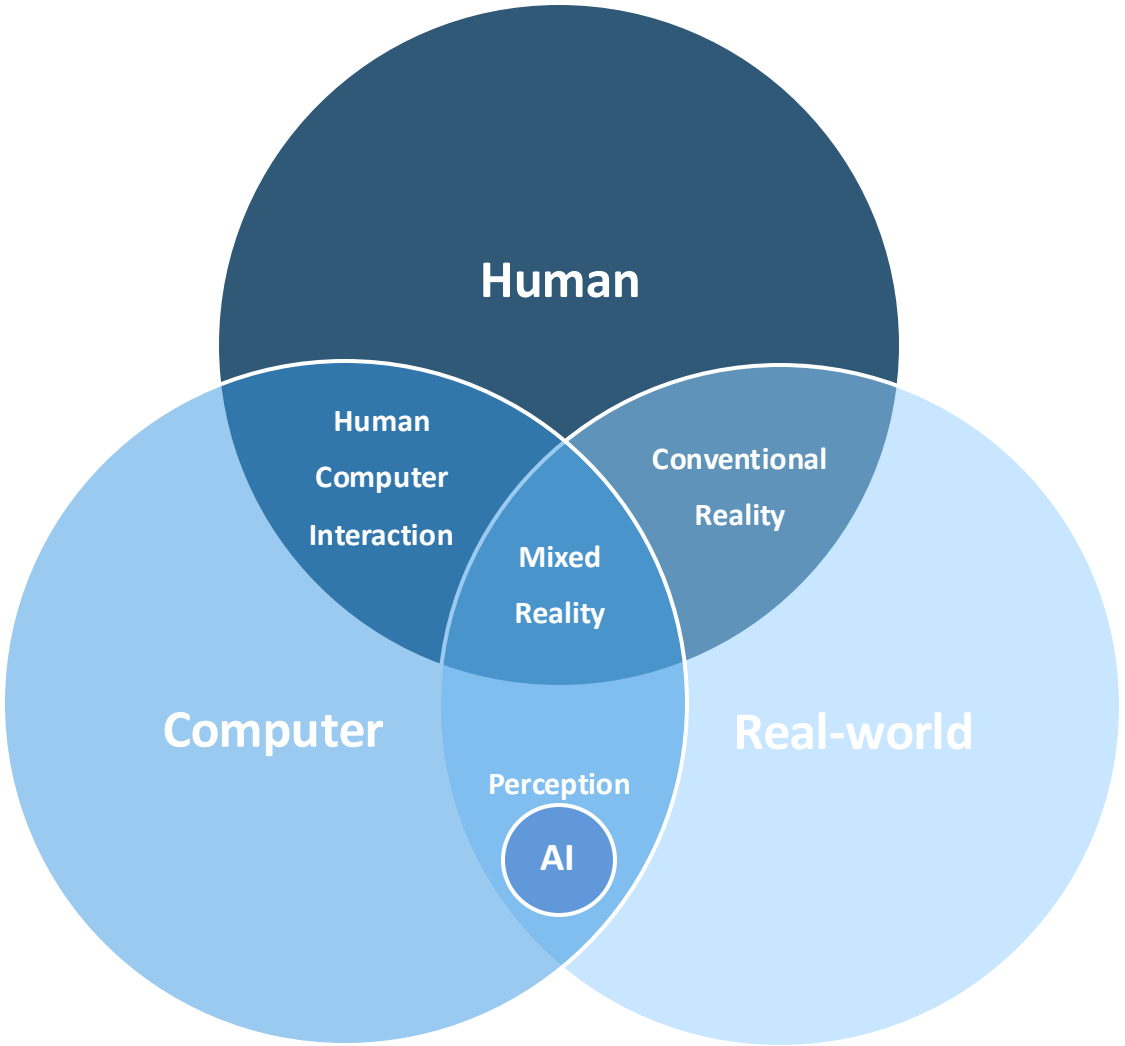


What’s next?

Digital + Real-World = **Mixed Reality**

Introduction

Relationships



Introduction

What happens in AI in last 3 years

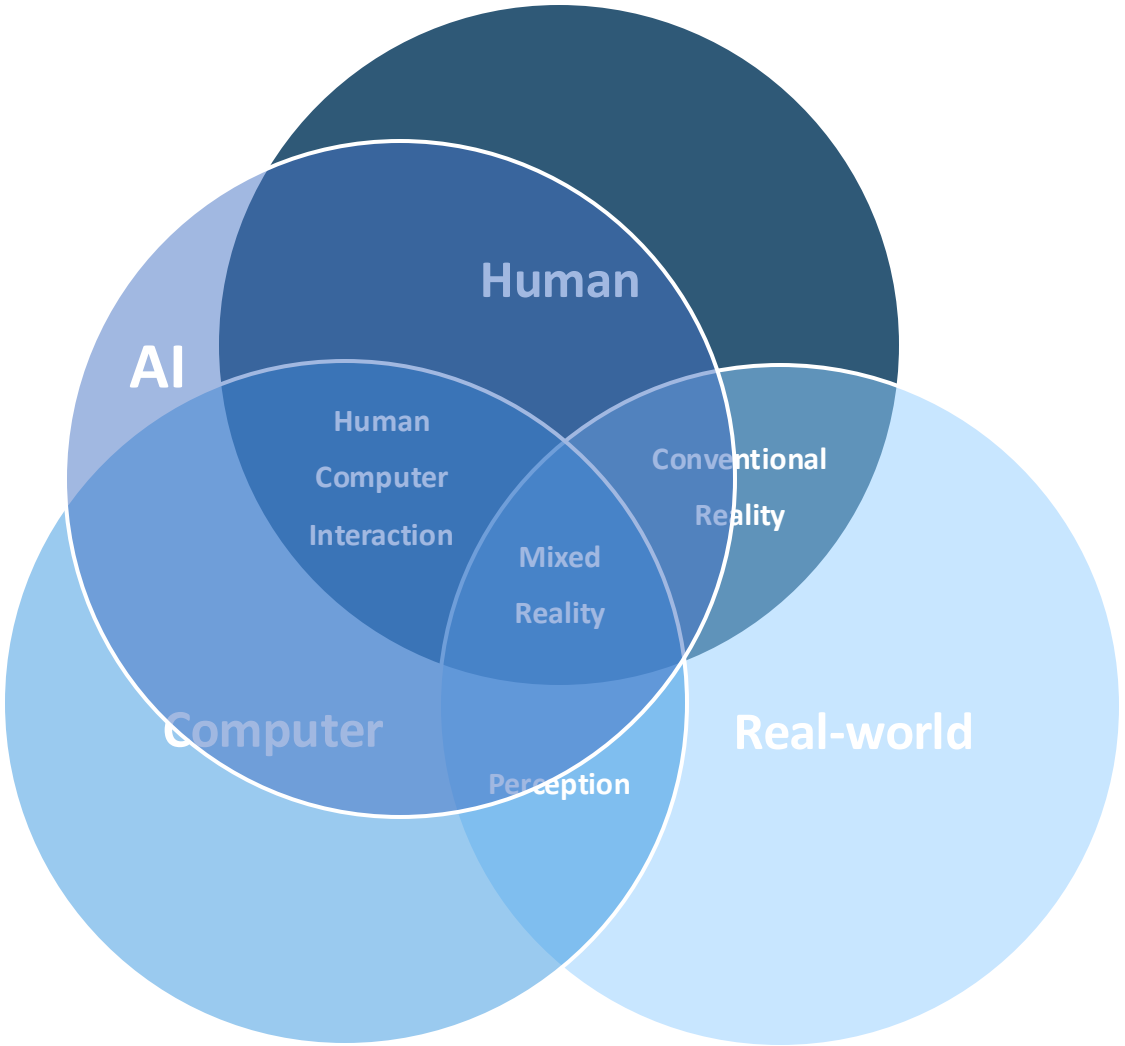
- Emergence of **LLMs** (*e.g.*, ChatGPT, Gemini)
  - We can communicate with AI
  - Multi-modalities
- Emergence of **Foundation models** (*e.g.*, Diffusion models, Segment Anything)
  - Generalized well in diverse scenarios

ChatGPT, OpenAI

Gemini, Google

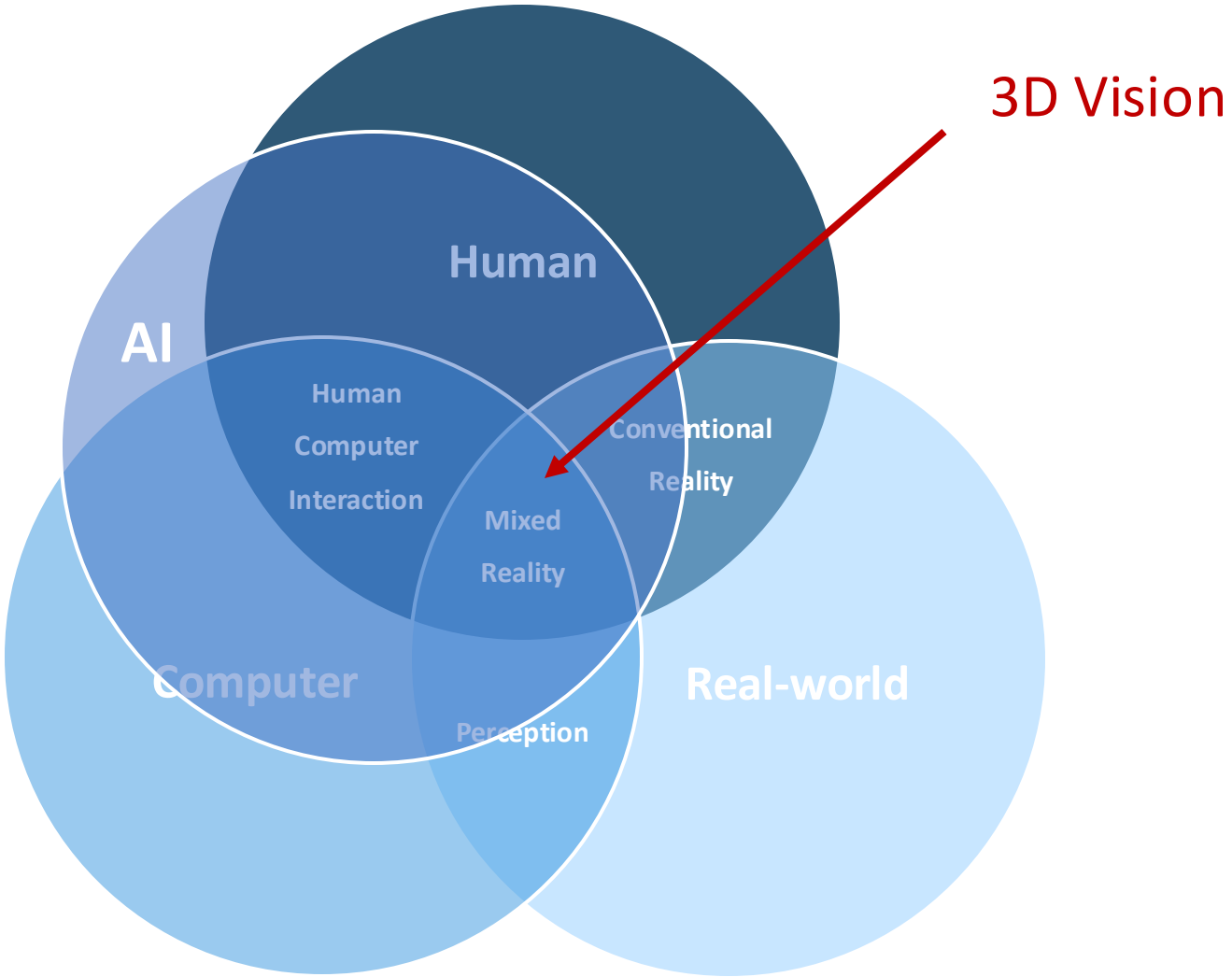
Introduction

Relationships



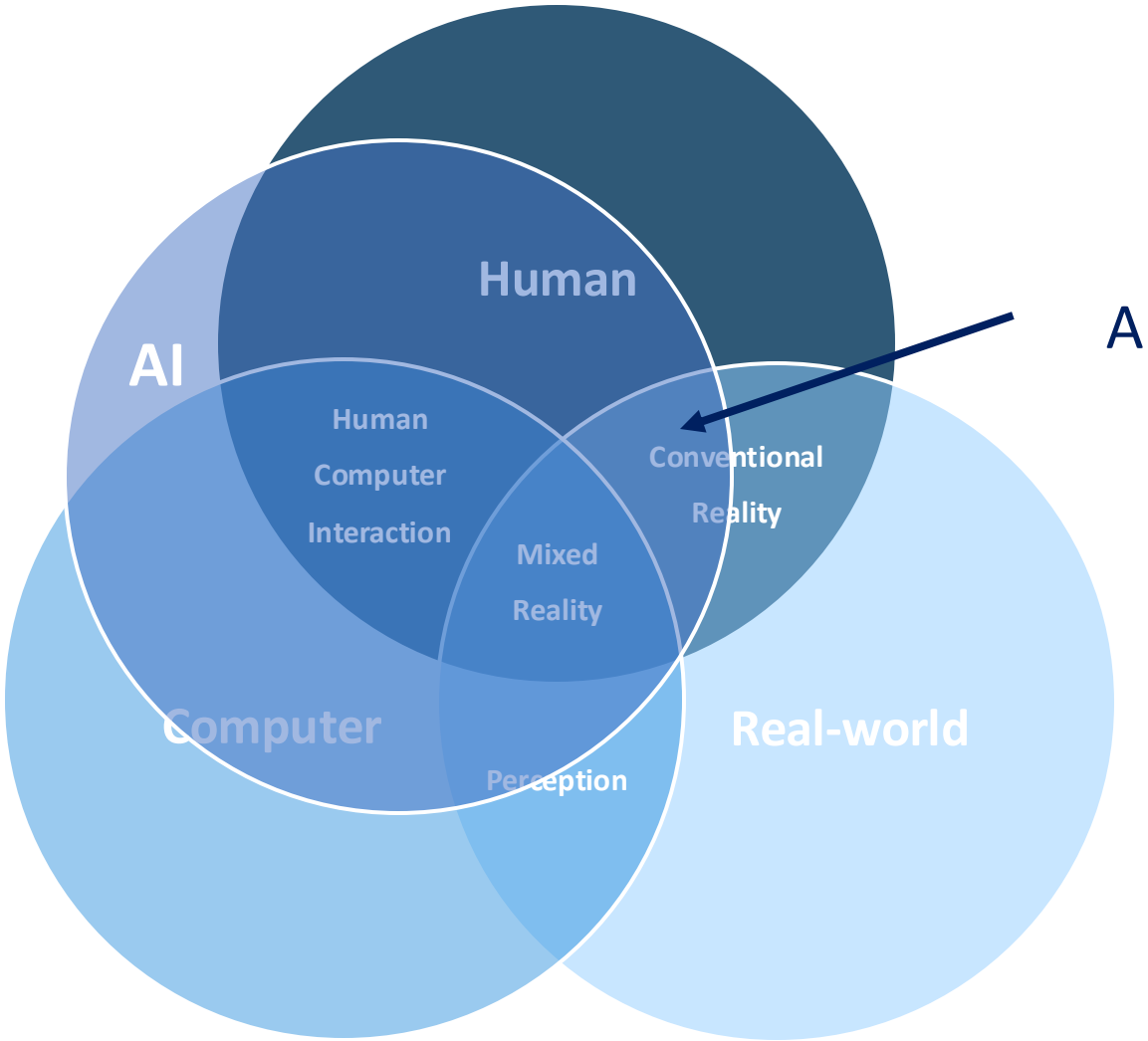
Introduction

Relationships



Introduction

Relationships



Robotics
Autonomous Driving

3D Vision

Interactive 3D with AR/VR

The model (VR-GS) can interactively manipulate the 3D scene with plausible dynamics

Input

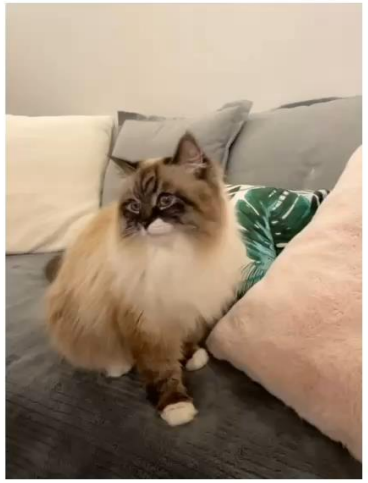
Jiang et al., VR-GS: A Physical Dynamics-Aware Interactive Gaussian Splatting System in Virtual Reality, SIGGRAPH 2024

3D Vision

Dynamic 3D scene editing

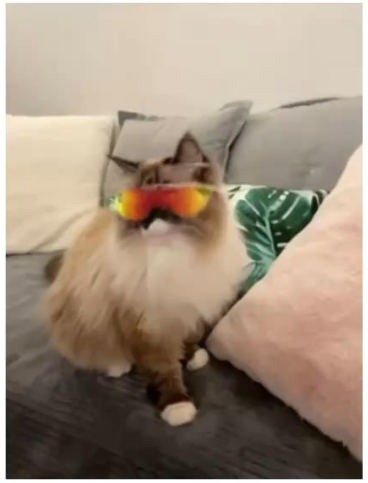
The model (Ctrl-D) can edit dynamic 3D scenes

Original



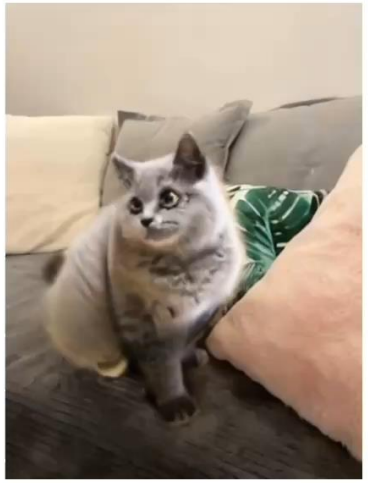
The original scene is rendered by trained dynamic 3DGS.

Editing I



“Give the cat a pair of  sunglasses”

Editing II



“Turn the cat into a  Shorthair”

He et al., Ctrl-D: Controllable Dynamic 3D Scene Editing with Personalized 2D Diffusion, Arxiv 2024

3D Vision

Dynamic 3D scene editing

Using a reference, AniGS generates reference-followed Gaussian Avatar

Reference

Motion

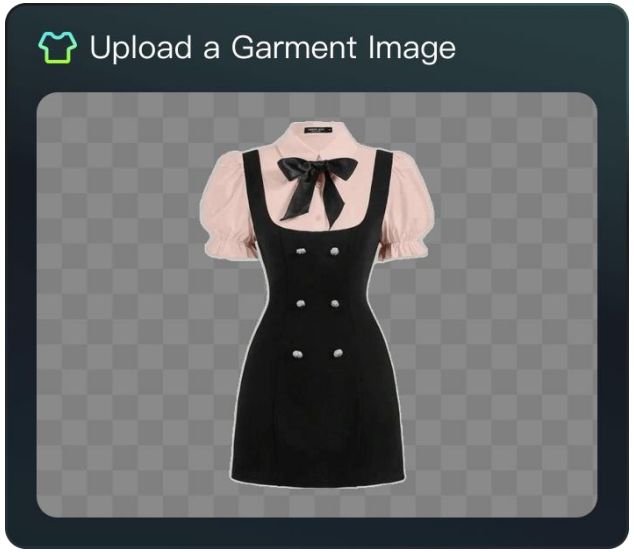
Reenactment

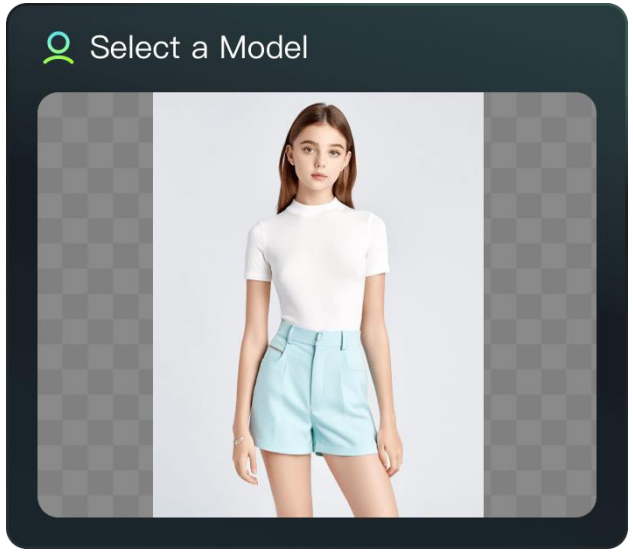
Qiu et al., AniGS: Animatable Gaussian Avatar from a Single Image with Inconsistent Gaussian Reconstruction, Arxiv 2024

3D Vision

3D Virtual Try-on

We can see the results using 3D virtual try-on



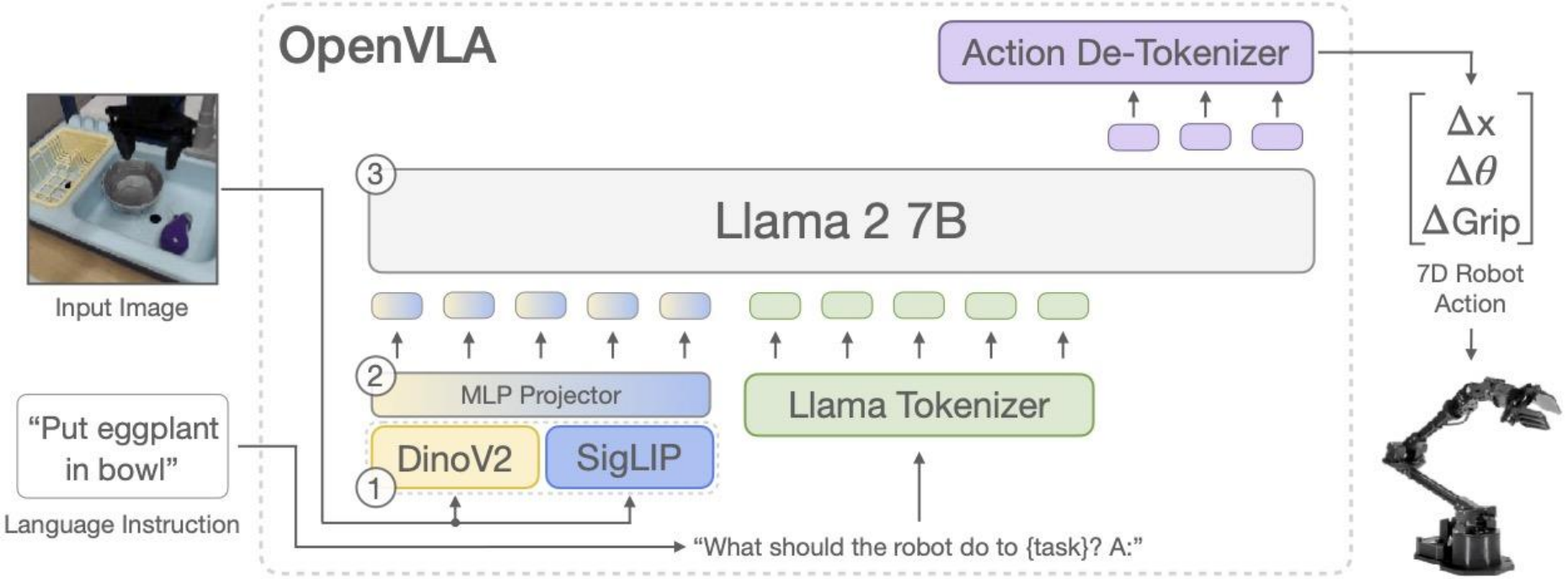


https://klingai.com, Kling AI Virtual Try-on

Robotics

Embodied AI

VLA model integrates visual and language information and performs robot actions

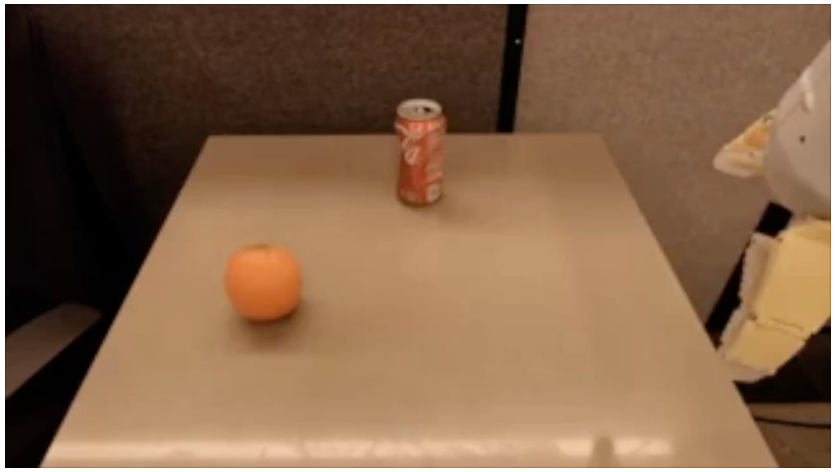
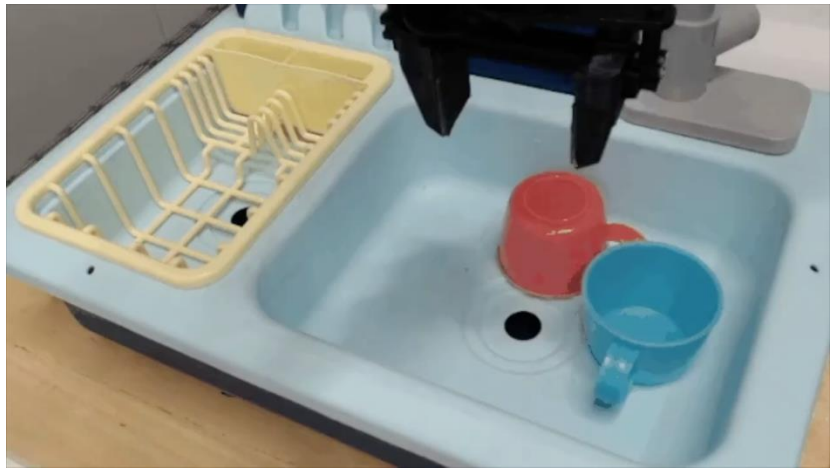


Kim et al., OpenVLA: An Open-Source Vision-Language-Action Model, CoRL 2024

Robotics

Embodied AI

VLA model integrates visual and language information and performs robot actions

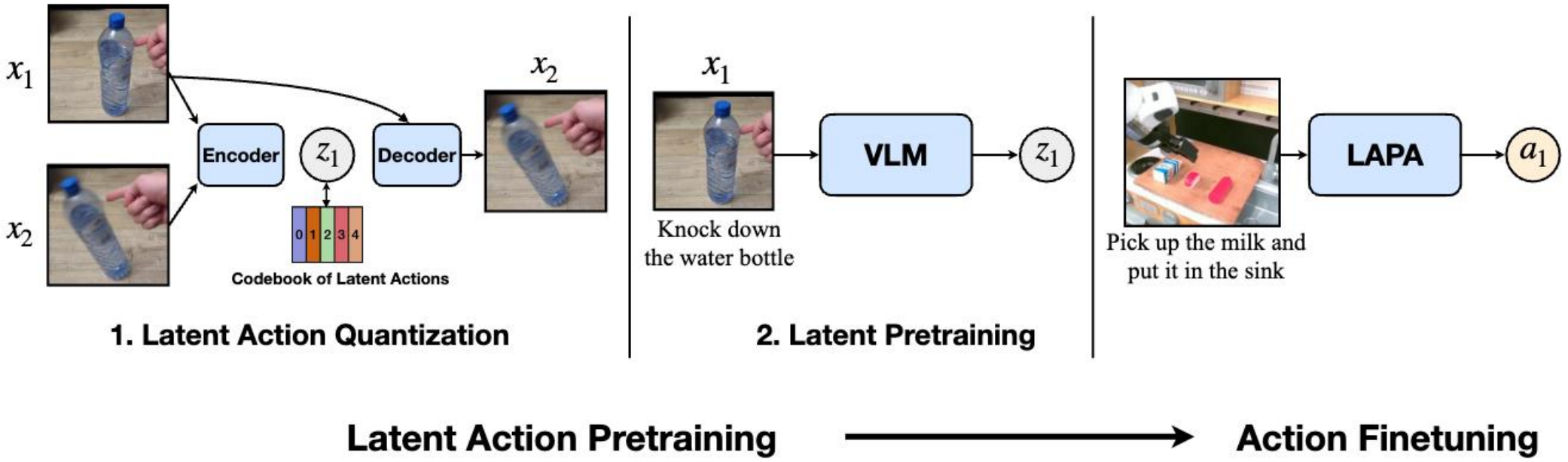




Kim et al., OpenVLA: An Open-Source Vision-Language-Action Model, CoRL 2024

Robotics

Embodied AI

LAPA achieves a robotics foundation model in unsupervised manner (scalable)



Ye et al., LAPA: Latent Action Pretraining from Videos, Arxiv 2024

Robotics

Embodied AI

LAPA achieves a robotics foundation model in unsupervised manner (scalable)

Scratch

❌

OpenVLA

❌

LAPA

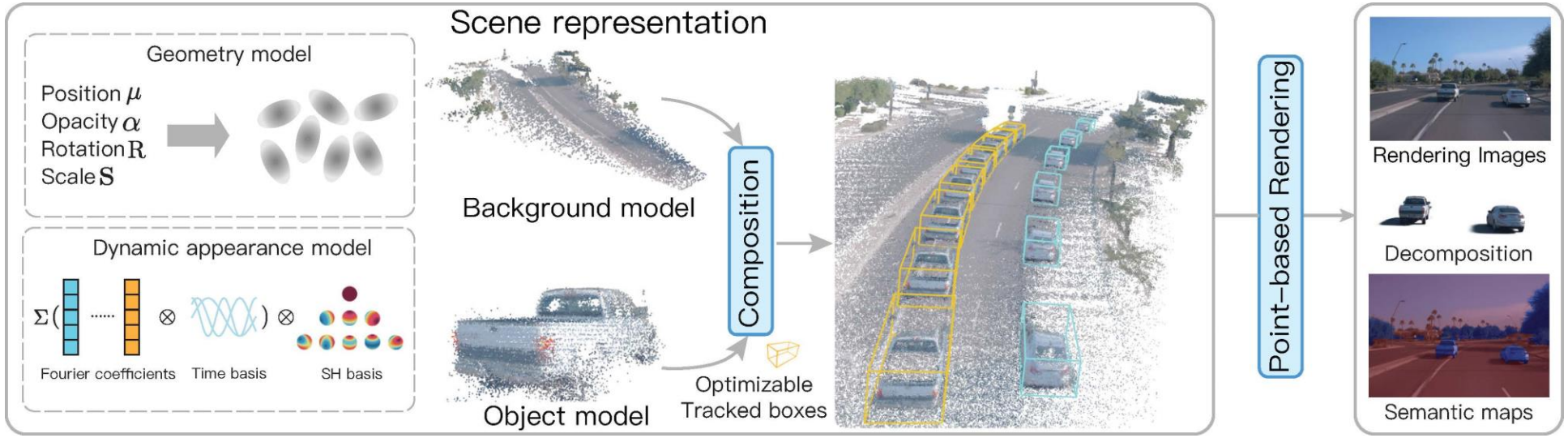
✅

Ye et al., LAPA: Latent Action Pretraining from Videos, Arxiv 2024

Autonomous Driving

Modeling urban scene

StreetGaussians reconstruct urban scene in real-time



Yan et al., Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting, ECCV 2024

Autonomous Driving

Modeling urban scene

StreetGaussians reconstruct urban scene in real-time

Yan et al., Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting, ECCV 2024

Discussions & Summary

We expend into Real-world

- AI is NOT at perception-level anymore
- We have a new target space in real-world
  - Mixed reality
    - AR/VR
- We have new target tasks in terms of AI
  - Robotics
  - Autonomous driving
- Next paradigm?
- Just be a User?