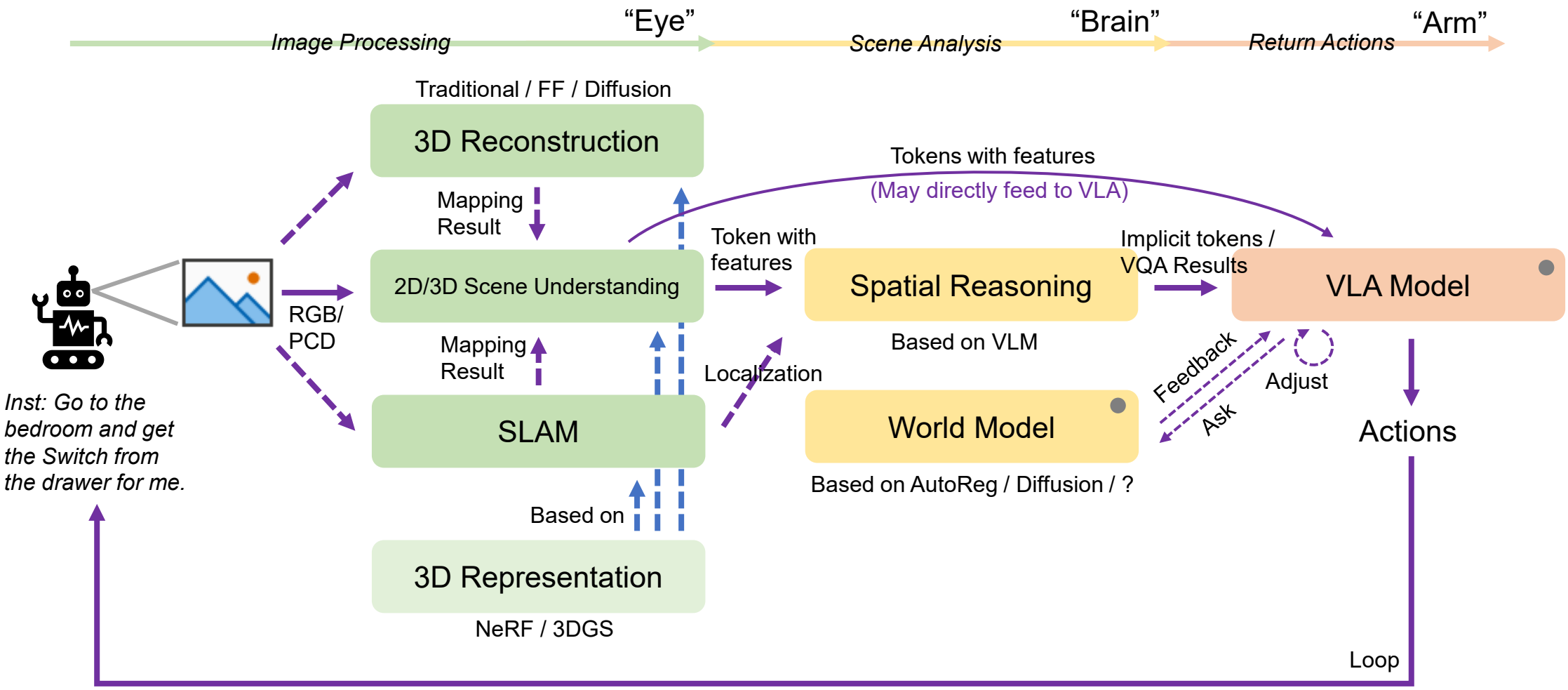
Research Interest

Yuzhuo Tian

sqrtyz.github.io

My View of Spatial Intelligence (Board Ver.) & Research Interest

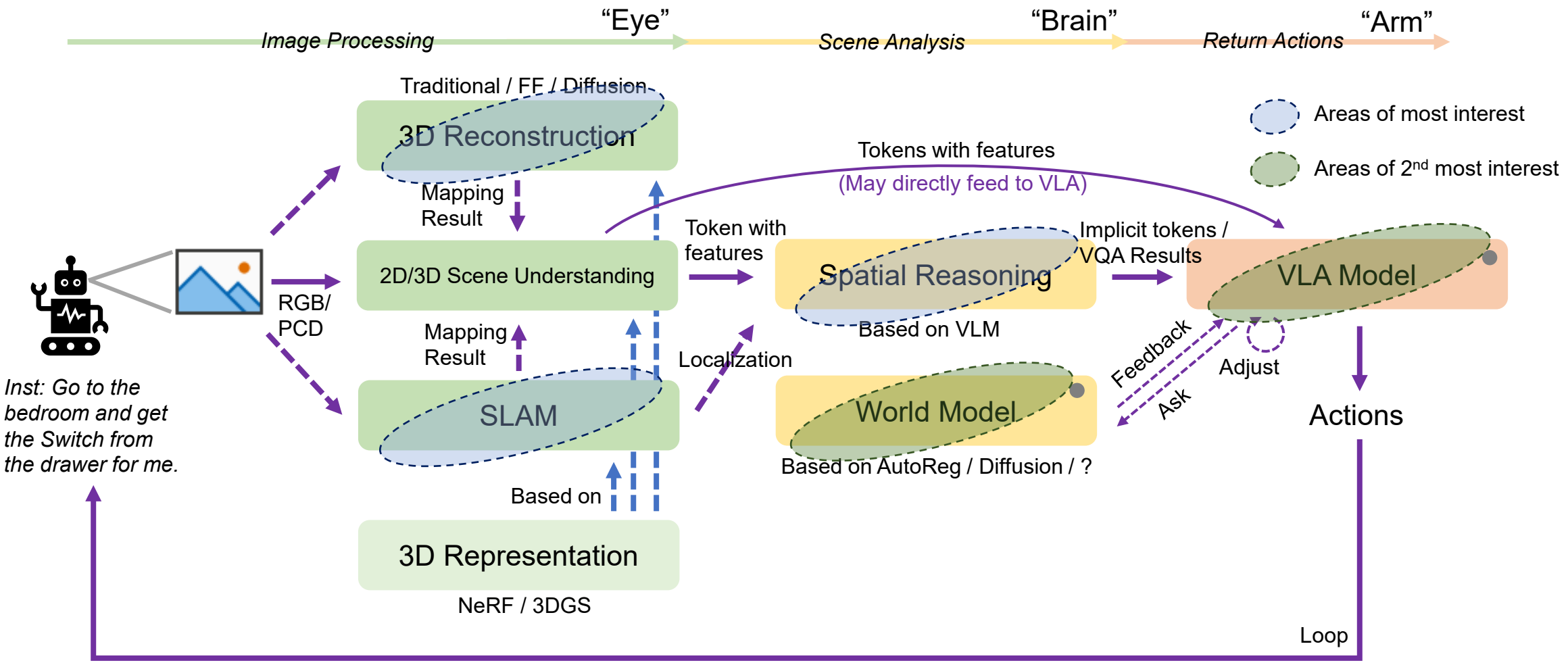


This pipeline is just my primary idea to demonstrate my view of “spatial intelligence” and introduce my interests. Interestingly, the viewpoint expressed in this figure may contradict the “bitter lesson”, which we will discuss later.

Datapath I consider optional

Datapath I consider necessary

My View of Spatial Intelligence (Board Ver.) & Research Interest



This pipeline is just my primary idea to demonstrate my view of "spatial intelligence" and introduce my interests. Interestingly, the viewpoint expressed in this figure may contradict the "bitter lesson", which we will discuss later.

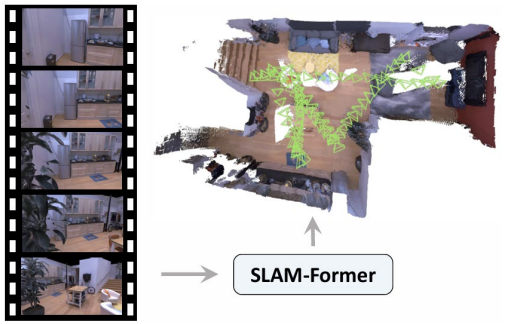
Datapath I consider optional

Datapath I consider necessary

Potential Specific Topics Interested

Inst: Go to the bedroom and get the Switch from the drawer for me. ?

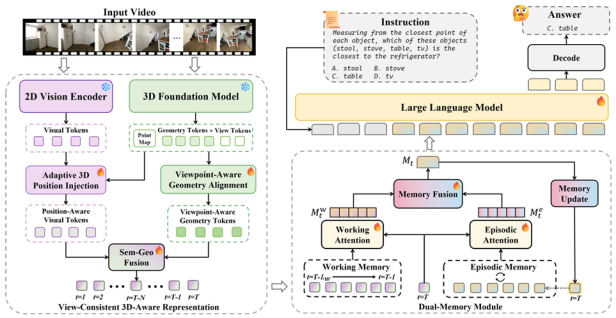
SLAM / 3D Reconstruction



Leverage feed-forward / diffusion-based model on **large-scene** reconstruction (RGB input version of my 3rd research project)

- SLAM-Former / Stream-VGGT
- While these work are able to handle longer scene compared to VGGT, the scales of scene they demonstrated in Experiment Section are still one or two rooms.
- We need robots that can map the **entire house** or even more complex indoor cases (which means larger scene) in order to accurately perform tasks within that area.

Spatial Reasoning

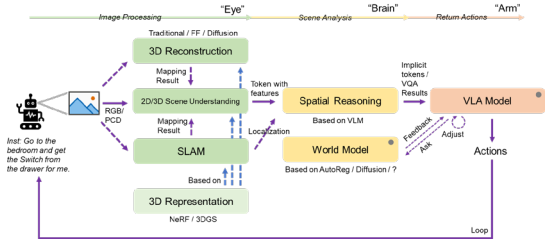


Design **memory mechanism** for long-horizon spatial reasoning

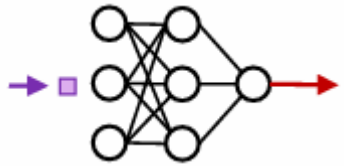
- **<u>Actually two possible pipelines to model large scene:</u>** 1) High-quality SLAM result -> guide the robot; 2) Current RGBD + Memory -> guide the robot.
- If no global SLAM result is provided, memory is needed to model the whole scene. Or, consider answering questions like *I'm in the living room right now. How do I get to the bedroom? (GT: Enter the door on your right.)* assuming the robot has explored the entire house.

My Understanding of “Bitter Lesson”

- **Simpler is better: To find meta methods.**
- Take 3D Reconstruction as the example. Traditional methods build the pipeline based on human knowledge prior: SFM + MVS. Recent work (from 2024) there are lots of “Feed-Forward” methods like Dust3r, Mast3r, or VGGT.
- Intuitive: Just like Dust3r mentioned in their paper, abundant pipeline may introduce accumulated errors.



- **More is better: To construct high-quality dataset.**
- AI is a data-driven science
- Hands-on practice: the huge impact of batch size and dataset size during training!



- **This may contradict the viewpoint I made in last slide,** as the pipeline I demonstrated introduces human prior of how we act for the given instructions (although there is no explicit knowledge). Meanwhile, long pipelines can also introduce cumulative errors.
- I have no idea which one will lead us further. The only thing I believe is that only experiments and practice can tell the truth :)

Motivation Based on Project Features

- During talk with some professors, a consensus is they treasures students' motivation.
- In my perspective: Field interested != Motivation. I may be motivated by projects with following features:

- Specific tasks with interesting background or statement（有趣的小任务场景）
  - e.g. Use robot to push boxes to specific positions; Guide agent to play Minecraft

- Methods that introduce amazing tricks（神乎其技的创新）
  - e.g. Use diffusion to replace linear layer in LLM output; RoPE embedding

- Methods that bring intuition to solution（符合直觉的通用解决方案）
  - e.g. Develop a robot that can learn to do something with one-shot video guidance; Introduce memory mechanism to SLAM

- Methods that related to basic modules or "essence"（触及基础或本质）
  - e.g. Gated Transformer