

Let's build something  
awesome!



# ImagineAR

## AI-Assisted In-Situ Authoring in Augmented Reality

Jaewook Lee<sup>\*1</sup>, Filippo Aleotti<sup>\*2</sup>, Diego Mazala<sup>2</sup>, Guillermo Garcia-Hernando<sup>2</sup>, Sara Vicente<sup>2</sup>, Oliver James Johnston<sup>2</sup>, Isabel Kraus-Liang<sup>2</sup>, Jakub Powierza<sup>2</sup>, Donghoon Shin<sup>1</sup>, Jon E. Froehlich<sup>1</sup>, Gabriel Brostow<sup>2,3</sup>, Jessica Van Brummelen<sup>2</sup>

University of Washington<sup>1</sup>; Niantic Spatial, Inc.<sup>2</sup>; University College London<sup>3</sup>



UNIVERSITY of  
WASHINGTON



















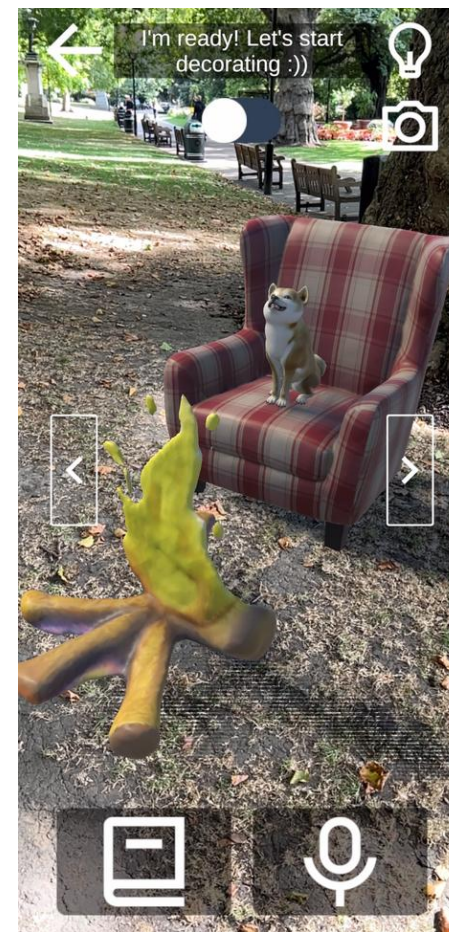
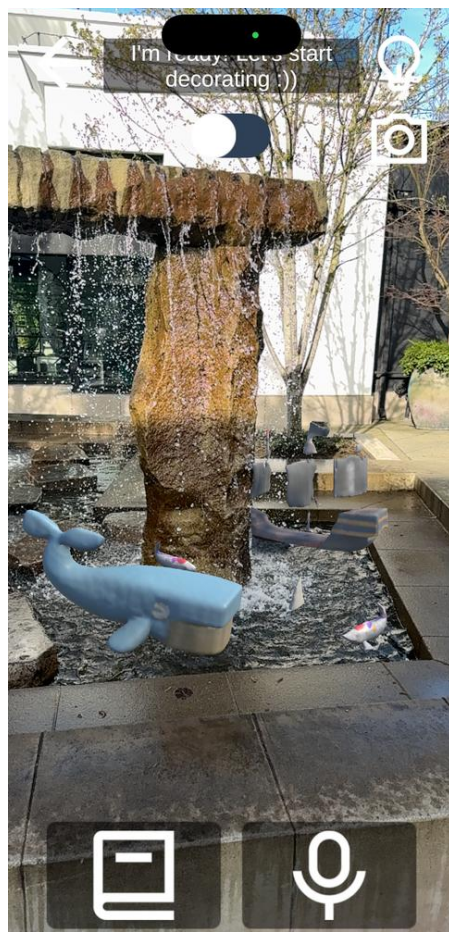


**Place a two-headed  
giraffe next to the tree**

?







We introduce ***ImagineAR***, a mobile AI-Assisted AR authoring tool that lets non-experts create AR scenes simply through speech.



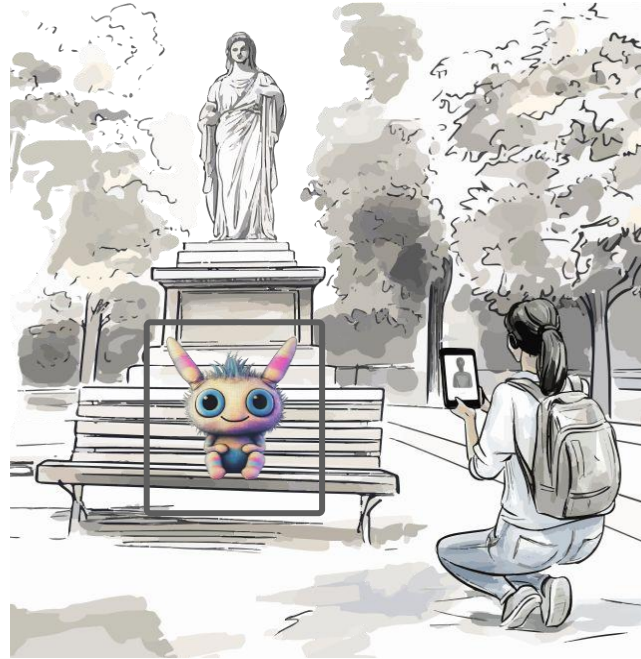




# The ImagineAR System



01 A custom outdoor scene understanding pipeline.



02 A fast, remote asset generation server.



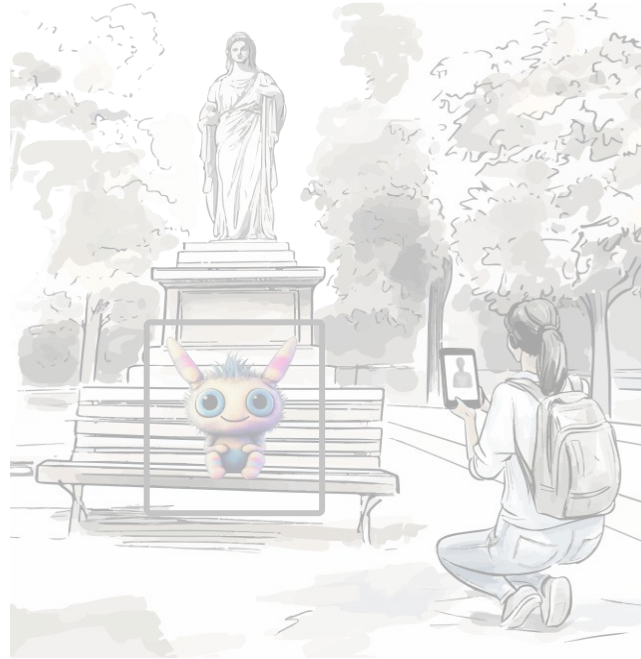
03 A mobile interface for LLM-driven speech interactions.



# The ImagineAR System



01 A custom outdoor scene understanding pipeline.



02 A fast, remote asset generation server.



03 A mobile interface for LLM-driven speech interactions.



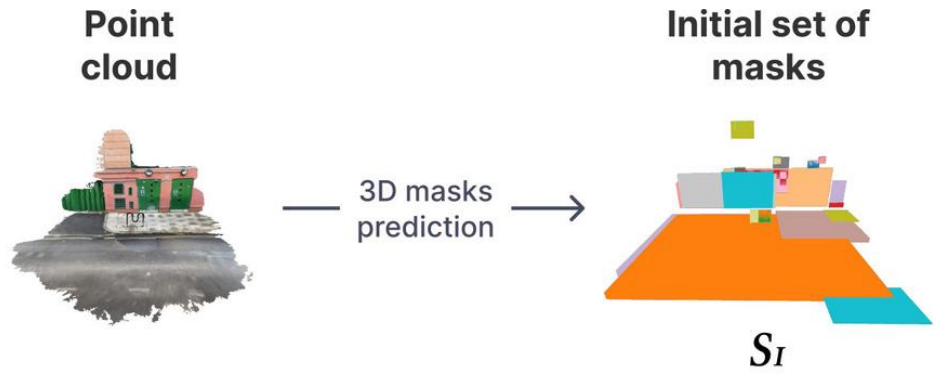
# Scene Understanding

Point  
cloud



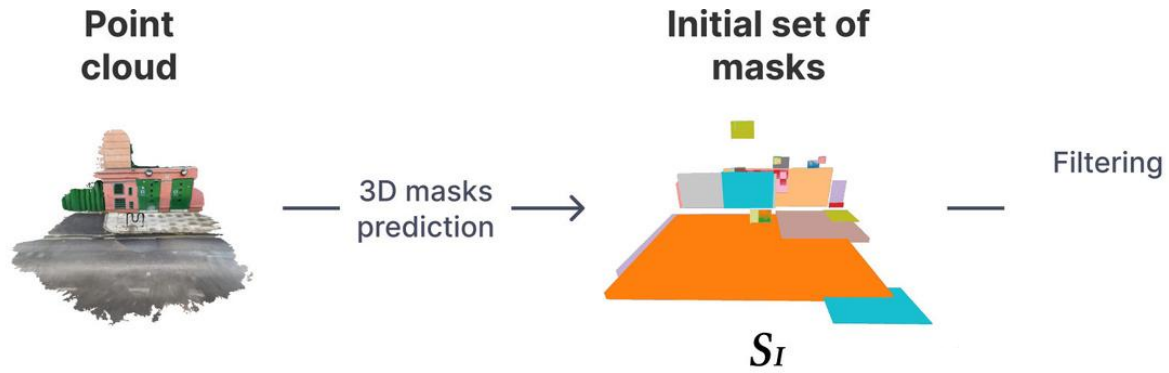


# Scene Understanding



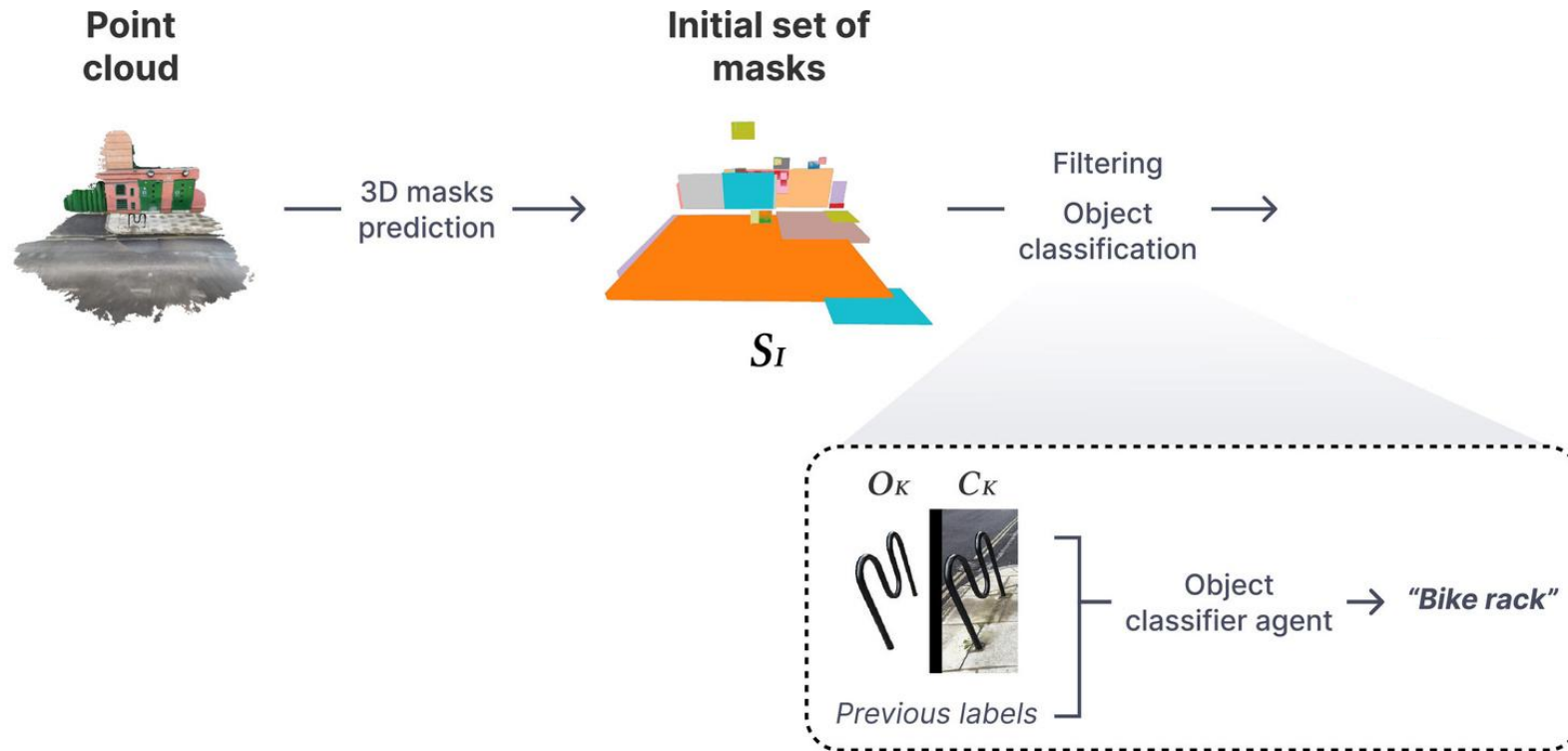


# Scene Understanding



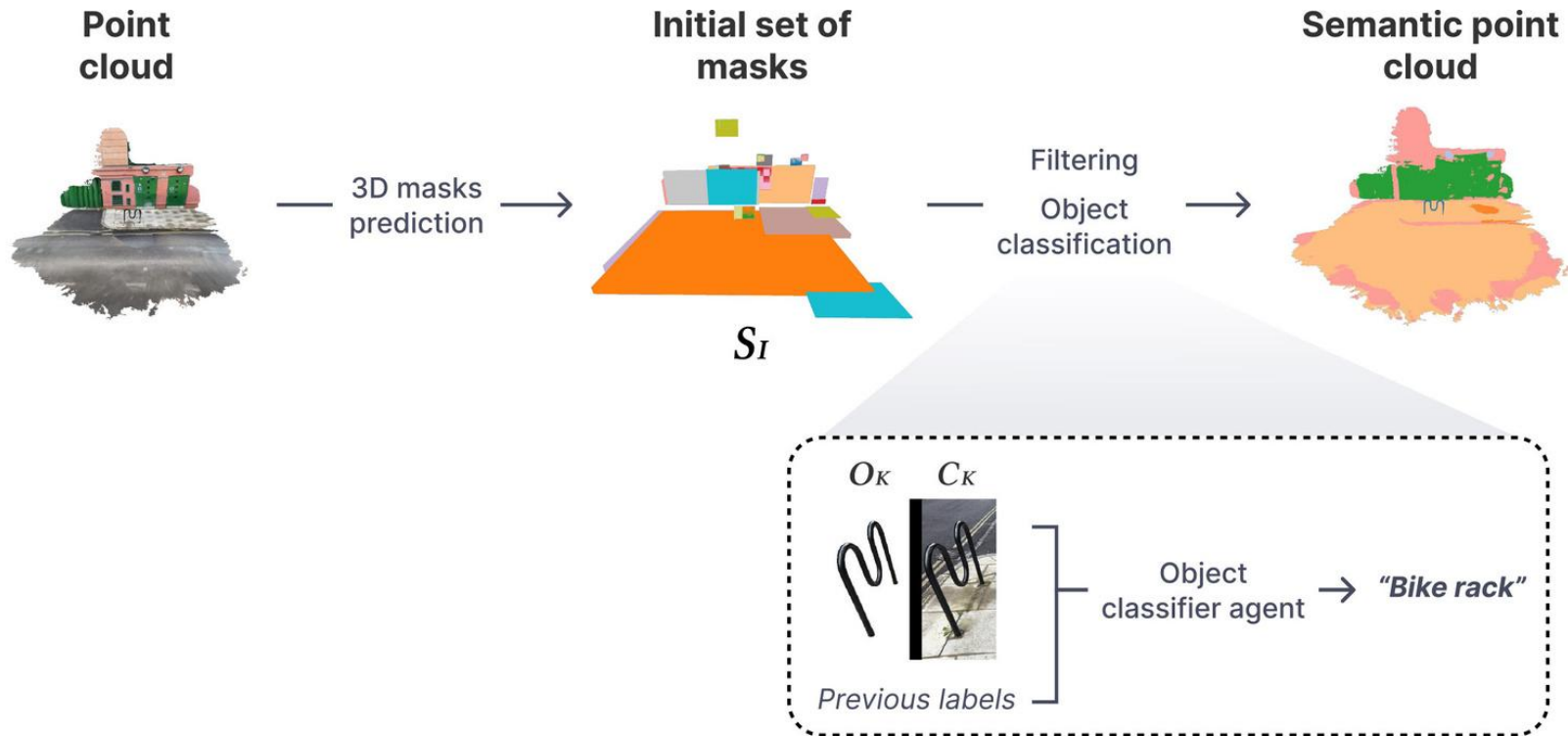


# Scene Understanding



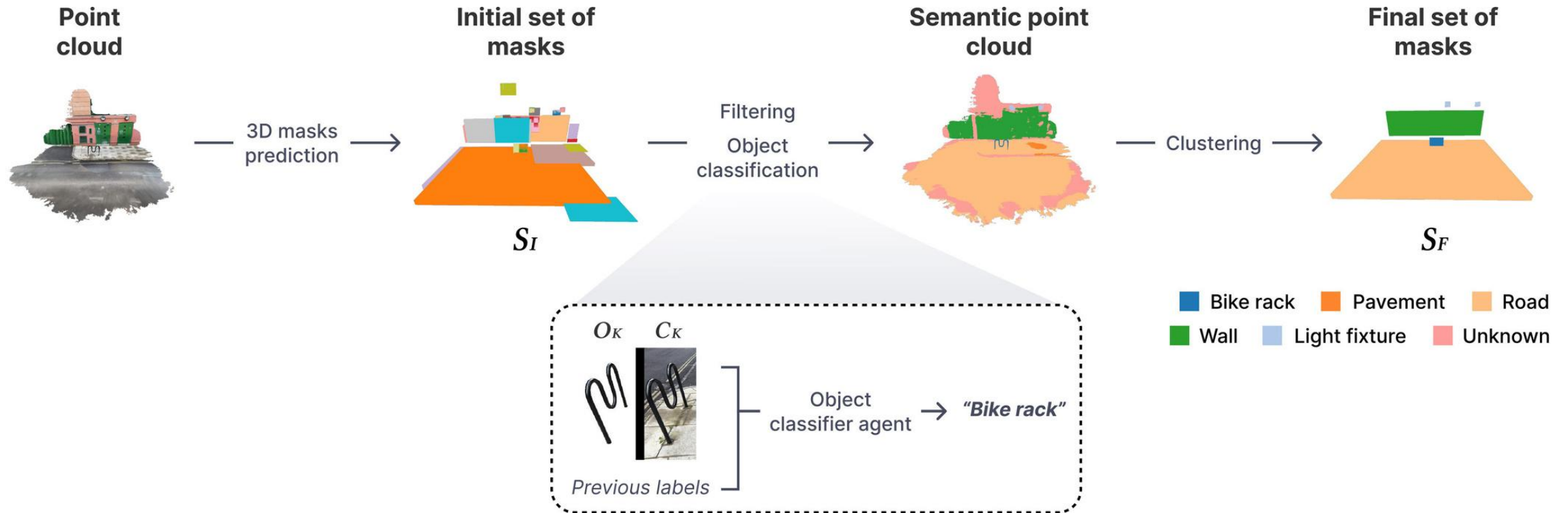


# Scene Understanding





# Scene Understanding





# Scene Understanding





# Scene Understanding

```
{
  "objects": [
    {
      "min_x": "-3.62",
      "min_y": "-0.05",
      "min_z": "4.60",
      "max_x": "-1.21",
      "max_y": "0.01",
      "max_z": "5.81",
      "probable_surface_height": "-0.01",
      "class_name": "sidewalk"
    },
    {
      "min_x": "0.26",
      "min_y": "0.02",
      "min_z": "4.17",
      "max_x": "2.87",
      "max_y": "0.15",
      "max_z": "5.43",
      "probable_surface_height": "0.04",
      "class_name": "sidewalk"
    },
    {
      "min_x": "-4.11",
```



# Scene Understanding

**Vase**

17 mins on Our Method



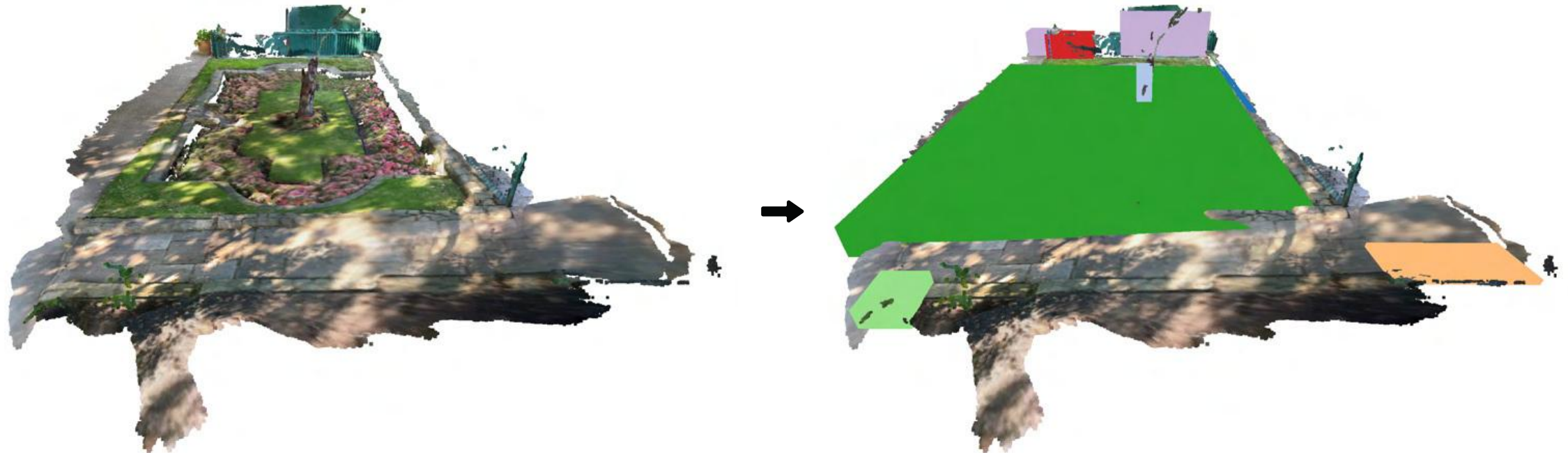
■ Sidewalk    ■ Planter    ■ Flowers



# Scene Understanding

## Garden

13.4 mins on Our Method



- |                      |            |              |            |       |            |
|----------------------|------------|--------------|------------|-------|------------|
| Stone wall           | Grass      | Rock         | Path       | Fence | Tree trunk |
| Advertisement banner | Fence post | Potted plant | Flower bed |       |            |



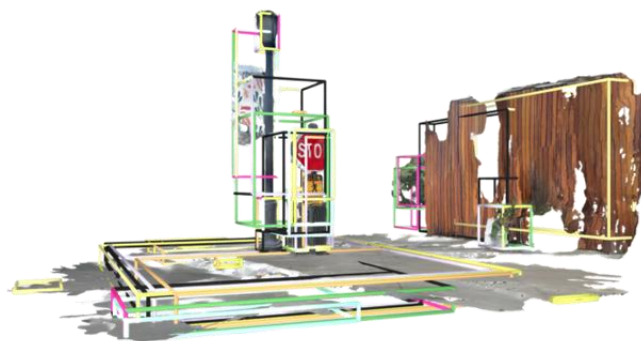
# Scene Understanding

## Stop Sign



**Ground Truth**

8 Masks



**OpenMask3D**

152 Masks



**Ours**

12 Masks



# Scene Understanding

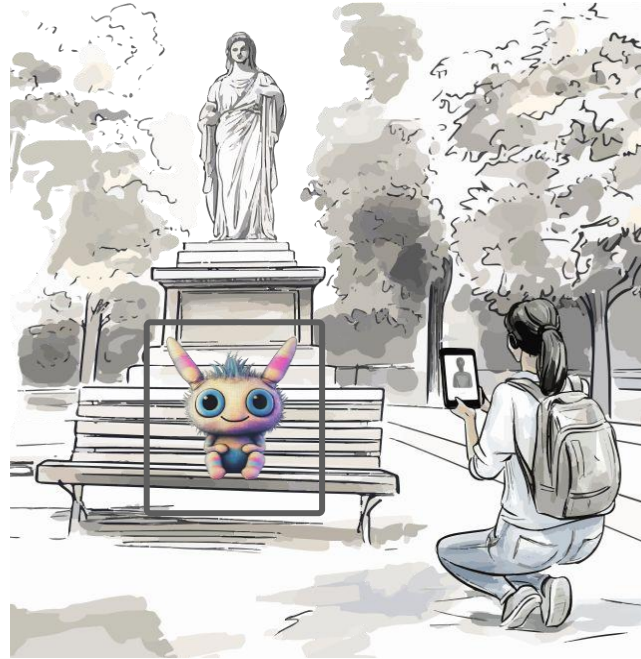
Method	Components Used				N	Evaluation Metrics	
	Filtering	Monocular Depth	Labeling	Clustering		mean Recall $\uparrow$	mean SS $\uparrow$
OpenMask3D [101]			CLIP		752	0.800	0.738
Ablation A	✓		GPT-4o		59	0.508	0.659 ( $\pm$ 0.008)
Ablation B	✓	✓	GPT-4o		60	0.558	0.791 ( $\pm$ 0.010)
Ablation C	✓	✓	CLIP		60	0.558	0.730
<b>Ours</b>	✓	✓	GPT-4o	✓	49 ( $\pm$ 1)	0.622 ( $\pm$ 0.087)	0.791 ( $\pm$ 0.073)



# The ImagineAR System



01 A custom outdoor scene understanding pipeline.



02 A fast, remote asset generation server.



03 A mobile interface for LLM-driven speech interactions.



# Asset Generation

User prompt

*"a roman statue"*

# Asset Generation

User prompt

*"a roman statue"*

Visual examples



Good

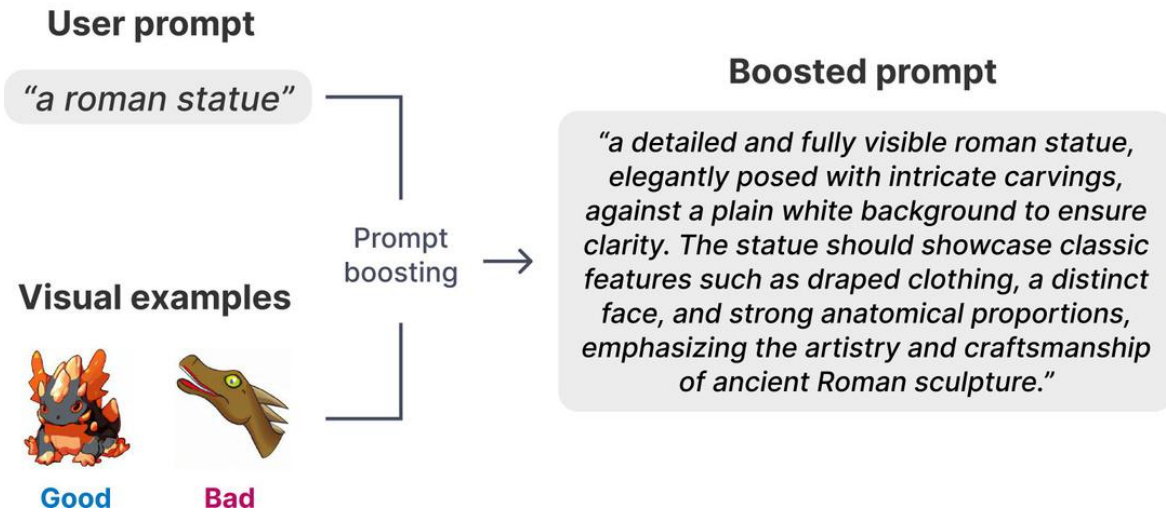


Bad

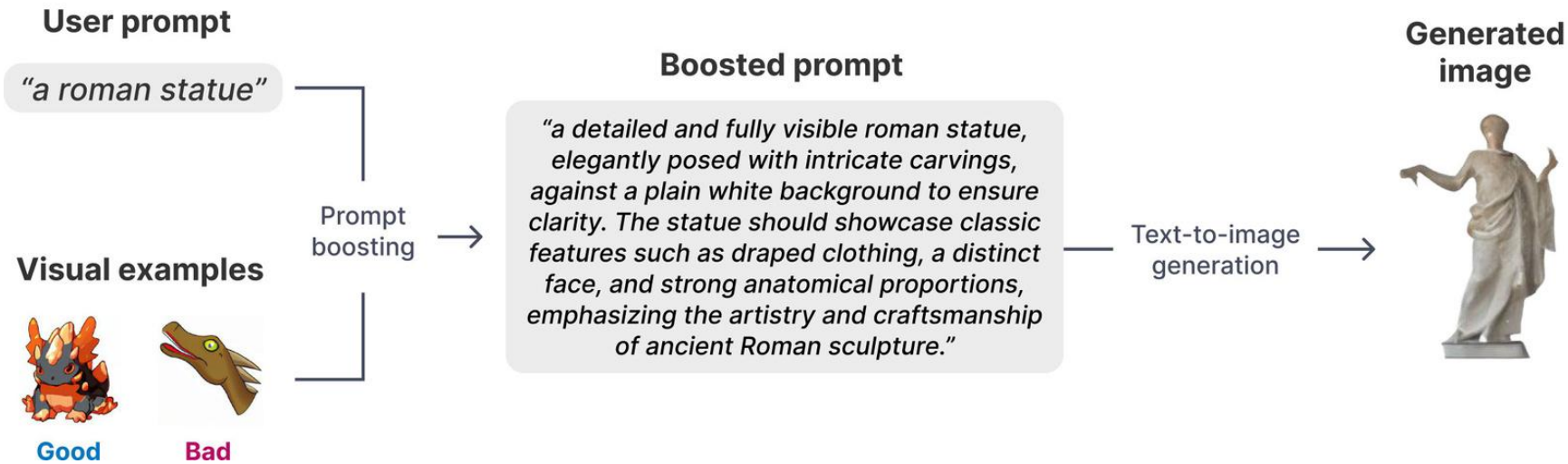
Prompt  
boosting



# Asset Generation



# Asset Generation

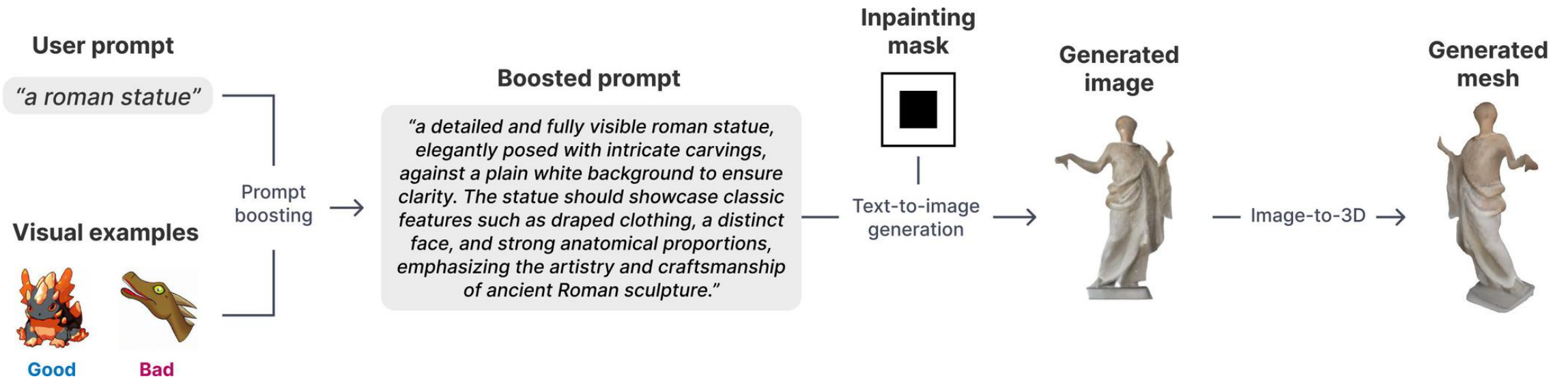




# Asset Generation



# Asset Generation





# Asset Generation



**Sandcastle**



**Teapot**



**Orchid**



**Paper Plane**

# Asset Generation

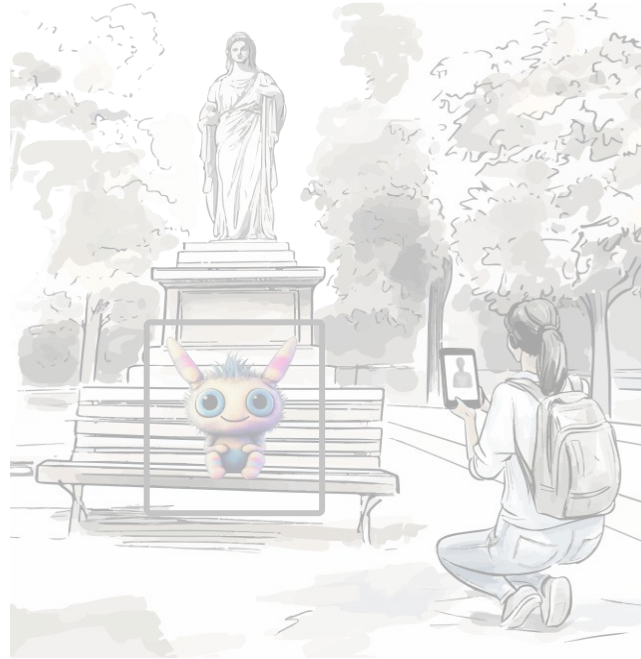
Model Name	Time	Quality $\uparrow$
DreamFusion [81]	30 min	24.9
ProlificDreamer [110]	240 min	51.1
MVDream [93]	30 min	53.2
DreamGaussian [102]	7 min	19.9
InstantMesh [117] + Dall-E 2 [75]	< 1 min	32.6
InstantMesh + Dall-E 2 + Prompt Boosting (Ours)	< 1 min	34.8



# The ImagineAR System



01 A custom outdoor scene understanding pipeline.

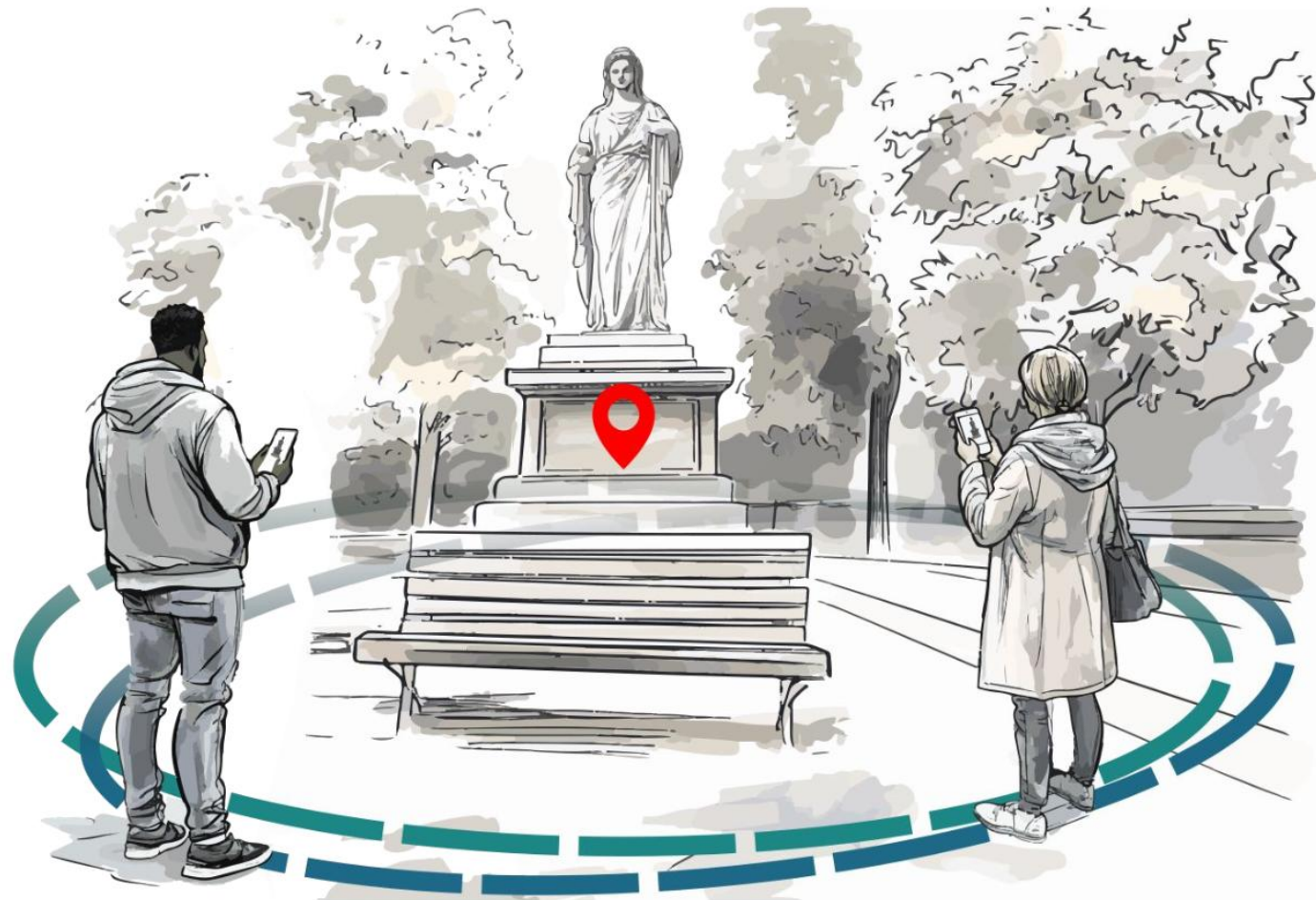


02 A fast, remote asset generation server.



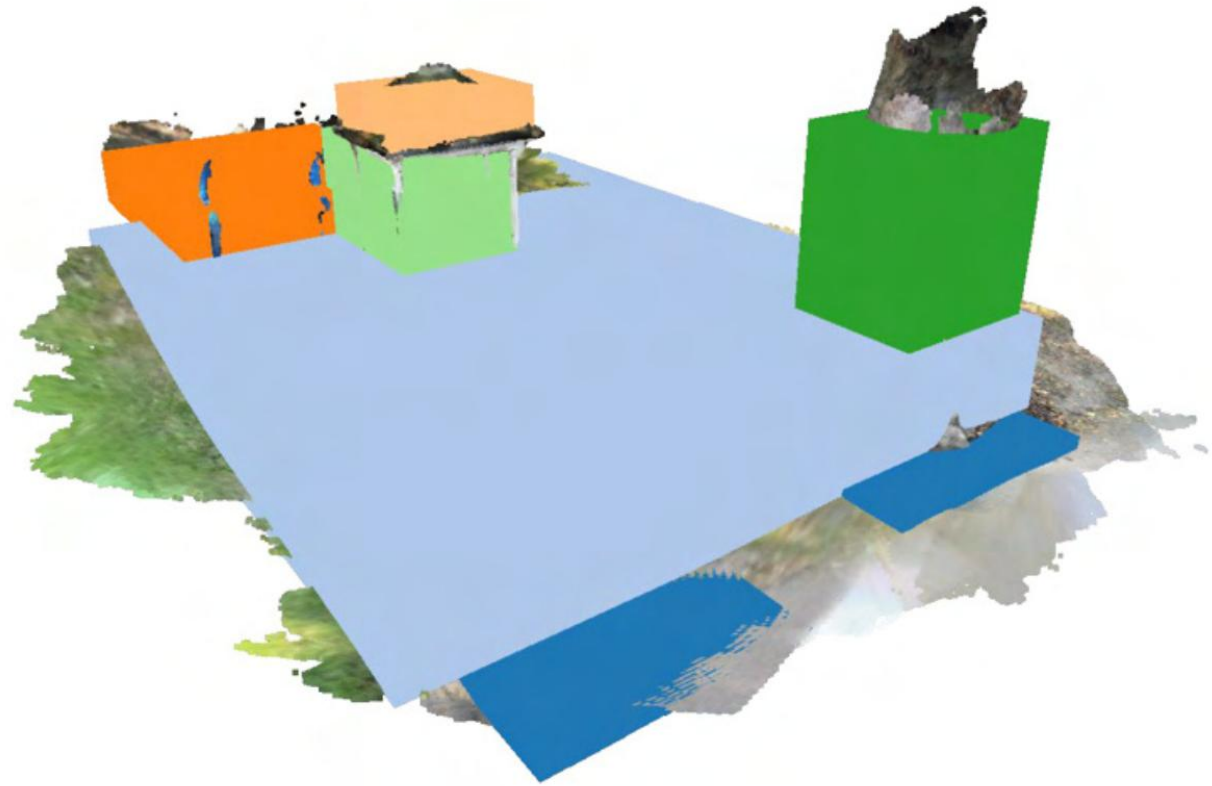
03 A mobile interface for LLM-driven speech interactions.

# LLM Interaction



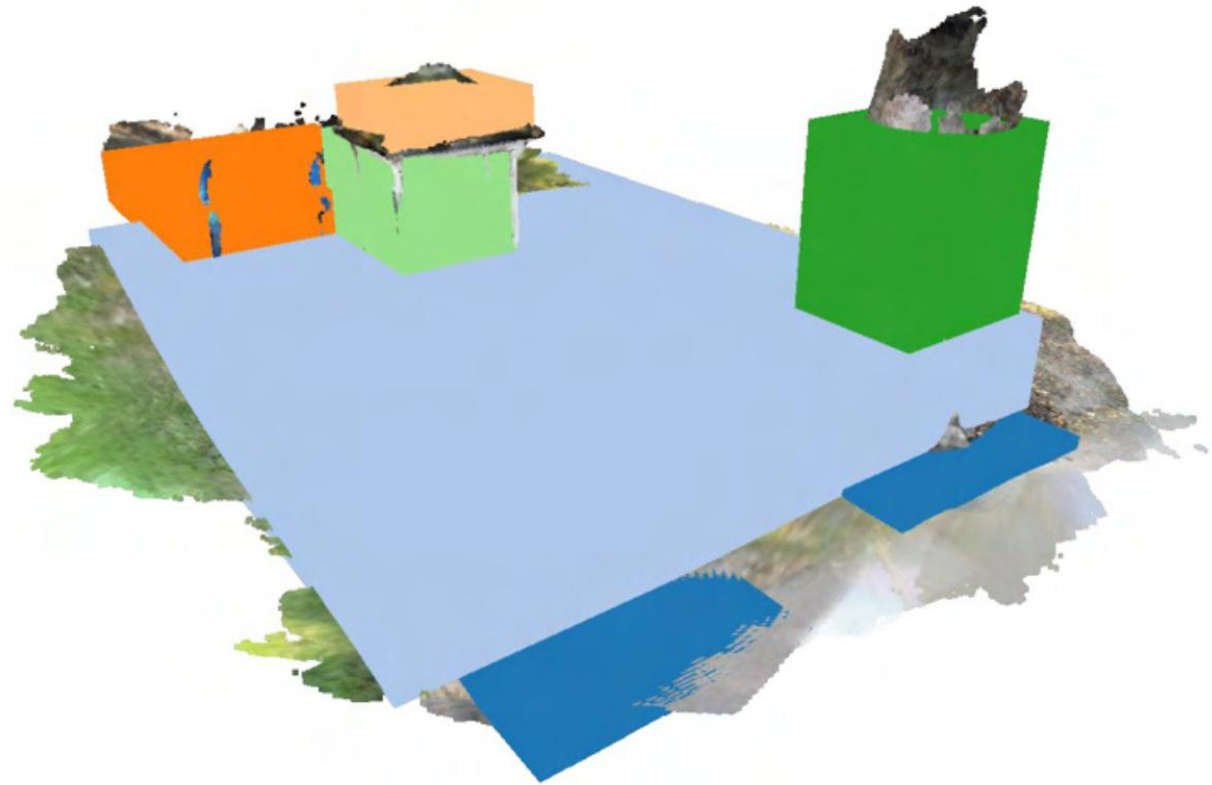


# LLM Interaction



■ Sidewalk ■ Tree ■ Roof ■ Shed ■ Handrail ■ Grass

# LLM Interaction



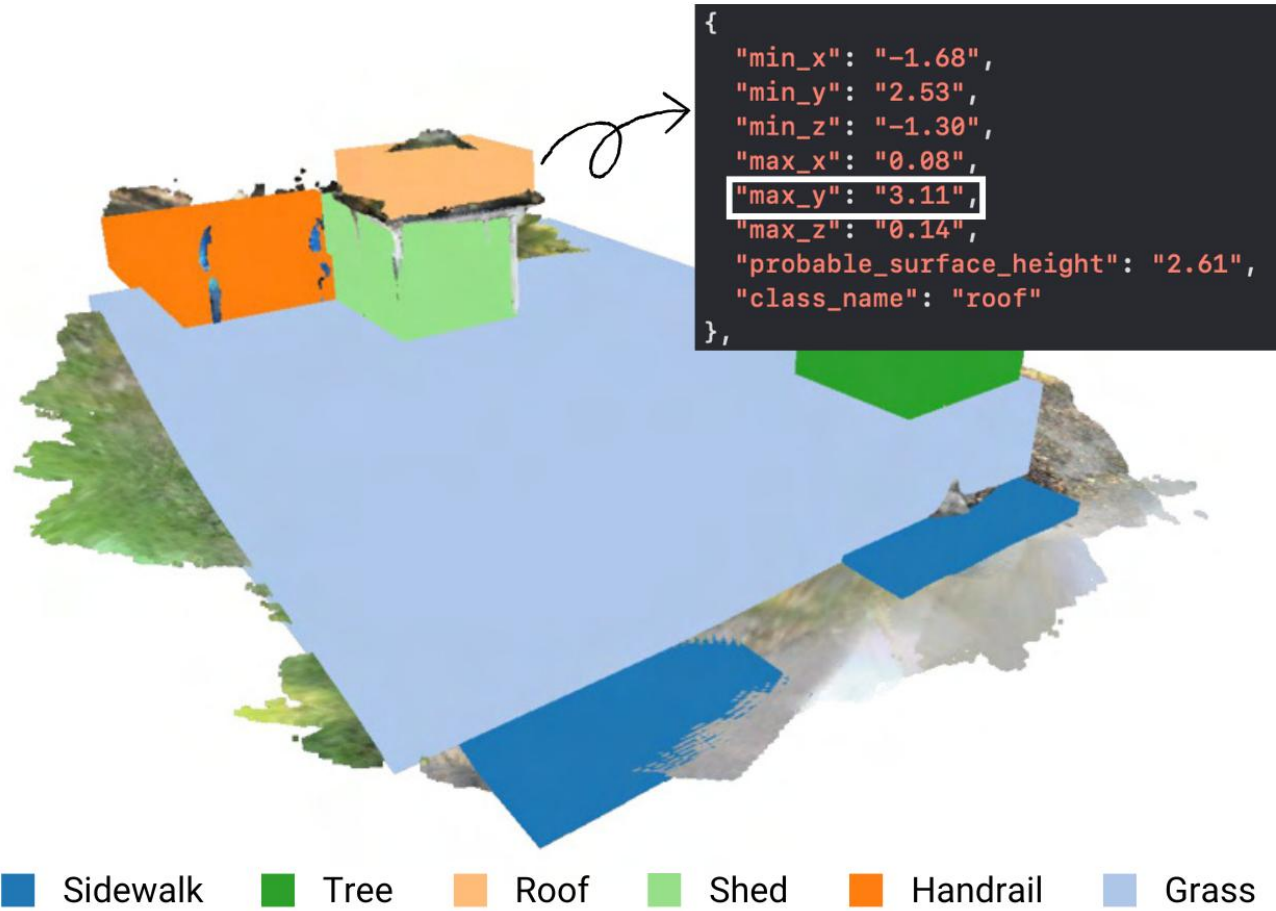
■ Sidewalk ■ Tree ■ Roof ■ Shed ■ Handrail ■ Grass



# LLM Interaction

```
{
  "objects": [
    {
      "min_x": "-3.62",
      "min_y": "-0.05",
      "min_z": "4.60",
      "max_x": "-1.21",
      "max_y": "0.01",
      "max_z": "5.81",
      "probable_surface_height": "-0.01",
      "class_name": "sidewalk"
    },
    {
      "min_x": "0.26",
      "min_y": "0.02",
      "min_z": "4.17",
      "max_x": "2.87",
      "max_y": "0.15",
      "max_z": "5.43",
      "probable_surface_height": "0.04",
      "class_name": "sidewalk"
    },
    {
      "min_x": "-4.11",
```

# LLM Interaction

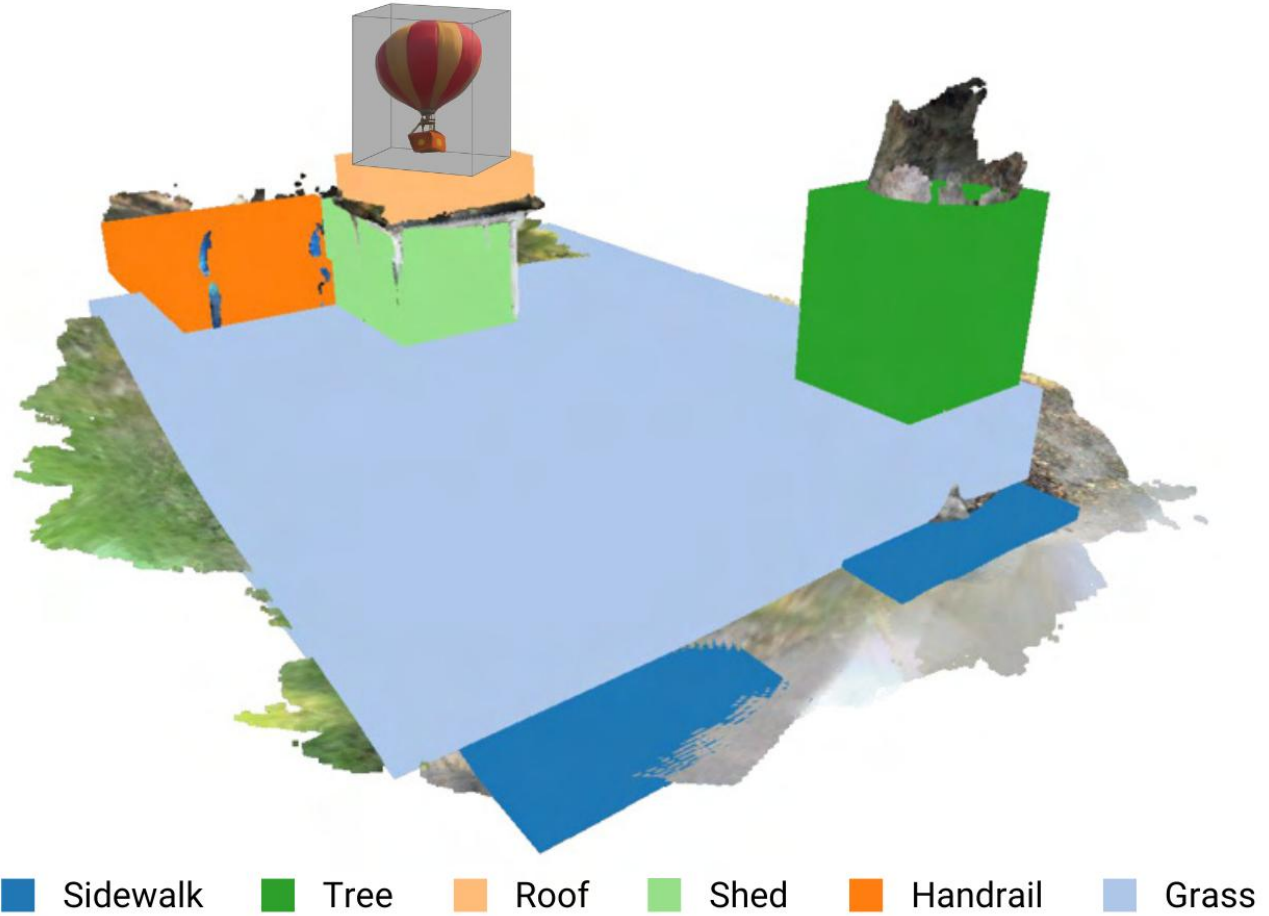




# LLM Interaction



# LLM Interaction





# LLM Interaction

Component	Time
Prompt Boosting	2.53s $\pm$ 0.91s
Image Generation	12.53s $\pm$ 2.48s
Background Removal	0.04s $\pm$ 0.002s
Image to Mesh	9.14s $\pm$ 0.08s
In-App LLM Agents	9.68s $\pm$ 1.24s
Total	33.92 $\pm$ 5.83s





# User Study

We conducted a three-part study in a public park with 20 participants.

Part 1 – comparing AI's role across different authoring stages.

## **3 levels of AI role:**

1. Manual
2. AI-Assisted
3. AI-Decided

**X**

## **5 AR authoring tasks:**

1. Brainstorming
2. Object Creation
3. Placement
4. Modification
5. Removal

Part 2 – free-form authoring with ImagineAR.

Part 3 – brainstorming future designs and applications of ImagineAR.





# Hybrid Workflows are Essential

- They want to co-create with AI, not just rely on it
- AI boosted creativity and speed
- Manual gave control and precision
- 18/20 participants combined both

# AI Creativity is a Double-Edge Sword

- AI's creativity was inspiring and fun but also frustrating
- Some loved playful, unexpected outputs
- Others wanted closer alignment to their creative intent
- Users need ways to "control" AI creativity



# Toward Conversational Co-Creation

- Users want dialogue, not command loops
- Iteratively refine ideas with the AI through conversations
- “Let’s make an aquarium!”
  - What fish?
  - What color?
  - Any vegetation?

# User Creations



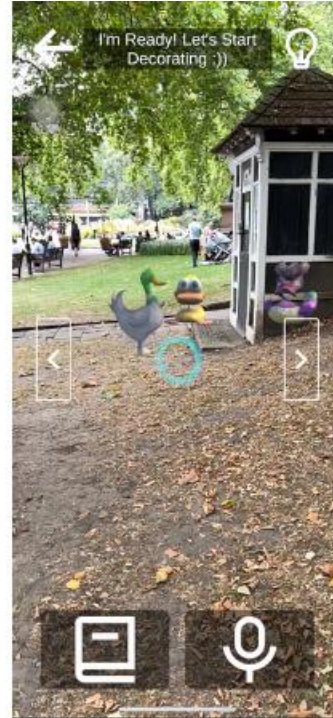
PP1: A hot air balloon on top of the shed.



PP2: A rainforest environment.



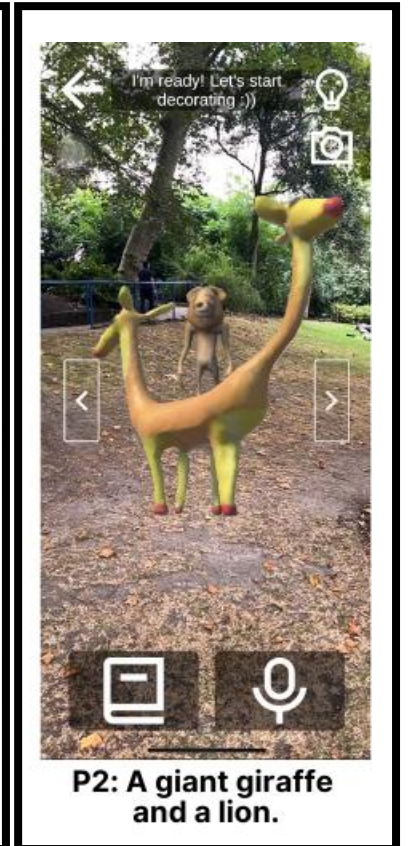
PP3: A food truck and a dog.



PP4: A flock of ducks.



P1: A dancing t-rex on grass.



P2: A giant giraffe and a lion.



# User Creations



**P3: A snowman with a carrot stuck on its head.**



**P4: A giant orange banana on top of shed.**



**P5: A pink fountain with surrounding plants.**



**P6: A sphinx and a pyramid on the ground.**



**P7: A dragon on a log around a campfire.**



**P8: Animal kingdom with a cat, horse and fox.**



# User Creations



**P9: A dog enjoying a campfire.**



**P10: A royal crown on my head.**



**P11: A spider chasing a cat chasing a bird and a mouse**



**P12: An animal pool party.**



**P13: A duck with a pinwheel during Halloween**



**P14: A house with a helicopter and truck.**



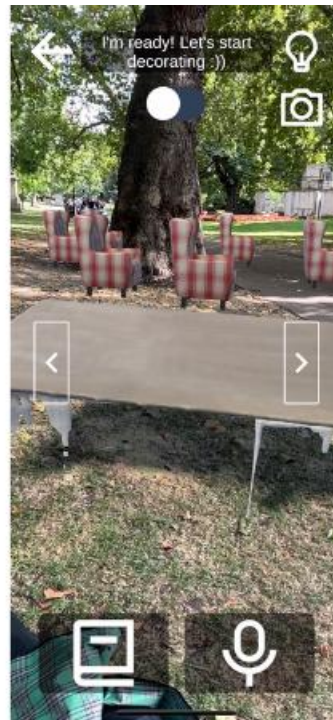
# User Creations



**P15: A "Thank You" decoration.**



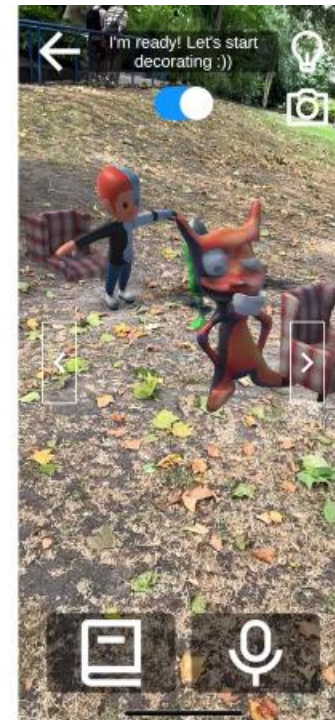
**P16: A cat chasing a row of yellow ducks.**



**P17: A table with many seats.**



**P18: A magic tree on a rock.**



**P19: Animals drinking coffee and looking at a spaceship.**



**P20: Sci-Fi scene with robots and aliens.**



Let's build something awesome!

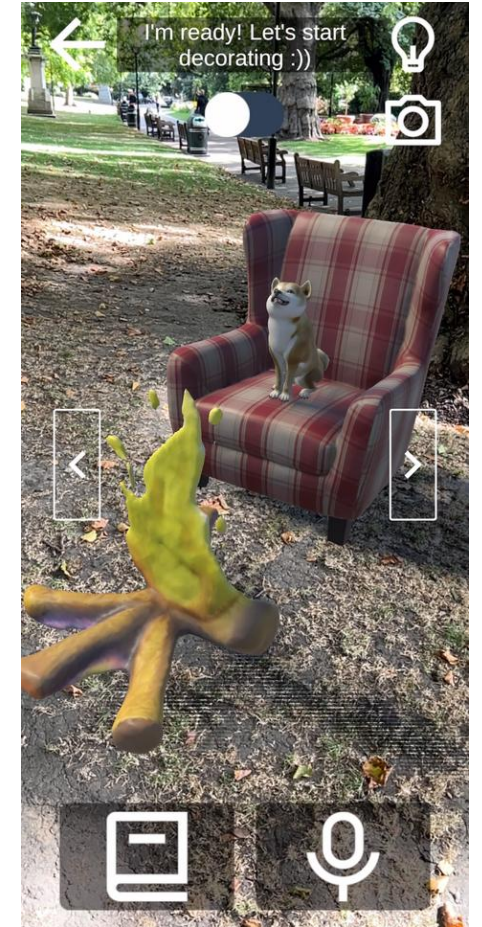
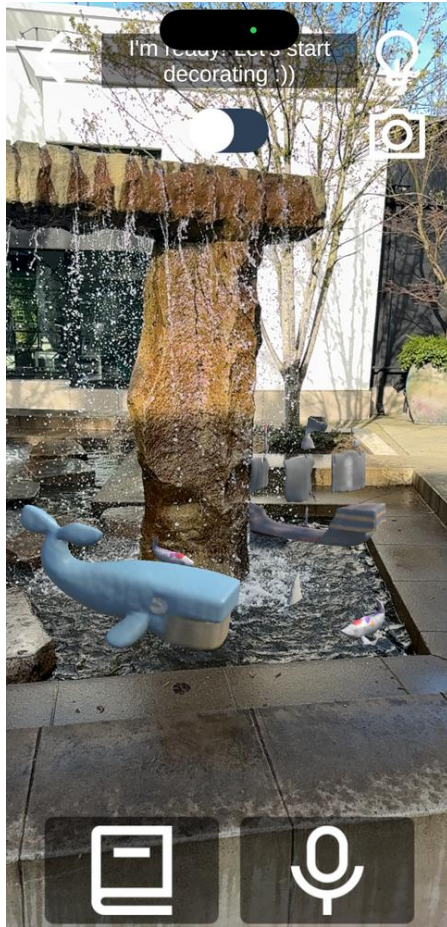


3

6







With ***ImagineAR***, we take a leap toward empowering *anyone* to create *anything, anywhere* --- simply by speaking their *imagination*.



# Thanks for listening!

Let's Connect!

Email [jaewook4@cs.washington.edu](mailto:jaewook4@cs.washington.edu)

Website <https://jaewook-lee.com>

Twitter [https://twitter.com/jaewook\\_jae](https://twitter.com/jaewook_jae)

