ManipVQA: Injecting Robotic Affordance and Physically Grounded Information into Multi-Modal Large Language Models

Siyuan Huang*[1,3], Iaroslav Ponomarenko*[2], Zhengkai Jiang[5], Xiaoqi Li[2], Xiaobin Hu[6]
Peng Gao[1], Hongsheng Li[1,4], Hao Dong[2]

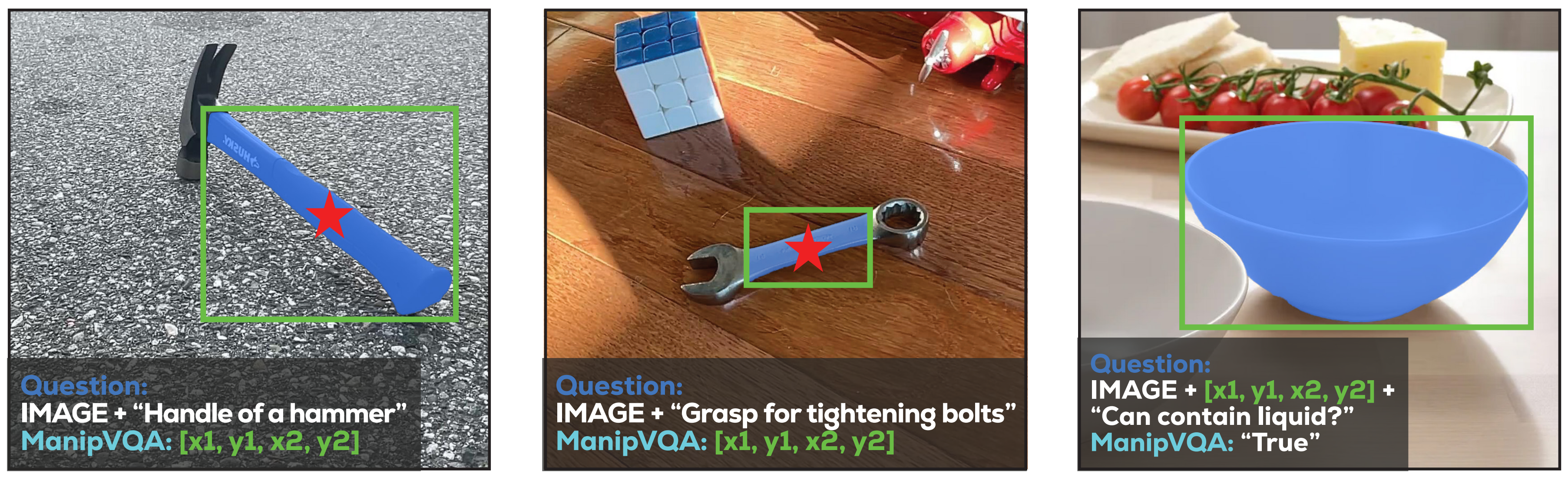
[1]Shanghai AI Laboratory, [2]Peking University, [3]Shanghai Jiao Tong University, [4]CUHK, [5]UCAS, [6]TUM

Introduction

Current MLLMs, while proficient in general vision tasks, encounter significant challenges in robotic manipulation. These limitations arise from their struggle to recognize affordances and physical properties of objects, which are essential for robotic manipulation.

ManipVQA overcomes these limitations by infusing MLLMs with robotics-specific knowledge.

Demonstration

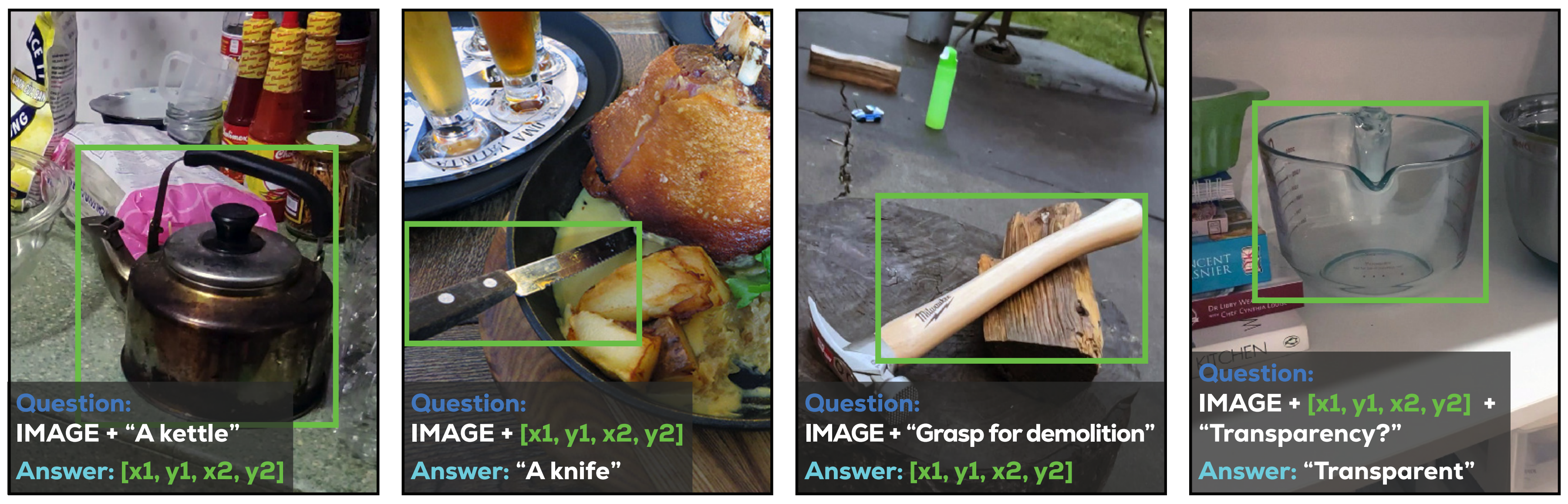

Question:
IMAGE + "Handle of a hammer"
ManipVQA: [x1, y1, x2, y2]
Question:
IMAGE + "Grasp for tightening bolts"
ManipVQA: [x1, y1, x2, y2]
Question:
IMAGE + [x1, y1, x2, y2] +
"Can contain liquid?"
ManipVQA: "True"

Affordance Recognition

Physical Properties Understanding

Method & Key Contributions

1 Unified VQA Format



Referring Expression Comprehension (REC)

Referring Expression Generation (REG)

REC-Grounding-Affordance

REG-Physical

General Vision Tasks

Robotics Specific Tasks

2 Training on Specialized Datasets

General Visual Reasoning Datasets

PACO, RefCOCO, and Visual Genome
Rich sources of information on parts and attributes of common objects

Physically Grounded Dataset

PhysObjects Dataset
We use annotations for liquid storage suitability, seal-ability, and transparency

Robotic Affordance Datasets

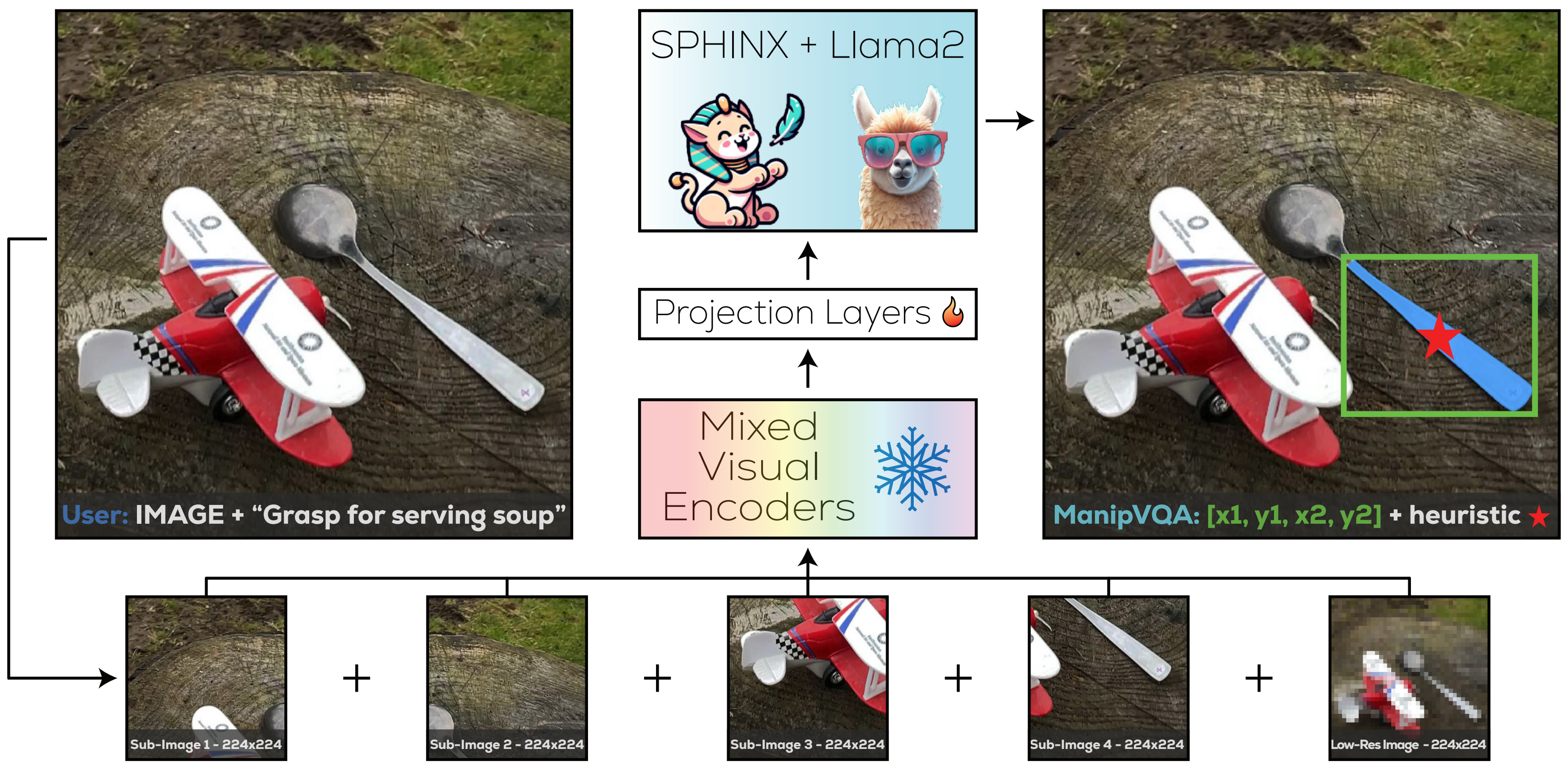
HANDAL Dataset
212 hardware and kitchen tools with annotated handle locations

RGB-D Part Affordance Dataset
105 kitchen, workshop, and garden tools with 7 pre-defined affordances

×

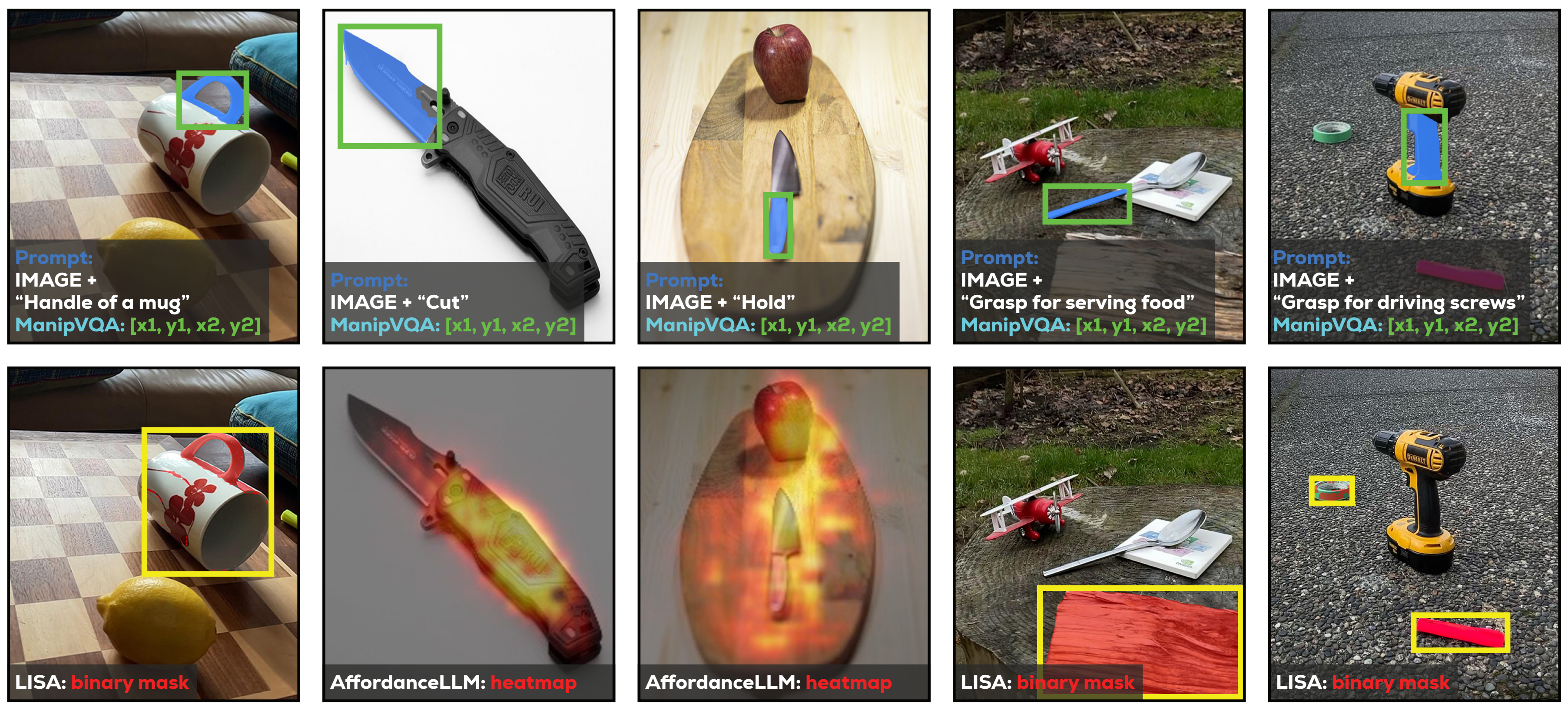
Augmented Instructions with GPT-4
Contextually rich affordance-based tasks

3 Built on SPHINX and Llama2



Experiments

ManipVQA Outperforms Previous Models in Robotic Specific Vision Tasks



上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

北京大學 PEKING UNIVERSITY

香港中文大學 The Chinese University of Hong Kong

Technical University of Munich

IROS '24 ABU DHABI