澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

DALL-E-Bot: Introducing Web-Scale Diffusion Models to Robotics

DOI: 10.1109/LRA.2023.3272516

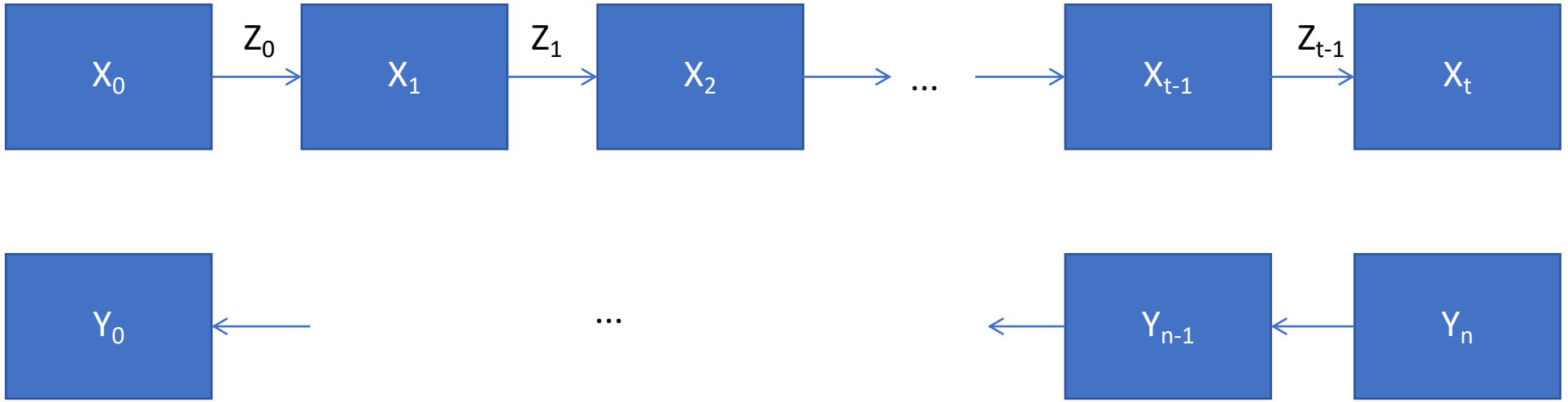
Authors: Kapelyukh, Ivan, Vosylius, Vitalis, Johns, Edward

speaker: Xiongyi Li

Dec 18th 2023

Introduction

- Diffusion Models

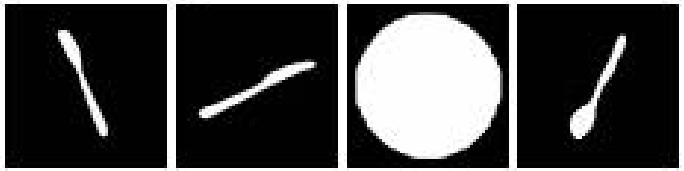


Methods

- A single RGB image
- Object-Level Representation
  - Mask R-CNN: detect objects in an image and generate segmentation masks and provide a bounding box, a segmentation mask and a class label
  - Through an OFA image-to-text captioning model to get text descriptions of the objects
  - CLIP provides each object crop and gives each object a 512-dimensional visual-semantic feature vector

Initial Observation

Segmentation Masks

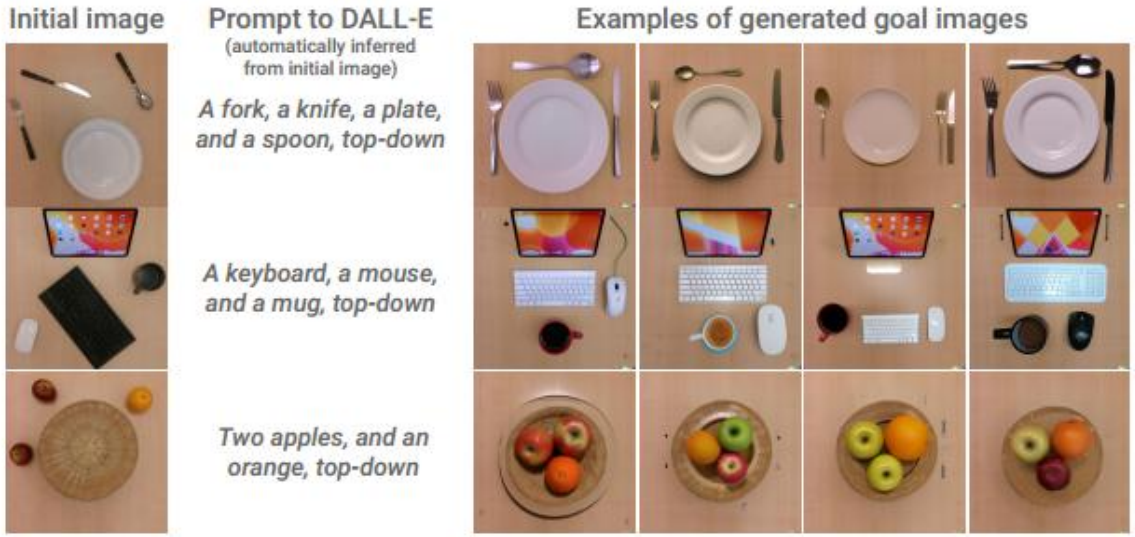


Object Captions

a **fork** with a black handle on a wooden table
a **knife** on top of a wooden table
an empty white **plate** on a wooden table
a **spoon** with a black handle on a wooden table

Methods

- Goal Image Generation
  - Web-Scale Diffusion Model (DALL-E2) provides the conditional distribution $p\theta(I_G|L; I_M)$ & image mask
  - L(text prompt): adding Constraints (top to down)
  - Fixed objects or allowed to be placed
  - Add edge to prevent objects from being placed on the edge



Methods

- Image Selection
  - Select the image whose objects best match those in initial image
  - Use Object-Level Representation&CLIP
- Object Matching & Pose Estimation
  - Use clip to match the objects's vector gained in initial image and generated image
  - Hungarian Matching algorithm

Generated Image

Object-Level Representation

Target Poses

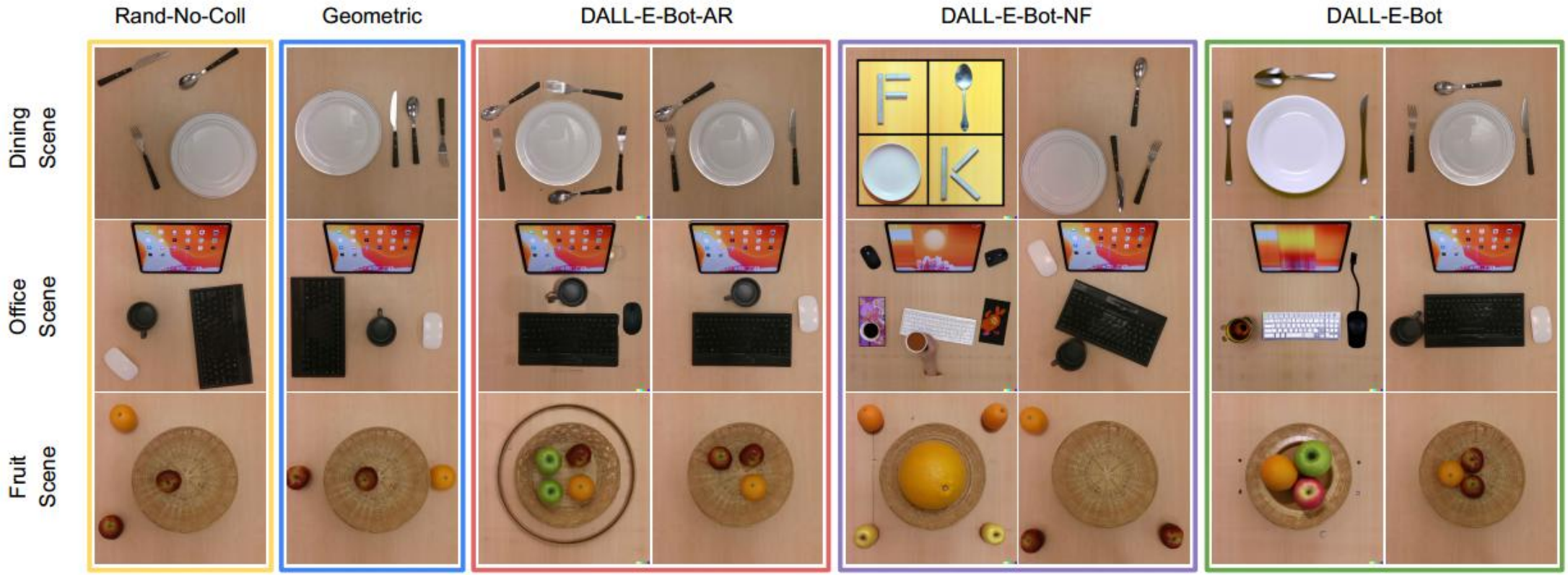
Methods

- Object Pose Estimation
  - Use Iterative Closest Point algrorithm to align two masks

Created Arrangement

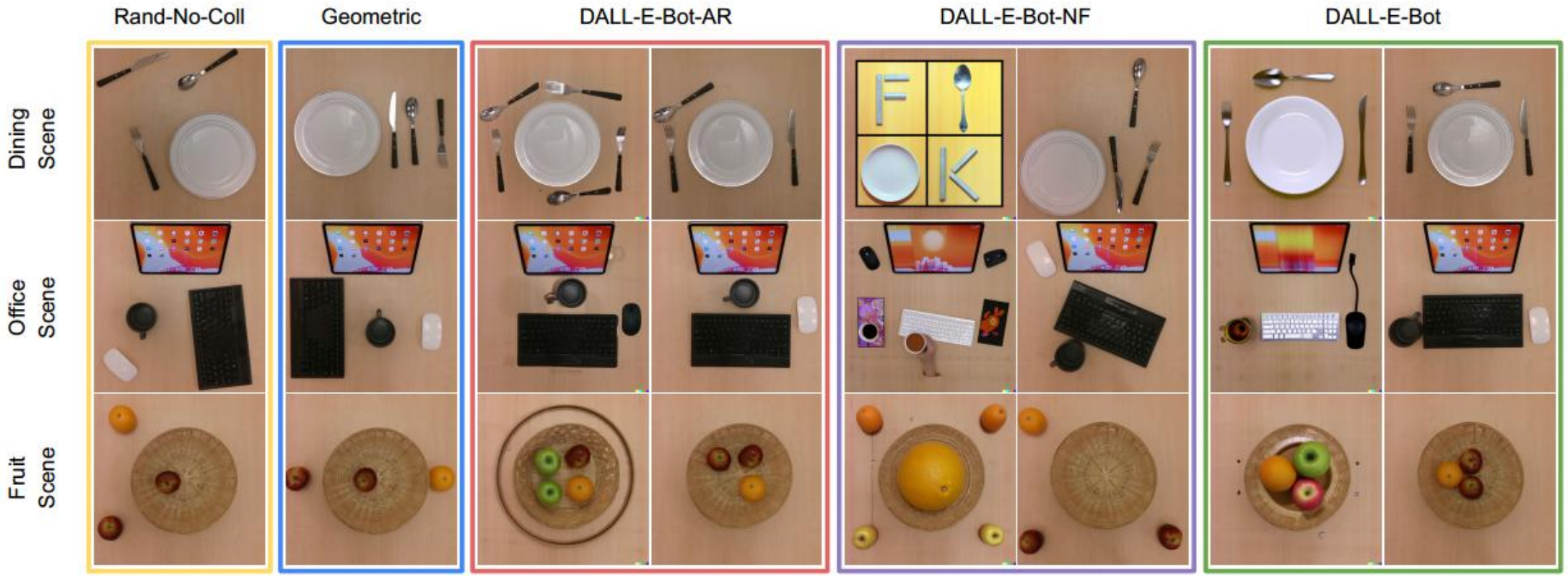
Experiment

- Zero-shot



- Ask human for feedback

Experiment



- Placing Missing Objects with Inpainting

Discussion

- Limitation
  - Top-down pick-and-place
  - Overlap between objects
  - Robustness of cross-domain object alignment
- Future Work
  - Personal preferences
  - Prompt engineering
  - Language-conditioned rearrangement
- https://www.robot-learning.uk/dall-e-bot

Thank You!

Avenida da Universidade, Taipa, Macau, China
Tel : (853) 8822 8833 Fax : (853) 8822 8822
Email : xiongyilee@outlook.com Website : www.um.edu.mo