# Evaluating Estimated Depth for Robotic Manipulation: Comparing Monocular and Stereodepth with 3D-LOTUS++

Kazybek Askarbek

High-quality depth sensors are costly and noisy. Estimated depth models provide a cheaper, promising alternative for robotic manipulation.

#### The problem

To investigate the feasibility of replacing expensive depth sensors with estimated depth models for robotic manipulation.

### The method needs a proper name here

We compare different depth sources (sensor, monocular, stereo) as inputs to the 3D-LOTUS++ model, testing manipulation performance, noise robustness, and generalization.

#### The solution

- Set up a systematic comparison of estimated vs. sensor depth,
- 2) Put mixed training in place to reduce noise sensitivity,
- 3) Get a robust, cost-effective solution for depth sensing.

## Graphical highlights as a form of your message<sup>1</sup>

1) Architecture of SoTA 3D-LOTUS++ model

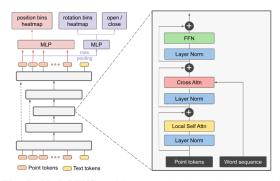


Fig. 3: 3D-LOTUS architecture. It takes point cloud and text as input to predict the next action.

<sup>&</sup>lt;sup>1</sup> Garcia, Ricardo et al. "Towards Generalizable Vision-Language Robotic Manipulation: A Benchmark and LLM-guided 3D Policy." (2024).