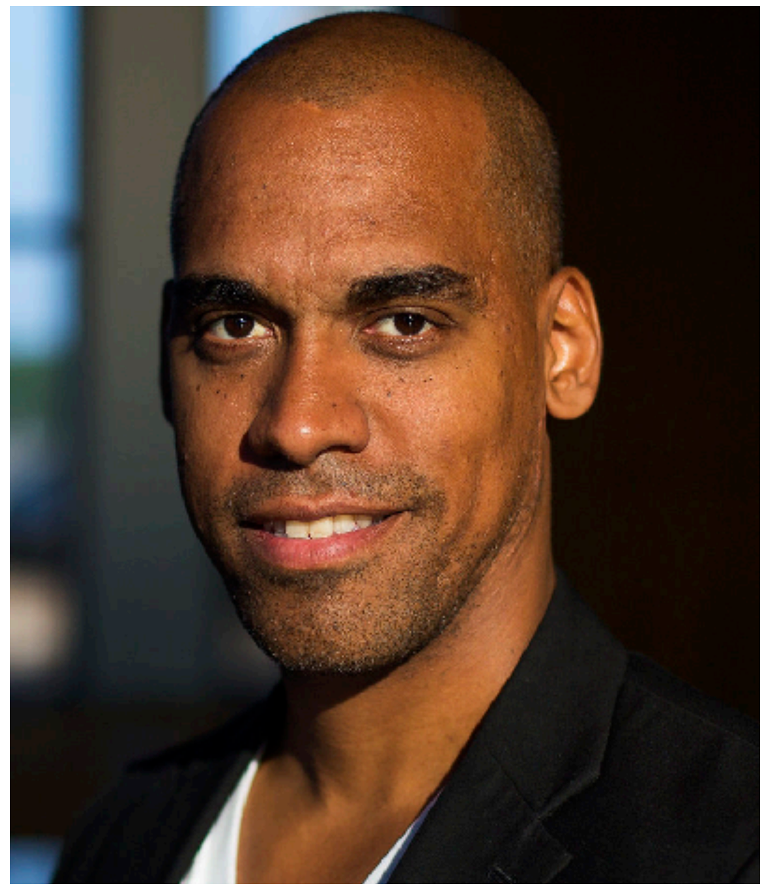
# Shared Autonomy for Unstructured Underwater Environments through Vision and Language


**Matthew R. Walter**
Toyota Technological Institute at Chicago


**Richard Camilli**
Woods Hole Oceanographic Institution



**Matthew Johnson-Roberson**
University of Michigan

# Overview

## Motivation

Underwater intervention is limited by its reliance on humans to perform low-level teleoperation and simultaneously interpret sensor data via manned submersibles or remotely operated vehicles

Challenges of the underwater environment make marine science a surrogate for the many domains that would benefit from shared autonomy in complex, unstructured environments

## Research Goals

Learn representations that capture the complementary nature of diverse modalities including language, vision, and human control for manipulation in unstructured environments

Realize a shared autonomy framework that reasons over these multiview representations to achieve safe, efficient, and effective human-collaborative manipulation in challenging environments

## Current Progress

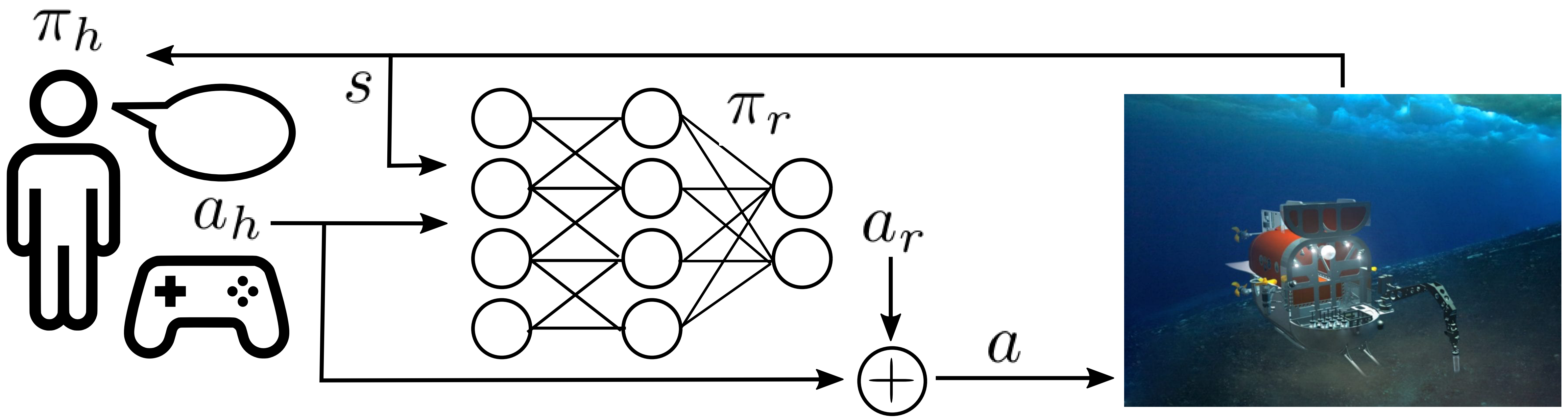
Developed an algorithm that accurately estimates an object's 6DOF pose from underwater monocular images

Developed residual control-based model-free shared autonomy algorithm that requires no knowledge of the task or the goal

Developing a framework that achieves accurate position control of power-efficient hydraulic arms in the face of hysteresis

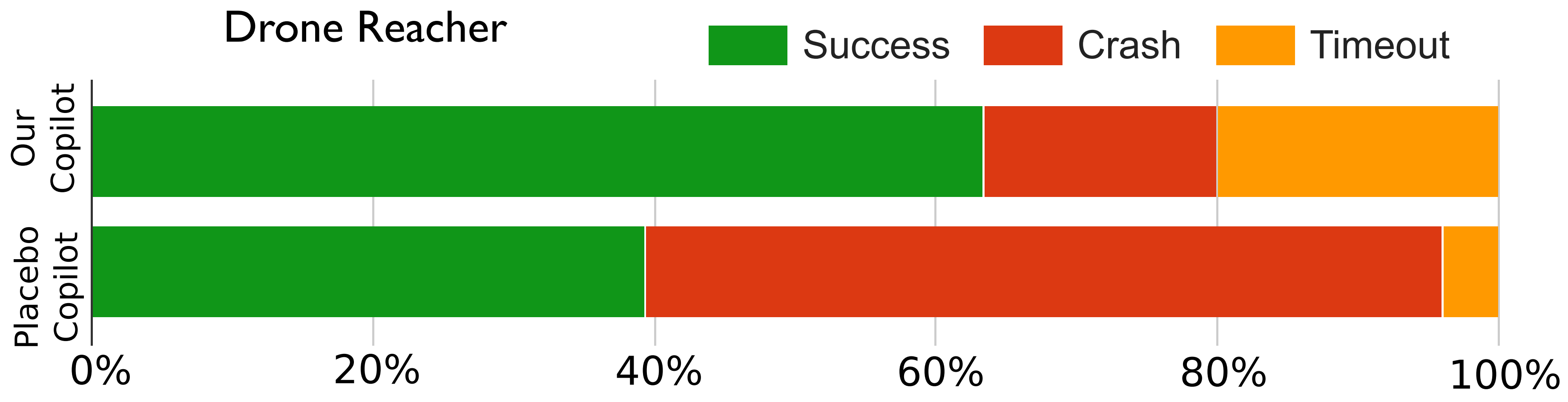
# Current Progress

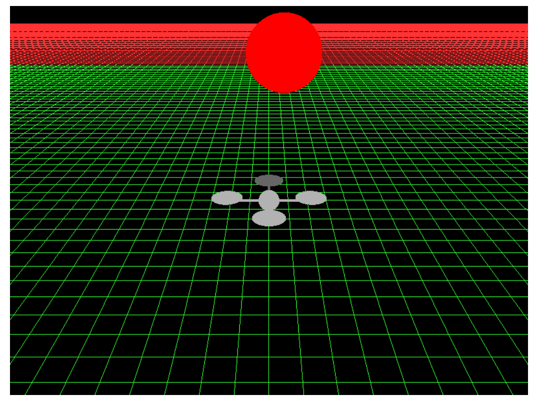
## Residual Policy Learning




Shared autonomy algorithm learns a residual policy without knowledge of the task or of the model

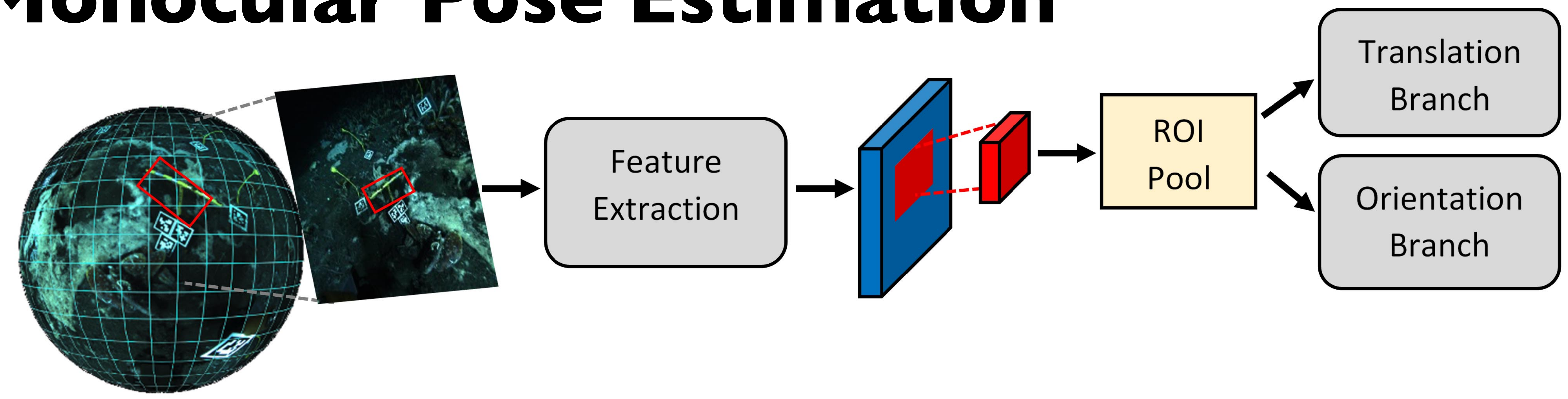
The residual policy minimally corrects human input as necessary to ensure that a task-agnostic reward (or safety) constraint is satisfied



Algorithm significantly improves quantitative and qualitative performance on challenging control tasks

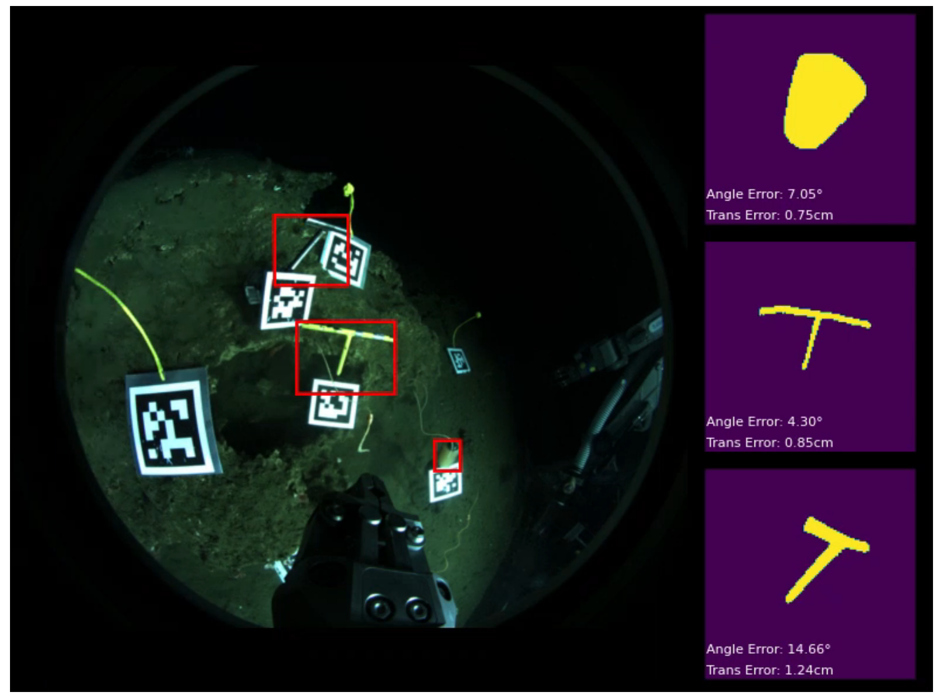
## Monocular Pose Estimation

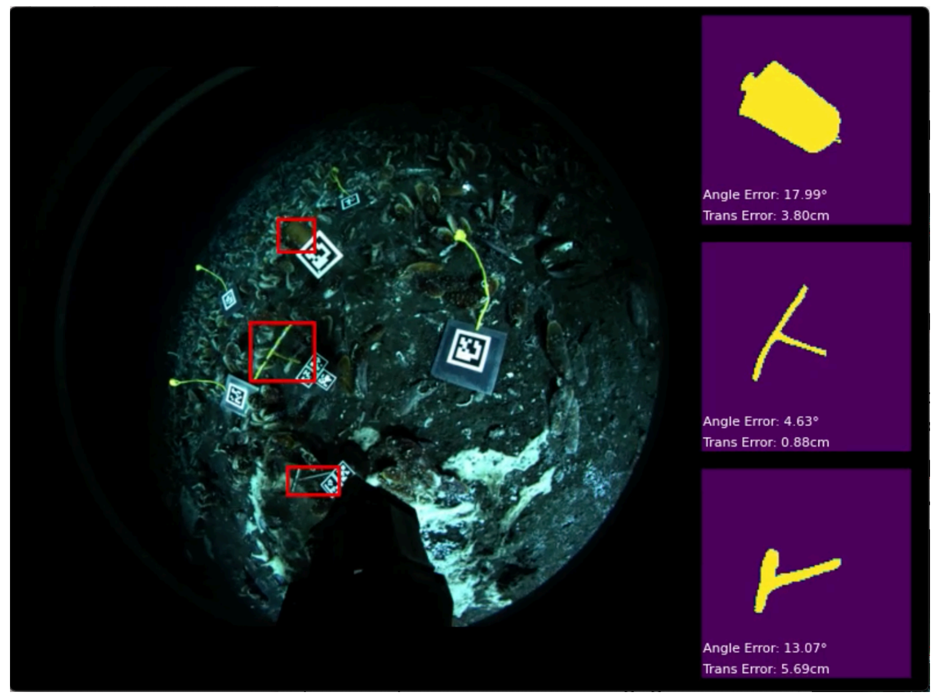



SilhoNet-FishEye estimates the 6DOF pose of objects based upon full fisheye images, despite the significant distortion

Projects region-of-interest in fisheye image onto virtual perspective image prior to position and orientation regression

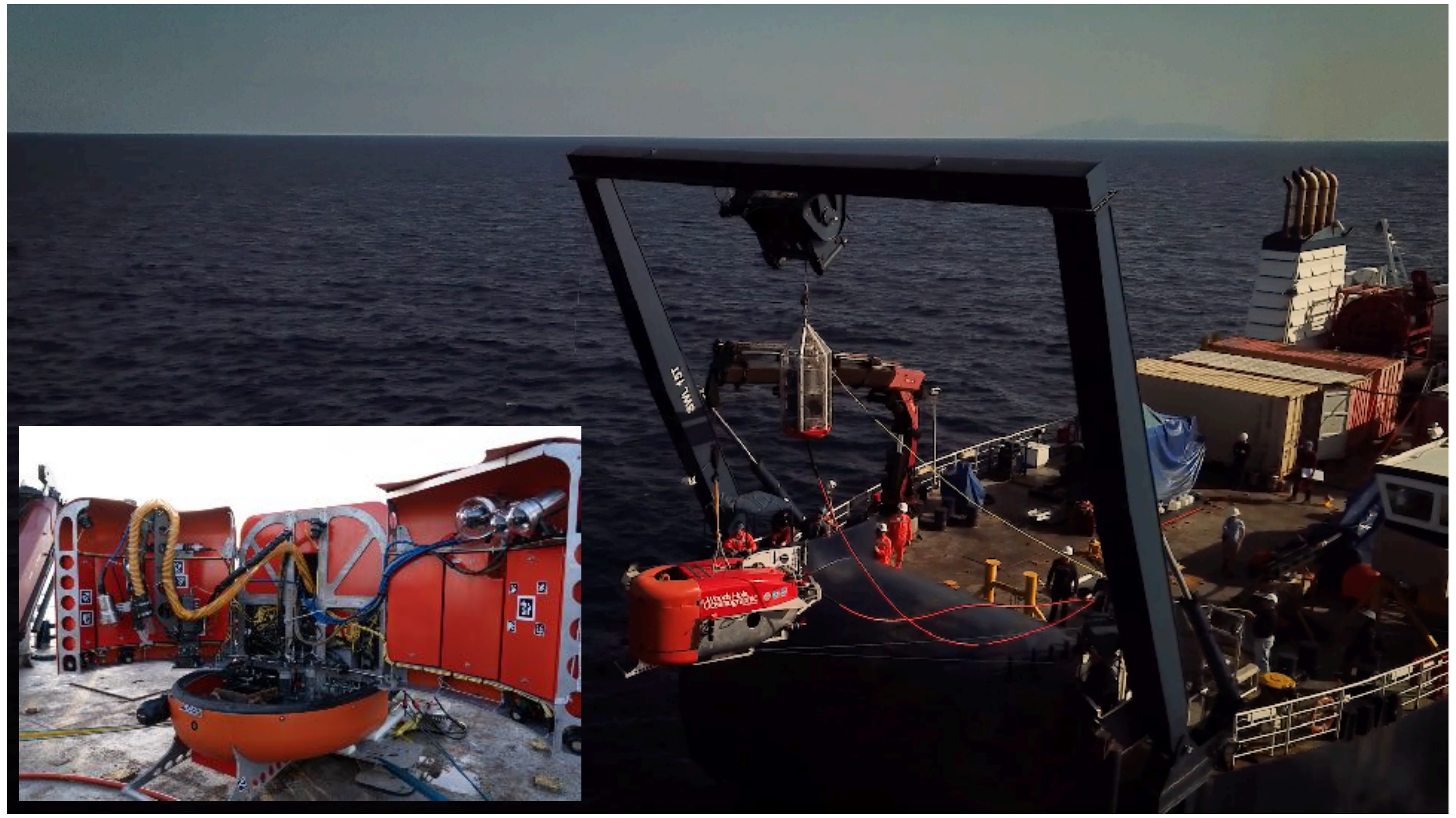
Key to SilhoNet-FishEye is its intermediate prediction of the object's silhouette, which is used to estimate orientation





# Current Progress and Next Steps

## Real-World Demonstration

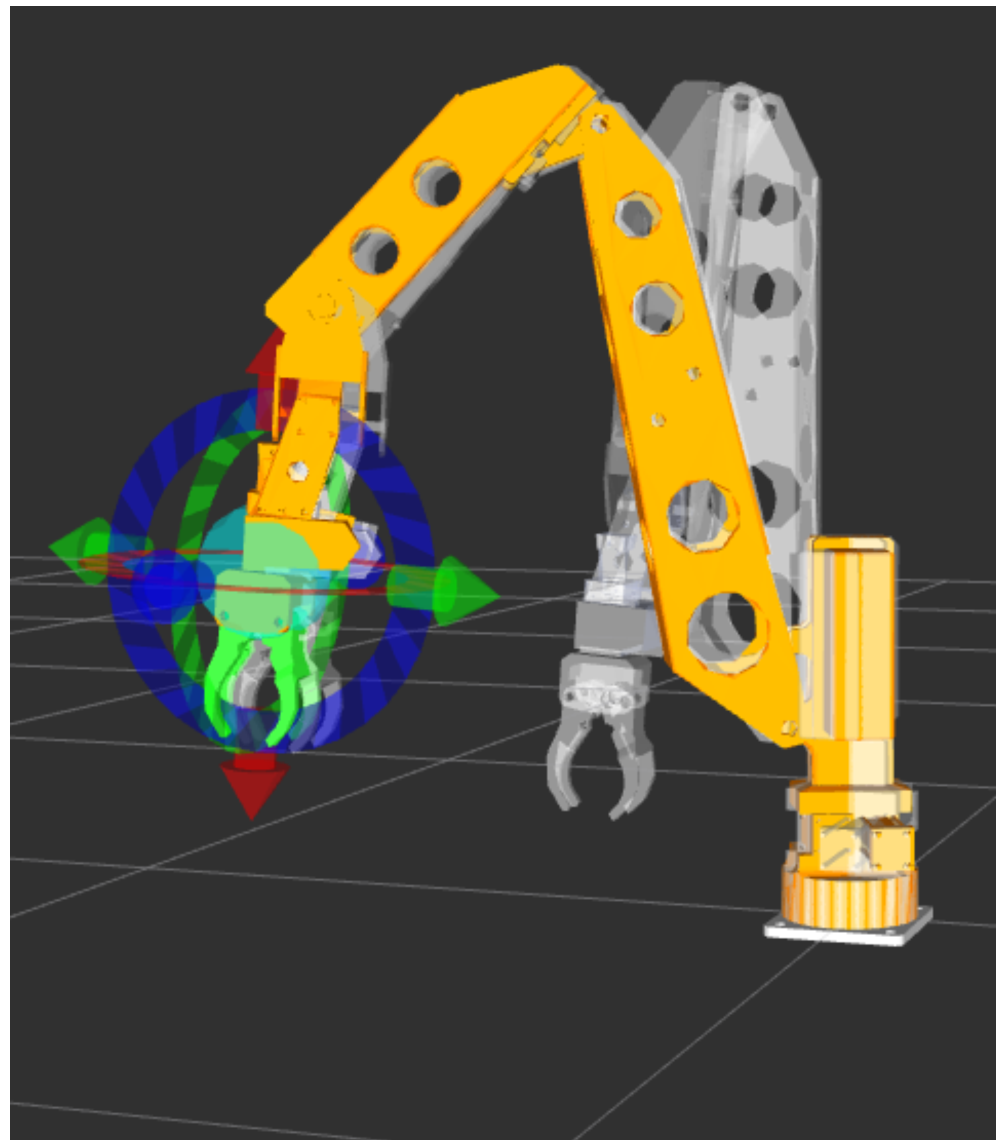


Science cruise in Aegean Sea studying hydrothermal venting sites at the Kolumbo submarine volcano off Santorini, Greece

Nereid Under Ice HROV equipped with hydraulic arm, fisheye camera, and stereo pair, with all power carried on-board

Successfully demonstrated shared autonomy framework that allows pilots to command remote sampling using high-level control, including natural language

## Hydraulic Arm Position Control



Hydraulic arms are far more energy efficient than electric arms, but suffer from hysteresis, which makes position control difficult

Developing a vision-based articulation learning algorithm that provides pose estimates with accuracy similar to electric arms

By tracking compensation curves, the algorithm provides information useful for predicting failure (e.g., when under load)