Physical AI Models for Autonomous Placement and Retrieval of Land-Seismic Nodes Using Robotic Arms

*Pavel Golikov and Ezzedeen Alfataierge**
EXPEC Advanced Research Center Saudi Aramco, Dhahran, Saudi Arabia

ABSTRACT

Land-based seismic acquisition continues to be the dominant method for imaging subsurface structures in desert environment, due to its extensive coverage and proven reliability. 3-D seismic projects typically demand substantial capital, forcing operators to constantly weigh costs against the quality of the data obtained. In recent years, the sector has moved away from conventional, time-intensive shooting techniques toward more rapid acquisition strategies that often produce blended datasets. Concurrently, the count of sensors per receiver line has been lowered, with wireless seismic nodes become the standard technology (Feed 2008; Dean et al. 2025).

Recent advances in robotics have addressed several logistical challenges and opened the door to further improvements in land-based nodal technology. While autonomous seismic drones (Yashin et al., 2023) may offer a viable solution for extremely difficult terrain, a more conventional autonomous ground vehicle equipped with a robotic arm for sensor placement and retrieval provides a practical balance between cost and efficiency for relatively flat areas.

In this work, we present a laboratory setup and initial evaluation results for Vision-Language-Action (VLA) models for the tasks of land seismic node placement and retrieval. The experiments are conducted using a teleoperation framework built around two open-source SO-101 robotic arms (Figure 1).

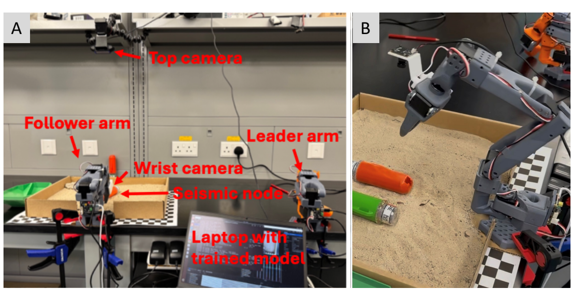

Figure 1. A general overview of the laboratory setup for the teleoperation training (A) and side view of the follower arm with wrist camera (B).

We tested 3 popular VLA models ranged in size and generalization ability: Action Chucking with Transformers (ACT) (Zhao et. Al, 2023), SmolVLA (Shukor, 2025) and Pi05 (Black et al., 2025). The trainable parameter counts vary significantly, from approximately 80 million for ACT to nearly 3 billion for Pi05. Larger and more general models typically require less task-specific data for fine-tuning. However, for a specialized application such as seismic node placement and retrieval in sand, a dedicated training dataset is still necessary.

The dataset was collected using our laboratory teleoperation setup, with an example shown in Figure 2. We recorded multiple episodes of node retrieval and placement, each featuring different initial node positions in the sandbox. Each episode lasted approximately 20 seconds, resulting in a total dataset duration of about 30 minutes. Fine-tuning was performed in a PyTorch environment on a single V100 GPU, with the number of training steps ranging from 100,000 for ACT to 20,000 for Pi0.

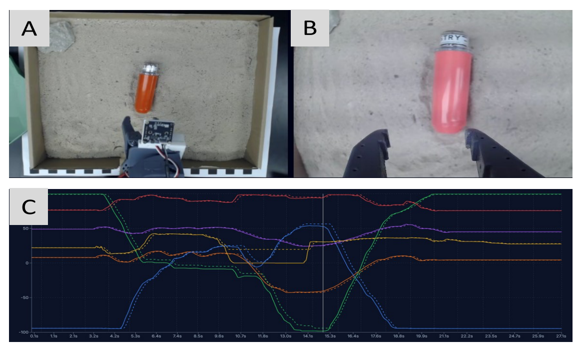

Figure 2. An example of the recorded dataset: video from the top view camera (A), wrist mounted camera (B) and positions of each actuator as function of recording time.

After fine-tuning, each policy was evaluated on the same laboratory setup, with the node placed in varying positions to measure success rates. We also tested several stress scenarios with different objects to assess robustness under noise. ACT achieved about 80% accuracy on simple, well-defined tasks but degraded quickly when unfamiliar objects were introduced. Pi0, the most general model, required the least training and remained stable under noise, though its accuracy on our task was the lowest at 72%. SmolVLA offered the best balance, reaching a 78% success rate with solid stability.

In conclusion, this paper presents a novel approach to evaluate various VLA models for the land seismic node placement and retrieval using laboratory teleoperation setup with two robotic arms.