Speak, Start, See, Sense: How NLP, Robotics, and Computer Vision can Improve Automated Experimentation in Self-driving Labs

Self-driving laboratories (SDLs) are poised to transform and accelerate scientific discovery by enhancing efficiency and reproducibility in experimental workflows. Broadly, SDLs combine automation and autonomy to select and conduct experiments without human intervention. SDLs demonstrate the potential to explore vast parameter spaces systematically and continuously with machine consistency, enabling round-the-clock experimentation that is less prone to fatigue and attention lapses. Although significant strides have been made in advancing the field of SDLs, automated workflows and robotic programming are often constrained by meticulously pre-defined instructions. This restricts real-time monitoring and hinders an SDL from adapting to dynamic and unstructured environments, resulting in rigid automated workflows. Advances in fields like natural language processing (NLP) and computer vision (CV) provide powerful machine learning (ML)-based techniques capable of enhancing the flexibility of automated workflows. Toward that aim, we present a framework that integrates NLP, robotics, and CV for improving the flexibility of automated workflows in SDLs for complex, real-world scenarios, applying the approach to one of the most ubiquitous laboratory tasks: pick-and-place for vial transport.

On one front, we develop the ability to *speak* instructions for a robotic arm to *start* conducting tasks using several natural language processing strategies. First, we combine the Natural Language Toolkit (NLTK), a basic rule-based text processor, with speech recognition in Python. Next, we consider a range of OpenAI Whisper models, based on an end-to-end transformer-based architecture, with varying parameter sizes. Finally, we explore the usage of Google DeepMind Gemma 3, another lightweight, transformer-based, open-source model. For each strategy, the audio is tokenized, and the parsed tokens are then mapped to robotic commands. We benchmark the effectiveness of each approach's ability to process spoken instructions related to laboratory equipment (e.g., vial, beaker, scale) and actions (e.g., weigh, stir, transfer). Furthermore, by incorporating sequence matching using a curated reference vocabulary, we observe an improvement in the performance of these approaches for spoken words, achieving accuracies >90%. Sentence transcription demonstrated superior performance for all NLP strategies, as the built-in sentence context could more easily resolve phonetic ambiguity and auditory misperceptions. These results highlight the feasibility of using NLP approaches in automated workflows, especially in scenarios where local approaches are preferred over cloud-based services that rely on application access keys, depend on rate limits, and may face institution compliance challenges when processing sensitive or proprietary data.

In parallel, we incorporate CV models to *see* experimental tasks and *sense* experimental readiness. Here, we consider varying images of standard 20 mL glass vials in vial racks. The ubiquity of these vials are ideal test objects for evaluating CV models for monitoring handling labware. We trained object detection models on a small dataset (<50 images) covering a minimum of four possible experimental conditions, including one acceptable scenario and three failure scenarios caused by missing/misconfigured lab objects. Image augmentation techniques (e.g., flipping, rotating, blurring, contrasting) expand the dataset to 205 images to achieve class balance and simulate worst-case detection scenarios. YOLO v11 achieves the highest precision (96%) and mAP50-95 score (82%) detecting multiple classes using a minimal training dataset, highlighting its suitability for resource-constrained laboratory settings. We also evaluated the ability of the OpenAI GPT-4 model to discover features and generalize about each class of image, which had comparable performance to the YOLO model suite. Our studies provide evidence that further CV integration could achieve real-time quality control and experimental validation without human intervention.

To demonstrate our pipeline's ability to *speak*, *start*, *see*, and *sense*, we consider a laboratory workflow where a UR3 robotic arm handles labware such as standard 20 ml glass vials, which are widely adopted across disciplines such as materials science, biology, and chemistry for sample preparation. We consider an experimental protocol that simulates routine lab operations involving vial transport such as loading and unloading vials from vial racks. While monitoring successful transfers, we introduce disturbances that simulate errors to be detected such as missing or misconfigured vials. Once an error is detected, the system alerts the human operator by a binary truth value and provides reasoning to inform the user about corrective actions. Together, this integration of NLP and CV demonstrates a step toward improving automated experimentation in SDLs by enhancing adaptiveness and flexibility, which can provide new freedom for designing and conducting experiments in any domain.