Title: Active learning for Autonomous Colorimetric assays design using RGB Dye mixtures and Computer Vision (camera) feedback

Group Members: Alis Bao, David Huang, Thomas Wang, Yeritmary Rodriguez Delgado

Project description: The aim of this project is to develop an autonomous experimental system that demonstrates how active learning can be used to efficiently explore and optimize scientific experiments. For this project, we plan to use RGB dye mixtures as a representation of biological reagents and a computer vision model as the measurement instrument to establish a reproducible process that resembles real-world colorimetric and concentration dependent biological assays. Our system will autonomously generate experiments using a liquid-handling robot such as OT-2, interpret the quantitative measurements from the image frame and select the next experiments or patterns to dispense based on model uncertainty and anticipated performance. This project allows us to bridge the gap between experimental processes, computer vision and machine learning to demonstrate the fundamentals of robotic systems and their potential to lead imaging and data-driven decisions.

What process are you modeling?

We are modeling an autonomous experimental design process for a colorimetric assay using RGB dye mixtures and image frame spectra measurements. We want to simulate how various combinations of three input reagents (represented by red, green and blue dyes) generate a measurable signal in a well, such as absorbance or color distance from a target color mixture. The control reagents will be mixed at the same scale ratio to maintain standardization of dye concentrations when using camera feedback.

This process is a simplified but realistic approximation for numerous biological and chemical assays where reagents are mixed in different proportions, a camera feedback based model can then measure quantitative signal and absorbance.

Our approach, when applied in real biology, allows us to optimize RGB absorbance spectrum in assays, fine tune drug combinations in screening experiments, find the spectra optimal conditions to explore unknown to identify buffers or reagents. Active learning will allow our system to decide which patterns are most beneficial to learn the most information regarding absorbance spectrum and reduce the number of trials and cost, ultimately allowing us to reach good conclusions and autonomous accessible science using an image.

What are the inputs to your model and what are you predicting from those inputs?

The inputs to our model are the volumes of three dyes added to each well:

$$(V_r, V_g, V_b)$$

V_r = volume of red dye

V_g = volume of green dye

V_b = volume of blue dye

Outputs:

Deep Learning Model: A camera to take a photo from a plate and send it to a deep learning model to read RGB values from an image.

Loss Function: An error value that measures if the mixture is near or further than the desired target absorbance value.

Active Learning: The model learns a function determined by us and uses it to decide which mixtures to test next for optimizing absorbance.

Can you find existing data for this process? If not, how can you generate representative fake data?

At present, we are not aware of publicly available datasets that directly capture this calibration process with controlled dye concentrations and corresponding RGB responses under our specific experimental setup. Existing RGB-related datasets typically focus on image classification or color constancy rather than precise concentration-response calibration.

We can generate representative data by simulating the relationship between dye concentration and RGB channel responses using physically informed assumptions (e.g., monotonic attenuation, channel-specific sensitivity, and noise models). These synthetic samples can serve as an initial training set for model validation and pipeline testing.

Describe one iteration of your experiment.

For our setup one iteration may consist of the following steps:

1. Experiment selection

Our model uses active learning to choose a set of dye volume combinations that may be the most efficient for optimizing and standardizing absorbance spectrum.

2. Automatic experiment dispense (OT-2)

The robot dispenses the selected volumes of red, green and blue dyes into the assigned wells of a plate in different patterns and ratios.

3. Data generation

We use an image and computer vision to obtain quantitative values for each well regarding RGB volumes and mix ratios.

4. Data interpretation

The quantitative data obtained is added to the training dataset and the predictive model is continuously updated.

5. Next experiment selection

We use the updated model and uncertainty value estimates so Active Learning can select the next set of dye combinations to run and test.

These steps will be repeated over multiple cycles (minimum of 5) to achieve better results.