Using computer vision techniques to track and quantify features of external representations of biochemical phenomena

Shushruth Kallutla and Thomas Bussey
Department of Chemistry and Biochemistry, University of California San Diego

UC San Diego

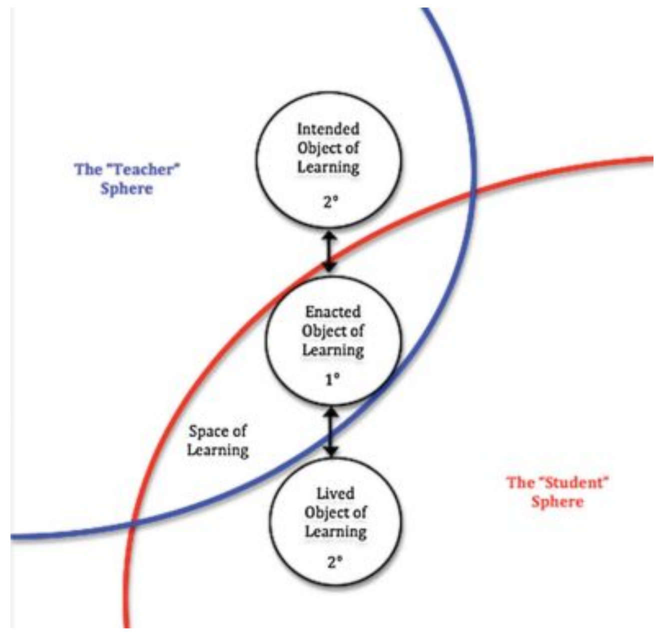
Motivating Change

- Visual representations of biochemical processes are important tools for biochemistry educators to build mental representations of various phenomena.[1]
- One key metric to assess the effectiveness of such representations in building understanding is saliency, i.e., the features to which an observer would likely pay attention. [2]
- Analyzing saliency has previously been done manually by identifying and tracking the different critical features of an object of learning within a representation, and applying a tracking methodology to quantify the saliency of each feature.[2]
- This projects aims to use prevalent computer vision and machine learning techniques to automate this process and create an analytical tool that allows educators to identify regions of interest in representational media.
- We will explore a few algorithmic approaches to achieve this and try to answer which method is most appropriate and achieves satisfactory results.

Theoretical Framework

Variation Theory

- Variation Theory offers a theoretical framework from which to explore:
  - What is intended for students to learn? - Intended Object of Learning
  - What is possible for students to learn - Enacted Object of Learning
  - What students actually learn about an object of learning - Lived Object of Learning[3]



- An individual's experience of a given phenomenon depends on the particular set of features to which they attend. In order to experience a phenomenon in a particular way, an individual discerns and assigns meaning to certain aspects of that phenomenon.
- This project aims to create a resource educators can use to gauge the enacted object of learning by identifying salient objects in educational media. Thus bridging the gap between what educators want to students to learn and what students actually learn.

Tools and Techniques

python OpenCV

OpenCV is a highly popular open source Computer Vision Library with high functionality. This project uses OpenCV implemented in Python for various image processing steps.[4]

YOLO

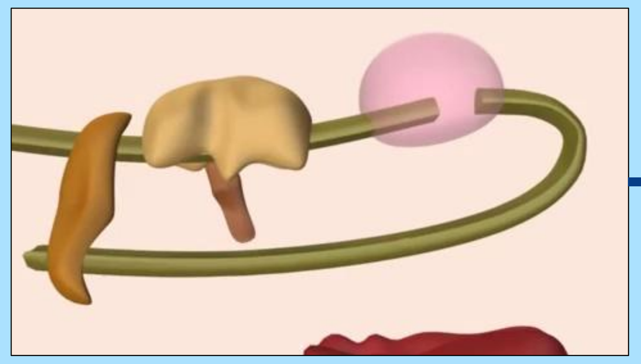
YOLO - "You Only Look Once" is a popular object detection algorithm that utilizes an end-to-end trained neural network to predict and detect object in an image[5]

Methodological Development

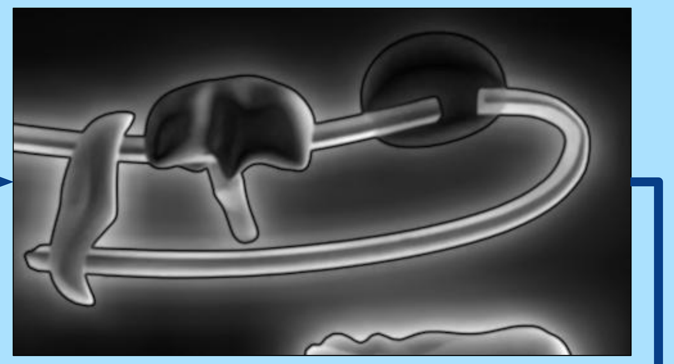
Preprocessing

We first use OpenCVs Saliency API to construct a map of the salient features on screen. The brightness of a pixel maps how salient it is.

Original Video/Image



Fine Grained Saliency Map



Detect Regions of Interest in Frame

We must then identify the specific regions of interest in frame. We have two approaches of doing this.

Approach 1 Identifies Contours of the frame and fits bounding boxes over sufficiently large blobs.

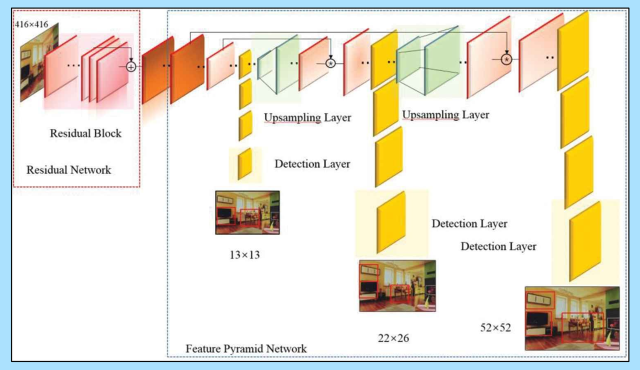
Approach 2 Uses a neural network trained on biochemistry educational videos to identify objects and also computes a confidence score for each object.

Approach 1
Contour Detection

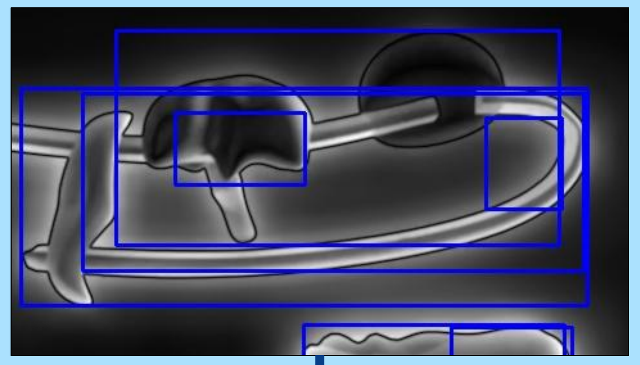
Approach 2
YOLO v3 Object Detector

Contours In Frame

YOLOv3 Network



Regions of Interest detected using Contours



Regions of Interest detected using YOLO

Analyze Saliency of Deteted Regions

Now that we have a collection of regions of interest, we assign a saliency score to each identified region.

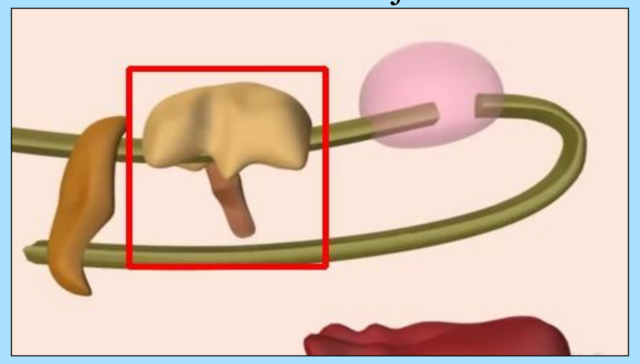
We take the region of interest and analyze it for certain features. We then compute a score based on these features.

Analyze features using OpenCVs built in functionality → Color / Size / Motion / Position → Multiply raw scores with a set of predefined weights → Obtain final saliency score

Identify Most Salient Object

We select the object with the highest score and present as the most salient feature in frame

Most Salient Object



Discussion

- Our two approaches achieve satisfactory results but each has their own set of advantages and disadvantages:

	Approach 1	Approach 2
Pros	• Computationally more efficient and thus a lot faster	• Great accuracy
Cons	• Compromises on accuracy as it tends to falsely detect objects in frame	• Significantly more computationally expensive and thus is slower in detecting objects on screen

- Approach 2 uses a neural net object detection system that has been trained on a large dataset of chemistry educational media, it's accuracy tends to depend on how similar the media we are analyzing is to the media that the dataset is trained on.
- One of the goals of this project is to expand this dataset with more diverse images to improve the accuracy of this neural net
- The project currently is built in Python but could be ported over to other interfaces such as Java or C++ or increase speed or performance.

Future Work

- To increase the ecological validity of the program we hope to incorporate real world data into our saliency score algorithm
- We plan to collect eye tracking data from students watching biochemistry educational videos

- We shall then create a neural network to finetune the weights of the saliency score algorithm by training it on this collected dataset.

References

1. Schönborn, K. J., & Anderson, T. R. (2009). A model of factors determining students ability to determine external representations in biochemistry. International Journal of Science Education. 31(2), 193-232.
2. Bussey, T. J., & Orgill, M. (2015). What do biochemistry students pay attention to in external representations of protein translation? The case of the Shine–Dalgarno sequence. Chemistry Education Research and Practice, 16(4), 714-730.
3. Bussey, T. J., Orgill, M., & Crippen, K. J. (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. Chemistry Education Research and Practice, 14(1), 9-22.
4. Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
5. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Acknowledgements

I would like to thank Dr. Thomas Bussey for all his support and guidance in the development of this project I would also like to thank The George C. and Carol A. Lattimer Faculty Research Fellowship and the UC San Diego Academic Senate for funding this project.