Project Insight:

The challenges of developing research tools in active and responsive open source communities

TY BUCKLEY[1], HATTIE PIMENTEL[2], AND DIRK COLBRY[3]
[1]COMPUTER SCIENCE AND [2]COMPUTATIONAL MATHEMATICS, SCIENCE AND ENGINEERING, MICHIGAN STATE UNIVERSITY, EAST LANSING, MI 48824

ABSTRACT

Scientists often gather observations of the world using images and need efficient image annotating tools to extract data from these images. Project Insight seeks to make this process easier and quicker. The project consists of two parts: first, an image annotation package, the Graphical User Interface (GUI); and second, behind-the-scenes machine learning. When Project Insight is completed, data from the GUI will be sent to the machine learning component, which will attempt to anticipate the user's next choices. The machine-learning component will search through algorithm space for the most suitable algorithm.

Jupyter Notebooks are an interface used by researchers to combines code, pictures, notes and other multimedia. Using Jupyter Notebooks, several image annotation tools were created. Currently, the image annotation tools are being moved to a future replacement of Jupyter Notebooks: JupyterLab. Citing security concerns, the JupyterLab team blocked key JavaScript execution in the beta version of JupyterLab. This means that many programs that worked in the original Jupyter Notebooks, including most of the Project Insight programs, are not functional in JupyterLab.

This presentation will cover multiple attempts to retaining functionality in JupyterLab and discuss why the JupyterLab's security block hinders functionality and contrasts the accessible theme of JupyterLab.

BACKGROUND

Researchers often use many different tools to gather data. For some researchers, the data will all come from images. For example, a scientist studying the motions of a chameleon's eye will use image annotation tools to determine the mechanics behind each motion. A widely-used platform to annotate images is Javascript code. This code is efficiently run using Jupyter Notebooks; however, Jupyter Notebooks is becoming less popular because of the creation of JupyterLab.

The JupyterLab team has blocked JavaScript execution. Some developers using JupyterLab have been upset with this decision, since their projects use JavaScript code to obtain results from image data. However, JupyterLab has not changed its decision and continues to frustrate developers who are accustomed to using accessible platforms to annotate images.

OBJECTIVES

Our objectives in creating this project are as follows:

- Move a Jupyter Notebooks image analysis project to JupyterLab
- Determine a replicable method to run JavaScript code in JupyterLab
- Convince JupyterLab developers to consider enabling JavaScript functionality without circumventing the intended usage of JupyterLab

RESEARCH APPLICATION

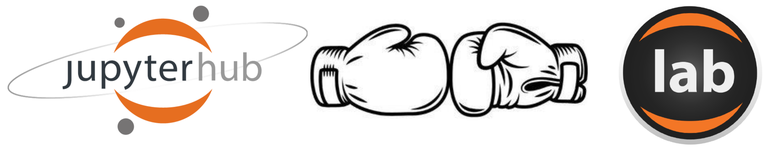

Figure 1. JupyterLab is the new, updated version of Jupyter. Most developers are moving from the older version, Jupyter Notebooks, to JupyterLab. However, Jupyter Lab is not nearly as accessible for projects that use JavaScript as Jupyter Notebooks. The two main issues are the difficulty of using projects containing JavaScript in JupyterLab, and the costs and benefits of switching from Jupyter Notebooks to Jupyter Lab.

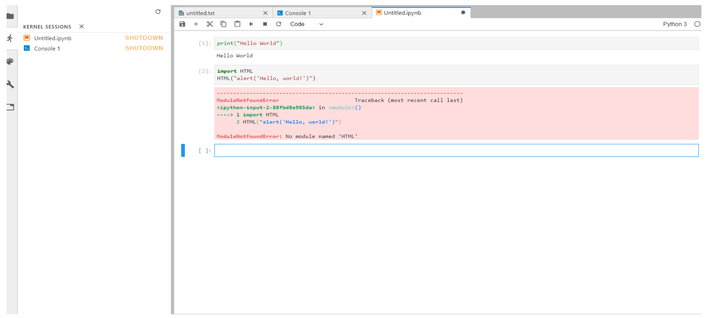

Figure 2. A simple-to-use platform like JupyterLab is ideal for a project using multiple programming languages. Although visually appealing, JupyterLab is not accessible for users who need to use JavaScript. In this figure, JupyterLab does not recognize HTML, the structure needed to run JavaScript code.

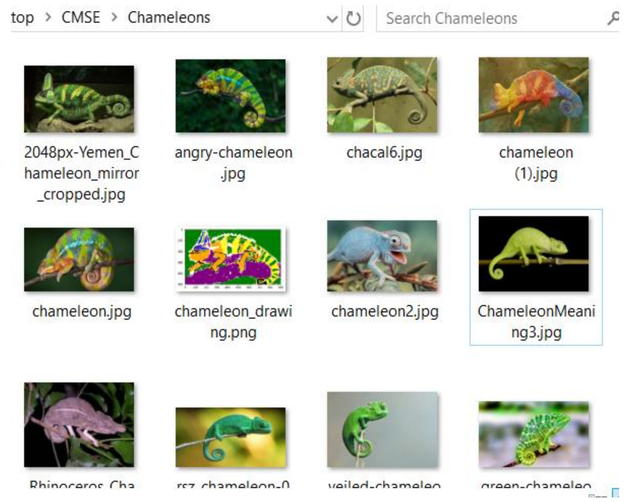

Figure 3. Using an annotation tool, a scientist may be able to measure and compare body part lengths of multiple chameleons.. Using Python and JavaScript code in conjunction with a genetic algorithm, this could become a much more efficient process.

FUTURE WORK

The goal of Project Insight is to create a Python package which makes it easy for scientists to annotate images efficiently. One aspect of this work is to find the most suitable image analsys algorithm to do annotations for the scientists. This would be advantageous because it would save a large amount of time for scientists who need to process many images.

As researchers begin to annotate a series of similar field images, then a genetic algorithm can be used to learn what the they are trying to annotate. Then, the learned algorithm will try to replicate the annotation on a different image, letting the scientist decide if it is replicated correctly or not. If it is done correctly, the scientist would let the algorithm annotate the rest of the images; otherwise, the genetic algorithm will continue to learn.

Figure 4. [1] Genetic Algorithms work much like biological evolution, changing over time to become more suitable to the situation.

A genetic algorithm is a very useful tool to search the algorithm space for image analysis. The concept of a genetic algorithm is very similar to biological evolution. The algorithm starts with many basic replications of the scientist's annotations and saves the ones that represent the original changes for the next generation. Then, after the scientist retrieves data the second time, the best "individuals" from the last generation will "mate". This means that their offspring are many other individuals which combine parts of each of the parent individuals' methods. The "fittest" algorithm is chosen out of the parents and children, and this process repeats until the algorithm is suitable enough for the scientist to let it annotate images by itself.

ACKNOWLEDGEMENTS AND REFERENCES

- Michigan State University Undergraduate Professorial Assistantship Program
- Michigan State University Honors College
- Michigan State University Computational Mathematics, Science and Engineering Department
- Michigan State University Computer Science Department

[1] "DNA double helix." *psypost.org*, https://d2ck0sxsjau14o.cloudfront.net/wp-content/uploads/2015/07/DNA-double-helix-by-Wellcome-Images.jpg