

Overview

Cancer diagnosis is based upon taking tissue from a patient and identifying irregular cells. The practice of using the tissue for a diagnostic purpose is called pathology. Cells are translucent, so stain is utilized as a means to identify structures. Different cell organelles and structures uptake the stain differently. A popular stain used for pathology is Hematoxylin and Eosin (H&E).

The biggest revolution to occur in pathology is the advent of digital imaging, or the ability to use digital versions of these slides for diagnostic purposes. However, the intrinsic variability of biological tissue and the vast differences in tissue preparation protocols lead to significant image variability that can hamper the effectiveness of computational tools.

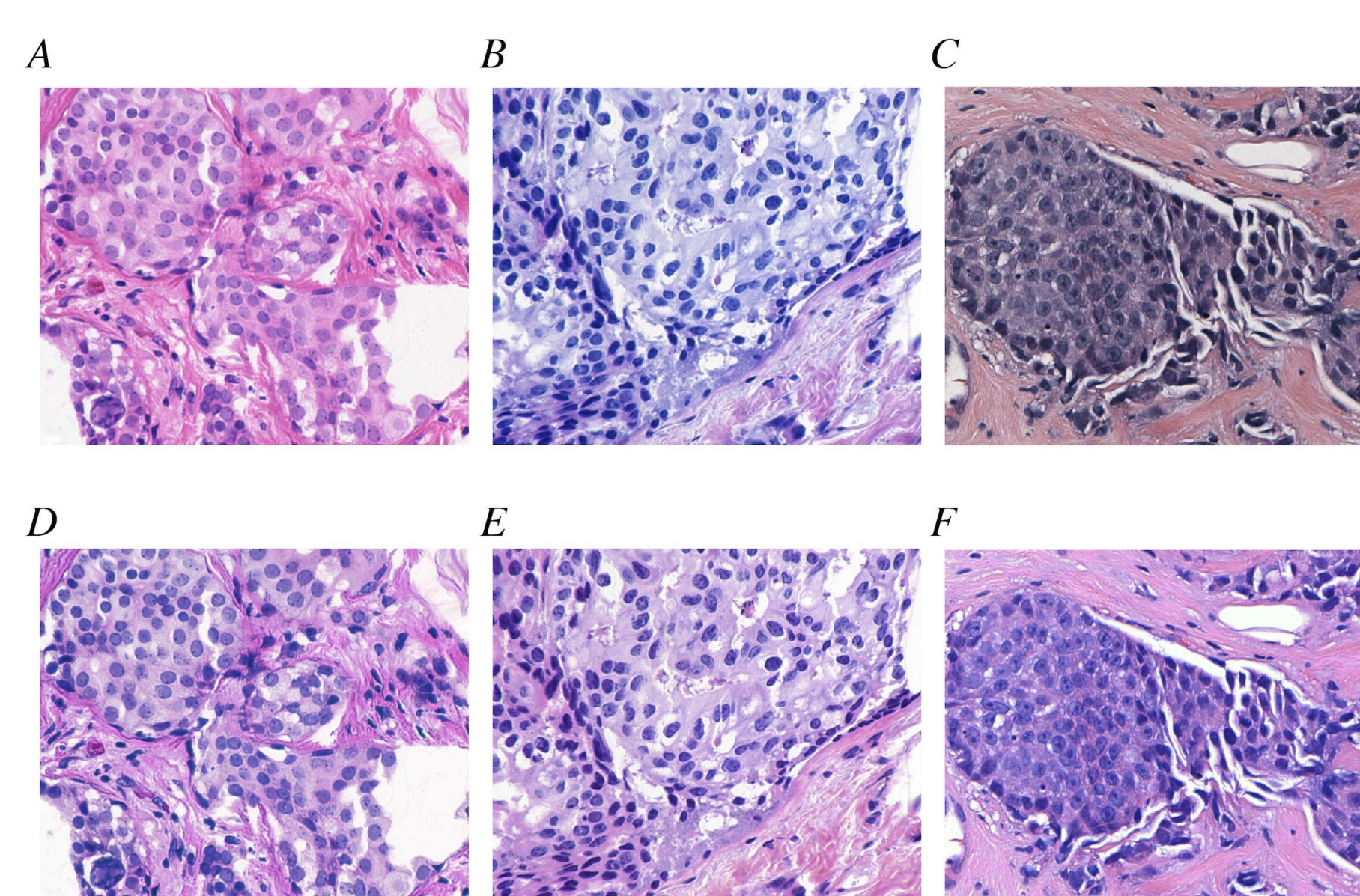


Fig 6. Color normalization for visualization from the 2017 Paper. (A,B) Two images from the UCSB data set exhibited different staining properties. A strong red hue was observed in (A) while the image in (B) had a greater proportion of blue. (C) Likewise, a representative image from the Drexel breast cancer databank exhibited yet a different color property. (D-F) Color normalization produced visually similar color properties

Our laboratory is interested in transformations to normalize pathology images for improved standardization and calibration in the anatomic pathology laboratory. At present, many algorithms have been developed, but there are limitations to their widespread usage, mainly due to documentation.

Key Concepts

Histology: The scientific study of the structure of tissue from plants, animals, and other living things.

Hematoxylin & Eosin: A stain that produces color variation in different tissue structures, which would otherwise be transparent.

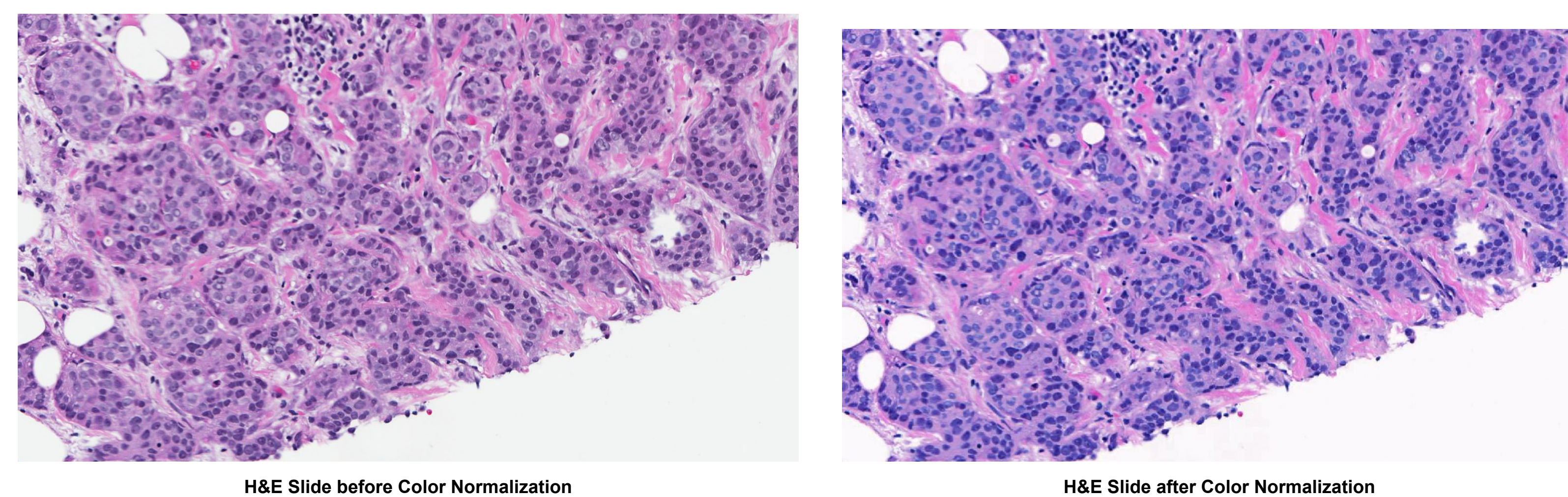
Project Description

Dr. Zarella and Dr. Breen have developed algorithms to identify tissue structures in H&E stained histology images and to normalize the colors in these images. These algorithms accompany several published papers. We are working on several tasks in order to promote the usage of the work and aid pathologists and other researchers in the incorporation of these tools into current workflows and future research.

- reproduce the original results
- updated the software to work with current MATLAB versions
- provide documentation and usage instructions for the algorithms
- integration the workflow with other pathology tools
- produce Python versions of the software

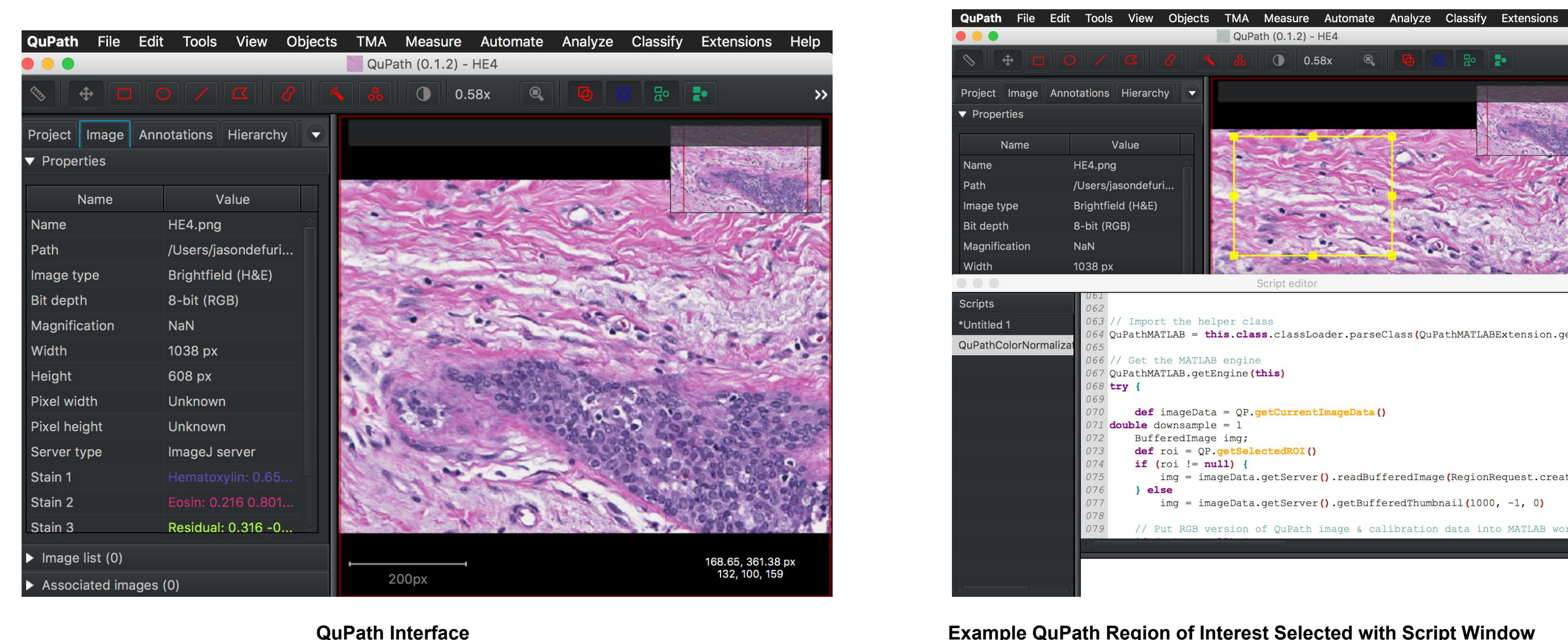
Activities

The first step of the project was reproducing the algorithms to work using the latest version of MATLAB. One of the machine learning algorithms was completely removed, so it became necessary to verify results amongst multiple versions of MATLAB. Below are the results of the rewritten and verified color normalization algorithms:



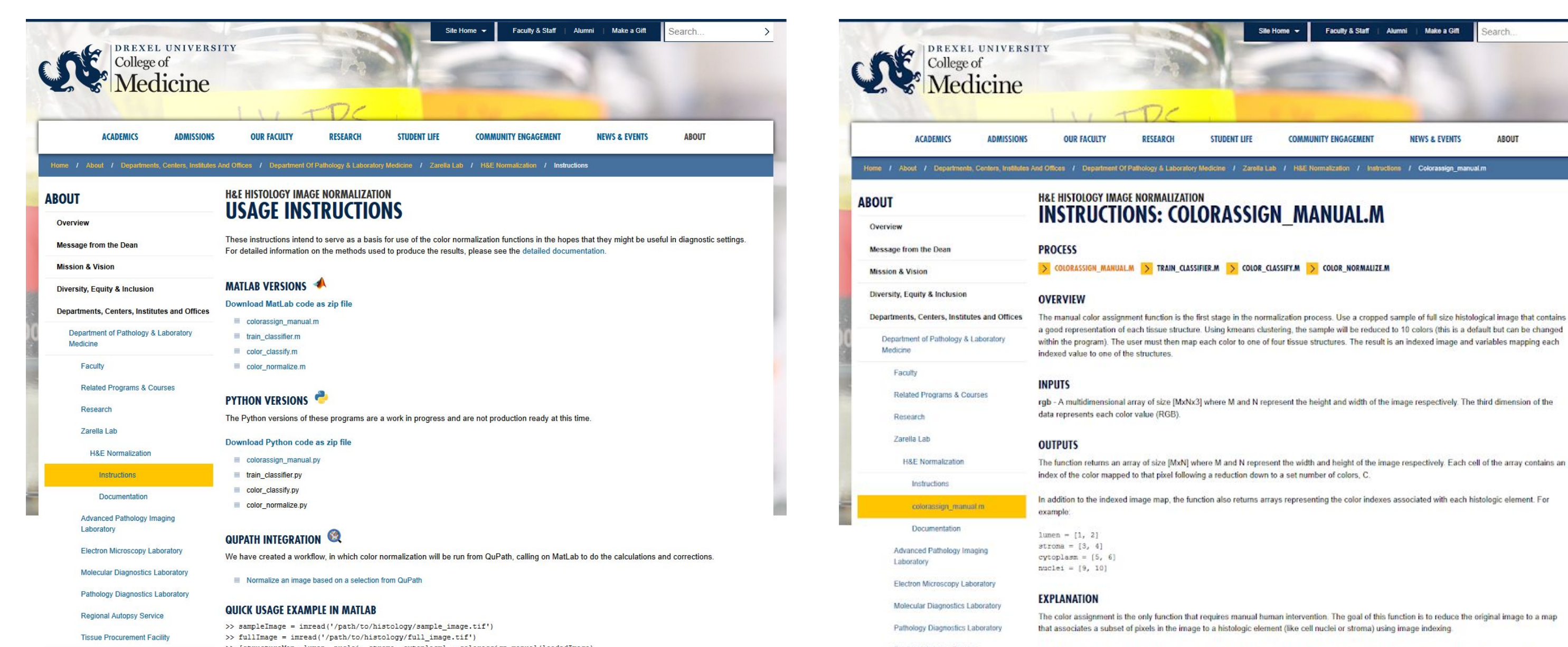
The project then veered off in two directions, but with the same end goal of increasing usage by making the algorithms more accessible.

Jason focused on integrating the algorithms with pathology software known as QuPath. QuPath is an open-source pathology analysis tool that is gaining momentum in the digital pathology world. To achieve this, he wrote QuPath scripts into a workflow so that users of this popular digital pathology tool would be able to easily use the color normalization scripts and output a color normalized image.



Python is one of the most widely used languages for scientific computing, data analysis, and machine learning. Laurence focused on reproducing the software in Python to allow experimentation and investigation by those without access to MATLAB. The goal was to allow new students and programmers to quickly begin using and adapting the programs.

The entire process and the inner workings of each program were documented to a website hosted on the Drexel School of Medicine domain.



Course Materials

Jason:

In my Biology classroom, I will be sharing this algorithm with my students to demonstrate the cell cycle and the process of detecting cancerous cells. It will hinge upon the importance of being able to standardize imagery and the fact that the scientific method is meant to be replicated.

In my Forensics classroom, I will be adding a digital analysis portion to my crime scene unit. I will be filling a few virtual hard drives with data, and students will have to develop an algorithm for sorting the files to be able to determine which are important for discovering the murderer. The victim of the murder will be a pathology student, and images will be hidden in irregular nuclei in the victim's computer.

Laurence:

I intend to produce data sets that can be used with machine learning approaches such as kmeans and SVM in order to demonstrate training an algorithm and using the results to classify further data. In this way, students will have a chance to see that the field (like much of the rest of computer science) is in the understanding and correct application of established tools and principles.

Future Work

Jason:

Currently, QuPath will export the normalized image. I investigated adding the image back to QuPath as an overlay, but unfortunately time is a limiting factor. A further goal would be to remove the dependence on MATLAB. This could be accomplished by either rewriting the scripts in Python or by writing native extensions that can be installed within QuPath. In my classroom, I am also hoping to eventually add an image analysis unit to my Forensics class, specifically of various fibers and hairs with digital image analysis. I need to spend some time refining and researching an appropriate way of incorporating this.

Laurence:

I have successfully reproduced one of four MATLAB functions needed to complete the color normalization of an histology image. Future work would include the completion of the remaining three functions. Additionally, the first stage of the normalization process relies on a manual process to identify tissue structures. Ideally, this identification would be completed algorithmically. Machine learning techniques would be a good fit for this type of problem and the automation of the process would be both a non-trivial and an extremely useful feature.

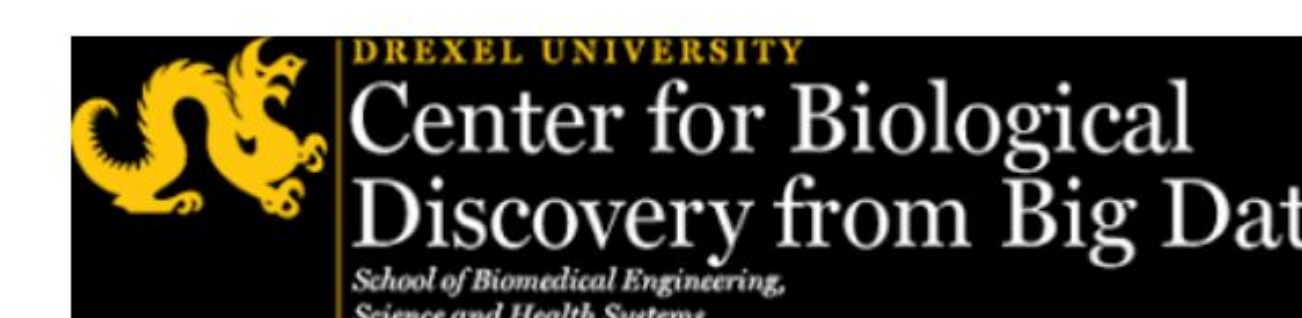
References

1. An alternative image representation for the reduced impact of H&E staining variability*
Zarella MD, Yeoh C, Breen DE, Garcia FU
PLoS One 12(3):p. e0174489 (2017)
2. *An optimized color transformation for the analysis of digital images of Hematoxylin and Eosin stained slides*
Zarella MD, Breen DE, Riegov A, Garcia FU
J Pathol Inform 6(1): 33-41 (2015)

Acknowledgements



Cancer Treatment
Centers of America



REthink@Drexel

Research Experiences for Teachers Site for Machine Learning to Enhance Human-Centered Computing

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1711773. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

