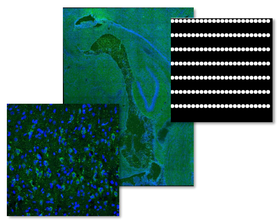
Protein Characterization and Classification in Pre-Clinical Neural Images

Julia Boese, Hawley Helmbrecht, Sage Scheiwiller, David Shackelford
DIRECT, Department of Chemical Engineering, University of Washington Seattle

Motivation



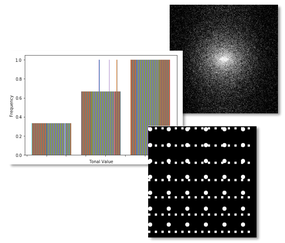
Purpose: Neural and protein images hold a wealth of information. Python packages can be used in order to maximize the amount of quantitative information obtained from every image. Data science methods can also be used in order to classify images into categories.
Goals: The Rockstar Lifestyle package hopes to classify proteins based on their patterning using neural networks and image transforms.

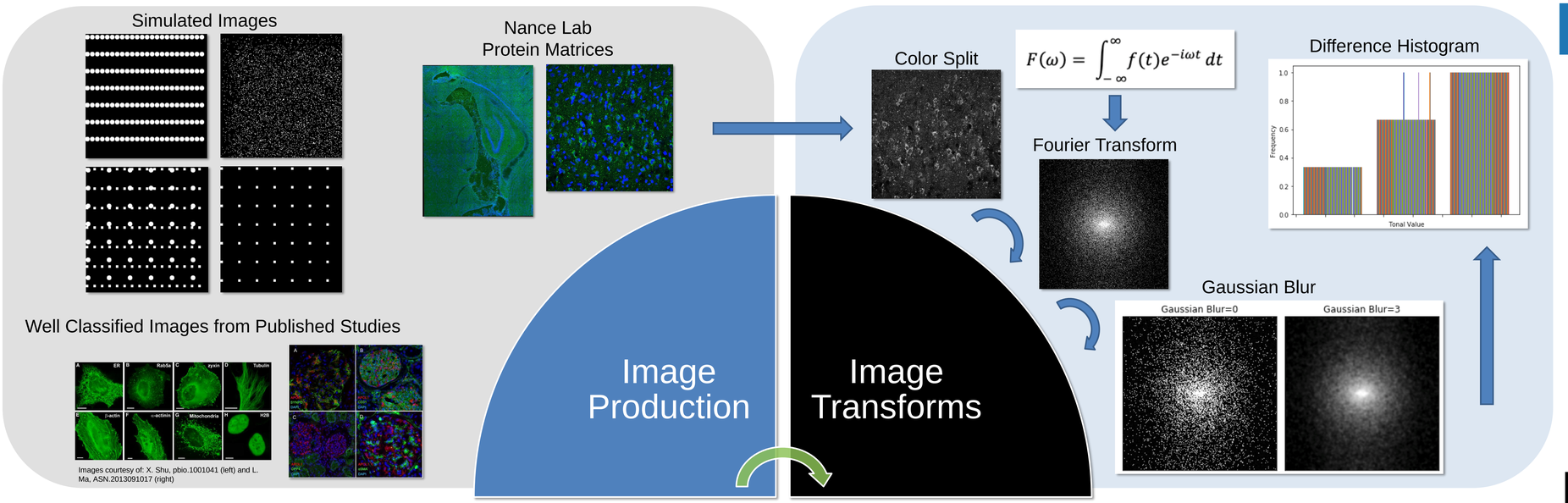
Use Cases

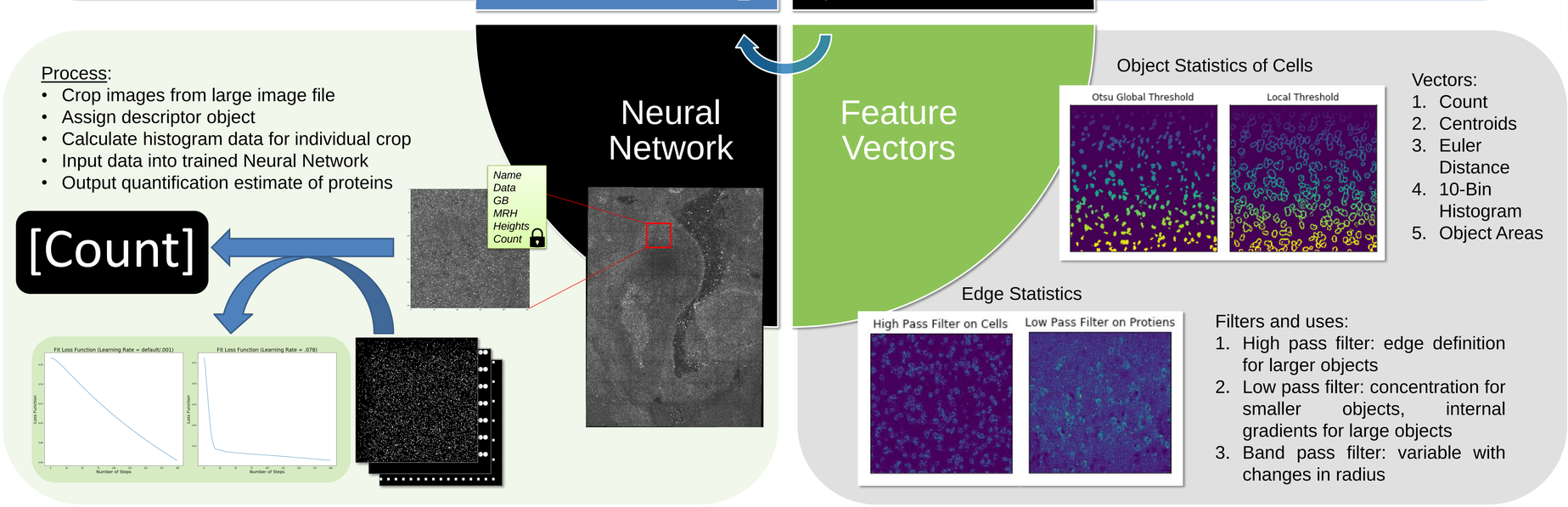
- Filter images for use in Fourier Transforms for localization data.
- Creation of multi-resolution histogram to describe further protein patterning.
- Use edge and object statistics to classify objects and their localization patterns.
- Generate a set of images to train the neural network in charge of classifying images.
- Classify protein domains for a given image supplied by the user based on the trained model.

Data Science Methods and Components

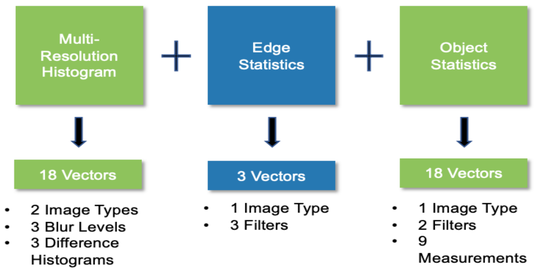
Image processing: Through the use of Fourier transforms which allows us to access the raw data behind each pixel, we can manipulate images to return images with greater qualitative impact. We can also take this data and produce quantitative data by using multiresolution histograms.
Neural Networks: The framework of algorithms work together to automatically process images and help researchers quantify images with ease.







Conclusions



Neural Network: Accuracy of quantifying images was about 14%. The reasoning for this may be due to the size of the training data to create a proper fit. Since the neural network only uses one-dimensional statistics (the protein count of the image) for learning, a large number of training images needs to be supplied with corresponding counts to produce an accurate output. Due to the computational time needed to collect training data, the neural network may not have been trained to a proper degree.

Future Work and Modifications

Future Component Additions:
1. Wavelets
2. Image Moments
3. Advanced Patterning from the Multi-Resolution Histograms
4. Improvements on neural network (implement monte carlo, morphological analysis, variational optimization)

Future Modifications:
1. Multi-Resolution Histogram move from for loops to better use of NumPy arrays and indexing
2. More characterization efforts
3. Object Stats on Proteins
4. Computation time improvements
5. Neural Network accuracy improvements

Acknowledgements

Mike McKenna and Nance Lab for generously donating some of their protein matrix images.
Dave Beck, Chad Curtis and the CHEM E 545 and 546 TAs that answered our endless questions and helped us fix many problems.
All packages used within this project are open source.
More information and documentation can be found on our GitHub at http://https://github.com/hhelmbre/Rockstar-Lifestyle

DIRECT
Data Intensive Research Enabling Clean Technologies

CLEAN ENERGY INSTITUTE

Rockstar Lifestyle

UNIVERSITY of WASHINGTON