2-D Cerebral Blood Vessel Segmentation

Bryan Ho
Computer Science
University of California - Los Angeles
Los Angeles, USA
bryanho1996@g.ucla.edu

Kelvin Wong
Computer Science
University of California - Los Angeles
Los Angeles, USA
kelvinwong623@gmail.com

I. Introduction

Strokes are prevalent in today's society. A stroke occurs when the blood supply to part of your brain is interrupted or reduced, depriving brain tissue of oxygen and nutrients. Within minutes, brain cells begin to die. In 2006, 1 out of every 17 deaths in the United States is caused by a stroke (StrokeCenter). This is 140,000 Americans that die from a stroke every year. To treat a stroke patient, the doctor must know what type of stroke the patient is having and the areas of the brain that has been affected (Mayo Clinic). They must do this becauses the treatments vary depending on the location and type of the stroke. One thing that a doctor can do to properly diagnose the patient is to have a 2D segmentation of the patient's brain. With this the doctor can easily analyze the patient's brain and see the affected areas.

Current technology includes the use of computerized tomography (CT) scan and magnetic resonance imaging (MRI) to identify strokes. CT scans take multiple cross-sectional x-ray images of the desired area and can be used to identify irregular blood flow in the brain. MRI uses a series of magnetic and radio waves to detect brain tissue. Both CT and MRIs can use a contrast dye to better identify the flow of blood in the generated images (Mayo Clinic).

The idea of 2D Cerebral Blood Vessel Segmentation involves taking cross-sections created from CT or MRI scans and highlighting blood vessels for better visibility. The reasoning for this is to make the doctor's life easier when analyzing the visual of the 2D segmentation of the brain. This will highlight the blood flow of the brain and they can easily notice the locations of the brain that is affected by a stroke. This will help in increasing the accuracy of the doctor's diagnosis on a stroke patient.

To implement this we will be using machine learning in our program. Machine learning will be used to analyze the image that the program was given and filter out any unnecessary visuals and present what the user wants. The machine learning is needed because the 2D segmentations that the program will received will be different each time. the program must be able to take that into account and adapt to the slight differences to make sure the accuracy does not diminish.

II. Methods

Our machine learning was accomplished with Python's Scikit-Learn. The base code of our program references the a sample face completion algorithm that was available on the SciKit-Learn website and cited below (Face Completion with a Multi-Output Estimators). Using brain angiography scans, we then annotated by hand the blood vessels. Using machine learning, we attempt to complete the image and try to get as close to the annotated one as possible. Images that we used were square to be easily comparable. The algorithm was then trained on a certain portion of the data and tested on another portion. Using the results, we were able to compare the performance of our different machine learning algorithms.

The different machine learning algorithms that we have used comes from the list of algorithms we covered in class. All the algorithms came from Python's `sklearn` package. The algorithms used are as follows:

- **Extra Trees Algorithm**
  This algorithm was imported from the `sklearn` package. The Extra Trees algorithm constructs decision trees based on the input data and chooses the best one. In the end, it chooses the tree that occurs the most, leading to the most accurate result.
- **Nearest Neighbors Algorithm**
  This algorithm was imported from the `sklearn` package. Nearest Neighbors provides a result given the closest `x` number of neighbors to the instance it is trying to solve for.
- **Linear Regression Algorithm**
  This algorithm was imported from the `sklearn` package. Linear Regression uses the basic linear equation $y = ax + b$ to solve for
- **Ridge Regression Algorithm**
  This algorithm was imported from the `sklearn` package. Ridge Regression is very similar to Linear Regression, however in the case that there is not a good match found, it will attempt to bias the variables to find a better match.

III. Evaluation

Evaluation for our results is fairly straightforward. The initial image has a lot of static and extra information that is not essential to identifying blood vessel location. On the other hand, the annotation is free of all static and easy to identify where blood vessels are.

To judge how good a machine learning algorithm is doing for a given input, we visually examine the original image and compare it to the results of each of the four algorithms. The more static is in the result, the less effective the algorithm is. On the other hand, the less static there is in the result, the more effective the algorithm is. As we tested our machine learning, we realized that effectiveness of the algorithm increased given a larger sample of training data provided. This makes sense since it provides more information for the algorithm to learn what output image to return given the input image.

IV. Results

Due to the lack of training and testing data we were able to acquire for cerebral angiograms, we reached out to Professor Scalzo for more images and annotations. Professor Scalzo provided us with 40 original images with their corresponding annotated images. With the retinal scans provided to us by Professor Scalzo, we were able to run our program properly. Our implementation is general enough that it can handle multiple types of scans and learn how to model the blood vessels in it. It is dependent on the data that is given to the program to learn on. If it was given sufficient angiography cerebral scans, it would be able to handle the scans and map out the blood vessels for the given angiography cerebral scans.

V. Discussion

Given the 40 images, we set 35 of the images to be training data and 5 of the images to be testing data. We initially tested on different ratios of training to testing data, then concluded that more training data would allow the algorithm to better predict results on the test data. However, even with our best predictions, the results are not very clear because of the number of training data that we had.

We had also tested with the amount of total data available and realized that with more data available, the accuracy of our images would increase. With a larger set, maybe of several hundred or even a few thousand retinal scans, we would be able to provide much better estimations for the results.

Extra Trees Algorithm was able to provide the best prediction because of its nature. It introduces a randomness factor into the inputs which lead to better accuracy on the output. Nearest Neighbors Algorithm has some inherent issues. With a low value for `k`, the algorithm will likely underfit on the training data and with a high value of `k`, the algorithm will likely overfit on the training data. Both of the regression models, Linear Regression and Ridge Regression did not perform very well and could likely not do much better even with more training data to work with. This is simply because these two algorithms do not perform well on datasets with noise.

In the future, if we were able to expand our Python program to allow for better predictions we would train with a larger data set and explore algorithms that handle noise better.

VI. Conclusion

Our Python program was able to properly generate a blood vessel drawing given an input image but not perfectly due to imperfections in the machine learning algorithms and lack of enough test data. By giving our program more test data, we can produce a more accurate depiction of the blood vessels in a given scan. This technology can be used in aiding doctors when diagnosis a specific patient. Whether it being diagnosing a stroke victim or issues with the blood vessels in the eye, our technology can be used to lend a helping hand.

VII. References

"Face Completion with a Multi-Output Estimators." *Face Completion with a Multi-Output Estimators*, scikit-learn.org/stable/auto_examples/plot_multioutput_face_completion.html.

"The Internet Stroke Center." *The Internet Stroke Center. An Independent Web Resource for Information about Stroke Care and Research.*, www.strokecenter.org/patients/about-stroke/stroke-statistics/.

"Stroke." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 24 Apr. 2018, www.mayoclinic.org/diseases-conditions/stroke/diagnosis-treatment/drc-20350119.

"Stroke." Mayo Clinic, Mayo Foundation for Medical Education and Research, 24 Apr. 2018, www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113.