Project Name
SaVANNA: Sparse Vectors Applied to Neural Net Architecture

Team members

1) Bijan Varjavand
2) Ryan Lu
3) Matt Figdore
4) Alex Fiallos
5) Stanley Wang
6) Victor Wang

Description of the project and the problem to be solved:

Novel and meaningful neuroscientific data sets are being developed at a high rate thanks to recent improvements in neuroimaging methods. The task of processing data such as brain MRI volumes to detect signs of disease or investigate brain anatomy and structure is becoming increasingly critical to progress understanding. A significant data set which has been recently published by Janelia, known as Mouselight, has great potential for neuroscientific progress as it maps axonal projections in the mouse brain on the scale of single-neurons as opposed to populations of neurons[1]. Our group has gained permission from Janelia to work with their Mouselight data, and we hope to improve the performance of automatic axon detection. We will accomplish this by building an open source Python package, which will allow other groups to use the methods we develop for their own research.

Mammalian brains contain projection neurons that connect across regions. With these mappings, neuroscientists can better understand information flow and test hypotheses relating brain function to network structure. While Mouselight contains mappings for 1,000 individual mouse neurons, a full characterization of connectivity in the mouse brain will require mapping out at least 100,000 neurons. The rate limiting step in producing such a dataset is the manual task of segmenting neurons. Janelia hopes that their Mouselight data will serve as a training set for machine learning based neuron segmentation.

While deep learning models have been developed to solve problems like these, large data sets (e.g. >30TB per mouse brain) pose computational challenges. This is especially important considering that the size of data sets will likely grow in the future - as larger brains are investigated and as resolution increases. An improvement over the efficiency of effective image classification and object detection is very desirable.

Convolutional neural networks (CNNs) are the gold standard for computer vision and image segmentation tasks[2]. CNNs require gradient descent and backpropagation, both very expensive operations; high amounts of training data and computations are necessary to train the network.

Recently, improvements in alternative, cheaper, methods for image classification such as Random Forest make them more viable as a possible contender for deep models. We have already seen these improved methods outperform CNNs on smaller data sets, but they are still limited by the level of complexity that they can capture[3]. Currently, these methods employ random forests (RF) or sparse oblique random forests (SPORF) on either the entire image or predetermined local areas of the image. The output is an array of probabilities, which is treated as an another image and RF/SPORF is applied again.

We aim to develop algorithms to more efficiently classify and detect significant objects in brain volumes. We believe that combining deep learning techniques with recently developed manifold oblique random forests (MORF) will improve classification and detection ability, creating a niche of lower-cost object detection methods which avoid significant computation. Unlike SPORF, MORF preserves the structure of the data, which we believe will improve outcomes. We hope to build on the previous work done on the conjunction of RF's and CNN's to create a package (SaVANNA: Sparse Vectors Applied to Neural Net Architecture) that accepts 3D volumes and can perform both image detection and localization.

We will then apply these algorithms to answer the following questions:

1) How does SaVANNA compare to SPORF/MORF and CNN in existing datasets (e.g. ImageNet, MNIST)?
2) Assuming SaVANNA outperforms CNN on datasets with few samples, above what sample size does CNN outperform SaVANNA?
3) How well does SaVANNA fully automate the task of segmenting neurons from two-photon tomography? Compared to ground truth?

Timeline and tasks for entire team

Sprint 1:

1) Setup the SaVANNA package Github Repository by organizing code from Satish
    a) Create the repository and set up user privileges, Python tests, Travis CI, codecoverage, and Sphinx
    b) Refactor code to be organized and sklearn compliant
    c) Fully document naive version of SaVANNA (2D classification)
    d) Fully test code in naive version of SaVANNA (codecov >80%)
2) Minimum viable product (MVP) package development:
    a) Basic Random Forest Classification
        i) Write a function to classify labelled brain images using MORF
        ii) Extend SPORF(/MORF) to accept 3D volumes as input and submit a pull request to SPORF
        iii) Write a function to classify labelled 3D volumes using MORF
        iv) Publish package on PyPi
3) Real data applications
    a) General applications

        i) Write code to appropriately process MRI images into a normalized format that can be inputted to MORF
        ii) Write code to appropriately process 3D brain volumes into a normalized format that can be inputted to MORF
    b) Mouselight
        i) Assist Tommy Athey et. al. with data retrieval functions from Mouselight data such as
            (1) Crop out image volume centered at a spatial coordinate with specified dimensions
            (2) Convert SWC neuron morphology files into pixelwise neuron segmentations
4) Additional package functionality
    a) Extend SaVANNA to handle segmentation tasks in addition to classification tasks
    b) Extend MORF to work on a sort of convolution on 2D images and 3D volumes (convolutional MORF)
    c) Layer MORF/convolutional MORF with connected layers (akin to a neural net architecture) to potentially extract more complicated features from the images/volumes (SaVANNA)
    d) Run SPORF/MORF on different levels of pooled images to potentially extract more complicated features from the images/volumes (SaVANNA)

Sprint 2:

1) Continue developing SaVANNA
    a) Continue with existing experiments and any new ideas
    b) After SaVANNA can effectively classify and segment existing machine learning data sets (eg ImageNet, MNIST) for some number of samples, apply SaVANNA to preprocessed brain images and volumes. Ideally, SaVANNA should outperform MORF
    c) Submit the updated SaVANNA package to PyPI and Anaconda repositories for public use and easy installation
2) Real data applications
    a) Apply SaVANNA to the Mouselight data set for fully automatically segmenting projection neurons from two-photon tomography data
    b) Find additional data (optionally: process it with *ndmg* to obtain structural connectomes) and use SaVANNA on the registered volumes
3) Graph Input
    a) Extend SPORF/MORF/SaVANNA to take graphs as input
    b) Use graph inputs for classification and segmentation tasks on connectomes

Sprint 3:

1) Publication
    a) Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
    b) Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

Timeline and tasks for individuals on team

Bijan

Sprint 1:

1) Setup the SaVANNA package Github Repository by organizing code from Satish
    a) Create the repository and set up user privileges, Python tests, Travis CI, codecoverage, and Sphinx
    b) Refactor code to be organized and sklearn compliant
    c) Assist documenting naive version of SaVANNA (2D classification)
    d) Assist fully testing code in naive version of SaVANNA (codecov >80%)
2) Minimum viable product (MVP) package development:
    a) Basic Random Forest Classification
        i) Extend SPORF(/MORF) to accept 3D volumes as input and submit a pull request to SPORF
        ii) Write a function to classify labelled 3D volumes using MORF
        iii) Publish package on PyPi
3) Additional package functionality
    a) Extend SaVANNA to handle segmentation tasks in addition to classification tasks
    b) Extend MORF to work on a sort of convolution on 2D images and 3D volumes (convolutional MORF)
    c) Layer MORF/convolutional MORF with connected layers (akin to a neural net architecture) to potentially extract more complicated features from the images/volumes (SaVANNA)
    d) Run SPORF/MORF on different levels of pooled images to potentially extract more complicated features from the images/volumes (SaVANNA)

Sprint 2:

1) Continue developing SaVANNA
    a) Continue with existing experiments and any new ideas
    b) After SaVANNA can effectively classify and segment existing machine learning data sets (eg ImageNet, MNIST) for some number of samples, apply SaVANNA to preprocessed brain images and volumes. Ideally, SaVANNA should outperform MORF
    c) Submit the updated SaVANNA package to PyPI and Anaconda repositories for public use and easy installation
2) Real data applications
    a) Apply SaVANNA to the Mouselight data set for fully automatically segmenting projection neurons from two-photon tomography data
3) Graph Input
    a) Extend SPORF/MORF/SaVANNA to take graphs as input
    b) Use graph inputs for classification and segmentation tasks on connectomes

Sprint 3:

1) Publication
    a) Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html

    b) Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

Matt

Sprint 1:

1) Extend SaVANNA to handle segmentation tasks in addition to classification tasks
2) Extend MORF to work on a sort of convolution on 2D images and 3D volumes (convolutional MORF)

Sprint 2:

1) Ensure the SaVANNA package code meets a high quality standard and submit SaVANNA to PyPI and Anaconda repositories for public use and easy installation
2) Apply SaVANNA to the Mouselight data set for fully automatically segmenting projection neurons from two-photon tomography data

Sprint 3:

1) Write a paper about SaVANNA package functionality, applied to an interesting brain image data set (eg Mouselight), according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
2) Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

Victor

Sprint 1:

1) Setup the SaVANNA package Github Repository by organizing code from Satish
    a) Refactor code to be organized and sklearn compliant
2) Real data applications
    a) General applications
        i) Write code to appropriately process 3D brain volumes into a normalized format that can be inputted to MORF
    b) Mouselight
        i) Assist Tommy Athey et. al. with data retrieval functions from Mouselight data such as
            (1) Crop out image volume centered at a spatial coordinate with specified dimensions
            (2) Convert SWC neuron morphology files into pixelwise neuron segmentations
3) Additional package functionality
    a) Layer MORF/convolutional MORF with connected layers (akin to a neural net architecture) to potentially extract more complicated features from the images/volumes (SaVANNA)

Sprint 2:

1) Continue developing SaVANNA

    a) After SaVANNA can effectively classify and segment existing machine learning data sets (eg ImageNet, MNIST) for some number of samples, apply SaVANNA to preprocessed brain images and volumes. Ideally, SaVANNA should outperform MORF
2) Real data applications
    a) Apply SaVANNA to the Mouselight data set for fully automatically segmenting projection neurons from two-photon tomography data

Sprint 3:
1) Publication
    a) Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
    b) Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

Alex
Sprint 1: Repo setup
1. Further flesh out Savanna repo by adding test scripts for run_experiment using CNN model functions
2. Fully test code in naive version and deep_conv_rf of SaVANNA (codecov >80%)
3. Layer MORF/convolutional MORF with connected layers (akin to a neural net architecture) to potentially extract more complicated features from the images/volumes (SaVANNA)

Sprint 2:
1. Write a function to classify labelled brain images using MORF
2. General applications
    a. Write code to appropriately process MRI images into a normalized format that can be inputted to MORF
    b. Write code to appropriately process 3D brain volumes into a normalized format that can be inputted to MORF

Sprint 3:
1. Publication
    a. Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
    b. Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice

Ryan

Sprint 1:

1) Write a function to classify labelled brain images using MORF
2) Write a function to classify labelled 3D volumes using MORF
3) Extend MORF to work on a sort of convolution on 2D images and 3D volumes (convolutional MORF)
4) Layer MORF/convolutional MORF with connected layers (akin to a neural net architecture) to potentially extract more complicated features from the images/volumes (SaVANNA)
5) Run SPORF/MORF on different levels of pooled images to potentially extract more complicated features from the images/volumes (SaVANNA)

Sprint 2:

1) MORF Parameter turning
    a) Apply SaVANNA to a variety of different 2D images and 3D volumes to better understand how varying MORF properties affects the overall
2) Graph Input
    a) Extend SPORF/MORF/SaVANNA to take graphs as input
    b) Use graph inputs for classification and segmentation tasks on connectomes

Sprint 3:

1) Publication
    c) Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
    d) Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

Stanley

Sprint 1:

1. Extend SPORF(/MORF) to accept 3D volumes as input and submit a pull request to SPORF
2. Real data applications

    2a. Write code to appropriately process MRI images into a normalized format that can be inputted to MORF

    1b. Write code to appropriately process 3D brain volumes into a normalized format that can be inputted to MORF
3. Additional package functionality

    3a. Extend SaVANNA to handle segmentation tasks in addition to classification tasks

Sprint 2:

1. Continue developing SaVANNA
    a. After SaVANNA can effectively classify and segment existing machine learning data sets (eg ImageNet, MNIST) for some number of samples, apply SaVANNA

to preprocessed brain images and volumes. Ideally, SaVANNA should outperform MORF

2. Real data applications
   a. Apply SaVANNA to the Mouselight data set for fully automatically segmenting projection neurons from two-photon tomography data
3. Graph Input
   a. Extend SPORF/MORF/SaVANNA to take graphs as input

Sprint 3:

1. Publication
   a. Write a paper about SaVANNA package functionality according to the following: https://bitsandbrains.io/2019/02/10/how-to-write-a-paper.html
   b. Submit the paper to the Journal of Machine Learning Research (JMLR), or another peer-reviewed journal based on Jovo's advice.

References

[1] J. Winnubst *et al.*, “Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain,” *Cell*, vol. 179, no. 1, pp. 268–281.e13, Sep. 2019.

[2] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions,” *J. Digit. Imaging*, vol. 30, no. 4, pp. 449–459, Aug. 2017.

[3] T. M. Tomita *et al.*, “Random Projection Forests,” 10-Jun-2015.