Towards explainable automated Neuroanatomy

Kui Qian, Beth Friedman, David Kleinfeld, Yoav Freund

No Institute Given

Abstract. A fundamental goal of neuroanatomy is the identification of brain structures. Manual identification of structures is based on the spatial distribution of cell shape, size, orientation and density. With new technology it is possible to image entire brains at high resolution. However, manual identification of structures in these massive datasets is prohibitively time consuming. We present a machine learning method for automatic detection of brain structures. Our approach is based on diffusion maps, in combination with hand-picked features, and Adaboost for combining the features. Our method is robust against brain to brain variation and details of neuronal staining. Our method produces structure detections together with their explanation. This that the human anatomist and the computer interact to improve the detection of structures and the addition of new structures.

1 Main

1.1 Summary

We propose a system for detecting anatomical structures in the mouse brain. Our system takes as input high-resolution images of aligned sections. It produces as output the estimated center of mass (COM) for each structure. Each detection is assigned a confidence. High confidence structures are associated with a visual explanation.

Significance for Neuroanatomy Does not have to be very detailed. Target audience is ML.

- Importance and History
- Manual detection of brain structures has two major drawbacks, the first is the amount of time needed by an anatomist. The second is inconsistencies between anatomists. Automatic detection has the potential of reducing the time and improving the consistency for this task.
- Challenges: Variability of images
  - biology: animal to animal, gender, generally alien across species, transgenic (differentions to genome due to gene expression)
  - technical: staining, sectioning (mechanical, optical)

Explainable AI Black box approaches focus on finding an accurate end-to-end mapping. As such, these approaches are hard to interrogate when they fail. In contrast, we take an explainable-detector which allows the anatomist to understand why the ML made particular decisions. This benefits both sides ...

Cell based approach Most CNN-based method of image analysis are based on the concept of a sliding window. This implies that an y explanation provided by the CNN is based on the content of one or more windows. Windows need to be large enough to capture the image statistics. This means that a typical window contains 10-100 cells. On the other hand, anatomist base their analysis on cytoarchitecture, which, in turn, is based on the shapes of individual cells and the relationshp between them. This makes it harder for the anatomist to understand the decisions of the detector.

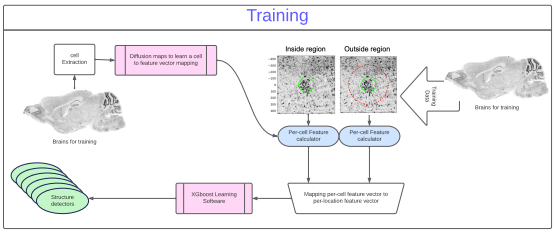

Fig. 1. Training

To remedy this problem and make the detections explainable, we use individual cells as our basic unit.

1.2 Experimental design

The trained system is a collection of structure detectors. Each structure is a known anatomical entity. To train the detectors we combine three sources of information:

- **Images of aligned sections:** The Nissl images of XX brains are the primary information source. They are also by far the largest at XXX GB per brain.
- **The Atlas** is a representation of the shape of each structure and the relative locations of the structures. The atlas was constructed through a concensus ....
- **COMs for individual brains:** For XX brains we had an anatomist locate the COM ....

1.3 System design

The overall design of the system is shown in Figure **??**. The main components of our system are the following (details and descriptions of the smaller components are given in the methods section)

Dimensionality reduction An image of a single cell is typically around 50×50 pixels, or a 2500 dimensional vector. The dimension of this representation is too high for effective machine learning. We therefor seek a dimensionality reduction mapping. This mapping consists of three parts:

1. **Normalization:** We normalize the image of the cell in three ways. We shift the grey-levels so that the mean is XXX, scale the grey levels so that the standard deviation is YYY, and rotate the image so that the long axis of the cell is in angle zero. The three parameters associated with the normalizations define three features.
2. **Standard shape features**: we use the Hu features: x,y,z...
3. **Diffusion maps**: Diffusion Mapping [**?**,**?**] is a non-linear dimensionality reduction method that is based on a graphical representation of the data and on the Laplace operator on graphs.

Characterizing Cytoarchitecture uses Difference between the CDFs for inside and outside of each structure (based on manually annotated structure boundaries).

Structure detectors Combine the difference-of-CDF features using boosted trees (XGBoost). Each structure has a corresponding detector.

Structure detection confidence The confidence of structure detections is measured by the prediction margin and by the sharpness of the detection peak.

Structure detection explanation

2 Results

- Detection results, with confidence levels, consistency with Beth.
- **Explaining detections** The explanation of the detection is expressed in terms of cells whose shape gives evidence to the structure.
- **Cell shape parametrization** Uses a combination of Hu moments and dimensionality reduction using eigen-maps.
  - Eigen-maps learn a dimensionality reducing mapping cell shape to a ten dimensional representation. As it is an unsupervised method it requires no human labeling. We take advantage of the very large number of cells in single brain.
  - Each brain creates a different mapping, however, the mappings can be made consistent by adding a linear transformation. This creates a stable parametrization and makes the detections more consistent.

3 Methods

In this section we provide more details on the algorithms we use in our system. The full code is available from github...

- Cell based features.
- Diffusion mapping.
- Calibration of the diffusion features.
- Region features and CDFs.
- Boosting and XGBoost.
- Localizing structures. Rough alignment, computing detections scores, computing autocorrelation and asigning confidence.
- Using the system for brain-to-atlas alignment.
- Generating explanations.