# **Computational Neuroanatomy (Project 5)**

**Aim 1: Creation of the Trainable Texture-based Digital Atlas.**

This is a poorly defined Aim. I don’t know what to write here.

**Aim 2: Construction of a fluorescence-based atlas to include spectrally distinct specific cell markers.**

Automatic detection of marked cells.

One of the more labor intensive tasks faced by anatomists in our group is identifying individual marked cells. (David and Beth to fill in biology and estimates of the amount of work involved in detecting cells in ¼ of the sections). We use machine learning to create a marked cell detector as depicted in Figure XX. Both human and computer detection are based on two channels, a channel for the fluorescent marker of projected cells and a reference channel (neuro-trace). Positive detections are based on having a bright blob in the first channel and a matching weaker blob in the second channel. The detector combines a large number of features that quantify the shapes of the blobs using XGBoost.

Our approach is incremental, rather than a wholesale replacement of the human anatomist by an AI, we use the computer to detect the easily marked cells, and solicit human feedback on the hard ones. Using the concept of boosting margins, we partition the detections into confident, or easy, detections vs. unconfident or hard examples. This scheme is successful if the number of marked cells that it misses, and the number of confident detections that are wrong are both small.

The initial training set was collected by manually locating the marked cells in one out of every four sections, this took XXXX work hours and generated 2805 human detections.

After training a detector on this data (400 sections; human ¼?; blocked?), we applied it to all sections (including the training sections) the following table summarizes the results

|  | Total | Training detections | Human + | Human - | No feedback |
| --- | --- | --- | --- | --- | --- |
| Computer confident | 12224 | 1599 | 0 | 50 | 12220 |
| Computer unconfident | 1420 | 111 | 402 | 247 | 634 |
| No detection |  | 121 | 299 |  |  |

The main conclusion from this table is that the confident predictions by the detector are very reliable, only 50 out of the more than 12,000 confident detections were marked by the human as false detection. (not sure how that relates to the cases where the number of detections was right but the exact locations were wrong). Detecting that number of marked cells would have taken XXX time. Instead, the human only needs to label the 1420 unconfident predictions. A more significant problem is the number of marked cells that were missed. (299). Reducing this number is a priority.



**Aim 3. Improving the computer vision used to detect landmarks.** 

In Year 4 we have further improved our structure detectors and made the detections “explainable”. To test the accuracy of those detectors we collected manual annotations of 20 structures from 10 brains.

Detection is done in two steps: Rough alignment followed by structure detection. The structure detector uses the locations estimated by the rough alignment as starting points for detecting the structure.

* **Rough Alignment** is based on grey-level matching at a resolution of 20 microns per pixel. For this operation we use public code from itk.org. The advantage of this approach is that it uses large grey level features, such as high contrast boundaries and does not require any training. The disadvantages are (1) it requires a reference brain that is aligned to the atlas that is imaged in the same way as the target brain. (2) limited accuracy.
* **Structure Detection** uses a separate trained detector for each structure. This detector is cell-based and was described in the Year 3 progress report.

Two important features were added to the detector in Year 4.

1. **Confidence Score:** Some structures are easier to detect than others. We define a confidence score which is based on the auto-correlation of the detection score. When the confidence score is high the detection is more likely to be at the correct location. When performing alignment we use the rough detections and the high confidence detections first, and then perform corrections using the less confident detections.
2. **Explanation for detections:** Detectors are complex adaptive algorithms. The generated detections are 3D locations (COMs) in the sectioned brain. The confidence score identifies the structures for which the computer’s detection is confident. When these detections are vetted by an anatomist it is very helpful to know **why** the detection is confident. (this is part of a general trend called “Explainable AI”). Using the fact that the detector is cell-based and that Boosting tends to select a sparse representation of features, we identify the cells are evidence for the inside and the outside of the structure.

To quantify the accuracy of our detector we collected manual structure centers (COMs) for 12 in 10 brains.

We then compared the automatic detections, after rough alignment and after rough alignment + detection, with the human detection. The figures demonstrate several facts: (1) The detections are significantly better than the rough alignment. (2) Most of the detections have error smaller than 200 micron. (3) most detections whos confidence score is higher than 3000 have error smaller than 100 micron.

**Plans for Project 5 during year 5:**

1. Continue improving structure detectors and identify more structures that can be detected with high confidence
2. Continue improving the marked cell detector and apply it to other tasks.