Robust Landmark Detection for Alignment of Mouse Brain Section Images

Yuncong Chen, David Kleinfeld, Martyn Goulding and Yoav Freund

Department of Computer Science and Engineering,
University of California, San Diego, La Jolla, CA 92122, USA

Abstract. In this paper we describe a method for the automatic detection of landmarks in histology images of mouse brain sections. The novelty of our method is in the use of unsupervised learning to find regions with distinct texture as well as clear boundaries. Region proposals are grown from superpixels, based on which significant regions such as nuclei and fiber tracts are identified using clustering. Detected regions are shown to correspond well with real anatomical structures. To complement region landmarks, robust boundary segments are also found by consensus voting. Landmarks from different sections can be matched using a distance function that is shown to be robust under distortion. This approach enables us to create a library of landmark detectors based on texture and shape. They can be applied to identify landmarks from new images and also be updated by incorporating new data. This aims to reduce human supervision in the registration and atlas building process.

Keywords: landmark detection, atlas building, mouse brain, registration, automated annotation

1   Introduction

In this paper we describe a method for the automatic detection of landmarks in histology images of mouse brain sections. The purpose is to facilitate image registration and atlas generation. By aligning images of nissl and fluorescent-stained brain sections, we aim to create a 3D digital atlas for the mouse brainstem, which incorporates cell type, tract tracing and other physiological data. The novelty of our method is in the use of unsupervised learning to find regions with distinct texture and clear boundaries, which reduces time-consuming human labelling.

Figure 1 shows a typical image of nissl-stained mouse brainstem section. Although the brainstem does not have salient edges like the cortex, it contains many compact neuron clusters (nuclei) and striated regions (fiber tracts). Both types of structure have distinct texture that can be detected and modelled.

We use semi-supervised learning to create a library of landmark detectors based on texture and shape. They can be applied to identify landmarks from new images and also be updated by incorporating new data. This is at the center of our vision of integrating atlas-guided annotation, landmark-based registration and atlas generation into an iterative framework.

The remaining of this paper is organized as the follows. Section 2 describes how we represent textures using Gabor filters and superpixels. Section 3 and 4 explains how we identify potential landmark regions based on texture distinctiveness. Section 5 describes how we detect stable boundary segments to complement region landmarks. Section 6 describes how landmarks are matched between sections.

2 Related Work

An atlas is typically constructed by mapping image series of different studies to a common coordinate system, and aggregating the information. Images of different modalities or specimens stained for different targets can be superimposed. For example, the Allen Mouse Brain Atlas [13] computes the expression energy of different genes by collecting statistics from in-situ hybridization images of all the studies. In other cases, inter-specimen variation is interested. For example, the Waxholm Space atlas [12] provides an anatomical probability map that gives the probability of each voxel belonging to a particular anatomical structure. The probabilities are computed from a large number of co-registered specimens. The atlas that we aim to construct will combine images of different specimens and stained for different targets, and serve as the underpinning for circuits analysis and quantitative analysis of phenotypical variations.

To alleviate the bottleneck of manually selecting landmarks for registration, there has been much work on automatically reconstructing 3D volume from histology sections. Intensity-based methods such as [14, 15] use block-matching to compute a regularized displacement field. In the Allen Brain Atlas [13, 17], the Waxholm Space [10] and many earlier work [2, 12], histology section registration is based on maximizing mutual information.Other methods such as [3, 5, 8] rely on detecting gradients in the image, and try to align high-gradient edges. The level of details in histology images also motivates the use of more elaborate descriptors for describing landmarks, including SIFT [16], binarized gradient orientation histogram [11] and bag of multiple types of features [7]. In our work, we use Gabor filters combined with superpixelation to represent texture, and detect significant regions with distinct texture as well as stable boundary segments as landmarks.

3 Representing Texture using Histograms of Gabor Textons

Images are first filtered using Gabor filters [6, 9] with 9 orientations and 11 scales, resulting in 99 dimensional feature vectors. We reduce the data by quantizing these feature vectors into 100 clusters using K-Means. Clusters with close centroids are then merged, and the final cluster centroids are the textons. In our experiment, we have 14 textons.

We further reduce the data by describing texture at the level of superpixels. For each superpixel, texture is represented by the histogram of the textons it contains (Figure 1). We denote the texton histogram of the i'th superpixel by h_i. Superpixels are obtained using the SLIC algorithm [1].

Since different sections may be oriented at different angles, we need the features to be rotation-invariant. For this purpose, before doing K-Means, we compute the energy

distribution over orientations for each Gabor feature vector, and shift all feature vectors such that the modes of the directional energy distributions are aligned. In this way, the texture is decoupled from directionality.
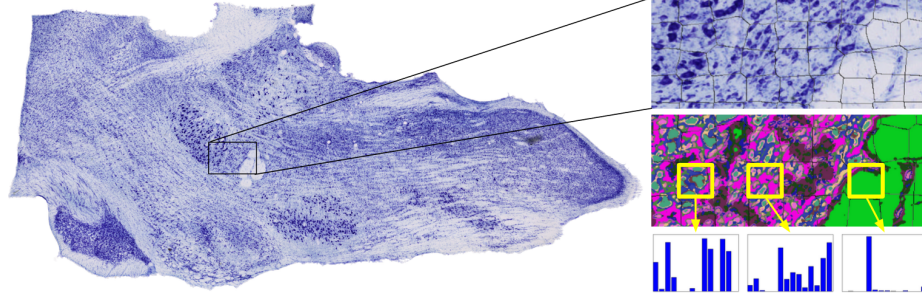

Fig. 1. Left: A typical image of mouse brainstem section. Right top: part of the original image with superpixels overlaid. Middle: texton map of the same area. Bottom: texton histograms of three superpixels with different textures.

4 Evaluating Region Significance using Statistical Test

We define a region landmark as a connected set of superpixels. Before describing how to find significant regions, we need a scoring function that evaluates the significance of a region. We propose the significance score, $F(S)$, of a region S, to be a linear combination of three terms.

The first term is the distinctiveness of this region's texture, or in other words, how different its texture is from the surrounds. Since textures are represented as histograms, we formulate this problem as a statistical test on whether the region's average texton histogram and the texton histograms of the surrounding superpixels are sampled from the same distribution. We employ the chi-squared test for independence and the resulting p-value serves as a quantitative measure of the distinctiveness. The smaller the p-value, the more significant the region is. Since a region often has neighbours with different textures, the test is made between the region's histogram and that of every surrounding superpixel. To be conservative, the largest p-value among all tests is used. Denote the set of surrounding superpixels by $T(S)$, the first term is,

$$F^{cont}(S) = - \max_{j \in T(S)} pval(h_S, h_j)$$

where $pval(\cdot, \cdot)$ is the p-value of applying the chi-square test for independence to two histograms.

The second term is the texture homogeneity within the region, defined as the mean p-value of chi-square tests between all superpixels in the region and the region's aver-

age:

$$F^{coh}(S) = \frac{1}{|S|} \sum_{i \in S} pval(h_S, h_i)$$

The third term stems from the fact that most region landmarks have compact shapes. A common measure for the compactness of a closed curve is the isoperimetric quotient, defined as the enclosing area divided by the square of the circumference. We use the number of surrounding superpixels as a proxy for the circumference, and the number of superpixels in the region as that for the area. Small

$$F^{comp}(S) = \frac{|T(S)|}{|S|^2}$$

The overall significance score $F(S) = F^{cont}(S) + F^{coh}(S) + F^{comp}(S)$.

5 Finding Significant Regions with Region Growing and Clustering

To find potential regions, we perform region growing for every superpixel. Starting with a singleton containing the seed superpixel, this greedy procedure considers all neighbours of the current set, and iteratively adds the one whose texture is closest to the current region's texture, measured by the χ^2 distance between texton histograms. The growing stops when the region reaches 10% of the total area. Eventually, the region at the point when the significance score is the largest is returned. Figure 2 shows examples of this procedure.
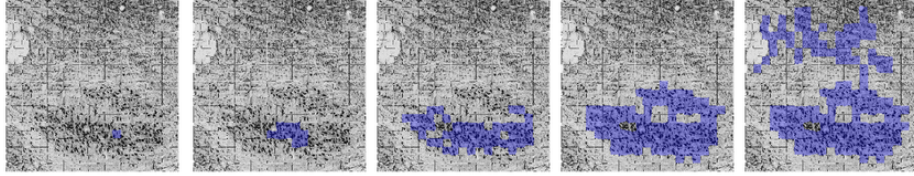

Fig. 2. Regions during growing, from left to right, at iteration 1, 10, 50, 96 (most significant), 150 (last iteration).

We call the set of superpixels that grows from a seed superpixel k the *region proposal* of the seed, denoted by S_k. If a region has distinct texture, the proposals of superpixels inside this region should be very similar, while proposals from within an inhomogeneous region are random (see Figure 3). In other words, the region proposals form clusters. The denser a cluster is, the more distinctive the region it represents. Using Jaccard index as pairwise distance, we group the region proposals using hierarchical clustering. Then within each group whose size is large enough, we select the proposal with the highest significance score to be the representative of that group. All representative proposals are ranked according to their significance scores.
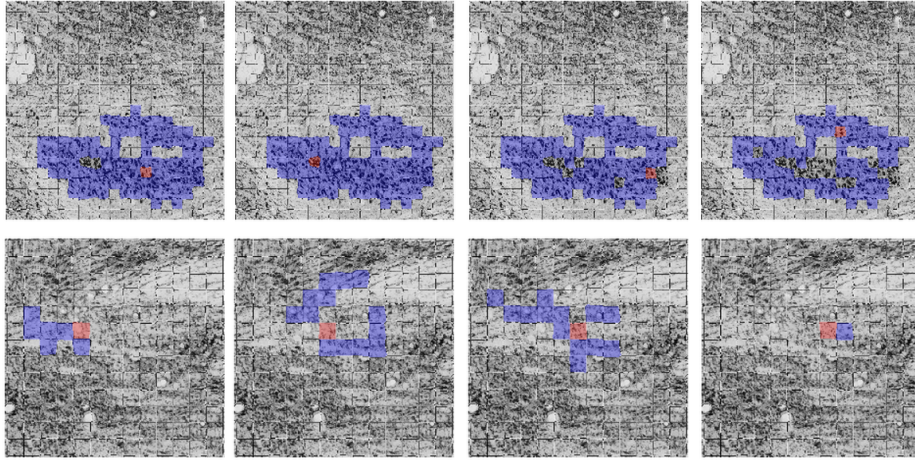
Fig. 3. The first row shows the region proposals of four different seeds (in red) in the facial motor nucleus. They are very similar. The second row shows the proposals of four neighboring seeds in an inhomogeneous region. They are very different.

6 Identify Robust Boundaries by Region Consensus

Sometimes a region gradually transitions into neighboring texture on one side, but has a clear boundary on the other side. The method described in the previous section may not capture this inhomogeneous region, but the open boundary is nonetheless a perfect landmark. Figure 4 shows such an example. Notice how much the proposals from seeds in such region vary, but many of them still agree on the clear boundary. This motivates a consensus-based approach for evaluating boundary robustness, in which each region proposal votes for the segments on its boundary.
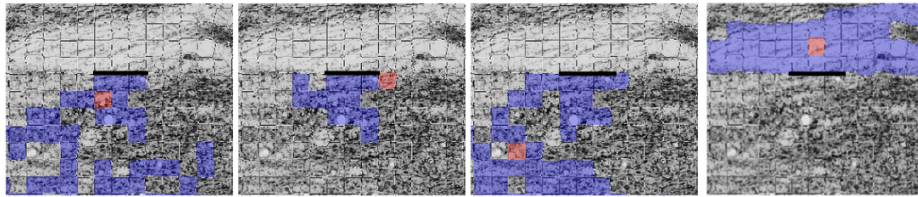

Fig. 4. The three region proposals in an inhomogeneous area are not consistent as a whole, but they all agree on a robust boundary segment (highlighted).

We represent a boundary segment between two superpixels by an ordered tuple (i, j), where i is the interior superpixel and j is the exterior superpixel. We also denote by δS_k, the set of segments on the boundary of a region proposal S_k. The vote a region proposal casts to a boundary segment depends on how contrasty the segment

is, measured by the texture histogram distance between the region's average and the segment's exterior superpixel. We call the set of superpixels that vote for a segment the *supporter set* of the segment, and denote it by $R_{(i,j)} = \{k : (i,j) \in \delta S_k\}$. The total score received by a segment (i,j) is then:

$$b_{(i,j)} = \sum_{k \in R(i,j)} \chi^2(h_{S_k}, h_j)$$

We discard segments whose vote is lower than a threshold. Figure 5 shows a vote map. Instead of modelling individual segments, we combine segments with similar supporter sets into groups, again using hierarchical clustering with Jaccard index between supporter sets as pairwise similarity. Each segment group represents a boundary. All boundaries are ranked according to the total vote received by its segments.
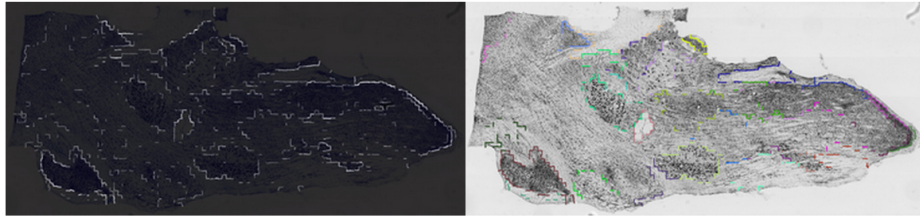

Fig. 5. Left: thresholded boundary vote map. Right: grouped boundary segments

7 Matching Landmarks from Different Sections

By representing region landmarks using closed boundaries and merging coinciding boundaries, we unify the two types of landmarks into a single set consisting of both open and closed boundaries. In order to use the landmarks for registration, correspondences must be made between them. In this section, we design a distance function for comparing two boundaries. This function is a weighted combination of the differences in four aspects, with weights chosen empirically:

$$D(\mathbf{B}_1, \mathbf{B}_2) = w^{int} D^{int}(\mathbf{B}_1, \mathbf{B}_2) + w^{shape} D^{shape}(\mathbf{B}_1, \mathbf{B}_2)$$
$$+ w^{ext} D^{ext}(\mathbf{B}_1, \mathbf{B}_2) + w^{loc} D^{loc}(\mathbf{B}_1, \mathbf{B}_2)$$

The first term is the difference of interior textures. For a region landmark S, the interior texture is simply the average texton histogram h_S. For a boundary B, the interior texture is the average texton histogram of the union of all segments' supporter sets. Denote this union by $Q_B = \cup_{(i,j) \in B} R_{(i,j)}$, then $h_B^{int} = h_{Q_B}$. The difference is computed using the χ^2 distance:

$$D^{int}(\mathbf{B}_1, \mathbf{B}_2) = \chi^2(h_{B_1}^{int}, h_{B_2}^{int})$$

The second term measures the similarity of boundary shapes. We reduce the boundary to a point set consisting of midpoints of the segments. The shape distance between two point sets can be computed using shape context descriptors [4]. The shape context descriptors characterize the organization of other points around each point using a histogram. Two sets of points are matched by finding the minimum bipartite matching, where the edge weights are the chi-square distances between shape context descriptors. The Hungarian algorithm is used to find the minimum matching. The average cost of this matching, is used as the second term:

$$D^{shape}(\mathbf{B}_1, \mathbf{B}_2) = \frac{1}{|M|} \sum_{(i,j),(p,q) \in M(B_1, B_2)} \chi^2(c_{i,j}, c_{p,q})$$

where $M(B_1, B_2)$ is the minimum matching, $c_{i,j}$ and $c_{p,q}$ are the shape context descriptors of segment (i, j) and segment (p, q) respectively.

After the matching is made, we compute the third term, defined as the total distance between exterior textures of all matched segments:

$$D^{ext}(\mathbf{B}_1, \mathbf{B}_2) = \sum_{(i,j),(p,q) \in M(B_1, B_2)} \chi^2(h_j, h_q)$$

The fourth term measures the spatial proximity. This is the thresholded Euclidean distance between the center of mass of the boundaries' midpoint sets:

$$D^{loc}(\mathbf{B}_1, \mathbf{B}_2) = \max(0, \|m_{B_1}, m_{B_2}\|_2 - l)$$

where m_{B_1} and m_{B_2} are the center of mass of boundaries B_1 and B_2, and l is a tolerance within which the position deviation is not penalized (set to 1 mm in our experiment).

Using this distance function, we compute the pairwise distances for two sets of landmarks detected from different sections. Two landmarks are matched if they are simultaneously the closest landmark to each another.

8   Experiments

8.1   Comparison to Human Labelling

We test our algorithm on a series of 30 section images of a nissl-stained mouse brainstem. The images are scanned at 2 microns per pixel, showing individual neuronal cell bodies. Both types of landmarks are detected on all images. For each image we use the top 20 closed boundary and the top 10 open boundary. Figure 6 shows the landmarks detected from one image. Also shown is a version created by a human labeller who annotated for nuclei and fiber tracts with the help of a printed atlas. Most significant structures from the human labelling are detected by the algorithm.
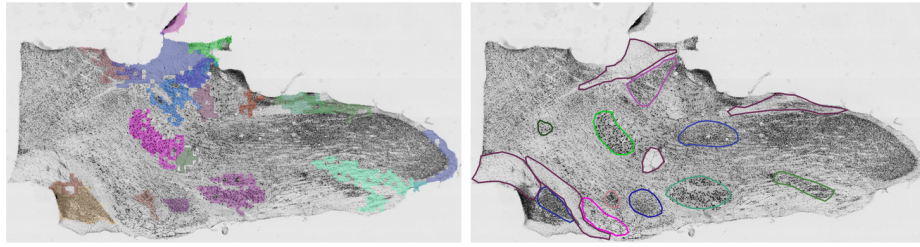
Fig. 6. Left: top 20 landmarks detected by the algorithm. Right: human labelling for recognizable nuclei and white matter.

8.2 Robustness of Matching

The landmark matching algorithm is applied to all 30 pairs of consecutive images. In order to test the robustness of matching under large displacement, we remove the spatial proximity term from the landmark distance function, leaving only the texture and shape terms. A human evaluator then judges whether a matching is correct, partially correct, or incorrect. Among all 166 matchings returned by the algorithm, 106 are correct (63%), 34 are partially correct (22%) and 26 are wrong (15%). One example is shown in Figure 7. This demonstrates the effectiveness of texture modelling. Even though some landmarks change shape significantly, the algorithm still finds the correct matchings.

9 Conclusion and Future Work

In this paper we described algorithms for detecting region and boundary landmarks from histology images of mouse brainstem. Region proposals are grown from superpixels, based on which significant regions such as nuclei and fiber tracts are identified using clustering. Detected regions are shown to correspond well with real anatomical structures. To complement region landmarks, robust boundary segments are also found by consensus voting. Landmarks from different sections can be matched using a distance function that is robust under distortion.

Our next step is to use the matched landmarks for intra-specimen registration, and further co-register multiple specimens to generate the atlas. Meanwhile, we also plan to include human feedback in the learning loop. Although the advantage of our approach is reducing human supervision, it is important for human experts to correct wrong labellings made by the algorithm in order to learn accurate models for landmarks. Instead of requiring them to label every region of interest, our method only needs them to validate our library of detectors. Once those have been validated, the process is completely automatic.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(11), 2274–2282 (2012) 3
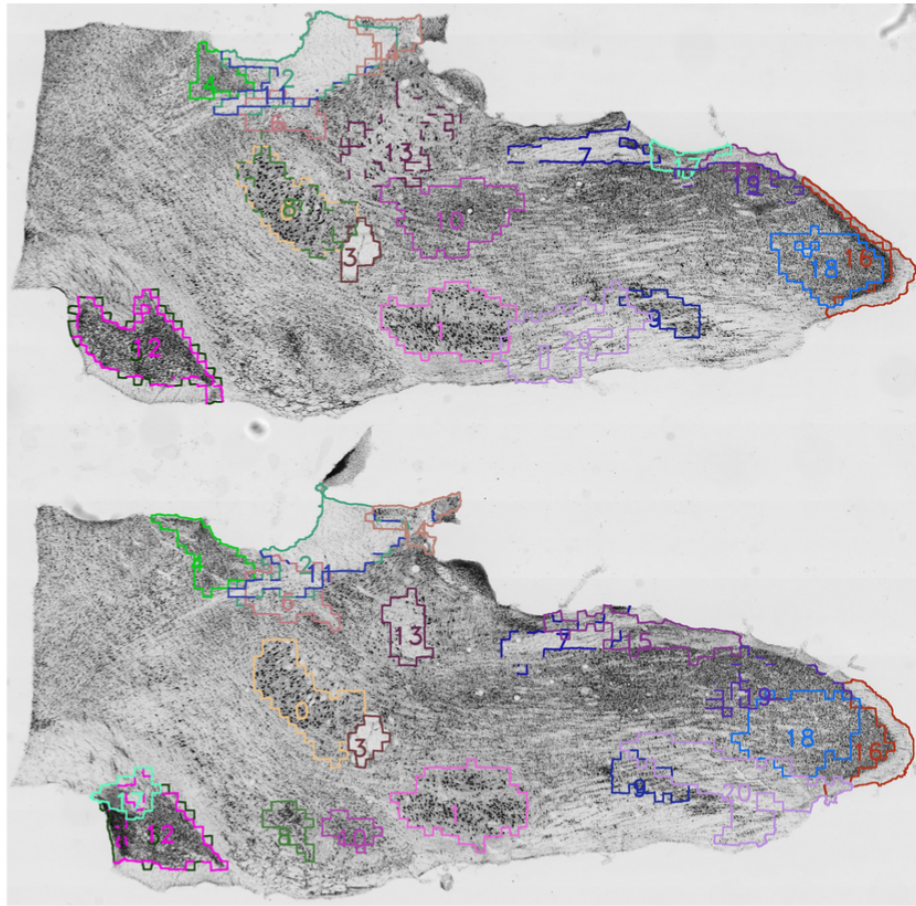
Fig. 7. Landmark matching example. Matched landmarks are marked with the same color and number. Note that matchings are found based only on texture and shape. Matching 13 is made possible by modelling both open boundaries and close boundaries. Landmarks such as 1 and 9 show considerable shape change, but are still matched.

2. Ali, A.A., Dale, A.M., Badea, A., Johnson, G.A.: Automated segmentation of neuroanatomical structures in multispectral mr microscopy of the mouse brain. Neuroimage 27(2), 425–435 (2005) 2

3. Bagci, U., Bai, L.: Automatic best reference slice selection for smooth volume reconstruction of a mouse brain from histological images. Medical Imaging, IEEE Transactions on 29(9), 1688–1696 (2010) 2

4. Belongie, S., Malik, J., Puzicha, J.: Shape context: A new descriptor for shape matching and object recognition. In: NIPS. vol. 2, p. 3 (2000) 7

5. Cifor, A., Bai, L., Pitiot, A.: Smoothness-guided 3-d reconstruction of 2-d histological images. Neuroimage 56(1), 197–211 (2011) 2

6. Clausi, D.A., Jernigan, M.E.: Designing gabor filters for optimal texture separability. Pattern Recognition 33(11), 1835–1849 (2000) 3

7. Cruz-Roa, A., Caicedo, J.C., González, F.A.: Visual pattern mining in histology image collections using bag of features. Artificial intelligence in medicine 52(2), 91–106 (2011) 2

8. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, pp. 726–733. Springer (2006) 2

9. Jain, A.K., Farrokhnia, F.: Unsupervised texture segmentation using gabor filters. In: Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on. pp. 14–19. IEEE (1990) 3

10. Johnson, G.A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., Nissanov, J.: Waxholm space: an image-based reference for coordinating mouse brain research. Neuroimage 53(2), 365–372 (2010) 2

11. Kurkure, U., Le, Y.H., Paragios, N., Carson, J.P., Ju, T., Kakadiaris, I.A.: Landmark/image-based deformable registration of gene expression data. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 1089–1096. IEEE (2011) 2

12. Lee, E.F., Jacobs, R.E., Dinov, I., Leow, A., Toga, A.W.: Standard atlas space for c57bl/6j neonatal mouse brain. Anatomy and embryology 210(4), 245–263 (2005) 2

13. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al.: Genome-wide atlas of gene expression in the adult mouse brain. Nature 445(7124), 168–176 (2007) 2

14. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3d structure from serial histological sections. Image and vision computing 19(1), 25–31 (2001) 2

15. Roberts, N., Magee, D., Song, Y., Brabazon, K., Shires, M., Crellin, D., Orsi, N.M., Quirke, R., Quirke, P., Treanor, D.: Toward routine use of 3d histopathology as a research tool. The American journal of pathology 180(5), 1835–1842 (2012) 2

16. Sun, S., De Magalhaes, N., Cahill, N.D.: Nearly rigid descriptor-based matching for volume reconstruction from histological sections. In: SPIE Medical Imaging. pp. 831417–831417. International Society for Optics and Photonics (2012) 2

17. Yushkevich, P.A., Avants, B.B., Ng, L., Hawrylycz, M., Burstein, P.D., Zhang, H., Gee, J.C.: 3d mouse brain reconstruction from histology using a coarse-to-fine approach. In: Biomedical Image Registration, pp. 230–237. Springer (2006) 2