Automatic Atlas-Based Segmentation of NISSL Stained Mouse Brain Sections Using Convolutional Neural Network

Jing Xiong[1], Feiran Wang[1], Jian Zhang[2]
[1]Department of Electrical Engineering, Stanford University
[2]Department of Computer Science, Stanford University
{jxiong1, feiran, zjian}@stanford.edu

Abstract

In this paper, we focus on segmenting NISSL stained mouse brain images into main brain regions - grey (cerebrum, brainstem, and cerebellum), fiber tracts, and ventricular systems. We feed mouse brain experimental slices acquired by the biology department and mouse brain reference slices from Allen Brain Institute into a fully convolutional neural network which is capaable of producing per-pixel predictions for each image. The neural nets generates a label for each pixel. We experimented on the reference mouse brain atlas from Allen Brain Institute and experimental images from the biology department at Stanford University. We achieved 96.1% accuracy on test atlas slices and 92.1% accuracy on the test experimental slcies. The work can be potentially extended to the automatic segmentation of more salient features that neuroscientists use to determine finer brain regions in an attempt to facilitate the manual brain annotation process.

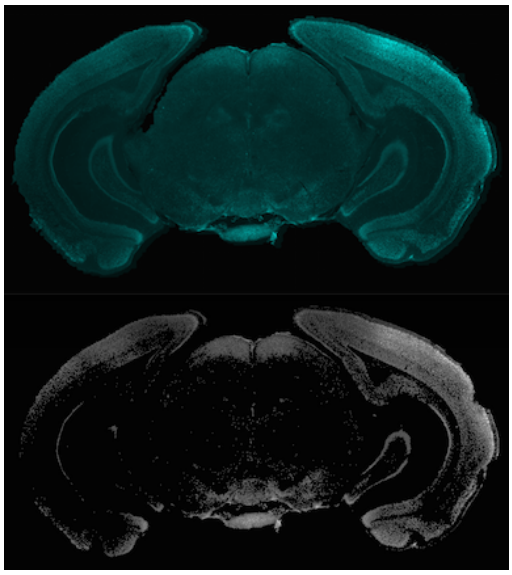

Figure 1. Intensity thresholded experimental slice. Top: original experimental slice. Down: Intensity thresholded experimental slice in an attempt to get the hippocampus

1. Introduction

A longstanding problem in neuroscience research is to manually annotate brain regions. Due to imaging constraints, it is hard to generate human brain images at neuron size resolution. Biologists have been using light field microscopic mouse brain sections as a starting point to study the mammalian brain neuron circuitry. The study of the neuron circuitry often requires finding fine regions of each labeled neurons.

This project focuses on segmenting NISSL stained brain images into main brain regions - grey (cerebrum, brainstem, and cerebellum), fiber tracts, and ventricular systems. The outcome can be potentially used to facilitate automatic non-rigid registration of an experimental brain slice to a reference section as a term in the energy function or to augment the process of 2D/3D localization of experimental slice to a reference volume that are both crucial to automatic annotation of the mouse brain slices. Moreover, fiber tracts and ventricular systems are key features to refer to when biologists decide the brain regions that a neuron lies in. This project can also be extended to train on more detailed features in the future.

The slice preparation process always introduce artifacts to the brain images, for example, uneven illumination caused by uneven staining. Moreover, the main contrast of images are from NISSL staining where only neuronal cell bodies are stained. The images are low-contrast and noisy. Therefore the traditional segmentation on medical images based on intensity cannot be directly used on the experimental mouse brain slices acquired by the biology labs. fig. 1 shows a segmentation result on an experimental slice. The right side of the image is heavily stained and is much

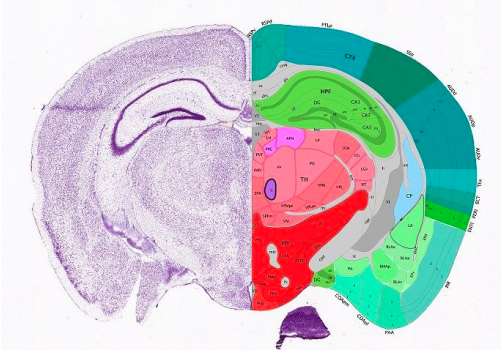

Figure 2. An example slice of the annotated reference atlas. Left side shows the original NISSL stained slice; right side shows the accompanying annotation slice at the lowest hierarchy.

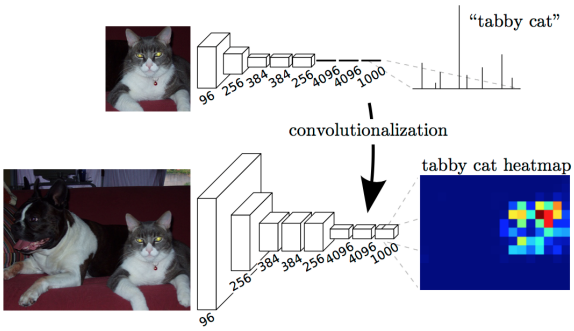

Figure 3. From fully connected to fully convolutional nets.

brighter than the left side. Trying to segment the hippocampal region in this slice purely based on intensity will only give the user the heavily stained area, and the hippocampus on the left side is eliminated. In addition to the uneven illumination, the experimental slices are often nonrigidly deformed during the sliding and staining process, and there could also be teared tissues, air bubbles, and tissue folding.

In this project we use multilevel convolutional network to train the reference mouse brain slices as a starting point. The input are the atlas mouse brain slices that are accurately labeled by neuroscientists in details. We parse the region hierarchy and got the labels for the five main regions, namely background, cerebrum, brainstem, cerebellum, fiber tracts, and ventricular systems. The output are the integers representing the region name for each pixel in the input image. We then input some of the labeled experimental images from the biology department into the system and finetune the top-layer hyperparameters on these experimental data.

2. Related Work

Image segmentation has been a long standing vision problem. There are two main stream of segmentation approach. The non-semantic methods [12, 1] generate regions where visual cues are consistent within the segments. The second stream is semantic methods where the aim is to assign labels to pixels in the images. Existing works either predict label for the non-semantically segmented regions or directly predict the category of each pixels. The segments-based approach is typically more efficient as the number of random variables are reduced from number of pixels to the one of segments. The direct approaches [10], on the other hand, provides finner-grain results.

The conventional semantic segmentation approach requires hand crafted or directly learned features. The inference approaches [9, 4] are then performed on these features. These approach are not end to end and requires considerable efforts in feature engineering. In [7], end-to-end approach are proposed for image classification tasks where raw image are feed into deep neural networks and no separate feature engineering are required. The end-to-end convolutional neural network approach is proven the state-of-art in different high level vision tasks including classification, localization and etc. However, it is not directly applied to low level tasks such as pixel-wise labeling. To adapt to the highly effective end-to-end approach to the pixel-wise semantic segmentation, Long et. al proposed fully convolutional neural nets [8] to simultaneously perform classification of all the pixels within an image. It has been successfully applied to natural scene and indoor scene segmentation tasks on VOC and NYU Depth V2 datasets [13].

Specifically for the segmentation problem of the mouse brain slices, Allen Brain Institute [5] has been manually annotating fine regions of the brain and produced a fine detailed volume. An reference slice image is shown in fig. 2. The reference atlas consists of two brains of the same shape - one is the original histological image, and the other is the manually annotated regions. [3] used extended markov random field to segment mouse brain MR slices. [2] uses a proabblistic approach to segment MR microscopic mouse brain images. However NISSL stained images are much noiser than MR images, therefore the method is not quite applicable. The closest work is given by [11] where an random forest is trained on the illumination equalized NISSL stained slices. The result is very good, but the boundaries of each region are still rough even after edge smoothing.

In this project, we applied the fully convolutional approach to the mouse brain NISSL stained histological slice segmentation problem. It is a first attempt to solve pixel-wise neural image segmentation problem using the new end-to-end deep learning approach.

3. Methods

3.1. Fully convolutional network

In conventional convolutional neural networks, the last few layers are typically fully connected layers. Assume the activation output from the previous layer is $\boldsymbol{x}$ =

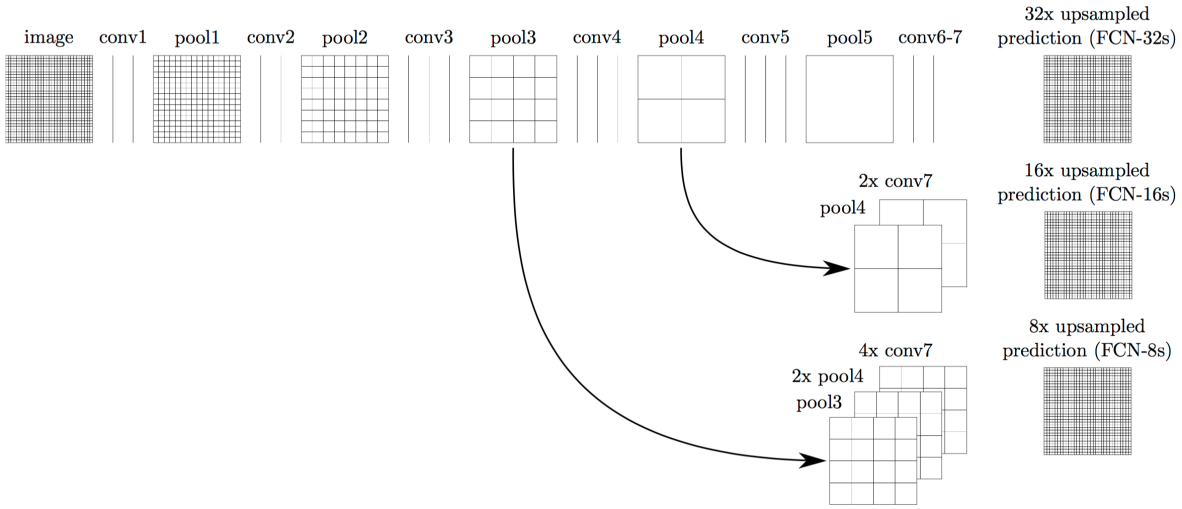

Figure 4. Upsampling for finer-grain prediction.

$(x_1, x_2, ..., x_M)$ while the activation output of the current layer is $\boldsymbol{y} = (y_1, y_2, ..., y_N)$. The fully connected is a general mapping $\boldsymbol{y} = f(\boldsymbol{x})$. It is non-linear if relu or other non-linear operations are added on top of linear mapping. The output activation of a fully connected layer can be considered as a 1D vector. More specifically for the last fully connected layer, each entry of the 1D vector can be treated as the score of the sample being assigned to a specific category. E.g. in fig. 3, the "tabby cat" entry in the 1000-dimensional vector indicates how likely the input image describes a "tabby cat".

As shown in the lower part of fig. 3, we want to produce a score for each pixel indicating how likely the pixel is from a specific category. Thus we need to produce a 2-dimensional map instead of a single entry for each category. As an direct generalization to fully connected layers, we define $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_M\}$ as the output from the previous layer where $\boldsymbol{x}_i$ is a $w \times h$ matrix. The output of the current fully convolutional layer is then computed as $\boldsymbol{y} = (y_1, y_2, ..., y_N)$ with $\boldsymbol{y}_i$ being also a $w \times h$ matrix. These output activations are now 3D volumns where each slice corresponds to a single category. As now the input and output are all 3D volumes, the mapping can be achieved using conventional convolutional operations. This fully convolutional layer provides us a easy-to-plug-in tool to do 2D prediction built upon convolutional nets.

3.2. Upsampling for dense prediction

As multiple pooling layers are employed in convolutional nets, the spatial scale of the activation are substantially reduced compared to the original input. Thus we only get very coarse prediction after. The finer-grain outputs with the same scale as input can only be achieved using upsampling mechanisms. The most simple approach may be bilinear interpolation which is an linear operations. However we can also learn the weights of filters in a data-driven way. A non-linear mapping can even be learnt when these linear interpolations are stacked together with non-linear operation layers.

The computation flow of a deconvolutional layer just exactly reverse the flow of a convolutional layer. In the deconvolutional forward pass, the activation are forwarded from a single node to a spatial range of nodes in the next layer. In the deconvolutional backward pass, gradient are concentrated from a 2D spatial range to a single node, which is exactly the way how convolutional layer perform forward pass. As shown in the first flowing in fig. 4, the data pass through multiple convolutional layers combined with max-pooling. From the last conv layer, deconvolutional operation are performed to recover the original scale using a $32\times$ upsampling.

3.3. Skipping structure for upsampling

As convolutional operations have blurring effects on images and max-pooling layers summarize information within receptive fields. The details in the original images gradually vanish when more convolutional layers are deployed. It is intuitive that better details are preserved in lower pooling layers while the higher pooling layers providing more global information. In the first flowing in fig. 4 which we call 32x model, upsampling is only performed on the final results of the fully convolutional layer. Thus the results lose details and shows an blob effects with blurring. In order to incorporate more details, we may consider the modifications in fig. 4.

In the 16x model, we pool out the pool4 layer and merge it with the upsampled pool5 layer. A upsampling factor of 2 is used to match the scale of pool4 and pool5 layer. After

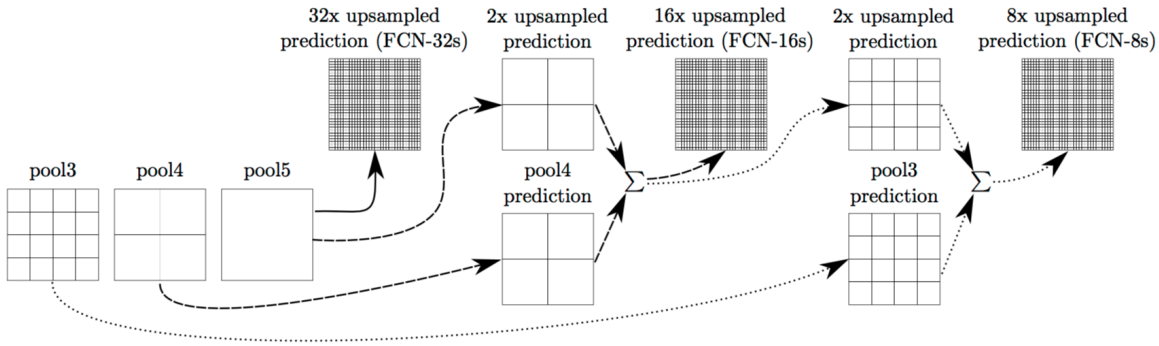

Figure 5. Details in merging layers for upsampling.

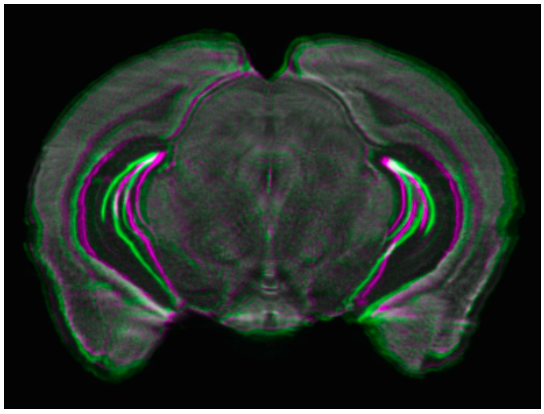

Figure 6. An overlay image of an experimental slice registered to its corresponding atlas slice. Purple slice is the experimental slice warped to the atlas slice; green slice is the corresponding atlas slice.

merging, another overall 16x upsampling is performed to recover the original scale.

As an augmentation to 16x model, we upsample the merging results of pool4 and pool5 with a factor of 2 and merge it with pool3. The overall merging results will be upsampled with a factor of 8 which constructs the 8x model. The detailed merging strategy is illustrated in fig. 5 as a zoom in to fig. 4

Another benefit of using skip architecture is to enhance the gradients in the early layers which is also reported in res-net [6]. The gradient flows into the early layers using the branches in case the normal flow of gradient diminishes too fast after passing though too many layers in the backward pass.

4. Dataset

The training data is the reference mouse brain volume from Allen Institute. The whole brain consists of 528 accurately annotated slices. Each slice is about 300 pixels in height and about 400 in width. We also got data from the Luo Lab in the biology department referred to as experimental data in this report. The reference mouse brain volume is well annotated with regions. The labels are of hierarchical manner and were processed to get the higher level region name. The experimental data is not annotated. Due to manual work needed for accurate labeling, we used freeform based nonrigid registration on the experimental images to register each experimental image to the corresponding atlas slice. The result of registering an experimental slice to its corresponding atlas slice is shown in fig. 6. Even though the hippocampus is not very accurately registered, the result will not be affected much since we are only segmenting the main regions. We then used the annotation of the corresponding atlas slice as the label of the experimental slice. The 2D to 3D localization of finding the corresponding atlas slice in the atlas volume for an experimental slice is based on the result given by an unpublished algorithm, and the accuracy of this 2D to 3D localization algorithm has been confirmed by the neuroscientists in the biology department. We obtained roughly accurate labeling through the above method for 35 experimental slices each of which is about 1000 pixels in height and 1500 in width. The images are all originally one-channel and preprocessed to be 3-channel images. They are also resized to 500x500 pixels images with bilinear interpolation. Labels are obtained on the original image and downsampled corrspondingly.

5. Experiments

We randomly select 428 images from atlas as training set. Each pixel in the image is assigned to one of the 6 possible regions: main brain regions - cerebellum, brain stem, cerebrum, ventricle systems, and fiber tracts - and background. The rest of altas dataset (100 images) is used as test set. The atlas dataset has ground truth label for each pixel. To test our generalization, we also test the trained model on an unseen data set - experimental data set. The experimental data are from worse imaging conditions which makes it harder to perform segmentation. The labels for the experi-

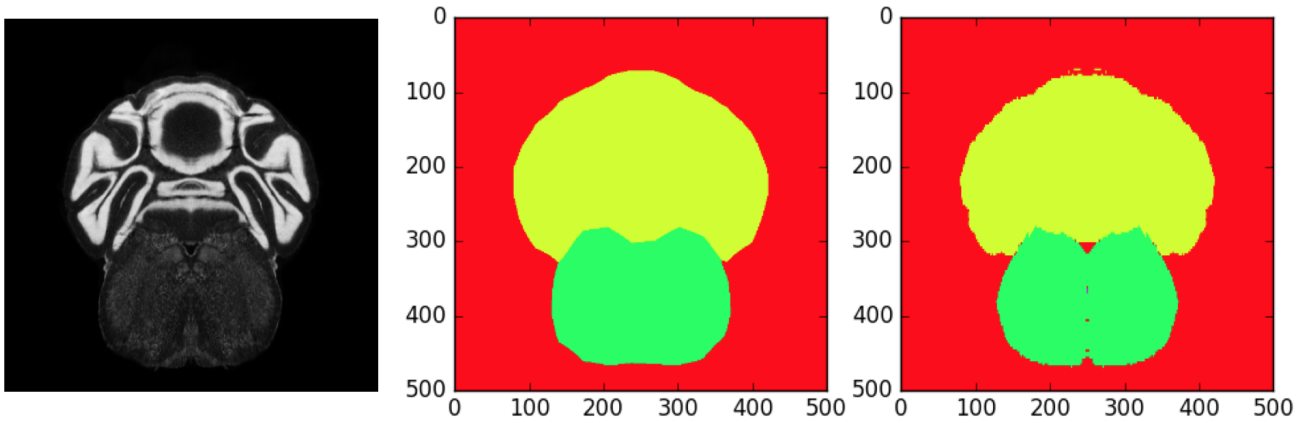

Figure 7. Visualization of the segmentation of an image from atlas test set. Left: original image. Middle: predicted segmentation. Right: ground truth segmentation.

mental images are obtained using the method described in the dataset section.

learning rate	1e-10
learning rate policy	fixed
momentum	0.99
number of epochs	10
regularization	L2
weight decay	0.0005
weight update rule	SGD + momentum

Table 1. Solver Setup

We use the 32-stride version of fully convolutional semantic segmentation network. The network is pre-trained on PASCAL-context 59-class task. We fine-tune the network for 10 epochs on our training set. The solver settings are shown in table 1. We evaluate the predictions using pixel-wise label accuracy:

$$\text{Accuracy} = \frac{\text{\#pixel labels correctly predicted}}{\text{\#pixels in the image}}.$$

The results are as follows:

- Test accuracy on atlas dataset: 96.2%
- Test accuracy on experimental dataset: 91.2%

We visualize our results along with original image and ground truth below. fig. 7 shows the visualization of an image from altas test set. We can see that the predicted segmentation is mostly in accordance with the ground truth. The predicted result get the main regions correct, but misses some details around the boundaries. For example, the the ground truth, it is clear that the lower green regions are composed of two smaller regions, but in the predicted result was not able to differentiate the two regions.

We also visualize one result from the experimental dataset in fig. 8. The performance on the experimental dataset is worse than the performance on atlas dataset. This is especially true with the boundaries. The predicted result does not show enough resolution on the boundaries. For example, the blue regions are all connected as one region in the predicted result. The green region does not have the correct shape in the predicted result.

Since the experimental dataset differs from the atlas dataset, we hand labeled another 25 images from the experimental dataset, and put them into the traininig set. We use the same protocol, and fine-tune the network for 10 epochs. The results are

- Test accuracy on atlas dataset: 96.1%
- Test accuracy on experimental dataset: 92.1%

By fine-tuning on the experimental dataset, we see a increase in the performance on the experimental test set. Since there is not enough training data for experimental images, the performance increase is minor.

6. Conclusion

In this project we have used an existing convolutional network to segment mouse brain NISSL stained slices into main brain regions. We attempted to use this project as a starting point to test the potential of using deep learning techniques to solve the medical annotation problem. Even though the regions we tested are mainly large regions, ventricular systems and fiber tracts are salient small features that biologists and neuroscientists use to make decision on region annotation for other less salient regions. We used Allen reference mouse brain atlas as the main training data and part of the experimental data to train the network. Since the experimental data is not very accurately labeled - labels were automatically generated by nonrigid registration, we

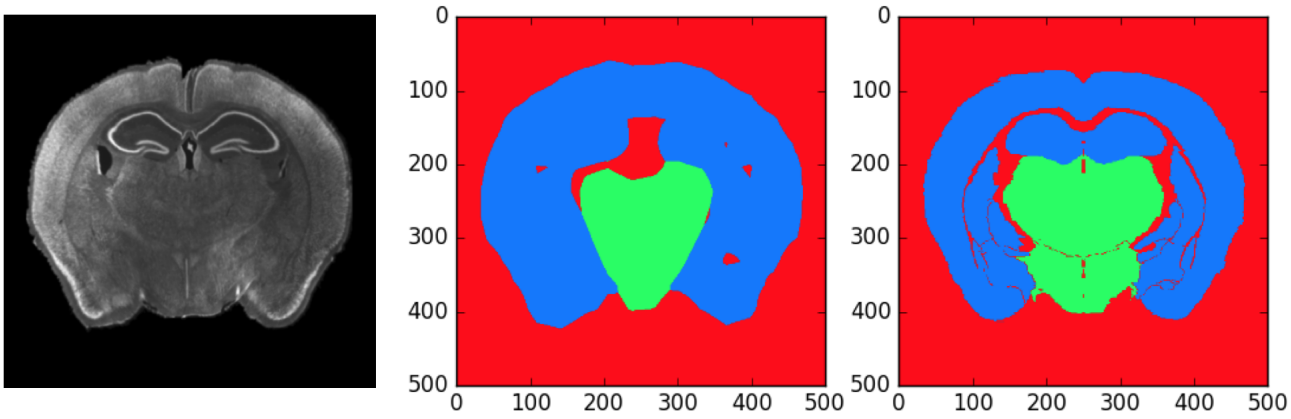

Figure 8. Visualization of the segmentation of an image from experimental test set. Left: original image. Middle: predicted segmentation. Right: ground truth segmentation.

expect the result to be more accurate if an accurately labeled experimental data could be obtained. Even with the roughly labeled experimental data, we obtained an accuracy of 92.1%. The result is generally descent, however the fine-grain details can be further improved.

7. Future work

In our experiments, we observed fully convolutional neural nets achieving high overall accuracy in the brain segmentation tasks. We only predict five major regions which can be further splitted into much more functional areas. Thus we plan on investigating the performance when we scale the number of possible labels to a few hundreds. In order to make the prediction richer in details, we also tried 8s model in addition to the most basic 32s one. Due to the time limit and improper learning rate selection, we did not achieve better results with the 8s model than the 32s model. However we do observed finer-grain prediction from 8s, though the result is typically contaminated by noisy regions resulting from small learning rate with insufficient number of iterations. Thus we also plan to redo the training for 8s model and get a reliable assessment on the upgraded architechture.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. Technical report, 2010.

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

[3] M. H. Bae, R. Pan, T. Wu, and A. Badea. Automated segmentation of mouse brain images using extended mrf. *Neuroimage*, 46(3):717–725, 2009.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[5] H. W. Dong. *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse.* John Wiley & Sons Inc, 2008.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[8] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[9] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 467–475. Morgan Kaufmann Publishers Inc., 1999.

[10] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[11] O. Senyukova, A. Lukin, and D. Vetrov. Automated atlas-based segmentation of nissl-stained mouse brain sections using supervised learning. *Programming and Computer Software*, 37(5):245–251, 2011.

[12] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[13] N. Silberman and R. Fergus. Indoor scene segmentation using a structured light sensor. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 601–608. IEEE, 2011.