Effects of scale on segmentation of Nissl–stained rat brain tissue images via convolutional neural networks

Alexandro Arnal
The University of Texas at El Paso
El Paso, Texas

Olac Fuentes
The University of Texas at El Paso
El Paso, Texas

Abstract

There are numerous efforts to automate the delineation of rat brain regions from rat brain histology. A leading approach uses convolutional neural networks which model anatomical variability and determine cytoarchitectonic boundaries. Currently, it is not clear what scale of the input tissue images offers the most information for these models to exploit. In this work, we test a fully convolutional architecture, U–Net, with Nissl–stained rat brain tissue images of different scales. We show that the networks obtain a lower precision and higher recall when trained on large-scale images. Conversely, networks trained with small-scale images produce fewer false-positive predictions and more false-negative predictions. Our work provides valuable insight into the optimal scale needed for convolutional neural networks to segment brain regions from Nissl-based images of the brain.

Introduction

Segmentation of biological images is often required before any further analysis can occur. In particular, research groups studying the rat brain rely on image segmentation to delineate brain regions and enable analysis, such as cell counting and density estimation of colocalized data. Delineating brain regions also offers the means to validate the results of probing procedures, such as injection deposits and electrode placement. In sum, the segmentation of rat brain regions opens up a world of possibilities for analysis.

Delineating brain regions is often guided by image modalities like MRI and histology. Consequently, a given modality will dictate the kinds of brain segments, anatomical or otherwise, a human or computer can delineate. Currently, histology micrographs of postmortem tissue stained to reveal microanatomy are the state-of-the-art modality for rat brain segmentation. Nissl–stained brain tissue, for example, offers information of cellular patterns used to draw boundaries that group areas with similar cellular or fiber patterns. However, microanatomy to delineate brain regions requires significant human expertise and time.

The task of segmenting rat brain histology from a human annotator’s perspective requires knowledge on several features, two of which include tissue patterns (e.g., orientation, density of staining, cellular morphology) and their relative positions across the brain. From a digital perspective, tissue and position features could translate to information found at different scales. Like other semantic segmentation problems such as remote sensing of earth or segmentation of cancer cells, segmenting brain regions could benefit from data at different scales. Still, no one has addressed how the scale of micrographs affects the task of semantic segmentation. In this work, we investigate how the scale of rat brain micrographs affects semantic segmentation via convolutional neural networks. We hypothesize that a large scale can provide detailed microanatomy information, while a small scale can inform on the relative position of brain regions.
# Related Work

The scientific community studies the brain across different scales. Consequently, researchers have developed several semantic segmentation methods tailored to different scales of the brain. Some researchers delineate individual cells while others segment cell distributions often called brain regions. Here we review the literature to provide context on brain region segmentation across different scales.

We begin with the work of Wagstyl et al. who trained a convolutional neural network to segment cortical layers of a human brain from histology, specifically the BigBrain dataset (Wagstyl et al. 2020). BigBrain is a volumetric reconstruction with an isotropic resolution of 20µm, or about 1 pixel per cell length, and consists of tissue sections stained to reveal cell bodies and other microstructures. Although the species and the brain regions of interest differ from our work, this publication highlights the superiority of histology over an MRI modality for brain region segmentation; the data allows the distinction of cell body distributions providing the necessary information to delineate cortical layers. Moreover, they do not address the problem of brain regionalization as a segmentation problem but a classification one, where the network’s input belongs to only one class. This choice allows them to increase the number of unique training samples while producing coarse predictions. And finally, they show that decreasing the resolution of the dataset to 40, 100, 200, 400, and 1,000 µm per pixel gradually impairs the segmentation performance.

Next, Xiong et al. trained a fully convolutional network to segment larger brain structures, such as gray and white matter and ventricular systems (Xiong, Wang, and Zhang 2016). They use histology tissue sections of a mouse brain, specifically the Nissl–stained sections from the Allen Brain Atlas. Compared to the work of Wagstyl et al., where they use information of cell body patterns to delineate human brain regions, Xiong et al. focuses on mouse brain images of small scale. The network inputs here are images with dimensions 500×500 pixels which cover an entire mouse brain tissue section yielding a resolution of about 25µm or less than 1 pixel per cell length. Even though they use histology images with a similar scale as Wagstyl et al., the small mouse brain regions suffer more from loss of cellular detail. As a result, they can only segment the coarse gray and white matter regions instead of more granular brain regions in the Allen Brain Atlas.

Moreover, Senyukov et al. uses Random Forest and Markov Random Fields to segment large brain structures in mice (Senyukova, Lukin, and Vetrov 2011). Similar to Xiong et al., they train with images of entire brain tissue sections from the Allen Brain Atlas, but with section dimensions 270×204 pixels, or a resolution of much less than 1 pixel per cell length. Since their focus is on larger structures, the dimensions are appropriate, as cellular details are unnecessary. In sum, this work shows how classic methods, like Random Forest and Markov Random Fields, can be applied to brain regionalization in a setting where the brain structures are homogeneous in pixel space, such as in small-scale brain images.

Lastly, Spitzer et al. trained a fully convolutional network to segment 13 areas from the histology of the human visual system (Spitzer et al. 2017). Compared to the other works mentioned, the input data for this work has the highest resolution of 2µm, or about 10 pixels per cell length. However, using large-scale images introduces a new problem; two different brain regions could be classified as the same class if they share similar cellular features. Human annotators usually overcome the issue by considering the patches’ relative position in the brain. This work uses a probabilistic atlas prior that the fully convolutional network exploits to disambiguate two distinct brain regions with similar cellular information. Although this seems promising, we will show it may not be necessary to rely on additional information if we transform the data so a smaller scale.

These investigations support histology as the modality of choice for the segmentation of brain regions. Histology offers fine-grained information not present in other modalities like MR. Additionally, histology can also provide coarse-grained information when the resolution is decreased (>20µm) by reducing the scale. Our work aims to provide a complete picture of the segmentation performance of brain regions when the scale is modified. Scientists have already shown the effects of manipulating image scale in other fields, such as remote earth sensing. Ikokou et al., for example, showed there is an optimal scale for their dataset (Ikokou and Smit 2013). However, Hao et al. showed there is no one optimal scale for their dataset due to variation in object size. They argue that multiple scales are necessary to represent different image objects (Hao, Cui, and Wang 2021). We will test if there is a single optimal scale or a benefit of segmenting at various scales.

# Data

We obtained the data for this experiment from the UTEP’s Systems Neuroscience Laboratory in 10 Adobe Illustrator files. Each file is approximately 5700×6500 pixels and contains the tissue section micrograph and its corresponding manual annotation. The micrographs are coronal sections of a rat brain at the level of the hypothalamus. The laboratory photographed the sections with the BX53 Olympus microscope at 4× magnification. The final images have a resolution of about 1µm, or about 20 pixels per cell length. The manual annotations for each brain region are the ground truth segmentations we used during training.

In this work, we focus on seven brain regions: fornix (Vesalius, 1543a), the suprachiasmatic nucleus (Spiegel & Zwieg, 1919), anterovernal thalamic nucleus (>1840), paraventricular thalamic nucleus (>1840), lateral globus pallidus (>1840), magnocellular nucleus (Swanson, 2004), and lateral hypothalamic area, anterior region, intermediate zone (Swanson, 2004).

We split the ten tissue sections into a testing and training set, five micrographs for training and five for testing. Since the dataset is an ordered series of tissue sections, adjacent ones are similar. Thus, we alternate the sections during the data split to ensure the proper distribution of brain regions between train and test sets. Figure 2 illustrates the data for this work as well as the data split.
Figure 2: Original dataset with human annotation of seven brain regions shown front to back of the brain from left to right. Alternating sections of the tissue series were split into the testing set and training set to balance out the number of brain regions for each set. Abbreviations: \(\text{fx}\) = fornix, \(\text{sch}\) = suprachiasmatic nucleus, \(\text{av}\) = anteroventral thalamic nucleus, \(\text{pvt}\) = paraventricular thalamic nucleus, \(\text{gpe}\) = lateral globus pallidus, \(\text{ma}\) = magnocellular nucleus, and \(\text{lhaai}\) = lateral hypothalamic area, anterior region, intermediate zone.

Figure 3: Data transformations for new datasets. **First row** shows an image of the new datasets after applying the corresponding transformations to the original dataset. **Second row** shows 512×512 frames from the images above to highlight the effects of the transformations. The input dimension for U-Net is 512×512 pixels, so the images in the second row also show what a training example would look like.
Figure 4: We trained one U–Net model per scale, per brain region, for a total of 28 models. We collected individual validation IoU curves during training and averaged them per scale.

Finally, to test our hypothesis, we generated three additional datasets by reducing the scale from the original dataset. We resized the original dataset by 70%, 40%, and 10%, and used all four datasets individually to train respective CNN models. Moreover, we perform our experiments per brain region for a final total of 28 CNN models. Figure 3 illustrates the different scales.

Data Sampling
We sampled image patches with equal dimensions as the CNN model inputs. In this case, the input dimensions are 512×512 pixels. Note that we cannot sample 512x512 patches from images less than 9% of the original scale. The following sampling scheme is identical for all datasets (training and validation sets) and across the seven brain regions.

For a given image–label pair, the sampling program uses the label to find the center of the brain region and its height and width. The program then uses these values to set a valid sampling region. Specifically, we first get the bounding box of the brain region. We then construct a circular sampling area centered with the brain region and a radius equal to half the euclidean distance of the bounding box’s top left and bottom right corners. This scheme ensures that most training samples contain the brain region in question. It also prevents training with the entire micrographs for some datasets depending on the brain region size. Therefore, the models are only aware of the portion of the training set containing their assigned brain regions. Finally, we purposefully avoid any data augmentation other than the random sampling with replacement; our goal is not to obtain the best possible segmentation or prevent overfitting but rather study the effects of scale and resolution.

Model
For this work, we use the U–Net architecture first introduced by Ronneberger (Ronneberger, Fischer, and Brox 2015).

Since its debut, U-Net has been applied to several domains and incorporated into other architectures. U-Net is a fully convolutional network, and convolutions have shown to be excellent tools to extract edges, corners, and other features from images. Our intuition is that a convolutional network is appropriate for extracting the necessary information to delineate brain regions (Haralock and Shapiro 1991) (LeCun et al. 1998).

We implemented all experiments with the Python library, Keras (Chollet and others 2015). We trained the models using the Adam optimizer with a learning rate of 1E-5 and set both the gradient clipping parameter and the label smoothing parameter to 0.01. The loss function for all models was binary cross-entropy, a standard loss function for semantic segmentation tasks. We trained each model for 100 epochs with 1000 training samples and ten training samples per batch.

Results
Our goal was to explore what scales are essential for the automatic segmentation of brain regions. To this end, we generated different datasets that alter the information available to a convolutional network as seen in Figure 3.

The original dataset has a resolution of 1µm, or about 20 pixels per cell length. A 512×512 pixel sample from this dataset spans a distance of half a millimeter and contains a few cells. Reducing the scale loses cellular details, but it also gains spatial coverage. A 512×512 frame from these datasets spans a length of more than half a millimeter and contains more cells. Specifically:

- 100% scale = resolution of 1µm = 20 px per cell length
- 70% scale = resolution of 1.4µm = 14 px per cell length
- 40% scale = resolution of 2.5µm = 8 px per cell length
- 10% scale = resolution of 10µm = 2 px per cell length

Figure 5: An individualized comparison of the maximum IoU achieved during training across brain regions and scales.
Evaluating on Relative Validation Sets

A standard metric used to measure model performance in semantic segmentation problems is the intersection over union (IoU). The IoU calculates how well the model predictions overlap with the ground truth segmentations, where an IoU of 1 is a perfect overlap, and an IoU of 0 means no overlap. We first consider the average intersection over union (IoU) of the validation set during training to assess scale effects. Specifically, we average the IoU curves of all models trained on individual brain regions for a given scale and plot the average curve for each scale. Figure 4 shows that a scale of 2.5 µm/px yields the highest IoU.

Next, in line with Hao et al.’s findings, we also compared scales for each brain region. Since the brain regions vary in size, we wanted to see if a scale of 2.5 µm/px was, in fact, optimal for all models. Figure 5 shows that different scales yield a better segmentation for different brain regions. Moreover, we ordered the brain regions ascending from left to right on the figure to uncover trends between brain region size and scale. However, it is not evident that object size directly determines an optimal scale.

Evaluating on Complete Validation Set

As mentioned in the data sampling section, each model is trained and validated on a neighborhood of their assigned brain region, guaranteeing each sample contains part of the region. We widen the analysis here by evaluating each model on complete images. Effectively, we examine the models’ performance on a larger portion of negative samples.

We first consider the global IoU for each model across the four scales. Figure 6 shows the IoU significantly declines when evaluating the models with complete images. The model trained on the fornix data, for example, achieves a higher IoU with our largest scale of 1 µm/px. In comparison, the model trained on the Lateral Globus Pallidus data achieves a higher IoU with our smallest scale of 10 µm/px. Moreover, we ordered the brain regions ascending from left to right on the figure to uncover trends between brain region size and scale. However, it is not evident that object size directly determines an optimal scale.
Still, some brain region and scale combinations offer little to no reduction indicating better generalization. We then calculated the average precision per scale to measure the extent of false-positive predictions. **Figure 7** shows there is a slight negative relationship between the scale and producing false-positive predictions from 70% to 10% scale. We repeated the steps to calculate recall. **Figure 9** shows there is a slight negative relationship between the scale and producing false-negative predictions.

Although we see similar results in the literature, where a large scale can provide detailed information and a small scale can offer context, we also see high variability across brain regions. For some brain regions, it is better to have context over local detail. Some prefer the opposite, and other brain regions obtain similar segmentation results across scales. Our intuition of sorting the plots based on brain region size mostly aligns with this "local" and "context" phenomena, but it is evident that other factors are involved.

**Conclusion**

In this work, we studied the effects of scale on the semantic segmentation of Nissl–stained coronal tissue sections of the rat brain via the fully convolutional neural network, U–Net. We found that large-scale images provide detailed cellular information, enabling a higher recall for the average brain region. Moreover, decreasing the scale of histology images results in fewer false-positive predictions. We also showed that this trend is absent or opposed for a few brain regions, and unlike previous results in the literature, brain region size is not a good indicator of these deviations.

Looking forward, initiatives to automate brain region segmentation should not assume a single scale will be optimal for all brain regions. Further, future efforts should also consider how we define brain regions from Nissl-stained coronal tissue. It may be the case that the underlying data alone is not sufficient to distinguish between similar brain tissue with different labels under the current labeling framework. For example, in the results, the fornix is a white matter tissue that looks almost identical at a large scale with other white matter regions across the brain. As we decrease the scale, the network improves and produces fewer false-positive predictions. However, even at the lowest scale studied, the network incorrectly classifies other regions as fornix. Even if future efforts consider multiple scales, it may also be necessary to incorporate additional information that better aligns with the labeling framework. For instance, including spatial context could help distinguish between similar regions positioned in different locations. Nonetheless, our work is sufficient to highlight the effects of rat brain tissue semantic segmentation across scales.

**References**

Chollet, F., et al. 2015. Keras. https://keras.io.

Eickhoff, S.; Yeo, B. T.; and Genon, S. 2018. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience* 19:672–686.

Hao, S.; Cui, Y.; and Wang, J. 2021. Segmentation scale effect analysis in the object-oriented method of high-spatial-resolution image classification. *Sensors (Basel, Switzerland)* 21.

Haralock, R. M., and Shapiro, L. 1991. Computer and robot vision.

Huang, B.; Reichman, D.; Collins, L.; Bradbury, K.; and Malof, J. M. 2018. Tiling and stitching segmentation output for remote sensing: Basic challenges and recommendations. *arXiv: Computer Vision and Pattern Recognition*.

Ikokou, G. B., and Smit, J. 2013. A technique for optimal selection of segmentation scale parameters for object-oriented classification of urban scenes. *South African Journal of Geomatics* 2:358–369.

Iqbal, A.; Khan, R.; and Karayannis, T. 2018. Developing a brain atlas through deep learning. *Nature Machine Intelligence* 1:277–287.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition.

Liu, Y.; Ma, C.-Y.; He, Z.; Kuo, C.-W.; Chen, K.; Zhang, P.; Wu, B.; Kira, Z.; and Vajda, P. 2021. Unbiased teacher for semi-supervised object detection. *ArXiv* abs/2102.09480.

Paxinos, G., and Watson, C. 1983. The rat brain in stereotaxic coordinates.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *ArXiv* abs/1505.04597.

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

Spitzer, H.; Amunts, K.; Harmeling, S.; and Dickscheid, T. 2017. Parcellation of visual cortex on high-resolution histological brain sections using convolutional neural networks. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 920–923.

Swanson, L. 2018. Brain maps 4.0—structure of the rat brain: An open access atlas with global nervous system nomenclature ontology and flatmaps. *The Journal of Comparative Neurology* 526:935 – 943.

Wagstyl, K.; Larocque, S.; Cucurull, G.; Lepage, C.; Cohen, J. P.; Bludau, S.; Palomero-Gallagher, N.; Lewis, L.; Funck, T.; Spitzer, H.; Dickscheid, T.; Fletcher, P.; Romero, A.; Zilles, K.; Amunts, K.; Bengio, Y.; and Evans, A. C. 2020. Bigbrain 3d atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS Biology* 18.

Wang, Q.; Ding, S.-L.; Li, Y.; Royall, J.; and Ng, L. 2020. The allen mouse brain common coordinate framework: A 3d reference atlas. *Cell* 181:936–953.e20.

Watts, A. G.; Khan, A.; Sanchez-Watts, G.; Salter, D.; and Neuner, C. 2006. Activation in neural networks controlling ingestive behaviors: What does it mean, and how do we map and measure it? *Physiology Behavior* 89:501–510.

Xiong, J.; Wang, F.; and Zhang, J. 2016. Automatic atlas-based segmentation of nissl stained mouse brain sections using convolutional neural network.