

AUTOMATED GREY AND WHITE MATTER SEGMENTATION IN DIGITIZED $A\beta$ HUMAN BRAIN TISSUE SLIDE IMAGES

Zhengfeng Lai^{1*} Runlin Guo^{1*} Wenda Xu¹ Zin Hu¹ Kelsey Mifflin¹
Brittany N. Dugger¹ Sen-ching Cheung² Chen-Nee Chuah¹

¹University of California, Davis ²University of Kentucky

{lzhengfeng, kolguo, wedxu, zinh, kmmifflin, bndugger, chuah}@ucdavis.edu
sccheung@ieee.org

ABSTRACT

Neuropathologists assess vast brain areas to identify diverse and subtly differentiated morphologies. Alzheimer's disease pathologies have different density distributions in grey matter (GM) and white matter (WM), making the task of separating GM and WM necessary to neuropathologic deep phenotyping. Standard methods of segmentation typically require manual annotations, where a trained observer traces the boundaries of GM and WM on digitized tissue slide images using software like Aperio ImageScope or QuPath. This method can be time-consuming and can prevent the analysis of large amounts of slides in a scalable way. In this paper, we propose a CNN-based approach to automatically segment GM and WM in ultra-high-resolution whole slide images (WSIs) by transforming the segmentation problem into a classification problem. Contrary to the traditional image processing segmentation method, our technique is flexible, robust, and efficient with the accuracy of 77.43% in GM and 79.42% in WM on our hold-out WSIs.

Index Terms— Neuropathology, Image Segmentation, Convolutional Neural Networks, Ultra-high Resolution

1. INTRODUCTION

Neuropathologists assess postmortem brain tissue slides to identify diverse and subtly-differentiated morphologies vital for the diagnosis of Alzheimer's disease [1]. The process of using microscopes to assess slides individually and compare between them can be time-consuming. Recently, with the help of digital slide scanners, all the minute details of the the physical tissue slides can be scanned into high-resolution whole slide images (WSIs). Using software (such as Aperio ImageScope and QuPath [2]), trained experts can make annotations of morphologies on WSIs and assess them using algorithmically-computed scores [3, 4]. The analysis of these brain tissue images is essential to study Alzheimer's disease

as it may aid in deeper phenotyping and analysis of the neuroanatomic progression of the disease [5].

One of the pathological hallmarks of Alzheimer's disease is the presence extracellular neuritic plaques in the brain [6–8]. These plaques can be found predominantly in GM but have also been reported in WM [9, 10]. By segmenting GM and WM in WSIs, neuropathologists can study the density distributions of neuritic plaques in them separately. In fact, the need to segment GM from WM in medical imaging has already been well-established, for example, in identifying spinal cord GM from MRI to predict disability in multiple sclerosis [11]. For this research, we focus on the segmentation problem of WSIs obtained using a common histological stain (Hematoxylin and Eosin) in addition to immunohistochemical stained for one of the main protein aggregates in Alzheimer's disease, $A\beta$. Manually segmenting these WSIs is a time-consuming and expensive process that cannot be performed on a large scale [12]. Another issue with manual segmentation is that it could be subjective and have inter-rater reliability issues, as evidenced by the two boundaries (green and yellow) from Figure 1 between GM and WM drawn by two domain experts.

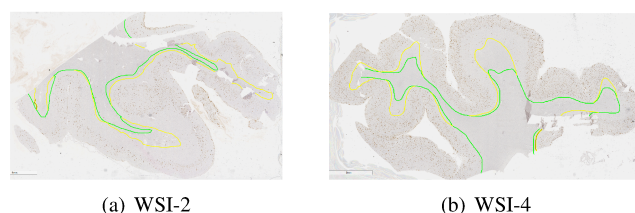


Fig. 1: WSIs Annotated by Two Trained Personnel

As such, there is a strong need to develop automated segmentation algorithms that can run across a database of WSIs efficiently to provide robust and unbiased annotation. Different traditional image processing methods and deep learning-based approaches have been studied for WSI segmentation [13, 14]. However, these studies focused on tissues primarily from breast, lymph node, rectum, tongue [13] and skin [14], or using images obtained from Magnetic Resonance

*The first two authors contributed equally to this paper.

Imaging (MRI) or Computerized Tomography [15, 16].

Segmenting GM from WM in A β stained WSIs is a challenging problem different from these related works. First, there is the prevalence of unwanted artifacts in these WSIs: for example, tissue residues, tissue folds, bubbles, dust, and other items that may result from imperfect tissue cutting and slide placement. Second, they are scanned at ultra-high resolution to retain details down to the cellular level. Typical brain tissues WSIs can resolve down to 0.5 micrometers per pixel, resulting in a single brain slide image with resolutions exceeding 50,000 by 50,000 pixels. Current approaches either downsample an ultra-high resolution image or crop it into small patches for separate processing [17]. These approaches cannot be used for our application because we need to preserve the distribution and characteristics of the minute neuritic plaques in our WSIs, with sizes ranging from 20×20 to 100×100 . These plaques are not only the pathological hallmarks of Alzheimer’s disease but the important features of GM and WM.

In this paper, we investigate the application of region-based CNN segmentation on ultra-high resolution WSIs of immunohistochemically stained brain samples. Our goal is to perform segmentation of GM and WM regions in WSIs with acceptable speed and accuracy. To address the ultra-high resolution challenge, we transform the segmentation problem into a region-based classification problem. To the best of our knowledge, this is the first attempt at automating the separation of GM and WM in ultra-high resolution WSIs of immunohistochemically stained archival brain samples with various shapes and contrasts. The main contribution of the paper is a novel application of deep learning based segmentation techniques to this medical problem. Our results show that the region-based CNN approach can segment GM, WM with the accuracy of 77.43% and 79.42% separately. The rest of the paper is organized as follows: Section 3 describes our proposed approach, Section 4 presents our results, and Section 5 concludes the paper.

2. RELATED WORK

Tissue segmentation is an important prerequisite for diagnosing diseases efficiently and accurately. Many traditional image processing methods have been proposed for whole slide images (WSI) segmentation. In [18], Hiary et al. developed an automated algorithm based on k-means clustering using pixel intensity, color, and texture features. In [14], Bug et al. used global thresholding at the mean value of the Gaussian blurred Laplacian of the greyscale image. These approaches are computationally efficient but the segmentation performance is usually quite poor. In addition, these methods have only been tested on non-brain WSIs such as breast, tongue, and skin so their applicability to A β stained WSIs is questionable.

In recent years, deep learning methods have been in-

creasingly applied for medical image segmentation. In [15], Havaei et al. developed a fully automatic brain tumor segmentation methods using a highly efficient Convolutional Neural Network (CNN) architecture to simultaneously capture both local features and global context. In [19], Milletari et al. proposed Hough-CNN, a CNN architecture to perform Hough voting to simultaneously localize and segment deep brain regions in MRI and ultra-sound images. A drawback of these works is that their deep networks are specially design for specific medical applications (tumor detection) or image modality (MRI and CT), making them difficult to transfer to our A β stained WSIs. Unlike MRI or CT scans that capture a global view of the whole brain, our WSIs are local representations of one region of brain and contains subtle and gradual changes of pathological features that can be easily missed by these networks. Also, image artifacts including large variance in staining color and the presence of tissue residues, dust, tears, folds and bubbles can significantly impact the performance of these classifiers.

Besides CNN, another type of deep neural network architecture commonly used in medical imaging is U-Net [20]. For example, in [21], Dong et al. proposed a fully automatic method for brain tumor segmentation. A drawback of U-net is that it is computationally intensive and could take much processing power with little gain in performance [22]. In addition, most of the deep-learning based approaches have only been tested on low to medium resolution images of up to a few megapixels, while the resolutions of our WSIs exceed 2 gigapixels. Downsampling used in [17] is not applicable as we need to preserve minute plaque features in our WSIs. Recently, we proposed a region-based CNN classifier robust against gradual changes in WSI to distinguish three different plaques [23]. In this paper, we followed a similar approach for WM/GM segmentation so as to build a complete pipeline in support of deeper phenotyping of the neuroanatomic progression of Alzheimer’s disease.

3. PROPOSED APPROACH

In this section, we will introduce our region-based CNN approach for classifying GM, WM, and tissue slide background regions.

3.1. Data

Our dataset consists of 18 WSIs (approximately $60,000 \times 50,000$ pixels each on average) from the temporal cortex anatomical region annotated by two trained neuropathology personnel. These 18 WSIs are postmortem human brain tissues, and all of them have been de-identified. All WSIs used in this paper were stained with an Amyloid-beta (A β) antibody [24] and were digitized using Aperio AT2 at $20\times$ magnification. All cases had a clinical-neuropathological diagnosis of Alzheimer’s Disease, with 10 males and 8 females,

and an average age at death of 84 ± 7 years. To protect data confidentiality, we will refer to them as WSI-1 to WSI-18 in the rest of our paper. 5 randomly selected WSIs are used for training and validation while the remaining WSIs are used for hold-out testing

Considering the variations of original WSIs, we applied color normalization to diminish the effect of slide-color [23]. As the native resolution of a WSI is too high to use as the direct input for CNNs, we divide our WSIs into smaller fixed-size 256×256 pixels image tiles for training with three labels: WM, GM, and background. Damaged areas in the image are excluded. The entire tile image dataset consists of 63,875 local region tiles and they are randomly split into two sets: training (51,100) and validation (12,775). To mitigate class imbalance, our training set includes a compatible number of tiles from three different regions: 20,100 tiles from GM, 21,000 tiles from WM and 10,000 tiles from background area. We apply several data augmentation methods, including random horizontal/vertical clip, rotation, color jitter and affine transformations, to enhance the heterogeneity of the training dataset.

To handle the ultra-high resolution, we use the PyVips library to build image processing pipelines on the original image instead of manipulating it directly [25]. This mechanism allows us to process WSIs without loading the entire image into memory at once. After the full processing pipeline is completed, the entire pipeline will execute at once by streaming the image in parallel from the first pipeline to the last pipeline simultaneously.

3.2. Architecture

There are a myriad of different CNN architectures that have been proposed in the past few years for image classification, including AlexNet, LeNet, VGG, ResNet, Inception, SqueezeNet to name just a few [26]. While the designs of different architectures have very different motivation, there are increasing evidence that the features extracted by these state-of-the-art architectures are quite similar and their performances are starting to converge [27]. As such, we want to experiment with a relatively simple architecture to minimize training and testing time while producing reasonably good performance. We have selected ResNet-18 as our target network as it has fewer filters and lower complexity than other architectures with similar performance [28].

We modified the ResNet-18 model by redefining the last fully connected layer to output three classes: GM, WM, and tissue slide background. We used pretrained parameters of ResNet-18 except for the last layer as our initialization because the pretrained model has learned rich features from ImageNet and was already able to extract useful features for natural images. We adapted the whole CNN using all training data for 20 epochs and subsequently trained on clearly identifiable image tiles far away from the boundary of WM and

GM with more distinctive features for 10 epochs. Using the pretrained model not only guarantees the convergence of our model but also improves our training speed to save computing sources and energy. In the training process, we selected Adam optimizer and the initial learning rate is 0.001. The batch size is set to 16. We trained the whole network for 20 epochs. Our model is implemented with PyTorch and ran on Google Colab.

While the training of ResNet-18 is fairly straightforward, the inference process used in actual WSI segmentation is still quite complex and not entirely suitable for interactive visualization. As such, we compare the ResNet-18 with two simple CNN models with only 3 and 5 layers respectively. These simpler models have the advantage of far lower complexity in both training and testing. For 3-layer model, there are two convolutional layers with the kernel size of 5×5 and one max-pooling layer with the kernel size of 2×2 , subsequently connecting to the fully connected layer. For 5-layer model, there are two convolutional layers with the kernel size of 5×5 , one with 3×3 , and two max-pooling layers with the kernel size of 2×2 as well as the fully connected layer. The learning rate used here is 0.01 and the optimizer is Adam.

4. EXPERIMENTAL RESULTS

4.1. Region-based Accuracy and Segmentation

Table 1 summarizes the GM and WM classification accuracy and the combined accuracy of the different CNN models we studied. As expected, the models with fewer convolutional layers (3-layer and 5-layer) do not perform well in distinguishing GM and background as GM is similar visually to tissue slide background, which indicates the limitation of classification ability of small CNN models although they run faster. By employing transfer learning with ResNet-18, the accuracy increased to 81.12%. Augmenting ResNet-18 with 10-epoch retrain further improves the performance to 90.00% in overall accuracy and over 93% for WM classification while the accuracy of GM classification (88.10%) is well-balanced compared to that of WM.

CNN Model	Overall	GM	WM	Back
3-Layer	70.21%	65.49%	85.32%	65.12%
5-Layer	74.00%	67.10%	86.91%	69.20%
ResNet-18	81.12%	77.99%	89.50%	80.20%
Res-Retrain	90.00%	88.10%	93.85%	93.10%

Table 1: Accuracy of Different CNN Models

To visualize the distribution and location of GM and WM in a broader context, we apply a sliding window approach with the step of 128 to generate a prediction matrix of original WSI. In essence, each element of matrix denotes the category (GM/WM/Background) of the corresponding area of

WSI. We plot the binary image of GM and WM based on

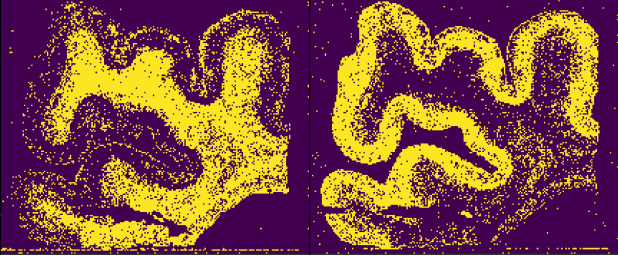


Fig. 2: Binary Image of GM and WM based on Prediction of WSI-16

the prediction matrix generated by CNN model, which will be visualized from the subtitle resolution up to the full WSI view. Figure 2 is the binary image generated by our prediction, which is the shape of GM and WM separately. In Figure 2, yellow pixels indicate GM in the image on the left, and WM in the image on the right.

We also test our CNN model on the 13 hold-out WSIs from different patients not used in the preparation of the training set (Section 3.1). Figure 3 shows the global view of GM/WM’s shape and boundary in different WSIs after we plot the prediction matrix where each element refers to the individual local region tiles cropped from the original WSI. Figure 3(a) shows the manually annotated GM/WM regions for WSI-10, while Figure 3(b) shows the segmentation results of our region-based CNN model. Our proposed approach successfully segments the whole WSI into three areas: background, GM and WM, which are indicated by black, yellow, and cyan, respectively. Our segmentation is almost the same as annotations by trained personnel. Figure 3(c)-(f) show our segmentation results for some of the other WSIs. Figure 3 clearly shows the inherent heterogeneity of the brain images in the different WSIs, where the WM/GM regions have a variety of shapes and sizes.

In our approach, the data preparation and training phase consumes the most computation time. The testing phase, due to use ResNet-18 which is light and of fewer layers, including generating prediction matrix and recovering of the boundary, only takes 30-45 minutes if we use Google Colab’s resource, which is faster compared to 1-2 hours of manual annotation [29].

4.2. WSI Evaluation

Besides using region-based accuracy and heat maps to evaluate our models, we generated binary masks of GM and WM, respectively. Using the contour line annotated by an experienced neuropathologist as the gold standard, we generated two ground-truth masks for GM and WM. Sample masks for one of the WSIs (WSI-18) is shown in Figure 4.

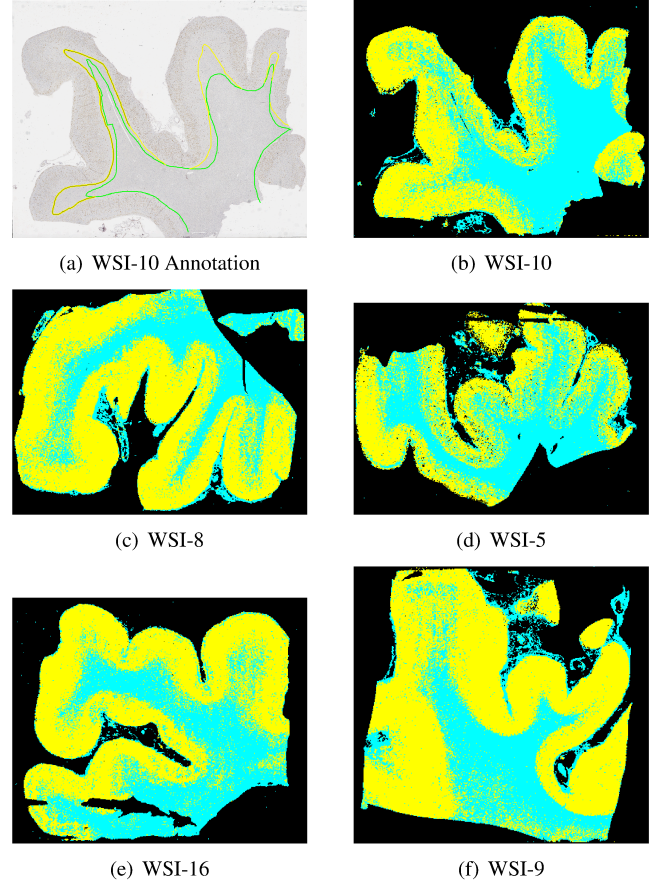


Fig. 3: CNN Segmentation of GM/WM and Background

By comparing these binary masks, we calculated the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for both GM masks and WM masks. Using Equation 1 below,

$$Accuracy = \frac{TP + TN}{total} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

we calculated the accuracy for both masks. The results are summarized in Table 2 below.

From Table 2, we can see that the average accuracy for GM masks is 77.43%, slightly lower than the accuracy for WM masks (79.42%), which indicates that our model has almost the same distinguishing ability in GM and WM. Table 3 is the recall, precision and F1 score of GM and WM. Although WM has relatively higher accuracy, it also has much lower precision and F1 score.

Category	Recall	Precision	F1 Score
GM	87.43%	79.63%	0.8335
WM	83.18%	44.54%	0.5801

Table 3: Measure the performance on 18 WSIs

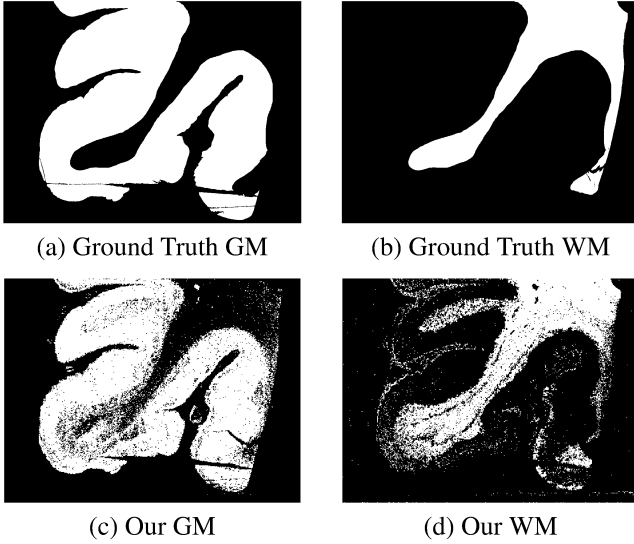


Fig. 4: WSI-18 GM Masks and WM Masks

WSI ID	GM Masks	WM Masks
WSI-1	66.77%	80.75%
WSI-2	75.04%	78.06%
WSI-3	73.12%	74.39%
WSI-4*	84.81%	87.35%
WSI-5*	81.14%	82.58%
WSI-6*	83.01%	83.36%
WSI-7	66.60%	71.22%
WSI-8	85.87%	85.20%
WSI-9	86.24%	86.62%
WSI-10*	86.89%	86.83%
WSI-11	68.83%	71.53%
WSI-12	84.19%	85.02%
WSI-13	81.36%	80.48%
WSI-14*	42.7%	45.96%
WSI-15	78.33%	78.31%
WSI-16	84.22%	84.79%
WSI-17	79.13%	79.49%
WSI-18	85.53%	87.18%
Average	77.43%	79.42%

Table 2: Accuracy for GM and WM Masks of 18 WSIs (* - The 5 training WSIs)

5. CONCLUSION

In this paper, we present an automatic WSI segmentation method based on convolutional neural networks. This method is tailored to ultra-high resolution histological images and has the potential to be more cost-effective than manual segmentation. However, we only tried three light CNN architectures to pursue a trade-off between running speed and accuracy. A

more comprehensive performance analysis on a wider range of CNN architecture is currently underway.

Another limitation of our study is the variety of data sources. The 18 WSIs contain images from one brain region (temporal cortex) and from one type of stain ($A\beta$) as well as from a limited group of cases with similar diagnoses (Alzheimer’s disease). A future direction is to investigate if the same pipeline is applicable to a more diverse dataset from different parts of the brains with various clinical conditions.

6. REFERENCES

- [1] B. N. Dugger and D. W. Dickson, “Pathology of neurodegenerative diseases,” *Cold Spring Harbor perspectives in biology*, vol. 9, no. 7, p. a028035, 2017.
- [2] P. Bankhead, M. B. Loughrey, and J. A. Fernández. *et al*, “QuPath: Open source software for digital pathology image analysis,” Dec 2017.
- [3] C. Higgins, “Applications and challenges of digital pathology and whole slide imaging,” *Biotechnic & Histochemistry*, vol. 90, no. 5, pp. 341–347, 2015.
- [4] L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan, “Review of the current state of whole slide imaging in pathology,” *Journal of pathology informatics*, vol. 2, 2011.
- [5] N. Iwamoto, E. Nishiyama, J. Ohwada, and H. Arai, “Distribution of amyloid deposits in the cerebral white matter of the alzheimer’s disease brain: relationship to blood vessels,” *Acta neuropathologica*, vol. 93, no. 4, pp. 334–340, 1997.
- [6] J. H. Neltner, E. L. Abner, F. A. Schmitt, S. K. Denison, S. Anderson, E. Patel, and P. T. Nelson, “Digital pathology and image analysis for robust high-throughput quantitative assessment of alzheimer disease neuropathologic changes,” *Journal of Neuropathology & Experimental Neurology*, vol. 71, no. 12, pp. 1075–1085, 2012.
- [7] B. N. Dugger, M. Tu, M. E. Murray, and D. W. Dickson, “Disease specificity and pathologic progression of tau pathology in brainstem nuclei of alzheimer’s disease and progressive supranuclear palsy,” *Neuroscience letters*, vol. 491, no. 2, pp. 122–126, 2011.
- [8] D. Selkoe, “Alzheimer’s disease: Genes, proteins, and therapy,” *Physiological reviews*, vol. 81, pp. 741–66, 05 2001.
- [9] D. W. Dickson, “The pathogenesis of senile plaques,” *Journal of Neuropathology & Experimental Neurology*, vol. 56, no. 4, pp. 321–339, Apr 1997.
- [10] B. N. Dugger and D. W. Dickson, “Pathology of neurodegenerative diseases,” *Cold Spring Harbor Perspectives in Biology*, vol. 9, Jan 2017.
- [11] C. S. Perone, E. Calabrese, and J. Cohen-Adad, “Spinal

- cord gray matter segmentation using deep dilated convolutions,” *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018.
- [12] A. de Brebisson and G. Montana, “Deep neural networks for anatomical brain segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 20–28.
- [13] P. Bándi, R. van de Loo, M. Intezar, D. Geijs, F. Ciompi, B. van Ginneken, J. van der Laak, and G. Litjens, “Comparison of different methods for tissue segmentation in histopathological whole-slide images,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 591–595.
- [14] D. Bug, F. Feuerhake, and D. Merhof, “Foreground extraction for histopathological whole slide imaging,” in *Bildverarbeitung für die Medizin 2015*. Springer, 2015, pp. 419–424.
- [15] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” *Medical image analysis*, vol. 35, pp. 18–31, 2017.
- [16] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep learning for brain MRI segmentation: state of the art and future directions,” *Journal of digital imaging*, vol. 30, no. 4, pp. 449–459, 2017.
- [17] W. Chen, Z. Jiang, Z. Wang, K. Cui, and X. Qian, “Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [18] H. Hiary, R. S. Alomari, and V. Chaudhary, “Segmentation and localisation of whole slide images using unsupervised learning,” *IET image processing*, vol. 7, no. 5, pp. 464–471, 2013.
- [19] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel *et al.*, “Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound,” *Computer Vision and Image Understanding*, vol. 164, pp. 92–102, 2017.
- [20] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [21] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic brain tumor detection and segmentation using u-net based fully convolutional networks,” in *annual conference on medical image understanding and analysis*. Springer, 2017, pp. 506–517.
- [22] R. K. Pandey, A. Vasan, and A. Ramakrishnan, “Segmentation of liver lesions with reduced complexity deep models,” *arXiv preprint arXiv:1805.09233*, 2018.
- [23] Z. Tang, K. V. Chuang, C. DeCarli, L.-W. Jin, L. Beckett, M. J. Keiser, and B. N. Dugger, “Interpretable classification of alzheimer’s disease pathologies with a convolutional neural network pipeline,” *Nature communications*, vol. 10, no. 1, p. 2173, 2019.
- [24] J. F. Poduslo, G. L. Curran, J. A. Peterson, D. J. McCormick, A. H. Fauq, M. A. Khan, and T. M. Wengenack, “Design and chemical synthesis of a magnetic resonance contrast agent with enhanced in vitro binding, high blood–brain barrier permeability, and in vivo targeting to alzheimer’s disease amyloid plaques,” *Biochemistry*, vol. 43(20), pp. 6064–6075, 05 2004.
- [25] K. Martinez and J. Cupitt, “Libvips: A fast image processing library with low memory needs,” 2007.
- [26] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *arXiv preprint arXiv:1901.06032*, 2019.
- [27] D. McNeely-White, J. R. Beveridge, and B. A. Draper, “Inception and resnet features are (almost) equivalent,” *Cognitive Systems Research*, vol. 59, pp. 312–318, 2020.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [29] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European conference on computer vision*. Springer, 2014, pp. 818–833.