

Visual deep learning-based explanation for neuritic plaques segmentation in Alzheimer's Disease using weakly annotated whole slide histopathological images

Gabriel Jiménez, Anuradha Kar, Mehdi Ounissi, Léa Ingrassia, Susana Boluda, Benoît Delatour, Lev Stimmer, Daniel Racoceanu

▶ **To cite this version:**

Gabriel Jiménez, Anuradha Kar, Mehdi Ounissi, Léa Ingrassia, Susana Boluda, et al.. Visual deep learning-based explanation for neuritic plaques segmentation in Alzheimer's Disease using weakly annotated whole slide histopathological images. MICCAI 2022 - 25th International Conference on Medical Image Computing and Computer Assisted Intervention, Sep 2022, Singapore, Singapore. pp.336-344, 10.1007/978-3-031-16434-7_33 . hal-03810578

HAL Id: hal-03810578
https://hal.science/hal-03810578

Submitted on 11 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visual deep learning-based explanation for neuritic plaques segmentation in Alzheimer's Disease using weakly annotated whole slide histopathological images

Gabriel Jimenez[1 [0000−0001−6474−6272]], Anuradha Kar[1 [0000−0002−2543−1697]], Mehdi Ounissi[1 [0000−0003−0655−8112]], Léa Ingrassia[1 [0000−0002−4109−4870]], Susana Boluda[1 [0000−0002−8045−2706]], Benoît Delatour[1 [0000−0002−9910−9932]], Lev Stimmer[1 [0000−0003−2800−839X]], and Daniel Racoceanu[1 [0000−0002−9416−1803]]

Sorbonne Université, Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, Department of Neuropathology, DMU Neuroscience, F-75013, Paris, France
daniel.racoceanu@sorbonne-universite.fr

Abstract. Quantifying the distribution and morphology of tau protein structures in brain tissues is key to diagnosing Alzheimer's Disease (AD) and its subtypes. Recently, deep learning (DL) models such as UNet have been successfully used for automatic segmentation of histopathological whole slide images (WSI) of biological tissues. In this study, we propose a DL-based methodology for semantic segmentation of tau lesions (i.e., neuritic plaques) in WSI of postmortem patients with AD. The state of the art in semantic segmentation of neuritic plaques in human WSI is very limited. Our study proposes a baseline able to generate a significant advantage for morphological analysis of these tauopathies for further stratification of AD patients. Essential discussions concerning biomarkers (ALZ50 versus AT8 tau antibodies), the imaging modality (different slide scanner resolutions), and the challenge of weak annotations are addressed within this seminal study. The analysis of the impact of context in plaque segmentation is important to understand the role of the micro-environment for reliable tau protein segmentation. In addition, by integrating visual interpretability, we are able to explain how the network focuses on a region of interest (ROI), giving additional insights to pathologists. Finally, the release of a new expert-annotated database and the code (`https://github.com/aramis-lab/miccai2022-stratifiad.git`) will be helpful for the scientific community to accelerate the development of new pipelines for human WSI processing in AD.

Keywords: Alzheimer's Disease · Tau aggregates · Neuritic Plaques · Deep Learning · Visual explanation · Whole Slide Images · Segmentation.

1 Introduction

Accumulations of Amyloid-β and tau protein aggregates, such as plaques in the brain gray matter, are well-known biomarkers of the neurodegenerative

Alzheimer's disease (AD) [2]. Quantitative estimation of plaques is typically done by pathologists manually or semi-automatically, using proprietary black-box software from histopathological images of the brain – a time and effort-intensive process prone to human observation variability and errors. In recent times, deep learning (DL) based methods have shown promising results in digital pathology [4] and incredibly high accuracy segmentation of digital whole slide images [6]. In [13], three different DL models were used to segment tau aggregates (tangles) and nuclei in postmortem brain Whole Slide Images (WSIs). The three models included a fully convolutional neural network (FCN), UNet, and Segnet, the latter achieving the highest accuracy in terms of IoU. In [10], an FCN was trained on a dataset of 22 WSIs for semantic segmentation of tangle objects from postmortem brain WSIs. Their model can segment tangles of varying morphologies with high accuracy under diverse staining intensities. An FCN model was also used in [12] to classify morphologies of tau protein aggregates in the gray and white matter regions from 37 WSIs representing multiple degenerative diseases. In [7], tau aggregate analysis was done on a dataset of 6 WSIs with a combined classification-segmentation framework which achieved an F1 score of 81.3% and 75.8% on detection and segmentation tasks, respectively. Several domains in DL-based histopathological analysis of AD tauopathy remain unexplored. Firstly, most existing studies have used DL to segment tangles rather than plaques, which are harder to identify against the background gray matter due to their diffuse/sparse appearance. Secondly, annotations of whole slide images are frequently affected by errors by human annotators. In such cases, a DL preliminary model may be trained using weakly annotated data and used to assist the expert in refining annotations. Thirdly, contemporary tau segmentation studies do not consider context information, which is essential in segmenting plaques from brain WSIs as these emerge as sparse objects against an extended background of gray matter. Finally, DL models with explainability features have not yet been applied in tau segmentation from WSIs. This is a critical requirement for DL models used in clinical applications [1] [14]. The DL models should not only be able to identify regions of interest precisely but also give clinicians and general users the knowledge about which image features the model found necessary that influenced its decision. Based on the above, a DL pipeline for the segmentation of plaque regions in brain WSIs is presented in our study. This pipeline uses context and explainability features with a UNet-based semantic segmentation model to identify plaque features from WSIs.

2 Methodology

2.1 Dataset characteristics

In this work, we analyzed eight whole slide images containing histological sections from the frontal cortices of patients with AD, which were provided by the French national brain biobank Neuro-CEB. Signed informed consent for autopsy and histologic analysis was obtained in all cases from the patients or their family members. The present cohort represents a common heterogeneity of AD cases,

including slides with variable tau pathology (e.g., different object densities), variable staining quality, and variable tissue preservation. Sections of the frontal lobe were stained with AT8 antibody to reveal phosphorylated tau pathology, using a standardized immunohistochemistry protocol. Obtained slides were scanned using two Hamamatsu slide scanners (NanoZoomer 2.0-RS and NanoZoomer s60 with 227 nm/pixel and 221 nm/pixel resolution, respectively) at 40x initial magnification. The slides were used for human-CNN iterative object annotation resulting in about 4000 annotated and expert-validated Neuritic plaques. The labels, extracted in an XML format, constitute the segmentation ground truth.

2.2 Data preparation

From the WSIs, at 20x magnification, patches with two levels of context information were generated using an ROI-guided sampling method. The larger patches (256x256 pixels) capture a broader context containing object neighborhood and background pixels, whereas the smaller (128x 128 pixels) mainly focus on the plaque region without much context information. The amount of context present in each patch is quantified using a ratio of the area of annotated ROI to the total area of the patch. The plaque example in different patch sizes is shown in Fig 1 (note that the bigger patch has additional objects-plaques). In addition, two different normalizations are used and compared: Macenko [5] and Vahadane [11] methods.

A new scheme for data augmentation was implemented based on ROI-shifting to prevent the networks' bias from focussing on the center location of plaques in the patches. Accordingly, the annotated plaque ROIs are shifted to four corners of a patch, producing a four-fold augmentation of each patch containing an object. This augmentation aims to train the UNet models robustly in the presence of variable neighborhood context information, especially when closely-spaced plaque objects are present. An example of this augmentation is shown in Fig 2.

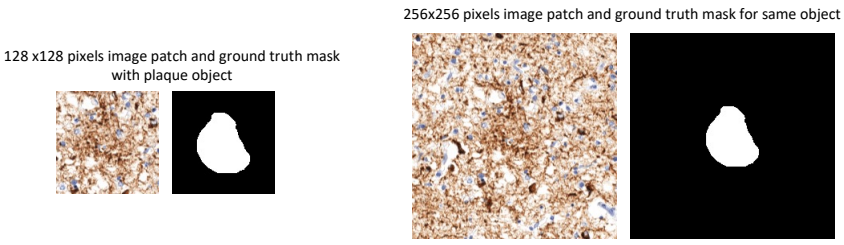


Fig. 1. Example of plaque image for different levels of context.

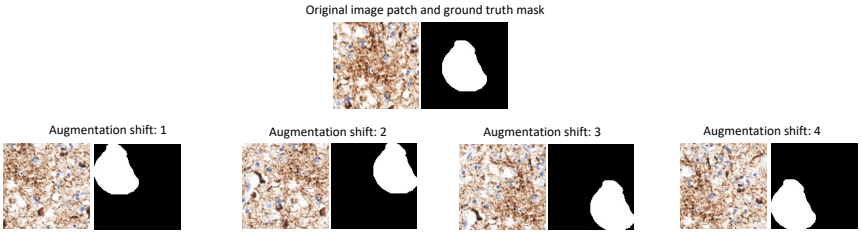


Fig. 2. Example of ROI shifting augmentation.

2.3 Deep learning architecture for segmentation

In order to segment the neuritic plaques, a UNet model adapted from [9] is used with modifications for accommodating context information within the WSI patches during training and testing. The model architecture is modified to work with the two datasets containing different patch sizes – i.e., 128×128 (having low context information) and 256×256 pixels (having more information about the plaque neighborhood). For the first dataset, the UNet architecture consists of 3 downsampling and 3 upsampling convolutional blocks, in addition to the convolutional middle block. For the 256×256-patch-size dataset, we added a downsampling and upsampling convolutional block to the previous UNet model. For the downsampling block, we used a leaky ReLU activation function and ReLU for the upsampling block. In both blocks, we used batch-normalization following the suggestions in [3] and [7]. Dropout was used in each convolutional block with a probability of 0.5.

2.4 Deep learning architecture for visual interpretation

In addition to the segmentation, we focus on deriving locations within the patches where the DL model found significant features from the plaque objects. Therefore, we used an attention UNet described in [8], which allows us to visualize the activated features at each iteration and evaluate qualitatively where the network focuses during training. The attention UNet architecture was also modified for the two different patch-size datasets following a configuration similar to the one described for the UNet.

3 Experiments and results

Data preparation and UNet experiments were executed on an 12-core Intel(R) Core i9-9920X @ 3.5GHz CPU with 128 GB RAM and two 12 GB RAM Nvidia GeForce RTX 2080 Ti GPUs. The attention UNet experiments run on a cluster (1 GPU Tesla V100S-PCIe-32GB, 12 CPU cores Intel(R) Xeon(R) Gold 6126 CPU @ 2.60GHz, and 80 GB of RAM). The average training and evaluation time of the UNet per epoch is approximately 2 minutes for the 128×128 patch-size database and 5 minutes for the 256×256 patch-size database. Meanwhile, for

the attention UNet, approximately half the time is needed. On the other hand, data preprocessing takes 2 to 5 hours to process using parallel computation. Regarding memory consumption, we used at most 6 GB of GPU RAM for the larger patch dataset. In order to increase the performance, we cache the data and annotations first in CPU RAM and then move them to the GPU.

We randomly divided the 8 WSIs into 4 folds for the DL experiments. Then, we tested the network using a 4-fold cross-testing scheme, and with the remaining data from each test fold, we also performed a 3-fold cross-validation. In addition, we run a series of tests (using these folds) to select the loss function and the best optimizer for the UNet and attention UNet. We tested 4 loss functions (i.e., Focal loss, BCEwithLogits, Dice, and BCE-Dice loss) and 4 different optimizers (i.e., SGD, Adam, RMSProp, and AdaDelta). After the hyperparameter tuning, we obtained the best performance using the BCE-Dice loss with a 50% balance between Dice and BCE (Binary Cross Entropy) and the Adadelta optimizer with $\rho = 0.9$ and a varying learning rate based on the evolution of the validation loss. Also, we implemented early stopping for training with a patience value of 15 epochs.

3.1 Results from UNet architecture

The segmentation evaluation metric used for all of the experiments regarding the UNet is the Dice score which is equivalent to the F1 score for binary segmentation problems. In the first experiment, the UNet model was trained with two datasets having different patch sizes: 128×128 and 256×256 pixels. The mean and standard deviations of the Dice coefficient for cross-validation and cross-testing are reported in Table 1. The patches were previously normalized using the Macenko method and then separated in their corresponding fold for training, validation, and testing following the scheme described above. We observe a decrease in the Dice score for larger patches having additional environmental context from the neuritic plaque.

Table 1. UNet results (Dice score) for 4-fold cross testing and 3-fold cross validation for different patch sizes.

Patch size	Normalization	Cross validation	Cross testing
128×128	Macenko	0.6954 ± 0.0289	0.6852 ± 0.0260
256×256	Macenko	0.6600 ± 0.0420	0.6460 ± 0.0330

As described, the WSIs were acquired using two different scanners. Therefore, to study the impact of its properties, we divided the entire cohort into two independent datasets: 4 WSIs belonging to the NanoZoomer 2.0-RS and 4 WSIs scanned with the NanoZoomer s60. For both datasets, we only evaluate the performance of the DL architecture using 4-fold cross-validation and patches of 128×128 pixels size. Additionally, we normalize each dataset independently

(i.e., using two reference patches: one for the NanoZoomer 2.0-RS and one for the NanoZoomer s60) using the Macenko method. The Dice score obtained using the images from the higher resolution Hamamatsu NanoZoomer S60 scanner was 0.6345 ± 0.0243, whereas that from the NanoZoomer 2.0-RS was 0.7342 ± 0.0063.

We also study the effect of normalization in the entire dataset (8 WSIs). We normalized the patches from the 128×128 dataset using Macenko and Vahadane methods, and we selected the best fold (i.e., highest Dice score in testing for the first experiment) to train, validate and test the UNet under different input color properties. Opposite to the results reported in [7], the Dice score obtained was higher using the Macenko method (0.7248 in testing) than the Vahadane (0.7098 in testing), even in validation (0.72313 for Macenko and 0.6864 for Vahadane). For a full list of results, see supplementary material.

3.2 Visual deep learning interpretation

The attention UNet model was trained using the 128×128 and the 256×256 patch size dataset, and the results are summarized in Table 2. All images were normalized using the Macenko method, and we observed a similar trend as the UNet: better performance using patches containing less background information.

Table 2. Attention UNet results (Dice score) for 4-fold cross testing and 3-fold cross validation for different patch sizes.

Patch size	Normalization	Cross validation	Cross testing
128×128	Macenko	0.7516 ± 0.0334	0.6920 ± 0.0254
256×256	Macenko	0.6931 ± 0.0447	0.6342 ± 0.0301

An example segmentation result from the attention UNet model in a 128×128 patch containing a plaque object and its corresponding ground-truth mask is shown in Fig 3. We observe that the attention UNet model finds significant activation features around the plaque object initially annotated by experts (see ground truth mask in Fig 3). We also notice that the loss at iteration 100 increases over iteration 1; however, we clearly distinguish the region of the object (dark red color). After 1000 iterations, the loss decreases 50% due to the fact that the Dice part of the BCE-Dice loss function influences the network into detecting a pattern very similar to the given ground truth.

Another result from attention UNet is in Fig 4. Here, the attention UNet focuses on 2 plaques initially annotated by a human expert. It also identifies strong activation features in regions with no ground truth annotations, which could indicate missed ROIs by human experts during the annotation process. Thus with the attention UNet, it is not only possible to segment the plaque objects but also to improve or refine the manual annotations by experts.

Weak and imprecise annotations are frequently observed in histopathology arising from human or software errors. In such cases, deep learning attention

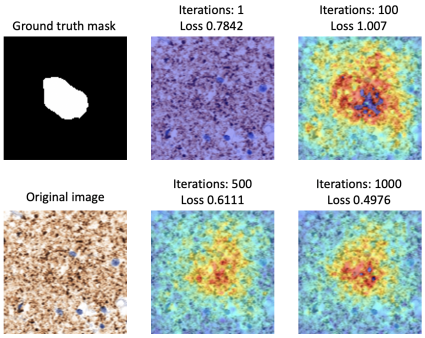


Fig. 3. Global coherence of attention-UNet result with human annotation.

maps could be useful to provide pathologists and biologists with refined annotations (e.g., precision on the boundaries of ROIs). An example is shown in Fig 5 where DL attention maps are closer to the shape of actual ROIs compared to human-made annotations.

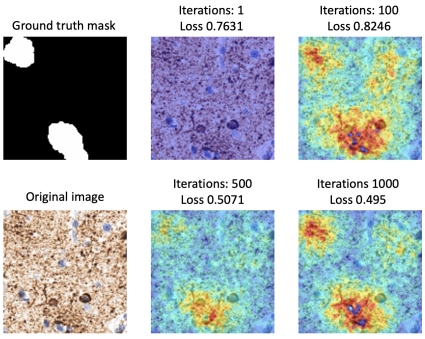


Fig. 4. Focus progression using successive activation layers of attention-UNet.

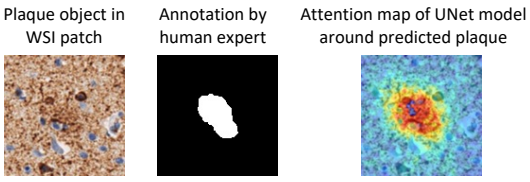


Fig. 5. Improving human annotations using attention-based DL models.

4 Discussion and conclusion

In the presented work, we studied/evaluated a number of factors that contribute to the segmentation of plaques from whole slide images using DL models. The key observations are the following:

1. Use of biomarkers: the study in [7] uses the ALZ50 (used to discover compacted structures) biomarker, while our study uses the AT8 (majorly used in clinics, helps to discover all structures). We focus on AT8 in order to stay close to clinical protocols. The drawback is that this biomarker creates less compact structures meaning a slightly more difficult segmentation of the plaques, as our results support.
2. Use of different modalities: using the AT8 biomarker, we analyzed 2 types of WSI scanners (see Section 2.1) with different resolutions. High-resolution scanners amplify the annotation errors (human-software). Accordingly, some results concerning the high-resolution scanners have been affected, generating lower dice scores.
3. Context effect on results of DL models: We noticed that increasing the background information in the patches negatively affects the segmentation results, which can be explained by the imbalance between the foreground and background pixels. In future works, this will be addressed using adaptive loss functions to take advantage of context information around the ROIs.
4. Attention maps: We observed that using the attention UNet model helps us see the weakness in the human-made annotations (see Fig 5), generating precious insights about the segmentation DL protocol, which can be used to refine the annotations by improving the border of the detected objects. These refined patterns can be used for a morphology and topology pipeline toward a robust AD patient's stratification proof. In addition, quantitative results show better performance of the same UNet architecture with attention blocks.
5. Comparison with state-of-the-art commercial software: We compared our WSI segmentation results with those generated by a commercial software. This software uses a UNet architecture with a VGG encoder which is different from our model. Our system outperforms this software (Dice score 0.63 for test), using the same WSI as the ones used in this paper. Besides, in this software, neither information about how patches are generated nor the type of normalization or pre-processing perfomed on the dataset is available.

Whole slide histopathology images whose sizes range in giga-pixels often contain thousands of objects per image. As seen for plaques in this study, it becomes more challenging when the objects being annotated do not have clear boundaries separating them from their surrounding environments, which may give rise to errors in human-made annotations. We saw an example of how DL models with visual explanation properties can help pathologists refine the ROI identification process. Our future challenge is to create deep learning assistive tools that can improve human-made few and weak annotations, a generic problem of a wide range of biomedical applications.

Acknowlegements

This research was supported by Mr Jean-Paul Baudecroux and The Big Brain Theory Program - Paris Brain Institute (ICM). The human samples were obtained from the Neuro-CEB brain bank (`https://www.neuroceb.org/en/`) (BRIF Number 0033-00011), partly funded by the patients' associations ARSEP, ARSLA, "Connaître les Syndromes Cérébelleux", France-DFT, France Parkinson and by Vaincre Alzheimer Fondation, to which we express our gratitude. We are also grateful to the patients and their families.

References

1. Border, S.P., Sarder, P.: From what to why, the growing need for a focus shift toward explainability of ai in digital pathology. Frontiers in Physiology **12** (2022)
2. Duyckaerts, C., Delatour, B., Potier, M.C.: Classification and basic pathology of alzheimer disease. Acta Neuropathol. **118**(1), 5–36 (Jul 2009)
3. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv **abs/1502.03167** (2015)
4. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J. Pathol. Inform. **7**, 29 (2016)
5. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1107–1110 (2009)
6. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: Challenges and opportunities. Med. Image Anal. **33**, 170–175 (Oct 2016)
7. Maňoušková, K., Abadie, V., Ounissi, M., Jimenez, G., Stimmer, L., Delatour, B., Durrleman, S., Racoceanu, D.: Tau protein discrete aggregates in Alzheimer's disease: neuritic plaques and tangles detection and segmentation using computational histopathology. In: Tomaszewski, J.E., Ward, A.D., M.D., R.M.L. (eds.) Medical Imaging 2022: Digital and Computational Pathology. vol. 12039, pp. 33 – 39. International Society for Optics and Photonics, SPIE (2022)
8. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: Learning where to look for the pancreas (2018)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
10. Signaevsky, M., Prastawa, M., Farrell, K., Tabish, N., Baldwin, E., Han, N., Iida, M.A., Koll, J., Bryce, C., Purohit, D., Haroutunian, V., McKee, A.C., Stein, T.D., White, C.L., Walker, J., Richardson, T.E., Hanson, R., Donovan, M.J., Cordon-Cardo, C., Zeineh, J., Fernandez, G., Crary, J.F.: Artificial intelligence in neuropathology: deep learning-based assessment of tauopathy. Laboratory Investigation **99**(7), 1019–1029 (Jul 2019)
11. Vahadane, A., Peng, T., Albarqouni, S., Baust, M., Steiger, K., Schlitter, A.M., Sethi, A., Esposito, I., Navab, N.: Structure-preserved color normalization for histological images. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). pp. 1012–1015 (2015)

12. Vega, A.R., Chkheidze, R., Jarmale, V., Shang, P., Foong, C., Diamond, M.I., White, C.L., Rajaram, S.: Deep learning reveals disease-specific signatures of white matter pathology in tauopathies. Acta Neuropathologica Communications **9**(1), 170 (Oct 2021)
13. Wurts, A., Oakley, D.H., Hyman, B.T., Samsi, S.: Segmentation of tau stained alzheimers brain tissue using convolutional neural networks. Annu Int Conf IEEE Eng Med Biol Soc **2020**, 1420–1423 (Jul 2020)
14. Yamamoto, Y., Tsuzuki, T., Akatsuka, J., Ueki, M., Morikawa, H., Numata, Y., Takahara, T., Tsuyuki, T., Tsutsumi, K., Nakazawa, R., Shimizu, A., Maeda, I., Tsuchiya, S., Kanno, H., Kondo, Y., Fukumoto, M., Tamiya, G., Ueda, N., Kimura, G.: Automated acquisition of explainable knowledge from unannotated histopathology images. Nature Communications **10**(1), 5642 (Dec 2019)