Articles Describing Code

VesselBoost: A Python Toolbox for Small Blood Vessel Segmentation in Human Magnetic Resonance Angiography Data

Marshall Xu, B.Eng., M.Eng.[1][a], Fernanda L. Ribeiro, Ph.D.[1][b], Markus Barth, Ph.D.[1], Michaël Bernier, Ph.D.[2,3], Steffen Bollmann, Ph.D.[1,4], Soumick Chatterjee, Ph.D.[5,6,7], Francesco Cognolato, Ph.D.[8,9], Omer F. Gulban, Ph.D.[10,11], Vaibhavi Itkyal, M.Sc.[12], Siyu Liu, B.Eng.[1,13], Hendrik Mattern, Ph.D.[5,14,15], Jonathan R. Polimeni, Ph.D.[2,3,16], Thomas B. Shaw, Ph.D.[1], Oliver Speck, Ph.D.[5,14,15], Saskia Bollmann, Ph.D.[1c]

[1] School of Electrical Engineering and Computer Science, University of Queensland, [2] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, [3] Department of Radiology, Harvard Medical School, [4] Queensland Digital Health Centre, University of Queensland, [5] Department of Biomedical Magnetic Resonance, Institute of Experimental Physics, Otto-von-Guericke-University, [6] Data and Knowledge Engineering Group, Faculty of Computer Science, Otto-von-Guericke-University, [7] Genomics Research Centre, Human Technopole, [8] Centre for Advanced Imaging, University of Queensland, [9] ARC Training Centre for Innovation in Biomedical Imaging Technology, University of Queensland, [10] Faculty of Psychology and Neuroscience, Maastricht University, [11] Brain Innovation, [12] Department of Biotechnology, Indian Institute of Technology, [13] Australian eHealth Research Centre, CSIRO, [14] German Center for Neurodegenerative Diseases, [15] Center for Behavioral Brain Sciences, [16] Harvard-MIT Program Health Sciences and Technology, Massachusetts Institute of Technology

Aperture Neuro

Magnetic resonance angiography (MRA) performed at ultra-high magnetic field provides a unique opportunity to study the arteries of the living human brain at the mesoscopic level. From this, we can gain new insights into the brain's blood supply and vascular disease affecting small vessels. However, for quantitative characterization and precise representation of human angioarchitecture to, for example, inform blood-flow simulations, detailed segmentations of the smallest vessels are required. Given the success of deep learning-based methods in many segmentation tasks, we explore their application to high-resolution MRA data and address the difficulty of obtaining large data sets of correctly and comprehensively labelled data. We introduce VesselBoost, a vessel segmentation toolbox, which utilizes deep learning and imperfect training labels for accurate vasculature segmentation. To enhance the segmentation models' robustness and accuracy, VesselBoost employs an innovative data augmentation technique, which captures the resemblance of vascular structures across scales by zooming in or out on input image patches—virtually creating diverse scale vascular data. This approach enables detailed vascular segmentation and ensures the model's ability to generalize across various scales of vascular structures.

1. INTRODUCTION

Here, we introduce *VesselBoost*, a Python-based software utilizing deep learning techniques to segment high-resolution time-of-flight MRI angiography data with high sensitivity towards small vessels. The software suite encompasses three functional modules: (1) *predict*, (2) *Test Time Adaptation* (*TTA*), and (3) *booster*. By leveraging these modules, users can efficiently segment high-resolution time-of-flight data (using either the *predict* or the *TTA* modules) or conveniently improve (or 'boost') segmentations for other vascular MRI contrasts (using the *booster* module).

One of the distinguishing features of *VesselBoost* lies in the idea of incorporating imperfect training labels for vessel segmentation[1-3] to improve segmentation of the **same** image data. At the core of *VesselBoost* is a data augmentation strategy that exploits the similarities between large and small vessels. Training examples for small vessels are created by zooming out from large, easily segmented vessels. Using a segmentation model trained this way improves coarse (less detailed) segmentations and increases the number of segmented small vessels.

[a] Co-first author

[b] Co-first author

[c] Corresponding author:
saskia.bollmann@uq.edu.au

2. METHODOLOGY

2.1. OVERVIEW

VesselBoost comprises three modules: 1) *predict*, 2) *TTA*, and 3) *booster*. These modules are designed to capture different levels of similarity between the original training image data and the new image data. Briefly, *predict* consists of our pre-trained network that can be readily applied to magnetic resonance angiography (MRA) data; if the properties of the new image data are close to the original training data, *predict* can be directly applied to the new image. Next, the *TTA* module can fine-tune the pre-trained network; *TTA* will be useful if the new image's characteristics are somewhat similar, but network adaptation is needed. *Booster* utilises the same pre-processing and data augmentation strategies as the other two modules but trains a new network from scratch. Thus, *booster* is intended for cases where the new image is significantly different from the original training data, for example, when using a different vascular MRI contrast.

2.2. DATA

2.2.1. TRAINING DATA

All pre-trained models were trained on the SMILE-UHURA challenge dataset,[4] which uses the data collected in the StudyForrest project.[5] The dataset consists of 3D multi-slab time-of-flight MRA data acquired on a 7T Siemens MAGNE-TOM magnetic resonance scanner with an isotropic resolution of 300 μm.[5] Twenty right-handed individuals (21-38 years, 12 males) participated in the original study, but we used the 14 samples for model training where corresponding segmentations were made available through the SMILE-UHURA challenge. Moreover, we use the 14 labelled samples in Experiment 2 and Experiment 4, described below, for quantitative evaluations.

We pre-trained three models, each using a specific set of labels, as provided for the SMILE-UHURA challenge: OMELETTE 1, OMELETTE 2, and manually corrected labels. The two sets of OMELETTE labels were generated in an automated fashion[6] using two sets of parameter values. The manually corrected labels were initially generated by intensity thresholding, followed by manual annotation of these segmentations to remove noise and delineate missing small vessels.[4]

2.2.2. EVALUATION DATA

We use a diverse range of image data resolutions—for which ground-truth segmentations are not available—to evaluate all *VesselBoost* modules qualitatively. All evaluation data are 3D multi-slab time-of-flight MRA data acquired on 7T Siemens MAGNETOM whole-body scanners (Siemens Healthcare, Erlangen, Germany) from different participants. They consist of one image acquired at **150 μm** isotropic resolution (TR/TE = 35 ms/6.63 ms, flip angle = 23 degrees),[7] one image at **160 μm** isotropic resolution (TR/TE = 20 ms/6.56 ms, flip angle = 18 degrees),[8] one image

acquired at **300 μm** isotropic resolution (TR/TE = 24 ms/3.84 ms, flip angle = 20 degrees), taken from the Study-Forrest[5] for which no manually corrected segmentation was available, and one image acquired at **400 μm** isotropic resolution (TR/TE = 20 ms/4.73 ms, flip angle = 18 degrees).[8] These four images were used across all experiments.

2.3. DATA AUGMENTATION

Before model training, all MRA data were bias-field corrected and denoised, as described below. Data augmentation was performed to increase the amount of training data and to leverage the similarities between large and small vessels. In our first step, four 3D patches of random sizes and at random locations were first extracted from the input image at each training epoch. Then, each patch was resized to 64×64×64 using nearest-neighbour interpolation (we refer to the resized patch as 'copy 1'). To capture a diverse range of scales for the vessel structures, the minimum crop size was 32×32×32, and the maximum was the dimension size of the original image. This procedure is equivalent to **zooming** in or out for patches smaller or larger than 64×64×64. We generated multiple copies (5 more copies from the original copy, i.e., copy 1) of each of these patches and applied rotation by 90°, 180°, and 270° (copies 2-4) or flipping horizontally and vertically (copies 5 and 6), totalling six copies per patch at each epoch. Therefore, each training image contributed 4 patches × 6 copies per training epoch. Like a conventional training iteration, our training epoch is defined as a complete traverse through the training dataset, meaning all samples (image files) in the training set contributed with augmented patches described above for model training in a given training epoch. By increasing the number of unique patches per training sample and setting the minimum size for each dimension of the cropped patch to 32, we found that the segmentation models were more stable across a range of random seeds used to initialize model weights.

2.4. MODEL ARCHITECTURE

Our segmentation model consists of a 3D U-Net model.[9] We performed several modifications to the 3D U-Net architecture. Previous works[10,11] have shown that adding more convolutional layers to the 3D U-Net model architecture can help capture more complex features, and that deeper networks have improved ability to learn more details and hierarchical features from volumetric data. As such, we increased the default 3D U-Net depth, expanding the encoder and decoder blocks from 3 to 4 layers each. Additionally, the number of input channels was equal to 1 to accommodate a single MRI contrast, e.g., an angiogram, which is a grayscale image, and the number of output channels was equal to 1 for the binary segmentation task. Furthermore, to decrease the total number of parameters, we changed the base number of convolution filters to 16. We implemented these modifications to reduce training time and make the trained model more lightweight while retaining its prediction performance. The models were implemented using Python 3.9 and Pytorch 1.13.[12] This modified 3D U-

Net architecture is provided as a starting point. However, users can easily tune the hyperparameters or change the model architecture based on the image attributes of their data.

2.5. TRAINING PROCEDURE

Each pre-trained model was trained for 1000 epochs at an initial learning rate of 0.001, and a learning rate scheduler, ReduceLROnPlateau (available through PyTorch[12]), was used to reduce the learning rate and make the training process more efficient when the loss reached a 'plateau' during training. The Tversky loss determined the learning objective, where the hyperparameters were set to be $\alpha = 0.3$ and $\beta = 0.7$, to make our models more sensitive to small vessels (potentially leading to higher number of false positives).[1,13]

2.6. VESSELBOOST MODULES

2.6.1. MODULE 1: PREDICT

The prediction pipeline includes input image pre-processing (Figure 1a, step i), image segmentation using a pre-trained model (Figure 1b, step ii), and post-processing (Figure 1b, step iii). The pre-processing step includes bias-field correction using N4ITK[14] and non-local means denoising[15] to increase the signal-to-noise ratio (SNR) (Figure 1a, step i). The model output is post-processed to appropriately convert the predicted probabilities to binary classes by setting the threshold to 0.1. Finally, any connected components with a size smaller than 10 voxels are removed[16] to clean up the final segmentation. The *predict* module can be used to simply segment classical MRA data using our pre-trained model. In addition, it provides extra flexibility for users to manipulate post-processing parameters to obtain a more suitable proxy (an initial segmentation) before, for example, using it for TTA.

2.6.2. MODULE 2: TEST-TIME ADAPTATION

TTA involves adapting pre-trained weights using an initial segmentation (a proxy) to guide parameter optimization (Figure 1b, step ii). The user has the flexibility to define the number of epochs, initial learning rate, type of optimizer, and loss function for model adaptation. The initial learning rate, optimiser, and loss function have default configurations equal to 0.001, Adam,[17] and the Tversky loss (α = 0.3 and β = 0.7), respectively. The default learning rate scheduler is the ReduceLROnPlateau, which automatically reduces the learning rate when the loss reaches a plateau.

The user can either provide a proxy segmentation, in which case *TTA* will enhance the details, or a proxy segmentation is generated using *predict*, which is then further improved. This module is particularly useful for users with access to a small number of labelled MRA data who want to improve small vessel segmentation.

2.6.3. MODULE 3: BOOSTER

Booster (Module 3) allows users to train a segmentation model from scratch using imperfect training labels from a new data set. This module benefits users with access to a small amount of labelled data who want to improve small vessel segmentation for data other than classical MRA images. This module shares the general training settings previously described, but the user can specify the parameters for training (e.g., number of epochs, learning rate) similar to the *TTA* module.

2.7. EXPERIMENTS

2.7.1. EXPERIMENT 1: USING PREDICT TO SEGMENT NEW, OUT-OF-DISTRIBUTION MRA DATA

In our first experiment, we tested the generalizability of the pre-trained segmentation model using *predict*. Specifically, this model was trained using all SMILE-UHURA challenge data with manually corrected labels. We ran *predict* on 3D MRA image slabs with a diverse range of image resolutions (from 400 μm to 150 μm[4,7,8]), and evaluated prediction performance qualitatively.

2.7.2. EXPERIMENT 2: USING TTA TO IMPROVE A PROXY SEGMENTATION

To qualitatively evaluate *TTA* and determine whether *TTA* can improve upon less detailed (or less accurate) initial segmentation, we used a segmentation model pre-trained on all SMILE-UHURA challenge data using OMELETTE 2 labels (see section 2.2. for more details) to generate the initial segmentation. To better capture the potential of TTA for improvement, we used the OMELETTE 2-based model for initial segmentation, as we expected this model to provide less accurate segmentation than the model pre-trained on the manually corrected labels. An initial segmentation was generated for each evaluation data, and the pre-trained model was adapted for each data separately for over 200 epochs using *TTA*, generating data-specific models.

To quantitatively evaluate *TTA*, we trained a new model using 13 MRA images from the SMILE-UHURA challenge and the OMELETTE 2 labels, leaving one sample out for evaluation. This pre-trained model was used for proxy generation (analogous to running *predict*) and adapted for 200 epochs using our *TTA* module and the holdout test sample. Using the manually corrected segmentation as the ground truth, we determined prediction performance by computing the Dice score.[18]

2.7.3. EXPERIMENT 3: USING BOOSTER TO TRAIN A SEGMENTATION MODEL FROM SCRATCH WITH IMPERFECT TRAINING LABELS

In this experiment, we tested whether our *booster* module can provide improved and more accurate segmentation than what is captured in the training data—e.g., an initial segmentation based on simple intensity thresholding—even when only a single image volume is available. We
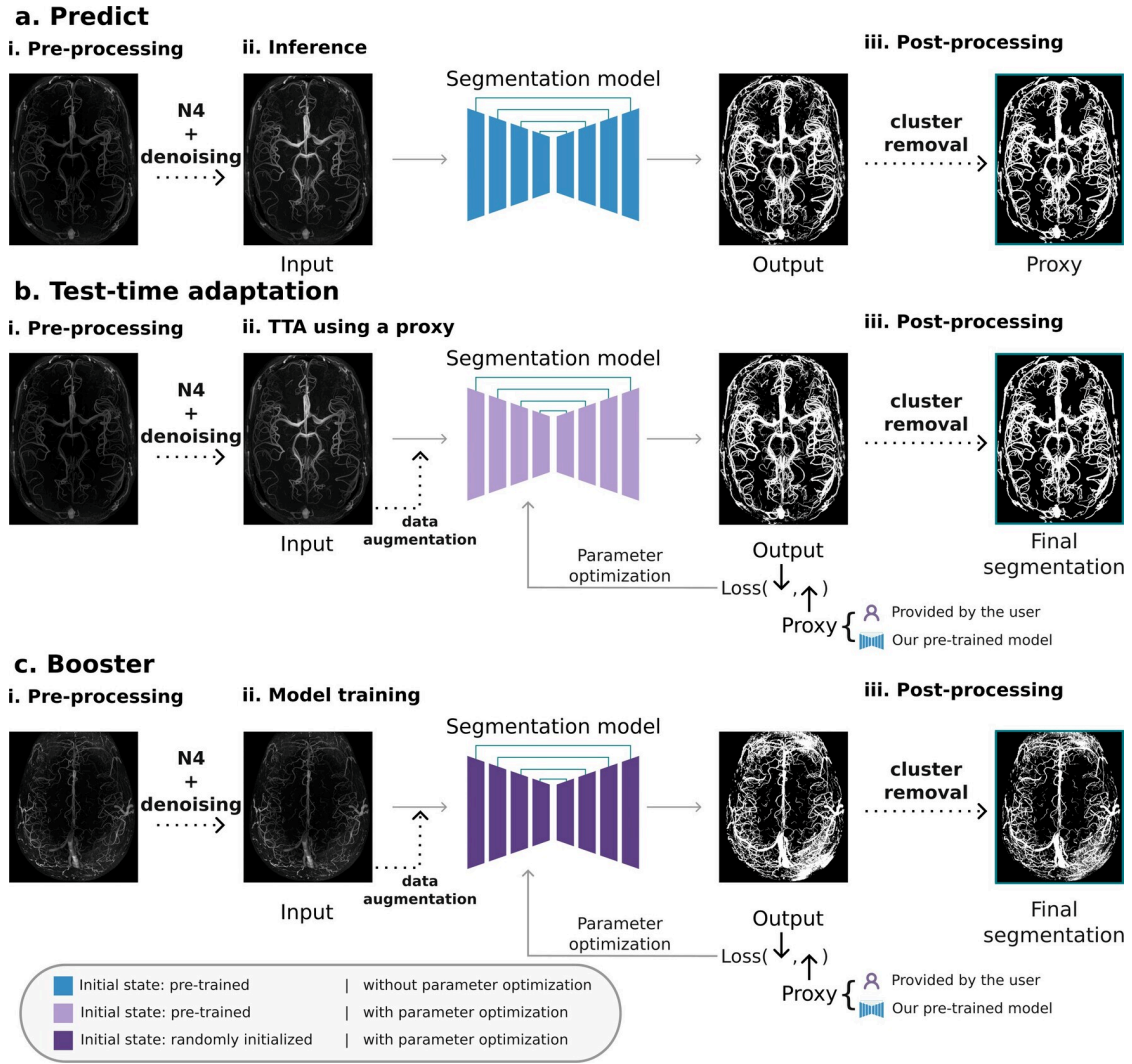
Figure 1. *VesselBoost* overview.

(a) *Predict* allows users to segment high-resolution time-of-flight data using our pre-trained models. (b) The *TTA* module allows users to provide a proxy segmentation to drive further adaptation of the pre-trained models. (c) *Booster* allows users to train a segmentation model on data with imperfect training labels.

trained new models from scratch using *booster* for the 160 μm and 150 μm MRA data, with initial segmentations generated by setting intensity thresholds. These models were trained for 1,200 epochs, keeping it the same as the total epoch number set for *TTA* (1000 epochs for training + 200 epochs for fine-tuning) and setting all other parameters to default values.

2.7.4. *EXPERIMENT 4: INVESTIGATING THE EFFECT OF OUR DATA AUGMENTATION STRATEGY ON SEGMENTATION PERFORMANCE*

Finally, we performed an ablation study to determine whether our data augmentation strategy improved segmentation results beyond the training data. We focused on five distinct settings: **(1)** without data augmentation, **(2)** zoom only, **(3)** zoom with one transformation of the image data that could be either rotation or blurring, **(4)** zoom with copied patches and both rotation and blurring as data transformation, and **(5)** our proposed augmentation setting—zoom with copied patches and both rotation and flipping as data transformation. For settings 1-3, the number of copies per patch was equal to 1, while for settings 4 and 5, the number of copies was equal to 6 (see section 2.3. for more details). Using the SMILE-UHURA challenge data, we evaluated each of these settings in a leave-one-out cross-validation fashion. Specifically, we trained models using 13 MRA images from the SMILE-UHURA challenge and their corresponding OMELETTE 2 labels at each cross-validation fold and evaluated segmentation performance on the held-out sample, resulting in 14 distinct models. Models were trained for 1,000 epochs using default settings.

To test the robustness of each of these settings, we trained models using all 14 MRA images and OMELETTE 2 labels, and ran *predict* on the high-resolution 3D MRA images (150 μm and 160 μm isotropic resolution). Specifically, we trained five distinct models (different initial random seeds) for each augmentation setting. Then, we computed the percentage of voxels that differed between predicted segmentations, for which higher values indicate more variability.
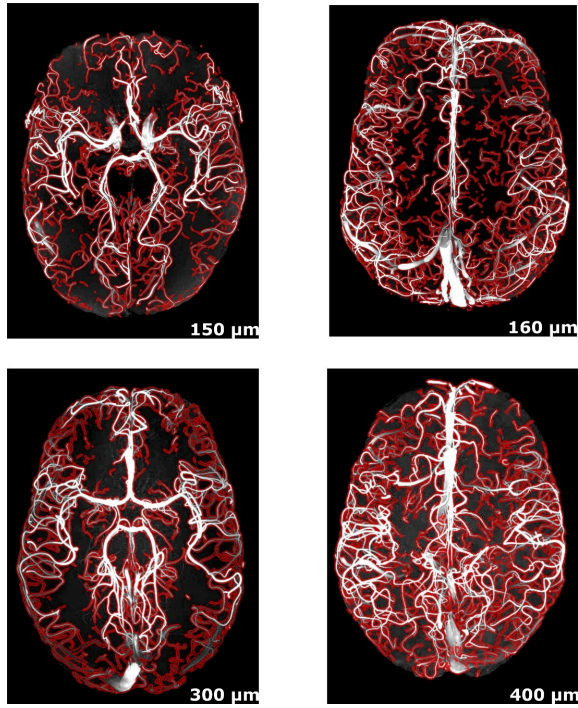
Figure 2. Qualitative evaluation of *Module 1 (predict)*.

Maximum intensity projection of predicted segmentations (red contour) overlaid on the input images with diverse resolutions.

2.8. HARDWARE SETTINGS

The *VesselBoost* segmentation pipeline has been tested with AMD EPYC 'Milan' processors @ 3.70 GHz, and an NVIDIA Tesla H100 graphics card, with 80 GB total requested RAM. During training, we used a dataset comprising 14 images, each with dimensions of [480, 640, 163], for 1000 epochs, an initial learning rate of 0.001 and a batch size of 4. The entire training duration amounted to 2 hours and 35 minutes. For *booster*, the training and inference time for each evaluation data (150 μm and 160 μm resolution data) was 23 minutes. The image dimensions were [312, 972, 1296] and [312, 1090, 1277] for 150 μm and 160 μm data, respectively.

3. RESULTS

3.1. EXPERIMENT 1

To qualitatively evaluate the generalizability of our pre-trained segmentation model used in *predict*, we used 3D MRA image slabs with a diverse range of image resolutions (from 400 μm to 150 μm[4,7,8]). Figure 2 shows the maximum intensity projections (MIP) of predicted segmentation overlaid on the original input images. We found that even though our model was trained on time-of-flight MRA data acquired at a single isotropic resolution (300 μm), it can generalize to MRA images with varying resolutions.

3.2. EXPERIMENT 2

3.2.1. QUALITATIVE EVALUATION

Figure 3 demonstrates that *TTA* can offer additional segmentation improvement when pre-trained models are fine-tuned on automatically generated, imperfect proxy segmentations— for example, generated with *predict*. Figure 3a shows the MIP of the original input images, and Figure 3b shows the initial segmentation of the OMELETTE 2-based pre-trained model. Note how this pre-trained model cannot segment the smallest vessel shown in the 'zoomed-in' patches. Despite being imperfect, these segmentations can be leveraged as proxy segmentations for *TTA*. Accordingly, we found improved segmentation of the smallest vessels (see 400 μm and 300 μm images) and improved segmentation continuity (see 160 μm and 150 μm images) (Figure 3c).

3.2.2. QUANTITATIVE EVALUATION

Figure 4 shows the OMELETTE 2 segmentation (coarse segmentation), the ground-truth (i.e., manually corrected) segmentation, the initial segmentation generated with *predict* and the OMELETTE 2-based pre-trained model, and the final segmentation after *TTA* using the initial segmentation as a proxy. Note that, to remove unwanted false-positive voxels from outside the brain, a semi-automatically derived brain mask was applied. Using the ground-truth segmentation as a reference, the final segmentation after *TTA* showed an increase in the Dice score of 0.04 compared to the initial proxy segmentation. This result demonstrates that imperfect segmentation can be leveraged as proxy segmentations for *TTA*, and *TTA* can improve segmentation performance.

3.3. EXPERIMENT 3

Figure 5 shows the utility of *booster* to train a segmentation model from scratch using imperfect training labels (middle), which were generated simply by thresholding the original image. Note that in contrast to *TTA*, the network has not seen any (MRA) images beforehand and, in this case, is only trained on the example image shown in Figure 5. By leveraging the similarities between large and small vessels through data augmentation, it is possible to train a model from scratch using imperfect labels to improve the segmentation of the smallest vessels, thus improving the segmentation results beyond the training data.

3.4. EXPERIMENT 4

Figure 6 shows the distributions of Dice scores across folds for each of the augmentation settings: **(1)** without data augmentation, **(2)** zoom only, **(3)** zoom with one data transformation that could be either rotation or blurring, **(4)** zoom with copied patches and both rotation and blurring as data transformation, and **(5)** our proposed augmentation setting—zoom with copied patches and both rotation and flipping as data transformation. We also include the Dice score between the OMELETTE 2 labels of each
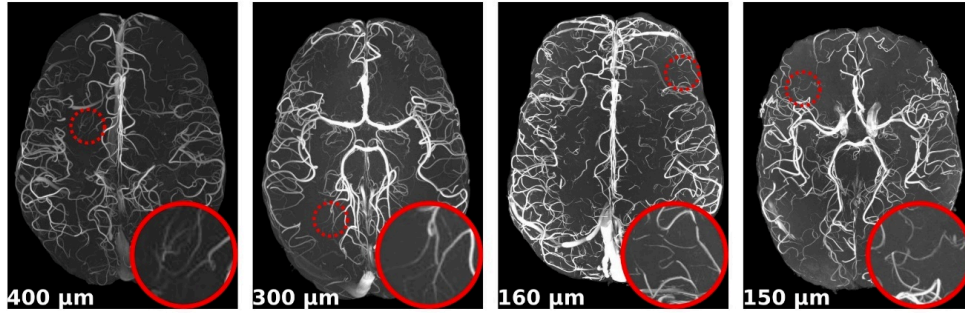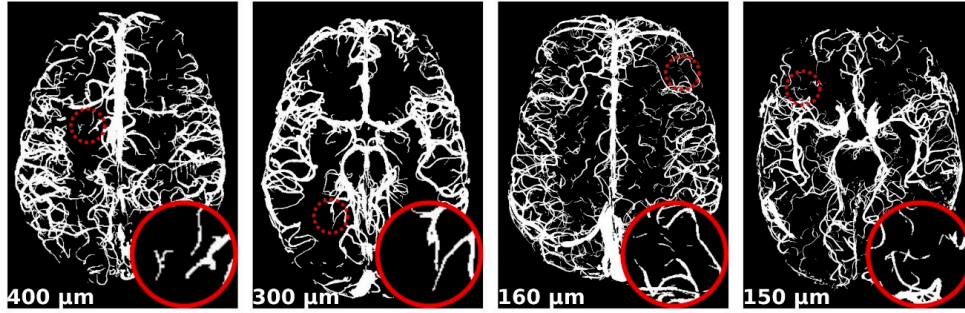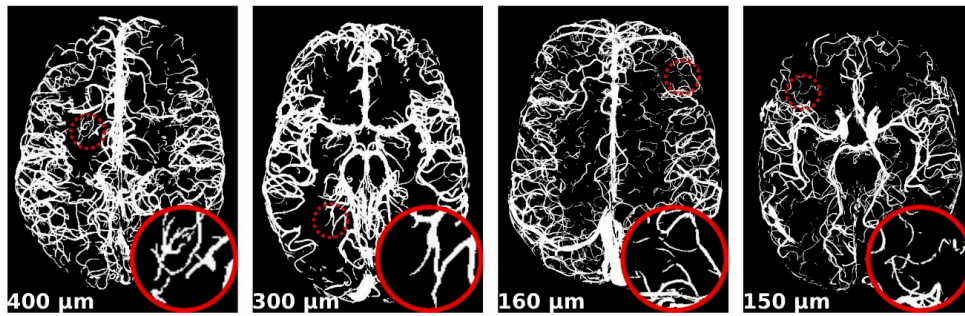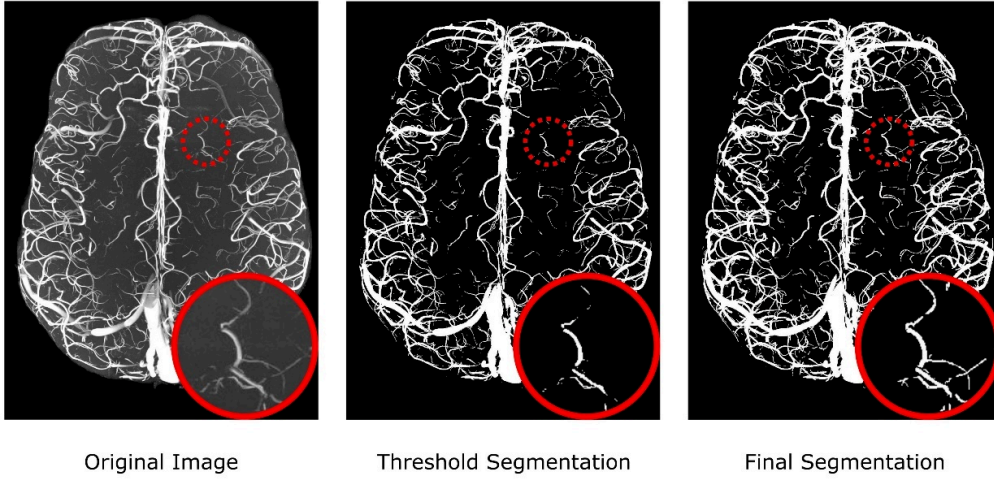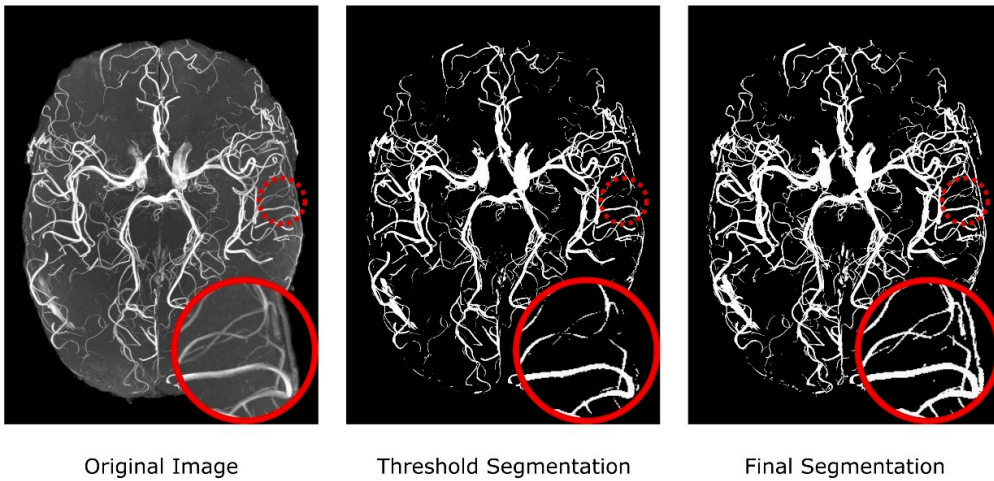
Figure 3. Qualitative evaluation of *Module 2 (TTA)*.

(a) Maximum intensity projection of input images with a diverse range of image resolutions. (b) Predicted segmentations of the OMELETTE 2-based pre-trained model. (c) *TTA*-based segmentation using the segmentation shown in (b) as proxy segmentation to guide model adaptation. The 'zoomed in' patches show the missing small vessels in panel (b) that are recovered with *TTA*.
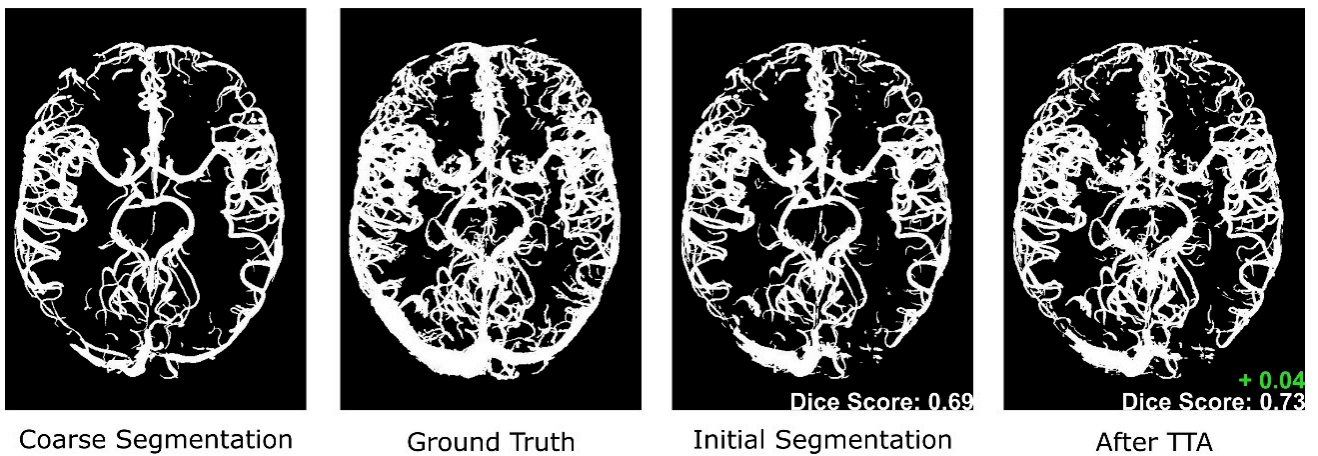
Figure 4. Quantitative evaluation of *Module 2 (TTA)* on 3D MRA image slab with an isotropic resolution of 300 μm.

Dice scores were estimated for the initial segmentation (proxy) and the final segmentation (after TTA), with the ground truth image as a reference. In green, we show the boost in Dice score after Test-Time Adaptation.

a. 160 μm

Original Image   Threshold Segmentation   Final Segmentation

b. 150 μm

Original Image   Threshold Segmentation   Final Segmentation

Figure 5. Qualitative evaluation of *Module 3 (booster)*.

Maximum intensity projection of input images (left), imperfect segmentation (middle), and predicted segmentation (right) for 160 μm (panel **a**) and 150 μm data (panel **b**) using models trained from scratch. The 'zoomed in' patches show the segmentation boost afforded with the *booster* module.

subject and the corresponding manually corrected segmentation for reference **(OML)**. We performed a one-way repeated measure ANOVA using Jamovi[19] to examine the effect of augmentation setting (including OML) on prediction performance across folds. The sphericity assumption was tested using Mauchly's test and met ($p = 0.166$). There was a significant main effect of augmentation setting on segmentation accuracy, $F(5, 65) = 15.8$, $p < 0.001$. This means that segmentation accuracy varied depending on the augmentation strategy. Post-hoc test (using the Bonferroni correction to adjust p; Supplementary Material) indicated that predictions from all settings with zoom were significantly more accurate than the ones generated without zoom or the OMELETTE-based labels ($p < 0.05$). No statistically significant difference was found for the remaining comparisons. Thus, we performed an equivalence test with the smallest effect size of interest (SESOI) of 0.05.[20] We tested the groups previously found not to be statistically different and found that the observed differences between these groups are statistically equivalent (SESOI < 0.05; supplementary material; Figure 6), i.e., we found no further improvement

in Dice scores when adding one additional data augmentation step (rotation or blurring), when increasing the number of copied patches and applying rotation and blurring, and when changing the data augmentation strategy from blurring to flipping. In addition, the Dice scores between the model trained without zoom and the OMELETTE 2 labels were also found to be equivalent.

Although the different augmentation settings provided similar results, we found that occasionally (between 5% to 10% of cases), artefacts were present in the segmentation. In other words, when training a model from scratch, we found that the quality of model performance varied. Using a smaller batch size (setting **(2)** and **(3)**), we observed false positive voxels as shown in Figure 7. Further, the augmentation strategy including blurring (setting **(4)**) resulted in segmentation occasionally degraded by noise. Indeed, when computing the percentage of voxels that differed between predicted segmentations from models trained using different initial states (random seeds) but the same training data, we observed that segmentation setting **(4)** led to the highest inconsistency in predicted segmentations (Figure
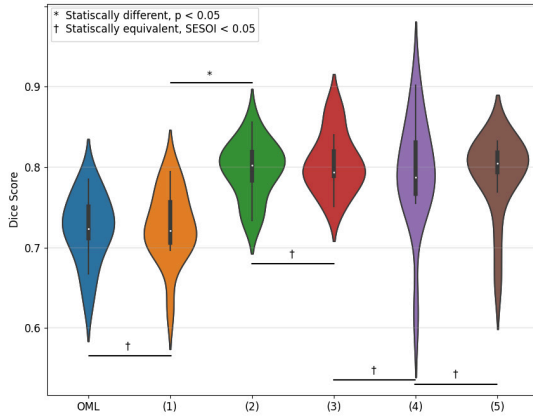
Figure 6. The cross-validation results where each value corresponds to the Dice score between the resultant segmentation (or the OMELETTE 2 labels used for training) and the manually corrected labels.

Shown are: **(OML)** OMELETTE 2 segmentation; **(1)** model trained without data augmentation or transformation; **(2)** model trained with zoom only; **(3)** model trained with zoom and one data transformation that could be either rotation or blurring; **(4)** model trained with zoom and six copied patches on which rotation or blurring was applied; **(5)** model trained with zoom and six copied patches on which rotation or flipping is applied.
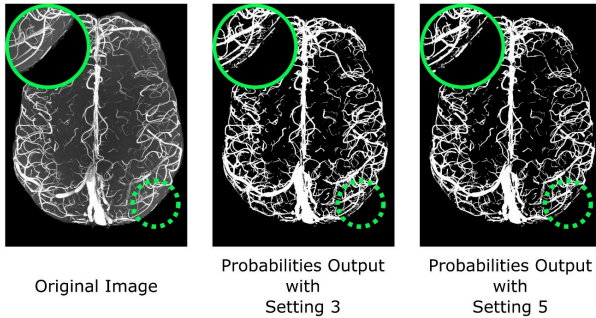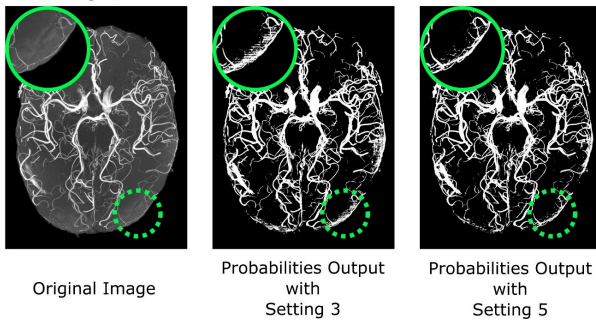
Figure 7. Ablation study of data augmentation technique.

Maximum intensity projection of input images (left), probabilities output of the model implemented with augmentation setting 3 (middle) and our proposed setting (right) for 160 μm (panel **a**) and 150 μm data (panel **b**) are shown. The 'zoomed in' patches show fewer artifacts in our proposed augmentation technique.

8). Accordingly, we recommend setting **(5)**, i.e., zoom, rotation, flipping and larger batch size, as the preferred augmentation strategy.

4. DISCUSSION AND CONCLUSION

We have introduced *VesselBoost*, a software toolbox that allows the segmentation of small vessels in high-resolution magnetic resonance angiography data. We found that our pre-trained models (available in *predict*) perform well on unseen data, that our test-time adaptation (*TTA*) module provides a fully automatic workflow to improve vessel segmentation, and that *booster* allows for the training of new segmentation models on single images, which generate segmentations with more detail than the training data.

We adapted the idea of training with imperfect labels[1-3] to vessel segmentation. This might considerably reduce the need for manually segmented training data, potentially restricting this laborious process to evaluating the final segmentation performance. Moreover, we used a comparatively simple model architecture, which showed good generalizability when tested on higher-resolution time-of-flight data.
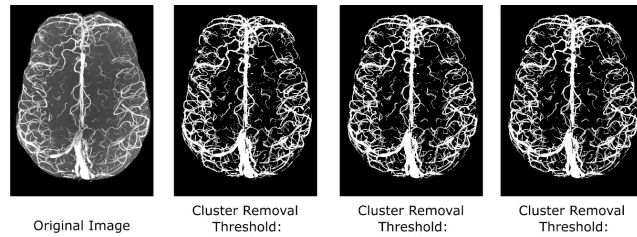
We showed that our data augmentation is crucial to increase segmentation sensitivity and model training stability. In detail, we found that zooming in and out of patches significantly increased prediction accuracy, whereas image rotation, flipping, and a larger batch size increased the stability of the model training process. Interestingly, the blurring of the training image had a detrimental effect on the model stability. Accordingly, we have chosen the recommended augmentation strategy to minimize the occurrence of segmentation artefacts (Figure 7). In cases where they do occur, they are best removed by repeating the training procedure. Alternatively, post-processing strategies, such as increasing the cluster-removal threshold (Figure 9), can be applied. Note that the released models have been thoroughly tested on all data sets, and no segmentation artefacts were observed. Noteworthy, we always trained and predicted on images with skulls but applied a brain mask afterwards to remove unwanted false positive segmentations stemming from the skin or fat around the skull. Thus, to obtain the best performance of *VesselBoost*, we recommend predicting on original, unmasked images first and then applying a brain mask. Note that *VesselBoost* was designed to be as sensitive as possible to small vessels, as we intended to increase small vessels detection rate, and, as a side effect, *VesselBoost* may capture the dura mater as a vessel-like structure, seen as vertical lines in the leftmost image in Figure 3 of panel b. Thus, if *VesselBoost* were to be used in clinical applications, it is paramount to implement additional checks to ensure segmentation accuracy.

Based on these insights, we developed an augmentation strategy and training pipeline (*booster*), which can leverage imperfect segmentation, for example, obtained through simple thresholding, and automatically boost these to capture smaller vessels within the same image. This strategy illustrates a way for leveraging deep learning on single example data sets. Importantly, *VesselBoost* can also be combined with various segmentation tools. The initial segmentation used for *TTA* and *booster* can be generated using pattern recognition methods, such as the 3D elevation maps used to detect intensity ridges,[21,22] or a surface cur-

a. 150 μm

b. 160 μm

(1) (2) (3) (4) (5)
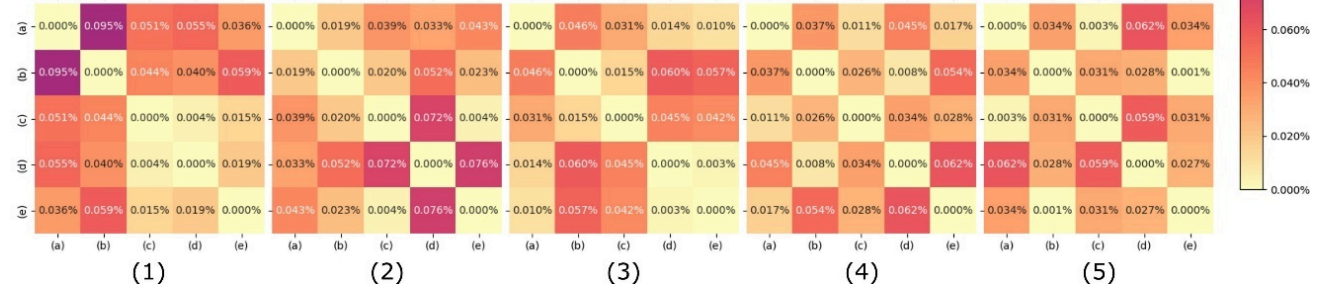
Figure 8. Predicted segmentation consistency across models trained with different random seeds.

We show the percentage of voxels that differed between predicted segmentations (sum of absolute difference divided by the total number of voxels) from models trained using different initial states (labelled as (a), (b), (c), (d), (e)) for each augmentation setting (1-5).

Original Image   Cluster Removal Threshold: 10   Cluster Removal Threshold: 20   Cluster Removal Threshold: 30

Figure 9. Prediction results for the 160 μm data with different cluster removal threshold in post-processing.

vature method.[23] Alternatively, it can utilize alternative model architectures for segmentation model training, including the well-known U-Net and its modifications, such as UNet++,[24,25] nnU-Net,[26] and VM-UNet,[27] as well as recent techniques focused on preserving the topological[28,29] or geometric features[30] of vascular structures.

VesselBoost is a toolbox with significant potential for advancing research in vascular morphology and could also be valuable in identifying new biomarkers for cerebrovascular diseases.[31-34] *VesselBoost* enables precise segmentation of vascular structures from angiograms, facilitating a deeper understanding of the differences in vascular morphology across human brain regions and different individuals. It also has the capabilities to extend its segmentation strategy to new image contrast such as T_2^*-weighted imaging to depict venous vasculature. This can lead to new insights into the geometrical and topological properties of brain vasculature.[35]

CODE AVAILABILITY

VesselBoost is freely available at https://osf.io/abk4p/, https://github.com/KMarshallX/VesselBoost and via Neurodesk,[36] a reproducible data analysis environment (https://neurodesk.org/), as Docker and Singularity containers.

CONFLICTS OF INTEREST

The authors declare no competing interests.

REFERENCES

1. Chatterjee S, Prabhu K, Pattadkal M, et al. DS6, Deformation-Aware Semi-Supervised Learning: Application to Small Vessel Segmentation with Noisy Training Data. *Journal of Imaging.* 2022;8(10):259. doi:10.3390/jimaging8100259

2. Fan S, Bian Y, Chen H, Kang Y, Yang Q, Tan T. Unsupervised Cerebrovascular Segmentation of TOF-MRA Images Based on Deep Neural Network and Hidden Markov Random Field Model. *Frontiers in Neuroinformatics*. 2020;13. doi:10.3389/fninf.2019.00077

3. Lucena O, Souza R, Rittner L, Frayne R, Lotufo R. Convolutional neural networks for skull-stripping in brain MR imaging using silver standard masks. *Artificial Intelligence in Medicine*. 2019;98:48-58. doi:10.1016/j.artmed.2019.06.008

4. Chatterjee S, Mattern H, Dubost F, Schreiber S, Nu¨rnberger A, Speck O. SMILE- UHURA Challenge. Published online 2023.

5. Hanke M, Baumgartner FJ, Ibe P, et al. A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific Data*. 2014;1(1):140003. doi:10.1038/sdata.2014.3

6. Mattern H. Openly available sMall vEsseL sEgmenTaTion pipelinE (OMELETTE). In: *29th Annual Meeting of International Society of Magnetic Resonance in Medicine (ISMRM)*. ; 2021:3745.

7. Mattern H, Sciarra A, Godenschweger F, et al. Prospective motion correction enables highest resolution time-of-flight angiography at 7T: Prospectively Motion-Corrected TOF Angiography at 7T. *Magnetic Resonance in Medicine*. 2018;80(1):248-258. doi:10.1002/mrm.27033

8. Bollmann S, Mattern H, Bernier M, et al. Imaging of the pial arterial vasculature of the human brain in vivo using high-resolution 7T time-of-flight angiography. *eLife*. 2022;11:e71186. doi:10.7554/eLife.71186

9. C¸ i¸cek Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Vol 9901. Springer International Publishing; 2016:424-432. doi:10.1007/978-3-319-46723-8_49

10. Ali S, Khurram R, Rehman KU, et al. An Improved 3D U-Net-based Deep Learning System for Brain Tumor Segmentation Using Multi-Modal MRI. *Multimedia Tools and Applications*. Published online 2024. doi:10.1007/s11042-024-19406-2

11. Aboussaleh I, Riffi J, Mahraz AM, Tairi H. Inception-UDet: An Improved U-Net Archi- tecture for Brain Tumor Segmentation. *Annals of Data Science*. 2024;11(3):831-853. doi:10.1007/s40745-023-00480-6

12. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. *NIPS 2017 Autodiff Workshop*. Published online 2017.

13. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *arXiv*. Published online 2017. doi:10.1007/978-3-319-67389-9_44

14. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908

15. Manj´on JV, Coup´e P, Mart´ı-Bonmat´ı L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels: Spatially Adaptive Nonlocal Denoising. *Journal of Magnetic Resonance Imaging*. 2010;31(1):192-203. doi:10.1002/jmri.22003

16. Silversmith W. cc3d: Connected components on multilabel 3D & 2D images. Published online 2021.

17. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980*. Published online 2014.

18. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409

19. jamovi. The jamovi project. Published online 2023.

20. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: A tutorial. *Advances in methods and practices in psychological science*. 2018;1(2):259-269. doi:10.1177/2515245918770963

21. Aylward S, Bullitt E, Pizer S, Eberly D. Intensity ridge and widths for tubular object segmentation and description. In: *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*. IEEE; 1996:131-138. doi:10.1109/MMBIA.1996.534065

22. Aylward SR, Bullitt E. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE transactions on medical imaging*. 2002;21(2):61-75. doi:10.1109/42.993126

23. Prinet V, Monga O, Ge C, Xie S, Ma S. Thin network extraction in 3D images: application to medical angiograms. In: *Proceedings of 13th International Conference on Pattern Recognition*. IEEE; 1996:386-390. doi:10.1109/ICPR.1996.546975

24. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*. Springer; 2018:3-11.

25. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: Redesigning skip connec- tions to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*. 2019;39(6):1856-1867. doi:10.1109/TMI.2019.2959609

26. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

27. Ruan J, Xiang S. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv preprint arXiv:240202491*. Published online 2024.

28. Qiu Y, Li Z, Wang Y, et al. Corsegrec: a topology-preserving scheme for extracting fully-connected coronary arteries from ct angiography. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2023:670-680. doi:10.1007/978-3-031-43898-1_64

29. Zhang X, Sun K, Wu D, et al. An Anatomy-and-Topology-Preserving Framework for Coronary Artery Segmentation. *IEEE Transactions on Medical Imaging*. Published online 2023.

30. Deng Z, Xu S, Zhang J, et al. Shape-Aware 3D Small Vessel Segmentation with Local Contrast Guided Attention. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2023:354-363. doi:10.1007/978-3-031-43901-8_34

31. Livne M, Rieger J, Aydin OU, et al. A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease. *Front Neurosci*. Published online 2019:13. doi:10.3389/fnins.2019.00097

32. Dai X, Huang L, Qian Y, et al. Deep learning for automated cerebral aneurysm detection on computed tomography images. *International Journal of Computer Assisted Radiology and Surgery*. 2020;15(4):715-723. doi:10.1007/s11548-020-02121-2

33. Paul JF, Rohnean A, Giroussens H, Pressat-Laffouilhere T, Wong T. Evaluation of a deep learning model on coronary CT angiography for automatic stenosis detection. *Diagnostic and Interventional Imaging*. 2022;103(6):316-323. doi:10.1016/j.diii.2022.01.004

34. Jiao Y, Zhang JZ, Zhao Q, et al. Machine Learning-Enabled Determination of Diffuseness of Brain Arteriovenous Malformations from Magnetic Resonance Angiography. *Translational Stroke Research*. 2022;13(6):939-948. doi:10.1007/s12975-021-00933-1

35. Hirsch S, Reichold J, Schneider M, Sz´ekely G, Weber B. Topology and Hemodynamics of the Cortical Cerebrovascular System. *Journal of Cerebral Blood Flow & Metabolism*. 2012;32(6):952-967. doi:10.1038/jcbfm.2012.39

36. Renton AI, Dao TT, Johnstone T, et al. Neurodesk: an accessible, flexible and portable data analysis environment for reproducible neuroimaging. *Nature Methods*. Published online 2024. doi:10.1038/s41592-023-02145-x

SUPPLEMENTARY MATERIALS

Supplementary Material

Download: https://apertureneuro.org/article/123217-vesselboost-a-python-toolbox-for-small-blood-vessel-segmentation-in-human-magnetic-resonance-angiography-data/attachment/244361.docx