

vesselFM: A Foundation Model for Universal 3D Blood Vessel Segmentation

Supplementary Material

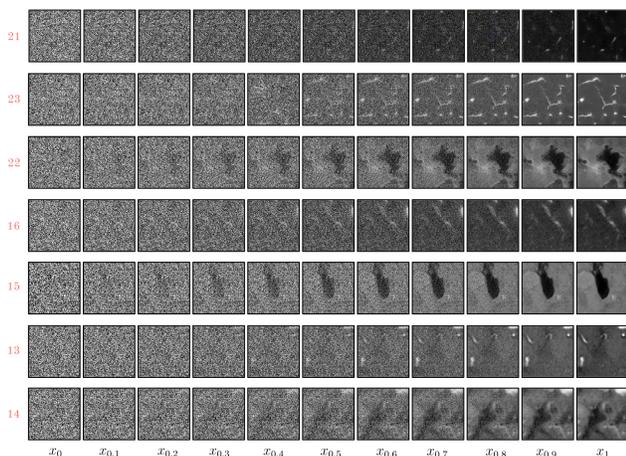


Figure 8. More trajectories of flow matching-based sampling. For improved visibility, we plot 2D slices. Given that the total number of time steps N is set to 100 in our experiments, Δt of 0.1 corresponds to 10 steps. Classes are indicated in red.

A. Pre-Processing of Datasets in $\mathcal{D}_{\text{real}}$

General pre-processing steps. We aim to ensure that all 23 datasets used to curate $\mathcal{D}_{\text{real}}$ (see Table 1) adhere to general vascular imaging characteristics and comply with our label quality standards. To this end, we apply carefully selected pre-processing steps (see Table 1, last column). First, we resample the MSD8 [1], BvEM [18], TubeTK [4], tUbeNet [10], TopCoW [21], and DeepVesselNet [16] datasets, establishing appropriate blood vessel scales (e.g., a single vessel should not occupy 90% of the patch) and voxel sizes. In the case of the BvEM dataset, only the labels are resampled, as the annotations were likely made in a downsampled version of the volume. As resampling may introduce label artifacts, we subsequently smooth affected labels (TubeTK and tUbeNet) using Gaussian smoothing followed by thresholding. Since the 3D-IRCADb-01 [15] dataset contains labels of multiple structures, we solely keep venous system, artery, and portal vein labels, converting them to binary labels. The original labels of the HR-Kidney [13] dataset are of relatively poor quality. However, enabled by the high signal-to-noise ratio of the volume representing the HR-Kidney dataset, we improve label quality by applying Algorithm 1. To further enhance the visibility of blood vessels, the intensities of the VesSAP [17] and LS [2] datasets are clipped at the 0% and 98% percentiles, while the intensities of the MSD8 dataset are clipped at the 20% and 98% percentiles. We crop the MSD8 and 3D-IRCADb-01 datasets to retain only fore-

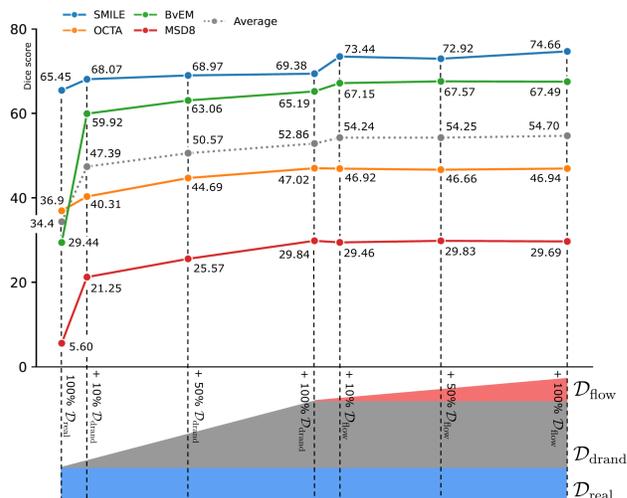


Figure 9. Experiment on the effect of our three proposed data sources on vesselFM’s performance. We gradually augment $\mathcal{D}_{\text{real}}$ with 10%, 50%, and 100% of $\mathcal{D}_{\text{drandid}}$, followed by adding 10%, 50%, and 100% of $\mathcal{D}_{\text{flow}}$ (see bottom part). We report *zero-shot* Dice scores on the four evaluation datasets. We generally find that augmenting $\mathcal{D}_{\text{real}}$ with $\mathcal{D}_{\text{drandid}}$ and $\mathcal{D}_{\text{flow}}$ results in increased segmentation performance (see average).

ground structures, as the images otherwise would primarily consist of non-annotated anatomical structures. Additionally, we crop the borders from the BvEM volume, given that they predominantly contain artifacts.

Algorithm 1 HR-Kidney label improvement.

Input: Image, Intensity Delta = 0.1, Threshold = 0.9, Filter Size = 11

| | |
|---|--|
| Median \leftarrow MedianFilter(Image, Filter Size) | \triangleright Apply filter to image. |
| Mask \leftarrow (Image - Median) > Int. Delta | \triangleright Include high local int. variations. |
| Mask \leftarrow Mask \vee (Image > Threshold) | \triangleright Include high int. values. |
| Mask \leftarrow Mask \bullet $\mathbf{1}_{3 \times 3 \times 3}$ | \triangleright Close small gaps. |
| Mask \leftarrow RemoveSmallObjects(Mask) | \triangleright Remove small connected components. |

\bullet denotes morphological closing.

Evaluation datasets. From each evaluation dataset (see Table 1, upper section), we extract three patches of size 128^3 for fine-tuning models in the *one-* and *few-shot* settings and use the remaining data for testing and validation. **OCTA** [8, 20]: We allocate three of the six samples provided in the OCTA dataset for *one-* and *few-shot* fine-tuning and reserve the remaining three for model evaluation: two samples for testing and one sample for validation. The three samples used for fine-tuning are center-cropped to adhere to our target shape of 128^3 . **SMILE-UHURA** [5]: The four-ten samples in the SMILE-UHURA dataset are divided

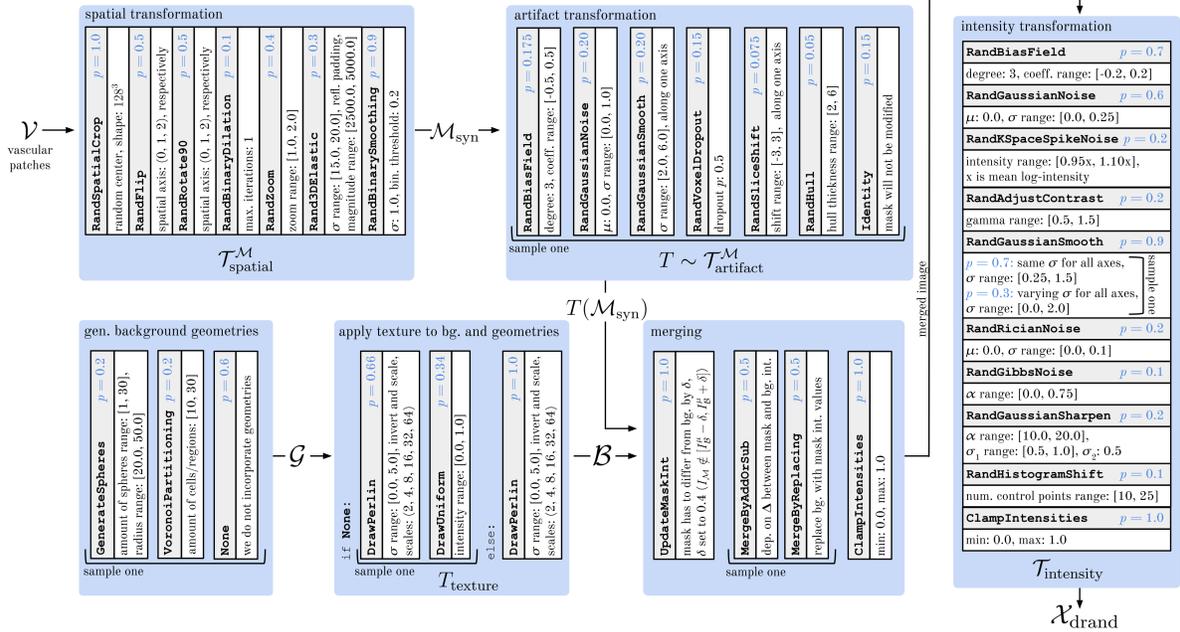


Figure 10. Parametrization of our domain randomized generative pipeline. All parameters were carefully tuned to ensure sufficient diversity while preserving key characteristics relevant to the general domain of vascular images. If not indicated otherwise, all the above transformations are applied consecutively, starting from the left-hand side. Probabilities associated with specific transformations are indicated in blue. Following common practices in medical image analysis, we utilize, whenever possible, transformations from the MONAI framework. Note that we exactly follow the notation from Fig. 4.

into one for validation, ten for testing, and three for extracting patches for fine-tuning. For the extraction of the fine-tuning patches, we pay special attention to extract patches highly representative of the characteristics of MRA scans contained in the SMILE-UHURA dataset (e.g., they contain vasculature, brain tissue, skull, and gyri/sulci). *MSD8* [1]: From the MSD8 dataset, we utilize one sample for validation, 296 for testing, and three to extract patches for fine-tuning. The patches for fine-tuning are chosen from representative regions and padded, if necessary, minimally in the z -dimension using reflective padding to conform to the target shape of 128^3 . *BvEM* [18]: The BvEM dataset contains solely a single volume of shape $3571 \times 5145 \times 2495$. We choose the first 130 slices to extract three 128^3 patches for fine-tuning and one for validation. Then, we leave a buffer of 120 slices to minimize information leakage between the patches used for fine-tuning and testing. Lastly, we extract three bigger test volumes of shape 500^3 from the remaining volume, limiting the overlap of these volumes with the fine-tuning patches in the x - and y -position as much as possible.

B. Parameters of Domain Randomization

An overview of the parametrization of transformations and operations in our proposed domain randomized generative pipeline, used to generate $D_{\text{drاند}}$, is shown in Fig. 10.

C. More Details on Experimental Setup

Parametrization of segmentation model. As already stated in the main manuscript, we opt for MONAI’s re-implementation of [12]’s UNet architecture, called DynUNet, to present our segmentation model. To be specific, we set `strides` to `[[1, 1, 1], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2]]`, `kernel_size` to `[[3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3]]`, `upsample_kernel_size` to `[[2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2]]`, `filters` to `[32, 64, 128, 256, 320, 320]`, and activate residual connection-based convolution blocks (`res_block`). We found that this parametrization performed best in our experiments.

Training of vesselFM. We employ a combination of Dice and cross-entropy loss functions, weighted by 0.9 and 0.1. VesselFM is trained on a single V100 GPU (32 GB) with a batch size of 8 until convergence. The learning rate is set to 10^{-4} . We utilize linear warm-up and a learning rate decay to 10^{-5} . During training, we sample classes near uniformly from D_{real} and also from D_{flow} . We solely apply data augmentation to samples from D_{real} . To optimize training efficiency, we perform data augmentation offline. Specifically, we apply, after extracting 128^3 patches, random flipping and rotation (angle in $[0^\circ, 10^\circ]$) along all axes followed by ran-

Table 6. More detailed ablation of supplementing $\mathcal{D}_{\text{real}}$ with $\mathcal{D}_{\text{drand}}$ and $\mathcal{D}_{\text{flow}}$, covering all four evaluation datasets and all tasks. We generally find that the combination of all of our three proposed data sources yields the best segmentation performance.

| Task | Model | OCTA [8, 20] | | BvEM [18] | | SMILE-UHURA [5] | | MSD8 [1] | |
|-----------|--|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|
| | | Dice \uparrow | clDice \uparrow |
| zero-shot | $\mathcal{D}_{\text{real}}$ | 36.94 | 57.23 | 29.44 | 52.71 | 65.45 | 63.53 | 5.60 | 8.60 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}}$ | 47.02 | 61.05 | 65.19 | 65.13 | 69.38 | 72.10 | 29.84 | 37.47 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}} + \mathcal{D}_{\text{flow}}$ | 46.94 | 67.07 | 67.49 | 62.04 | 74.66 | 75.27 | 29.69 | 36.14 |
| one-shot | $\mathcal{D}_{\text{real}}$ | 69.32 | 77.68 | 72.01 | 85.22 | 72.20 | 74.87 | 27.14 | 40.51 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}}$ | 70.63 | 81.11 | 75.77 | 78.48 | 71.77 | 71.90 | 35.35 | 49.39 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}} + \mathcal{D}_{\text{flow}}$ | 72.10 | 83.73 | 78.27 | 79.91 | 76.43 | 78.36 | 36.88 | 48.65 |
| few-shot | $\mathcal{D}_{\text{real}}$ | 73.01 | 80.14 | 67.18 | 81.41 | 77.63 | 77.24 | 38.65 | 48.71 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}}$ | 74.44 | 82.64 | 73.43 | 84.75 | 77.37 | 78.28 | 42.31 | 54.44 |
| | $\mathcal{D}_{\text{real}} + \mathcal{D}_{\text{drand}} + \mathcal{D}_{\text{flow}}$ | 75.70 | 84.03 | 78.11 | 84.54 | 78.77 | 79.37 | 45.04 | 57.25 |

dom elastic deformation (σ in [10, 20] and magnitude in [100, 500]) and random zooming (factor in [0.9, 1.3]).

One- and few-shot fine-tuning. In the *one-* and *few-*shot setting, we fine-tune vesselFM using a similar setup with a learning rate of 10^{-5} . We train vesselFM until convergence, selecting the checkpoint with the best Dice score on the respective validation volume. We apply lightweight data augmentation on the fly: random zooming (factor in [1, 1.3]), random shearing (shearing factors in [0, 0.4]), random flipping, random Gaussian smoothing (σ in [0, 0.5]), random Gaussian noise (μ of 0.3, σ in [0, 0.01]), and random histogram shifting (number of points in [5, 10]).

Following our general procedure, we fine-tune all baselines until convergence and select the checkpoint with the best Dice score on the validation volume for testing. *tUbeNet* [10]: We fine-tune tUbeNet using our training scheme described above, solely adapting the patch size (tUbeNet operates on patches of shape 64^3) and employing its linear learning rate decay. We further apply the same lightweight data augmentations used for fine-tuning vesselFM. *VISTA3D* [9]: We fine-tune VISTA3D with the script provided by the authors. As VISTA3D predicts 127 classes, we default to the only class representing blood vessels, the hepatic vessel class. Given that VISTA3D is designed specifically for CT images, we replace their interval-based intensity scaling scheme with a percentile-based scaling scheme and omit their resampling transformation. Further augmentations are left unchanged. During inference, we use the default "auto + point" configuration, which has been shown to yield the best results. *SAM-Med3D* [19]: We adopt SAM-Med3D's training and inference pipeline without major changes. We utilize their default setting, providing one query point during training and five during inference. Since their data augmentation pipeline closely resembles vesselFM's, we retain SAM-Med3D's without alterations. *MedSAM-2* [23]: As MedSAM-2 is trained on images of size 1024, we resample patches used for fine-tuning. Other than that, we keep their original setup, which fine-tunes the mask decoder of the SAM 2 model and the

memory layer, unchanged. We use the default configuration of one query point in every second slice for both training and inference.

D. VesselFM From Scratch

'VesselFM (from scratch)' (see Table 2) demonstrates a relatively strong performance compared to other baselines and even outperforms them on the OCTA and BvEM datasets in some metrics. We attribute this to their unique imaging artifacts and intensity patterns, which, unlike those in CT and MRA, have not been observed by any baseline during pre-training. Note that the human retinal OCTA sample present in tUbeNet's training dataset exhibits poor label quality and differs significantly from the characteristics [20] of the murine cerebral OCTA images used in our experiments. Further, architectural biases (SAM-Med3D and MedSAM-2 rely on large Transformers) and mismatches in already learned representations (VISTA3D is exclusively trained on CT images; tUbeNet is exclusively trained on images with stark contrast to background tissues) may impede *few-shot* fine-tunability of baselines. This, together with 'vesselFM (from scratch)' benefiting from our well-evaluated UNet architecture, explains its relatively strong performance.

E. More Detailed Ablation of Data Sources

A more detailed ablation study of the effect of supplementing $\mathcal{D}_{\text{real}}$ with $\mathcal{D}_{\text{drand}}$ and $\mathcal{D}_{\text{flow}}$, covering all evaluation datasets and tasks, is shown in Table 6. We find that the combination of all of our three proposed data sources generally yields the best segmentation performance.

To further substantiate this hypothesis, we investigate how the performance of vesselFM scales with the amount of available training data for *zero-shot* segmentation. To this end, we progressively augment $\mathcal{D}_{\text{real}}$ with 10%, 50%, and finally 100% of the data from our two synthetic data sources, $\mathcal{D}_{\text{drand}}$ and $\mathcal{D}_{\text{flow}}$. We scale weights assigned to data sources accordingly. Similar to Table 6, we first augment $\mathcal{D}_{\text{real}}$ with $\mathcal{D}_{\text{drand}}$, followed by $\mathcal{D}_{\text{flow}}$. Our findings are demonstrated in

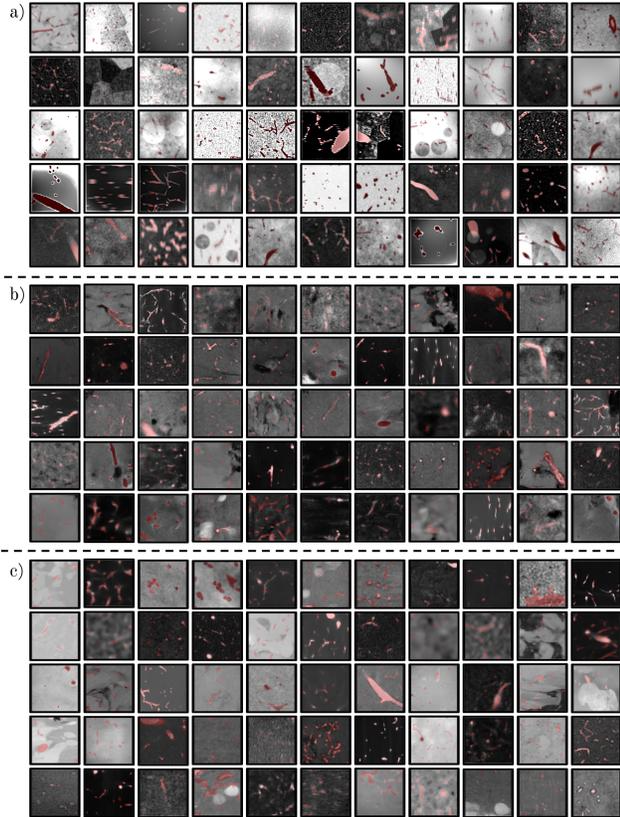


Figure 11. More slices of exemplary domain randomized images (a), images sampled from our flow matching-based generative model \mathcal{F} (b), and images sampled from Med-DDPM (c). Masks are shown in translucent red.

Fig. 9. We observe a significant performance increase introducing $\mathcal{D}_{\text{drand}}$, which flattens as it approaches 100%. On the SMILE-UHURA and BvEM datasets, performance additionally spikes after introducing $\mathcal{D}_{\text{flow}}$, while performance on OCTA and MSD8 stagnates. Averaged across all four evaluation datasets (see Fig. 9, average), we find that the additional diversity introduced by $\mathcal{D}_{\text{flow}}$ proves to be consistently beneficial for segmentation performance.

F. VesselFM’s Utility in (Pre-)Clinical Settings

In (pre-)clinical settings, high-quality annotations for emerging imaging technologies and novel, unstudied structures of interest are often not immediately available. Therefore, clinicians and researchers typically create voxel-level annotations from scratch through labor-intensive manual labeling to train supervised segmentation algorithms, a necessity for automated, large-scale, and accurate analysis. Voxel-level annotations, however, can be far more efficiently obtained via automated pre-segmentation and iterative label refinement. This process is often referred to as

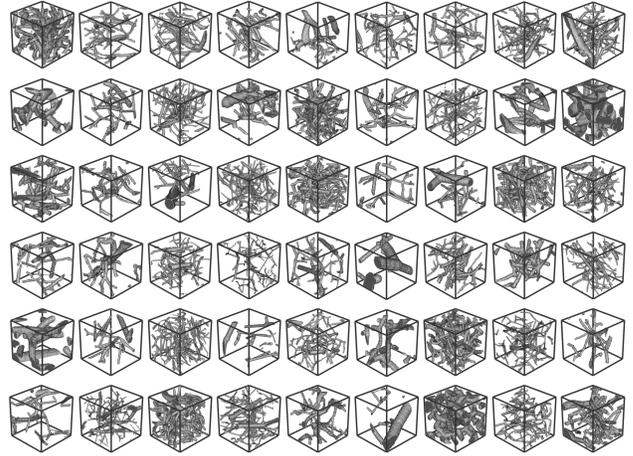


Figure 12. More exemplary synthetic masks \mathcal{M}_{syn} , generated by our proposed domain randomized generative pipeline (see Fig. 4a).

bootstrapping. With specialist models failing to bridge domain gaps, vesselFM’s exceptional *zero-shot* generalization and fine-tunability render it ideal for such applications.

G. More Samples From $\mathcal{D}_{\text{drand}}$ and $\mathcal{D}_{\text{flow}}$

We present additional samples from $\mathcal{D}_{\text{drand}}$ and $\mathcal{D}_{\text{flow}}$ in Fig. 11a and b, respectively. In conclusion, one can state that our domain randomized generative pipeline produces a wide variety of image-mask pairs with highly diverse foreground and background geometries and textures, while images sampled from our flow matching-based generative model \mathcal{F} exhibit intensity patterns closely mimicking those of real images in $\mathcal{D}_{\text{real}}$. Please note that Fig. 11c depicts images sampled from the Med-DDPM baseline for comparison.

H. More Masks \mathcal{M}_{syn}

To further showcase the wide variety of synthetic masks \mathcal{M}_{syn} produced by our proposed domain randomized generative pipeline, we present a comprehensive selection in Fig. 12. Masks contained in \mathcal{M}_{syn} encompass a broad range of realistic vascular patterns, capturing variations in blood vessel scale, density, curvature, and tortuosity.

I. Additional Qualitative *Zero-Shot* Results

To emphasize the exceptional *zero-shot* generalization of vesselFM, we present additional qualitative results achieved on all four evaluation datasets (see Figs. 13 to 16). Our findings demonstrate that vesselFM segments blood vessels very accurately across all evaluation datasets. Interestingly, vesselFM also segments tubular-appearing structures beyond blood vessels (*e.g.*, axons (see Fig. 15) or parts of the colon (see Fig. 14)). This highlights vesselFM’s strong inductive bias towards tubular shapes.

Table 7. Comparison to specialist models trained on individual blood vessel segmentation datasets.

| Task | Model | OCTA [8, 20] | | BvEM [18] | | SMILE-UHURA [5] | | MSD8 [1] | |
|------------------|--------------------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|
| | | Dice \uparrow | cIDice \uparrow |
| <i>zero-shot</i> | specialist model - TopCoW [21] | 39.14 | 24.03 | 55.09 | 52.42 | 38.71 | 34.35 | 4.18 | 3.45 |
| | specialist model - VesSAP [17] | 33.31 | 60.69 | 50.89 | 32.75 | 16.27 | 25.63 | 9.40 | 15.19 |
| | specialist model - CSD [3, 6] | 21.92 | 42.98 | 0.04 | 0.01 | 59.17 | 53.07 | 0.44 | - |
| | vesselFM (ours) | 46.94 | 67.07 | 67.49 | 62.04 | 74.66 | 75.27 | 29.69 | 36.14 |
| <i>one-shot</i> | specialist model - TopCoW [21] | 64.80 | 74.39 | 71.53 | 78.65 | 36.52 | 35.66 | 17.61 | - |
| | specialist model - VesSAP [17] | 69.26 | 76.67 | 72.31 | 75.94 | 27.40 | 30.50 | 28.94 | 35.40 |
| | specialist model - CSD [3, 6] | 64.62 | 77.56 | 68.64 | 78.96 | 75.96 | 78.20 | 16.09 | - |
| | vesselFM (ours) | 72.10 | 83.73 | 78.27 | 79.91 | 76.43 | 78.36 | 36.88 | 48.65 |
| <i>few-shot</i> | specialist model - TopCoW [21] | 70.28 | 76.79 | 65.89 | 77.76 | 52.97 | 47.54 | 26.25 | 37.16 |
| | specialist model - VesSAP [17] | 70.27 | 77.14 | 69.36 | 78.40 | 46.83 | 44.31 | 36.27 | 44.74 |
| | specialist model - CSD [3, 6] | 71.11 | 79.36 | 59.94 | 78.61 | 78.20 | 78.59 | 34.28 | 47.57 |
| | vesselFM (ours) | 75.70 | 84.03 | 78.11 | 84.54 | 78.77 | 79.37 | 45.04 | 57.25 |

By segmenting all tubular structures in the volume, vesselFM segments structures, which are, at least to some degree, not annotated in ground truth labels (*e.g.*, aorta or other components of the systemic arterial circulation in MSD8 (see Fig. 14)). We argue that this may artificially deflate vesselFM’s quantitative results reported for the *zero-shot* task in Table 2.

J. Additional Flow Matching Trajectories

Fig. 8 presents additional sampled flow matching trajectories, similar to Fig. 5. Specifically, we visualize the mapping from $x_0 \sim \mathcal{N}(0, I)$ to samples x_1 of exemplary classes indicated in red.

K. Statistical Analysis of Results

We conduct a statistical analysis of the quantitative results reported in the main manuscript using paired t-tests. To this end, we compare vesselFM’s base configuration to the respective runner-up. We find all results to be statistically significant ($p < 0.05$), except for the Dice score in the ablation on class conditioning (Table 4, 4th row; $p = 0.081$). The cIDice score in the ablation on class conditioning, however, remains significant ($p = 0.036$). Notably, values reported in Table 5 are statistically significant ($p = 0.019$ for Dice; $p = 2.39 \cdot 10^{-5}$ for cIDice).

L. Computational Resources

VesselFM comprises 31,418,977 parameters and requires minimal computational resources compared to other foundation models for 3D image segmentation. Processing a volume of shape 128^3 takes 335.7 ms on a T4 GPU, 95.8 ms on a V100 GPU, 30.6 ms on an A100 GPU, and 3.3 s on an AMD Epyc 7702 CPU, with a VRAM consumption of approximately 4.21 GB.

M. Comparison to Specialist Models

We further compare vesselFM to specialist models trained on individual blood vessel segmentation datasets, as shown in Table 7. As expected, vesselFM not only consistently outperforms state-of-the-art foundation models for medical image segmentation, but also specialist models, even when trained on exactly the same imaging modality (CSD and SMILE-UHURA both contain MRA images).

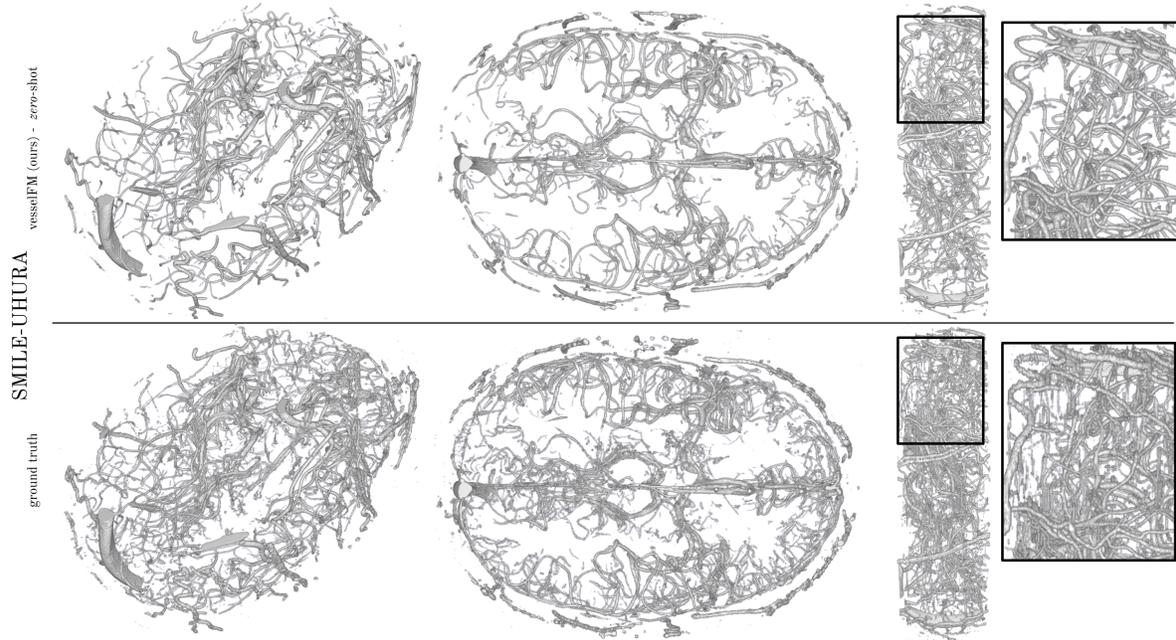


Figure 13. Qualitative results achieved on an exemplary test sample from the SMILE-UHURA dataset [5]. We compare vesselFM’s prediction in the *zero-shot* setting (top row) to the ground truth label contained in the SMILE-UHURA dataset (bottom row). VesselFM delivers remarkable results free of artifacts and accurately maintains the tubular appearance of blood vessels (see black box).

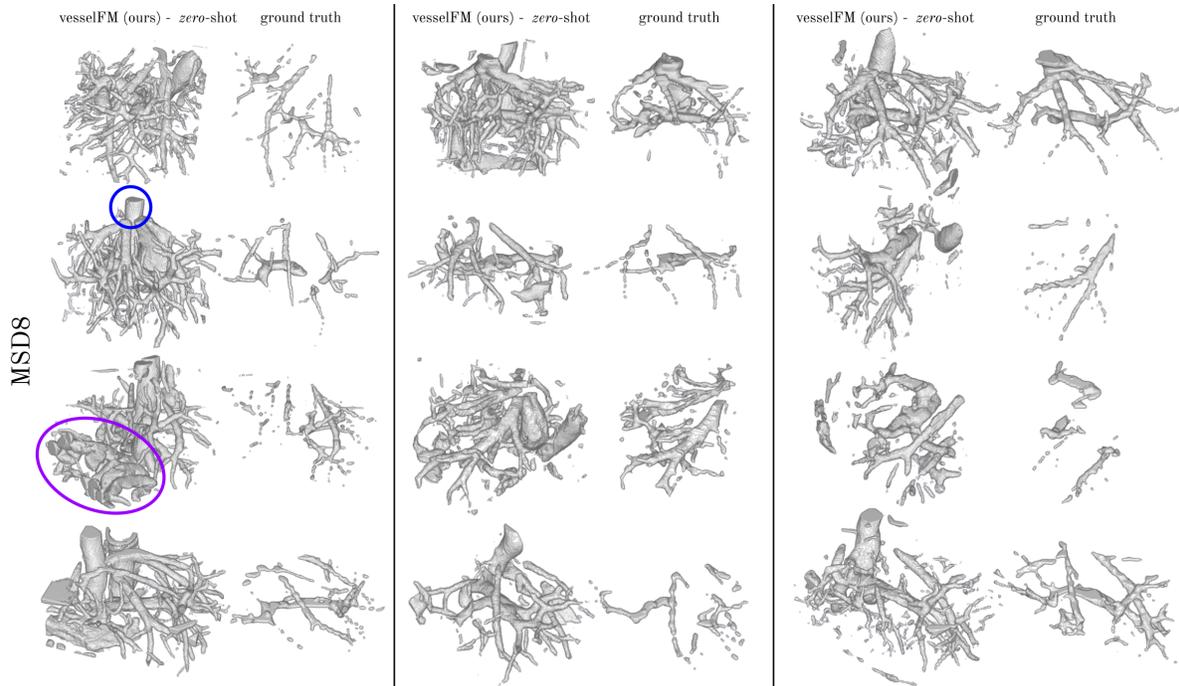


Figure 14. Qualitative results achieved on multiple test samples from the MSD8 dataset [1]. We compare vesselFM’s predictions in the *zero-shot* setting to ground truth labels for the task of hepatic vessel segmentation contained in the MSD8 dataset. VesselFM accurately segments all blood vessels (*e.g.*, aorta (marked in blue) and other major components of the systemic arterial circulation) and even other tubular structures (*e.g.*, the colon (marked in purple) and parts of the rib cage) present in CT scans.

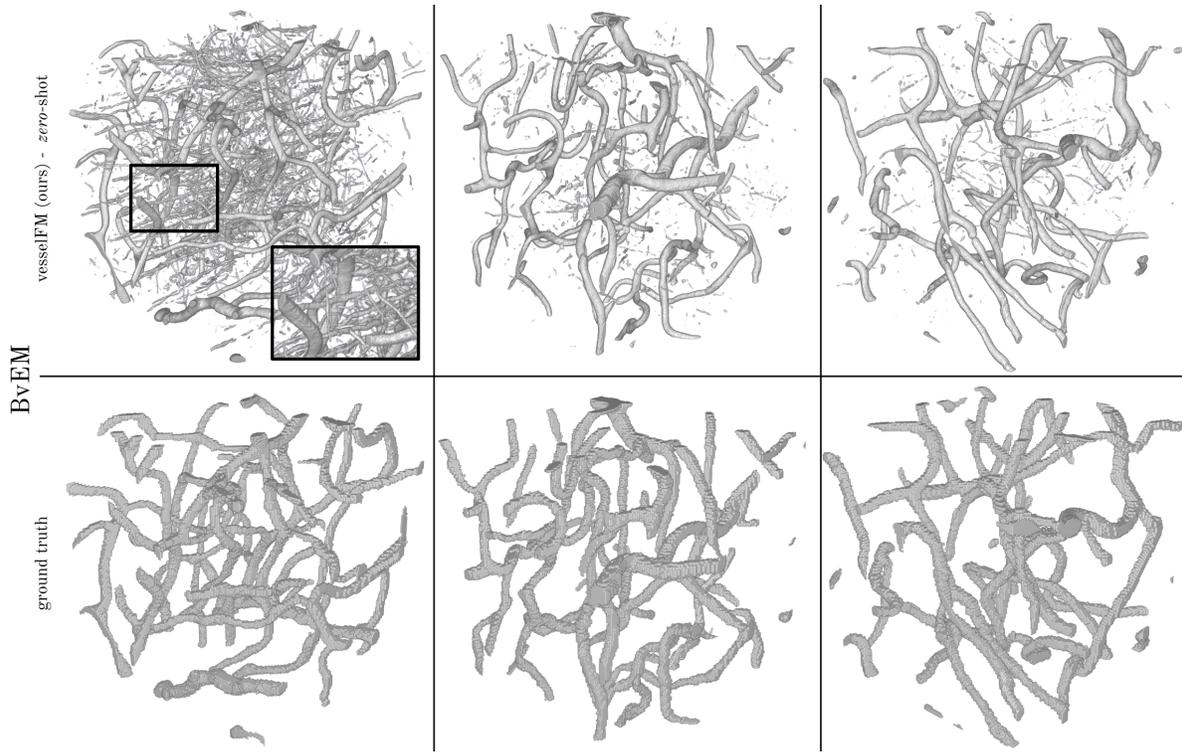


Figure 15. Qualitative results achieved on the three test volumes extracted from the BvEM dataset [18]. We compare vesselFM’s predictions in the *zero-shot* setting (top row) to ground truth labels contained in the BvEM dataset (bottom row). We find that vesselFM segments murine cortical vasculature contained in volume electron microscopy (vEM) images very precisely. Interestingly, vesselFM segments not only blood vessels but also tubular-appearing axons and even dendrites of pyramidal cells (see black box) visible in vEM images [7].

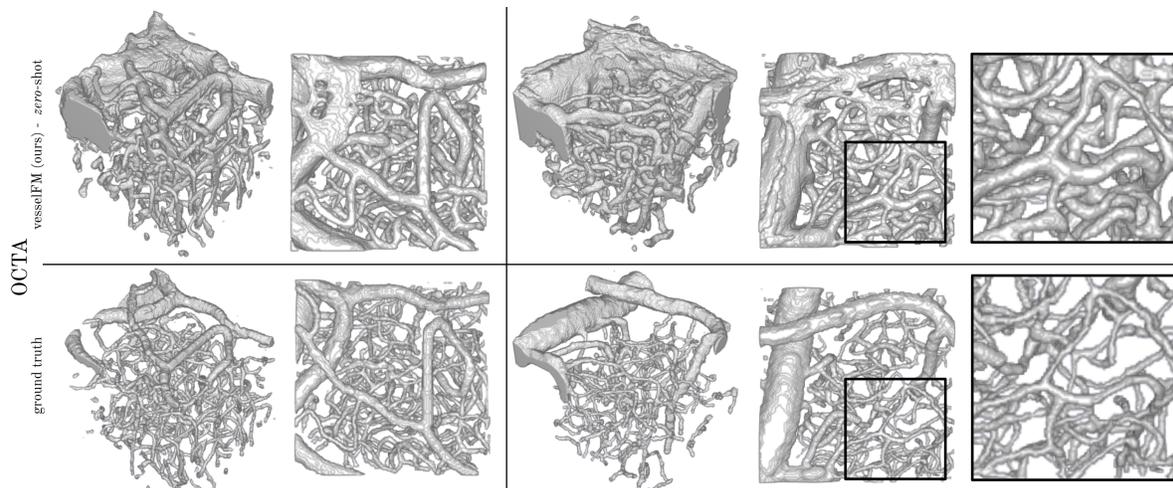


Figure 16. Qualitative results achieved on two test samples from the OCTA dataset [8, 20]. We compare vesselFM’s predictions in the *zero-shot* setting (top row) to ground truth labels contained in the OCTA dataset (bottom row). Although OCTA images are known for being plagued by dominant imaging artifacts [11, 14, 22], vesselFM still manages to segment densely connected vasculature (see black box).

References

- [1] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The Medical Segmentation Decathlon. *Nature communications*, 13(1):4128, 2022. 1, 2, 3, 5, 6
- [2] Nadine Felizitas Binder, Mohamad El Amki, Chaim Glück, William Middleham, Anna Maria Reuss, Adrien Bertolo, Patrick Thurner, Thomas Deffieux, Chryso Lambride, Robert Epp, et al. Leptomeningeal collaterals regulate reperfusion in ischemic stroke and rescue the brain from futile recanalization. *Neuron*, 112(9):1456–1472, 2024. 1
- [3] brain-development.org. IXI Dataset. <http://brain-development.org/ixi-dataset/>, n.d. 5
- [4] Elizabeth Bullitt, Donglin Zeng, Guido Gerig, Stephen Aylward, Sarang Joshi, J Keith Smith, Weili Lin, and Matthew G Ewend. Vessel Tortuosity and Brain Tumor Malignancy: A Blinded Study. *Academic radiology*, 12(10):1232–1240, 2005. 1
- [5] Soumick Chatterjee, Hendrik Mattern, Marc Dörner, Alessandro Sciarra, Florian Dubost, Hannes Schnurre, Rupali Khatun, Chun-Chih Yu, Tsung-Lin Hsieh, Yi-Shan Tsai, et al. SMILE-UHURA Challenge—Small Vessel Segmentation at Mesoscopic Scale from Ultra-High Resolution 7T Magnetic Resonance Angiograms. *arXiv preprint arXiv:2411.09593*, 2024. <https://doi.org/10.7303/syn47164761>. 1, 3, 5, 6
- [6] Ying Chen, Darui Jin, Bin Guo, and Xiangzhi Bai. Attention-Assisted Adversarial Model for Cerebrovascular Segmentation in 3D TOF-MRA Volumes. *IEEE Transactions on Medical Imaging*, 41(12):3520–3532, 2022. 5
- [7] MICrONS Consortium, J Alexander Bae, Mahaly Baptiste, Caitlyn A Bishop, Agnes L Bodor, Derrick Brittain, JoAnn Buchanan, Daniel J Bumbarger, Manuel A Castro, Brendan Celi, et al. Functional connectomics spanning multiple areas of mouse visual cortex. *bioRxiv*, pages 2021–07, 2021. 7
- [8] Lukas Glandorf, Bastian Wittmann, Jeanne Droux, Chaim Glück, Bruno Weber, Susanne Wegener, Mohamad El Amki, Rainer Leitgeb, Bjoern Menze, and Daniel Razansky. Bessel beam optical coherence microscopy enables multiscale assessment of cerebrovascular network morphology and function. *Light: Science & Applications*, 13(1):307, 2024. 1, 3, 5, 7
- [9] Yufan He, Pengfei Guo, Yucheng Tang, Andriy Myronenko, Vishwesh Nath, Ziyue Xu, Dong Yang, Can Zhao, Benjamin Simon, Mason Belue, et al. VISTA3D: Versatile Imaging Segmentation and Annotation model for 3D Computed Tomography. *arXiv preprint arXiv:2406.05285*, 2024. 3
- [10] Natalie Holroyd, Zhongwang Li, Claire Walsh, Emmeline Brown, Rebecca Shipley, and Simon Walker-Samuel. tUbe net: a generalisable deep learning tool for 3D vessel segmentation. *bioRxiv*, pages 2023–07, 2023. 1, 3
- [11] Tristan T Hormel, David Huang, and Yali Jia. Artifacts and artifact removal in optical coherence tomographic angiography. *Quantitative Imaging in Medicine and Surgery*, 11(3):1120, 2021. 7
- [12] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 2
- [13] Willy Kuo, Diego Rossinelli, Georg Schulz, Roland H Wenger, Simone Hieber, Bert Müller, and Vartan Kurtcuoglu. Terabyte-scale supervised 3D training and benchmarking dataset of the mouse kidney. *Scientific Data*, 10(1):510, 2023. 1
- [14] Yuntao Li and Jianbo Tang. Blood vessel tail artifacts suppression in optical coherence tomography angiography. *Neurophotonics*, 9(2):021906–021906, 2022. 7
- [15] Luc Soler, Alexandre Hostettler, Vincent Agnus, Arnaud Charnoz, J Fasquel, Johan Moreau, A Osswald, Mourad Bouhadjar, and Jacques Marescaux. 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. *IRCAD, Strasbourg, France, Tech. Rep*, 1(1), 2010. 1
- [16] Giles Tetteh, Velizar Efremov, Nils D Forkert, Matthias Schneider, Jan Kirschke, Bruno Weber, Claus Zimmer, Marie Piraud, and Björn H Menze. DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes. *Frontiers in Neuroscience*, 14:592352, 2020. 1
- [17] Mihail Ivilinov Todorov, Johannes Christian Paetzold, Oliver Schoppe, Giles Tetteh, Suprosanna Shit, Velizar Efremov, Katalin Todorov-Völgyi, Marco Düring, Martin Dichgans, Marie Piraud, et al. Machine learning analysis of whole mouse brain vasculature. *Nature methods*, 17(4):442–449, 2020. 1, 5
- [18] Jia Wan, Wanhua Li, Atmadeep Banerjee, Jason Ken Adhinarta, Evelina Sjostedt, Jingpeng Wu, Jeff Lichtman, Hanspeter Pfister, and Donglai Wei. TriSAM: Tri-Plane SAM for zero-shot cortical blood vessel segmentation in VEM images. *arXiv preprint arXiv:2401.13961*, 2024. 1, 2, 3, 5, 7
- [19] Haoyu Wang, Sizheng Guo, Jin Ye, Zhongying Deng, Junlong Cheng, Tianbin Li, Jianpin Chen, Yanzhou Su, Ziyang Huang, Yiqing Shen, Bin Fu, Shaoting Zhang, Junjun He, and Yu Qiao. SAM-Med3D: Towards General-purpose Segmentation Models for Volumetric Medical Images. *arXiv preprint arXiv:2310.15161*, 2024. 3
- [20] Bastian Wittmann, Lukas Glandorf, Johannes C Paetzold, Tamaz Amiranashvili, Thomas Wälchli, Daniel Razansky, and Bjoern Menze. Simulation-Based Segmentation of Blood Vessels in Cerebral 3D OCTA Images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 645–655. Springer, 2024. 1, 3, 5, 7
- [21] Kaiyuan Yang, Fabio Musio, Yihui Ma, Norman Juchler, Johannes C Paetzold, Rami Al-Maskari, Luciano Höher, Hongwei Bran Li, Ibrahim Ethem Hamamci, Anjany Sekuboyina, et al. Benchmarking the CoW with the TopCoW Challenge: Topology-Aware Anatomical Segmentation of the Circle of Willis for CTA and MRA. *arXiv preprint arXiv:2312.17670*, 2023. 1, 5
- [22] Jun Zhu, Marcel T Bernucci, Conrad W Merkle, and Vivek J Srinivasan. Visibility of microvessels in Optical Coherence

Tomography angiography depends on angular orientation.
Journal of biophotonics, 13(10):e202000090, 2020. [7](#)

- [23] Jiayuan Zhu, Yunli Qi, and Junde Wu. Medical SAM 2: Segment medical images as video via Segment Anything Model 2. *arXiv preprint arXiv:2408.00874*, 2024. [3](#)