tUbe net: a generalisable deep learning tool for 3D vessel segmentation

Natalie Holroyd[a], Zhongwang Li[a], Claire Walsh[b], Emmeline Brown[a], Rebecca Shipley[a,b], Simon Walker-Samuel[a,*]
[a]Centre for Computational Medicine, University College London, London, United Kingdom, WC1E 6BT
[b]Department of Mechanical Engineering, University College London, United Kingdom, WC1E 6BT

Abstract. Deep learning has become an invaluable tool for bioimage analysis but, while open-source cell annotation software such as cellpose are widely used, an equivalent tool for three-dimensional (3D) vascular annotation does not exist. With the vascular system being directly impacted by a broad range of diseases, there is significant medical interest in quantitative analysis for vascular imaging. However, existing deep learning approaches for this task are specialised to particular tissue types or imaging modalities. We present a new deep learning model for segmentation of vasculature that is generalisable across tissues, modalities, scales and pathologies. To create a generalisable model, a 3D convolutional neural network was trained using data from multiple modalities including optical imaging, computational tomography and photoacoustic imaging. Through this varied training set, the model was forced to learn common features of vessels cross-modality and scale. Following this, the general model was fine-tuned to different applications with a minimal amount of manually labelled ground truth data. It was found that the general model could be specialised to segment new datasets, with a high degree of accuracy, using as little as 0.3% of the volume of that dataset for fine-tuning. As such, this model enables users to produce accurate segmentations of 3D vascular networks without the need to label large amounts of training data.

Keywords: deep learning, vasculature, segmentation.

*Simon Walker-Samuel, simon.walkersamuel@ucl.ac.uk

1 Introduction

A broad range of diseases directly involve the vascular system, and as such the organisation and structure of blood vessels is of inherent medical interest. For example, blood vessel networks become highly disrupted following angiogenesis in cancer,[1] inflammation in response to infection induces changes in vascular structure,[2] and blockages causing loss of perfusion has serious consequences in cardiac infarction,[3] stroke[4] and diabetes.[5] Many bioimaging techniques have been developed to characterise both normal and diseased blood vessel networks, across a range of length scales, organs and pathologies. However, the ability to quantify and fully characterise the complex, hierarchical nature of blood vessels still presents a significant challenge, particularly when consideration of their three-dimensional structure is required.

Deep learning (DL) with convolutional neural networks has rapidly become the state of the art approach for semantic segmentation (labelling each pixel according to a predefined set of classes) in medical imaging due its accuracy and generalisability. However, it requires a large and varied array of training data, complete with paired ground truth labels. Whilst image data is generally widely available, acquiring manually-defined labels is a substantial and often challenging undertaking and introduces a significant bottleneck to implementation. Accordingly, most DL models to date have been trained using a limited set of data from a single pathology or organ, imaged with a single modality, at a single length scale. In particular, much of the field is focused on segmenting retinal vasculature from two-dimensional colour fundus images.[6–10] Models have also been created for segmenting brain vasculature in three dimensions.[11,12] Typically, each new application requires a new model, trained on new data.

In the field of cellular imaging, cellpose[13] and cellpose2[14] have become popular open-source tools that employ an alternative, a human-in-the-loop workflow: where a general model is trained most of the way and then fine-tuned on a small amount of the end-users' data in order to specialise the model to their needs. This approach minimises manual labelling and increases the accessibility of deep learning tools for users without the computational resources or sufficient labelled data to train a model from scratch. Similar approaches are also available for vessel segmentation in the form of proprietary software (for example, the Imaris 10.0 Filament Tracer and the Amira-Avizo XFibre extension) but no open-source tools exist for this purpose.

In this study, we aimed to leverage the ubiquitous tube-like structuring of blood vessels to create a model that can be used to segment three-dimensional image data containing blood vessels, acquired with multiple modalities, and from multiple organs. This is achieved by implementing a training strategy that utilises highly varied multi-modal data to develop a generalisable vessel-

finding model, which can be fine-tuned to new tasks with a minimal amount of labelled data.

1.1 Semantic Segmentation With U-Nets

Convolutional neural networks (CNN) consist of a sequence of convolutional layers (a deep network) with weights that can be trained to enable it to detect image features at multiple length scales. U-Nets are a class of CNN architecture consisting of encoding layers that feed into a symmetric set of decoding layers.[15] In the encoding arm, alternating convolutional and pooling layers downsample the data and convert the input data into a latent representation. Once encoded, the decoding arm converts the latent representation into a new image space, by sequential upsampling using deconvolutional (or transpose convolutional) layers. The downsampling and upsampling branches are symmetric, giving a 'U' shaped architecture. Skip connections enable direct communication between corresponding encoding and decoding layers, primarily to help overcome the vanishing gradient problem that affects very deep CNNs and has been found to improve the accuracy of segmentation of biological images.[16]

1.2 Bottlenecks In ML-based Segmentation

Collating sufficient data to accurately train a deep neural network is one of the major challenges in deep learning. In the field of medical imaging there is generally no shortage of raw data, but the corresponding ground truth labels are more challenging to obtain. Images must be manually segmented to create paired datasets, which is highly labour-intensive and potentially error-prone.[10] The training data must also comprise a representative sample of the range of the data the model will be expected to classify upon application, to enable sufficient generalisability.[17]

Multiple approaches have been developed to overcome these challenges. In addition to cre-

ating architectures that require less training data, such as the U-Net, researchers have attempted to make smaller datasets go further using data augmentation,[18,19] and have employed pre-training (or transfer learning) strategies.[20–22] Pre-training leverages the ability of CNNs to detect features which are characteristic of a broad range of image domains; as such, a CNN trained on one domain can be transferred to another domain using less data than would be required to train it from scratch. Similarly, synthetic data has been used to address training bottlenecks: models have been trained using either purely synthetic data,[23] or a combination of real and synthetic data,[24,25] for which the 'ground truth' labels are inherently known. However, the extent to which models trained on synthetic data may be meaningfully applied to physiological data remains an open question.[26]

1.3 Semantic Segmentation Of Blood Vessels

CNNs have been successfully used to label blood vessels in two-dimensional (2D) images and have been particularly widely applied in the segmentation of retinal vasculature from image data.[6–10] This is aided by the fact that several hand-labelled data sets are available to allow developers to benchmark new models and approaches (e.g. DRIVE,[27] Structured Analysis of the Retina (STARE)[28] and FIVES[29]).

Segmentation of three-dimensional (3D) vasculature, on the other hand, presents a greater challenge due in part to the much higher demands of manual segmentation, higher computational expense, and the lack of ready-labelled libraries against which to benchmark. Furthermore, much of the research in this area has been limited to brain vasculature; for example, Ref. 11 used a 3D CNN to segment the vascular network from dual-stained mouse brains, imaged using light-sheet microscopy, in their VesSAP pipeline. Likewise, DeepVesselNet[12] was trained with two sets of brain data: clinical MRA and rat brain micro-CT. There are few examples of applications outside

of the brain, for example Ref. 30, which demonstrated the effectiveness of an integrated U-net and graphical neural network on two clinical contract-enhanced liver CT datasets. To our knowledge, no work has been published in which DL is used to segment 3D microvasculature outside of the brain.

While these recent examples point to many potential applications for 3D CNN image analysis, both in research and in the clinic, there is clearly a need for models that can be applied to a wider range of tissues and data sources. In particular, there is a need for models that can segment vasculature at the micron scale in organs other than the brain and retina. Here, we explore the use of transfer learning to address bottlenecks in blood vessel segmentation, and aim to create a 'generalised' 3D vessel segmentation model that can be fine-tuned to multiple applications.

2 Methods

2.1 Model Architecture

A convolutional neural network was produced in tensorflow version ..... by adapting the U-Net architecture for 3D data. The model consists of six convolutional layers followed by five deconvolutional layers. The convolutional layers (purple, Fig. 1) consist of two 3D convolutions (kernel size $3 \times 3 \times 3$) followed by a Leaky ReLU activation function (alpha = 0.2) and group normalisation. A pooling function follows each layer, downsampling the data by a factor of two in each dimension - effectively halving the resolution. Deconvolutional layers (orange) increase the resolution of the data using a deconvolving function ('conv3DTranspose'), followed by two standard convolutions, again with a Leaky ReLU activation function. The output layer is a convolution with a single filter and a softmax activation function to give a pseudo-probabilistic output for each pixel belonging to each class (i.e. inside or outside a vessel). Concatenations are used to 'skip' layers by

transferring feature maps directly from the encoding branch to the symmetric layer of the decoding branch, which has been shown to improve convergence speed in deep networks.[16] A $25\,\%$ dropout is included after each layer to reduce the risk of overfitting to the training data.

2.2 Attention Layers

A second version of the model was created in which skip connections were replaced with attention layers. Attention mechanisms are used to capture correlations in the data across scales or channels. Here, an attention map was computed between feature maps of the same resolution in the encoding and decoding branches, similar to the method proposed in Ref. 31, but using the Luong model of attention[32] rather than additive attention. The attention map for a given layer of the U-net is produced by first transforming both the encoding branch feature map (q) and decoding branch feature map (k) via a convolution with a trainable 3x3x3 kernel. The resulting matrices are combined and then passed through another convolution layer with a $1 \times 1 \times 1$ kernel and sigmoid activation to produce the final attention map. The attention map was then multiplied element-wise with the decoding branch feature map.

2.3 Model Hyperparameters

Model hyperparameters were determined using a random search function (5 iterations, 3-fold cross validation) to compare a range of initial learning rates (0.0001, 0.001, 0.01), activation functions (ReLU, Leaky ReLU alpha=0.1, Leaky ReLU alpha=0.2, Leaky ReLU alpha=0.4), and dropout rates (0.1, 0.2, 0.4). The optimal combination of hyperparameters was found to be an initial learning rate of 0.0001, leaky ReLU activation function with alpha=0.2, and a dropout rate of 0.2.

2.4 Correcting For Class Imbalance

In classification problems where one class is more common than the other, ML algorithms can become biased towards the better represented class. This is common in image data containing blood vessels, as blood volume in biological tissue typically $3\,\%$ to $5\,\%$. To counter this, the training data generator was weighted towards batches of image subvolumes containing at least 0.1% vessels. This method meant that the model would still see some batches containing no vessels, but far less often than would otherwise be the case.

Three loss functions designed to handle imbalanced classes were compared: weighted categorical cross entropy (WCCE), DICE binary cross entropy (DICE BCE), and focal loss. WCCE used pixel-wise cross entropy, a measure of similarity between distributions, commonly used for segmentation and classification tasks, and scales the contribution of each pixel based on the ground truth class. Class weights were assigned in proportion to the inverse of the class size. DICE-BCE is a sum of the DICE coefficient and Binary Cross Entropy (BCE). The DICE coefficient is robust to imbalanced classes as it takes into account recall and precision, while BCE provides a smooth gradient for backpropagation.[33] Focal loss is designed to more highly weight pixel classifications that the model is 'uncertain' about, allowing the training to focus on sparse examples over easily classified background pixels. This is achieved by scaling standard cross entropy by a factor of $(1-p)^y$, where p is the model-estimated probability of the pixel belonging to the ground truth class, and y is a tuneable hyperparameter.[34]

2.5 Training Data

A library of paired image data and manual labels was created from three-dimensional image data acquired using a range of modalities. Data from four sources were chosen to represent a range

of sources of contrast, imaging resolutions and tissues types. As such, the training set would encompass a huge amount of variability with relatively little data. A summary of dataset used in training and in validation can be seen in Table 1.

Data set 1: A high resolution episcopic microscopy (HREM) dataset, acquired in house by staining a healthy mouse liver with Eosin B and imaged using a standard HREM protocol.[35] *Data set 2*: X-ray microCT images of a microvascular cast, taken from a subcutaneous mouse model of colorectal cancer (acquired in house). *Data set 3*: Raster-Scanning Optoacoustic Mesoscopy (RSOM) data from a subcutaneous tumour model (provided by Emma Brown, Bohndiek Group, University of Cambridge). *Data set 4*: retinal angiography data obtained using Optical Coherence Tomography Angiography (OCT-A) (provided by Dr Ranjan Rajendram, Moorfields Eye Hospital).

Manual labelling of blood vessels was carried out using Amira (2020.2, Thermo-Fisher, UK). Fig. 1 shows representative 2D sections from each dataset. Volumes representing $25\,\%$ of both the dataset 1 and 2 (HREM and CT) were set aside for testing of the model.

2.6 Validation Data

Additional data sets from modalities not included in the training library were used to validate the cross-modality generalisability of the model. *Validation set 1*: a subcutaneous colorectal tumour mouse model was imaged in house using Multi-fluorescence HREM in house, with Dylight 647 conjugated lectin staining the vasculature.[35] *Validation set 2*: an MF-HREM image of the cortex of a mouse brain, stained with Dylight-647 conjugated lectin, was acquired in house.[23] *Validation set 3*: two-photon data of mouse olfactory bulb blood vessels, labelled with sulforhodamine 101, was kindly provided by Yuxin Zhang at the Sensory Circuits and Neurotechnology Lab, the Francis

Table 1 Summary of data used in training and validation.

Dataset	Description	Volume (pixels)
Training 1	Eosin HREM, mouse liver	$4080 \times 3072 \times 416$
Training 2	CT, mouse subcutaneous tumour	$1000 \times 1000 \times 682$
Training 3	OCT-A, human retina	$500 \times 500 \times 64$
Training 4	RSOM, mouse subcutaneous tumour	$191 \times 221 \times 501$
Validation 1	MF-HREM, mouse subcutaneous tumour	$2659 \times 2034 \times 1561$, from which a subvolume of $480 \times 480 \times 480$ was used in fine-tuning
Validation 2	MF-HREM, mouse cortex	$4080 \times 3072 \times 99$
Validation 3	Two photon microscopy, mouse olfactory bulb	$1536 \times 1536 \times 79$, from which a subvolume of $500 \times 250 \times 79$ was use in fine-tuning

Crick Institute.

2.7 Training

Training data were prepared by normalising pixel intensities between zero and one. Optionally, the data was cropped at this point, for example in datasets where only a small region was of interest. Pixel data and true labels were converted into numpy arrays and saved in numpy proprietary format (npy) files. This file type is optimised for faster loading and occupies less memory than alternative formats, but lacks a metadata header. Therefore, an additional header class was created for easier data curation (Fig. 1).

Due to the large volume of data used in training, memory constraints became a limiting factor. Therefore, instead of loading entire data sets into motherboard RAM for training, a mini-batch of $64 \times 64 \times 64$ pixel sub-volumes were read in from the disk at the start of each iteration (Fig. 1). This method of loading in pixel data increases the time taken for training but allows the model

to be trained on more data without running out of memory. It also has the additional effect of acting as a translation augmentation of the data, as the sub-volumes were taken from random sets of coordinates within the training data. Additional augmentations are then performed on the generated subvolumes, randomly selecting from the following: rotations in the xy plane between $-30°$ and $30°$; rescaling by a factor between 1.0 and 1.25; reflection in one or both of the in-plane axes.

A range of optimisers and learning rate schemes were compared to improve speed of convergence and final accuracy of the model, with the Adam optimiser[36] and an exponential decay learning rate schedule ultimately being chosen for training. The model was trained with a batch size of two, which was the maximum that could be stored in the VRAM of the GPUs used for training, and stopped when no further improvement was seen in the accuracy of the validation set. Training took approximately 24 hours.

2.8 Evaluation

The quality of the segmentation produced by tUbeNet was evaluated throughout training using volumes of paired data from Training sets 1 and 2 (HREM and microvascular cast CT) that had been withheld from training (i.e. data that had not previously seen by the model). Model performance was measured by calculating the area under the Receiver Operating Characteristic (AUROC) curve and DICE score. The ability of the model to generalised to new data was assessed by calculating AUROC and DICE scores for manually labelled sub-volumes of Validation sets 1, 2 and 3 (in the case of Validation 1 and 3, where fine tuning was performed, a manually labelled sub-volume was withheld for testing).

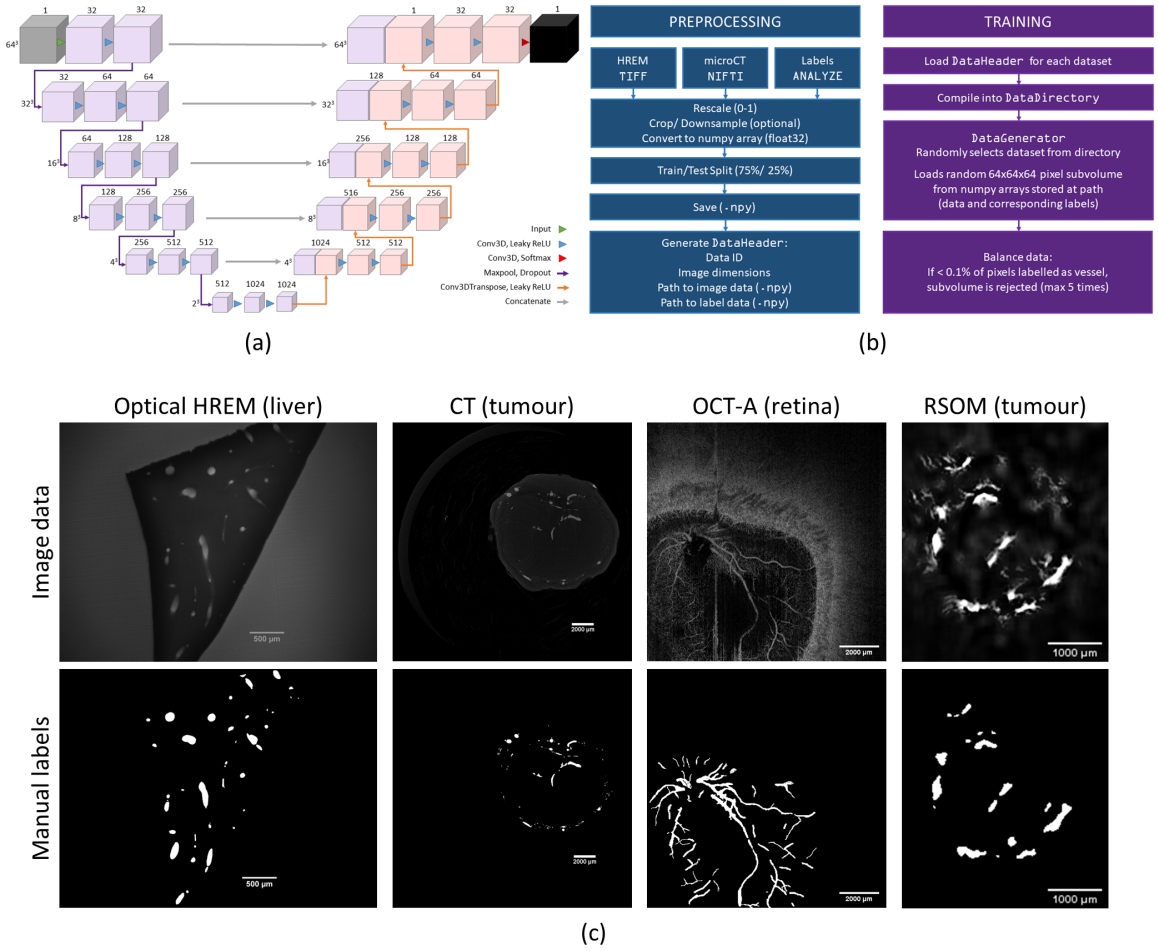

Fig 1 (a) This schematic shows the architecture of the tubenet model: the purple and blue boxes represent multi-channel feature maps, with the number of channels denoted by the number above the box. The dimensions to the side of each layer give the x-y-z dimensions of the feature maps in that layer. The network takes an image (grey) as an input and gives a map of probabilities as an output (black). The arrows represent operations, as summarised in the image key. (b) An overview of the data pre-processing pipeline, and method of loading data sub-volumes from the disk storage during training. The use of .npy file format and method of reading in data allows the model to train on larger datasets that cannot fit within the computer's RAM. (c) Example images taken from the three-dimensional image data used in training. 3D vessel images from four different modalities, and three different tissue types, were manually segmented to create a varied training set. Here, a 2D slice is shown from each dataset (top row), along with the corresponding manual labels (bottom row).

3 Results

3.1 Accounting For Class Imbalance

Three loss functions designed to address imbalanced classes were compared in order to identify the best loss function for the tUbenet model. Training with the weighted categorical crossentropy (WCCE) loss function produced the most accurate segmentation, achieving an AUROC of 0.9 on the validation data, but required iterative hyperparameter optimisation to identify the best combination of weights (in this case, 1:7). DICE-Binary crossentropy (DICE-BCE) outperformed focal loss (AUROC of 0.86 versus 0.80) and had the advantage of not requiring hyperparameter tuning. As such, DICE-BCE is used throughout this work unless otherwise stated.

3.2 Cross-branch Attention

The inclusion of a cross-branch attention mechanism sped up the initial rate at which the model learned the segmentation, with a AUROC score of 0.9 being achieved within 60 epochs, versus 200 epochs for the model without attention (Fig. 3). However, the attention model did not improve beyond 60 epochs, and both model versions converged at approximately the same level of accuracy. This result differs from the marked improvement in segmentation accuracy seen in Ref. 37. Despite converging in fewer epochs, the attention model had a much higher computational cost, meaning the training was slow and prone to memory errors. For these reasons the standard model structure without attention was preferred, and will be used throughout the rest of this work.

3.3 Cross-modality Generalisation

Through training on a library of data from different imaging modalities, the aim was to create a vessel-labelling model that would generalise to any volumetric dataset, even those acquired by pre-

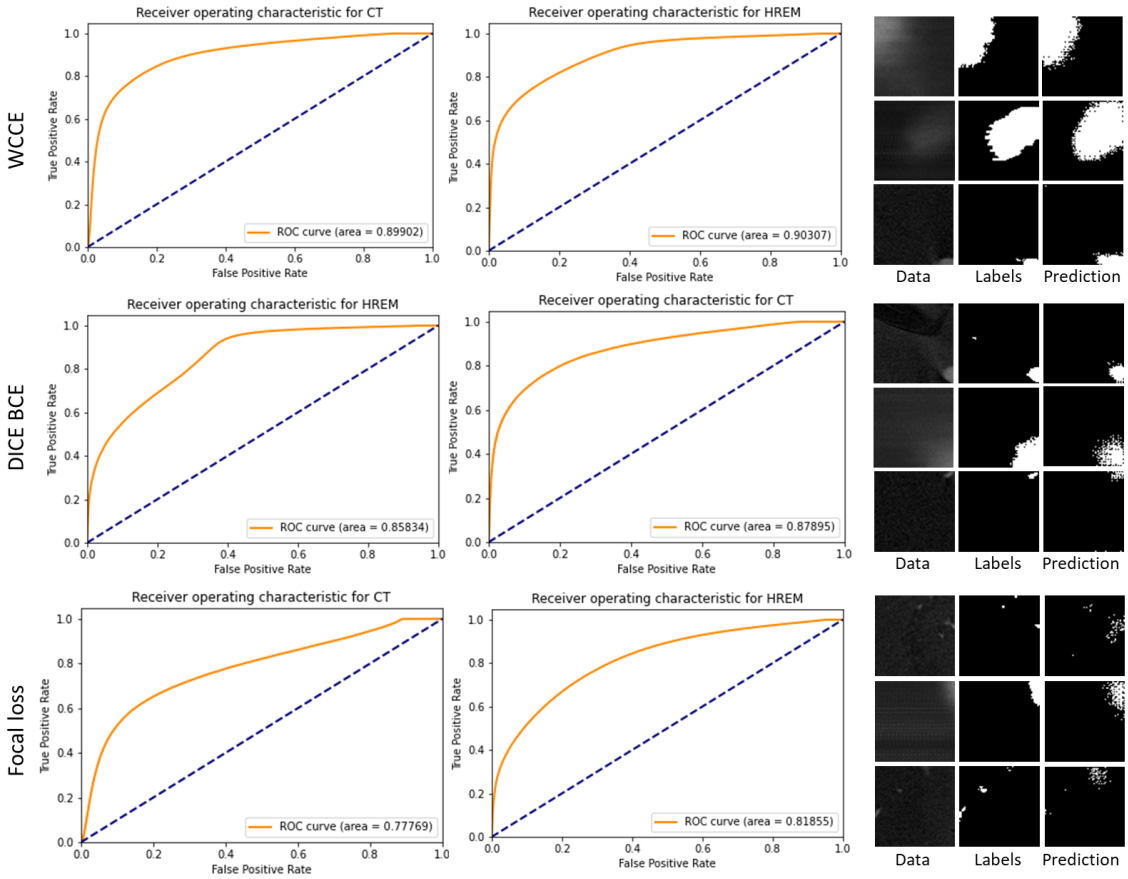

Fig 2 Comparison of ROC curves for three different loss functions (WCCE, DICE-BCE and focal loss), for two validation datasets. The left-hand panel shows the central slice of example $64 \times 64 \times 64$ volumes, along with the ground truth and predicted labels for each model. WCCE outperformed both the DICE BCE and focal loss functions, with focal loss producing the lowest AUROC score and visibly worse segmentations.

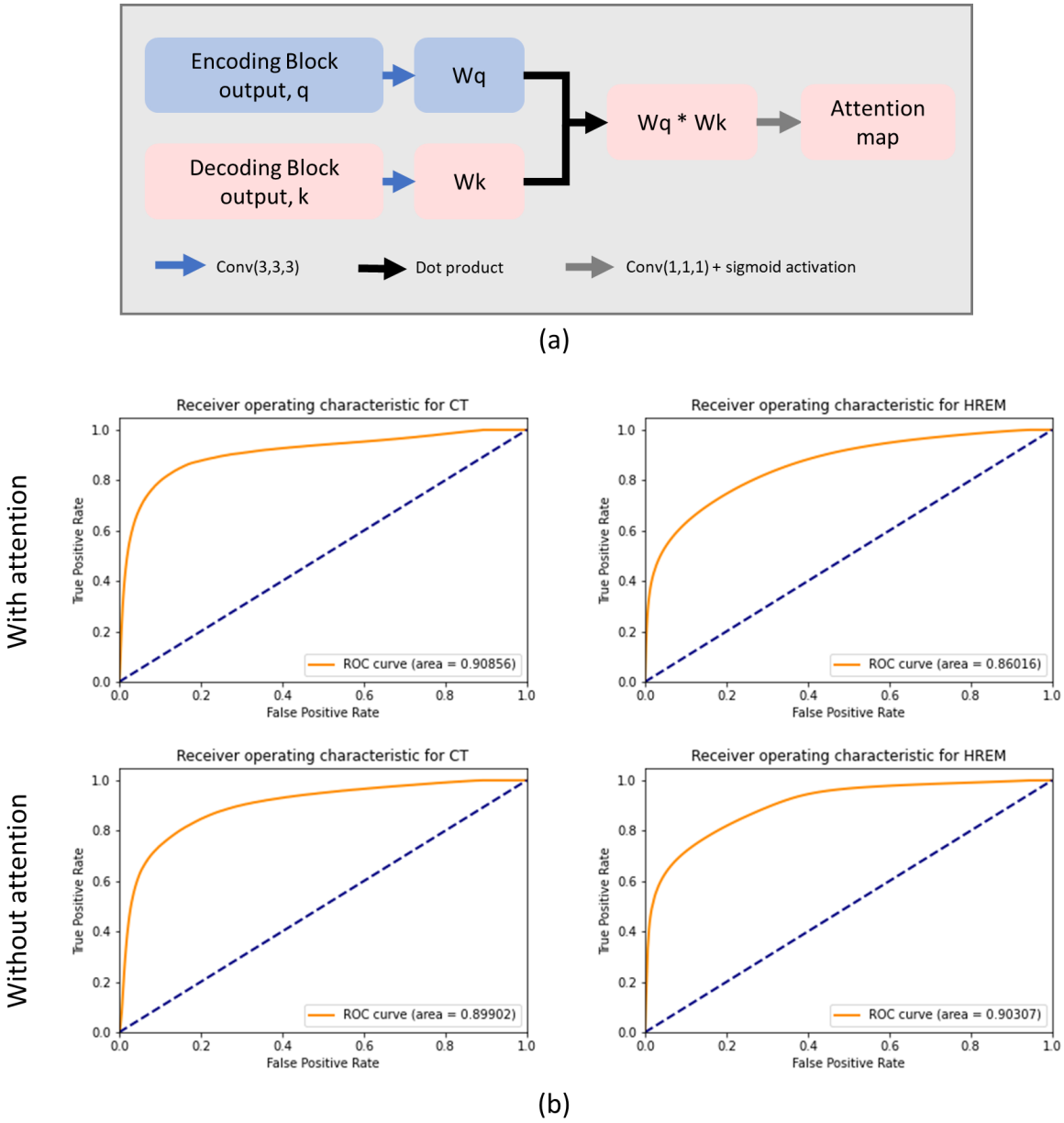

Fig 3 (a) A schematic illustration of the attention method used in this model. Concatenation operations, represented by the grey arrows in Fig. 1, were replaced with an attention layer as shown here. The feature maps from corresponding encoding and decoding layers (q and k respectively) are convolved with trainable kernels to produce Wq and Wk. These outputs are combined to find the dot product. A final convolution and sigmoid activation function produces the attention map for this layer, which is multiplied by the decoding feature map. (b) Receiver operating characteristic curves for the model with (top) and without (bottom) the inclusion of attention layers. The inclusion of attention did not improve segmentation quality when tested on the two validation datasets.

viously unseen modalities, with minimal fine tuning. To test whether this aim had been achieved, tUbenet was evaluated on positive-contract MF-HREM images of a subcutaneous hypopharyngeal cancer model.[35] This modality was chosen due to the relative newness of MF-HREM: there is very little labelled data available for training, meaning existing deep learning approaches would not be suitable for analysing this data.

Applying the trained model without fine-tuning to this new data type, we produced a segmentation of the vasculature with high recall of the vessel class ($96\,\%$) but low precision ($23\,\%$) (2 s.f.). Erroneous labelling of out-of-plane fluorescence as vasculature was noted, leading to the low precision score (Fig. 4). However, after two hours of fine-tuning the model using a labelled subvolume of the data ($480 \times 480 \times 120$ pixels) representing just $0.3\,\%$ of the total image volume, the recall and precision were improved to 0.98 and 0.97 respectively (DICE score = 0.97) (2 s.f.). Increasing the train set size to $480 \times 480 \times 480$ pixels ($1.2\,\%$ of the total image volume) further improved the segmentation accuracy (precision 0.99, recall 0.97, DICE 0.98 (2 s.f.)). This sub-volume was manually annotated in a matter of hours, compared to the days/weeks that would be required to label the entire dataset. This result demonstrates that our pre-trained model can be fine-tuned to new data with a minimal amount of manually annotation, making it a practical option for applications where large libraries of labelled data do not exist.

Moreover, the fine-tuned model was then applied to another MF-HREM dataset from a different tissue: the cortex of a mouse brain stained with DyLight-647 conjugated lectin (Fig. 4). Despite differences in cortical and tumour vasculature, the model was able to generalise to the cortex dataset with a high level of accuracy and no additional fine-tuning (recall=0.90, precision=0.88, DICE score=0.89 (2 s.f.)). This suggests that fine-tuning may only need to be done once per image type, further saving on manual labelling.

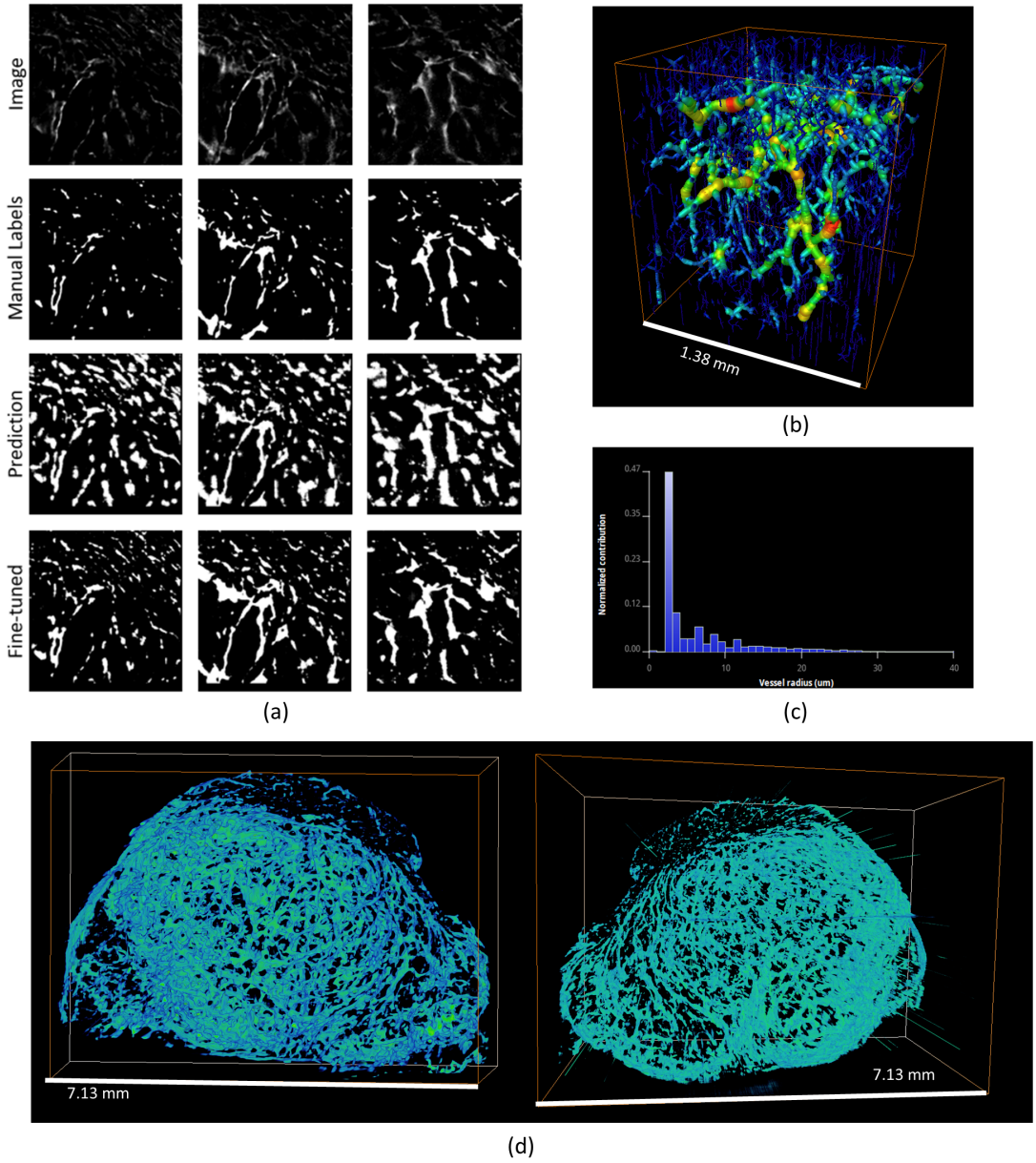

Fig 4 (a) 2D sections from an MF-HREM dataset of a fluorescence-stained subcutaneous tumour (top row) are shown alongside the corresponding manual labels (second row) and the predicted labels generated by the pre-trained model prior to fine-tuning (third row) and following fine-tuning (bottom row). (b) A skeletonised subvolume of the tumour and corresponding vessel radii distribution (c). (d) 3D views of the entire dataset segmented using the fine-tuned model.

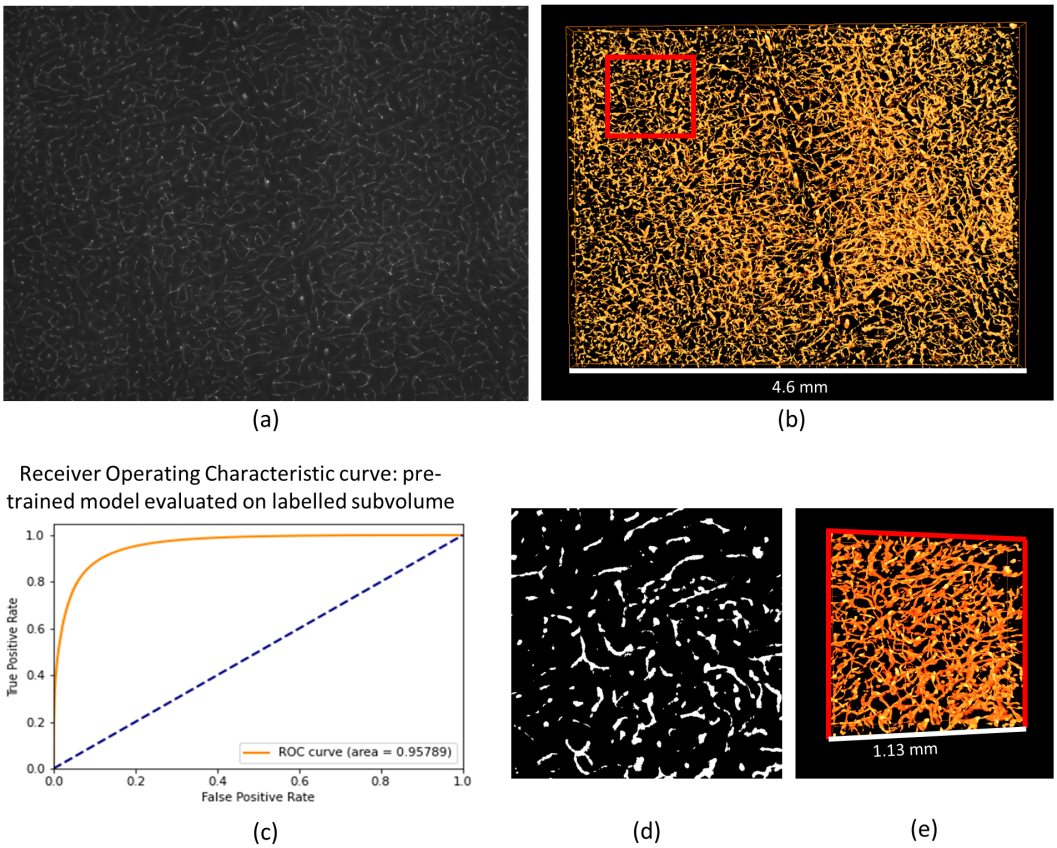

Fig 5 (a) 2D section of an MF-HREM image of the cortical vasculature of a mouse. (b) 3D vessel segmentation produced by the tubenet model fine-tuned using the tumour MF-HREM dataset seen in Fig 4. The red square highlights a region of the data that was manually labelled to provide a ground truth for validation. (c) ROC curve analysis performed using the manually labelled ground truth (d) and model predictions for this subvolume (e) shows that the prediction is highly accuracy (AUROC = 0.96 (2 s.f.)).

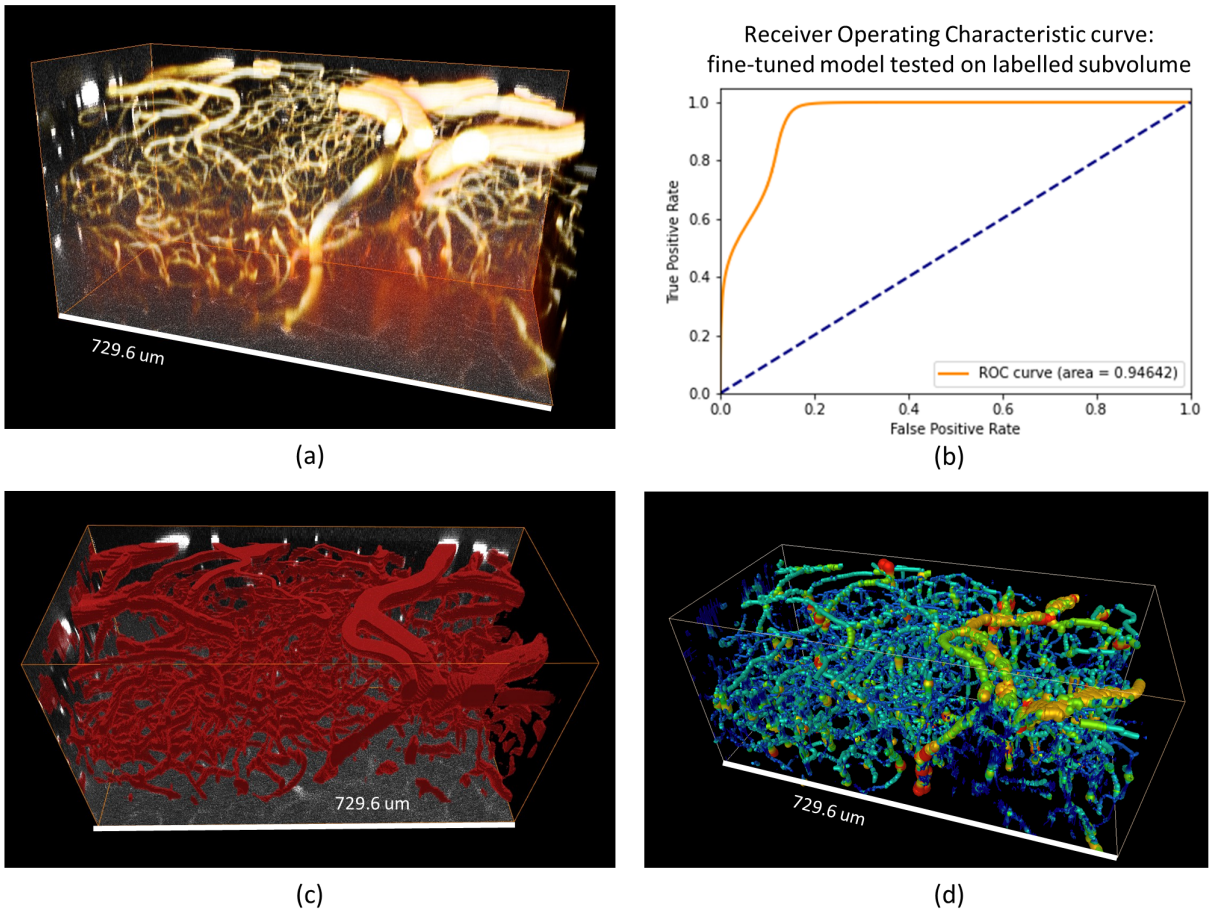

Fig 6 (a) Two photon fluorescence microscopy data showing labelled vasculature (yellow) in the mouse olfactory bulb. (b) ROC curve analysis of the model prediction following fine-tuning on a manually labelled subvolume of the data shows that the model is accurately labelling vasculature (AUROC = 0.95 (2 s.f.)). (c) A 3D visualisation of the predicted vessel labels (red). (d) Skeletonised vasculature, colour coded by vessel radius.

Finally, the model was applied to images of the blood vessels of the mouse olfactory bulb, acquired by two photon microscopy. This dataset provided an additional challenge as, not only it an imaging modality that the model had not previously seen, but it exhibits spatially varying imaging artefacts due to resolution drop off with depth. In order to specialise the model to this data, a sub-volume that spanned the whole z-axis (79 virtual sections) was manually labelled for fine-tuning. Training on a volume of $79 \times 500 \times 250$ pixels ($5\,\%$ of the whole image volume) resulted in a model that labelled vasculature with a precision of 0.84 and recall of 0.98 (2 s.f.) (Fig. 6).

4 Discussion

In this work a 3D U-net for vessel segmentation was trained on a highly varied dataset, comprised of four imaging modalities and three tissue types. The general model was then specialised for new applications - segmenting data from previously unseen modalities - with minimal fine-tuning. Our results show that this two-step training approach to deep learning for image analysis is a viable strategy. Fine-tuning the general model with a small sub-volume of the new data of interest (as little as $< 1\%$ of the data volume) was sufficient to generate accurate label predictions for the entire dataset. This approach to deep learning is especially advantageous when analysing data from new or lesser-used imaging modalities, where a back catalogue of data is not available to train a model from scratch.

In the future this model could be made more accessible to non-computational researchers creating a complementary graphical user interface for manual labelling and fine-tuning, or by integrating it into existing open-source data visualisation programs (for example, as a Napari plug in). Overall, the move towards an open-source, generalisable DL tool for vessel segmentation expands

the options available to researchers looking to quantify vaculature in 3D and overcomes many of the bottlenecks preventing wider use of DL in the 3D imaging community.

Disclosures

The authors have no financial interests or other potential conflicts of interest to disclose.

Acknowledgments

Thank you to Dr Ranjan Rajendram (Moorfields Eye Hospital), Yuxin Zhang and Dr Carlos Bosh Pinol (Francis Crick Institute), Dr Emma Brown and Dr Sarah Bohndiek (University of Cambridge) for kindly providing data that was used in model training and validation.

This research was funded by Cancer Research UK (C44767/A29458 and C23017/A27935) and EPSRC (EP/W007096/1). Dr Walsh was supported by an MRC Skills Development fellowship (MR/S007687/1).

Code, Data, and Materials Availability

The code used in this paper is available at https://github.com/natalie11/tUbeNet. Model weights and training data are available upon request.

References

1 D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation.," *Cell* **144**, 646–74 (2011).

2 L. Zanoli, M. Briet, J. P. Empana, *et al.*, "Vascular consequences of inflammation: a position statement from the ESH Working Group on Vascular Structure and Function and the ARTERY Society," *Journal of hypertension* **38**, 1682 (2020).

3 K. Thygesen, J. S. Alpert, A. S. Jaffe, *et al.*, "Fourth Universal Definition of Myocardial Infarction (2018)," *Journal of the American College of Cardiology* **72**, 2231–2264 (2018).

4 G. J. Del Zoppo and J. M. Hallenbeck, "Advances in the Vascular Pathophysiology of Ischemic Stroke," *Thrombosis Research* **98**, 73–81 (2000).

5 T. Dinh, S. Scovell, and A. Veves, "Review: Peripheral Arterial Disease and Diabetes: A Clinical Update," *http://dx.doi.org/10.1177/1534734609336768* **8**, 75–81 (2009).

6 L. Luo, D. Chen, and D. Xue, "Retinal blood vessels semantic segmentation method based on modified U-Net," in *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, 1892–1895, IEEE (2018).

7 A. Chowdhury, D. V. Dylov, Q. Li, *et al.*, "Blood vessel characterization using virtual 3D models and convolutional neural networks in fluorescence microscopy," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 629–632, IEEE (2017).

8 P. Liskowski and K. Krawiec, "Segmenting Retinal Blood Vessels With Deep Neural Networks," *IEEE Transactions on Medical Imaging* **35**, 2369–2380 (2016).

9 H. Fu, Y. Xu, D. W. K. Wong, *et al.*, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 698–701, IEEE (2016).

10 S. Moccia, E. De Momi, S. El Hadji, *et al.*, "Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics," *Computer Methods and Programs in Biomedicine* **158**, 71–91 (2018).

11 M. I. Todorov, J. C. Paetzold, O. Schoppe, *et al.*, "Machine learning analysis of whole mouse brain vasculature," *Nature Methods* **17**, 442–449 (2020).

12 G. Tetteh, V. Efremov, N. D. Forkert, *et al.*, "DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes," *Frontiers in Neuroscience* **14**, 1285 (2020).

13 C. Stringer, T. Wang, M. Michaelos, *et al.*, "Cellpose: a generalist algorithm for cellular segmentation," *Nature Methods 2020 18:1* **18**, 100–106 (2020).

14 M. Pachitariu and C. Stringer, "Cellpose 2.0: how to train your own model," *Nature Methods 2022 19:12* **19**, 1634–1641 (2022).

15 O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **9351**, 234–241 (2015).

16 M. Drozdzal, E. Vorontsov, G. Chartrand, *et al.*, "The importance of skip connections in biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **10008 LNCS**, 179–187 (2016).

17 S. Candemir, X. V. Nguyen, L. R. Folio, *et al.*, "Training strategies for radiology deep learning models in data-limited scenarios," *Radiology: Artificial Intelligence* **3** (2021).

18 D.-X. Xue, R. Zhang, H. Feng, *et al.*, "CNN-SVM for Microvascular Morphological Type Recognition with Data Augmentation," *Journal of Medical and Biological Engineering* **36**, 755–764 (2016).

19 R. Ma, P. Tao, and H. Tang, "Optimizing data augmentation for semantic segmentation on small-scale dataset," in *ACM International Conference Proceeding Series*, 77–81, Association for Computing Machinery, (New York, New York, USA) (2019).

20 A. van Engelen, A. C. van Dijk, M. T. B. Truijman, *et al.*, "Multi-Center MRI Carotid Plaque Component Segmentation Using Feature Normalization and Transfer Learning," *IEEE Transactions on Medical Imaging* **34**, 1294–1305 (2015).

21 N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Transactions on Medical Imaging* **35**, 1299–1312 (2016).

22 H. Azizpour, A. S. Razavian, J. Sullivan, *et al.*, "From generic to specific deep representations for visual recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 36–45, IEEE (2015).

23 P. W. Sweeney, C. Walsh, S. Walker-Samuel, *et al.*, "A three-dimensional, discrete-continuum model of blood flow in microvascular networks," *bioRxiv* , 2022.11.23.517681 (2022).

24 X. Dong, Y. Lei, S. Tian, *et al.*, "Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network," *Radiotherapy and Oncology* **141**, 192–199 (2019).

25 I. Sirazitdinov, H. Schulz, A. Saalbach, *et al.*, "Tubular Shape Aware Data Generation for Semantic Segmentation in Medical Imaging," *arXiv* (2020).

26 J. Jordon, L. Szpruch, F. Houssiau, *et al.*, "Synthetic Data – what, why and how?," (2022).

27 J. Staal, M. D. Abràmoff, M. Niemeijer, *et al.*, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging* **23**, 501–509 (2004).

28 A. Hoover, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging* **19**(3), 203–210 (2000).

29 K. Jin, X. Huang, J. Zhou, *et al.*, "FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation," *Scientific Data 2022 9:1* **9**, 1–8 (2022).

30 R. Li, Y.-J. Huang, H. Chen, *et al.*, "3D Graph-Connectivity Constrained Network for Hepatic Vessel Segmentation; 3D Graph-Connectivity Constrained Network for Hepatic Vessel Segmentation," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS* **26**(3), 1251 (2022).

31 O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," (2018).

32 M. T. Luong, H. Pham, and C. D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing* , 1412–1421 (2015).

33 J. Ma, J. Chen, M. Ng, *et al.*, "Loss odyssey in medical image segmentation," *Medical Image Analysis* **71**, 102035 (2021).

34 T.-Y. Lin, P. Goyal, R. Girshick, *et al.*, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2980–2988 (2017).

35 C. Walsh, N. A. Holroyd, E. Finnerty, *et al.*, "Multifluorescence High-Resolution Episcopic Microscopy for 3D Imaging of Adult Murine Organs," *Advanced Photonics Research* **2**, 2100110 (2021).

36 D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR (2015).

37 O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," (2022).

List of Figures

1 (a) This schematic shows the architecture of the tubenet model: the purple and blue boxes represent multi-channel feature maps, with the number of channels denoted by the number above the box. The dimensions to the side of each layer give the x-y-z dimensions of the feature maps in that layer. The network takes an image (grey) as an input and gives a map of probabilities as an output (black). The arrows represent operations, as summarised in the image key. (b) An overview of the data pre-processing pipeline, and method of loading data sub-volumes from the disk storage during training. The use of .npy file format and method of reading in data allows the model to train on larger datasets that cannot fit within the computer's RAM. (c) Example images taken from the three-dimensional image data used in training. 3D vessel images from four different modalities, and three different tissue types, were manually segmented to create a varied training set. Here, a 2D slice is shown from each dataset (top row), along with the corresponding manual labels (bottom row).

2 Comparison of ROC curves for three different loss functions (WCCE, DICE-BCE and focal loss), for two validation datasets. The left-hand panel shows the central slice of example $64 \times 64 \times 64$ volumes, along with the ground truth and predicted labels for each model. WCCE outperformed both the DICE BCE and focal loss functions, with focal loss producing the lowest AUROC score and visibly worse segmentations.

3 (a) A schematic illustration of the attention method used in this model. Concatenation operations, represented by the grey arrows in Fig. 1, were replaced with an attention layer as shown here. The feature maps from corresponding encoding and decoding layers (q and k respectively) are convolved with trainable kernels to produce Wq and Wk. These outputs are combined to find the dot product. A final convolution and sigmoid activation function produces the attention map for this layer, which is multiplied by the decoding feature map. (b) Receiver operating characteristic curves for the model with (top) and without (bottom) the inclusion of attention layers. The inclusion of attention did not improve segmentation quality when tested on the two validation datasets.

4 (a) 2D sections from an MF-HREM dataset of a fluorescence-stained subcutaneous tumour (top row) are shown alongside the corresponding manual labels (second row) and the predicted labels generated by the pre-trained model prior to fine-tuning (third row) and following fine-tuning (bottom row). (b) A skeletonised sub-volume of the tumour and corresponding vessel radii distribution (c). (d) 3D views of the entire dataset segmented using the fine-tuned model.

5 (a) 2D section of an MF-HREM image of the cortical vasculature of a mouse. (b) 3D vessel segmentation produced by the tubenet model fine-tuned using the tumour MF-HREM dataset seen in Fig 4. The red square highlights a region of the data that was manually labelled to provide a ground truth for validation. (c) ROC curve analysis performed using the manually labelled ground truth (d) and model predictions for this subvolume (e) shows that the prediction is highly accuracy (AUROC = 0.96 (2 s.f.)).

6 (a) Two photon fluorescence microscopy data showing labelled vasculature (yellow) in the mouse olfactory bulb. (b) ROC curve analysis of the model prediction following fine-tuning on a manually labelled subvolume of the data shows that the model is accurately labelling vasculature (AUROC = 0.95 (2 s.f.)). (c) A 3D visualisation of the predicted vessel labels (red). (d) Skeletonised vasculature, colour coded by vessel radius.

List of Tables

1 Summary of data used in training and validation.