Segmentation of 3D blood vessel networks using unsupervised deep learning

Paul W. Sweeney[1,2], Lina Hacker[1,2], Thierry L. Lefebvre[1,2], Emma L. Brown[1,2], Janek Gröhl[1,2], and Sarah E. Bohndiek[1,2]

[1]Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK
[2]Department of Physics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK

Keywords— Unsupervised, deep learning, segmentation, blood vessels, photoacoustics

Abstract

Innovations in imaging hardware have led to a revolution in our ability to visualise vascular networks in 3D at high resolution. The segmentation of microvascular networks from these 3D image volumes and interpretation of their meaning in the context of physiological and pathological processes unfortunately remains a time consuming and error-prone task. Deep learning has the potential to solve this problem, but current supervised analysis frameworks require human-annotated ground truth labels. To overcome these limitations, we present an unsupervised image-to-image translation deep learning model called the *vessel segmentation generative adversarial network* (VAN-GAN). VAN-GAN integrates synthetic blood vessel networks that closely resemble real-life anatomy into its training process and learns to replicate the underlying physics of an imaging system in order to learn how to segment vasculature from 3D biomedical images. To demonstrate the potential of VAN-GAN, the framework was applied to the challenge of segmenting vascular networks from images acquired using mesoscopic photoacoustic imaging (PAI). With a variety of *in silico*, *in vitro* and *in vivo* pathophysiological data, including patient-derived breast cancer xenograft models, we show that VAN-GAN facilitates accurate and unbiased segmentation of 3D vascular networks from PAI volumes. By leveraging synthetic data to reduce the reliance on manual labelling, VAN-GAN lower the barriers to entry for high-quality blood vessel segmentation to benefit imaging studies of vascular structure and function.

Introduction

The past decade has brought about a revolutionary change in our ability to visualise vascular networks in 3D at a spatial resolution never before possible. Advances in photoacoustic mesoscopy [1–3] and ultrafast ultrasound [4,5] provide *in vivo* images of vessel form and function at the micron scale in living subjects, while tools such as X-ray histology [6,7] and light sheet fluorescence microscopy [8,9] yield unprecedented insight into vascular architectures at the sub-micron scale *ex vivo* in intact specimens. The segmentation of microvascular networks from these 3D image volumes and interpretation of their meaning in the context of physiological and pathological processes is now a crucial task in biological research, yet currently lags behind the phenomenal pace of development in imaging hardware and scale of data output.

With an impressive ability to extract biological features across vast datasets accurately and rapidly, deep learning is a catalyst for advances in the field of semantic segmentation, where pixels are automatically classified into classes (e.g. vessels and background tissue) based on features of interest [10]. Current supervised analysis frameworks [11–13] typically require matched pairs of image volumes with human-annotated ground truth labels [14–16]. Human annotation is time-consuming, so typically limited to just a few image pairs. Furthermore, the presence of artefacts that depend on the applied imaging modality, as well as the varied interpretation by users of what constitutes as vascular networks, can make human annotation error-prone, which is especially evident in complex pathological tissues like tumours [17,18].

Given these challenges associated with manual labelling, semi-supervised and unsupervised image-to-image translation models have naturally gained special attention for image segmentation tasks [19]. Here, we introduce an unsupervised image-to-image translation deep learning model called *vessel segmentation generative adversarial network* (VAN-GAN), which segments 3D vascular networks from biomedical image volumes without any hand-annotated labels. Generative adversarial networks (GANs) can be used to transform images from a source domain to a target domain, for example, between a biomedical image volume and its corresponding segmentation mask. VAN-GAN is based on a cycle-consistent generative adversarial network (CycleGAN) [20], which performs style transfer between unpaired images by allowing two neural network translators to be trained in an unsupervised and competitive manner, assuming an underlying relationship between each domain. CycleGAN, and modified formulations, have been shown to accurately segment a wide range of different biomedical images [21–25], including cell nuclei from 3D fluorescence microscopy [26] and 2D X-ray coronary angiograms [27]. Our model builds upon CycleGAN in several ways in order to target segmentation of blood vessel networks from 3D images, specifically, by: adopting 3D deep residual U-Net generators to improve image generation, including additional reconstruction loss functions to enhance segmentation performance, and applying random noise to discriminator inputs and convolution layers for improved training stability and regularisation.

Prior studies using unsupervised methods for 3D vascular segmentation are dependent on image enhancement algorithms [28] or active contours [29]. Furthermore, currently available models may be hindered by: sensitivity to choice of parameters, signal-to-noise ratio (SNR) inhomogeneities,

differing vessel morphologies, class imbalance between vessel features and background, or a lack of 3D ground truth segmentation examples[20,28,29]. Our VAN-GAN model alleviates these limitations. Drawing from supervised methods that integrate synthetic biomedical data[22,23,26,27], VAN-GAN learns to segment 3D vasculatures from a dataset of vessel labels stochastically created using mathematical formalisms. These mathematical descriptions replicate the natural features of blood vessels (such as diameters, lengths and bifurcation angles), as well as vascular pathologies (such as aneurysms and stenoses). The resulting rich and complex image dataset of synthetic morphological examples with known segmentation labels enables VAN-GAN to learn how to segment real, unpaired image volumes in an unsupervised manner.

To demonstrate the potential of VAN-GAN, we tackle the challenge of segmenting blood vessel networks from images volumes acquired *in vivo* using mesoscopic photoacoustic imaging (PAI). Mesoscopic PAI is an exciting emerging modality that provides non-invasive label-free images of vascular networks at a a spatial resolution of around ~ 20 μm at up to 3 mm depth based on the absorption of light by haemoglobin. PAI has shown promise in a range of biomedical applications, from diagnosis of skin disorders to grading of inflammation[30–33]. The formation of photoacoustic images involves laser excitation and acoustic detection, which introduces a diverse and unique set of imaging artefacts[34]. In the context of 3D vascular imaging, these artefacts include: misrepresented vessel diameters, false network discontinuities, and segmentation of feature echoes[35]; all of which can hinder vessel segmentation and make it challenging to perform quantitative analysis required for diagnostic purposes.

First, we use our synthetic vascular dataset in a physics-driven simulation framework to create realistic PAI image volumes. This paired dataset allows us to highlight respective strengths and weaknesses of VAN-GAN. Next, we confront the issue of user-bias when producing hand-annotations, showing significant variability between expert users in generating ground truth labels. Then, through a combination of *in silico*, *in situ*, and *in vivo* datasets, we show that the observed performance enhancement seen in these supervised evaluations is attributable to VAN-GAN overcoming artefacts intrinsic to PAI. Finally, we show that VAN-GAN advances our ability to quantify tumour vascular networks imaged using PAI. This work has important implications in PAI, and for the biomedical imaging field more generally, by facilitating accurate and unbiased 3D vascular segmentations.

Results

VAN-GAN (Vessel Segmentation Generative Adversarial Network) builds upon the CycleGAN model[20] in several ways (see Methods and Supplementary Methods for full implementation details). Firstly, we improve information propagation by combining U-Net[11] and deep residual learning[36] into a deep residual U-Net generator architecture[37]. Secondly, we apply random Gaussian noise to discriminator inputs and convolution layers for improved training stability and regularisation. Thirdly, we extend the design to 3D for image volume generation using 3D convolutions. Finally, we use a modified objective loss function that: 1) excludes identity loss, as it only preserves the tint of generated im-

ages[20] which is unnecessary for grayscale images; 2) introduces a structure similarity reconstruction loss between real and cycled biomedical images[38]; and 3) applies a spatial and topological constraint between real and cycled blood vessel labels by measuring the overlap of these tubular-based structures[39] (Fig. 1A). The network architecture consists of three parts: encoder, bridge and decoder[37] (detailed in Supplementary Table 1). Residual units with a full pre-activation design[40] were used, each composed of two convolutional blocks and an identity mapping, which connects the unit input and output (Fig. 1B). The encoder and decoder consist of four layers with residual units connected via skip connections (Fig. 1C). The convolutional block consist of a sequence of instance normalisation[41], ReLU activation[42] and a 3D convolutional layer. Prior to applying convolutions, reflection padding was used to reduce boundary artefacts in feature maps[20,43]. VAN-GAN is trained on image subvolumes with predicted segmentations patched together in a sliding window approach for an entire image volume (see Fig. 1D).

To make a comparison between VAN-GAN and prior art using supervised learning-based segmentation, we also train a random forest pixel classifier (RF)[35,44]. RFs are preferred for this task as they do not require a large amount of training data, avoiding difficulties with volumetric annotation and imaging artefacts (see Methods and Supplementary Methods).

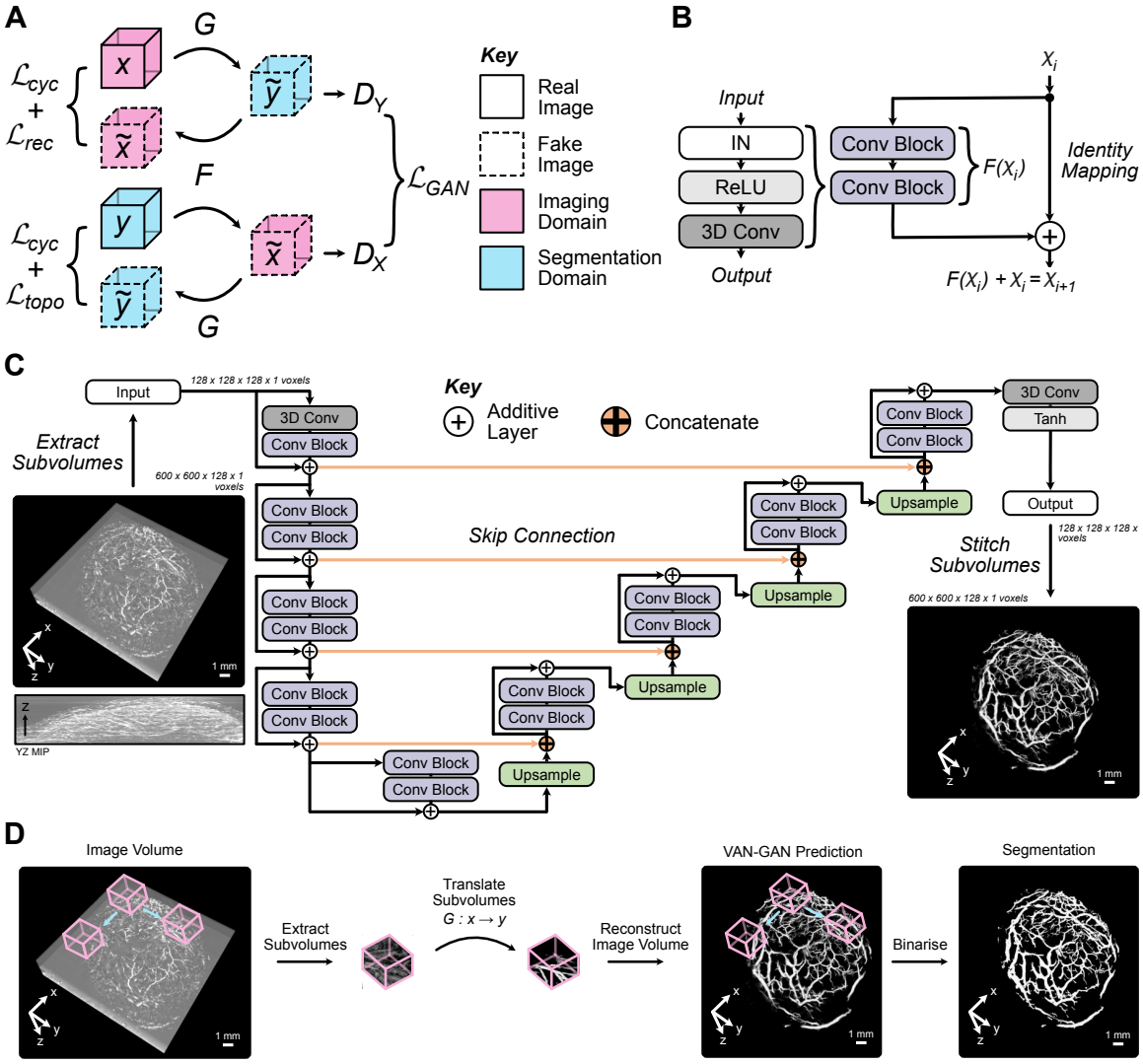

Figure 1: **Vascular segmentation generative adversarial network (VAN-GAN) architecture for unsupervised segmentation of blood vessels in image volumes.** (A) VAN-GAN modifies the cycleGAN model for image-to-image translation of image volumes to automate segmentation of blood vessels. VAN-GAN learns mappings between the imaging, x and segmentation, y, domains ($G : x \rightarrow \tilde{y}$ and $F : y \rightarrow \tilde{x}$) by applying additional cycle-consistency constraints in the form of a reconstruction loss, $\mathcal{L}_{rec}$, for forward-consistency ($x \rightarrow \tilde{x}$), and a topological loss, $\mathcal{L}_{topo}$, for backward-consistency ($y \rightarrow \tilde{y}$). (B) Residual units with a full pre-activation design are integrated into (C) a 3D deep residual U-Net generator architecture where image subvolumes ($128 \times 128 \times 128 \times 1$ voxels corresponding to $depth \times width \times height \times channel$ or $Z \times X \times Y \times C$) are used as input. A photoacoustic mesoscopy example is shown at the input image (C, left) and output segmentation (C, right). (D) A subvolume window is strided across the image to extract subvolumes that

are then segmented and stitched together to form a probablility map for the entire image volume, which is then binarised to create a segmentation mask. Note, VAN-GAN also uses an equivalent architecture to generate fake images of the imaging domain. MIP = maximal intensity projection.

VAN-GAN segments realistic vascular network structure and topology

Training and validation of unsupervised methods for vessel segmentation is challenging without 3D ground truth labels. Motivated by this issue, we created a fully-synthetic paired dataset as a performance baseline to compare methods. The paired dataset is comprised of image volumes of computer-generated 3D vascular networks, where each has undergone a physics-driven photoacoustic simulation (see Methods, Supplementary Methods, Supplementary Fig. 1 and 2). The paired dataset was extremely powerful, enabling VAN-GAN to perform unsupervised learning between the imaging domain and segmentation domains and then providing validation of segmentation predictions in a supervised manner (Fig. 2A). The paired dataset was also used to train a supervised RF model (in ilastik, see Methods) using the exact 3D locations of all vessel labels, enabling comparisons to a supervised segmentation approach using complete annotations.

Training on these paired data showed that VAN-GAN was able to learn and generate fake photoacoustic image volumes that included typical photoacoustic artefacts, arising from limited illumination and detection views, as well as shadowing from overlying absorbers[34] (Fig. 2B). The ability to capture these instrument-specific artefacts enabled VAN-GAN to predict segmentation masks of the simulated photoacoustic domains with overall improved accuracy compared to the RF when comparing common segmentation metrics (Fig. 2C-F and Supplementary Table 2). Qualitative examination of VAN-GAN segmentation predictions showed errors were largely confined to the smallest vessels, which exhibited relatively low signal-to-noise ratios, as these vessels were furthest from the simulated light-source in the tissue (Fig. 2G).

Regardless of the imaging modality, pertinent metrics are necessary to validate performance of a segmentation model appropriately[45]. Standard machine learning metrics, such as F1 Score and Intersection over Union (IoU), do not provide a complete picture of segmentation performance in biomedical images, and so alternative metrics were employed. To evaluate segmentation more robustly, we computed established descriptors of network structure and topology on the skeletonised 3D vessel masks[35,46] for both VAN-GAN and RF segmentations (see Supplementary Table 3). VAN-GAN predicted greater network connectivity and maintained performance with increasing network complexity (Fig. 3B). Further, VAN-GAN represented vascular density more accurately (Fig. 3C). VAN-GAN tended to under-predict the number of vessels in more complex networks (< 0.3 mm in length and branching orders ≥ 12), while RF tended to overestimate the number of vessels for decreasing network complexity (Fig. 3D). These quantitative findings thus support the qualitative observation that VAN-GAN errors tend to be concentrated among the smallest vessels.

On average, VAN-GAN overestimates the lengths of vessels and underestimates the number

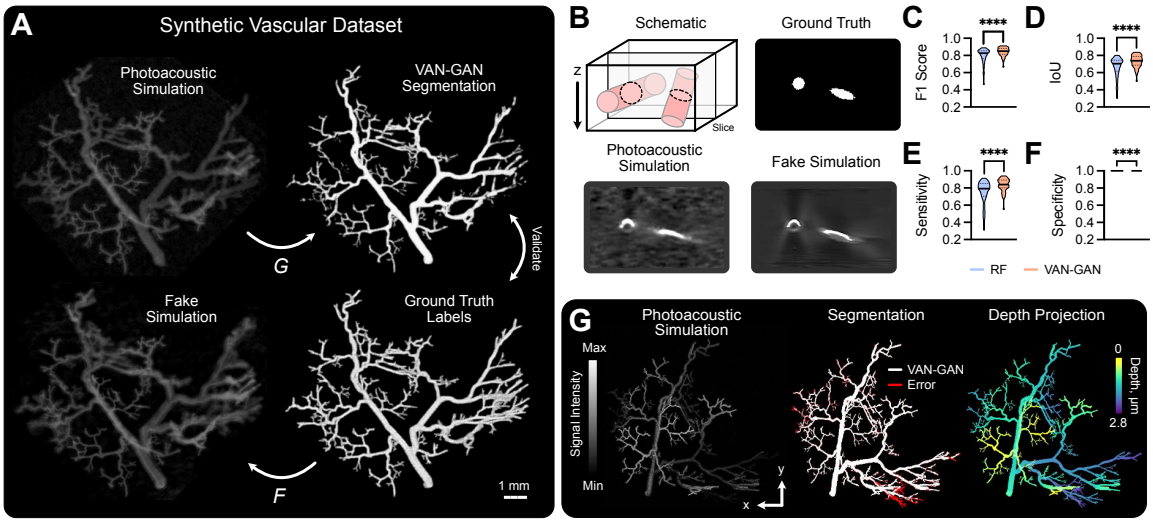

Figure 2: **VAN-GAN performance on synthetic photoacoustic data.** (A) We developed a paired dataset consisting of synthetic vascular labels with corresponding physics-driven photoacoustic simulation image volumes. Here, two neural networks, G and F, are trained to map computer−generated images to the segmentation and fake simulation domains, respectively. The paired dataset enables validation of the VAN-GAN segmentation mask, via G, in a typical supervised manner. (B) Fake simulation images generated by generator F reproduce artefacts inherent to photoacoustic mesoscopy (present in our simulations) that influence vessel shape and continuity. The example illustrates an artefact arising from limited angular coverage of the illumination fibres, which means the circular cross-section appears as an arc. Segmentation metric distributions between a random forest pixel classifier (RF) and VAN-GAN: (C) F1 Score; (D) Intersection over Union (IoU); (E) Sensitivity; and (F) Specificity. (G) An example of (left) a photoacoustic simulation, (middle) the corresponding VAN-GAN segmentation labels and error with respect to the ground truth, and (right) a colour-coded depth projection of the network. Data in (C-F) are represented by truncated violin plots with interquartile range (dotted) and median (solid) shown. Note, the standard deviation in (F) is 2.3×10^{-4}. Statistical significance indicated by * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) and **** ($P < 0.0001$).

of branching nodes (Fig. 3E,F), which would occur if a daughter vessel was not segmented at a bifurcation. In such a case, the remaining parent and daughter vessels are skeletonised as a single vessel, resulting in an overestimation of vessel length and the removal of a bifurcation node. On the other hand, the supervised RF classifier overestimates the number of vessels in less complex networks (Fig. 3D) and also had issues with network discontinuities, leading to both under- and over- estimation of vessel lengths and nodes (Fig. 3E,F). A repercussion of this is a less accurate representation of network looping structures, which were well preserved by VAN-GAN (Fig. 3G). Moreover, the supervised model used for comparison here benefits from a fully annotated dataset;

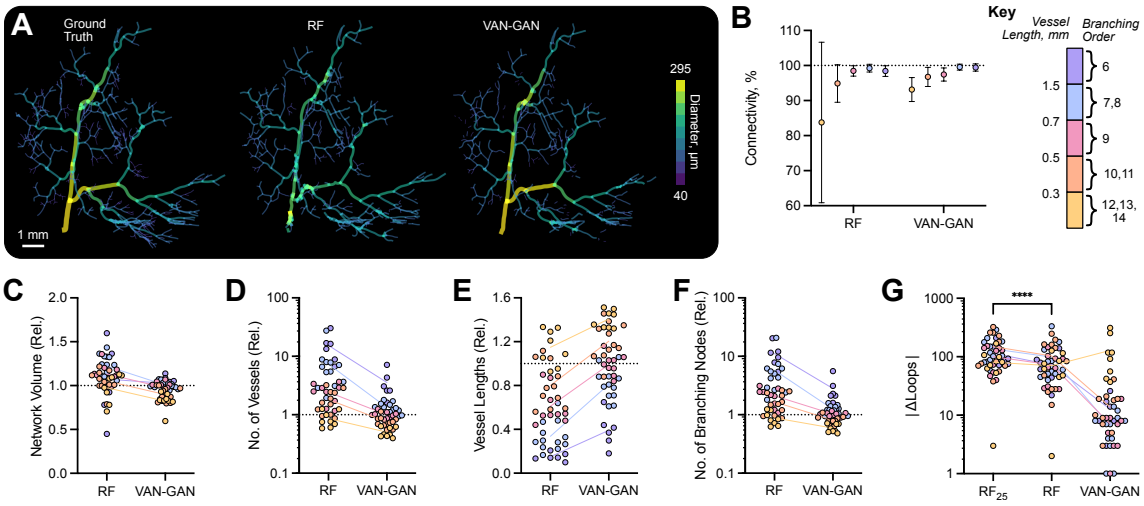

Figure 3: **Analysing vascular skeletons to assess segmentation performance against ground truth.** (A) Vascular skeletons were computed from RF and VAN-GAN segmentation masks for a direct comparison against the ground truth skeleton. (B) Mean and standard deviation of network connectivity (the volume percentage of the largest vascular subnetwork with respect to the sum of all subnetwork volumes, %) for decreasing network complexity (based on number of branching orders). Vascular descriptors normalised against known ground truth values: (C) network volume; (D) number of vessels; (E) mean vessel length; and (F) number of branching nodes. (G) The absolute difference in number of network loops compared to ground truth (normalisation was not applied since some networks have zero loops). RF_{25} indicates an RF model training with 25% of known labels. Note, data for (D), (F) and (G) is shown on a $\log_{10}$ scale. Significance asterisks excluded to due significant differences compared ground truths for all methods. See Supplementary Table 3 for absolute values. Statistical significance between RF_{25} and RF shown in (G) by **** ($P < 0.0001$).

in reality human-annotations typically generate labels on only a small subset of the available data due to time limitations. Performance of the supervised RF model degrades when only 25% of the data are annotated (Fig. 3G and Supplementary Fig. 3). VAN-GAN thus provides more realistic representations of network structure and topology.

VAN-GAN avoids user bias compared to human annotations

Having established the relative performance of VAN-GAN and a supervised segmentation method in our paired dataset, we then sought to determine the influence of the human annotator in a realistic *in vivo* scenario. We first compared two experts independently annotating blood vessels in maximum intensity projections (2D, MIPs) in the XY-plane created from 3D images of mouse ears (n=1) and skin (n=4) (Fig. 4A). As expected, there was considerable variability between the two experts

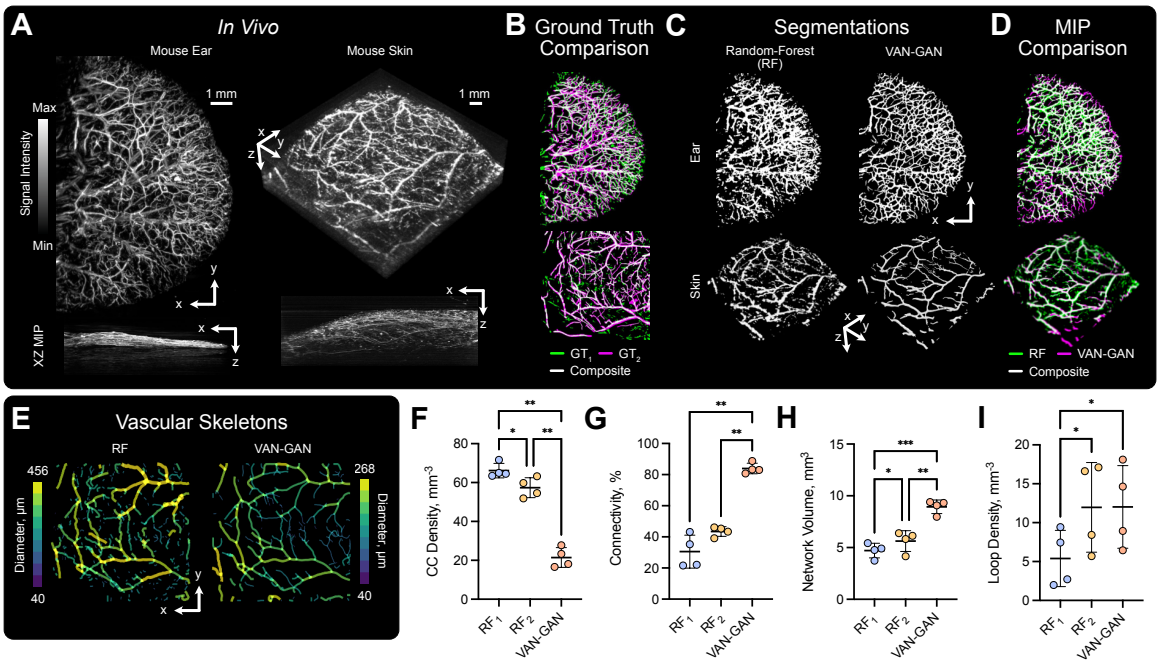

Figure 4: **Unsupervised vascular segmentation overcomes user variation in human annotations.** (A) Mouse ear and skin imaged *in vivo* using PAI. (B) Ground truths for the XY maximal intensity projections (MIPs) were generated for each dataset by two experts (indicated by subscripts *1* and *2*). Visualisations indicate variability between each expert for the ear (top) and skin (bottom). (C) 3D segmentation masks predicted using RF (left) and VAN-GAN (right) for datasets shown in (A,B). (D) Comparison of MIPs generated from the 3D segmentation labels shown in (C). (E) 2D projections of 3D vascular skeletons created from the two segmentation methods applied to mouse skin. Variability between individual users RF classifiers and VAN-GAN on the skeletonised mouse skin is indicated for a range of statistical and vascular descriptors: (F) connected component (CC) density (the number of subnetworks normalised against tissue volume, mm^{-3}); (G) connectivity (the volume of the largest subnetwork normalised against total network volume, %); (H) network volume (mm^3); and (I) loop density (the number of vessel network loops normalised against tissue volume, mm^{-3}). Mean and standard deviation of data are shown in (F-I). Statistical significance indicated by * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) and **** ($P < 0.0001$). MIP = maximal intensity projection.

(Fig. 4B) across standard metrics (F1 Score: 0.667/0.569, IoU: 0.501/0.397, Specificity: 0.916/0.892 and Sensitivity: 0.754/0.659 calculated between users' ear and skin ground truths, respectively) underscoring the difficulty of creating consistent ground truth labels.

Next, we had each expert (*1* and *2*) independently train separate 3D RF models for vessel segmentation (termed RF_1 and RF_2). These models, along with VAN-GAN, were used to predict

3D vessel segmentation labels for the mouse ear and skin datasets (performance summarised in Supplementary Table 2). MIPs of 3D predictions from each model were then used to evaluate performance against our 2D ground truth annotations. There was no indication of bias towards ground truths generated by the associated expert. For example, for the ground truth created by expert *2*, the F1 Score of RF_1 is higher than RF_2. Qualitatively, VAN-GAN appears more robust to segmentation of background noise (Fig. 4C, D), which was confirmed through improved specificity (see Supplementary Table 2). RF shows higher F1 score, IoU and sensitivity across both datasets and expert-generated ground truths. However, upon closer inspection, the results are revealed to be sub-optimal due to, for example, the segmentation of reflection artefacts. This highlights the difficulties involved in creating 3D hand-annotated labels, which are more challenging to produce than their 2D counterparts.

To analyse more comprehensively the 3D morphology of vascular networks predicted by each model, we skeletonised the 3D vessel masks (Fig. 4E) and again calculated a set of vascular descriptors. Focusing our analysis on mouse skin, we found that the number of vascular subnetworks (connected components) between RF_1 and RF_2 were significantly different ($P < 0.05$, Fig. 4F). VAN-GAN predicts more interconnected and larger vascular networks, with the largest subnetwork on average forming 84.0% of the total subnetwork volume (Fig. 4G). The computed network volume was also significantly larger compared to both RF classifiers ($P < 0.001$ and $P < 0.01$ for RF_1 and RF_2, respectively - see Fig. 4H). These results indicate that VAN-GAN segments larger and more well-connected networks compared to RF, which would align with our expectation of the biological network in healthy skin, and is consistent with observation that VAN-GAN is more robust to imaging artefacts such as network discontinuities.

These analyses underscore how using equivalent but separately trained supervised segmentation models can influence the biological outcome of a study by: 1) highlighting that supervised models are limited by the quality of training data; and 2) showing that supervised segmentation models trained independently by different users can create systematically different segmentation masks. For example, network volume exhibited significant differences between all three segmentation methods (Fig. 4H), which could have ramifications for studies that track transient modulation of blood volume. Further, RF_2 and VAN-GAN have a higher density of loops when compared to RF_1 ($P < 0.05$ in both cases; Fig. 4I). Significant differences between users of the supervised models highlights the importance of implementing unsupervised methods, such as VAN-GAN, which limit the introduction of human biases, produce consistent predictions and improve the reliability of downstream biological hypothesis testing.

VAN-GAN is more robust to artefacts compared to human annotations

Here, we specifically examine the influence of photoacoustic illumination artefacts and depth-dependent signal-to-noise ratio (SNR) on the performance of VAN-GAN and the supervised RF models. The illumination artefact in PAI (Fig. 5A) arises from the limited coverage of the imaged object with the excitation light. As a result, objects can appear flattened in reconstructed

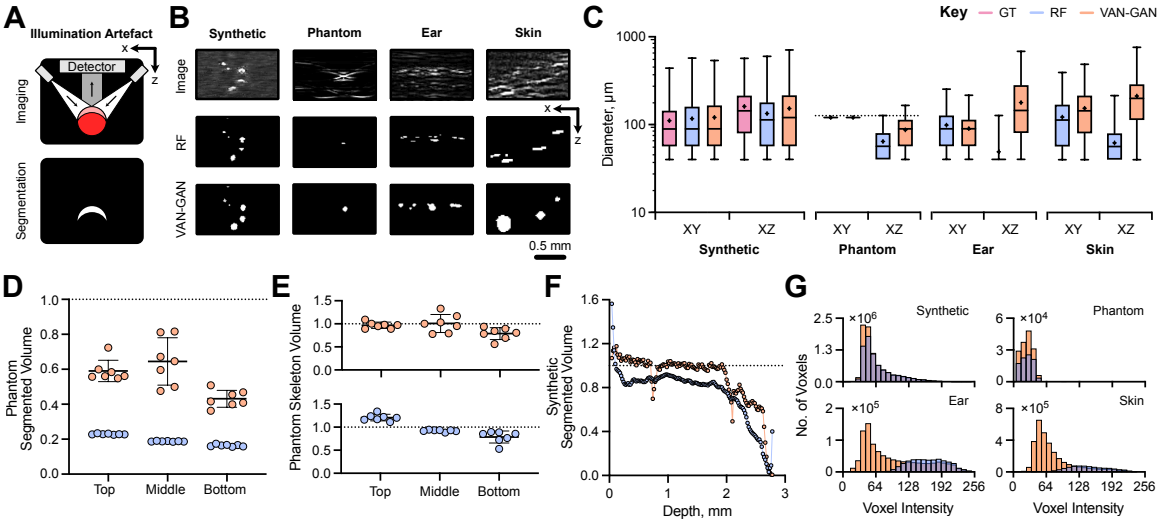

Figure 5: **VAN-GAN overcomes image artefacts in photoacoustic mesoscopy.** (A) The illumination artefact in photoacoustic mesoscopy results in the incorrect segmentation of the top surface of blood vessels due to lower signals from underlying regions. (B) Example image subsections highlighting the illumination artefact in each imaging domain dataset and corresponding random forest pixel classifier (RF) and VAN-GAN segmentations. (C) Diameters in the XY and XZ directions calculated for each dataset. For the string phantom, the known string volume is indicated by the dotted line. Computed string phantom volumes from (D) 3D segmented images and (E) skeletonised strings normalised against the ground truth value. (F) Normalised segmented volume for our synthetic dataset with respect to tissue depth. (G) Histograms of signal intensities for each voxel segmented by each method and for each dataset: synthetic vasculatures, *in situ* string phantoms and *in vivo* mouse ear and skin datasets. Data in (C) is given by box and whisker plots which display the 25 th, 50 th and 75 th quartiles, and the interquartile range. Mean and standard deviation of data are shown in (D) and (E).

images, which typically results in the brighter top surface of blood vessels being annotated by users, leaving the underside labelled as background. We can input ground truth labels into the supervised RF for our computed generated paired dataset, allowing the RF to learn the full range of feature signal intensities leading to more axisymmetric segmentation of vessels with respect to local vessel coordinates (Fig. 5B,C). This is not the case for the *in vivo* mouse ear and skin datasets where it can be challenging to distinguish the lower intensity underside of a vessel wall or vessel lumen from background noise so the hand-annotations bias the segmentations. Consequently, vessels appear as thin slices in the XZ-plane of predicted segmentation masks (Fig. 5B) and diameters in the XZ-plane are underrepresented (Fig. 5C, see Supplementary Methods for more details). In comparison, VAN-GAN accounts for the spatial range of signal intensities across the entirety of a vessel wall and lumen across both *in silico* and *in vivo* datasets (Fig. 5B,C). These observations were further validated

using photoacoustic string phantoms that consist of three non-overlapping strings of known size located at three separate depths parallel to the imaged surface[34]. Our results show that both RF and VAN-GAN can accurately predict the XY diameter of strings, however, the VAN-GAN output is closer to the ground truth when predicting diameters in the XZ-plane (Fig. 5B,C). We note that this is the case even though no string phantom images were used in training VAN-GAN whereas an RF classifier was specifically trained for this task.

The synthetic and string datasets were also used to understand the relationship between SNR and imaging depth. For the string phantom, VAN-GAN consistently provides more accurate predictions of string volumes at all depths (Fig. 5D). Predictions of network volume improved for both models following skeletonisation, which assumes axisymmetric vessels by using the largest axial diameter (Fig. 5E). Quantifying segmented network volume with respect to depth for our synthetic dataset predictions showed that VAN-GAN is consistently more accurate compared to RF, with performance relative to the ground truth declining from 2 mm (Fig. 5F). Upon analysing the signal intensity of segmented voxels, we discovered that VAN-GAN segments a higher number of low intensity voxels across all our datasets and the proportion of these voxels increases for our *in vivo* data (see Fig. 5G). This may be due to the presence of more complex background noise in these images that could impede a user's ability to perform accurate vessel labelling, or due the ability of neural networks to learn more complex feature representations compared to random forest classifiers, which typically do not have a fixed or pre-defined receptive field. Consequently, these findings indicate that VAN-GAN outperforms RF by more effectively analysing the spatial arrangement of signal structure that varies with depth and is more robust to background noise, enabling it to overcome photoacoustic artefacts that limit the performance of supervised methods.

Quantifying tumour vascular architecture using VAN-GAN

A key application of PAI is in evaluation of the longitudinal development of blood vessels in tumours[2,47]. Tumours demonstrate a high level of heterogeneity, which impacts depth-dependent SNR and can bias supervised segmentation. To that end, we apply VAN-GAN to a dataset of 3D images of oestrogen receptor positive (ER+) and negative (ER-) breast cancer patient-derived xenograft tumours[35,48], as well as tumours derived from MCF7 and MDA-MB-231 cell lines, to quantify differences in vascular descriptors.

Improvements made by VAN-GAN allowed us to distinguish structural differences between the ER+ and ER- subtypes (Fig. 6A,B and Supplementary Notes). In comparison to ER+ tumours, the ER- subtype exhibited a significant increase in surface area density with respect to tumour volume ($P < 0.05$, Fig. 6C). Our *ex vivo* immunohistochemistry (IHC) analysis of CD31 staining, an endothelial cell marker, cross-validated this finding, as the ER- subtype showed significantly greater CD31 positivity ($P < 0.05$, Fig. 6D). Additionally, ER- tumours displayed a higher density of vascular looping structures ($P < 0.05$, Fig. 6E), which could indicate a more immature vascular network compared to ER+. This finding was supported by our IHC staining of α-smooth muscle actin (αSMA), a pericyte and smooth muscle marker, colocalised with CD31, where ER+ tumours

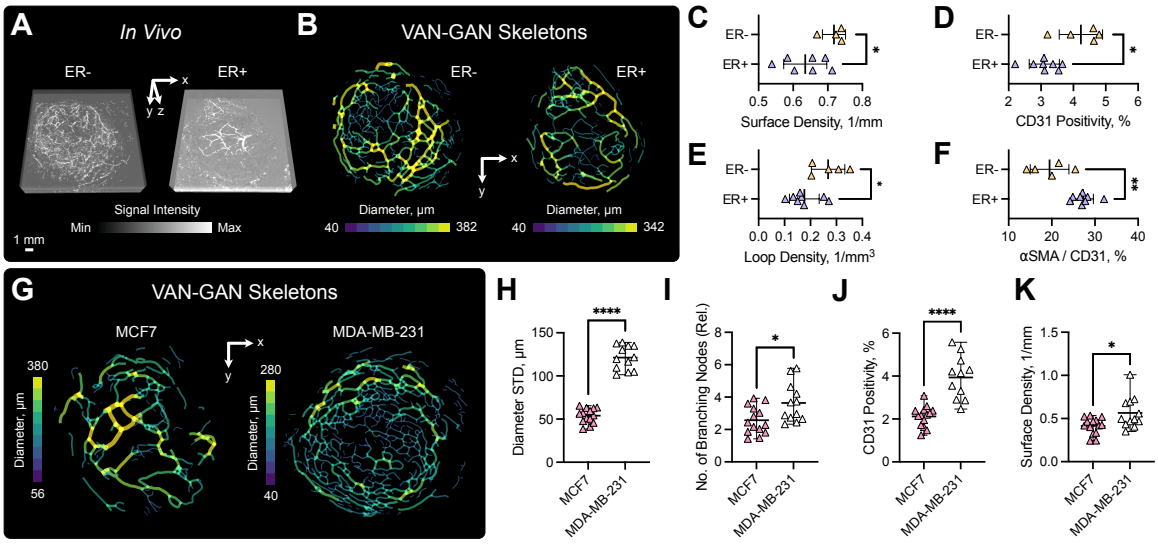

Figure 6: **VAN-GAN improves quantification of pathological vascular architecture.** (A) *In vivo* signal intensity and (B) vascular skeletons for oestrogen receptor negative (ER-, left) and positive (ER+, right) breast cancer patient-derived xenograft tumours. A comparison of metrics for ER- (top) and ER+ (bottom): (C) vessel surface density (1/mm), (D) the percentage of CD31 staining area with respect to tumour area (%), (E) vessel loop density (1/ mm^3) and (F) α-smooth muscle actin (αSMA) colocalisation with CD31 staining (%). (G) Vascular skeletons of tumours grown from MCF7 (left) and MDA-MB-231 (right) breast cancer cell lines. The vascular metrics for MCF7 and MDA-MB-231 tumours: (H) standard deviation (STD) of vessel diameters (μm) and (I) the number of branching nodes relative to the tumour volume. (J) The percentage of CD31 staining area with respect to tumour area. (K) The total surface area of blood vessels normalise against the tumour volume (surface density, 1/ mm). (D), (F) and (J) were quantified via immuno-histochemistry of tumour sections. Mean and standard deviation of data are shown in (C-F) and (H-K). Statistical significance indicated by * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) and **** ($P < 0.0001$) (ns = not significant).

showed significantly larger positivity ($P < 0.01$, Fig. 6).

VAN-GAN also enabled us to differentiate between MCF7 and MDA-MB-231 breast cancer subtypes (Fig. 6E) by comparing key vascular descriptors. Our findings indicate that vessel diameters in MDA-MB-231 tumours exhibit a significantly wider range compared to those in MCF7 tumours ($P < 0.0001$, Fig. 6F). We also showed that MDA-MB-231 tumour vasculature display significantly more branching network architecture due to increased numbers of bifurcation nodes ($P < 0.05$, Fig. 6G) which was supported by our analysis of loop density where no significance was found ($P = 0.262$). VAN-GAN predictions are supported by our immuno-histochemistry analysis where we stained 2D tumour sections for CD31 in order to quantify the percentage of vessel wall surface with respected to intra- and extra-vascular area. Here, MDA-MB-231 tumours display significantly elevated CD31

levels ($P = 0.0001$, Fig. 6H), with similar trends observed when quantifying blood vessel network surface area with respect to the tumour volume from our 3D VAN-GAN segmentations ($P < 0.05$, Fig. 6I).

Discussion

In this work, we have demonstrated the performance of an unsupervised deep learning framework, VAN-GAN (Vessel Segmentation Generative Adversarial Network), to facilitate the segmentation of vascular networks from biomedical image volumes. VAN-GAN is a novel approach that builds on the foundation of CycleGAN but is specifically used for the task of vascular segmentation. For example, we introduce a vesselness loss term that focuses on preserving tubular structures, and incorporate random noise to discriminator input and convolutional layers for improved regularisation and training stability (as a consequence of class imbalance). Our approach is conceptually distinct from past unsupervised models that necessitate pre-processing or contouring of images[28,29]. Instead, our method incorporates synthetic vascular architectures that closely resemble real-life anatomy directly into the training process and learns to replicate the underlying physics of an imaging system by learning to map 3D images between the imaging and segmentation domains.

Standard supervised segmentation models are limited by the ability of the human operator to make ground truth vessel annotations. PAI-specific artefacts make this process challenging due to distortions in the vessel structures and a high level of depth-dependent background noise. These are, for example, introduced by the light delivery, limited detected area or image reconstruction algorithms. We highlight significant differences between network structures annotated by different experts, even in 2D projections, and showed that these biases propagate through to training of a popular supervised RF classifier. In contrast, VAN-GAN directly encodes the information required to segment PAI images in an unbiased manner, limiting the human interaction that led to contrasting statistical and topological output metrics derived from supervised RF models trained separately by two expert users.

Our results show that training based on physics-driven simulations applied to synthetic 3D vascular networks generated stochastically from mathematical formalisms imparts robustness to the segmentation process. Quantifying the accuracy of VAN-GAN using the paired dataset showed that it outperformed the RF supervised segmentation method, preserving important topological metrics. By evaluating segmentation masks predicted across *in silico*, *in situ* and *in vivo* datasets, we further showed that VAN-GAN was able to overcome challenges introduced by intrinsic artefacts of the imaging physics and system engineering. In particular, VAN-GAN predicts accurate vascular structures despite the presence of illumination artefacts and also addresses limitations associated with depth-dependent SNR.

Having established the performance characteristics of VAN-GAN, we apply it as a tool to segment a dataset of breast cancer patient-derived xenograft tumours and tumours derived from MCF7 and MDA-MB-231 cell lines. These data confirm the consistent trend in extracting well-connected

vessel structures from which it is possible to quantify distinct topological features between these pathological vascular architectures. VAN-GAN thus offers a time-efficient and accurate means for life scientists to segment volumetric images of blood vessels without the need for the error-prone and time-consuming task of creating hand-annotated feature labels. By utilising image-to-image translation and synthetically-generated vascular architectures, our tool enables more consistent and robust analyses of vascular morphology from 3D images and facilitates reliable comparisons across multiple users' data through standardising segmentation. This further highlights the potential for more robust longitudinal quantitative analyses of biomedical imaging data where VAN-GAN mitigates the potential confounding effect of vascular delineation on images acquired at multiple time points. This is essential especially for the long-term monitoring of cancer therapeutics effectiveness, and experimental repeatability, of drug delivery or radiation therapy, for instance where the tumour vasculature play an essential role[49].

Future development of VAN-GAN could explore optimising training schemes and loss function weightings. For example, a semi-supervised model that integrates paired synthetic data could be created for PAI to enhance segmentation performance on real images. Although this work focused on the use of PAI data, where haemoglobin acts as an endogenous vascular label, we envision that the proposed approach could be extended across a range of biomedical imaging modalities. For example, VAN-GAN is likely to be directly applicable to fluorescently labelled vasculature, such as in light sheet or two-photon laser scanning microscopy, where background has similar noise levels as in PAI. It could also be tested across 2D images, such as the DRIVE[50] and STARE[51] datasets where retinal images are paired with ground truth segmentation masks. Application to other methods where vascular features appear on a visible tissue background, such as in X-ray histology, may require preprocessing or modifications to the cycle reconstruction loss functions. Accessibility of modern image analysis techniques to users continues to improve due to the development of several open-source bioimage platforms[44,52–60], which allow life scientists to intuitively apply advanced algorithms to their own data. Following testing across other modalities, future integration of VAN-GAN with such platforms would be valuable for widespread uptake.

In conclusion, our work has demonstrated the potential of VAN-GAN as a valuable tool to advance the segmentation standard of 3D microvascular networks in biomedical images, using examples from photoacoustic mesoscopy. By reducing the reliance on manual labelling and leveraging synthetic data, our approach lowers the barriers to entry for high-quality blood vessel segmentation, leading to more robust and consistent characterisation of vascular structures. This could pave the way for the discovery of novel therapeutics and improve diagnostic accuracy across innumerable clinical applications.

Methods

The following details architecture and training methodology of VAN-GAN, in addition to describing image synthesis and preprocessing. VAN-GAN was implemented using Tensorflow[61] with Keras

backend[62] and Tensorflow Addons, along with Tensorflow-MRI[63]. The model was trained on either: 1) a Dell Precision 7920T with a Dual Intel Xeon Gold 5120 CPU with 128GB RAM and two NVIDIA Quadro GV100 32GB GPUs with NVLink; 2) a custom built workstation with a Intel Xeon Gold 5220 CPU with 128GB RAM and four NVIDIA RTX A6000 48GB GPUs with NVLinks. The optimal model for each dataset was selected for application based on qualitative image evaluation of generated images from the test set and a comprehensive analysis of minima for each loss function (see Supplementary Methods and Supplementary Figs. 5-7).

3D Deep Residual U-Net Generators. Our generators use a modified version of a deep residual U-Net architecture[37], which integrates residual units into a U-Net architecture to ease training and facilitate information propagation through the network without degradation. For the latter, it is important to use low-level details and retain high-resolution semantic information to improve segmentation performance[11,37,64]. Both the generator input and output tensor shapes are $128 \times 128 \times 128 \times 1$ ($depth \times width \times height \times channel$).

3D PatchGAN Discriminators. The discriminators use a five layer PatchGAN architecture[65]. Each layer is composed of a 3D convolutional layer, instance normalisation[41], leaky ReLU[66] and spatial dropout[67] (a rate of 20% and excluded from first and final layers). Similarly to our generators, reflection padding was also used prior to convolution layers to reduce feature map artefacts[43]. Further, for additional regularisation and to limit unstable behaviour of VAN-GAN during training, we added random Gaussian noise to real or fake inputs to the discriminator[68,69] and for every proceeding layer prior to convolution blocks[69] (see Supplementary Table 4).

Loss Functions. Our goal is to learn the mapping functions G and F between the source and targets domains, X and Y. Here, X and Y represent the 3D domains of real biomedical images and segmentation labels of synthetic vasculature, respectively. In CycleGAN, the mapping spaces $G : X \rightarrow Y$ and $F : Y \rightarrow X$ are constrained by introducing a cycle-consistency loss[20], which ensures forward and backward cycle-consistency i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. Here, we modify the vanilla implementation of cycle-consistency loss to:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F, x, y) = {} & \mathbb{E}_{x \sim p_{data}(x)} \left[||F(G(x)) - x||_1 \right] \\ & + \mathbb{E}_{y \sim p_{data}(y)} \left\{ \sum_{i \in I} -y_i \log \left[G(F(y_i)) \right] - (1 - y_i) \log \left[1 - G(F(y_i)) \right] \right\},\end{aligned} \tag{1}$$

where L_1-norm is used for forward cycle-consistency and binary cross-entropy is used to compare the real and reconstructed segmentation labels for backward cycle-consistency.

To mitigate problems with vanishing gradients, the adversarial loss is expressed via least-squares adversarial loss[70]. In the case of the mapping $G : X \rightarrow Y$, this is given by

$$\mathcal{L}_{GAN}(F, D_X, x, y) = \frac{1}{2} \mathbb{E}_{y \sim p_{data}(y)} \left[\left(D_X(x + \epsilon_Y) - 1 \right)^2 \right] + \frac{1}{2} \mathbb{E}_{x \sim p_{data}(x)} \left[\left(D_X \left(F(y) + \epsilon_X \right) \right)^2 \right], \tag{2}$$

where ϵ_X and ϵ_Y are the randomly sampled Gaussian noise[69]. Here the discriminator D_X aims to discern translated images, $F(y)$, from real images y by minimising the objective function, whereas F aims to maximise it against its adversarial rival. Adversarial loss is similarly used for the mapping $G : X \to Y$. Following Ihle et al.[22], we do not impose an identity mapping loss as this constrains the tint of an image[20] and so is not required for our greyscale and binary image volumes.

In addition to $\mathcal{L}_{cyc}$, we impose two additional constraints on cycle-consistency: 1) a structural similarity index measure (SSIM) loss; and 2) a topology-preserving loss (centrelineDice or clDice[39]) for backward cycle-consistency. SSIM is used for forward-consistency to ensure structural and perceptive features in biomedical images are retained when generating fake images. Similarly, we are interested to preserve the morphological characteristics of vascular networks when segmenting blood vessels, and so we apply a constraint on backward-consistency that seeks to minimise differences in network structure and topology.

SSIM is a perceptually motivated model composed of three comparison functions: luminance, contrast and structure. Typically SSIM is used to assess image quality, so as a loss function can be used for image restoration[71], such as denoising and super-resolution[72,73] and pose-guided person image generation[74]. Here, SSIM adopts a sliding Gaussian window to calculate the SSIM index between two local patches centred around a pixel coordinate. For image patches I_p and J_p, SSIM is defined as

$$\mathbf{SSIM}(I_p, J_p) = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{(\mu_I^2 + \mu_J^2 + C_1)(\sigma_I^2 + \sigma_J^2 + C_2)}. \tag{3}$$

Here, μ_I, μ_J, σ_I^2 and σ_J^2 are the mean and variance of I_p and J_p, respectively, σ_{IJ} is the covariance of I_p and J_p, and $C_1 = (0.01 \cdot L)^2$ and $C_2 = (0.03 \cdot L)^2$, where L is the dynamic range of the pixel-values. To maximise the SSIM of biomedical images we form the reconstruction loss:

$$\mathcal{L}_{rec}(G, F, x, y) = 1 - \mathbb{E}_{x\sim p_{data}(x)}\mathbf{SSIM}\Big(x_p, F(G(x))_p\Big), \tag{4}$$

where the subscript p indicates image patches.

Cycle-consistency alone does not provide sufficient spatial constraint on the network topology of segmented images[39]. Consequently, we added a spatial and topological constraint on backward-consistency to act as additional regulatory loss function term. Here, the segmentation labels of synthetically-generated 3D vascular networks, y, are compared to $G(F(y))$ to ensure differences in topology are minimised. This is achieved using the connectivity-preserving metric **clDice**, which enforces topology preservation up to homotopy equivalence for binary segmentation[39]. Following Shit et al.[39], the loss function is a combination of soft-Dice loss and the differentiable form of **clDice**, $soft$**clDice**:

$$\begin{aligned}\mathcal{L}_{topo}(G, F, y) = \mathbb{E}_{y\sim p_{data}(y)} \Big\{ &(1-\alpha)\Big[1 - soft\mathbf{Dice}\Big(y, G(F(y))\Big)\Big] \\ &+\alpha\Big[1 - soft\mathbf{clDice}\Big(y, G(F(y))\Big)\Big]\Big\},\end{aligned} \tag{5}$$

where the weighting, α, is set to 0.5.

Thus, the objective function of VAN-GAN is given by

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, x, y) + \mathcal{L}_{GAN}(F, D_X, x, y) \\ + \lambda\mathcal{L}_{cyc}(G, F, x, y) + \eta\mathcal{L}_{rec}(G, F, x) + \mu\mathcal{L}_{topo}(G, F, y),\end{aligned} \tag{6}$$

where the hyperparameters λ, η and μ control the relative importance between the objectives are set to 10, 5 and 5, respectively. We aim to solve:

$$G^*, F^* = \arg\min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y). \tag{7}$$

Synthetic Vasculature. We utilise a language-theoretic model, called a Lindenmayer system (L-System)[75], to generate 3D branching vascular networks (termed here as *V-System*). This approach is ideally suited to our segmentation task as these systems have been shown to create realistic, computer-generated 3D vascular branching structures[76,77] quickly and at scale[35] ($O(10^2)$ networks in $O(1)$ minutes). To generate a synthetic branching network, we use a new stochastic grammar to create a string, which defines the complexity (in our case, the number of branching orders) of the vascular network, for example, branching order and angle, vessel diameter and tortuosity, and aneurysms or branching vessel shrinkage (see Supplementary Methods for mathematical descriptions). These strings are translated to graph form using a lexical and syntactic analyser and subsequently converted into volumetric binary segmentation masks[35], forming a synthetic dataset of 459 images for network training.

Photoacoustic Simulations. To generate a paired image dataset, we performed photoacoustic simulations on the segmentation volume of each image in our synthetic vascular dataset. Each image pair consists of a physics-driven image volume and its corresponding known segmentation labels. Simulations followed the method of Brown et al.[35] who used SIMPA[78] (v0.1.1 with MCX v2020, 1.8) with the k-Wave MATLAB toolbox[79] (v1.3, MATLAB v2020b, MathWorks, Natick, MA, USA) to predict photoacoustic signals across synthetic vasculatures under the assumption that they are embedded in muscular tissue. In brief, vascular planar (XY) illumination was achieved on an isotropic resolution with optical forward modelling assuming an absorption spectrum of 50% oxygenated haemoglobin in blood vessels to mimic tumours[80]. 3D acoustic forward modelling was then performed with the signal detected by a planar array of sensors positioned at the tissue surface, mimicking our PAI instrument (see below). The resulting photoacoustic wave-field was then reconstructed using a fast Fourier transform[79]. While the PAI instrument raster-scans, we approximated this process with a planar illumination due to computational restrictions (reducing simulation time by a factor of 600^2).

Experimental Imaging. PAI was performed using a commercial system (Raster-scan optoacoustic mesoscopy RSOM Explorer P50, iThera Medical GmbH), as described previously[35]. Briefly, string

phantoms were composed in agar mixed with intralipid (both Merck, UK) to mimic tissue-like scattering with red-coloured synthetic fibres (Smilco, USA) embedded at three different depths. PAI data were acquired at 100% laser energy with a 2kHz repetition rate. All animal procedures were conducted in accordance with project (PE12C2B96) and personal licenses (I544913B4, IA70F0365) issued under the United Kingdom Animals (Scientific Procedures) Act, 1986 and approved locally by the Cancer Research UK Cambridge Institute Animal Welfare Ethical Review Board.

To generate *in vivo* vascular tumour models, breast PDX tumour fragments were cryopreserved in a freezing media consisting of heat-inactivated foetal bovine serum (10500064, Gibco^{TM}, Fisher Scientific, Göteborg Sweden) and 10% dimethyl sulfoxide (D2650, Merck). The fragments were then defrosted at 37°C, washed with Dulbecco's Modified Eagle Medium (41965039, Gibco^{TM}), mixed with matrigel (354262, Corning, NY, USA), and surgically implanted into the flank of 6–9 week-old NOD scid gamma (NSG) mice (#005557, Jax Stock, Charles River, UK), following standard protocols[35,48]. The implantation involved one estrogen receptor negative (ER-, n=6) PDX model and one estrogen receptor positive (ER+, n=8) PDX model. After the tumours had reached an average diameter of ~ 1 cm, the mice were imaged and then sacrificed, with the tumours collected in formalin for IHC analysis.

For the remaining breast cancer cell lines, seven-week old immuno-deficient female nude (BALB/c nu/nu) mice (Charles River) were inoculated orthotopically in the mammary fat pad of both flanks $1 \cdot 100^6$ cells (either MCF7, n=7, or MDA-MB-231, n=6, random group assignment) in a final volume of 100 μL of 1:1 phosphate-buffered saline (PBS, Gibco) and matrigel (BD). For MCF7, oestrogen implants (E2-M - 127 β-estradiol 90 days release, daily dose: 41.2-105.6 pg/ml, Belma Technologies) were implanted subcutanaously in the scruff of the neck 3 days before tumour cell injection.

For animal imaging, were mice anaesthetised using 3-5% isoflurane in 50% oxygen and 50% medical air. Mice were shaved and depilatory cream applied to remove hair that could generate image artefacts; single mice were placed into the PAI system, on a heat-pad maintained at 37°C. Respiratory rate was maintained between 70-80 bpm using isoflurane ($\sim 1 - 2\%$ concentration) throughout image acquisition. PAI data were acquired at 80% laser energy at 1kHz.

For string phantom imaging, phantoms were prepared following standard protocols[81] using agar mixed with intralipid (both Merck, UK) to mimic tissue-like scattering with red-coloured synthetic fibres (Smilco, USA) embedded at three different depths (top: 0.5 mm, middle: 1 mm, bottom: 2 mm).

Immunohistochemistry. The tumour tissues, obtained for *ex vivo* validation, were processed by sectioning formalin-fixed paraffin-embedded (FFPE) samples. After deparaffinisation and rehydration, IHC analysis was performed on the tissues using the following antibodies: CD31 (anti-mouse 77699, Cell Signalling, London, UK), αSMA (anti-mouse ab5694, Abcam, Cambridge, UK), and carbonic anhydrase-IX (CAIX) (anti-human AB1001, Bioscience Slovakia, Bratislava, Slovakia), at concentrations of 1:100, 1:500, and 1:1000, respectively. The analysis was carried out using a BOND automated stainer, with a bond polymer refine detection kit (Leica Biosystems, Milton Keynes,

UK), and 3,3'-diaminobenzadine as a substrate. The stained FFPE sections were scanned at a magnification of 20x using an Aperio ScanScope (Leica Biosystems, Milton Keynes, UK) and analysed with either ImageScope software or HALO Software (v2.2.1870, Indica Labs, Albuquerque, NM, USA). Regions of interest (ROIs) were drawn over the entire viable tumour area, and the built-in algorithms were customised to analyse the following: CD31 positive area (μm^2) normalised to the ROI area (μm^2) (reported as CD31 positivity (%)), area of CD31 positive pixels (μm^2) colocalised on adjacent serial section with αSMA positive pixels/CD31 positive area (μm^2).

Datasets and Preprocessing. Image volume datasets are split into two categories: synthetic or experimental datasets (see Supplementary Table 5 for summary). Our synthetic datasets are comprised of binary labels of 3D mathematically-generated vascular networks, which are paired with computer-simulated photoacoustic image volumes (n=449 for each). Here, physics-driven predictions are performed on the spatial architecture of the synthetic vasculature provided by the V-System in each binary image volume. The experimental datasets consist of a string phantom (n=7), mouse ears (n=32), mouse skin (n=41) and breast cancer patient-derived xenograft (PDX) tumours in mice (n=445), and MCF7 and MDA-MB-231 tumours derived from bread cancer cell lines (n=204), all imaged *in vivo*. These datasets represent the *imaging domain* in VAN-GAN where n indicates the number of image volumes used. Note, the string phantoms were not used for training.

All photoacoustic datasets were stored as 32-bit grayscale $600 \times 600 \times 140$ μm^3 (real) and $512 \times 512 \times 140$ μm^3 (simulated) voxel tiff stacks with an isotropic voxel size of $20 \times 20 \times 20$ μm^3 in the X-, Y- and Z-directions, where the Z-axis is perpendicular to the surface. All synthetic images are stored in an 8-bit format and generated with dimensions $512 \times 512 \times 140$ μm^3 and an equal isotropic voxel size. As VAN-GAN trains on image subvolumes of size $128 \times 128 \times 128 \times 1$ voxels, all datasets were downsampled to 128 voxels in the Z-axis using a combination of maximum and bicubic downsampling, to ensure that depth-dependent SNR information is retained for training. All real and simulated photoacoustic images were normalised by performing XY slice-wise Z-score normalisation followed by thresholding of the top and bottom 0.05% of pixel intensities to correct for uneven illumination with respect to depth. Finally, datasets were normalised to a pixel intensity range of $[-1, 1]$. Datasets were partitioned with 10% assigned for testing and with the remaining 90% split 80/20 between training/validation.

Network Training. Image patches were randomly sampled with a global batch size of two. Due to the sparsity of vessels with respect to the background for all datasets, a mapping function was applied when images were retrieved from the synthetic segmentation dataset. Here, the function detected whether a sampled volume contained any vessels to ensure VAN-GAN learnt how to segment vessels rather than just background. For 90% of sampled volumes, if no vessel was detected, a new image volume was sampled. In addition, to artificially extend the size of our datasets, all images were augmented via a rotation about the Z-axis randomly sampled from the set $\{0, \pi/2, \pi, 3\pi/2, 2\pi\}$.

All convolutional kernels were initialised using a He-normal initialiser and our loss functions

hyperparameters were set to default values[20,39] $\alpha = 0.5$, $\lambda = 10$ and $\eta = \mu = 5$. Following Zhu et al.[20], training was performed for 200 epochs using the Adam optimizer[82] with a learning rate of 2×10^{-4} and 1st and 2nd moment estimates of 0.5 and 0.9, for both generators and discriminators. Each generator and discriminator was given a linear learning rate decay to 0 from the 100th epoch.

Training was stabilised in two ways. Firstly, noise was applied to the real and fake input images to the discriminators[69]. Noise was sampled from a Gaussian distribution with a mean and variance, σ, of 0.0 and 0.8^i, where i was the ith epoch, and so σ is annealed during training. Secondly, to avoid exploding gradients, a gradient clip normalisation strategy[83] where all gradients of each weight was individually clipped so that its norm is no higher than 100.

Postprocessing. To reconstruct whole image volumes from generator output a sliding-window approach was employed[84]. In summary, a sliding-window of size $128 \times 128 \times 128$ voxels was strided across each inputted image with a stride length of 25 voxels in each XY direction. Output intensities were summed and the mean value calculated for each voxel location by tracking the number of times the window passed across a given voxel. To reduce edge artefacts, symmetric padding was used to ensure the sliding-window passed over each voxel for an equal number of instances. For segmenting an image, this reconstruction method results in a 32-bit 3D greyscale image where intensities indicate a probability that a voxel is a blood vessel. Consequently, following bicubic upsampling to an isotropic voxel size (140 voxels in the Z-axis), images were then thresholded based on the histogram of voxel intensities to binarise the image.

Supervised Random Forest (RF) Pixel Classifier. Using the open source ilastik tool, users can pre-define a set of possible image features and manually annotate a subset of voxels in training images (either 2D or 3D images) for a random forest model to learn to recognise complex image features. The trained model can then be applied to the remaining unlabelled data and its accuracy improved through retrospective interactive training. Photoacoustic images were preprocessed for application to ilastik's random forest toolbox by applying a slice-wise median filter with a kernel size of 3×3 to remove noise and improve contrast. For each dataset, a sample of images were manually annotated with two label categories: blood vessel and tissue background. Here, approximately 10 XY images were labelled from each training image volume. All 3D colour / intensity, edge and texture features were selected for all Gaussian mean variational scales (0.3 to 10.0) to extract fine details at a small scale as well features from larger voxel neighbourhoods. RF models were trained by authors PWS, LH, TLL and ELB. For synthetic data, all vessel labels were provided for training along with a proportion of surrounding background labels (detailed in Brown et al.[35]). Trained models were applied to unlabelled images with predictions thresholded using default settings to create binary segmentation masks.

Evaluation Metrics. Common machine learning metrics do not provide a complete picture of image segmentation performance for tubular-like structures[45]. To evaluate our results, we calculated both

standard segmentation metrics and a set of vascular descriptors[35,46] to provide a deeper insight into how well network morphology is predicted. The standard segmentation metrics used comprised of: F1 Score $= 2 \cdot (precision \cdot sensitivity)/(precision + sensitivity)$; Intersection over Union (IoU) $= TP/(TP + FP + FN)$; Sensitivity $= TP/(TP + FN)$ and Specificity $= TN/(TN + FP)$, where TP = true positive, TN = true negative, FP = false positive and FN = false negative. To calculate vascular descriptors, all segmentations were skeletonised using the open-source package Russ-learn[85,86]. This allowed us to perform structural and topological data analyses on the vascular skeletons[46,87]. The metrics were use are: number of vessels and branching nodes, vessel mean and standard deviation of diameters and lengths, network volume, surface density (the surface area of the vascular network normalised against the tissue volume), whereas topological descriptors consisted of connected components (or subnetworks, Betti-0) and looping structures (Betti-1) and network connectivity (the volume of the largest vascular subnetwork normalised against total network volume).

Statistical Analysis. Statistical analyses were conducted using Prism (v9, GraphPad Software, San Diego, CA, USA). Comparisons of F1 Scores and vascular descriptor between synthetic segmentations were computed using paired, non-parametric t-tests (Wilcoxon tests). Comparisons of vascular descriptors for *in vivo* ear datasets were performed using either paired parametric or non-parametric t-tests depending on data satisfying normality. Comparison of vascular descriptors for *in vivo* tumour datasets were performed using Wilcoxon tests, with comparisons between ER- and ER+ or MCF7 and MDA-MB-231 tumour types made using unpaired non-parametric t-tests (Mann-Whitney tests). Statistical outliers were identified by five non-parametric tests: 1) Tukey's fences; 2) Median Absolute Deviation (MAD); 3) Modified Z-Score; 4) percentiles (5th and 95th percentile cuttoffs) and 5) Hampel identifier. All P-values < 0.05 were considered statistically significant.

Code Availability

All our software are open-source and available in Github repositories. VAN-GAN (2023, Version 1.0) [Computer software - https://github.com/psweens/VAN-GAN]. V-System (2022, Version 2.0) [Computer Software - https://github.com/psweens/V-System]. Vascular Topological Data Analysis (2022, Version 2.0) [Computer Software - https://github.com/psweens/Vascular-TDA].

Data Availability

Scientific data supporting the findings of this study will be made available upon publication via the University of Cambridge Research Data Repository at: https://doi.org/10.17863/CAM.96379.

Acknowledgements

PWS, TLL, LH, ELM, and SEB acknowledge the support from Cancer Research UK under grant numbers C14303/A17197, C9545/A29580, C47594/A16267, C197/A16465, C47594/A29448, and Cancer Research UK RadNet Cambridge under the grant number C17918/A28870. PWS acknowledges the support of the Wellcome Trust and University of Cambridge through an Interdisciplinary Fellowship under grant number 204845/Z/16/Z. TLL is supported by the Cambridge Trust. LH was funded from NPL's MedAccel programme financed by the Department of Business, Energy and Industrial Strategy's Industrial Strategy Challenge Fund. JG acknowledges funding from the Walter Benjamin Stipendium of the Deutsche Forschungsgemeinschaft. We thank the Cancer Research UK Cambridge Institute Biological Resources Unit and Imaging Core for their support in conducting this research. For the purpose of open access, the corresponding author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Competing Interests

The authors have no conflict of interest related to the present manuscript to disclose.

Author Contributions

Conceptualisation: PWS, SEB
Methodology: PWS, JG
Software: PWS
Validation: PWS, SEB
Formal Analysis: PWS
Investigation: PWS, LH
Resources: SEB
Data Collection: PWS, LH, TLL, ELB
Data Curation: PWS
Writing - original draft: PWS
Writing - review & editing: PWS, LH, TLL, ELB, JG, SEB
Visualisation: PWS
Project Administration: PWS, SEB
Funding Acquisition: PWS, SEB

References

[1] Murad Omar, Juan Aguirre, and Vasilis Ntziachristos. Optoacoustic mesoscopy for biomedicine. *Nature Biomedical Engineering*, 3(5):354–370, 5 2019. ISSN 2157846X. doi: 10.1038/s41551-019-0377-4.

[2] Katja Haedicke, Lilach Agemy, Murad Omar, Andrei Berezhnoi, Sheryl Roberts, Camila Longo-Machado, Magdalena Skubal, Karan Nagar, Hsiao Ting Hsu, Kwanghee Kim, Thomas Reiner, Jonathan Coleman, Vasilis Ntziachristos, Avigdor Scherz, and Jan Grimm. High-resolution optoacoustic imaging of tissue responses to vascular-targeted therapies. *Nature Biomedical Engineering*, 4(3):286–297, 3 2020. ISSN 2157846X. doi: 10.1038/s41551-020-0527-8. URL `https://pubmed.ncbi.nlm.nih.gov/32165736/`.

[3] Colin A. Goebel, Emma Brown, Fabian B. Fahlbusch, Alexandra L. Wagner, Adrian Buehler, Thomas Raupach, Martin Hohmann, Moritz Späth, Neal Burton, Joachim Woelfle, Michael Schmidt, Andrea Hartner, Adrian P. Regensburger, and Ferdinand Knieling. High-resolution label-free mapping of murine kidney vasculature by raster-scanning optoacoustic mesoscopy: an ex vivo study. *Molecular and Cellular Pediatrics*, 9(1), 12 2022. ISSN 21947791. doi: 10.1186/s40348-022-00144-0. URL `https://pubmed.ncbi.nlm.nih.gov/35788444/`.

[4] Claudia Errico, Juliette Pierre, Sophie Pezet, Yann Desailly, Zsolt Lenkei, Olivier Couture, and Mickael Tanter. Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature*, 527(7579):499–502, 11 2015. ISSN 0028-0836.

[5] Charlie Demené, Justine Robin, Alexandre Dizeux, Baptiste Heiles, Mathieu Pernot, Mickael Tanter, and Fabienne Perren. Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients. *Nature Biomedical Engineering*, 5 (3):219–228, 3 2021. ISSN 2157846X. doi: 10.1038/s41551-021-00697-x. URL `https://www.nature.com/articles/s41551-021-00697-x`.

[6] C. L. Walsh, P. Tafforeau, W. L. Wagner, D. J. Jafree, A. Bellier, C. Werlein, M. P. Kühnel, E. Boller, S. Walker-Samuel, J. L. Robertus, D. A. Long, J. Jacob, S. Marussi, E. Brown, N. Holroyd, D. D. Jonigk, M. Ackermann, and P. D. Lee. Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nature Methods*, 18(12):1532–1541, 11 2021. ISSN 15487105. doi: 10.1038/s41592-021-01317-x. URL `https://www.nature.com/articles/s41592-021-01317-x`.

[7] J. Brunet, C. L. Walsh, W. L. Wagner, A. Bellier, C. Werlein, S. Marussi, D. D. Jonigk, S. E. Verleden, M. Ackermann, Peter D. Lee, and Paul Tafforeau. Preparation of large biological samples for high-resolution, hierarchical, synchrotron phase-contrast tomography with multimodal imaging compatibility. *Nature Protocols*, pages 1–27, 3 2023. ISSN 17502799. doi: 10.1038/s41596-023-00804-z. URL `https://www.nature.com/articles/s41596-023-00804-z`.

[8] Christoph Kirst, Sophie Skriabine, Alba Vieites-Prado, Nicolas Michalski, Marc Tessier-Lavigne, and Nicolas Renier Correspondence. Mapping the Fine-Scale Organization and Plasticity of the Brain Vasculature. *Cell*, 180(4):780–795, 2020. doi: 10.1016/j.cell.2020.01.028.

[9] Claudia Prahst, Parham Ashrafzadeh, Thomas Mead, Ana Figueiredo, Karen Chang, Douglas Richardson, Lakshmi Venkatraman, Mark Richards, Ana Martins Russo, Kyle Harrington, Marie Ouarne, Andreia Pena, Dong Feng Chen, Lena Claesson-Welsh, Kin Sang Cho, Claudio Franco, and Katie Bentley. Mouse retinal cell behaviour in space and time using light sheet fluorescence microscopy. *eLife*, 9, 2 2020. ISSN 2050084X. doi: 10.7554/eLife.49779.

[10] Erik Meijering. A bird's-eye view of deep learning in bioimage analysis, 1 2020. ISSN 20010370.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 2015. ISBN 9783319245737. doi: 10.1007/978-3-319-24574-4_28.

[12] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016.*, volume 9901. Springer, 2016. ISBN 9783319467221. doi: 10.1007/978-3-319-46723-8_49.

[13] Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings - 2016 4th International Conference on 3D Vision*, pages 565–571. Institute of Electrical and Electronics Engineers Inc., 6 2016. ISBN 9781509054077. doi: 10.1109/3DV.2016.79. URL `https://arxiv.org/abs/1606.04797v1`.

[14] Petteri Teikari, Marc Santos, Charissa Poon, and Kullervo Hynynen. Deep Learning Convolutional Networks for Multiphoton Microscopy Vasculature Segmentation. *arXiv*, 6 2016. doi: 10.48550/arxiv.1606.02382. URL `https://arxiv.org/abs/1606.02382v1` `http://arxiv.org/abs/1606.02382`.

[15] Mohammad Haft-Javaherian, Linjing Fang, Victorine Muse, Chris B. Schaffer, Nozomi Nishimura, and Mert R. Sabuncu. Deep convolutional neural networks for segmenting 3D in vivo multiphoton images of vasculature in Alzheimer disease mouse models. *PLoS ONE*, 14(3):e0213539, 3 2019. ISSN 19326203. doi: 10.1371/journal.pone.0213539. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213539`.

[16] Mohammad Haft-Javaherian, Martin Villiger, Chris B Schaffer, Nozomi Nishimura, Polina Golland, and Brett E Bouma. A topological encoding convolutional neural network for segmentation of 3D multiphoton images of brain vasculature using persistent homology. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2020-June, pages 4262–4271, 2020. ISBN 9781728193601. doi: 10.1109/CVPRW50498.2020.00503. URL `https://github.com/mhaft/DeepVess`.

[17] Angela D'Esposito, Paul W. Sweeney, Morium Ali, Magdy Saleh, Rajiv Ramasawmy, Thomas A. Roberts, Giulia Agliardi, Adrien Desjardins, Mark F. Lythgoe, R. Barbara Pedley, Rebecca Shipley, and Simon Walker-Samuel. Computational fluid dynamics with imaging of cleared tissue and of in vivo perfusion predicts drug uptake and treatment responses in tumours. *Nature Biomedical Engineering*, 2(10):773–787, 10 2018. ISSN 2157-846X. doi: 10.1038/s41551-018-0306-y. URL `http://www.nature.com/articles/s41551-018-0306-y`.

[18] A A Folarin, M A Konerding, J Timonen, S Nagl, and R B Pedley. Three-dimensional analysis of tumour vascular corrosion casts using stereoimaging and micro-computed tomography. *Microvascular research*, 80(1):89–98, 2010. ISSN 1095-9319. doi: 10.1016/j.mvr.2010.03.007.

[19] Yingxue Pang, Jianxin Lin, Tao Qin, and Zhibo Chen. Image-to-Image Translation: Methods and Applications. *IEEE Transactions on Multimedia*, 2021. ISSN 19410077. doi: 10.1109/TMM.2021.3109419.

[20] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-Octob, pages 2242–2251, 2017. ISBN 9781538610329. doi: 10.1109/ICCV.2017.244.

[21] Arnab Kumar Mondal, Aniket Agarwal, Jose Dolz, and Christian Desrosiers. Revisiting CycleGAN for semi-supervised segmentation. *arXiv*, 2019. URL `http://arxiv.org/abs/1908.11569`.

[22] Stephan J. Ihle, Andreas M. Reichmuth, Sophie Girardin, Hana Han, Flurin Stauffer, Anne Bonnin, Marco Stampanoni, Karthik Pattisapu, János Vörös, and Csaba Forró. Unsupervised data to content transformation with histogram-matching cycle-consistent generative adversarial networks. *Nature Machine Intelligence 2019 1:10*, 1(10):461–470, 9 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0096-2. URL `https://www.nature.com/articles/s42256-019-0096-2`.

[23] Faisal Mahmood, Daniel Borders, Richard J. Chen, Gregory N. Mckay, Kevan J. Salimian, Alexander Baras, and Nicholas J. Durr. Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images. *IEEE transactions on medical imaging*, 39(11):3257–3267, 11 2020. ISSN 1558254X. doi: 10.1109/TMI.2019.2927182.

[24] Thomas de Bel, John Melle Bokhorst, Jeroen van der Laak, and Geert Litjens. Residual cyclegan for robust domain transformation of histopathological tissue slides. *Medical Image Analysis*, 70:102004, 5 2021. ISSN 13618423. doi: 10.1016/j.media.2021.102004.

[25] Ricky Chen, Timothy T. Yu, Gavin Xu, Da Ma, Marinko V. Sarunic, and Mirza Faisal Beg. Domain Adaptation via CycleGAN for Retina Segmentation in Optical Coherence Tomography. *arXiv*, 7 2021. doi: 10.48550/arxiv.2107.02345. URL `https://arxiv.org/abs/2107.02345v1` `http://arxiv.org/abs/2107.02345`.

[26] Chichen Fu, Soonam Lee, David Joon Ho, Shuo Han, Paul Salama, Kenneth W Dunn, and Edward J Delp. Three dimensional fluorescence microscopy image synthesis and segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2018-June, pages 2302–2310, 2018. ISBN 9781538661000. doi: 10.1109/CVPRW.2018.00298.

[27] Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. Self-Supervised Vessel Segmentation via Adversarial Learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7516–7525. Institute of Electrical and Electronics Engineers (IEEE), 2 2022. ISBN 9781665428125. doi: 10.1109/iccv48922.2021.00744.

[28] Hengfei Cui, Chang Yuwen, and Lei Jiang. Unsupervised Three-Dimensional Tubular Structure Segmentation via Filter Combination. *International Journal of Computational Intelligence Systems*, 14(1):1–12, 12 2021. ISSN 18756883. doi: 10.1007/s44196-021-00027-8. URL `https://link.springer.com/article/10.1007/s44196-021-00027-8`.

[29] Shir Gur, Lior Wolf, Lior Golgher, and Pablo Blinder. Unsupervised microvascular image segmentation using an active contours mimicking neural network. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-Octob, pages 10721–10730, 2019. ISBN 9781728148038. doi: 10.1109/ICCV.2019.01082. URL `https://github.com/shirgur/UMIS`.

[30] Juan Aguirre, Mathias Schwarz, Natalie Garzorz, Murad Omar, Andreas Buehler, Kilian Eyerich, and Vasilis Ntziachristos. Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy. *Nature Biomedical Engineering*, 1(5):1–8, 5 2017. ISSN 2157846X. doi: 10.1038/s41551-017-0068.

[31] Yik Weng Yew, Dinish Unnimadhava Kurup Soudamini Amma, Amanda Hui Yu Kuan, Xiuting Li, Kapil Dev, Amalina Binte Ebrahim Attia, Renzhe Bi, Mohesh Moothanchery, Ghayathri Balasundaram, Juan Aguirre, Vasilis Ntziachristos, Malini Olivo, and Steven Tien Guan Thng. Raster-scanning optoacoustic mesoscopy imaging as an objective disease severity tool in atopic dermatitis patients. *Journal of the American Academy of Dermatology*, 84(4):1121–1123, 4 2021. ISSN 10976787. doi: 10.1016/j.jaad.2020.06.045. URL `https://pubmed.ncbi.nlm.nih.gov/32561371/`.

[32] Adrian P. Regensburger, Emma Brown, Gerhard Krönke, Maximilian J. Waldner, and Ferdinand Knieling. Optoacoustic imaging in inflammation. *Biomedicines*, 9(5), 2021. ISSN 22279059. doi: 10.3390/biomedicines9050483.

[33] Xosé Luís Deán-Ben and Daniel Razansky. Optoacoustic imaging of the skin. *Experimental Dermatology*, 30(11):1598–1609, 11 2021. ISSN 16000625. doi: 10.1111/exd.14386.

[34] Lina Hacker, Emma L. Brown, Thierry L. Lefebvre, Paul W. Sweeney, and Sarah E. Bohndiek. Evaluation of precision and sensitivity in mesoscopic photoacoustic imaging. *bioRxiv*, 2022.

[35] Emma L. Brown, Thierry L. Lefebvre, Paul W. Sweeney, Bernadette J. Stolz, Janek Gröhl, Lina Hacker, Ziqiang Huang, Dominique Laurent Couturier, Heather A. Harrington, Helen M. Byrne, and Sarah E. Bohndiek. Quantification of vascular networks in photoacoustic mesoscopy. *Photoacoustics*, 26:100357, 6 2022. ISSN 22135979. doi: 10.1016/j.pacs.2022.100357.

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 770–778. IEEE Computer Society, 12 2016. ISBN 9781467388504. doi: 10.1109/CVPR.2016.90.

[37] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 5 2018. ISSN 15580571. doi: 10.1109/LGRS.2018.2802944.

[38] Chongzhen Zhang, Yang Tang, Chaoqiang Zhao, Qiyu Sun, Zhencheng Ye, and Jurgen Kurths. Multitask GANs for Semantic Segmentation and Depth Completion with Cycle Consistency. *IEEE Transactions on Neural Networks and Learning Systems*, 32(12):5404–5415, 11 2021. ISSN 21622388. doi: 10.1109/TNNLS.2021.3072883.

[39] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P W Pluim, Ulrich Bauer, and Bjoern H Menze. clDice-a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16560–16569, 2021.

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 630–645, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46493-0.

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv*, 7 2016. doi: 10.48550/arxiv.1607.08022. URL `https://arxiv.org/abs/1607.08022v3` `http://arxiv.org/abs/1607.08022`.

[42] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve Restricted Boltzmann machines. In *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pages 807–814, 2010. ISBN 9781605589077.

[43] Bilal Alsallakh, Narine Kokhlikyan, Facebook Ai, Vivek Miglani, Jun Yuan, and Orion Reblitz-Richardson. Mind the Pad – CNNs can Develop Blind Spots. In *9th International Conference on Learning Representations (ICLR)*, Vienna, Austria, 10 2021. URL `https://arxiv.org/abs/2010.02178v1`.

[44] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N. Straehle, Bernhard X. Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, Kemal Eren, Jaime I. Cervantes, Buote Xu, Fynn Beuttenmueller, Adrian Wolny, Chong Zhang, Ullrich Koethe, Fred A. Hamprecht, and Anna Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16(12):1226–1232, 12 2019. ISSN 15487105. doi: 10.1038/s41592-019-0582-9. URL `https://pubmed.ncbi.nlm.nih.gov/31570887/`.

[45] Lena Maier-Hein, Annika Reinke, Patrick Godau, Minu D. Tizabi, Evangelia Christodoulou, Ben Glocker, Fabian Isensee, Jens Kleesiek, Michal Kozubek, Mauricio Reyes, Michael A. Riegler, Manuel Wiesenfarth, Michael Baumgartner, Matthias Eisenmann, Doreen Heckmann-Nötzel, A. Emre Kavur, Tim Rädsch, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, Peter Bankhead, Arriel Benis, M. Jorge Cardoso, Veronika Cheplygina, Beth Cimini, Gary S. Collins, Keyvan Farahani, Luciana Ferrer, Adrian Galdran, Bram van Ginneken, Robert Haase, Daniel A. Hashimoto, Michael M. Hoffman, Merel Huisman, Pierre Jannin, Charles E. Kahn, Dagmar Kainmueller, Bernhard Kainz, Alexandros Karargyris, Alan Karthikesalingam, Hannes Kenngott, Florian Kofler, Annette Kopp-Schneider, Anna Kreshuk, Tahsin Kurc, Bennett A. Landman, Geert Litjens, Amin Madani, Klaus Maier-Hein, Anne L. Martel, Peter Mattson, Erik Meijering, Bjoern Menze, David Moher, Karel G. M. Moons, Henning Müller, Brennan Nichyporuk, Felix Nickel, Jens Petersen, Nasir Rajpoot, Nicola Rieke, Julio Saez-Rodriguez, Clarisa Sánchez Gutiérrez, Shravya Shetty, Maarten van Smeden, Carole H. Sudre, Ronald M. Summers, Abdel A. Taha, Sotirios A. Tsaftaris, Ben Van Calster, Gaël Varoquaux, and Paul F. Jäger. Metrics reloaded: Pitfalls and recommendations for image analysis validation. *arXiv*, 6 2022. doi: 10.48550/arxiv.2206.01653. URL `https://arxiv.org/abs/2206.01653v3`.

[46] Bernadette J. Stolz, Jakob Kaeppler, Bostjan Markelc, Franziska Braun, Florian Lipsmeier, Ruth J. Muschel, Helen M. Byrne, and Heather A. Harrington. Multiscale topology characterizes dynamic tumor vascular networks. *Science Advances*, 8(23):2456, 6 2022. ISSN 23752548. doi: 10.1126/sciadv.abm2456. URL `https://www.science.org/doi/10.1126/sciadv.abm2456`.

[47] Gael Diot, Stephan Metz, Aurelia Noske, Evangelos Liapis, Barbara Schroeder, Saak V. Ovsepian, Reinhard Meier, Ernst Rummeny, and Vasilis Ntziachristos. Multispectral Optoacoustic Tomography (MSOT) of human breast cancer. *Clinical Cancer Research*, 23(22): 6912–6922, 11 2017. ISSN 15573265. doi: 10.1158/1078-0432.CCR-16-3200.

[48] Alejandra Bruna, Oscar M. Rueda, Wendy Greenwood, Ankita Sati Batra, Maurizio Callari, Rajbir Nath Batra, Katherine Pogrebniak, Jose Sandoval, John W. Cassidy, Ana Tufegdzic-Vidakovic, Stephen John Sammut, Linda Jones, Elena Provenzano, Richard Baird, Peter Eirew, James Hadfield, Matthew Eldridge, Anne McLaren-Douglas, Andrew Barthorpe, Howard Lightfoot, Mark J. O'Connor, Joe Gray, Javier Cortes, Jose Baselga, Elisabetta Marangoni, Alana L. Welm, Samuel Aparicio, Violeta Serra, Mathew J. Garnett, and Carlos Caldas. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell*, 167(1):260–274, 9 2016. ISSN 10974172. doi: 10.1016/j.cell.2016.08.041.

[49] Emma Brown, Joanna Brunker, and Sarah E Bohndiek. Photoacoustic imaging as a tool to probe the tumour microenvironment. *Disease Models and Mechanisms*, 12(7):dmm039636, 2019. ISSN 17548411. doi: 10.1242/dmm.039636. URL `https://dmm.biologists.org/content/dmm/12/7/dmm039636.full.pdf`.

[50] Joes Staal, Michael D. Abràmoff, Meindert Niemeijer, Max A. Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 4 2004. ISSN 02780062. doi: 10.1109/TMI.2004.825627.

[51] Adam 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. ISSN 02780062. doi: 10.1109/42.845178.

[52] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, Alexander Dovzhenko, Olaf Tietz, Cristina Dal Bosco, Sean Walsh, Deniz Saltukoglu, Tuan Leng Tay, Marco Prinz, Klaus Palme, Matias Simons, Ilka Diester, Thomas Brox, and Olaf Ronneberger. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16:67–70, 2019. ISSN 15487105. doi: 10.1038/s41592-018-0261-2.

[53] Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajadin, and Nasir Rajpoot. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Analysis*, 65, 10 2020. ISSN 13618423. doi: 10.1016/j.media.2020.101771.

[54] Christoph Sommer, Christoph Straehle, Ullrich Kothe, and Fred A. Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *Proceedings - International Symposium on Biomedical Imaging*, pages 230–233, 2011. ISBN 9781424441280. doi: 10.1109/ISBI.2011.5872394.

[55] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 6 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2019.

[56] Peter Bankhead, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D. Dunne, Stephen McQuaid, Ronan T. Gray, Liam J. Murray, Helen G. Coleman, Jacqueline A. James, Manuel Salto-Tellez, and Peter W. Hamilton. QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 12 2017. ISSN 20452322. doi: 10.1038/s41598-017-17204-5.

[57] Ignacio Arganda-Carreras, Verena Kaynig, Curtis Rueden, Kevin W. Eliceiri, Johannes Schindelin, Albert Cardona, and H. Sebastian Seung. Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics*, 33

(15):2424–2426, 8 2017. ISSN 14602059. doi: 10.1093/bioinformatics/btx180. URL https://academic.oup.com/bioinformatics/article/33/15/2424/3092362.

[58] A. Inés, C. Domínguez, J. Heras, E. Mata, and V. Pascual. DeepClas4Bio: Connecting bioimaging tools with deep learning frameworks for image classification. *Computers in Biology and Medicine*, 108:49–56, 5 2019. ISSN 18790534. doi: 10.1016/j.compbiomed.2019.03.026.

[59] Estibaliz Gómez-de Mariscal, Carlos García-López-de Haro, Wei Ouyang, Laurène Donati, Emma Lundberg, Michael Unser, Arrate Muñoz-Barrutia, and Daniel Sage. DeepImageJ: A user-friendly environment to run deep learning models in ImageJ. *Nature Methods*, 18(10):1192–1195, 9 2021. ISSN 15487105. doi: 10.1038/s41592-021-01262-9. URL https://www.nature.com/articles/s41592-021-01262-9.

[60] Matthias Arzt, Joran Deschamps, Christopher Schmied, Tobias Pietzsch, Deborah Schmidt, Pavel Tomancak, Robert Haase, and Florian Jug. LABKIT: Labeling and Segmentation Toolkit for Big Image Data. *Frontiers in Computer Science*, 4:10, 2 2022. ISSN 26249898. doi: 10.3389/fcomp.2022.777728.

[61] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Danelion Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems. In *12th USENIX Symposium on Operating Systems Design and Implementation*, pages 265–278, 2015. ISBN 978-1-931971-33-1. URL tensorflow.org.

[62] François Chollet and Others. Keras, 2015. URL https://keras.io.

[63] Javier Montalt Tordera. TensorFlow MRI, 12 2021. URL https://zenodo.org/record/5752336.

[64] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 6 2015. ISSN 1063-6919. doi: 10.1109/CVPR.2015.7298965.

[65] Phillip Isola, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 5967–5976, 11 2017. doi: 10.1109/CVPR.2017.632.

[66] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[67] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient Object Localization Using Convolutional Networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 648–656. IEEE Computer Society, 11 2014. ISBN 9781467369640. doi: 10.1109/CVPR.2015.7298664. URL `https://arxiv.org/abs/1411.4280v3`.

[68] Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc Huszár. Amortised map inference for image super-resolution. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017. URL `https://www.repository.cam.ac.uk/handle/1810/304913`.

[69] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.

[70] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least Squares Generative Adversarial Networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821, Venice, Italy, 2017. ISBN 9781538610329. doi: 10.1109/ICCV.2017.304.

[71] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss Functions for Image Restoration With Neural Networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 12 2016. ISSN 2573-0436. doi: 10.1109/TCI.2016.2644865.

[72] Abdul Rehman, Mohammad Rostami, Zhou Wang, Dominique Brunet, and Edward R. Vrscay. SSIM-inspired image restoration using sparse representation. *Eurasip Journal on Advances in Signal Processing*, 2012(1):1–12, 1 2012. ISSN 16876172. doi: 10.1186/1687-6180-2012-16/FIGURES/5. URL `https://asp-eurasipjournals.springeropen.com/articles/10.1186/1687-6180-2012-16`.

[73] Dam Ja Ba Nay. Single Image Super Resolution using ESPCN – With SSIM Loss. *SSRN Electronic Journal*, 1 2021. doi: 10.2139/SSRN.3898913. URL `https://papers.ssrn.com/abstract=3898913`.

[74] Haoyue Shi, Le Wang, Nanning Zheng, Gang Hua, and Wei Tang. Loss functions for pose guided person image generation. *Pattern Recognition*, 122:108351, 2 2022. ISSN 0031-3203. doi: 10.1016/J.PATCOG.2021.108351.

[75] Aristid Lindenmayer. Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *Journal of Theoretical Biology*, 18(3):280–299, 1968. ISSN 10958541. doi: 10.1016/0022-5193(68)90079-9.

[76] Miguel A. Galarreta-Valverde. *Geração de redes vasculares sintéticas tridimensionais utilizando sistemas de Lindenmayer estocásticos e parametrizados.* PhD thesis, Instituto de Matemática e Estatística, 2012.

[77] Miguel A. Galarreta-Valverde, Maysa M. G. Macedo, Choukri Mekkaoui, and Marcel P. Jackowski. Three-dimensional synthetic blood vessel generation using stochastic L-systems. In *Medical Imaging 2013: Image Processing*, page 86691I, 2013. ISBN 9780819494436. doi: 10.1117/12.2007532.

[78] Janek Gröhl, Kris K. Dreher, Melanie Schellenberg, Tom Rix, Niklas Holzwarth, Patricia Vieten, Leonardo Ayala, Sarah E. Bohndiek, Alexander Seitel, and Lena Maier-Hein. SIMPA: an open-source toolkit for simulation and image processing for photonics and acoustics. *Journal of Biomedical Optics*, 27(8):083010, 4 2022. ISSN 1083-3668. doi: 10.1117/1.JBO.27.8.083010.

[79] Bradley E. Treeby and B. T. Cox. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of Biomedical Optics*, 15(2):021314, 2010. ISSN 10833668. doi: 10.1117/1.3360308. URL `https://pubmed.ncbi.nlm.nih.gov/20459236/`.

[80] Isabel Quiros-Gonzalez, Michal R Tomaszewski, Sarah J Aitken, Laura Ansel-Bollepalli, Leigh Ann McDuffus, Michael Gill, Lina Hacker, Joanna Brunker, and Sarah E Bohndiek. Optoacoustics delineates murine breast cancer models displaying angiogenesis and vascular mimicry. *British Journal of Cancer*, 118(8):1098–1106, 2018. ISSN 15321827. doi: 10.1038/s41416-018-0033-x.

[81] James Joseph, Michal R. Tomaszewski, Isabel Quiros-Gonzalez, Judith Weber, Joanna Brunker, and Sarah E. Bohndiek. Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects. *Journal of Nuclear Medicine*, 58(5):807–814, 5 2017. ISSN 2159662X. doi: 10.2967/jnumed.116.182311.

[82] Diederik P. Kingma and Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014. URL `https://arxiv.org/abs/1412.6980v9`.

[83] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *30th International Conference on Machine Learning, ICML 2013*, pages 2347–2355, 2013.

[84] Cao Duong Ly, Van Tu Nguyen, Tan Hung Vo, Sudip Mondal, Sumin Park, Jaeyeop Choi, Thi Thu Ha Vu, Chang Seok Kim, and Junghwan Oh. Full-view in vivo skin and blood vessels profile

segmentation in photoacoustic imaging based on deep learning. *Photoacoustics*, 25:100310, 3 2022. ISSN 22135979. doi: 10.1016/j.pacs.2021.100310.

[85] Russell Bates. *Learning to Extract Tumour Vasculature: Techniques in Machine Learning for Medical Image Analysis*. PhD thesis, University of Oxford, 2017.

[86] Russell Bates. Russ-learn: set of tools for application and training of deep learning methods for image segmentation and vessel analysis, 2018. URL `https://ibme-gitcvs.eng.ox.ac.uk/RussellB/unet-test`.

[87] Moo K. Chung, Hyekyoung Lee, Alex Di Christofano, Hernando Ombao, and Victor Solo. Exact topological inference of the resting-state brain networks in twins. *Network Neuroscience*, 3(3): 674–694, 7 2019. ISSN 24721751. doi: 10.1162/netn_a_00091.