

PAPER · OPEN ACCESS

Image segmentation for neuroscience: lymphatics

To cite this article: N Tabassum *et al* 2021 *J. Phys. Photonics* **3** 035004

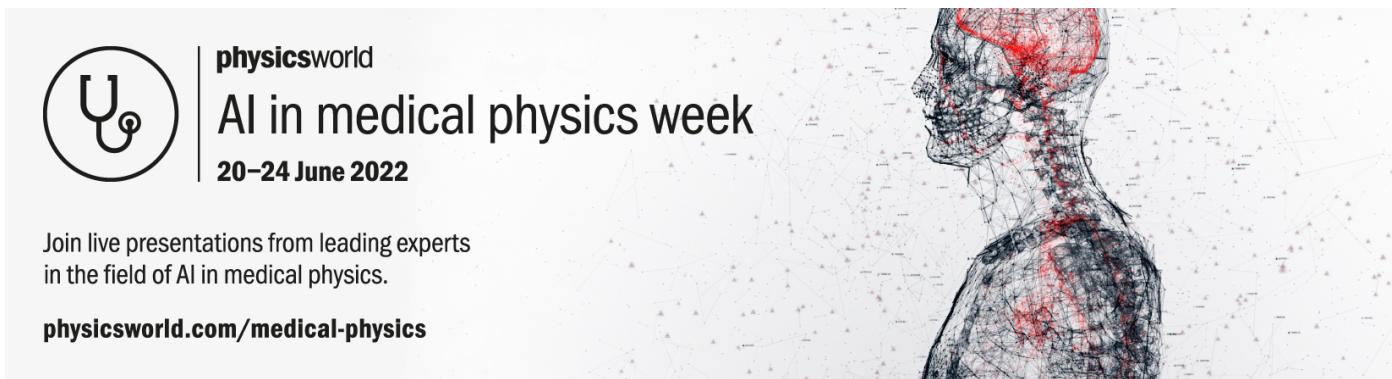
View the [article online](#) for updates and enhancements.

You may also like

- [A unified variational segmentation framework with a level-set based sparse composite shape prior](#)
Wenyang Liu and Dan Ruan

- [Fully automated renal parenchyma volumetry using a support vector machine based recognition system for subject-specific probability map generation in native MR volume data](#)
Oliver Gloger, Klaus Tönnies, Birger Mensel et al.

- [An Automatically Initialized Level Set Method for Ochotona Curzoniae Image Segmentation](#)
Haiyan Chen and Huaqing Zhang



The advertisement features a circular logo on the left containing a stylized 'U' and 'G' intertwined. To the right of the logo, the text 'physicsworld' is written in a lowercase, sans-serif font. Below this, the text 'AI in medical physics week' is displayed in a large, bold, sans-serif font. Underneath that, the dates '20–24 June 2022' are shown in a smaller, bold, sans-serif font. To the right of the text, there is a profile view of a human head and shoulders. The head is filled with a complex network of red and black lines, representing a medical image or a neural network. The background of the entire advertisement is a light gray with a subtle, faint network pattern.

Join live presentations from leading experts in the field of AI in medical physics.

physicsworld.com/medical-physics

**OPEN ACCESS****PAPER**

Image segmentation for neuroscience: lymphatics

N Tabassum^{1,6,*} J Wang¹, M Ferguson², J Herz³, M Dong⁴, A Louveau⁵, J Kipnis³ and S T Acton¹

RECEIVED
16 November 2020

REVISED
18 April 2021

ACCEPTED FOR PUBLICATION
25 May 2021

PUBLISHED
17 June 2021

Original Content from
this work may be used
under the terms of the
Creative Commons
Attribution 4.0 licence.

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



¹ Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, United States of America

² Department of Computer Science, University of Virginia, Charlottesville, VA, United States of America

³ Department of Pathology and Immunology, Washington University School of Medicine in St Louis, St Louis, MO, United States of America

⁴ Internal Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States of America

⁵ Lerner Research Institute, Cleveland Clinic, Cleveland, OH, United States of America

⁶ Analytic Server R&D Testing, SAS Institute, Cary, NC 27513, United States of America

* Author to whom any correspondence should be addressed.

E-mail: nt5rc@virginia.edu and Nazia.Tabassum@sas.com

Keywords: meningeal lymphatics, vessel segmentation, level-set segmentation, shape transform, data augmentation, manual segmentation, deformation transfer

Abstract

A recent discovery in neuroscience prompts the need for innovation in image analysis. Neuroscientists have discovered the existence of meningeal lymphatic vessels in the brain and have shown their importance in preventing cognitive decline in mouse models of Alzheimer's disease. With age, lymphatic vessels narrow and poorly drain cerebrospinal fluid, leading to plaque accumulation, a marker for Alzheimer's disease. The detection of vessel boundaries and width are performed by hand in current practice and thereby suffer from high error rates and potential observer bias. The existing vessel segmentation methods are dependent on user-defined initialization, which is time-consuming and difficult to achieve in practice due to high amounts of background clutter and noise. This work proposes a level set segmentation method featuring hierarchical matting, LyMPhi, to predetermine foreground and background regions. The level set force field is modulated by the foreground information computed by matting, while also constraining the segmentation contour to be smooth. Segmentation output from this method has a higher overall Dice coefficient and boundary F1-score compared to that of competing algorithms. The algorithms are tested on real and synthetic data generated by our novel shape deformation based approach. LyMPhi is also shown to be more stable under different initial conditions as compared to existing level set segmentation methods. Finally, statistical analysis on manual segmentation is performed to prove the variation and disagreement between three annotators.

1. Introduction

The meningeal lymphatics are present around major sinuses in the meninges, three membranes that line the skull and vertebral canal, enclosing the brain and spinal cord. Both the fluid part of the cerebrospinal fluid (CSF) and immune cells are drained through the surrounding lymphatic vessels, all the way to the deep cervical lymph nodes. This drainage occurs in steady state [1]. Recent work shows that in old age, lymphatic drainage is reduced by shrinking of the lymphatic vessels [2], leading to cognitive decline and loss of memory, which is common in Alzheimer's disease (AD) [3]. In the case of multiple sclerosis, it is evident that the lymphatic vessels are used to transport the immune cells responsible for autoimmune attack [4, 5]. Such attacks occur on the myelin sheaths of neurons, which can result in paralysis.

The underlying question for neuroscientists then remains: can we predict onset of these diseases or diagnose patients by studying an individual's lymphatic vasculature [6]? Through the study of lymphatics via image analysis, this paper takes a major step towards this goal. There exist no specialized image processing tools for meningeal lymphatics—this work proposes the first approach. We have developed a toolkit for analysis of meningeal lymphatic vessels that is two-pronged: a novel segmentation approach is proposed, and

then tested on real and synthetic data that has been generated using a unique application of shape deformation.

First, some review and background of the current findings concerning the meningeal lymphatics will be covered. Background of traditional segmentation methods will subsequently be included. Section 4 introduces our method, LyMPhi (Lymphatic Matted Phi), a level-set segmentation method powered by image matting, the first of its kind. Section 5 covers the image datasets, including how the manual segmentation is performed. This section is also devoted to a short literature review of shape analysis, as well as our innovative shape deformation based data augmentation method. Section 6 presents the segmentation results on all datasets using automated segmentation methods, including LyMPhi. Section 7 illustrates the fallibility of manual annotation by using statistical analysis. Afterwards, the final discussions and conclusions are presented.

2. Background on meningeal lymphatics

The meninges communicate to the brain periphery via the lymphatic system, because the meningeal compartment lacks a blood brain barrier. The meningeal lymphatic system is a central regulator of CNS homeostasis. Meningeal T-lymphocytes produce cytokines (such as interleukin-4 and interferon-gamma) that regulate cognitive and social behavior in mice. There are many different immune cells in the brain. During neuroinflammation (modeled by Experimental Autoimmune Encephalomyelitis in mice), lymphoid cells 1 (group 1 innate lymphoid cells, or ILC1) and mast cells play an important role. Lymphoid cells 1 generate a pro-inflammatory environment and encourage filtration of cells through the brain parenchyma. The immune cells in the meninges activate pathogenic T-cells and regulate their migration into the central nervous system (CNS) [7].

The lack of a blood brain barrier is important (the parenchyma has a blood brain barrier); however, more important is the lymphatic system which connects the CNS to the peripheral immune system [7].

The meningeal lymphatic system was first described in 1787 by Mascagni and Sanctius [8] but it was excluded from anatomical textbooks, and only recently re-discovered in 2014 [6]. Meningeal lymphatic vasculature (MLV) develops postnatally, unlike other lymphatic networks. Development of the MLV depends on the Vascular Endothelial Growth Factor c—Vascular Endothelial Growth Factor Receptor 3 (VEGFc-VEGFR3) pathway, just as peripheral lymphatic networks do (those outside the meninges). The MLV network remains dependent on this pathway well after the developmental stage. Also, transcriptomic analysis of the MLV shows that the genes regulating development and maintaining of the vasculature are unique from those of the peripheral lymphatic system. This implies that while the MLV and peripheral lymphatic endothelial cells (LECs) share molecular features, atypical pathways are involved in forming and maintaining the MLV [5]. This is extremely important. Therefore, it is reasonable to conclude that the structure could also be different from regular lymphatic networks, since the genes encoding development and maintenance are different [7].

Also, the spinal region contains lymphatic vessels; a network of collecting lymphatics is observed at each nerve root between the lumbar and cervical regions of the spinal cord. These are the lymphatic vessels we will analyze in our images of the spinal meninges. Analysis of these spinal lymphatics showcases a permeable lymphatic phenotype, which scales positively with complexity. Less complex junctions are found to be impermeable, such as the zipper-like junctions found along the transverse and superior sagittal sinuses [5, 7, 9, 10].

The MLV is capable of uptake of CSF, which was found in studies performed by [2], where the MLV was ablated, resulting in less CSF derived macromolecules found in the cervical lymph nodes. However, this did not cause buildup in CSF pressure, so unlike peripheral lymphatics, the MLV is not responsible for recycling of interstitial fluids—that amount of fluid would be too large for MLV uptake [7].

Still, the MLV is definitely a regulator of the glymphatic system, which removes wastes from the CSF into the brain parenchyma. During aging, restoring impaired MLV using VEGFc results in improvement in drainage, glymphatic function (waste or protein removal) and cognitive decline, which was observed in [2]. Amyloid buildup is reduced, and the behavioral effects can be demonstrated in mice [2]. These observations are not fully understood. When MLV are impaired, this might change the composition of the CSF, meaning cell concentrations will be shifted, changing the cells' ability to remove wastes. Parkinson's disease (PD), like AD, is another neurodegenerative disease that displays abnormal accumulation of proteins in the brain parenchyma. Modulating the MLV function does affect both PD and AD in terms of physiopathology: the reduced drainage leads to more protein buildup for both diseases, which further incurs neuronal damage and behavioral degeneration [7].

The MLV promote fluid homeostasis, but like other lymphatic networks, also aid in the recycling of immune cells. In multiple sclerosis (MS) models, ablating the MLV has a positive effect, because this allows

the antigens, including those that attack the myelin sheath of neurons, do not travel as freely [7]. There is a signal sent from the brain to the lymph nodes that requests immune cells to reenter the brain. These immune cells travel through the lymphatic vessels and are responsible for autoimmune attack, which can result in paralysis [4]. The positive effect is shown by lowering paralysis in mice [7]. Removing the vessels does decrease the number of immune cells present, but does not completely remove MS from a mouse. We need to further understand how the signal is sent (cellular, molecular, etc) before researching treatment options. Interestingly, contrary to in AD, during neuroinflammation, there is little change in the size or complexity in the lymphatic vessels [4].

Lymphangiogenesis is the remodeling of peripheral lymphatic networks during inflammation. During the inflammatory state, there is no change in the dorsal or spinal MLV. However, the cribiform plate MLV expands, similar to peripheral lymphangiogenesis [5, 11]. With age, the dorsal MLV degenerate and the basal MLV undergo hyperplasia [2, 10]. Hyperplasia is enlargement, a morphological change. In stroke, the method of stroke affects the morphology of the MLV. For example, photothrombosis (PT) induces lymphangiogenesis, while in transient middle cerebral artery occlusion (tMCAO), lymphangiogenesis is not observed. On the other hand, the absence of MLV during tMCAO worsens the stroke conditions, but this is not the case during PT [7].

Based on this review of what is currently known about the meningeal lymphatics, it is clear that much remains unknown. The current approaches for studying the MLV are not transferable to clinics and human patients as of yet. Thus, researching the mechanics and molecular dynamics of how the MLV change the meningeal and parenchymal brain compartments is of great necessity [7]. It is clear also to see that structurally, there are great changes in the MLV in different conditions, and that these structural changes result in great molecular changes in the central nervous system, which either worsen or improve disease prognosis. To study the vascular structure, and in the future hopefully model these dynamic causal effects, accurate segmentation of the MLV is of the utmost importance.

3. Current segmentation of MLV

Image segmentation is needed to isolate vessels from the background so they can be analyzed further and so that informatics can be computed. Microscopic images of lymphatic vasculature are currently segmented and quantified by way of manually drawing a boundary around the edges of a vessel. While performing manual quantification, we found that this manual segmentation tends to overestimate vessel content and cannot account for small holes or loops in vasculature. Evidence of overestimation is provided in figure 1. Another challenge with manual quantification is that it depends on the operator, and thus there is or can be significant variability, i.e. thirty percent difference in the measurements found of the same image by different operators [2]. Furthermore, manual quantification is a tedious and laborious process which can take a significant amount of time. For example, the segmentation of a single 2000×4000 pixel image can take over two hours.

Our method, LyMPHi (Lymphatic Matted Phi, where Phi is the level set variable), is the first automated method for segmentation of meningeal lymphatic vessels, developed to address the aforementioned problems of user-based segmentation and the following inadequacies of existing level-set segmentation methods.

Level set segmentation is a geometric adaptive technique that benefits from easily varying the topology of active contours by splitting and merging [12]. This geometric segmentation utilizes a higher-dimensional function ϕ where the zero level set represents the object boundary [13]. The motion of the object contour is performed by way of minimizing an energy functional $\mathcal{E}(\phi)$ that yields a zero level set boundary [12–15].

Level set segmentation methods that target tubular structures have been proposed. Zhao *et al* [16] uses a regional kernel for variational level set formulation [17] to maintain the continuity of segmented retinal vessels with variant intensity, where the kernel parameters are manually adjusted for optimal performance. L2S [18], or Legendre Level Set, models the inhomogeneity of tubular structures and intra-region illumination variation using Legendre Polynomials. L2S is heavily biased to the initialization of the level set function. Tubularity Flow Field (TuFF) [19] is a level set method developed for neuron segmentation. With TuFF, to compute curve evolution, an attraction force is proposed to reconnect the disjoint vessel components that are lost using traditional level set segmentation methods.

For the level set methods described above [16–19], initialization, or an offset/threshold value for the level set boundary, is critical to the final result. Most level set algorithms use intensity thresholding, combined with search of specific object scales within the image. Neither of these base methods are substantial especially for meningeal lymphatic segmentation because of the wide variability of intensity and scale within even one single vessel.

The following are the main roadblocks to successfully segmenting meningeal lymphatic vessel images: the variant thickness of vessels, the inhomogeneity of intensities within the vessels, noise, and background clutter. The background clutter is perhaps the most challenging aspect; it consists of blood vessels with which

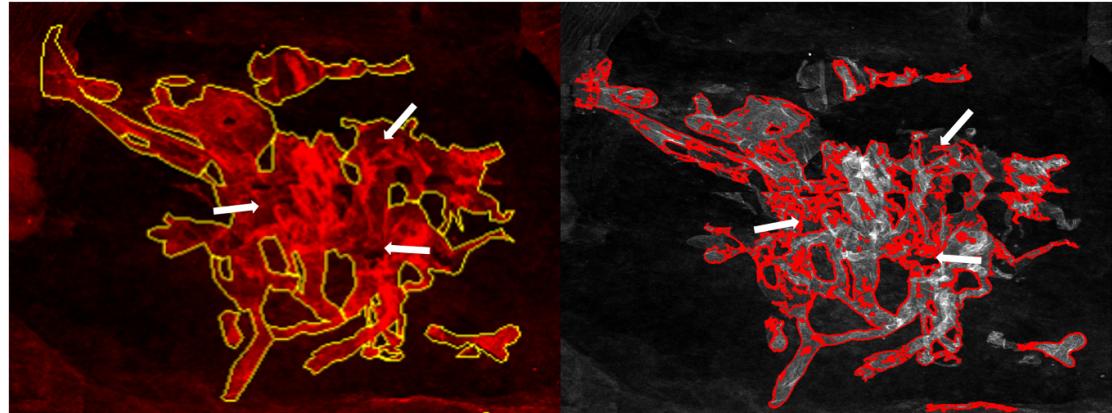


Figure 1. Evidence of over-segmentation when using manual annotation. On the left is a manual segmentation created by a single neuroscience expert, outlining the lymphatic vessels in yellow. This is the current standard for lymphatic vessel segmentation. On the right is an example of segmentation using the proposed method, LyMPhi—the segmentation boundary is displayed in red. Arrows indicate locations at which the user is unable to annotate holes, or gaps, in the vessels. The image depicts spinal lymphatics in a mouse after contusion injury.

the lymphatics vessels are intertwined, as well as other cell types that are also stained. The clutter cannot be removed with thresholding because the vessel staining varies widely in intensity. Existing level set methods do not accommodate these unique challenges of lymphatic vasculature images. In order to remove background clutter, TuFF, for instance, relies on consistent vessel size, which cannot be used in images of lymphatic vasculature. Lymphatic vessel structure varies widely in shape, thickness, and structure. In response to these challenges, we propose a new method which we call Lymphatic Matted ϕ , or LyMPhi to perform image analysis of lymphatic vasculature. LyMPhi utilizes a technique called matting which has been used by others as applied to retinal images [20]. However, to our knowledge, this is the first ever reported work that further develops and uses matting applied to the image analysis of lymphatic vasculature. Furthermore, LyMPhi is the first level-set method powered by matting to reduce clutter in segmentation.

Many of the major segmentation and analysis challenges are shown in figure 2. Even for a neuroscientist, determining which parts of the image belong to the lymphatic vessel set is difficult. In figure 2(c), there are small endothelial cells at the bottom middle cluttering the image. The boundaries of the vessels are not always well-defined, and are surrounded by noise. Hand annotation, thus, is unreliable as well as time-consuming. Removing these repeating noisy artifacts while accurately segmenting the vessel boundary manually is challenging, as the difference in intensity between the noise and the vessel boundary is small.

4. Segmentation method: LyMPhi

4.1. Matting

LyMPhi uses hierarchical image matting, a method developed for separation of foreground vessels from background in fundus images. Matting is a technique used to separate foreground pixels from background pixels, which is necessary to remove the background clutter mentioned above. The following matting algorithm is called hierarchical image matting [20] and is composed of three steps.

4.1.1. Initial trimap

Image matting begins with an initial trimap, composed of foreground vessel (V), background (B), and unknown (U) pixels. In the original matting method, the trimap is generated using size criteria on the blood vessels. Since lymphatic vessels vary widely in size and shape, and enough studies have not been performed to set a bound on maximum or minimum vessel width, LyMPhi uses intensity thresholds to generate the initial trimap. On datasets where the staining strength varies heavily (spinal and synthetic datasets), Otsu multithresholding is used to provide two threshold levels: low and high [21]. Otsu multithresholding has been shown to be a successful pre-segmentation step for iris segmentation, followed by geodesic active contours in [22]. For the partial and whole mount datasets, the lower threshold is 75, and the upper threshold is 100; out of 255 for a grayscale image (single channel). The lower threshold is used as the cutoff for B , while the upper threshold is used for demarking V . The pixels with intensity between the two threshold levels fall in the set U .

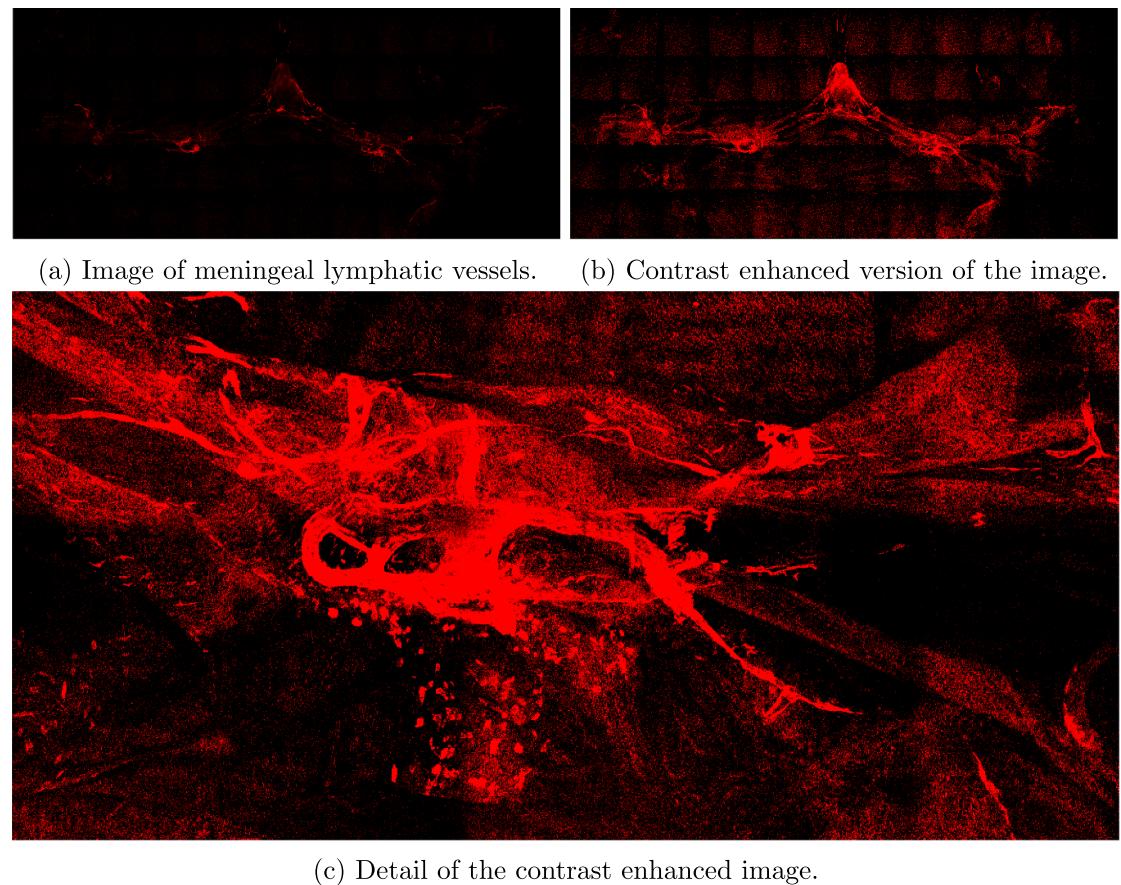


Figure 2. One example image. The original image is shown on the left (a), and the contrast enhanced image is shown on the right (b). The contrast enhancement displays the challenges of noise, artifacts, intensity variation, and background clutter. The final image (c) is a detail image from (b).

4.1.2. Stratification

The minimum distance from a vessel pixel in V , d_i , is calculated for each pixel in U . The pixels in U are stratified into hierarchies based on this minimum distance. Unknown pixels, (pixels that fall between the higher and lower thresholds), with the same distance, d_i are collected into one hierarchy. The hierarchies are ordered based on distance, with the first hierarchy containing all unknown pixels with the lowest distance from foreground pixels. The unknown pixels are labeled in the next step, in order of hierarchy.

4.1.3. Hierarchical update

The hierarchies are updated in order, from closest distance to V to farthest. In a grid centered on each unknown pixel, correlations are calculated between the unknown pixel and surrounding known pixels(already labeled as background or foreground from the initial trimap or lower hierarchies). In the following, a subscript of u refers to an unknown pixel, and the subscript k refers to a pixel in the known set. The correlation function is composed of a color cost function ($\beta_c(u, k)$) and a spatial cost function ($\beta_s(u, k)$). The color cost function is defined as $\beta_c(u, k) = ||c_u - c_k||$, where c_u and c_k are the intensities of the unknown and known pixels, respectively. The intensity is taken in the red channel of the original image. The spatial cost function is $\beta_s(u, k) = \frac{\|x_u - x_k\| - x_{\min}}{x_{\max} - x_{\min}}$, where x_u and x_k are the spatial coordinates of the pixels u and k , respectively. x_{\max} and x_{\min} are normalization terms; namely, the maximum and minimum distances between the unknown pixel u and any known pixel k in the surrounding grid.

The combined correlation function is

$$\beta(u, k) = \frac{1}{\beta_c(u, k) + \beta_s(u, k)} \quad (1)$$

where u is the unknown pixel at the center of the window and k is the known pixel within the window being used for comparison.

An unknown pixel is labeled as foreground if $\beta(u, V) > \beta(u, B)$, and vice versa. After all unknown pixels in a hierarchy have been labeled, the next hierarchy is updated using the initial foreground and the newly

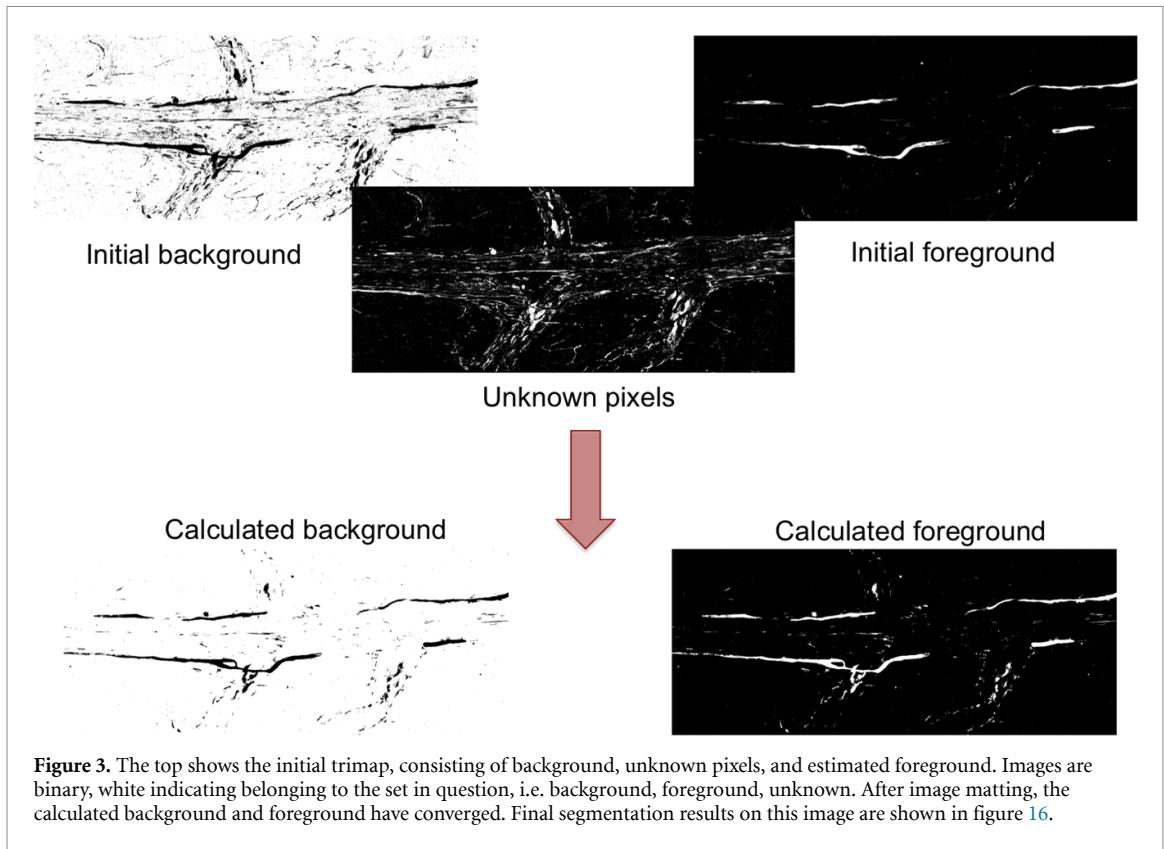


Figure 3. The top shows the initial trimap, consisting of background, unknown pixels, and estimated foreground. Images are binary, white indicating belonging to the set in question, i.e. background, foreground, unknown. After image matting, the calculated background and foreground have converged. Final segmentation results on this image are shown in figure 16.

labeled pixels from the previous hierarchies. The transition from initial trimap to the final matting result is shown in figure 3. The matting algorithm provides more separation between background and foreground elements than simple thresholding. However, the matting result is not smooth because the method is not iterative and has no smoothness constraints, which is why it is in combination with level-set segmentation.

4.2. Curve evolution

To address the challenge of segmenting lymphatic vessels, we leverage and extend TuFF [19] as the backbone of our proposed level set segmentation method. Minimization of the energy functional $\mathcal{E}(\phi)$ is solved by gradient descent, using

$$\frac{\partial \phi}{\partial t} = \mathcal{F}_{\text{reg}}(\mathbf{x}) + \mathcal{F}_{\text{evolve}}(\mathbf{x}) + \mathcal{F}_{\text{attr}}(\mathbf{x}) \quad (2)$$

from [19]. \mathcal{F}_{reg} , $\mathcal{F}_{\text{evolve}}$, and $\mathcal{F}_{\text{attr}}$ are forces, which are defined and applied with respect to the normal direction of the implicit function, $\frac{\partial \phi}{\partial t}$ (as described below). The \mathcal{F}_{reg} term regularizes the length of the zero level set, which controls the smoothness of the contour. $\mathcal{F}_{\text{evolve}}$ drives the evolution of the curve where a combination of axial and orthogonal vector field is used to achieve the curve propagation perpendicular to the vessel boundary and along the vessel axis, respectively. In TuFF, this vector field is created using a edgemap of the vesselness response map, produced by [23], and tuned by the scale parameter determining the thickness of the vessels. As tuning this parameter is difficult for lymphatic vessels, where the vessel scale varies greatly, in LyMPhi, the edgemap is instead computed on the original image.

$\mathcal{F}_{\text{attr}}$ acts for connecting the discontinuous fragments in a way of creating an attraction force field using vector field convolution (VFC) [24]. The vector \mathbf{x} is a position in the image. $\mathcal{F}_{\text{reg}} = \nu_1 \text{div}[\mathbf{n}(\mathbf{x})] \delta_\epsilon(\phi)$ while

$$\mathcal{F}_{\text{attr}}(\mathbf{x}) = \nu_2 \sum_{i=1}^p \sum_{j \neq i}^p \mathcal{F}_{\text{attr}}^{(i,j)}(\mathbf{x}), \forall \mathbf{x} \in \Omega. \quad (3)$$

The expression of $\mathcal{F}_{\text{evolve}}$ and the full forms and derivations of all forces can be found in [19].

For \mathcal{F}_{reg} , ν_1 is a smoothing parameter, $\mathbf{n}(\mathbf{x})$ is the inward normal unit vector to ϕ , and δ_ϵ is the regularized Dirac delta function [25]. For $\mathcal{F}_{\text{attr}}$, ν_2 decides the effect of the attraction force, and p is the number of disjoint connected components that can potentially be attracted to one another. Ω is the image, and d is the dimension of said image. The local attraction force is defined as follows: $\mathcal{F}_{\text{attr}}^{(i,j)}(\mathbf{x}) = \kappa_i \langle \Gamma_i(\mathbf{x}), -\mathbf{n}(\mathbf{x}) \rangle \theta_j(\mathbf{x})$.

i, j refer to two different connected-components. κ_i is the normalized mass of the ‘child’ component, which can be an isolated segment that is the smaller of the two. The indicator function $\theta_j(\mathbf{x})$ determines if a child component is within the convex hull of the ‘parent’, the larger of the two components, usually a branch of the neuron. Γ_i is the attraction force field computed via the VFC technique [24].

Level set evolution is an iterative process, where the level set boundary moves closer to the object boundary over many iterations. A novel modification we make to level set curve evolution takes place during the iteration process. We design a force field over the image that is made up of the component forces:

$$F(i, j) = F_{\text{reg}}(i, j) + F_{\text{evolve}}(i, j) + F_{\text{attr}}(i, j). \quad (4)$$

This force field is the velocity of the level set evolution, or $\frac{\partial \phi}{\partial t}$. The velocity at each point (i, j) has a magnitude and a direction. The magnitude is the speed at which we move the zero level set inwards or outwards towards the object boundary (moving the level set elevation up or down), calculated from the component forces. However, we modify the direction of the velocity according to the matting result, as explained below.

4.2.1. Modifying the velocity field

LyMPhi changes the sign of the velocity field at each pixel according to whether it is labeled foreground or background by matting. After computing $\frac{\partial \phi}{\partial t}$ from (2), LyMPhi uses the following decision rule

$$\frac{\partial \phi}{\partial t}_{i,j} = \begin{cases} \left| \frac{\partial \phi}{\partial t}_{i,j} \right| & \text{if } p_{i,j} \in V \\ -\left| \frac{\partial \phi}{\partial t}_{i,j} \right| & \text{if } p_{i,j} \in B \end{cases} \quad (5)$$

to change the sign of $\frac{\partial \phi}{\partial t}$ at each pixel location (i, j) . p is the matting result at location (i, j) and V, B are the foreground and background maps produced by matting. $\frac{\partial \phi}{\partial t}_{i,j}$ is the original level set propagation at each iteration before modification. If p at i, j belongs to the foreground map (V) created by matting, the sign on $\frac{\partial \phi}{\partial t}$ is made positive, to drive the zero level set towards the vessel boundary. The contour will evolve according to the reassigned velocity function to segment the true vessel boundary.

A cartoon illustration is shown in figure 4 at iteration t . At time t , the original velocity field, calculated from the TuFF forces, is shown by brown arrows. The yellow shape represents the calculated foreground, V , from hierarchical matting. In (A), the red arrows represent an incorrect direction for curve evolution. These arrows are driving ϕ inside the vessel. These directions are corrected in (B), using the matting label calculated. The now green arrows of the contour evolution speed have been changed in the normal direction.

5. Images of MLV

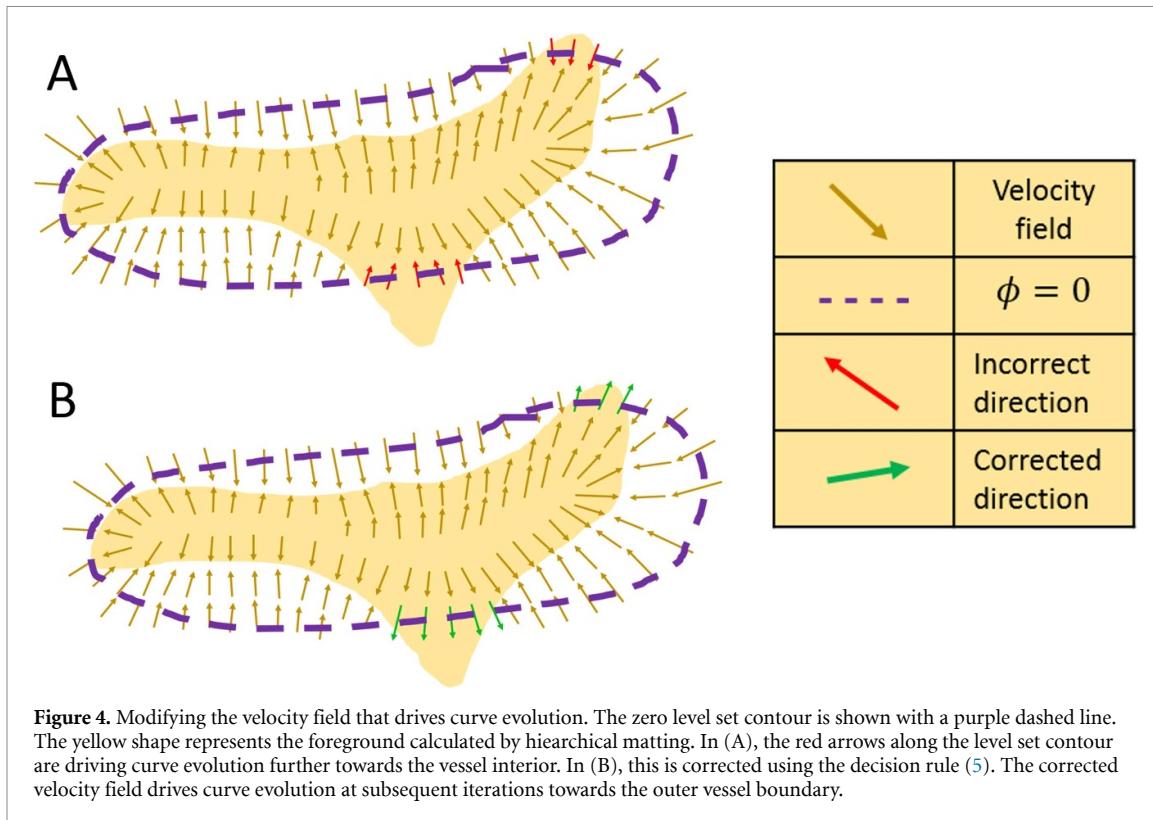
5.1. Description of the datasets used

There are three real image datasets used for the experiments performed. All three datasets contain confocal microscopy images taken of lymphatic vessels in mice, stained with LYVE-1. The images were acquired by the Kipnis laboratory at the Center for Brain Immunology and Glia (BIG), then at the University of Virginia School of Medicine, Department of Neuroscience. The current imaging technology produces 2D images, as the lymphatic vessels are laid on a slide and then imaged. In the future, we hope to extend LyMPhi, particularly the matting algorithm, to 3D images of lymphatics.

There are 39 images in total. Seventeen of these are images of the superior sagittal sinus in the mouse brain, with nine images taken in whole-mount, and eight in partial-mount. An example of the whole-mount sinus is shown in figure 15. The images are 16 bits per pixel for the spinal dataset, and 8 bits per pixel for the whole and partial mount datasets. The image size varies within each dataset. The whole-mount images are 2643×948 pixels, with one image having dimensions 2311×948 pixels. The resolution is 0.33 pixels per micron. The partial-mount images vary more widely in dimension, with the minimum width and height being 4631 and 1941 pixels, respectively. The maximum width and height are 4653 and 1993 pixels. Resolution is 1.61 pixels per micron.

The twenty-two remaining images are of lymphatic vessels in the mouse spinal cord post-injury. These images are referred to as the spinal dataset, an example of which is shown in figure 14. While these vessels do not originate from the spinal meninges, they are phenotypically related to the meningeal lymphatic vessels. The spinal dataset original images vary the most widely in dimension. The minimum width and height are 1344 and 702 pixels, and the maximum width and height are 4195 and 3114 pixels. The resolution for these images is 0.805 pixels per micron.

It is valuable to have three datasets with different resolutions because different levels of vessel details are captured. The vessel size and structure varies widely for each dataset. The whole-mount images contain the



most clutter, as they have the entire structure of the blood vessel network present in the sinus, showing the lymphatic vessels growing on top of and around the blood vessel network (shown in figure 15). The partial-mount images provide finer detail along within the sinus, generally displaying the overall vessel structure at finer detail—which necessarily removes higher level structure, such as branching or looping. In the spinal dataset, lymphatic vessels aggregate in rounder shapes, with a large amount of holes or gaps in the vessels, as shown in figure 1. The vessels in the spinal images are overall not as thin or elongated as the meningeal lymphatic vessels shown in the other two datasets.

The driving biological interest in this approach is in segmenting meningeal lymphatics in the sinus, but because of limited data (17 images), testing is also performed on the spinal images. High performance of LyMPhi on the spinal dataset may imply promising application to generalized lymphatic vessel images.

5.2. Creation of manual annotation

Manual segmentation was performed for comparison with automated segmentation methods. Three operators created separate annotations for each dataset, and the images were merged using majority voting to create one consensus segmentation image for comparison. The vessels are primarily distinguished by brightness or color, and also by shape and relation to surroundings. This manual quantification was performed using the software FIJI (ImageJ), a Java-based image editing software package. The manual annotation is generated by using the ‘Clear’ tool under ‘Edit,’ with ‘Freehand’ selection. The user draws regions around objects in the image that are not considered lymphatic vessel, and these are cleared successively, until only lymphatic vessel content remains in the image. Annotating in the vicinity of small holes and gaps in the vessel presents difficulty; zooming into these smaller regions means that the intensity variation between vessel and non-vessel are less easily detectable by the human eye, due to noise. Often, these holes or gaps are overlooked or improperly manually segmented. We thus add a synthetic dataset to our testing data, to provide unimpeachable ground truth.

5.3. Synthetic dataset

We also test performance on a synthetic dataset of 100 images. We choose to test using synthetic data in addition to real data, because there are errors in the hand annotation, due to not removing the many holes present in the vessels. Thus, only testing with real data means that instead of comparing our results to ground truth, we are comparing the similarity to another flawed measurement. Our method, LyMPhi, may in fact capture more accurately the true shape of lymphatic vasculature. Therefore, we test all methods on a

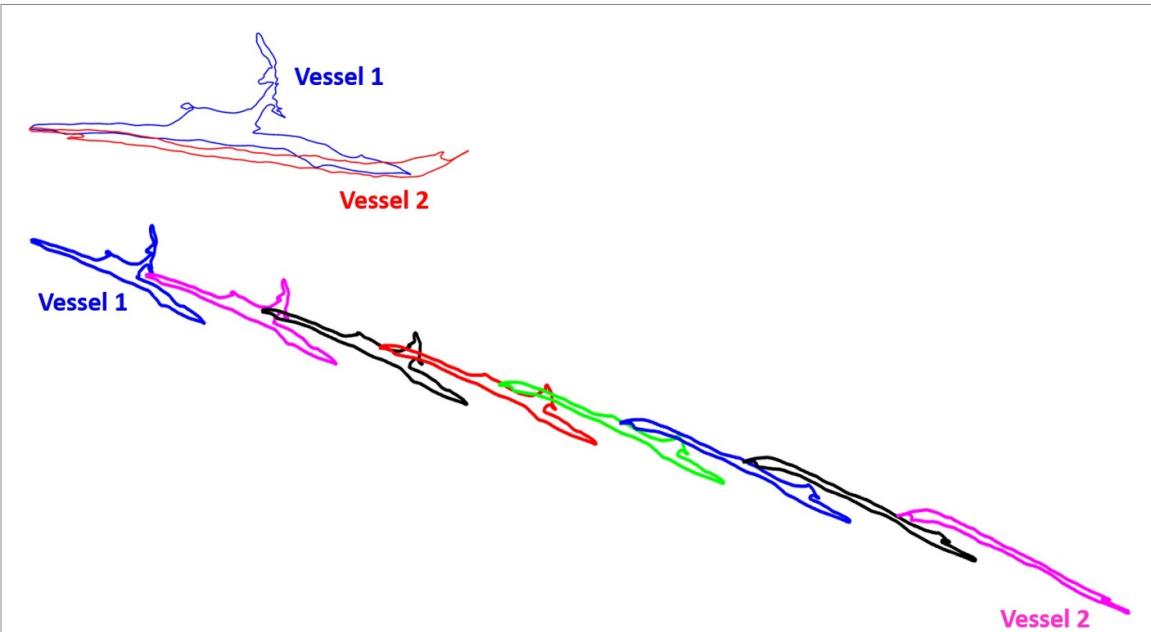


Figure 5. Showing the evolution of one vessel shape (vessel 1) to another shape (vessel 2) along a geodesic path, i.e. morphing between one shape to another. At the top left are shown the two registered shapes.

synthetic dataset of our own creation, where ground truth perfectly captures the true vessels so that any comparisons are more meaningful.

There is also a limited amount of data for analysis purposes. Data collection is expensive in three ways: time, money, and animal life—as each image of meningeal lymphatics requires the sacrifice of one mouse. In this section, we propose a method for creating synthetic vessel data so that everyone who wishes to study the meningeal lymphatics can have access to these data.

We build our own deformation model for lymphatic vessels, allowing us to morph the vessels into reasonable formations. All of our deformations can be run by neuroscientists to ensure they are in line with what changes the vessels naturally undergo. Elastic deformation of shapes is a much explored area of interest within image processing. Similar object morphing approaches are used throughout the image processing literature.

In [26], deformation is used to stretch and bend images of cells into new, biologically plausible shapes—thus creating more data. The authors build a convolutional neural network for segmenting closed loop structures, i.e. cells, in biological microscopy images. The second major contribution that the authors make is to use their own elastic deformation model to create new training data from the limited data available. The deformations are created by randomly displacing pixel intensities along a 3×3 window. The interior pixel displacements are then interpolated. Using the augmented data and the trained u-net, segmentation results are vastly improved compared to other competing methods.

The fundamental drawback of using deformation as described in the u-net paper is that the deviations take place in Euclidean space. These deformations are not guaranteed to be meaningful, i.e. have any similarity to real biological shapes. To ensure that the deformations we use to generate new data are biologically sound, we move to deformations along the *shape space*. In [27], the authors use this idea of shape spaces to show that shapes reside on high dimensional manifolds. Between different shapes on this manifold, a geodesic can be taken, which is the shortest path between two shapes on a manifold. Sampling points along this geodesic showcases the evolution of one shape to another shape. An example of this is shown in trajectory using lymphatic vessel data is shown in figure 5. The sample shapes that lie in between are deformations of the original shapes [28]. We propose to build a large lymphatic dataset by sampling these deformations found between the true vessel data available.

Srivastava *et al* [29] introduces the square root velocity (SRV) transform for shapes to remove translation, scaling, and rotation discrepancies. A shape space is rotation and scale invariant. The authors also introduce a path straightening approach to find the geodesics between shapes on the shape space. We propose using the square root velocity function (SRVF) to transform shapes of lymphatic vessels into new shapes, as a method for data augmentation.

The following workflow explains how the synthetic images were created. To begin, new vessel shapes must be generated. Elastic deformation is used to stretch existing vessel shapes into new vessels [29].

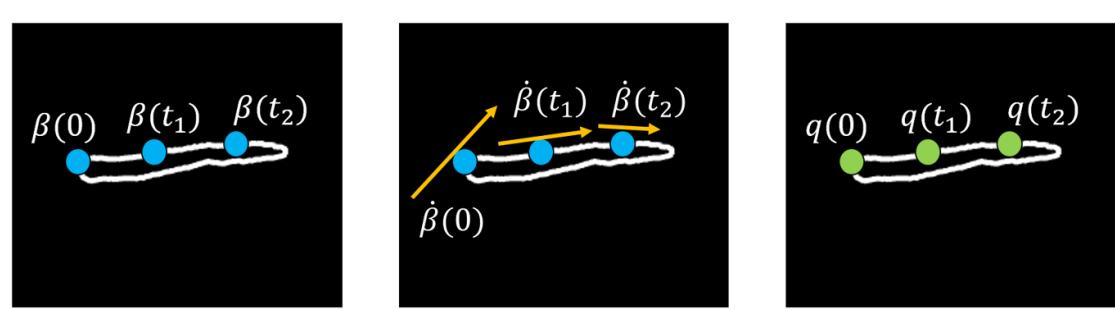


Figure 6. A representation of the SRVF computed at locations on the closed vessel boundary shown in white. The function allows the curve to stretch elastically into another shape. On the left, the original points are labeled $\beta(t)$, starting at $t = 0$ at one end of the shape and ending at $t = 1$ after traversing the whole shape. In the middle, the gradient, or tangent slope, at each point is calculated. The rightmost shape has the labeled points $q(t)$ after the SRV function (6), has been utilized.

5.3.1. Vessel shape generation

First, from the manual segmentation of partial mount images, isolate each individual vessel segment. The manual segmentation must be used in this step, to prevent bias from any automated segmentation methods. Then, use the square root velocity function (SRVF) to represent each shape, thus accounting for rotation, shift, and scale variance [29]. These are just the outer contours.

To use the SRVF, first, the shape boundary is sampled in x, y -space. Three hundred sample points are used to achieve a smooth shape boundary. Using fewer sample points will not allow for a smooth boundary, but using more than three hundred sample points will greatly slow the overall shape generation process without much benefit in terms of smoothness. These sample points, denoted β , are locations, which are transformed by

$$q(t) = \frac{\beta(t)}{\sqrt{\|\beta(t)\|}} \quad (6)$$

the SRV equation [29, 30]. $\beta(t)$ is the gradient at each point $\beta(t)$ on the original curve.

Prior to using SRV, unit length curves are enforced on closed curves (which begin and end at the same location) to remove scaling effects. Each location on the shape is a function of the parameter $t \in [0, 1]$, with $t = 0$ and $t = 1$ being the beginning and end of the curve. The process is illustrated in the schematic shown in figure 6. After the SRVF has been applied, the new points on the curve, $q(t)$, have been transformed from $p(t)$ into the shape space, where elastic bending, stretching, and shrinking is possible. SRV is a shape transform where we can still use Euclidean coordinates to describe the transformed shape. SRV supports elastic deformation, unlike previous methods, such as [27], which only include bending, and not stretching energy. The same deformation process is repeated to represent any holes, or capillary loops in a vessel.

From any two original vessel shapes, morph along a shape geodesic to find new interpolated vessel shapes. The interpolated vessels are stretched and bent versions of the original vessels, i.e. intermediate deformations [30]. Essentially, sample the geodesic path to get closed curve shapes, representing the vessel exterior. The sampling can be performed finely or sparsely to create any number of new shapes.

Because the manifold containing SRV transformed shapes is locally Euclidean, affine transformations are possible. Shape interpolation is performed by taking convex combinations of q : $\alpha q_1 + (1 - \alpha)q_2$, where q_1 and q_2 are two SRV transformed shapes, and $\alpha \in [0, 1]$ (intermediate algorithmic time-steps). q_1 and q_2 must have the same dimensions. For a two-dimensional shape, the dimension of each q is $n \times 2$, where n is the number of samples taken to characterize the shape (in our case, 300).

The SRV transform also enables transport of deformation from one shape to another shape. Using deformation transport [29], the hole shapes can be morphed with the same amount of stretching and bending as the interpolated vessel (as used for the exterior). Essentially, we use the same bending and stretching energy on a new shape. Given an SRV-transformed shape q_1 and an intermediate deformation to a second known SRV-transformed shape, the shooting velocity v_1 can be calculated. This v_1 signifies which direction and how far to travel along the manifold before reaching the desired deformation. ‘Transport’ means finding the parallel translation of v_1 (call it v_2) for morphing the new shape—in this case, a vessel hole. Figure 7 shows this intuitively.

Let $[q_1^a]$ and $[q_1^b]$ be the shapes of a lymphatic vessel, at two points along a geodesic path. $[q_1^a]$ is the original vessel shape in SRV-space (initial time τ_a), and $[q_1^b]$ is some arbitrary deformation of the original shape. We can think of this deformation having occurred after a certain amount of time, say after time τ_b .

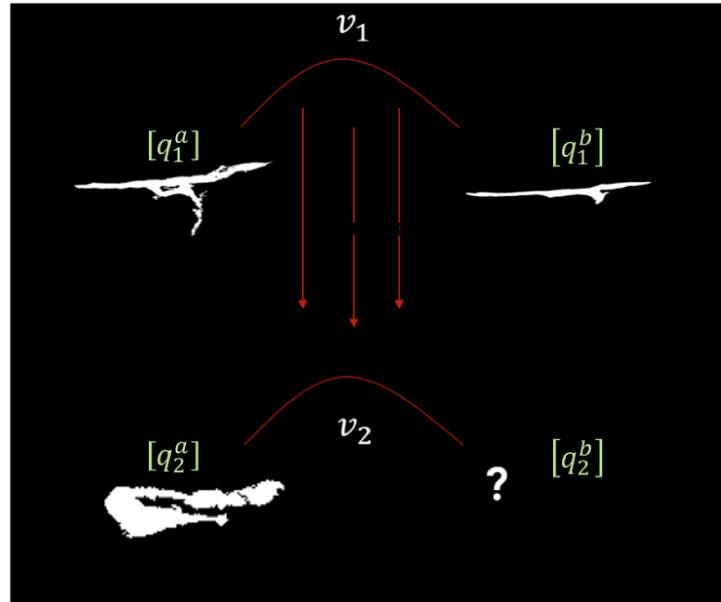


Figure 7. If v_1 is the velocity used to reach the intermediate deformation, v_2 is the velocity to apply the same amount of deformation on the new shape, in this case, a hole. The lower left panel is the large hole within the vessel in the upper right panel. Since shape spaces are overall nonlinear manifolds, the deformations of one shape cannot simply be applied to another. The manifold is only linear when the shape is of the same type.

This contour deformation depends on the geometry of the lymphatic vessel. Now, take a different object, a vessel hole from the original vessel shape, which is similar to the lymphatic vessel, but obviously not identical geometrically. Given its initial shape $[q_2^a]$, prior to any deformation, we wish to predict its shape $[q_2^b]$ after the same amount of time, τ_b as for the exterior lymphatic vessel. We thus take the deformation that deformed $[q_1^a]$ to $[q_1^b]$ and then apply this deformation to $[q_2^a]$, using the following.

- Let $\alpha_1(\tau)$ be a geodesic between the shapes $[q_1^a]$ and $[q_1^b]$ in the SRVF transformed shape space and $v_1 \equiv \dot{\alpha}_1(\tau_a)$ be its initial velocity. (q_1 and q_2 represent two different shapes, and a or b mean a deformation).
- Using forward parallel translation, we *transport* v_1 to $[q_2^a]$. Let $\alpha_{1-2}(\tau)$ be a geodesic from $[q_1^a]$ to $[q_2^a]$ in the same shape space. Construct a vector field ω_τ such that $\omega_0 = v_1$ and $\frac{D\omega}{d\tau} = 0$ for all the points along α_{1-2} . Please see [29] for additional details. Figure 7 shows the relationship between q_1 and q_2 for more clarity.
- Then, $v_2 = \omega(1)$ is a parallel translation of v_1 .
- Using v_2 as the initial velocity, form a geodesic starting at $[q_2^a]$, which at time-step τ_b will end in deformation $[q_2^b]$.

The original work [29] on deformation transport is for applications where viewing angles of objects change. Here, we substitute the change in ‘viewing angle’ for a moment in time, either before or after deformation. This is the first use of the theory for modulating the interior of an object, since a shape necessarily only retains the boundary.

A new, complete vessel shape includes an interpolated shape plus similarly deformed holes placed proportionally (semi-randomly) according to vessel size within the interpolated vessel. A vessel skeleton is generated using the methodology in [31], and the radius of the vessel at each skeleton point is calculated, using size-constrained inscribed spheres [32]. Considering the original size of the hole prior to deformation, the point within the vessel is found where the vessel diameter best matches the original hole size, and the deformed hole is placed at the best diameter match. This ensures the deformed hole fits within the boundary of deformed vessel. Placing the holes based on the original hole location unfortunately does not scale as the vessel may shrink at certain points to where the hole will no longer fit.

New (final) binary vessel images are produced at this stage. An example is shown in figure 8. Bioinformatics can be computed on these new shapes. The full diagram illustrating the shape interpolation process is shown in figure 9.

In summary, we can loop through combinations of vessels—to create a new vessel, morph between two vessel shapes on the shape space. The two original vessel shapes are picked from all possible vessel pairs. For



Figure 8. Output from the shape deformation process is an individual binary vessel, with fitted deformed holes.

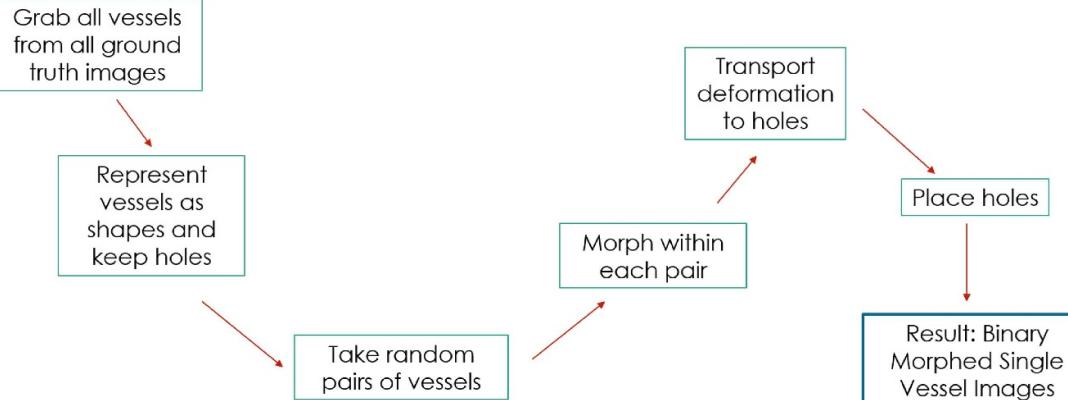


Figure 9. Shape interpolation—from an image to a shape. Streamlined, generalizable workflow from images to new, complete vessel shapes.

example, in a dataset of 18 images, where there are 451 original vessels, the total number of vessel pairs is 101 475!. To create a full meningeal lymphatic image, combinations of these new vessel shapes are used. As the total number of possible interpolated vessels is incredibly large (there are multiple interpolations possible for each of those 101 465 pairs), and new vessel shapes can be endlessly combined in different selections or groupings, truly, the amount of synthetic data that can be generated through this process is nearly infinite.

There is a possibility for using shape deformation to simulate lymphangiogenesis, thinning of vessels, or lymphatic regression, which is the shortening of the lymphatic vessels along the sinus with age. This could be formulated similarly to neurodegeneration, which has already been shown in [30].

5.3.2. Synthetic image formation

A flowchart of the image generation process is shown in figure 10. As both the shape generation and image generation processes are generalizable, other areas where synthetic data would be of benefit could use these procedures to augment existing datasets.

Selections of vessel shapes are used to create foreground and background layers of a new synthetic image. Each layer is 4649×1967 pixels, which is also the size of the final synthetic images. This is a similar size to the images found in the partial mount dataset. Select a random number of vessels to include (based on real images). These vessels are the generated vessels from the shape deformation discussed previously. The images containing these random assortments of vessels are layers for our final image. First, there are some steps needed to produce the foreground layers of the final images.

The foreground layer is used as ground truth, before being convolved with a Gaussian filter to attenuate the signal at vessel edges. This mimics the true images, and is the reason why simple thresholding will not

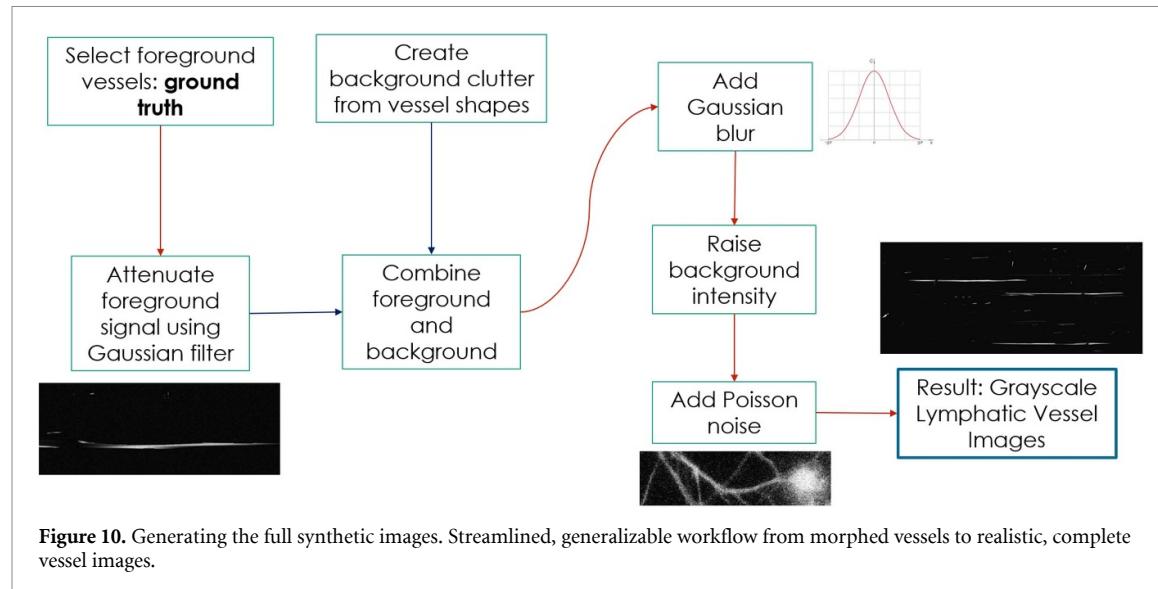


Figure 10. Generating the full synthetic images. Streamlined, generalizable workflow from morphed vessels to realistic, complete vessel images.

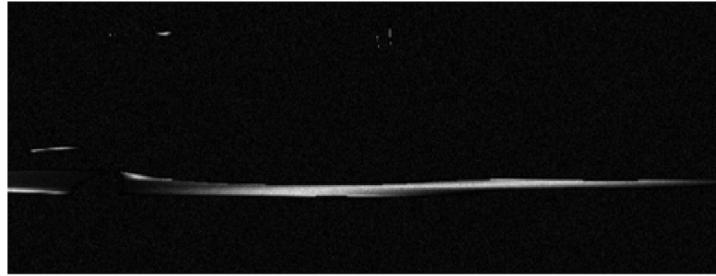


Figure 11. A trimmed Gaussian filter is used over the foreground image layer to attenuate signal intensity. This mimics the true variation in intensity in the microscopy images.

solve the segmentation problem. Set the intensity of the vessels in this layer to the maximum intensity, at 255. The maximum intensity vessel image is used as the ground truth.

Then, a trimmed Gaussian filter is applied over each row in the image where the vessel exists to attenuate the intensity of the vessel at boundaries. This mimics the true images, and is the reason why simple thresholding will not solve the segmentation problem. The result is shown in figure 11.

For each vessel in the foreground layer, the rows in the layer where the vessel exists must be identified. Then, for each of those identified rows, take the length (the number of pixels in that row that belong to vessel foreground) as n . A 1D Gaussian filter of size $2m + 1$, where $m = n$ if n is even, and $m = n + 1$ if n is odd, is used to attenuate the vessel at each row. Centered at the center pixel in the vessel row, this longer Gaussian filter gives the trimmed Gaussian effect.

Next, on the background layers, which simulate clutter, some additional perturbations need to be added. Create three background layers with the generated vessels but set these intensities at 20 percent of the possible higher foreground intensity (*5 to 1 signal to clutter ratio*), so that there is some overlap between the foreground and background within the lower intensities. The background layers represent the clutter of blood vessels in the original images.

We then combine the foreground and background layers together. The following are adding blur and noise to the combined image. Gaussian blur and Poisson noise [33] are added to complete the synthetic images. Gaussian blur is added with sigma between 0.2 and 0.5. The static background intensity is raised up to 19 out of 255, which is similar to the background levels in real data. This is done to make sure that when Poisson noise is applied to the image, the background also becomes noisy—if the background intensity is zero, the Poisson noise in the background will have no effect. Poisson noise is based on the signal intensity in the image.

We use the Poisson noise model on the synthetic data because in confocal microscopy using fluorescence emission to stain biological specimens, Poisson noise is a significant variable. The Poisson noise model is used to reflect the small number and extreme variation of detected photons [33]. Alternatively, using a spatial light modulator could also emulate the particle nature of light as observed using a real confocal microscope.

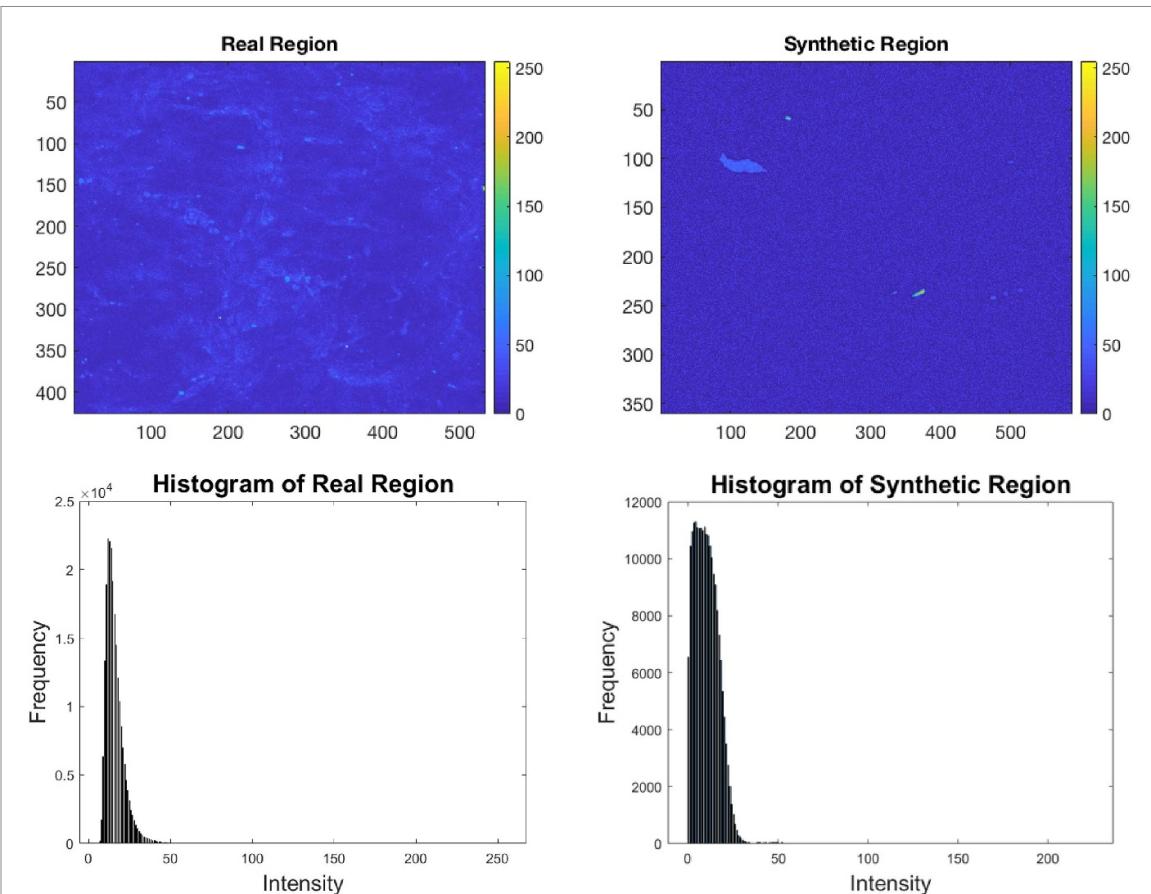


Figure 12. Plotting the small background region in a microscopy image (left) and a synthetic image (right). There is more faint background clutter in the original image that we do not simulate, but the background intensity is in a similar range. In the synthetic images, the background intensity cannot be lowered too much, otherwise the Poisson noise effects are no longer apparent; which is why the static background in the synthetic image is higher. Histograms of a near constant intensity region in both images are presented on the bottom panel: in a microscopy image (left) and a synthetic image (right).

There are some considerations that need to be made when designing such a system, however. Image borders, and image resolution may become distorted during the process, and care needs to be taken not to over-blur the image in the effort to simulate the noise.

Comparisons in terms of noise are shown in figure 12. The noise levels are not exactly the same, but are similar in level and spread given the variations in noise across the many images in the three microscopy datasets. In the histograms, the frequency count for the microscopy region is higher because the region is slightly larger. The intensity in the histograms has been scaled from 0 to 1.

Figure 13 showcases one synthetic image. The ground truth is shown as well to display where the clutter resides.

6. Automated segmentation results

Three level-set methods, including LyMPhi, are compared on the datasets described above. Each method is initialized by thresholding a small amount of noise and background clutter, in order to retain the majority of the foreground. For TuFF, the scale parameter is set to 10 for these images, which defines the width of the Gaussian functions used for the Frangi vessel enhancement [23]. This is the best choice for scale enhancement in these images, in order to enhance the most prominent vessels. The other level-set method tested is L2S. All three level-set methods are run for 200 iterations. Hierarchical image matting is tested as described in section 4.1.

The methods described above are run on each of the four datasets. Including creation of the matting initialization, LyMPhi run-time is approximately 2 h for a spinal lymphatics image with size 1500×2700 pixels, on a standard Windows \times 86 64-bit desktop with 8GB RAM. This is similar to the computing time for the other level-set methods tested. Each result was compared to either the manual majority voted annotation or the generated ground truth using the Sørensen–Dice coefficient to measure accuracy [34]. This measures the pixel overlap between the annotation and the experimental results (or for the synthetic case, ground

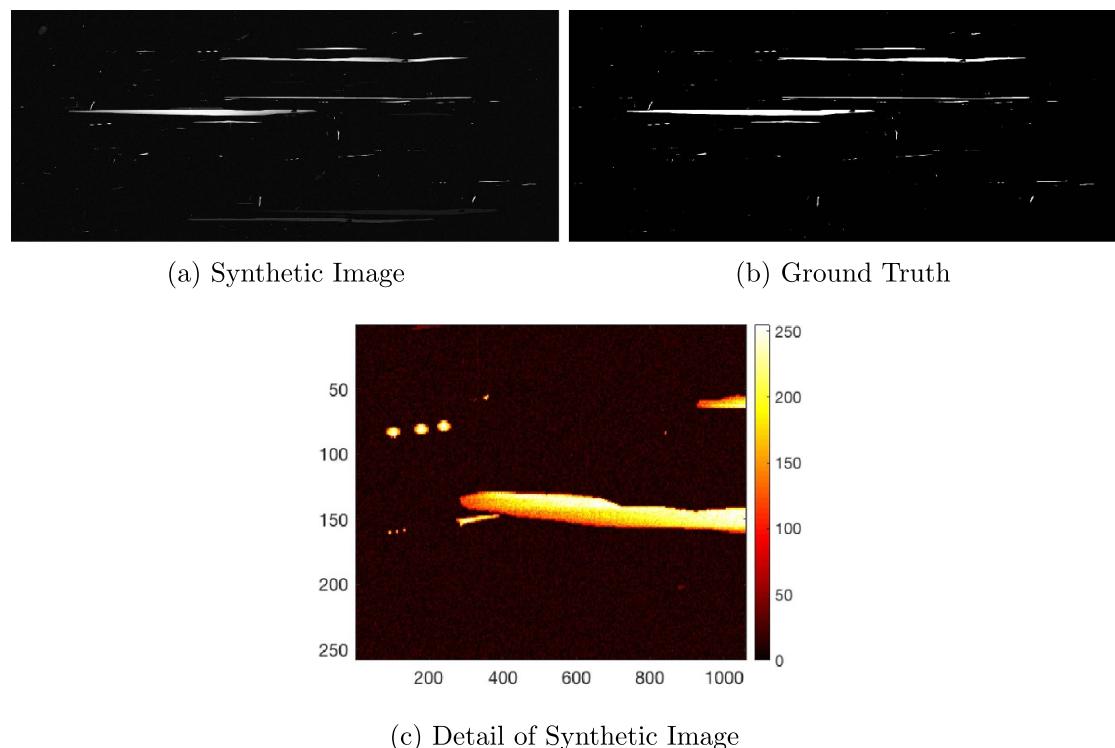


Figure 13. One example image from the synthetic dataset. The ground truth is shown on the right, and a close-up of the synthetic image in a red-based colormap in (c), in order to show the added noise, blur, intensity variation, and background clutter.

truth vs. experimental), and combines evaluation metrics such as sensitivity, specificity, and accuracy that are also often used for evaluation segmentation results [35]. The results are shown in table 1, with the Sørenson–Dice coefficient averaged over each dataset per method. The Dice coefficient has a range from zero to one, with one signifying a perfect match. The Dice coefficient is complemented by the Boundary F1-score, a measure computed from precision and recall on the segmented boundary [36]. This is denoted by *BF-score* in table 1. The Boundary F1-score also ranges from zero to one, one meaning a perfect boundary match. The BF-scores are also averaged over each dataset. Standard deviations for each dataset are provided in table 1.

For all four datasets, LyMPhi has the maximum average and median Dice coefficient measured from all four methods, as shown in tables 1 and 2. The median Dice coefficients are also shown as these values are less affected by outliers, such as images with overall weaker staining. TuFF performs well when vessel scales do not vary widely. However, TuFF performs poorly when the scale is highly variable: this method retains too little of the vessel in the segmentation result. As lymphatic vasculature varies so widely in size compared to neuron branches, the scale enhancement used in TuFF tends to discount larger more blob-like vessel regions. If a larger scale parameter is used, thinner vessel structures are omitted from the segmentation result. On synthetic data, any noise that is not eradicated in thresholding is magnified by TuFF, where the attraction force (even when used at a minimum) joins noisy elements together. Reduction of clutter appearing in the segmentation result, using existing techniques, inevitably removes vessel information that cannot always be fully recovered by region growing methods. L2S, when initialized finely, performs better, as on the spinal dataset. However, if initialization is coarser (and thus inaccurate), L2S retains much of the background clutter (because of the focus of the method on resolving high intensity variation within noisy foreground).

Mean BF-scores are also shown in table 1. Again, LyMPhi outperforms or has equivalent performance to the other methods tested. LyMPhi's BF scores are close to 1. When analyzing why the BF-score for LyMPhi is higher compared to the other methods, it is observed that LyMPhi has the highest precision values on the boundary for all datasets. This means there is a higher proportion of relevant boundary pixels returned by LyMPhi, because the other methods return too many positive pixels. Sometimes precision is more important than recall, which could be argued in this case. Precision is important for getting more accurate complexity measurements. Including too much background matter in the lymphatic segmentation could grossly overestimate the amount of lymphatics present.

Additionally, the standard deviation of Dice coefficient and BF-score is presented in table 1. LyMPhi has the lowest standard deviation for both Dice coefficient and BF-score. Alongside the average scores reported in table 1, the results show that LyMPhi is the most consistently high-performing method across all datasets.

Table 1. Average Dice coefficient and BF-score for all methods.

Method	Tuff		L2S		Mat.		LyMPhi	
	Dice	BF-score	Dice	BF-score	Dice	BF-score	Dice	BF-score
Partial	0.31 ± 0.15	0.43 ± 0.12	0.55 ± 0.18	0.58 ± 0.10	0.55 ± 0.10	0.60 ± 0.07	0.71 ± 0.10	0.84 ± 0.07
Whole	0.22 ± 0.10	0.14 ± 0.15	0.35 ± 0.13	0.48 ± 0.17	0.44 ± 0.10	0.74 ± 0.07	0.51 ± 0.10	0.86 ± 0.05
Spinal	0.39 ± 0.23	0.35 ± 0.27	0.59 ± 0.27	0.62 ± 0.28	0.61 ± 0.15	0.79 ± 0.18	0.67 ± 0.10	0.79 ± 0.14
Synthetic	0.06 ± 0.04	0.15 ± 0.08	0.58 ± 0.25	0.79 ± 0.20	0.81 ± 0.13	0.79 ± 0.12	0.89 ± 0.09	0.96 ± 0.05

Table 2. Median Dice coeff. for all methods.

Method	TuFF	L2S	Mat.	LyMPhi
Partial	0.30	0.59	0.59	0.74
Whole	0.20	0.34	0.48	0.56
Spinal	0.39	0.67	0.65	0.70
Synthetic	0.05	0.56	.91	0.91

For the spinal dataset, where hierarchical image matting also has a high BF-score, LyMPhi by comparison has a slightly lower standard deviation. One reason why hierarchical image matting performs well on the spinal dataset is because, in the spinal images, the vessel mass is larger than in meningeal images. This allows for more information when comparing unknown pixels to the higher intensity foreground—increasing the number of connected known foreground pixels improves the labeling of the unknown pixels. It is important to note, however, that the overall vessel boundary in the segmentation is not smooth as a result of image matting, and appears pixelated, even with these high resolution images. Additionally, although hierarchical image matting partially provides thin branches that are ignored by other methods, this inclusion comes at the expense of including irrelevant cell types.

The vessel size (area and diameter) extracted by LyMPhi most closely matches the manual annotation. TuFF results in an undersegmentation, or thinning of vessels, and L2S frequently oversegments, not adhering to the outer vessel boundary. Matting, while more correctly extracting vessel size, includes extraneous cellular matter, which obfuscates the measured amount of lymphatic vasculature. Thus, LyMPhi, due to its incorporation of matting, is the most robust method of the three for detecting thinning of meningeal lymphatic vessels, a precursor to AD.

Sample segmentation results are shown in figure 14. It can be seen that LyMPhi has captured less of the background clutter while preserving the thickness of the vessels. In figure 14, it is clear from the displayed results that LyMPhi has retained less of the clutter present at the top and bottom right of the image. LyMPhi has the most consistent results (maximum Dice coefficient) across all datasets.

The results are poorest, for all methods, on the whole-mount dataset (out of the real datasets) due to poor channel separation. In these images, the lymphatic vessels are present in the green channel, but the background clutter of blood vessels and endothelial cells also have strong representation in the green channel, as shown in figure 15. This leads to poor segmentation by all methods. However, LyMPhi still outperforms the others in this challenging circumstance.

6.1. Stability of LyMPhi

One important facet of LyMPhi is its stability and robustness to various starting points due to initialization. This robustness cannot be offered by other level-set methods tested on the lymphatic data. We have found that initialization plays an enormous part in the final accuracy of the results found by using L2S or TuFF on our data, and this initialization is difficult to tune from image to image. The staining varies in each vessel sample, leading to variable signal strength. Using too low an initial threshold retains too much clutter, while using too high a threshold removes part of the vessel that is lower in intensity. LyMPhi can segment lymphatic vessels even with a coarser initialization, which may be necessary to retain signal across a dataset.

The matting procedure has already produced a vessel foreground that has removed much of the background clutter, so the first iterations can quickly move the level set boundary close to the calculated vessel foreground. This is done by changing the force field, or velocity, of phi. Subsequent updates smooth the zero level set contour.

Two variants of initialization are shown in figure 16, with their respective segmentation results using TuFF and LyMPhi.

The TuFF result using the finer initialization indeed only captures a few elements of clutter. But the level set boundary tends to overall creep inside the true vessel boundary. This could be due to intensity inhomogeneity and lower signal strength at the vessel boundary, as well as difficulty tuning the vessel scale parameter. If the scale parameter is made larger to prevent such undersegmentation, this will lead to missing thinner vessel segments in the segmentation result. The TuFF result using coarser initialization picks up much more background clutter, as expected. The level set boundary has extended at more points outside the vessel boundary, leading to oversegmentation.

The LyMPhi result using the finer initialization still picks up some background clutter, with a bit more clutter retained when using the coarser initialization. However, the main advantage of LyMPhi is that it captures the true vessel width which TuFF cannot. Width is a measure of significant interest to neuroscientists, as they use it to quantify lymphatic presence.

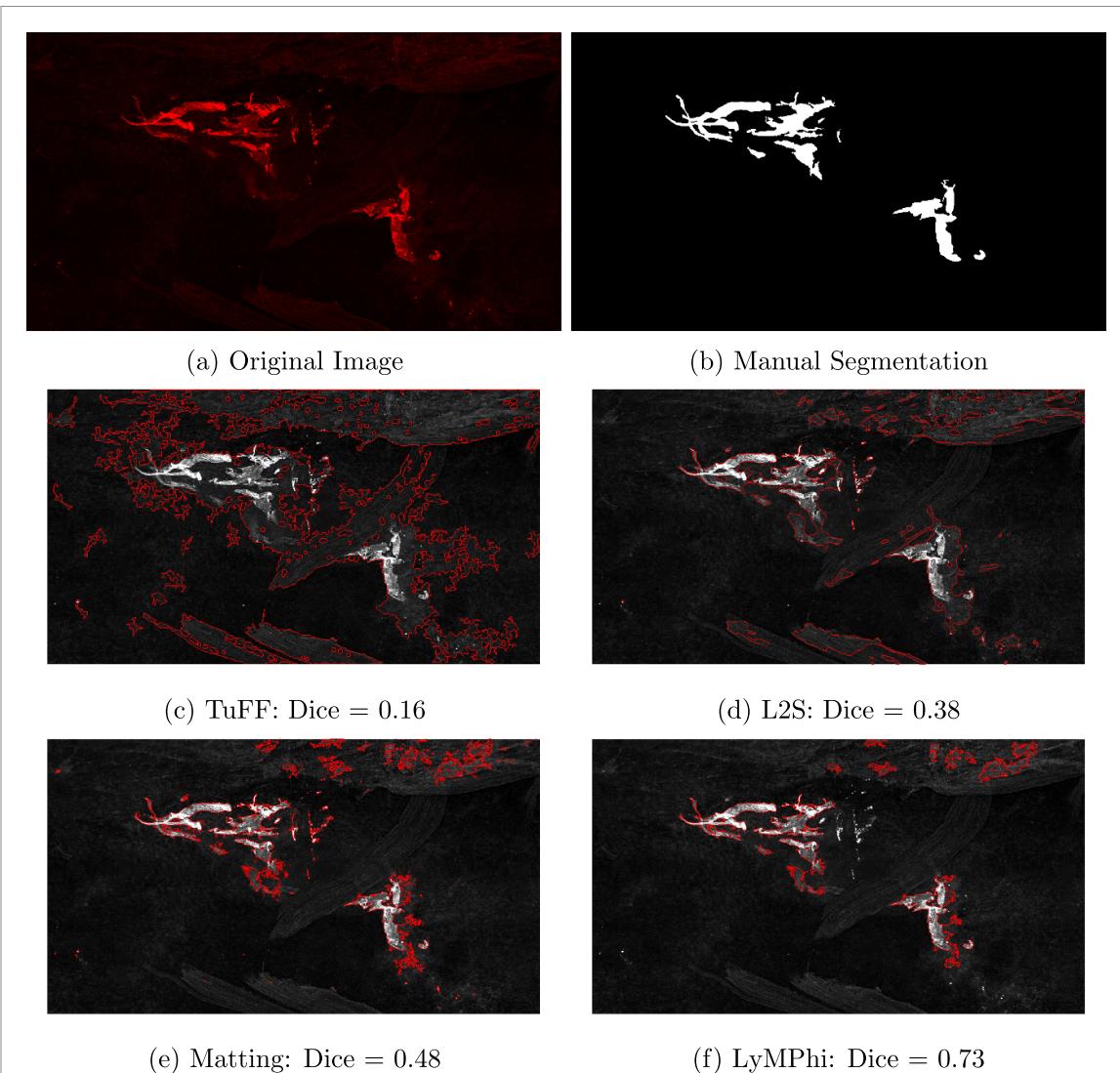


Figure 14. Segmentation results on one image from the spinal dataset. The results are displayed as gray-scale image data and overlays with the final segmentation contours in red. Captioned is the Dice score.

7. Analysis of manual segmentation

7.1. Analysis of agreement between annotators

The results shown prior only compare automated segmentation results between different segmentation algorithms. Since the manual annotation obtained is combined input from three annotators, in this section the agreement and kappa statistics between the annotators is compared, as well as the statistics between the chosen segmentation algorithms and the annotations.

Fleiss's Kappa is used for comparing the three annotations created by three separate operators for each vessel image. Fleiss's Kappa measures the reliability of the agreement between the three annotators in this case [37]. Instead of creating a voting matrix and storing annotator labels for each pixel in a dataset, which would require a matrix of extremely large size, running counts were kept of overall voting proportions and probabilities. Percent agreement and Fleiss's Kappa are shown in table 3.

The Fleiss's Kappa values between annotators are between 0.6 and 0.75, indicating moderate agreement between annotators on where vessel is present. For the spinal dataset, the Kappa value is highest, indicating better agreement than for the other two datasets. However, none of the Kappa values have reached .8 or .9, which are more preferred Kappa values—showing that there is still considerable disagreement between annotators, and that creating accurate ground truth for these images is not a trivial problem [38].

Agreement and the kappa statistic are calculated for all images in each dataset, between the segmentation algorithm and the majority voted groundtruth. Cohen's Kappa is used in this case because there are two raters being compared, the automated segmentation algorithm in question, and the consensus segmentation. Cohen's Kappa is considered to be more reliable of a measure than percent agreement, as the Kappa takes

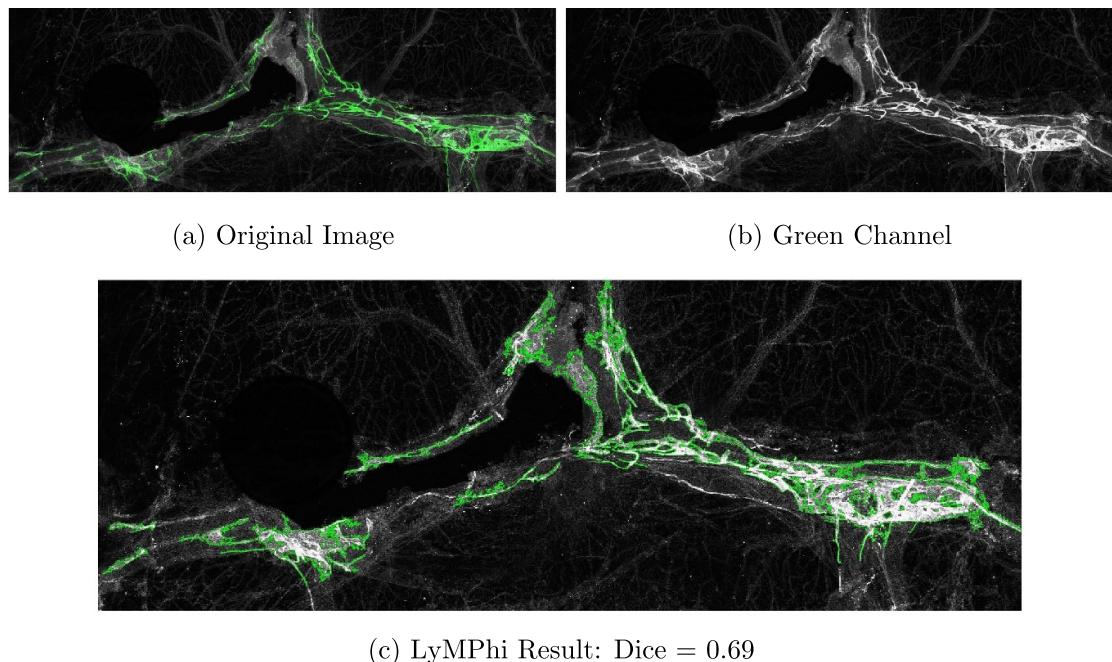


Figure 15. Segmentation results on one image from the whole-mount dataset. The results are displayed as gray-scale image data and overlays with the final segmentation contours in green.

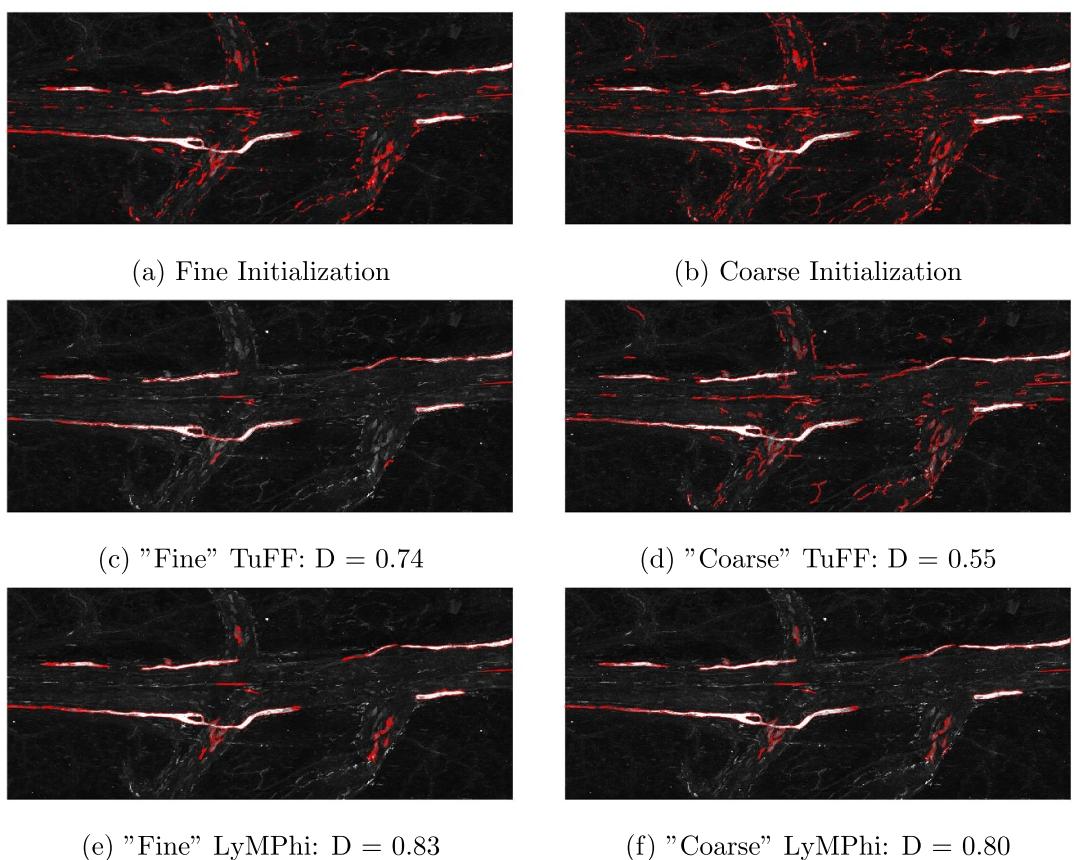


Figure 16. The first row shows two initial segmentation images of a partial mount image: (a) is fine-grained initialization, and (b) is a rougher thresholding keeping much of the background clutter. The second row shows two TuFF results after 200 iterations, using each initialization. The third row shows two LyMPhi results after 200 iterations. The respective Dice coefficient (denoted with 'D') is displayed underneath each result. There is a marked decrease in Dice coefficient of almost 0.2 depending on the initialization used, whereas LyMPhi changes in Dice coefficient by only 0.03 even when the initialization becomes much coarser.

Table 3. Agreement and Fleiss's Kappa between annotators.

Dataset	Agreement	Kappa
Partial	0.98	0.62
Whole	0.97	0.60
Spinal	0.96	0.75

Table 4. Agreement and Cohen's Kappa between algorithms and annotators.

Method	TuFF		L2S		Mat.		LyMPhi	
	Statistic	Agr.	Kap.	Agr.	Kap.	Agr.	Kap.	Agr.
Partial	0.87	0.23	0.94	0.44	0.97	0.56	0.98	0.70
Whole	0.68	0.14	0.84	0.28	0.93	0.43	0.95	0.52
Spinal	0.65	0.19	0.84	0.39	0.96	0.64	0.96	0.64

into consideration the probability of random agreement [39]. Percent agreement is a simple measure of true positive and true negative rate, unlike Dice coefficient or Boundary F1-score, which also consider false positive and false negative rate. Percent agreement and Cohen's Kappa are provided in table 4. Agreement is denoted by *Agr.* and Kappa by *Kap.*

From table 4, agreement between LyMPhi and annotators largely outperforms inter-annotator agreement in the partial mount dataset. Thus, for this dataset, it can be stated that LyMPhi is superior to manual segmentation. In the other datasets, the Kappa value is lower than the Kappa values shown in table 3, meaning that LyMPhi has not yet surpassed hand annotation for these datasets. However, the Kappa values for LyMPhi do outperform the agreement between annotation and the other segmentation algorithms tested. This further affirms that LyMPhi is better suited for segmentation of meningeal lymphatic vessels than other automated methods.

To further research how the resolution of the image affects the result of manual segmentation, the following details are provided. In order of highest resolution to lowest, the datasets are as follows: partial-mount, whole-mount, and spinal. Based on Fleiss's Kappa in table 3, the agreement, and Kappa values, between annotators does not have a linear relationship with resolution. The highest resolution images only have the median Kappa value of 0.62.

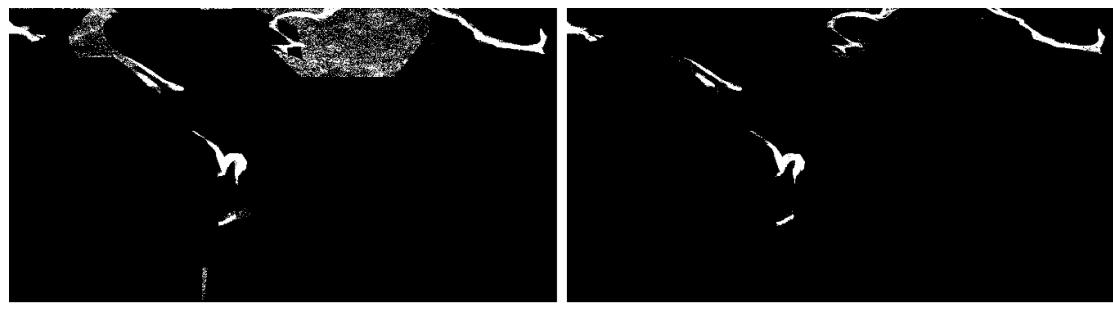
More than image resolution, image contrast and image noise or artifacts have a greater impact on the manual annotation reliability. From an imaging standpoint, if staining can be improved and the presence of artifacts can be decreased, this will greatly improve the quality of the majority voted annotation, as well as the automated segmentation results.

7.2. Performance of LyMPhi and each annotator as measured by STAPLE

An example of where disagreement can occur is shown in figure 17. When combining three annotations using the STAPLE algorithm [40], noisy regions, such as in the top of the image in figure 17(a) can still remain. The images are difficult to annotate due to low-level noise that must be removed. Due to the low image contrast, patches of noise may be missed by an individual annotator. Noise and other challenges lead to annotator disagreement. Using majority voting, as shown in figure 17(c), can reduce the effects of the disagreement due to noisy background; however, some small amounts of noise do remain, which create divergence when comparing with LyMPhi and other automated algorithms.

Although we use majority voted manual annotation for comparison instead of the STAPLE consensus segmentation, using the STAPLE algorithm can also provide an estimate of sensitivity and specificity of raters [40]. In table 5, the sensitivity and specificity of the three annotators as well as LyMPhi are compared. Sensitivity refers to the number of true positives that are measured, and specificity the number of true negatives [41]. This is in comparison to the STAPLE estimated consensus segmentation, which has been shown to have some flaws—so these sensitivity/specificity measurements are also only estimates.

Specificity is so high across all raters and datasets because the number of negatives, or non-lymphatic pixels, is so large, and it is easier to mark these correctly. High sensitivity, however, is more difficult to achieve. LyMPhi outperforms on sensitivity for the partial mount dataset, which is in accordance with the discussion of Kappa values in section 7.1. This means that LyMPhi is better at distinguishing the lymphatic vessels than either of the raters individually, or when they are considered together. Although this is not the case for the remaining two datasets, it is interesting to note that for the whole mount dataset, the sensitivity value for Rater 2 is lower than that of LyMPhi (0.56 compared to 0.74). So, in cases where multiple annotators cannot be reached to perform a consensus segmentation, there is great risk relying on a single



(a) Consensus Segmentation via STAPLE (b) Consensus Segmentation via Majority

Figure 17. Consensus segmentation results on a portion of an image from the spinal dataset. The results are displayed as binary. The left image shows a consensus segmentation obtained via the STAPLE algorithm. The right image is the manual annotation we used for comparison, obtained by majority voting.

Table 5. Sensitivity and specificity.

Dataset	Partial		Whole		Spinal	
	Annotator	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity
Annotator 1	0.77	0.99	0.91	0.98	0.82	0.99
Annotator 2	0.77	0.99	0.56	0.99	0.96	0.97
Annotator 3	0.76	0.98	0.81	0.98	0.93	0.99
LyMPhi	0.80	0.99	0.74	0.98	0.60	0.99

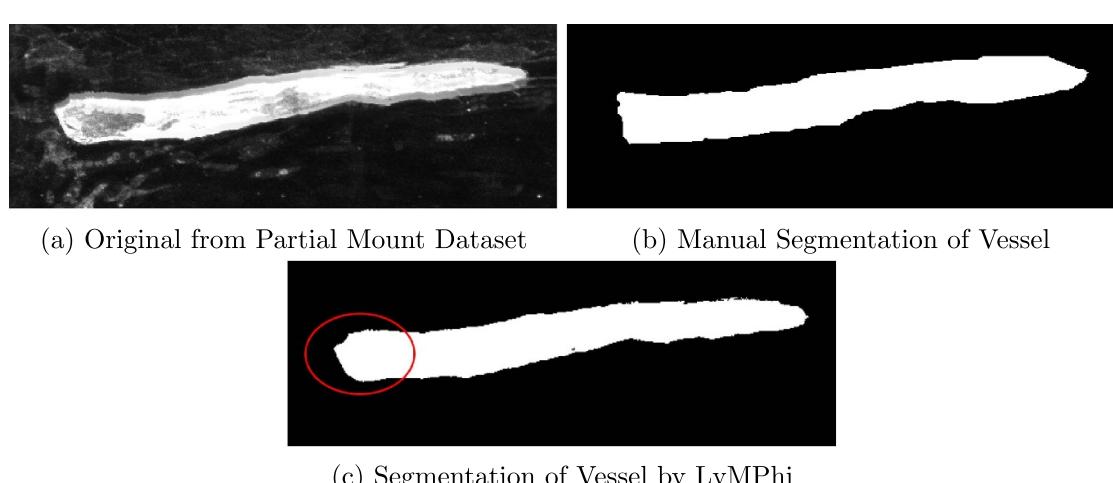


Figure 18. Consensus segmentation results on a portion of an image from the partial mount dataset. The segmentation results are displayed as binary. The top left image shows the original vessel from the cropped image. The top right image is the manual annotation, obtained by majority voting. The bottom image is segmentation performed by LyMPhi (Dice = 0.94).

annotator for a correct segmentation of the lymphatic vessels. A single annotator is the current standard for MIV segmentation, the dangers of which have been shown in figure 1 as well as in the above sensitivity analysis. As mentioned before, manual annotation takes a great deal of time; a more accurate segmentation may be found in less time by using LyMPhi, instead of relying on a potentially flawed hand segmentation.

7.3. Failure to annotate boundary by manual segmentation

Combined errors of sensitivity and specificity at the fine-grained level along the vessel boundary shows that our level-set method, LyMPhi, better captures the true vessel boundary than manual segmentation. This can be shown in figure 18, where the red circle in figure 18(c) shows that the boundary found by LyMPhi better matches that in figure 18(a)—the boundary in figure 18(b) has been flattened and does not provide accurate curvature of the vessel edge.

8. Conclusion

The main contributions of this paper can be summarized as follows. First, a matting based level-set segmentation approach, LyMPhi, is proposed to obtain robust lymphatic vessel segmentation by removing background clutter and retaining vessel smoothness. Essentially, LyMPhi has enabled the segmentation of a recently discovered anatomy (lymph in the brain), allowing quantification of the delicate vessels that vary in width and intensity. Second, LyMPhi is automated, unlike the current manual segmentation used by neuroscientists, or traditional level-set segmentation procedures, which may require difficult-to-tune intensity and scale thresholds. Third, extensive experiments are conducted on four types of lymphatic vessel datasets to validate the performance of LyMPhi compared to other state-of-the-art segmentation algorithms. A novel approach to synthetic data augmentation is proposed, based on shape deformation of the real lymphatic vessels. Segmentation output from LyMPhi has a higher overall Dice coefficient compared to that of competing algorithms as well as a higher BF-score, whilst being stable under different initial conditions. Finally, an in-depth analysis is performed of the manual annotation, showing the fallacy in holding manual segmentation as the gold standard in lymphatic analysis.

In the long-term, studying segmented vessels may lead to future discoveries on the role our meningeal lymphatic system plays in diseases of the central nervous system. Complexity measures built on segmentation can lead to fundamental understanding of these vessels: how they grow, decline, and ultimately, the intricacies of their function.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <http://dx.doi.org/10.17632/fcxzbpt3s.1>.

Acknowledgment

This work was partially supported by a Double Hoo Award from the Office of Undergraduate Research (OUR) at the University of Virginia.

Conflict of interest

The authors declare no competing financial interest.

ORCID iD

N Tabassum  <https://orcid.org/0000-0003-4390-7000>

References

- [1] Sweeney M D and Zlokovic B V 2018 Lymphatic waste disposal in the brain *Nature* **560** 172–4
- [2] Da Mesquita S *et al* 2018 Functional aspects of meningeal lymphatics in ageing and Alzheimer's disease *Nature* **560** 185
- [3] Barney J 2018 UVA brain discovery could block aging's terrible toll on the mind *UVAToday*
- [4] Barney J 2018 UVA identifies brain's lymphatic vessels as new avenue to treat multiple sclerosis *UVAToday*
- [5] Louveau A *et al* 2018 CNS lymphatic drainage and neuroinflammation are regulated by meningeal lymphatic vasculature *Nat. Neurosci.* **21** 1380
- [6] Louveau A, Harris T H and Kipnis J 2015 Revisiting the mechanisms of CNS immune privilege *Trends Immunol.* **36** 569–77
- [7] Frederick N and Louveau A 2020 Meningeal lymphatics, immunity and neuroinflammation *Curr. Opin. Neurobiol.* **62** 41–7
- [8] Mascagni P and Sanctius C 1787 *Vasorum Lymphaticorum Corporis Humani Historia et Ichnographia* (Sienna, Italy: Ex typographia Pazzini Carli)
- [9] Antila S *et al* 2017 Development and plasticity of meningeal lymphatic vessels *J. Exp. Med.* **214** 3645–67
- [10] Ahn J H *et al* 2019 Meningeal lymphatic vessels at the skull base drain cerebrospinal fluid *Nature* **572** 62–6
- [11] Hsu M, Rayasam A, Kijak J A, Choi Y H, Harding J S, Marcus S A, Karpus W J, Sandor M and Fabry Z 2019 Neuroinflammation-induced lymphangiogenesis near the cribriform plate contributes to drainage of CNS-derived antigens and immune cells *Nat. Commun.* **10** 1–14
- [12] Osher S and Sethian J A 1988 Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations *J. Comput. Phys.* **79** 12–49
- [13] Acton S T and Ray N 2009 *Biomedical Image Analysis: Segmentation* (*Biomedical Image Analysis: Segmentation*) vol 4 (Williston, VT: Morgan & Claypool Publishers) pp 1–108
- [14] Caselles V, Kimmel R and Sapiro G 1997 Geodesic active contours *Int. J. Comput. Vis.* **22** 61–79
- [15] Deng J and Tsui H-T 2002 A fast level set method for segmentation of low contrast noisy biomedical images *Pattern Recognit. Lett.* **23** 161–9
- [16] Zhao Y Q, Wang X H, Wang X F and Shih F Y 2014 Retinal vessels segmentation based on level set and region growing *Pattern Recognit.* **47** 2437–46

[17] Li C, Kao C-Y, Gore J C and Ding Z 2008 Minimization of region-scalable fitting energy for image segmentation *IEEE Trans. Image Process.* **17** 1940–9

[18] Mukherjee S and Acton S T 2015 Region based segmentation in presence of intensity inhomogeneity using legendre polynomials *IEEE Signal Proc. Lett.* **22** 298–302

[19] Mukherjee S, Condron B and Acton S T 2015 Tubularity flow field—a technique for automatic neuron segmentation *IEEE Trans. Image Process.* **24** 374–89

[20] Fan Z, Lu J, Wei C, Huang H, Cai X and Chen X 2019 A hierarchical image matting model for blood vessel segmentation in fundus images *IEEE Trans. Image Process.* **28** 2367–77

[21] Otsu N 1979 A threshold selection method from gray-level histograms *IEEE Trans. Syst. Man Cybern.* **9** 62–6

[22] Rapaka S and Kumar P R 2018 Efficient approach for non-ideal iris segmentation using improved particle swarm optimisation-based multilevel thresholding and geodesic active contours *IET Image Process.* **12** 1721–9

[23] Frangi A F, Niessen W J, Vincken K L and Viergever M A 1998 Multiscale vessel enhancement filtering *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 130–7

[24] Li B and Acton S T 2007 Active contour external force using vector field convolution for image segmentation *IEEE Trans. Image Process.* **16** 2096–106

[25] Chan T F and Vese L A 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77

[26] Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234–41

[27] Klassen E, Srivastava A, Mio M and Joshi S H 2004 Analysis of planar shapes using geodesic paths on shape spaces *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 372–83

[28] Wang J, Fu Z, Sadeghzadehyazdi N, Kipnis J and Acton S T 2018 Nonlinear shape regression for filtering segmentation results from calcium imaging *2018 25th IEEE Int. Conf. on Image Processing (ICIP)* (IEEE) pp 738–42

[29] Srivastava A, Klassen E, Joshi S H and Jermyn I H 2010 Shape analysis of elastic curves in Euclidean spaces *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 1415–28

[30] Batabyal T, Condron B and Acton S T 2020 NeuroPath2Path: classification and elastic morphing between neuronal arbors using path-wise similarity *Neuroinformatics* **18** 1–30

[31] Telea A, Sminchisescu C and Dickinson S 2004 Optimal inference for hierarchical skeleton abstraction *Proc. 17th Int. Conf. on Pattern Recognition 2004 (ICPR 2004)* vol 4 (IEEE) pp 19–22

[32] Zhang M, Zhang J, Wang Y, Wang J, Achimovich A M, Acton S T and Gahlmann A 2020 Non-invasive single-cell morphometry in living bacterial biofilms *bioRxiv Preprint*

[33] Spring K R, Fellers T J and Davidson M W 2020 Confocal microscopy—signal-to-noise considerations

[34] Sørensen T 1948 A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons *Biol. Skr.* **5** 1–34

[35] Yan Z, Yang X and Cheng K-T 2018 A three-stage deep learning model for accurate retinal vessel segmentation *IEEE J. Biomed. Health Inform.* **23** 1427–36

[36] Csurka G, Larlus D, Perronnin F and Meylan F 2013 What is a good evaluation measure for semantic segmentation? *BMVC* vol 27 p 2013

[37] Fleiss J L and Cohen J 1973 The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability *Educ. Psychol. Meas.* **33** 613–9

[38] Richard Landis J and Koch G G 1977 The measurement of observer agreement for categorical data *Biometrics* **33** 159–74

[39] McHugh M L 2012 Interrater reliability: the kappa statistic *Biochem. Med.* **22** 276–82

[40] Warfield S K, Zou K H and Wells W M 2004 Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation *IEEE Trans. Med. Imaging* **23** 903–21

[41] Yerushalmy J 1947 Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques *Public Health Rep.* **62** 1432–49