frontiers
in Computer Science

Analysis of Video Retinal Angiography With Deep Learning and Eulerian Magnification

Sumit Laha[1], Rodney LaLonde[1], Austin E. Carmack[2], Hassan Foroosh[1], John C. Olson[3], Saad Shaikh[2] and Ulas Bagci[1]*

[1] Department of Computer Science, University of Central Florida, Orlando, FL, United States, [2] College of Medicine, University of Central Florida, Orlando, FL, United States, [3] Central Florida Retina, Orlando, FL, United States

Objective: The aim of this research is to present a novel computer-aided decision support tool in analyzing, quantifying, and evaluating the retinal blood vessel structure from fluorescein angiogram (FA) videos.

Methods: The proposed method consists of three phases: (i) image registration for large motion removal from fluorescein angiogram videos, followed by (ii) retinal vessel segmentation, and lastly, (iii) segmentation-guided video magnification. In the image registration phase, individual frames of the video are spatiotemporally aligned using a novel wavelet-based registration approach to compensate for the global camera and patient motion. In the second phase, a capsule-based neural network architecture is employed to perform the segmentation of retinal vessels for the first time in the literature. In the final phase, a segmentation-guided Eulerian video magnification is proposed for magnifying subtle changes in the retinal video produced by blood flow through the retinal vessels. The magnification is applied only to the segmented vessels, as determined by the capsule network. This minimizes the high levels of noise present in these videos and maximizes useful information, enabling ophthalmologists to more easily identify potential regions of pathology.

Results: The collected fluorescein angiogram video dataset consists of 1, 402 frames from 10 normal subjects (prospective study). Experimental results for retinal vessel segmentation show that the capsule-based algorithm outperforms a state-of-the-art convolutional neural networks (*U-Net*), obtaining a higher dice coefficient (85.94%) and sensitivity (92.36%) while using just 5% of the network parameters. Qualitative analysis of these videos was performed after the final phase by expert ophthalmologists, supporting the claim that artificial intelligence assisted decision support tool can be helpful for providing a better analysis of blood flow dynamics.

Conclusions: The authors introduce a novel computational tool, combining a wavelet-based video registration method with a deep learning capsule-based retinal vessel segmentation algorithm and a Eulerian video magnification technique to quantitatively and qualitatively analyze FA videos. To authors' best knowledge, this is the first-ever development of such a computational tool to assist ophthalmologists with analyzing blood flow in FA videos.

Keywords: capsule networks, retinal vessel segmentation, Eulerian video magnification, haar wavelets registration, fluorescein angiogram

1. INTRODUCTION

Fluorescein angiography (FA) is a diagnostic imaging technique that aids ophthalmologists in diagnosing, treating, and monitoring retinal vascular pathology (Almotiri et al., 2018). In this procedure, a fluorescent dye is injected into the patient's bloodstream to highlight the blood vessels in the retina.

Ophthalmologists diagnose abnormalities in the retina and its vasculature using FA images and videos, which are captured by a fundus camera. A typical fundus camera is shown in **Figure 1**. The fundus camera can be considered a low power microscope for the purpose of capturing retinal images. Special excitation and barrier filters capture images of the fluorescent dye as it traverses through the patients' vasculature. The passage of dye through the vasculature from the choroidal vessels, to the retinal arteries and veins, corresponds to different phases of the retinal circulation, from which retinal vascular flow is approximated by the treating physician. An example FA video showing select frames is shown in **Figure 2** depicting this retinal vascular flow. Blockages or leakage of the dye into the retinal neural tissue are indicative of retinal pathology. Some of the more common retinal pathologies diagnosed using FA include diabetic retinopathy, age-related macular degeneration (AMD), retinal tumors, and other retinal vascular diseases. **Figure 3** shows a typical FA of a diabetic patient where many such abnormalities can be seen.
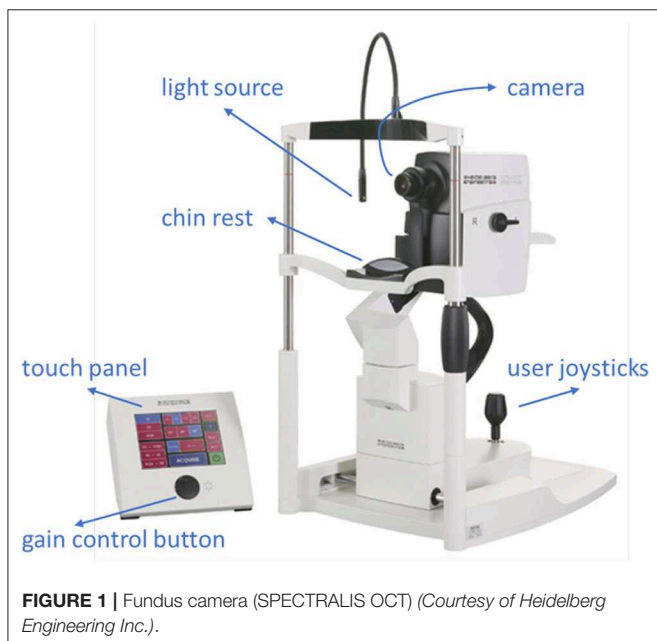
Currently, physicians approximate vascular transit times by "eyeballing" FA videos to determine retinal flow characteristics, and subsequently detect disease. Although generally effective, this approach is typically unreliable due to high inter- and intra-operator variability and requires extensive training in the interpretation of retinal angiographic images with clinical disease correlation. Approximation errors and oversights can lead to misinterpretation with significant visual consequences. To the best of our knowledge, there is no existing system for the automated analysis of retinal vascular flow in retinal angiography. Some critical challenges in automated FA video analysis exist, including motion, noise, the complex shape, and nature of retinal vessels, lack of validated ground truths (typically required for training supervised machine learning algorithms), and difficulty in accurately measuring blood flow. Specifically, one of the biggest challenges in FA analysis is the inevitable motion between frames of the FA videos. Because the camera is hand-operated, significant motion artifacts are typically present. This motion causes the frames to become translated, rotated, and/or skewed, relative to one another. Another significant challenge is noise, where most of the areas in FA video frames are extremely noisy. To develop an image analysis tool, such noise should be carefully minimized, especially when identifying the vessel structures for quantification purposes. **Figure 4** displays two consecutive frames of an FA video for a normal subject. The highlighted region shows red and green vessels with significant motion that caused non-overlapping vessels between subsequent frames. Additionally, in non-vessel image areas, the large amount of noise present.
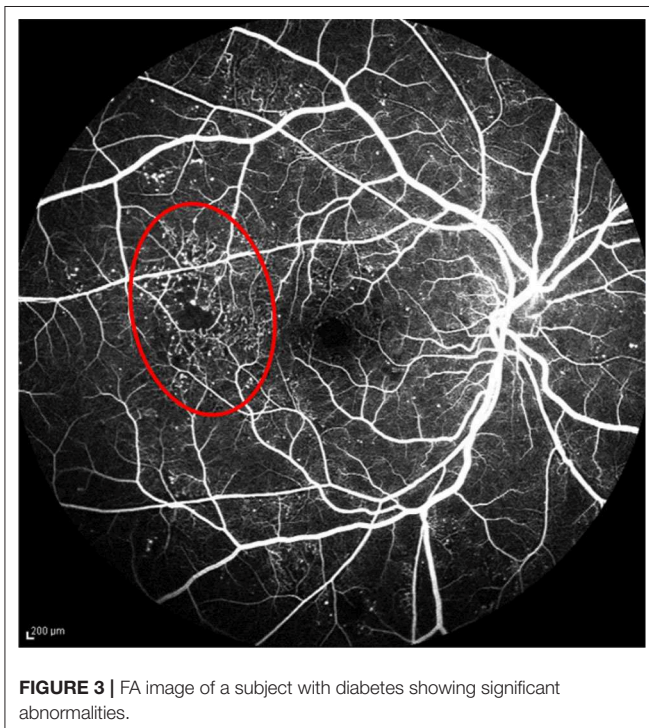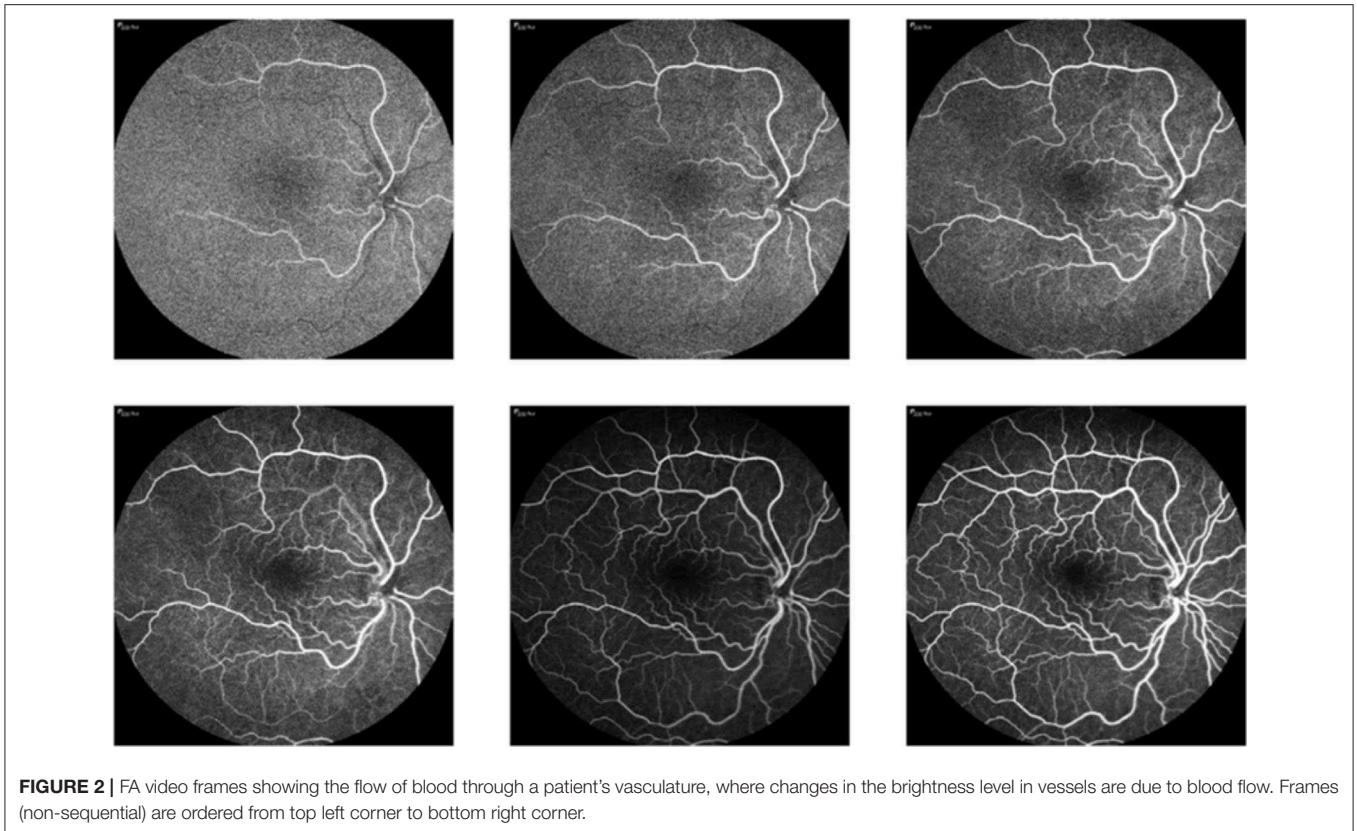
The arboriform structure of the retinal vasculature provides another major hurdle, similar to the tree-like appearances found in the bronchial airways (Xu et al., 2015). Due to the irregular structures within the same retinal fundus image, accurate segmentation at different scales is harder than for closed objects at the same resolution settings. This has been a well-known problem in computer vision for decades and although numerous multi-scale retinal segmentation methods exist in the literature, this problem is not yet globally solved. Other factors that make retinal vessel segmentation a challenging task are the varying size, shape, angles and branching patterns of vessels, non-uniform illumination, and anatomical variability between the subjects (Srinidhi et al., 2017). While these challenges to creating an automated analysis system are interesting, there are additional factors which motivate the proposed study. The spatiotemporal sensitivity of the human visual system is limited, and many signals fall below our ability to detect. It has been shown, these signals can be digitally enhanced to reveal important features that are otherwise not readily apparent. For instance, a slight variation in the redness of the skin due to blood flow can be amplified to extract a person's pulse rate (Poh et al., 2010), or the subtle motion of any structure or component can be magnified to reveal the mechanical behavior of materials (Liu et al., 2005). Recently, Wu et al. (2012) conducted a study to visualize blood flow in the face through a similar method of amplifying small, usually imperceptible, motion, and color variations in several real-world videos.

Although these studies have described the application of different approaches to reveal "invisible" signals in the video, the magnification of motion for possible clinical application has yet to be explored (particularly in FA videos). To analyze blood circulation, we propose a segmentation-constrained Eulerian video magnification method to obtain the subtle flow characteristics of retinal blood vessels, for the sake of aiding disease detection.

The structure of the remainder of this paper is as follows: section 2 reviews the existing literature; section 3 describes

FIGURE 1 | Fundus camera (SPECTRALIS OCT) *(Courtesy of Heidelberg Engineering Inc.).*

FIGURE 2 | FA video frames showing the flow of blood through a patient's vasculature, where changes in the brightness level in vessels are due to blood flow. Frames (non-sequential) are ordered from top left corner to bottom right corner.

FIGURE 3 | FA image of a subject with diabetes showing significant abnormalities.

our proposed approach; section 4 reports the results of the computational experimentation; and finally, section 5 summarizes our findings and proposes future research directions.

2. BACKGROUND AND RELATED WORKS

This section reviews previous studies in retinal image registration, retinal blood vessel segmentation, and Eulerian video magnification. A brief discussion of the main contributions and known shortcomings of these works is provided, and, in the end, we summarize our proposed innovations to overcome these shortcomings.

2.1. Retinal Image Registration

Retinal image registration deals with the process of constructing a pixel-to-pixel correspondence between two retinal images. Historically, the study of image registration has revealed several challenges, including variations frequently encountered in the following: subject-eye positions and movements, camera internal parameters, modalities of images, and retinal tissues in two retinal images of the same subject. Moreover, due to the limitation in image capture angle, smaller image regions are frequently considered instead of whole fundus images; thus, a sufficient number of small image regions are required to obtain full information of eye diseases (Guo et al., 2017). Due to these above difficulties in the alignment process, the problem of retinal image registration has been a major area of research.

Generally, retinal image registration can be categorized into feature-based, gradient-based, and correlation-based methods (Guo et al., 2017). Feature-based methods use either manual or automatic selection of some definite objects like edges and corners to estimate the transformation of the pair of images. On
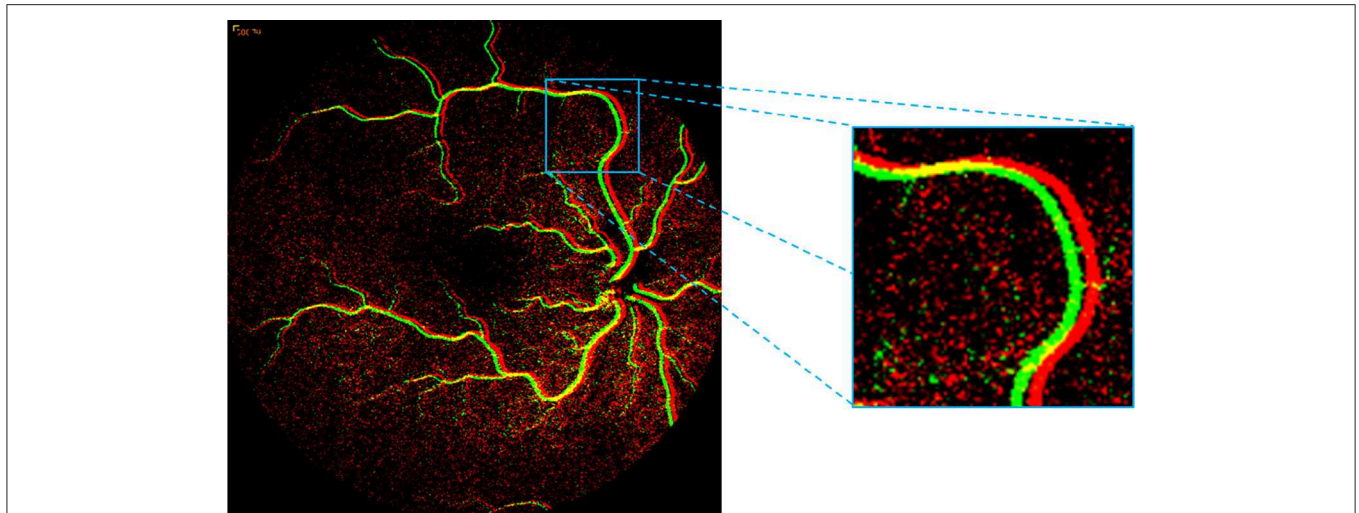
FIGURE 4 | Presence of motion highlighted within the square region by overlaying two successive FA frames. The high-level of noise can be seen in red and green dots in non-vessel image areas.

the other hand, gradient-based methods rely on the estimation of translation parameters following some linear differential equations. Whereas, correlation methods make use of translation from the normalized spectrum to estimate translation, scaling, and rotation using Fourier representation. For the last few decades, feature-based registration has been one of the most used in the literature with methods falling into one of the following subcategories: region-matching, point-matching, or structure-matching. Chen et al. (2015) proposed a retinal image registration approach based on a global graph-based segmentation, which utilized a topological vascular tree and bifurcation structure comprised of a master bifurcation point and its three connected neighboring pixels. Laliberté et al. (2003) presented a global point-matching-based method, which considered a particular point matching search, utilizing the local structural information of the retina. Ghassabi et al. (2016) implemented a structure-based region detector on a robust watershed segmentation for high-resolution and low-contrast retinal image registration. Kolar et al. (2016) presented a two-phase based approach for the rapid retinal image registration. In the first phase, they applied phase correlation to remove large eye movements and in the second phase, Lucas-Kanade tracking with the help of adaptive robust selection of tracking points.

Ramaswamy et al. (2014) successfully implemented the phase correlation method for removal of residual fundus image shift and rotation obtained from a non-scanning retinal device. They noted from the experiment was that the phase correlation approach could accomplish promising results with respect to retinal nerve fiber sharpness measured by texture analysis. Matsopoulos et al. (1999) applied an automatic global optimization scheme by combining affine-, bilinear-, and projective-based transformation models with different optimization techniques, such as a simplex method, simulated annealing, and a genetic algorithm. In the area of self-organizing maps, which is an unsupervised neural network, Matsopoulos

et al. (2004) implemented a successful multimodal retinal image registration method. However, manual registration remains the standard in clinical practices (Karali et al., 2004).

In this work, we propose a novel wavelet-based registration method to spatiotemporally align FA video frames to a central reference frame to compensate the global camera and patient motion. While being more robust to noise than other relevant algorithms defined above, the presented registration algorithm also provides better performance compared with the state-of-the-art registration algorithms in this application domain.

2.2. Retinal Image Segmentation

Over the last two decades, researchers have developed many different retinal vessel segmentation methods (Singh et al., 2015; Zhang et al., 2015; Liskowski and Krawiec, 2016; Maji et al., 2016; Almotiri et al., 2018). Recent surveys have chronicled the approaches to blood vessel segmentation in retinal fundus images (Fraz et al., 2012; Srinidhi et al., 2017; Almotiri et al., 2018), while the remainder of this section will highlight some of the more prominent of these methods.

In kernel-based modeling, Singh et al. (2015) presented a segmentation technique based on a combination of local entropy thresholding and a Gaussian matched filter for delineating the retinal vascular structure. The authors modified the parameters of the Gaussian function to improve the overall performance of the method for identifying both thin and large blood vessel segments on the DRIVE dataset (Staal et al., 2004). Kumar et al. (2016) proposed an algorithm considering two-dimensional kernels and the Laplacian of Gaussian (LoG) filter for detecting the retinal vascular structure. In their algorithm, the authors apply two-dimensional matched filters with LoG kernel functions. Kar and Maity (2016) developed an approach combined with curvelet transformation, matched filtering, and Laplacian Gaussian filter. Their experimental results reveal that the method performed well on both pathological and noisy

retinal images. However, matched kernels respond strongly to both vessel and non-vessel structures, leading to false-positive errors.

Considering model-based methods, Zhang et al. (2015) proposed a snake-based segmentation technique in the category of parametric modeling, to fit a parameterized curve to vessel boundaries in retinal images. However, snake-based techniques suffer from drawbacks due to not having the flexibility required to represent the complex and diverse arboriform structures present in the retinal vasculature. Additionally, snakes can also fail to converge to the correct vessel edges in the presence of high noise, empty blood vessels, and low contrast levels. Gongt et al. (2015) proposed a novel level set technique in the category of geometric deformable models, based on local cluster values through bias correction. This method provides more local intensity information at the pixel level and does not require initialization of the level set function. The method by Rabbani et al. (2015) is one of the few applied to FA images, where they determine vessel regions for 24 subjects by subtracting other fluorescing features then applying an opening operation utilizing a disk-shaped structuring element with a radius of 2 pixels followed by eroding the image utilizing disk-shaped structuring elements with radii of 5 and 3 pixels. This work then initializes an active contour method based on gray-level values in order to segment leakages and obtains a Dice score of 0.86%. However, the weakness of model-based techniques, either parametric or geometric, is that they require the generation of a set of seed points during initialization steps.

A mathematical morphology-based technique was applied in Jiang et al. (2017) using global thresholding operations. This technique was tested and validated on the publicly available DRIVE and STARE data sets (McCormick and Goldbaum, 1975). By implementing the good features of both matched filtering and morphological processing, this methodology not only achieves higher accuracy and better robustness but also require less computational effort. The main drawback of this method is that it fails to model the highly curved vessels, which are mostly seen in young individuals.

Imani et al. (2015) presented an approach based on morphological component analysis combined with Morlet wavelet transform to overcome the traditional problem of vessel segmentation in the presence of some abnormalities. The purpose of this method was to divide vessels from other lesion structures which plays vital role in clinical settings for the assessment of abnormal situations.

Multi-scale approaches deal with the property of multiple scales and orientations appearing in the vessel structures. Christodoulidis et al. (2016) proposed a scheme based on multi-scale tensor voting (MSTV) to improve the segmentation of small/thin vessels. Their technique consists of four major phases: pre-processing, multi-scale vessel enhancement, adaptive thresholding, and MSTV processing, and post-processing. The adaptive thresholding is used to segment large- and medium-sized vessel structures, whereas, the MSTV is applied for small-sized vessels. These multi-scale approaches, however, have some drawbacks, such as tending to over-segment large vessels and under-segmenting thin vessels. Similarly, Mapayi et al.

(2015) also proposed a local adaptive thresholding method, this one based on a gray level co-occurrence matrix energy transformation, to attempt to segment both large and thin retinal blood vessels.

Nguyen et al. (2013) modified the basic concept of multi-scale line detection in their approach for the retinal vessel segmentation to show improved segmentation, especially, near two closely parallel vessels and at crossover points.

Sharma and Wasson (2015) developed a fuzzy logic-based vessel segmentation method. The authors used fuzzy-logic processing, considering the difference between low-pass and high-pass filters in the retinal image. The fuzzy logic is comprised of different sets of fuzzy rules, where each fuzzy rule is constructed based on different thresholding values, to select and discard pixel values, leading to vessel extraction.

Recently, convolutional neural networks (CNNs) have emerged as powerful deep learning algorithms compared to other machine learning approaches, and perform remarkably well in a variety of fields. Liskowski and Krawiec (2016) presented a CNN-based algorithm for segmenting retinal vessels using a relatively simple network of four convolutional layers followed by three fully-connected layers. Maninis et al. (2016) implemented a fast and accurate deep CNN algorithm to simultaneously segment both the retinal and optic disc structures. The authors extracted feature maps from the $1st$–$4th$ and $2nd$–$5th$ layers of a pre-trained CNN, concatenating and adding a single additional convolutional layer to each set, outputting the segmentation of the vessels and optic disc, respectively. Maji et al. (2016) proposed a deep neural network-based technique by creating an ensemble of 12 simple five-layer CNNs to discriminate between vessel pixels and non-vessels using 20 raw color retinal images of DRIVE dataset.

Despite achieving astounding results, these deep CNNs suffer from their own set of shortcomings, such as a limited ability to model object-part relationships due to the scalar and additive nature of neurons in CNNs. Additionally, CNNs lose valuable information due to their use of a max-pooling component; while providing important features such as reducing the computational cost and number of parameters, improving invariance to minor distortions, and increasing the receptive field of neurons, max-pooling does so in a rigid and nearly-indiscriminate way, causing some important information to be lost and some unimportant information to be retained.

As an alternative method to the CNNs, a new type of deep learning network, called a capsule network, has attracted growing interest of researchers and is beginning to emerge as a competitive deep learning tool. These capsule networks, first introduced by Sabour et al. (2017) have shown remarkable performance compared to CNNs in many computer science applications including digit recognition, image classification, action detection, and in generative adversarial networks (Sabour et al., 2017, 2018; Deng et al., 2018; Duarte et al., 2018; Jaiswal et al., 2018). Recently, capsule networks have been successfully implemented in various medical image applications as well (Afshar et al., 2018; Iesmantas and Alzbutas, 2018; Jiménez-Sánchez et al., 2018; LaLonde and Bagci, 2018; Mobiny and Van Nguyen, 2018).

In summary, while a variety of CNN-based and other methods have been introduced, no work based on capsule neural networks have been introduced for the task of retinal vessel segmentation. Although the literature on capsule networks continues to grow, almost all previously introduced methods focus on classification-based tasks and virtually none, with the exception of LaLonde and Bagci (2018) are directed toward segmentation. Additionally, the previous study, (LaLonde and Bagci, 2018) focuses on segmenting a comparatively simpler object, lungs (large and closed objects), where the translation of such an approach to the more challenging arboriform structure present in retinal vessels is currently unclear. Toward this goal, this research is motivated to test the suitability of a capsule-based neural network in retinal blood vessel segmentation and to compare its performance with state-of-the-art CNN-based segmentation methods.

2.3. Video Magnification

Many signals that are frequently encountered in real-life environments fall below the spatiotemporal sensitivity of human visual systems. Signals such as the variation in skin coloration due to blood flow or the subtle mechanical motion of many structures or components typically fall within this group. Several recent studies have shown these signals can be artificially magnified to reveal important features of many objects around us (Liu et al., 2005; Wu et al., 2012).

Wu et al. (2012) proposed a Eulerian video magnification method using spatiotemporal processing to magnify imperceptible color changes and subtle motions in some real-world videos. They visualized the temporal variation of blood in a human face by amplifying the small color variations caused by blood flow. Wadhwa et al. (2016) implemented Eulerian video magnification to visualize subtle color and motion variations in real-world videos by amplifying the variations, namely, a person's pulse, breathing of an infant, sag and sway of a bridge, which is difficult to identify by human vision. Recently, Oh et al. (2018) applied deep CNNs to learn filters from the set of hand-designed kernels to overcome the drawback of the video magnification method due to the presence of noise and excessive blurring in small motion videos. The authors show their method is effective for producing high-quality results on real videos with less ringing artifacts and improved signal quality than the existing methods. Other related significant works include the studies by Liu et al. (2005) and Wu et al. (2018).

Although there have been several studies on the applications of different tools to reveal imperceptible signals in videos, the application of magnifying the subtle motion of blood circulation through retinal vessel structures in FA videos using Eulerian-based methods has not been reported in the literature. Pursuing to fill this literary gap, we introduce a Eulerian video magnification approach which is guided by our deep learning-based segmentation module to magnify the flow of blood through patients' vasculature. By constraining the magnification process to the segmented blood vessels, we maximize useful information and minimize the high levels of noise present in FA videos, thus providing ophthalmologists with a much needed AI-assisted tool for determining retinal flow characteristics and detecting potential disease.

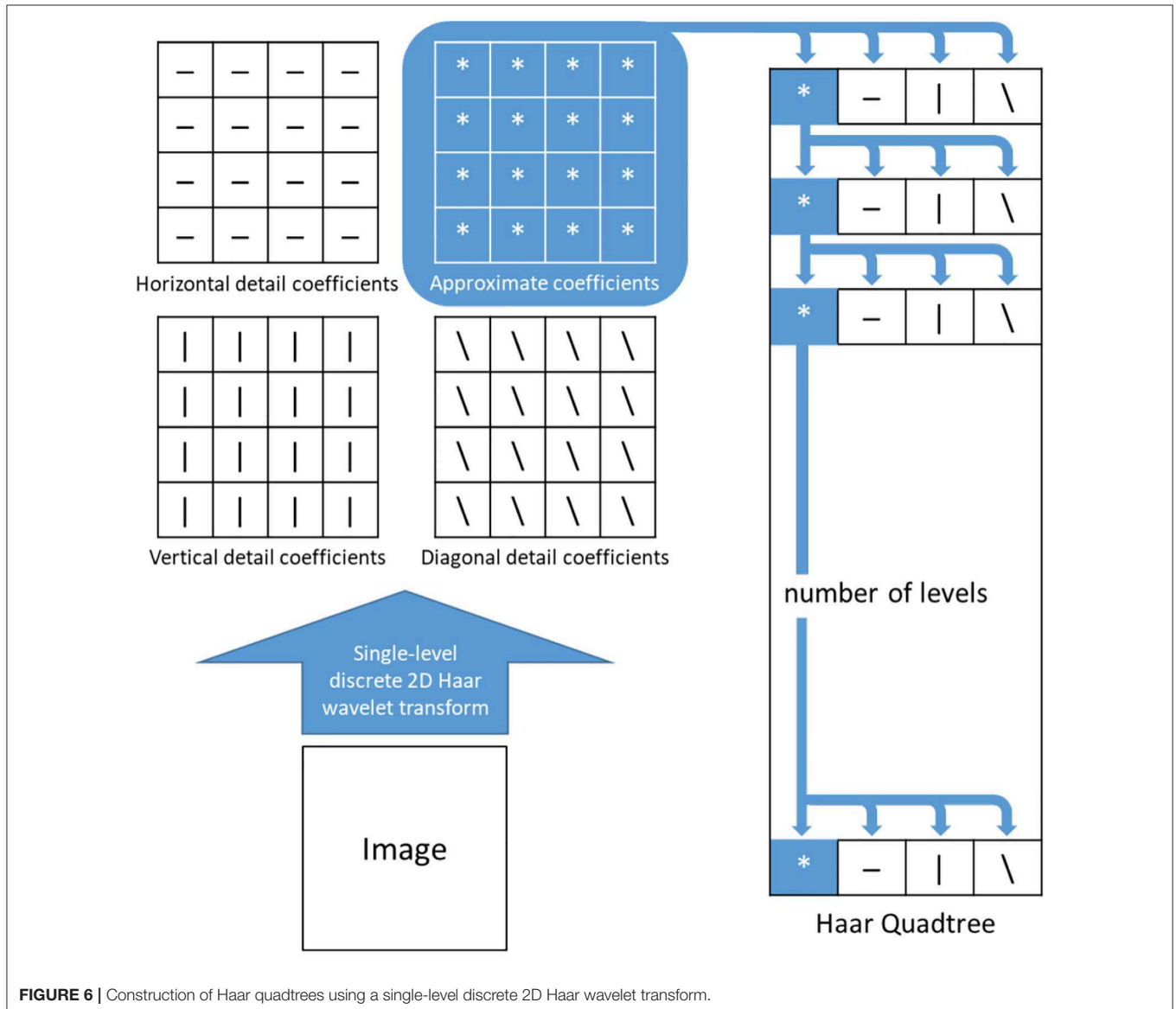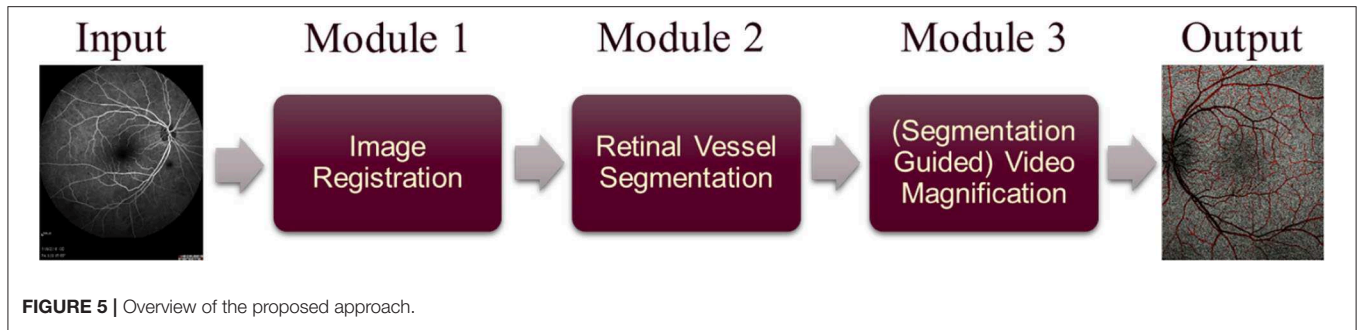## 3. METHODS

3.1. Overview of the Proposed System

We present a new tool that has three modules: image registration, retinal blood vessel segmentation, and segmentation-guided video magnification (an overview is shown in **Figure 5**). The main contribution of the first module is a novel wavelet-based registration algorithm using Haar coefficients. During registration, a middle frame is chosen as a reference to register all the remaining frames to that one. The objective of this module is to remove the camera and subject motion between all video frames, while also removing some of the noise.

The second module performs the segmentation of retinal vessels, which are later used to constrain our final module, greatly reducing the effect of noise, and providing quantitative analysis of blood flow patterns to ophthalmologists. To the best of our knowledge, there is no existing deep learning-based segmentation method for FA video in literature, and our work is the first such attempt. We propose to use a capsule-based deep learning approach (LaLonde and Bagci, 2018), where capsule networks are a relatively recent innovation within deep learning (discussed in section 2). We also perform a comparison with the popular *U-Net* (Ronneberger et al., 2015) architecture. In the final module, we introduce a segmentation-guided Eulerian video magnification method, as applied to retinal blood vessels during fluorescein angiogram imaging. Due to the high levels of noise present in FA videos, we constrain our video magnification to the vessel structures, guided by our deep learning-based segmentation module. In this way, instead of applying the Eulerian magnification to the entire frame, as done in previous methods, we use the segmentation as a prior to magnify motion and color changes in the blood vessels only, effectively amplifying our desired signal without simultaneously amplifying noise.

3.2. Image Registration

During FA video capture, video frames are not properly aligned due to several independent sources of motion, including patient movement, eye blinking, and the hand-operation of the fundus camera. Therefore, accurate image registration is required to properly analyze flow dynamics in FA videos. Moreover, our final module of the proposed solution, Eulerian video magnification, is designed to amplify subtle motions of interest, and errors in the registration can be greatly amplified in this stage to yield undesired results. In this paper, we utilize the Haar wavelets to introduce a new image registration algorithm - Haar wavelet domain multi-resolution image registration. This approach utilizes the Haar coefficient details of the images and optimization of the transformation matrix (T) using the Jacobian of Haar coefficient details. To utilize the Haar wavelets of the images, we built two quadtrees where each respective quadtree stores the coefficients for both the reference image and the image to be aligned.

The quadtrees are constructed in the following manner, where we define the height of the quadtree, N. First, the reference image is resized using an interpolation method by a factor of 2. Then 2D Haar wavelet transformation is applied on the resized images and their coefficients are stored in the first level of their respective

FIGURE 5 | Overview of the proposed approach.

FIGURE 6 | Construction of Haar quadtrees using a single-level discrete 2D Haar wavelet transform.

quadtrees. For the 2nd level, 2D Haar wavelet transformation is applied on the approximation coefficients of the previous level and the outputs, with the four coefficients, are stored at that level. This process is repeated for the rest of the levels of the quadtree (Q_1) of the reference image. The quadtree (Q_2) of the image to be aligned is constructed in the same manner. **Figure 6** depicts one such quadtree.

The main objective of this algorithm is to optimize the transformation matrix T that will align the moving image (frame) to the reference image. We first initialize the matrix T as a
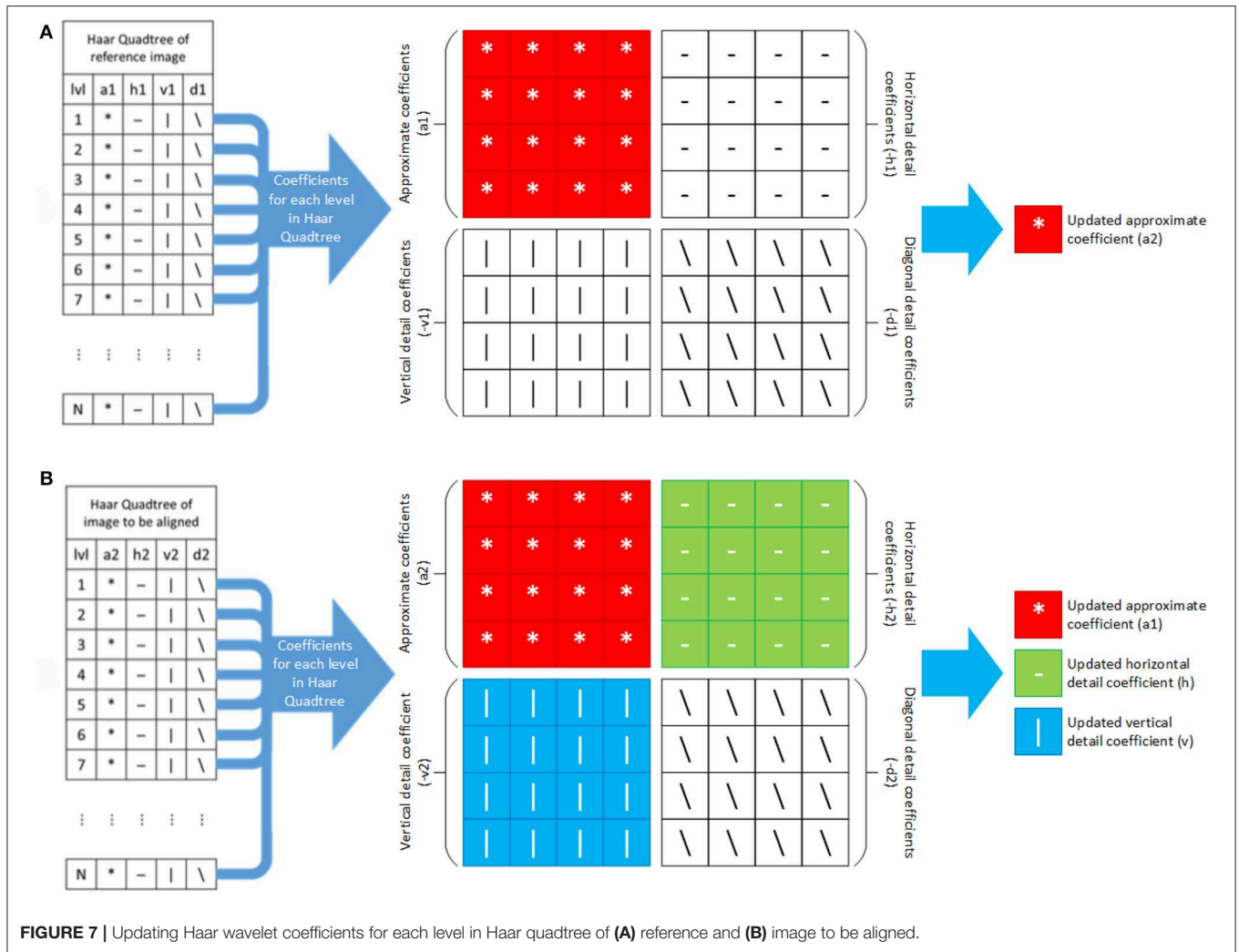
FIGURE 7 | Updating Haar wavelet coefficients for each level in Haar quadtree of **(A)** reference and **(B)** image to be aligned.

3×3 identity matrix, then create an outer loop to handle the coefficients of each level in the respective quadtrees and traverse both the quadtrees in bottom-up fashion (i.e., we deal with the lowest level first). For each level in Q_1, the first coefficient matrix is updated as the new approximate coefficients (a_2) for the reference image for that particular level as illustrated in **Figure 7A**. Similarly, for each level in Q_2, the first three coefficient matrices are updated as the new approximate (a_1), horizontal (h) and vertical coefficients (v) respectively, of the image to be aligned for the respective level (see **Figure 7B**).

The next step involves the Jacobian of Haar wavelet coefficients to optimize the transformation matrix T. Here we use an inner loop for P iterations. First, a 2D spatial transformation is used to update the variables a_1, h, and v using initial transformation matrix T. A mask is created to deal with pixels outside image borders. The mean of the approximation coefficients in the mask area, a_1 and a_2 are taken and are subtracted from their original values. The current transformation matrix T and the normalized cross-correlation values are computed and stored. Next, we compute the Jacobian of the warped approximation coefficients, J_h from the horizontal

and vertical coefficients h and v. The inverse of the auto-covariance of J_h, A_i, is used for computing the error matrix J_e. We define a correction matrix d_T as the product of A_i, J_e, and a Lagrange multiplier λ, defined as

$$\lambda = \frac{1}{\log_2 (1 + p)}, \qquad (1)$$

where, p is the iteration number of the inner loop. This Lagrange multiplier λ can be adjusted for better convergence. The transformation matrix T is updated using the correction matrix d_T. Both loops are closed and the updated transformation matrix T is used to align the moving image to the reference image. The procedure involving the inner loop is illustrated in **Figure 8**.

3.3. Segmentation Using Baseline SegCaps

There have been numerous studies applying deep learning to the task of image segmentation in various fields, including in biomedical imaging. Fully Convolutional Networks (*FCN*), (Long et al., 2015) *U-Net* (Ronneberger et al., 2015), and The One Hundred Layers Tiramisu (*Tiramisu*) (Jégou et al., 2017),

FIGURE 8 | Flowchart of Jacobian optimization of transformation matrix T.
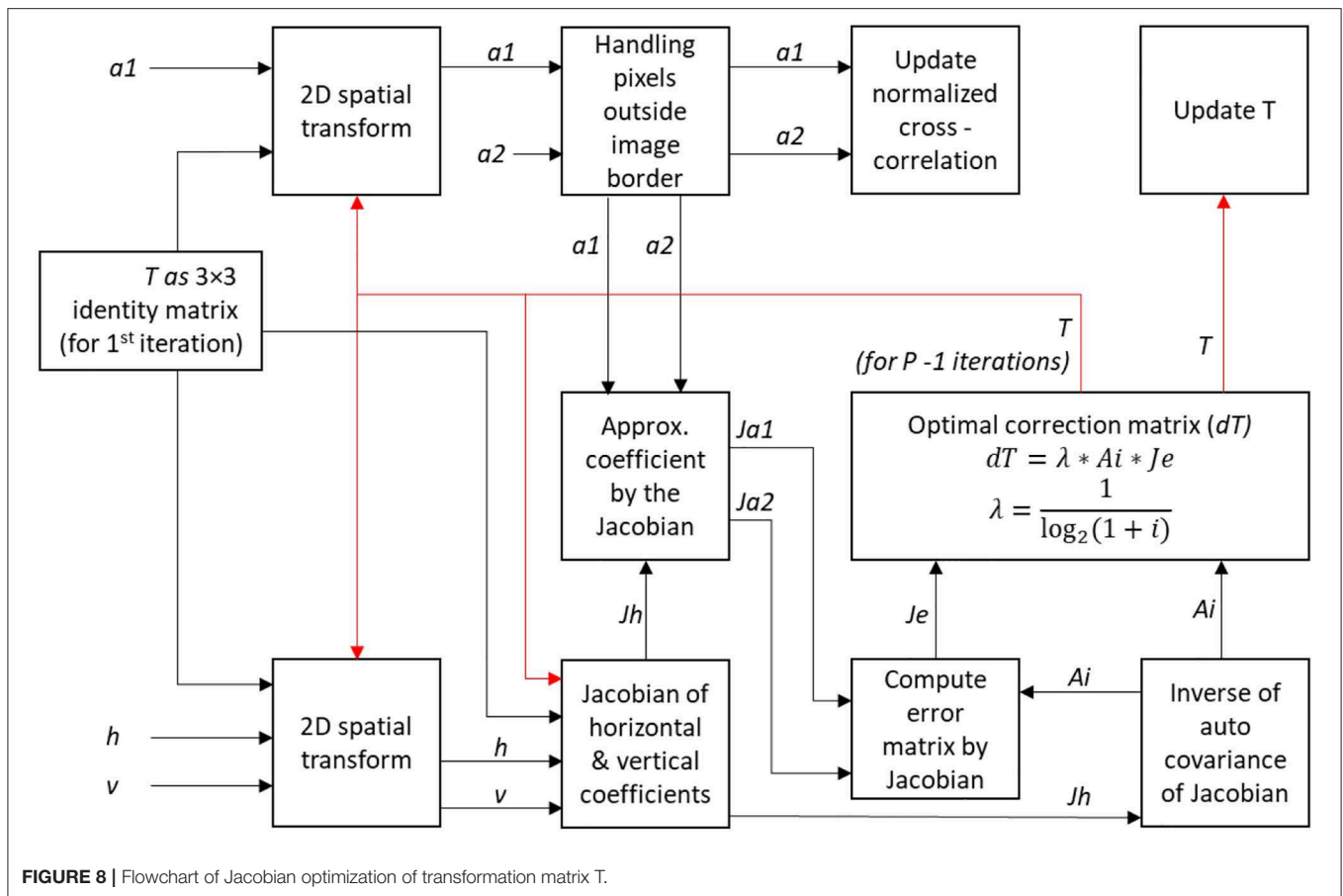
are some of the widely used deep learning algorithms for segmentation. The *U-Net* architecture is currently the most widely used image segmentation algorithm in biomedical imaging. Based on *FCN*, *U-Net* makes two modifications which improves the final segmentation results. First, *U-Net* continues the upsampling and combining present in *FCN* back to the original image resolution while keeping the number of feature maps symmetric at each level. Second, it replaces the summation over feature maps with a concatenation operation to perform the combining. These changes give *U-Net* its "U-like" shape, hence the name, and combines a downsampling path or "encoder" with an equally-shaped upsampling path or "decoder," hence networks of this type are typically referred to as "encoder-decoder" networks. These modifications allow *U-Net* to outperform the previous *FCN* and provide a finer segmentation. One of the major shortcomings of CNN based algorithms (including *U-Net*) is that higher-layer features only care if the neurons in the previous layer are giving a high value, and so on back through the network, without any regard to *where* those features are coming from, so long as they are within the receptive field of that higher-layer neuron. This makes the learning and preservation of part-whole relationships with a CNN extremely challenging. An additional shortcoming is introduced by the pooling operation, existing in virtually every successful CNN, that rigidly and nearly indiscriminately throws away information about the input

without any respect to the heterogeneity or importance of the information coming from one region relative to other regions.

To solve these major shortcomings, Sabour et al. (2017) introduced the concept of "capsule networks," in the form of *CapsNet* for small digit and small image classification. Capsules solve these major issues present in CNNs by changing the scalar representations of CNNs to vectors and removing the pooling operation in favor of a "dynamic routing" algorithm. Each capsule in a capsule network now stores information about the position and orientation of a given learned feature within its vector, and activations at the higher layer are computed as affine transformations on the lower layer capsules. In this way, capsule networks get an additional benefit; whereas CNNs attempt somewhat poorly to learn affine invariance for their feature values, capsule networks learn to be affine equivariant, thus allowing for much stronger robustness to affine transformations of the input. A CNN must learn how to map all possible affine transformations of an input to the same activation value for a given neuron, while capsule networks simply learn the input once and then can project all other possible affine transformations to the next layer. To best utilize these capsule vectors and remove the pooling operation, capsule network use strided, overlapping convolutions and a dynamic routing algorithm. The dynamic routing algorithm takes the affine transformed features from a previous layer and attempts to find the agreement between their

vectors at the next layer. For example. if all relevant features for a face (e.g., eyes, nose, mouth, ears) are present, but those features do not agree in their size and orientation about where the face should be, then the face capsule at the next layer will not be activated.

Capsule networks have shown some amazing initial promise, but there are very few works in the literature using this new style of neural networks. Since CapsNet and subsequent capsule works are proposed for small image classification, it was necessary to adapt it to segmentation. The *SegCaps* architecture (LaLonde and Bagci, 2018) is a heavily modified version of CapsNet for object segmentation that we adopted to use in our particular problem too. *SegCaps* changes the dynamic routing of CapsNet, constraining the routing to within a defined spatially-local window, and shares transformation matrices across each member of the grid within a capsule type. These modifications enable handling large image sizes, unlike previous capsule networks. Additionally, the concept of deconvolutional capsules was introduced, making the whole network a U-shaped network but with convolutional layers and pooling operations replaced with locally-constrained routing capsules. In this work, we conduct a comprehensive computational experiment to compare the baseline *SegCaps* (LaLonde and Bagci, 2018) with a state-of-the-art baseline network (*U-Net*), which itself has never been applied in this application domain. We chose to favor a three-layer convolutional capsule network, since global information obtained from an encoder-decoder structure is less important in our particular application, and we instead wish to introduce as little downsampling as possible to maintain the spatial resolution necessary to localize the small vessels present in FA videos. The schematics of the *SegCaps* network is illustrated in **Figure S1**.

3.4. Video Magnification

The final module of the proposed pipeline-based architecture is video magnification. This module enables the visualization of changes in intensity inside blood vessels for more easily analyzing flow dynamics and detecting disease. In the literature, it has been found that Lagrangian perspective-based approaches (with reference to fluid mechanics) can be used to amplify subtle motions and visualize the deformation of objects that would otherwise be undetectable. However, this perspective involves excessive computational times and it is difficult to make artifact-free, especially at regions of obstruction and areas of complex motion due to its exact motion estimation. On the other hand, Eulerian-based approaches do not rely on exact motion estimation, but instead, amplify the motion by changing temporal color at fixed positions (Wu et al., 2012). In this study, we create a Eulerian video magnification (EVM) method which is guided by our segmentation module and apply it to FA videos. The segmentation results are used as a prior, to constrain the magnification to only blood vessels, allowing effective analysis of blood flow while minimizing noise coming from non-vessel regions.

A Eulerian video magnification algorithm combines spatial and temporal processing to emphasize to minor changes in a video along the temporal axis. The process involves the decomposition of video frames into different spatial frequency bands. Then, the frames are converted from the spatial to the frequency domain using Fourier transformation. Considering the time series of pixel intensity in a frequency domain, a bandpass filter is applied to obtain the frequency bands of interest. The extracted bandpassed signal is amplified by a magnification factor of α. This factor is user-defined. Finally, the magnified signal is appended to the original and the spatial pyramid is collapsed to generate the output. The overall architecture of the Eulerian video magnification is illustrated in **Figure S2**.

We use a modified version of Eulerian video magnification as illustrated in **Figure 9**. The main requirements for the magnification algorithm to perform successfully are to have registered and noiseless frames, where the presence of these factors creates artifacts during magnification. In this study, we reduce the motion in FA frames using our proposed wavelet-based registration algorithm and reduce noise, which is mainly present in the background non-vessel regions, using the masks obtained from our proposed capsule-based segmentation algorithm. We apply a mean intensity projection on all the masks to obtain a single mask, which is used to generate frames that only contain all blood vessels. In this way, we apply the magnification algorithm to amplify only the motions of interest (blood flow through retinal vessels). Accordingly, the segmented frames obtained from the segmentation algorithm are taken as the inputs of the magnification algorithm. We apply Eulerian magnification with an amplification factor of 5, a frequency range of $0.5 - 10$ Hz, and spatial decomposition levels of 4. The output is a gray-scale video, where only the blood vessels are magnified in red color, showing the flow of blood in the vessels as the video progresses.

4. EXPERIMENTS AND RESULTS

A prospective study was conducted with an IRB approved at the University of Central Florida College of Medicine, Orlando, FL in collaboration with Central Florida Retina, Orlando, FL (IRB: MODCR00000068, Title: Automated Retinal Vascular Flow Analysis). FA videos of 10 normal subjects are considered for this study after obtaining the written consent of the participants. The details of the dataset are presented in **Table S1**. The dataset will be made available by the authors on request.

4.1. Wavelet-Based Image Registration

The wavelet-based registration algorithm was coded in MATLAB, using the wavelet and parallel computing toolboxes. We used four different transformation types namely, translation, Euclidean, affine, and projective to generate four videos. Two experiments were performed based on the number of iterations for each transformation type. In the first experiment, 2 levels of Haar quadtree with 50 iterations was used. In the second experiment, the number of iterations was increased to 150. We considered the middle frame as the reference image for registration. To evaluate the output quantitatively, the normalized cross-correlation of each frame was calculated with respect to the reference frame. As a brief note, we considered cross-correlation as a similarity matrix instead of mutual information because it has been shown in the literature that cross-correlation often gives better interpretation

FIGURE 9 | Data flow diagram of Eulerian video magnification for FA video.
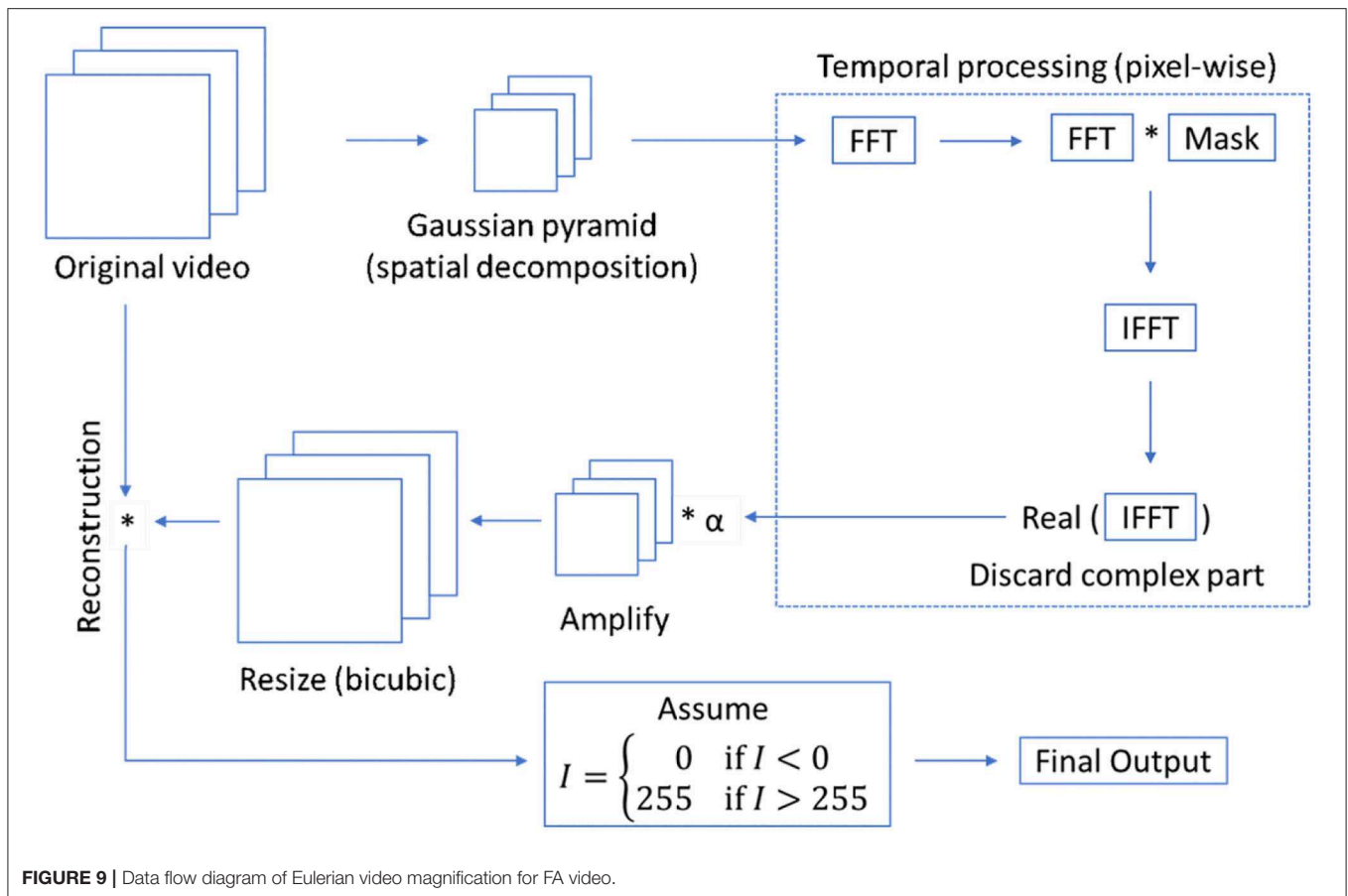
than mutual information on images with the same modality or source.

The *t*-test was used to validate the statistical significance of the improved results obtained by the second experiment over those given by the first experiment for all the four different transformation types individually. The video considered for this experiment had 182 frames. In the first experiment, four series of normalized cross-correlation were obtained for each transformation type. Similarly, another set of normalized cross-correlation was generated through the Experiment 2. Thus, each transformation type for the two experiments comprised 182 pairs of normalized cross-correlation values. For each transformation type, the mean and the standard deviation of the 182 differences in the cross-correlation values were determined. The difference corresponding to a transformation was computed by subtracting the cross-correlation of the particular transformation type of the second experiment from that of the first experiment. The results of statistical test (*P*-values) of the two different experiments for each transformation type are as follows, Translation: 0.0175, Euclidean: 0.00868, Affine: 0.0053, and Projection: 0.026, respectively.

At 95% confidence level, alpha (significance level) = 0.05. Since there were total of 182 frames, degrees of freedom were 181. Using a standard table of *t*-distribution, critical value t for 181 degrees of freedom was obtained for each transformation type.

The results depict that *t* Statistic is greater than *t* critical for one-tailed distribution. Based on *t*-test, both results in experiments were found to be statistically significantly different for all four transformation types.

As a further baseline for comparison, we also considered another standard image registration algorithms, rigid registration. The rigid registration algorithm was coded in C++, using the Insight Toolkit (ITK) library. The OpenCV library was also adopted for pre-processing FA video frames. The experiments for the rigid registration was carried out with binary images as this yielded far better results than operating on the gray-scale images. To optimize the registration parameters, we applied the gradient descent algorithm with a learning rate of 0.125 and a step size of 0.001. We run the optimizer for 200 iterations, and the parameters of the gradient descent algorithm were obtained through manual search.

The visual evaluations revealed that the images were far better aligned when binary image was used, while the image registration failed when a gray-scale image was used as a reference image. The overlapping of two successive frames using two different reference images (one of them gray-scale, while the other is a binary image) is shown in **Figure 10**. Further, visual evaluation of the registration outputs shows the wavelet-based registration algorithm to be superior to the rigid registration algorithm. The
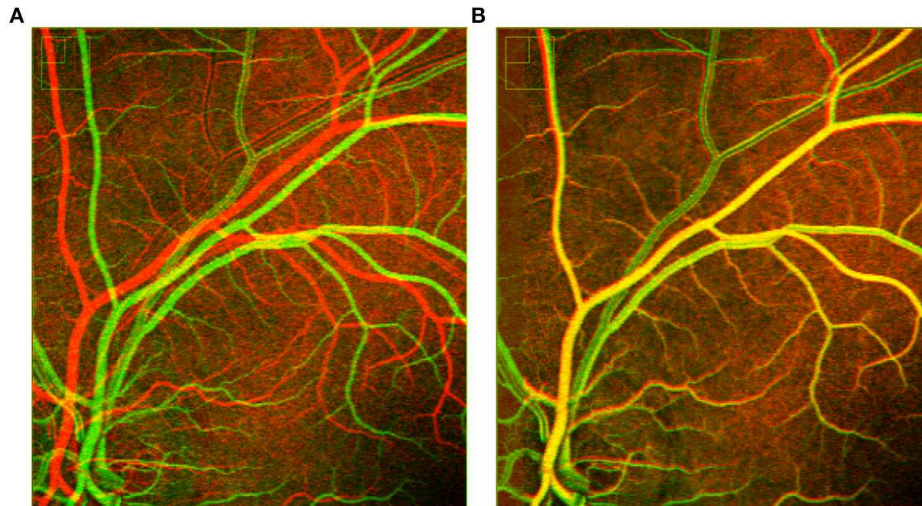
FIGURE 10 | Overlapping two successive frames shows rigid registration **(A)** fails with grayscale images and **(B)** successful with binary images.

main aim of this module is to remove the motion across frames, and the proposed algorithm greatly reduces the motion.

4.2. Segmentation Using Deep Learning

Manual annotations (ground truths) of the FA video frames were done using AMIRA software. Due to the difficulty of annotation, and since physicians primarily look for the changes in the blood flow in large vessels, only large and medium-size vessels were annotated, with the majority of very small vessels left absent. The networks *U-Net* and baseline *SegCaps* was coded in python using Keras and Tensorflow, and experiments were run on a single 12GB memory NVIDIA Titan X GPU. The networks were trained from scratch on the FA images with standard data augmentation techniques (i.e., random scaling, flip, shift, rotation, elastic deformations, and salt-n-pepper random noise). The Adam optimizer was used with an initial learning rate of 10^{-4}, with the learning rate decayed by a factor of 10^{-6} when the validation loss reached a saturation level. Due to the large image sizes and a loss value coming from every pixel, the batch size was taken as 1 for all the experiments, with weighted binary cross-entropy loss used as the loss function. The training continued for 200 epochs (1000 steps per epoch) and early stopping was applied with a patience value of 25 epochs. To validate the algorithm's performance, we used a 10-fold cross-validation (leave-one-out for our 10 patients videos), with 10% of the training video frames set aside for validation.

4.2.1. Quantitative Analysis

The results of our segmentation experiments were quantitatively evaluated in terms of the standard metrics: Dice coefficient, sensitivity, and specificity. The results of these experiments and the total number of parameters of the models are shown in **Table 1**. The presented *SegCaps* network outperforms the *U-Net* model, producing

TABLE 1 | Comparison of Dice, Sensitivity (Sens.), and Specificity (Spec.) for different neural network models are given in mean ± standard deviation (SD) format.

Methods	Params.	Dice (%)	Sens. (%)	Spec. (%)
U-Net	31 M	85.50 ± 3.78	90.74 ± 6.24	**98.56 ± 0.71**
SegCaps	**1.6 M**	**85.94 ± 3.05**	**92.36 ± 3.57**	98.38 ± 0.95

higher sensitivity while obtaining comparable dice and specificity scores. Additionally worth noting, the *SegCaps* achieved these results while using **94.84%** fewer parameters than *U-Net*.

Although *SegCaps* outperformed *U-Net* model, the difference was not found to be statistically significant with respect to the region-based evaluation metric (Dice coefficient). We performed a *t*-test on the Dice scores between the *SegCaps* and the *U-Net* models. The *p*-value for one-tail was 0.38, hence the result is not statistically significant. We argue this is somewhat due to limitations in the Dice metric where small differences in thin vessel structures are overshadowed by the large overlap in the more easily segmented vessels. This is demonstrated in **Figure 12** which show how much better *SegCaps* captures the vessel structures, especially for thin or close-together vessels, even though this results in very similar Dice scores. However, this is partially captured by the higher sensitivity in *SegCaps* over *U-Net*. It should be noted that these results for both the proposed *SegCaps* and our baseline *U-Net* represent the first application of deep learning-based methods in the literature to analyze FA videos, and both perform well ($> 90\%$ sensitivity) in our proposed pipeline.

Since there was no ground truth for small vessels, we were not able to analyze them quantitatively. Moreover, the small vessel quantification is not within the scope of our study due to the current clinical workflow that targets medium to large vessels only. However, we have let our participating
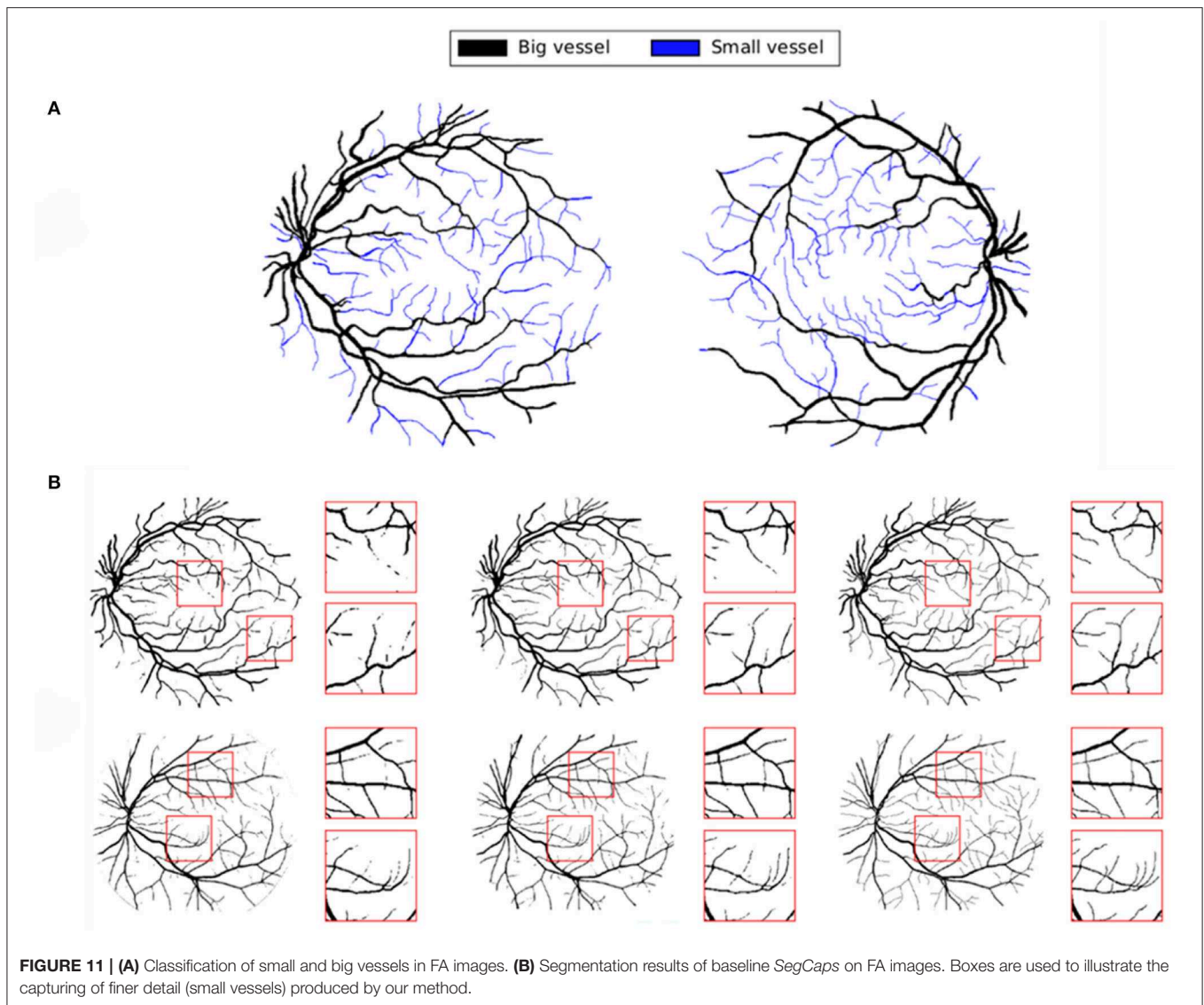
FIGURE 11 | (A) Classification of small and big vessels in FA images. **(B)** Segmentation results of baseline *SegCaps* on FA images. Boxes are used to illustrate the capturing of finer detail (small vessels) produced by our method.

ophthalmologists evaluate the proposed method's small vessel segmentation for their correctness, their visual evaluations did not reveal any concerns.

4.2.2. Qualitative Analysis

To truly capture the difference in performance between methods, visual evaluation of segmented objects is often necessary when the object exists in multiple scales, where finer details can get washed out in the quantitative metrics. These finer details (small vessels) are illustrated in **Figure 11A**. We show some results of our proposed baseline *SegCaps* in **Figure 11B**, illustrating the fine details of smaller vessels captured successfully by baseline *SegCaps*, shown on small patches from segmentation results of FA images for better visualization.

To further illustrate our proposed method's performance on finer details, we provide a qualitative comparison on a few examples between baseline *SegCaps* and *U-Net* in **Figure 12**. Two different frames are shown from two subjects, with the ground

truth colored in cyan, the predicted segmentation in magenta, and the white region represents the intersection of both regions. From the figures, we see far more white regions in the *SegCaps* patches than its *U-Net* counterpart. We mark some places using red (for *U-Net*) and green (for *SegCaps*) arrows to show regions of under- or over-segmentation by *U-Net*. While the quantitative results were extremely similar, we can see the baseline *SegCaps* performed far better in these challenging regions denoted by the arrows. As an interesting note, manual annotations are done for medium and large vessels only; however, it is found that baseline *SegCaps* achieves great segmentation of smaller blood vessels as well.

4.3. Eulerian Video Magnification

The results of our experiments on Eulerian video magnification are shown in **Figure 13**. In a first experiment (**Figure 13A**), we magnify the FA frames based on the corresponding masks obtained from the *SegCaps* network. Since at the beginning of the
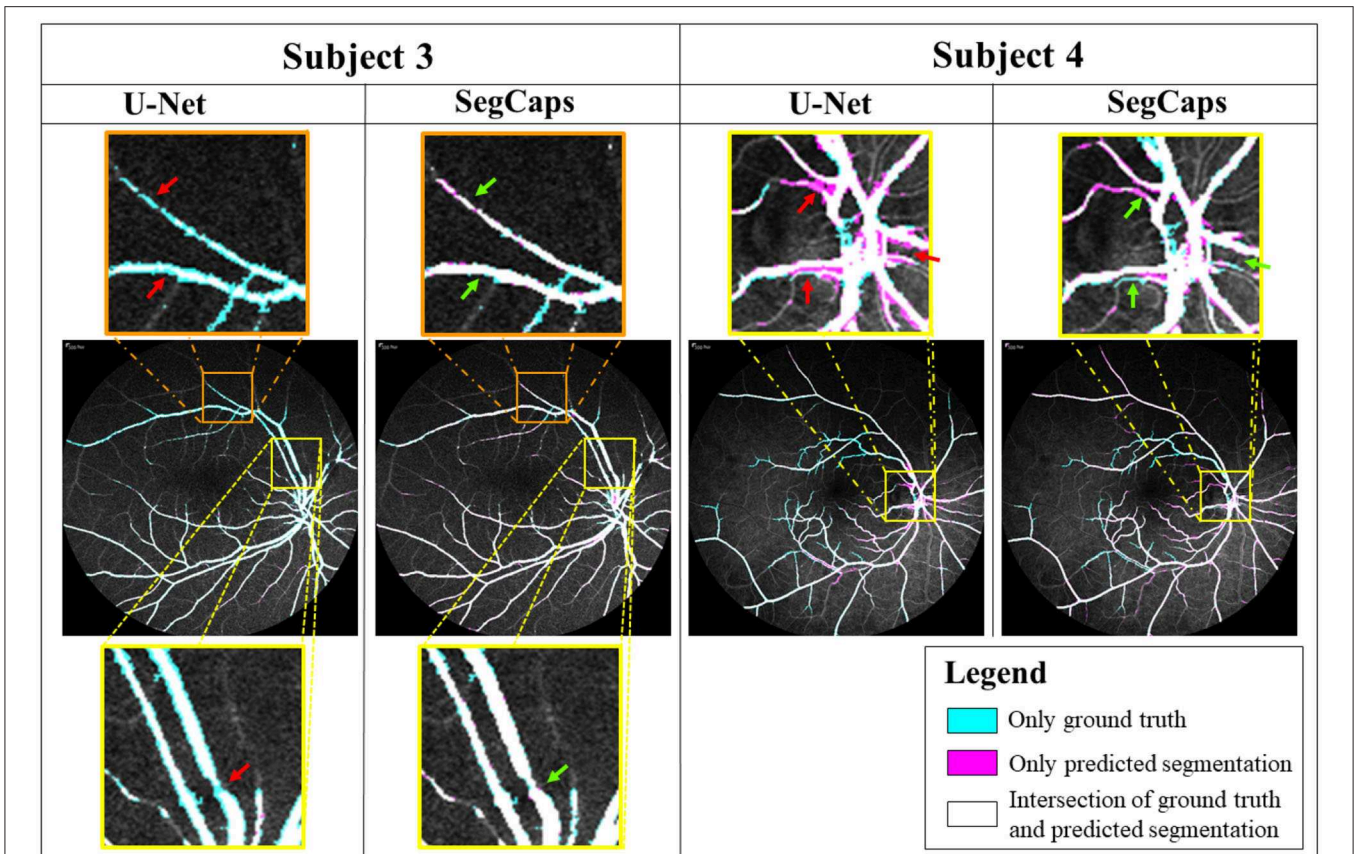
FIGURE 12 | Comparison of segmentation results on *U-Net* and baseline *SegCaps*. Despite achieving similar quantitative results, we can see *SegCaps* outperforms *U-Net* in challenging regions as denoted by the corresponding red (*U-Net*) and green (*SegCaps*) arrows. U-Net struggles with under-segmentations in Subject 3 and over-segmentations in Subject 4.
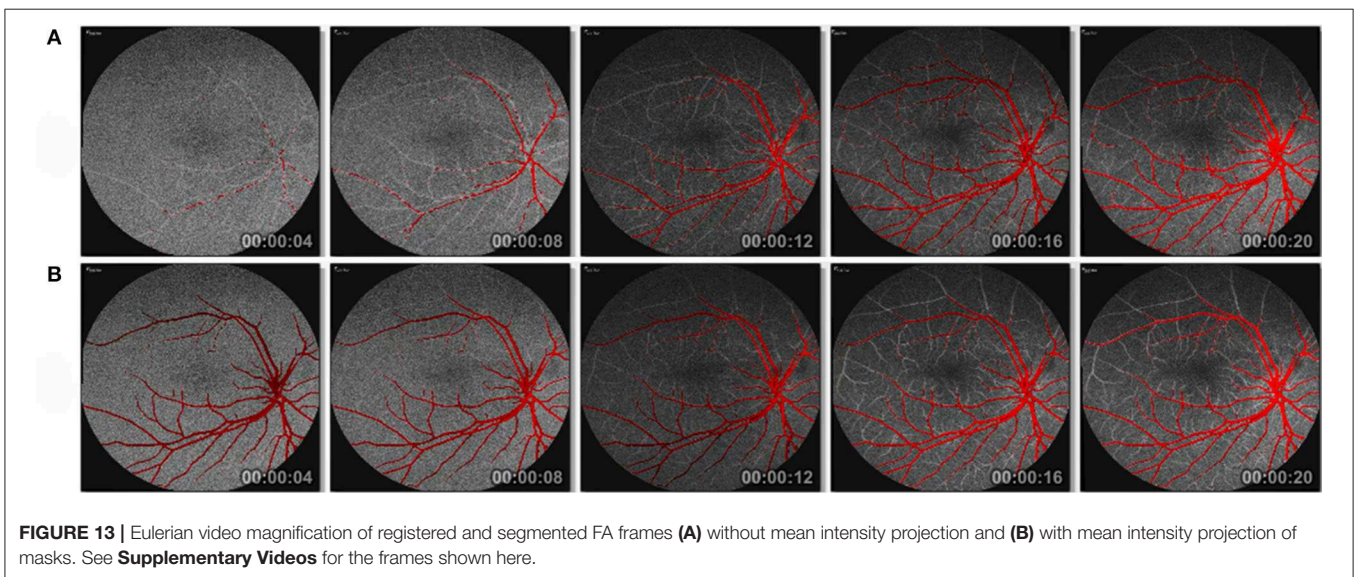
FIGURE 13 | Eulerian video magnification of registered and segmented FA frames **(A)** without mean intensity projection and **(B)** with mean intensity projection of masks. See **Supplementary Videos** for the frames shown here.

FA video, the blood flow has not started, we see incomplete red-colored vessels in the first few frames of the FA video. After a few frames, as the blood slowly makes its way into the arteries and

veins, we see a complete picture. To get a better understanding of the blood flow dynamics, we perform a second experiment. Instead of using the corresponding masks for each frame, we

apply a mean intensity projection on all the masks to get a single mask that will represent all the frames. This mask is then used as the prior for our segmentation guided magnify of the blood vessel regions. In this second experiment, we see much clearer blood flow dynamics. The blood vessels are highlighted from the first frame. We see gradual intensity change in the red color channel as the blood flows into the veins as illustrated in **Figure 13B**. We conducted some preliminary computational experiments to ascertain the best parameters of the proposed EVM method and consider the parameters for the EVM approach: amplification factor (α) of 5, a frequency range between 0.5 and 10 Hz, and spatial decomposition levels of 4.

5. DISCUSSIONS AND CONCLUSION

We present a novel computational tool that helps physicians analyze and diagnose fluorescein retinal angiography videos. The proposed pipeline-based approach consists of three phases: image registration, followed by retinal vessel segmentation, and lastly, segmentation guided video magnification. In the image registration phase, we apply a novel wavelet-based registration method that uses Haar coefficients to spatiotemporally align individual frames of the FA video. Moreover, in this phase, all the frames are registered to the middle frame to eliminate the motion between the frames and some of the noises. In the second phase, a baseline capsule-based neural network architecture performs segmentation of retinal vessels. We demonstrate this novel application of a capsule network architecture outperforms a state-of-the-art CNN *U-Net* in terms of both parameter use, as well as qualitative and quantitative results. Finally, we introduce a segmentation-guided Eulerian video magnification method to further enhance the intensities of the retinal blood vessels. The magnification is performed on the segmented retinal vessels, emphasizing the retinal vessel information in FA videos while minimizing the high levels of noise. The final output video from our proposed novel AI-assistant tool provides a better analysis of blood flow from fluorescein retinal angiography videos and should aid ophthalmologists in gaining a deeper understanding of the blood flow dynamics and enable more accurate identification of potential regions of pathology.

Limitations in the current study are to be noted. Image registration is essential for the removal of motion from the video. From a segmentation point of view, we have investigated the performance of the *SegCaps* based network architecture on a small dataset. The dataset is limited to only 10 subjects for analysis due to the extremely time-consuming nature of collecting annotations, which were analyzed using leave-one-out cross-validation. Although our data validated the superior performance of the algorithm, a larger dataset, once available, could be used for more accurate FA video image segmentation. Across all subjects the capsule segmentation network (SegCaps) outperformed U-Net but the difference may not be statistically significant from region-based evaluation metric (Dice). We argue this is somewhat due to limitations in the quantitative metrics and that the qualitative figures demonstrate how much better SegCaps captures the vessel structures, especially for thin or

close-together vessels, even though this results in very similar Dice scores. This is partially captured by the higher sensitivity in SegCaps over U-Net. The parameters used in the magnification algorithm can also be further optimized to get more precise amplification. Since we empirically find the parameters, a more thorough parameter search can potentially lead to further performance gains.

Future applications of this research can include automated flow calculation and detection at various segments of the retinal vasculature for more accurate disease detection. Theoretically, these can be integrated into existing clinical software applications for automated diagnoses. Lastly, one may re-organize the system by using the segmentation module first, followed by the registration and magnification modules. It may bring additional benefit in the registration step to first have a segmentation mask; however, potential mistakes in segmentation can affect the overall robustness of the system in such a scenario. Since we are currently using a 2D neural network, this would be an easy switch; however, as a future direction 3D segmentation methods could be investigated on the registered video frames.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the University of Central Florida (UCF). Patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SL and UB contributed conception and design of the whole study. SL and RL contributed in coding based on the study. AC and JO provided data acquisition of FA videos from Central Florida Retina, FL, USA. HF contributed in the design of a new method of image registration. SS provided medical expertise in this study. SL wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00024/full#supplementary-material

REFERENCES

Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2018). "Brain tumor type classification via capsule networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)* (Athens), 3129–3133. doi: 10.1109/ICIP.2018.8451379

Almotiri, J., Elleithy, K., and Elleithy, A. (2018). Retinal vessels segmentation techniques and algorithms: a survey. *Appl. Sci.* 8:155. doi: 10.3390/app8020155

Chen, L., Huang, X., and Tian, J. (2015). Retinal image registration using topological vascular tree segmentation and bifurcation structures. *Biomed. Signal Process. Control* 16, 22–31. doi: 10.1016/j.bspc.2014.10.009

Christodoulidis, A., Hurtut, T., Tahar, H. B., and Cheriet, F. (2016). A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images. *Computer. Med. Imaging Graph.* 52, 28–43. doi: 10.1016/j.compmedimag.2016.06.001

Deng, F., Pu, S., Chen, X., Shi, Y., Yuan, T., and Pu, S. (2018). Hyperspectral image classification with capsule network using limited training samples. *Sensors* 18:3153. doi: 10.3390/s18093153

Duarte, K., Rawat, Y., and Shah, M. (2018). "Videocapsulenet: A simplified network for action detection," in *Advances in Neural Information Processing Systems* (Montreal, QC), 7610–7619.

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. (2012). Blood vessel segmentation methodologies in retinal images-a survey. *Comput. Methods Prog. Biomed.* 108, 407–433. doi: 10.1016/j.cmpb.2012.03.009

Ghassabi, Z., Shanbehzadeh, J., and Mohammadzadeh, A. (2016). A structure-based region detector for high-resolution retinal fundus image registration. *Biomed. Signal Process. Control* 23, 52–61. doi: 10.1016/j.bspc.2015.08.005

Gongt, H., Li, Y., Liu, G., Wu, W., and Chen, G. (2015). "A level set method for retina image vessel segmentation based on the local cluster value via bias correction," in *2015 8th International Congress on Image and Signal Processing (CISP)* (Shenyang), 413–417. doi: 10.1109/CISP.2015.7407915

Guo, F., Zhao, X., Zou, B., and Liang, Y. (2017). Automatic retinal image registration using blood vessel segmentation and sift feature. *Int. J. Pattern Recogn. Artif. Intell.* 31:1757006. doi: 10.1142/S0218001417570063

Iesmantas, T., and Alzbutas, R. (2018). "Convolutional capsule network for classification of breast cancer histology images," in *International Conference Image Analysis and Recognition* (Póvoa de Varzim: Springer), 853–860. doi: 10.1007/978-3-319-93000-8_97

Imani, E., Javidi, M., and Pourreza, H.-R. (2015). Improvement of retinal blood vessel detection using morphological component analysis. *Comput. Methods Prog. Biomed.* 118, 263–279. doi: 10.1016/j.cmpb.2015.01.004

Jaiswal, A., AbdAlmageed, W., Wu, Y., and Natarajan, P. (2018). "Capsulegan: Generative adversarial capsule network," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich). doi: 10.1007/978-3-030-11015-4_38

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2017). "The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Honolulu, HI), 11–19. doi: 10.1109/CVPRW.2017.156

Jiang, Z., Yepez, J., An, S., and Ko, S. (2017). Fast, accurate and robust retinal vessel segmentation system. *Biocybern. Biomed. Eng.* 37, 412–421. doi: 10.1016/j.bbe.2017.04.001

Jiménez-Sánchez, A., Albarqouni, S., and Mateus, D. (2018). "Capsule networks against medical imaging data challenges," in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis* (Granada: Springer), 150–160. doi: 10.1007/978-3-030-01364-6_17

Kar, S. S., and Maity, S. P. (2016). Blood vessel extraction and optic disc removal using curvelet transform and kernel fuzzy c-means. *Comput. Biol. Med.* 70, 174–189. doi: 10.1016/j.compbiomed.2015.12.018

Karali, E., Asvestas, P., Nikita, K. S., and Matsopoulos, G. K. (2004). "Comparison of different global and local automatic registration schemes: an application to retinal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Saint-Malo: Springer), 813–820. doi: 10.1007/978-3-540-30135-6_99

Kolar, R., Tornow, R. P., Odstrcilik, J., and Liberdova, I. (2016). Registration of retinal sequences from new video-ophthalmoscopic camera. *Biomed. Eng. Online* 15:57. doi: 10.1186/s12938-016-0191-0

Kumar, D., Pramanik, A., Kar, S. S., and Maity, S. P. (2016). "Retinal blood vessel segmentation using matched filter and Laplacian of Gaussian," in *2016 International Conference on Signal Processing and Communications (SPCOM)* (Bengaluru), 1–5. doi: 10.1109/SPCOM.2016.7746666

Laliberté, F., Gagnon, L., and Sheng, Y. (2003). Registration and fusion of retinal images-an evaluation study. *IEEE Trans. Med. Imaging* 22, 661–673. doi: 10.1109/TMI.2003.812263

LaLonde, R., and Bagci, U. (2018). "Capsules for object segmentation," in *International Conference on Medical Imaging with Deep Learning (MIDL)* (Amsterdam), 1–9.

Liskowski, P., and Krawiec, K. (2016). Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* 35, 2369–2380. doi: 10.1109/TMI.2016.2546227

Liu, C., Torralba, A., Freeman, W. T., Durand, F., and Adelson, E. H. (2005). "Motion magnification," in *ACM Transactions on Graphics (TOG), Vol. 24 (ACM)*, 519–526. doi: 10.1145/1073204.1073223

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 3431–3440. doi: 10.1109/CVPR.2015.7298965

Maji, D., Santara, A., Mitra, P., and Sheet, D. (2016). Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. *arXiv preprint arXiv:1603.04833.*

Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., and Van Gool, L. (2016). "Deep retinal image understanding," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Athens: Springer), 140–148. doi: 10.1007/978-3-319-46723-8_17

Mapayi, T., Viriri, S., and Tapamo, J.-R. (2015). Adaptive thresholding technique for retinal vessel segmentation based on glcm-energy information. *Comput. Math. Methods Med.* 2015:597475. doi: 10.1155/2015/597475

Matsopoulos, G. K., Asvestas, P. A., Mouravliansky, N. A., and Delibasis, K. K. (2004). Multimodal registration of retinal images using self organizing maps. *IEEE Trans. Med. Imaging* 23, 1557–1563. doi: 10.1109/TMI.2004.836547

Matsopoulos, G. K., Mouravliansky, N. A., Delibasis, K. K., and Nikita, K. S. (1999). Automatic retinal image registration scheme using global optimization techniques. *IEEE Trans. Inform. Technol. Biomed.* 3, 47–60. doi: 10.1109/4233.748975

McCormick, B., and Goldbaum, M. (1975). "Stare: structured analysis of the retina: image processing of tv fundus image," in *USA-Japan Workshop on Image Processing, Jet Propulsion Laboratory* (Pasadena, CA).

Mobiny, A., and Van Nguyen, H. (2018). "Fast capsnet for lung cancer screening," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Granada: Springer), 741–749. doi: 10.1007/978-3-030-00934-2_82

Nguyen, U. T., Bhuiyan, A., Park, L. A., and Ramamohanarao, K. (2013). An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recogn.* 46, 703–715. doi: 10.1016/j.patcog.2012.08.009

Oh, T.-H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W. T., et al. (2018). "Learning-based video motion magnification," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich), 633–648. doi: 10.1007/978-3-030-01225-0_39

Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* 18, 10762–10774. doi: 10.1364/OE.18.010762

Rabbani, H., Allingham, M. J., Mettu, P. S., Cousins, S. W., and Farsiu, S. (2015). Fully automatic segmentation of fluorescein leakage in subjects with diabetic macular edema. *Invest. Ophthalmol. Vis. Sci.* 56, 1482–1492. doi: 10.1167/iovs.14-15457

Ramaswamy, G., Lombardo, M., and Devaney, N. (2014). Registration of adaptive optics corrected retinal nerve fiber layer (RNFL) images. *Biomed. Opt. Express* 5, 1941–1951. doi: 10.1364/BOE.5.001941

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Sabour, S., Frosst, N., and Hinton, G. E. (2017). "Dynamic routing between capsules," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 3856–3866.

Sabour, S., Frosst, N., and Hinton, G. E. (2018). "Matrix capsules with EM routing," in *6th International Conference on Learning Representations, ICLR* (Vancouver, BC).

Sharma, S., and Wasson, E. V. (2015). Retinal blood vessel segmentation using fuzzy logic. *J. Netw. Commun. Emerg. Technol.* 4, 1–5.

Singh, N. P., Kumar, R., and Srivastava, R. (2015). "Local entropy thresholding based fast retinal vessels segmentation by modifying matched filter," in *International Conference on Computing, Communication & Automation* (Noida), 1166–1170. doi: 10.1109/CCAA.2015.7148552

Srinidhi, C. L., Aparna, P., and Rajan, J. (2017). Recent advancements in retinal vessel segmentation. *J. Med. Syst.* 41:70. doi: 10.1007/s10916-017-0719-2

Staal, J., Abrámoff, M. D., Niemeijer, M., Viergever, M. A., and Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* 23, 501–509. doi: 10.1109/TMI.2004.825627

Wadhwa, N., Wu, H.-Y., Davis, A., Rubinstein, M., Shih, E., Mysore, G. J., et al. (2016). Eulerian video magnification and analysis. *Commun. ACM* 60, 87–95. doi: 10.1145/3015573

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* 31, 1–8. doi: 10.1145/2185520.2185561

Wu, X., Yang, X., Jin, J., and Yang, Z. (2018). Amplitude-based filtering for video magnification in presence of large motion. *Sensors* 18:2312. doi: 10.3390/s18072312

Xu, Z., Bagci, U., Foster, B., Mansoor, A., Udupa, J. K., and Mollura, D. J. (2015). A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT. *Med. Image Anal.* 24, 1–17. doi: 10.1016/j.media.2015.05.003

Zhang, J., Tang, Z., Gui, W., and Liu, J. (2015). "Retinal vessel image segmentation based on correlational open active contours model," in *2015 Chinese Automation Congress (CAC)* (Wuhan), 993–998.