

Automated Detection of Collagen Bundles in Second Harmonic Generation Microscopy Images

Cihan Bilge Kayasandik

Abstract— Collagen is one of the most abundant proteins in the body. It is essential for the structure, functionality, and strength of the connective tissue such as skin, bone, tendon, and cornea. It is known that a change in the arrangement or morphology of these fibrillar structures relates to multiple dysfunctions including corneal diseases and various cancer types. Due to their critical roles in wide-range abnormalities, there is an increasing interest in the pattern analysis of collagen arrangements. In recent years, Second Harmonic Generation (SHG) microscopy is proven to be an efficient imaging modality for visualizing unstained collagen fibrils. There are plenty of studies in the literature on the analysis of collagen distribution in SHG images. However, the majority of these methods are limited to detecting simple, statistical and non-local properties such as pixel intensity and orientation variance. There is a need for a method to detect the local structural properties of collagen bundles. This paper is to introduce an automated method to detect collagen bundles in 3-dimensional SHG microscopy images. The origin of the proposed method is based on multiscale directional representation systems. The proposed method detects the collagen bundles by measuring the dominant orientation of local regions and an orientation-based connected component analysis. Through more local analysis and the detection of collagen bundles separately, the proposed method would lead to the extraction of more detailed structural information on collagen bundle distribution.

Index Terms—collagen detection, cornea analysis, image analysis, machine learning, SHG.

I. INTRODUCTION

CORNEA is the transplant front region of the outer casing of the eye [1]. Although one of the main functions as protecting the interior content, the cornea has multiple roles in maintaining the human vision system properly. The highly complex and organized structure of the cornea is critically important for satisfying transparency and refraction [2]. Cornea is dominantly formed by structures called collagen. Collagen gives the cornea the ability to be strength required to fulfill its role of producing a tough container for the inner contents of the eye with precise curvature and a high level of transparency to visible wavelengths. The collagen fibers are arranged in parallel bundles called fibrils. These fibrils are packed in layers or lamellae. The stroma of the human eye contains 200–250 distinct lamellae, each layer arranged at right angles relative to fibers in adjacent lamella [3]. It is known that a change in the arrangement or morphological properties of these fibrils relates to dysfunction in the vision system [4]. Since collagen is one of the main structural proteins found in connective

● **Cihan Bilge Kayasandik** is with the Department of Computer Engineering Istanbul Medipol University, Istanbul, Turkey, e-mail: cbkayasandik@medipol.edu.tr

tissues, besides the vision system, collagen structures are known to be related to various abnormalities such as breast and ovarian cancer [5], [6]. Due to this critical relationship between collagen structures and wide-range abnormalities, there is an increasing interest in the pattern analysis of images of collagen arrangements [4], [5], [6].

Multiphoton microscopy is recently recognized as a powerful imaging technique to visualize unstained samples. With no need for staining and advantages for diagnostic procedures, Multiphoton microscopy and its variants have been used increasingly in biomedical imaging [7]. Multiphoton microscopy includes Second Harmonic Generation (SHG) microscopy which is based on an absorption-free process [8]. In recent years, SHG microscopy is proven to be an efficient imaging modality for visualizing unstained collagen fibrils. The Type I collagen which is found on tendons, skin, and cornea is visualized bright in SHG images without any staining [9]. Besides, it produces a better visualization of collagen structures than fluorescent imaging of stained samples [10], [11]. As a result of these advantages, SHG microscopy has become the most preferred imaging modality for the analysis of collagen structures.

There are plenty of studies in the literature on the analysis of collagen distribution in SHG images. When we focus on the computational method papers, majority of these methods work in 2-dimensional space and their abilities are limited to detect basic properties such as pixel value or non-local statistical properties. Some of these studies focus on global or windowed texture analysis like directional variance [12], [13], [14], Fourier Transform and grey level co-occurrence matrix (GLCM) [15]. However, these methods are unable to detect the significant morphological features like fiber size or number [16]. Ogura et al. [17] used Fourier transform to quantify 2-dimensional SHG corneal collagen images. Fourier transform is widely used in texture analysis and gives successful results [18]. However, due to the nature of this method, local information about the morphology of collagen fibers cannot be detected. Therefore, such a technique will not be sufficient to characterize the detailed structure of collagen fibrils. In another study, Hu et al. [19] used the co-occurrence matrix method to distinguish different collagen tissues from each other. The co-occurrence matrix is a statistical method to measure the texture complexity of the given image [20]. This method, which does not give detailed information about individual collagen fibers and only looks at the general image texture, will not be sufficient to measure the features of collagen bundles individually. Besides these non-local analysis studies, there is a limited number of attempts conducted to detect collagen

fibers and extract more local properties. Yong Park et al. [21] used an automatic thresholding method to detect individual collagen fibers. Although this method aims to collect local information despite the previous ones, the thresholding method fails in many object segmentation tasks and does not give an efficient result in complex and low-contrasted images such as an SHG image of collagen. After separating the fibers from the background with thresholding, the authors applied texture analysis to several predetermined patches. However, the failure of the segmentation obtained by thresholding, as well as the analysis of small patches instead of the whole image, makes this method unreliable. Besides, this analysis was applied in 2-dimensional images instead of 3-dimensional images, so, there is a loss of information.

Liu et al. [13] aimed to detect the 3-dimensional structure of collagen bundles. However, since 3-dimensional analysis is computationally expensive, they applied the method referred to as 2.5 convolutions in the literature to reduce the computational load. This method aims to calculate the 2-dimensional projections of 3-dimensional images on three perpendicular planes, and then work with these three 2-dimensional images. Although this technique is suitable for plain images (for example, a single neuron cell image, a single isotropic object image, etc.), it can cause significant information to be lost in the collagen image by detecting separate bundles as a whole. This causes miscalculation of quantification, and creates a possible bias between images and misleads the data analysis.

In a recent study [22] SHG quantitative properties are used to distinguish morphological changes in collagen distribution in different basal cell carcinoma sub-types. For the Quantitative analysis the study focuses on the texture and the directional distribution of collagen bundles. For that purpose, beside the non-local frequency-based analysis, CurveAlign and CT-FIRE methods were used [23], [24]. Curve Align aims to quantify all fiber angles within a specified region [25]. On the other hand, CT-Fire performs a local directional analysis. CT-Fire aims to determine vessel-like structures in the input. As a result, CT-Fire is theoretically applicable for the current collagen bundle detection purpose, and this method forms a comparable alternative for the presented method. However CT-Fire is very sensitive to the contrast in the images. As a

result of that, objects can be detected in several disconnected components and due to that it may result with high number of false positive detection.

Some other methods from the literature are out of focus of this paper since the tasks and/or data acquisition methods are not comparable [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36].

In brief, despite the significance of automated quantification of collagen bundles, current methods in the literature are insufficient to detect local properties of 3-dimensional of these structures. This paper aims to introduce a directional analysis-based quantification method to detect collagen bundles in the 3-dimensional SHG microscopy images. Through that, the 3-dimensional structure of collagen bundles will be preserved, and local geometrical properties of bundles will be possible to be collected.

II. METHOD

All numerical analysis was performed in MATLAB 2023a. The proposed method is specified for the analysis of 3-dimensional SHG images of collagen distribution. All presented numerical results are for healthy data. However, the proposed method can be applied for diseased samples as well through optimizing the parameters accordingly.

The analysis steps and associated sub-steps can be seen in Fig. 1. The proposed method can be grouped into two main steps:

- A. Preprocessing.
- B. Collagen bundle detection.

A. Preprocessing step

Preprocessing is designed to take full advantage of the capabilities of instrumentation by reducing sources of image degradation such as blurring and noise. However, SHG images of collagen bundles require more than that. The collagen bundles in SHG images seem like a cluster of noise, and even by eye, it is very difficult to determine the exact boundaries of bundle regions. Hence, before the application of

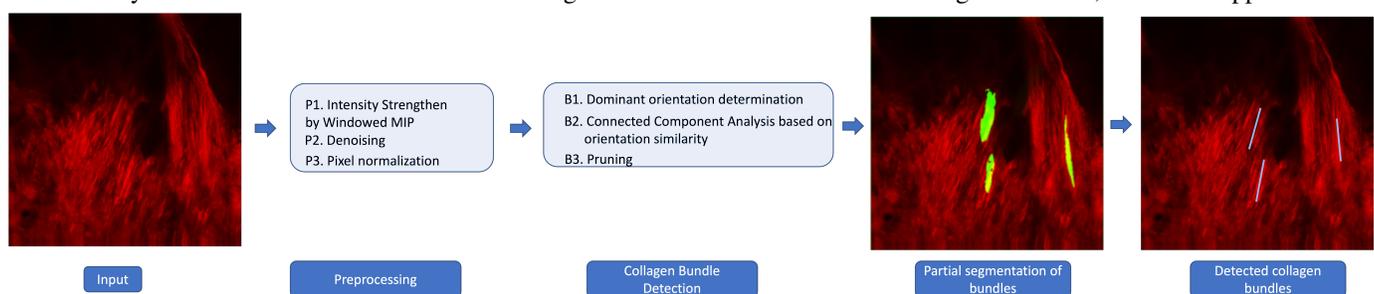


Fig. 1. The framework of the proposed method: The 3-dimensional input is firstly pre-processed by Windowed Maximum intensity projection (MIP), denoising and pixel intensity normalization. Then the pre-processed image is proceed to next step for collagen bundle detection. The output of this step is segmented regions which correspond to a part of collagen bundle. These segmentation are not guaranteed to be an accurate segmentation of bundles. However, they can be safely used for detection purpose. Hence, for next step, to form a better visualization of detected bundles, new output image is generated where detected bundles are represented with rectangular markers.

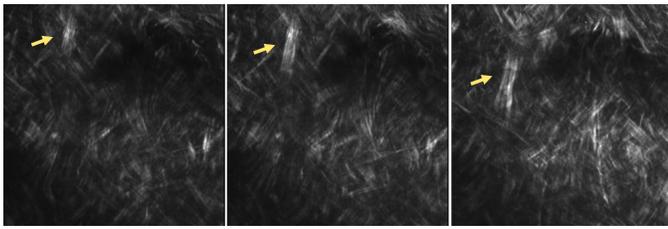


Fig. 2. The partition of a 3-dimensional bundle structure in the subsequent slices. As it is seen the bundle (marked with a yellow arrow) is oriented in the X-Z direction, as a result, it is not possible to detect the bundle in any 2-dimensional slice. However, by maximum filtering, it is possible to increase the visibility of bundles in individual slices (see Fig.3).

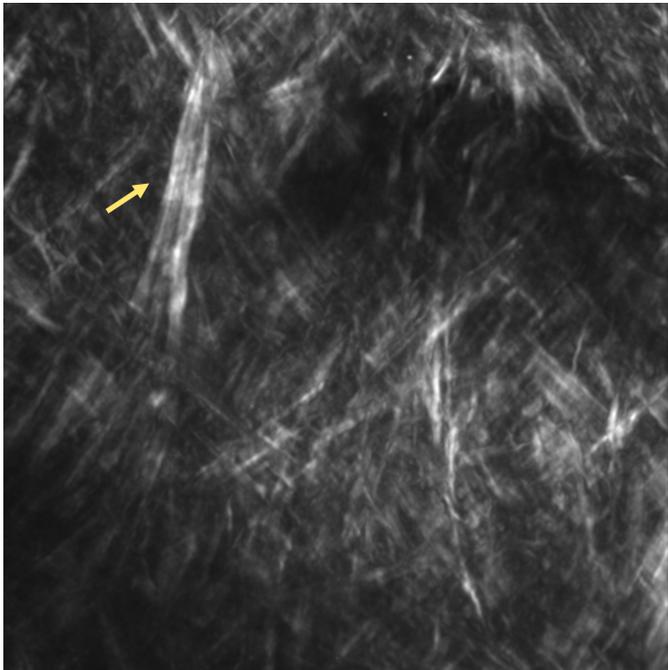


Fig. 3. After pre-processing of the slices in Fig. 2. As a result of maximum filtering on a certain window size, it is possible to visualize the 3-dimensional structure of collagen bundles more clearly in 2-dimensional slices. The bundle is marked with the yellow arrow as in Fig. 2.

noise reduction, an additional step to strengthen the pixel intensities on bundle regions is necessary. The 3-dimensional structure of collagen bundles helps to distinguish bundle regions from artifacts. Through that lead, as the first step of preprocessing, a maximum filter is applied in the X-Z-direction of input volume with a scale depending on the X-Z-direction resolution of the input. This process can be considered as a windowed maximum projection of the input volume. Through this strategy, the pixel brightness in bundle regions increases, and as a result, noise and blurring artifacts get fader. Besides, images are compressed in the X-Z-direction which benefits reducing the cost of the remaining computational analysis (see Fig. 2&3).

Through increased contrast of bundle regions, artifacts can be distinguished from collagen bundles. Thus, pre-processing continues with a denoising step. One of the most common denoising routines is Gaussian smoothing [37], which is based on convolving the input image with a Gaussian function. While

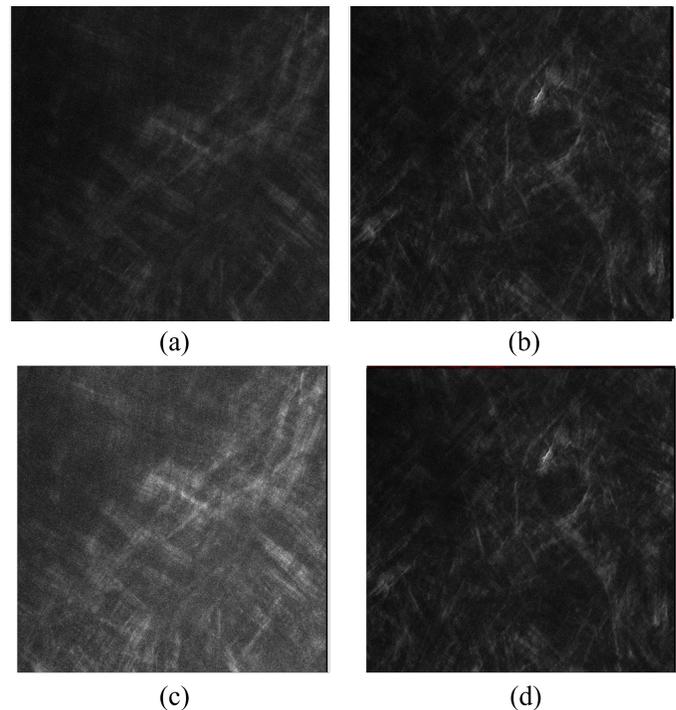


Fig. 4. Pixel normalization for each slice individually affects the visual quality significantly. The top images (a-b) are from the deeper, and the bottom images (c-d) are from surface regions. The first column (a-c) shows the volume-based normalization results, while the second column (b-d) shows slice-based normalization results. As seen slice-based normalization is significantly improving the contrast for deeper regions, while it has minimal effect on surface slices.

Gaussian smoothing is computationally highly efficient, it also has disadvantages such as edge surpassing due to the nature of the Gaussian function. However, since the SHG images of collagen bundles have weak edge information, and edges are not critical for the proposed method, that disadvantage is not a significant issue for the current task. Besides, due to the large image size, the low computational expense is a critical advantage. Hence 3-dimensional small-scale Gaussian filtering is used to smooth and denoise the images. As the last step of the pre-processing, the pixel intensity normalization is applied on each slice of the input volume, individually. Intensity normalization is significantly important when a threshold-based analysis or data clustering is included. In our analysis, besides its necessity for further steps, normalization is used to increase the contrast and have a more obvious view of bundles. The whole volume normalization is not efficient in SHG images since the pixel intensity constantly decreases while going deeper slices (while Z values are increasing). As a result, in the deeper regions, the detection of bundles is difficult even by eye (see Fig. 10). In order to handle that drawback and extract all possible information from each depth, the introduced method applies pixel-normalization for each 2-dimensional slice separately.

B. Collagen bundle detection

The proposed method is designed to detect and quantify the bundles only in a certain shape and visual condition.

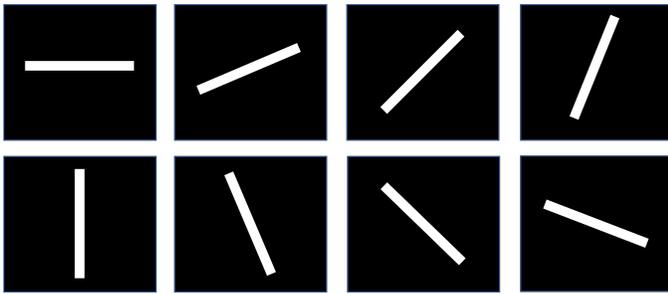


Fig. 5. The oriented rectangular filters evenly distributed into the X-Y plane.

In the SHG images, the collagen bundles are supposed to be in an anisotropic shape due to their structural nature. Hence, the proposed method ignores the isotropic high pixel intensity regions by accepting them as an artifact. Hence, the pixel intensity is not enough to detect bundles, but the intensity properties must be supported by geometric features. Since the images are noisy, and not all bundle regions are distinguishable, individual detection of each bundle is a challenging task. The proposed method is based on measuring the similarity of pixels based on their spatial and direction features. For that purpose, orientable unit filters are used to extract those geometrics features. Orientable unit filters are rectangular in shape and rotated in multiple orientations (see Fig. 5); thus, they can determine the dominant orientation of a close neighborhood of a point. Collagen bundle detection will be done in three steps:

- B1. Determination of dominant orientation for pixels in bundle region
- B2. Orientation similarity based Connected Component Analysis
- B3. Shape-constrained Pruning

Now, these three main steps will be explained in detail.

B.1. Dominant orientation determination of pixels in the bundle region: For that purpose, the input image is firstly convolved with the 3-dimensional rectangular prism directional filters. These filters are rotated in X-Y and X-Z plane with user determined angles (see Fig. 5 for rectangular filters rotated in X-Y plane.). The filter has size $a \times b \times c$, where a is the length, b is the height and c is the width. The filter size must be optimized according to the input image. After convolving the input image with filters in different orientations, the method determines the dominant orientation of the local neighborhood of each pixel by determining the direction of the maximum filtering response. As seen in Fig. 6, points have the maximum filtering response while the filter direction is close to the dominant orientation of the close neighborhood of the point. As a result, the orientation of a region can be determined by finding the general maxima of the filtering response. This idea is successfully used before for neurite orientation detection [38]. However, due to the high level of noise and complexity of the collagen images, further steps will be needed to use a similar idea for the current task. There are two major problems 1) intersecting bundle regions, and 2) large bundle regions which would give maximum filtering response at consecutive

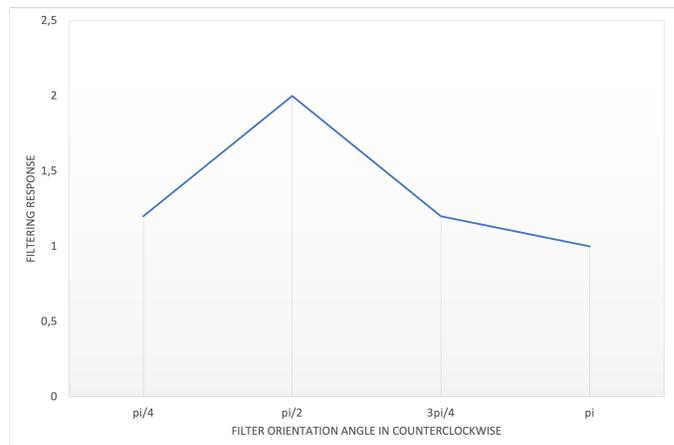
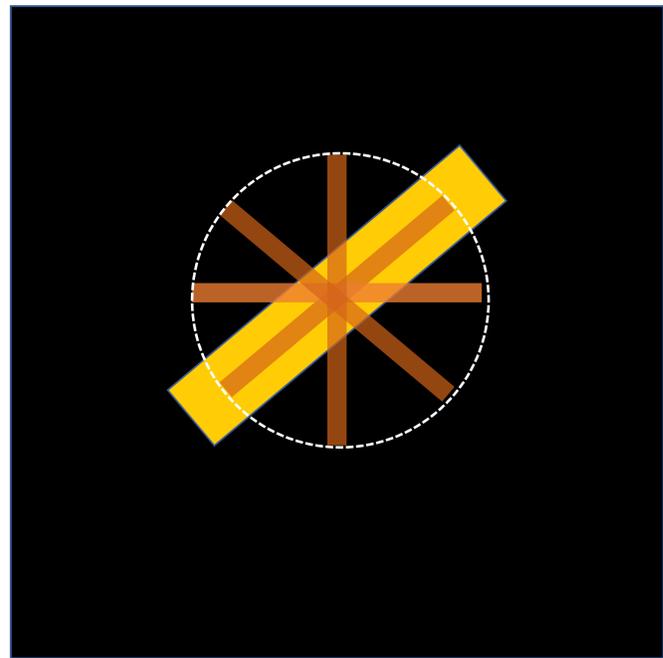


Fig. 6. The bundle orientation detection method on a synthetic image. (Top) A synthetic image where the collagen bundle is shown as the yellow anisotropic region on black background. Orange rectangular regions are the unit filters in four different orientations. (Bottom) The filtering response for each of the four filtering by (orange) oriented filters. The maximum filtering response is measured at the angle which is closest to the bundle direction ($\pi/4$). That angle gives the dominant orientation of the close neighborhood (circle with white dashed boundary) of the pixel in the center.

directions.

In the case of intersecting bundles, which is a common situation for SHG images of collagen structures, the pixels in the overlapping region have multiple dominant orientations (see Fig. 7). Subsequently, determining the general maxima of the filtering response would be misleading for such cases. By considering these problems, the proposed method is designed to detect at most up to two intersecting bundles. For that purpose, instead of detecting one direction as a dominant orientation, the method collects two directions with the largest filtering responses. If the angle between two dominant directions is larger than a certain threshold value, then that pixel is

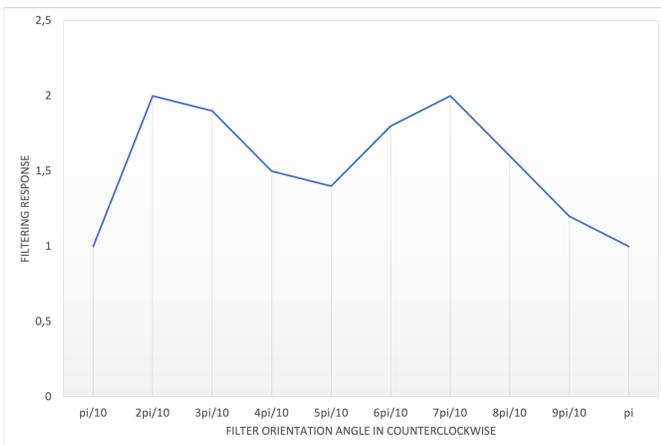
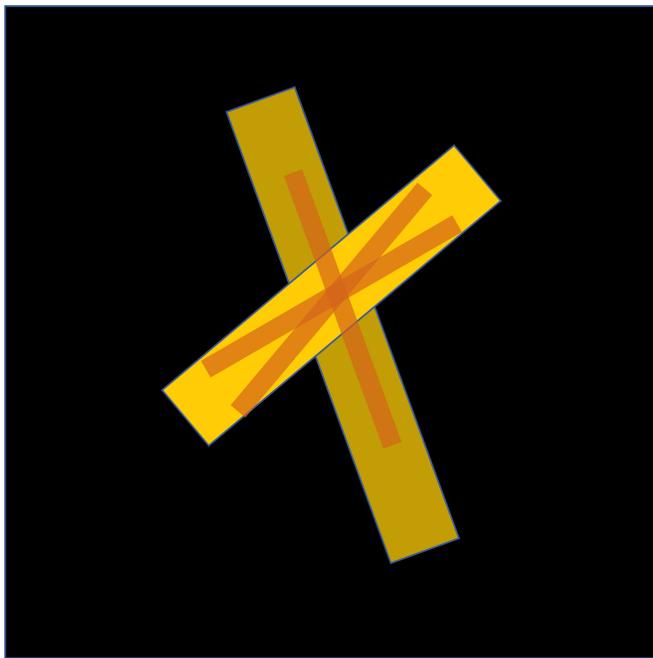


Fig. 7. The bundle orientation detection method on a synthetic image in the case of intersecting bundles or large bundle regions. (Top) A synthetic image where the collagen bundles are shown as the yellow anisotropic region on black background. Orange rectangular regions are the unit filters in different orientations. (Bottom) The filtering response of the centroid of the intersection of the bundles for ten oriented filters. As it is seen, the intersection point has two dominant orientations through two bundles at the angles $2\pi/10$ and $7\pi/10$. Besides the filtering response at angle $3\pi/10$ is close to the maximum although it is misleading due to the wide bundle region.

recorded with two dominant orientations. Otherwise, only one direction with the largest response is kept as the orientation of the neighborhood of the pixel. By examining the images in the dataset, more crowded intersection scenarios are neglected. The aim of using a lower threshold for the angle distance is to distinguish intersecting bundles from the single extra wide bundles. The extra wide bundles might have maximum filtering response in two consecutive angles even though they have a single dominant orientation (see Fig. 7). To avoid such problems, the algorithm assigns multiple dominant orientations to a pixel only if these orientations are distant from each other.

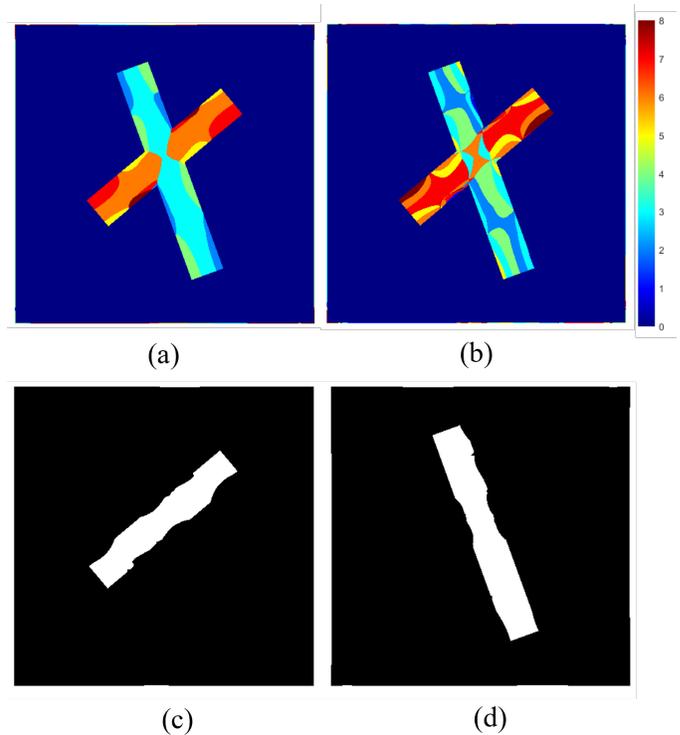


Fig. 8. Detection and segmentation of collagen bundles on a 2-dimensional synthetic image. Heat maps are showing the directions which yield maximum (a) and second maximum (b) filtering responses. These values show the dominant orientation of points in their local neighborhood. Due to the intersection of two bundles, there is a problem with assigning dominant orientations in their intersecting region. However, by considering these two filtering responses together and getting the connected components merged, both bundles are segmented with minimum error on boundaries (c-d).

B.2. Connected Component Analysis based on dominant orientation similarity: After the dominant orientation of each pixel is determined, pixels with the same dominant orientation are connected as being part of a collagen bundle. Through this process, the points with multiple dominant orientations are preserved in their nature and all possible connection paths are examined. As a result of this process, even the intersecting bundles are segmented separately by the proposed algorithm (Fig. 8).

B.3. Pruning: Since the nature of the collagen bundles is anisotropic, the post-processing step starts with determining the anisotropy level of segmented bundle regions through the Directional Ratio (DR). DR is a multiscale geometric descriptor to measure the isotropy level of the close neighborhood of a pixel. DR is calculated as the proportion of minimum to maximum filtering responses of a pixel over filters in multiple orientations [39]. DR approaches 1 while regions get more isotropic, and it approaches 0 while regions get more anisotropic [40], [41], [42]. It is used in previous studies to distinguish blob-like objects from vessel-like objects [40], [38], [39], [41]. For pruning, the algorithm calculates the DR of each connected component as the average DR of each pixel contained in the inner part of the region. For any region, even if the region is perfectly isotropic, the DR values of pixels will approach 0 while getting closer to the boundary [42]. Hence,

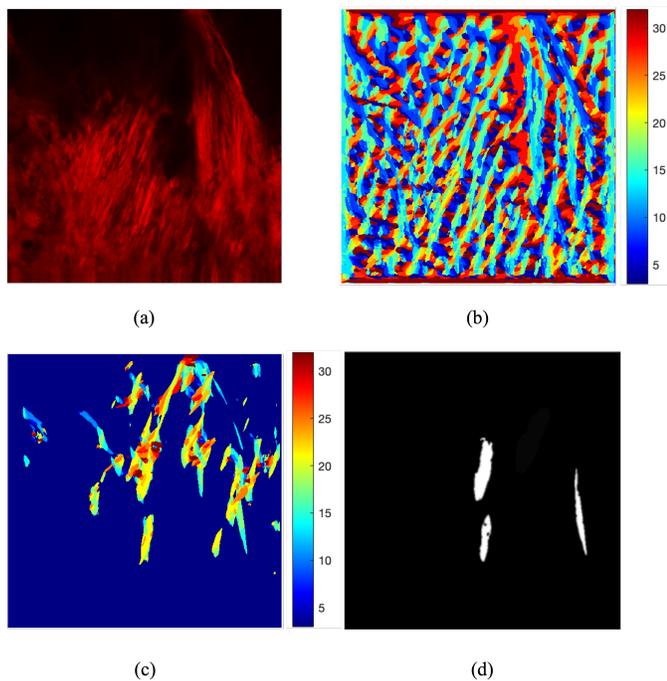


Fig. 9. Sample outputs of the intermediate steps in collagen bundle detection: a) Preprocessed image, b) Output of step B1. Dominant orientation determination of pixels in the bundle region. Since the orientations are enumerated as 1-32, this image has pixel intensities between 1-32. c) Output of step B2. Connected Component Analysis based on dominant orientation similarity. d) output of step B3. Pruning: Components in c) are eliminated if detections are isotropic and smaller than a certain volume threshold. Please note that these outputs are actually 3-dimensional images, however for visualization 2dimensional projections are given.

pixels close to the boundary are misleading to calculate the isotropy level of a region. Therefore, these misleading pixels are eliminated through removing the close-boundary parts of regions by morphological erosion. The remaining pixels' DR values are averaged to assign a DR value for each connected component. This DR value shows the anisotropy level of the whole region. To eliminate the isotropic regions a lower DR value threshold is applied. If the DR value is larger than the threshold that means the component is more isotropic than a usual collagen bundle. Hence, it will be accepted as an artifact and eliminated. As a result, the output image contains the segmented anisotropic collagen bundles. The computational cost of this part is neglectable since the filtering responses were already calculated in part B.1. for Dominant orientation determination of pixels.

III. RESULTS

The performance of the proposed method is measured on a publicly available dataset of SHG volume collected from three healthy rats [43]. The dataset contains around 4000 scans, but some of these images do not show the orientated nature of collagen bundles. Such images are eliminated and as a result, four volumes are selected to measure the performance of the method. A few slices of the samples which are ignored for the analysis can be seen in Fig. 10. As it is seen in the figure, these slices do not have oriented anisotropic patterns. Hence, such images are out of the scope of the presented method.

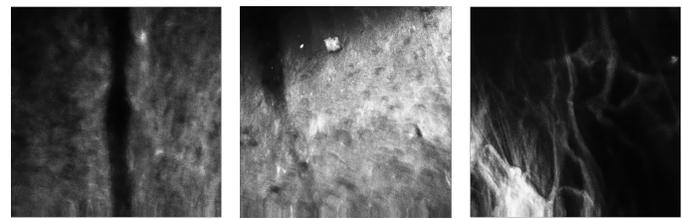


Fig. 10. Example images from the dataset that cannot be processed by the proposed method. These images do not have orientation information, bundles are not detectable by the eye and there are artifacts. Such images are ignored for sample selection for analysis.

The ground truth of the dataset was for semantic segmentation of similarly oriented regions. The pixels in the ground-truth images are in three categories as similar orientation, dissimilar orientation, and not interest. There was no label for direction or bundle detection. The presented method generates more local information as the pattern and orientation detection. Hence, the given ground truth was not useful for measuring the performance of the method. For that purpose, the detection and orientation labels are determined by the experts at Istanbul Medipol University, and the performance of the method is measured according to that.

The method has several parameters to be tuned for the given input. For the presented results, those parameters are optimized through the whole dataset, and these values are set as Default in the method. Filter size is set as 40x4x4 (with respect to x,y,z coordinates) pixels to catch the large anisotropic bundles efficiently. The analysis is performed in 8 directions in the X-Y-plane and 4 directions in the X-Z-plane, a total of 32 orientations. Since the images are so crowded and the same collagen bundle is visible in multiple slices, ground truth was collected for random 80 slices from samples by considering their 3-dimensional environment. The connected components whose volume is smaller than 3000 pixels are accepted as artifacts and eliminated. For each of the remaining connected components, the DR value is calculated, as explained in the Method section. The components with DR larger than 0.9 are eliminated since they are more isotropic than an accepted bundle. Then, the centroid of each connected component is calculated and detection results are demonstrated with a line that is oriented with the same angle of the bundle and centered at the centroid of the segmented region. As a result, large and vessel-like detections are collected as collagen bundles. The results of a few slices can be seen in Fig. 11, 14 and 15.

To measure the performance of the method Recall, Precision, and F1 score are used. Recall shows the performance of avoiding False negative detections, Precision shows the performance of avoiding False positive detections; while the F1 score shows the overall performance considering False Positives and Negatives. All these measurements range between 0 and 1, where 1 is the perfect score for each [44].

The method correctly detected 42 collagen bundles, out of 51 (Table I). The Recall score of the method is measured as 0.82. The Precision rate is measured low due to high False Positive detections. However, about half of the False detections were due to multiple detections of wide bundles. If such

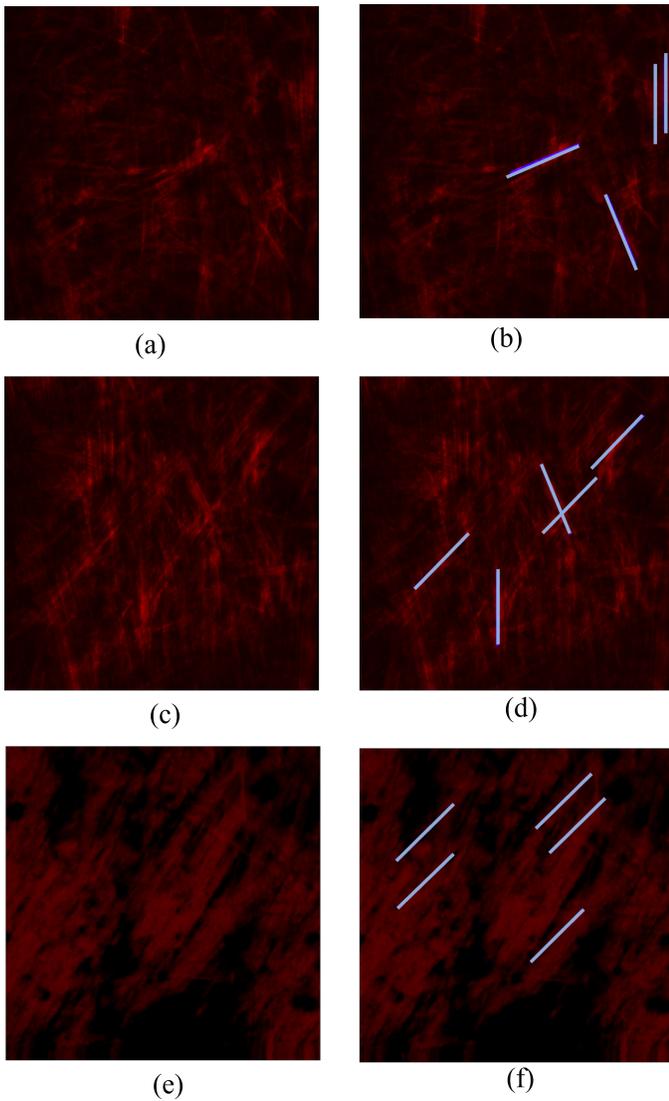


Fig. 11. Results of proposed method for collagen bundle detection. The raw image slices are shown in the right column (a, c, e), and associated detection results are given on the left (b, d, f) as an overlay on the original slices. Detected bundles are visualized with blue lines. The orientation of these lines indicates the dominant orientation of the detected collagen region.

	TP	FP	FN
multiple detections are included	42	19	9
multiple detections are excluded	42	11	9

TABLE I

DETECTION RESULTS OF THE METHOD WITH AND WITHOUT THE ELIMINATION OF FALSE POSITIVES DUE TO MULTIPLE DETECTIONS OF WIDE BUNDLES. TP STANDS FOR TRUE POSITIVE, FP STANDS FOR FALSE POSITIVE, AND FN STANDS FOR FALSE NEGATIVE DETECTIONS.

false detections were eliminated the Precision rate increased to 0.79 from 0.69 (Table II). Since orientation is the core of the proposed detection method, the accuracy of orientation detection for detected bundles was 100%.

IV. DISCUSSION

The SHG images of collagen bundles are highly complex and low-contrasted. Since the edge information is almost

	P	R	F1
multiple detections are included	0.69	0.82	0.75
multiple detections are excluded	0.79	0.82	0.81

TABLE II

PERFORMANCE RESULTS OF THE METHOD WITH AND WITHOUT ELIMINATION OF FALSE POSITIVES DUE TO MULTIPLE DETECTIONS OF WIDE BUNDLES. P STANDS FOR PRECISION, R STANDS FOR RECALL, AND F1 STANDS FOR F1 SCORE.

missing, collagen bundles seem like regions with uncertain boundaries. As a result, most detection methods fail for such cases. However, the presented method gives promising results through its advantage of using a connected component analysis through orientation similarity. This idea skips the edge or boundary information and focuses on local texture, which results in successful detection performance for the current task. As it is seen in Table II, the method has a high Recall score. False Positive detections are high despite low False Negative detections, which results in a lower Precision score (Table I & II). It is observed that some of the False Positives are due to multiple detections of one wide collagen bundle region. For example, in Fig. 11 (b), the rightmost two vertical detections are indicating the same region. Although this is not a significant error of the method, it could cause a problem if the detected bundle number is in focus since it will be overcalculated. In order to handle that an additional post-processing step can be included. When these multiple detections of wide bundles are ignored, the False Positive count reduces by almost 50% (Table I) and the F1 score increases from 0.75 to 0.81 (Table II).

On the other hand, False Positive detections are higher for the images where bundles are crossing each other wildly (Fig. 14). In such images, patterns are formed by short anisotropic regions, and large-scale filters cause the mixing of the information of multiple bundles. As a result, high False Positive detections occur. In order to handle that, a small-scale filter or multi-scale analysis could be used. By using multi-scale filters, short patterns can be detected as well as longer patterns.

The collagen bundles close to each other may have different orientations. As a result, in the SHG images, multiple collagen bundles crossing each other can be seen. Since this is a common issue, the proposed method is specifically designed to handle such cases. For that purpose, the method always examines if the points have two dominant orientations, as explained in detail in Section II. However, this may be misleading in some cases. For example, when the contrast is critically low, the method can be faulted to see noisy regions as if they are intersection regions of multiple bundles. This causes False Positive detections (Fig. 14 (d)). In order to handle that, a lower threshold for pixel intensity can be used. If the determined bundle region has an average pixel intensity lower than the threshold, it can be considered a noisy region instead of a bundle.

Another important aspect of the method is that not all bundle regions are targeted by the presented method. The bundles with certain shape and intensity properties can be detected. However, the method gives a consistent highly local texture

measurement. As a result, the number of detections as well as the orientation of detections can be significant indicators of multiple disorders.

The proposed method has multiple steps and for each step there are parameters to be adjusted for the input. Throughout the analysis it is observed the method's sensitivity to the selection of these parameters are different (see Table III). In the pre-processing step the method is sensitive to one parameter which is "Maximum filtering projection (MIP) window scale". That parameter must be adjusted according to the X-Z direction resolution of the input image. If the resolution is high default value can be increased to optimize the computational cost. But, the detection result is not highly sensitive to that parameter. Similarly, method is sensitive or highly sensitive to the several threshold values. When the input image has wildly crossing and thin bundle regions, "Threshold to determine intersecting bundles" in step B1 can be reduced. Otherwise multiple distinct bundles in close orientation can be detected as a whole (This would be a risk as well if the number of orientations is increased. So this threshold parameter is closely related with the user-determined number of orientation.). "The threshold to eliminate small regions" in step B2 can be decreased if the input image is highly noisy. "Threshold to eliminate isotropic regions" must be adjusted by observing the results. The complete list of these parameters and their default values are given in Table III.

During the analysis, it is observed that the method is highly sensitive to the number of orientations of the analysis and the scale of the filters. This is due to high noise and uncertain boundaries of collagen bundles. It is observed that when the number of orientations are increased the method tends to detect the same bundle multiple times with slightly different orientations. As a result of that, False Positives increase and the Precision rate decreases; or bundles are detected with incorrect orientation (see Fig. 12 (a)). On the other hand, when the number of orientations is increased a highly sensitive analysis results with one bundle region detected in several disconnected and differently oriented partitions. These partitions with different orientations can not be attached to each other by the connected component analysis (step B2). As a result of that, some large bundle regions can not be detected. As it is seen in (Fig. 12 (c)) two bundle regions are skipped by the method when the number of X-Y plane orientation is set to 20. For the current dataset, 8-10 orientations in the X-Y plane and 4 orientation on the X-Z plane was optimal.

Similarly, it was observed that method is sensitive to the selection of filter scale. When the filter scale is small the case of multiple detection of the same bundle is more probable. As a result of that, the number of False Positives increase and the Precision rate decreases. However, when the bundles are crossing with each other widely smaller-scale filters would work better. In order to measure the proposed method's sensitivity to the filter scale selection, 40 slices from two volumes are selected to reduce the computational burden. These samples were chosen carefully to cover all different situations to be fair on the sensitivity analysis. These two volumes were analysed with varying filter scales and the performance metrics are compared with each other. As it is seen in Fig. 13, for any

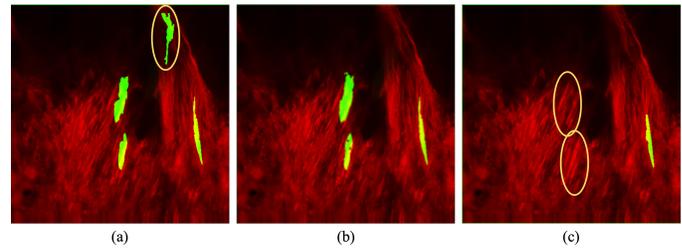


Fig. 12. Detected collagen bundles outputs when the proposed method is performed with different number of orientations. A) X-Y plane orientation is 4, B) X-Y plane orientation is 8 and C) X-Y plane orientation is 20. For all cases X-Z plane orientation number was fixed to 4. Since the resolution of volumes in X-Z plane is smaller than X-Y, the parameter sensitivity analysis on X-Z plane is ignored. Errors are shown by yellow ellipses.

Method Step	Parameter	Sensitivity of method to the selection	Default
Preprocessing	MIP window size	Sensitive	3 pixels in XZ direction
	Denosing filter scale	Low sensitivity	1
Collagen Bundle Detection B1	Threshold to determine intersecting bundles	Highly sensitive	$\pi/8$
	Filter scale	Highly sensitive	40 x 4 x 4
	Orientation number	Sensitive	8 x 4
Connected Component Analysis B2	Threshold to eliminate small regions	Sensitive	5000 pixels
Collagen Bundle Detection B3	Threshold to eliminate isotropic regions	Sensitive	0.93
	Erosion filter scale	Low sensitivity	4 pixels

TABLE III
LIST OF PARAMETERS TO BE ADJUSTED FOR INPUT

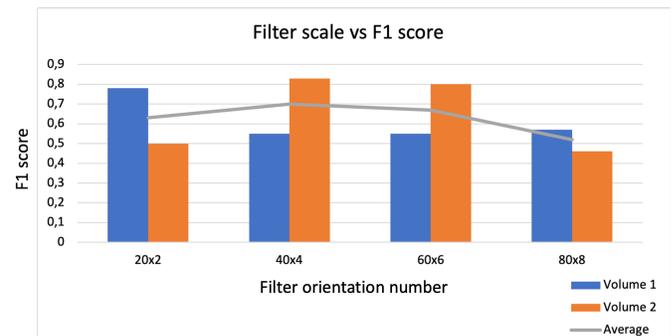


Fig. 13. The change in F1 score with the analysis with different filter scales. X-Z plane analysis is ignored since the resolution in X-Z direction is small. Given filter scales are for the length and width respectively. For each case height was set as same with width. As it is seen the average F1 score is maximized when 40x4 scale filters are used. However, for volume 1 the performance is maximized with the smallest scale 20x2.

filter scale the variation of F1 scores between the analysis of different volumes were critically high. Besides, the results show that the accuracy of the method is highly sensitive to the selection of filter scale. As it is seen in Fig. 13, in Volume 1 the maximum performance is obtained with smallest scale filter, while that scale yielded the worst performance result in the Volume 2. As a result, in order to show the overall performance of the method with optimal effort, parameters are numerically optimized and set to a one value for the whole dataset. Hence, the reported numerical results can be improved by optimizing parameters individually for each input volume. For the further studies, an automated parameter optimization step could be included.

The dataset originally had ground-truth labels for semantic

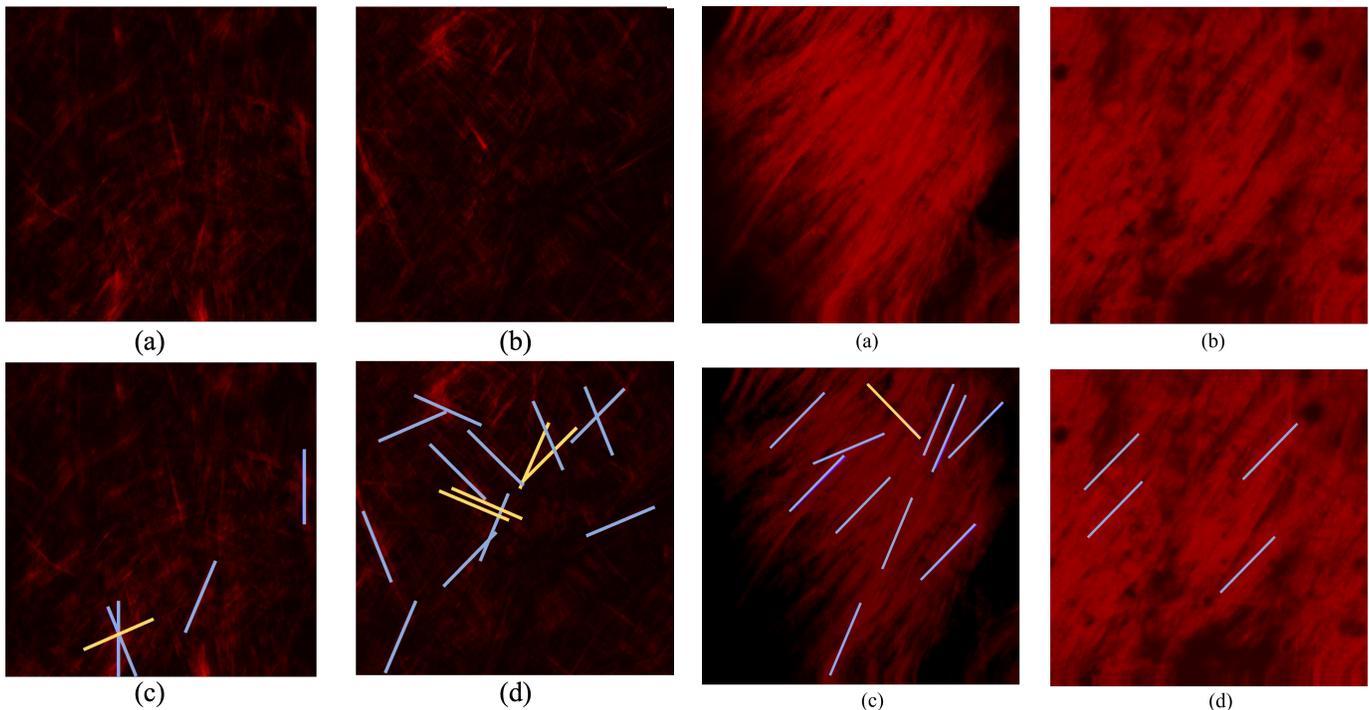


Fig. 14. Some examples for False Positive detection: Raw images are given in (a), (b), and the corresponding detection are given (c), (d). True Positives are represented with blue lines while False Positives are represented with yellow.

segmentation. Since the proposed method has no application for pixel segmentation, the given labels could not be used. However, they can still be used to show the consistency of the presented method. As it is seen in Fig. 15 (e-f), the correctly detected bundles in the green region, which indicates the ground-truth labels for similarly oriented regions, are in highly close orientations.

In order to compare the performance of the proposed method with the state of the art methods, CT-FIRE is selected as a reference. CT-FIRE is the best candidate since it has the similar objectives with the presented method. Besides CT-FIRE has applications in the literature of collagen bundle analysis [22]. Since the CT-FIRE is analyzing volumes slice by slice the 3-dimensional analysis was computationally highly expensive. As a result, instead of analyzing the whole dataset, the performance comparison is done only on the half of the dataset. For that purpose, 40 slices from each volume is collected instead of 80. By observing the results, it can be said that comparison is fair due to the consistent behavior of CT-FIRE based on errors and successes. To have a fair comparison, parameters of CT-FIRE are optimized numerically for the all test samples. When we compared the performance of CT-FIRE with the proposed method, it is observed that CT-FIRE tends to have high False Positive detections (see Table IV). One major reason is that CT-FIRE fails to detect regions close to the image boundaries as a bundle due to high contrast change (Fig. 16 (a), (c)). As a result, for all test images, CT-FIRE yield a very large False Positive detection value when it is compared with the proposed method. Besides, CT-FIRE detected some bundle regions with wrong angle.

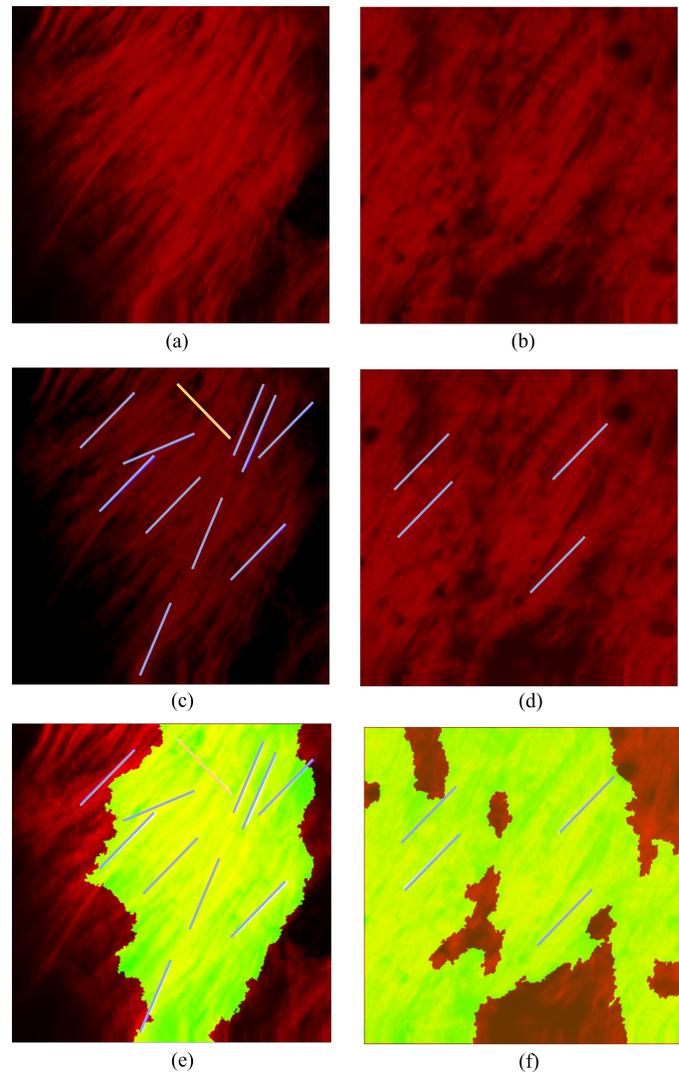


Fig. 15. Comparison of detection results with the ground-truth pixel-wise orientation similarity labels. a-b) original raw images, c-d) overlay of the presented method's detections in blue with the raw images. There is only 1 False Positive detection, which is represented with the orange line. e-f) overlay of ground-truth labels for similarly oriented regions in the green channel, detected bundles in blue, and bundle regions in red.

These detections were counted as False Positive again, since the bundle orientation is a significant morphological feature. As it is seen in the Table IV the proposed method has F1 score 94 % while CT-FIRE reaches 44 %. Although CT-FIRE is successful to detect vessel-like structures, this method is so sensitive to contrast change. As a result, in the noisy images such as SHG images of collagen bundles, it could not reach the performance of proposed method.

The performance analysis of the method is done on samples from healthy rats. This could be a limitation since the method has not been applied to any sample with an abnormality. The structural properties of fibrillar collagen bundles tend to vary due to countless diseases and abnormalities. One of these properties is known as the curvature of the bundle both on small and large scales [45], [46]. Although the presented method is designed only to detect linear bundles, it is observed

Filter scale	TP	FN	FP	P	R	F1
CT-FIRE	11	11	17	0.39	0.5	0.44
Presented Method	21	2	1	0.96	0.91	0.94

TABLE IV

COMPARISON OF PERFORMANCE METRICS RESULT FOR CT-FIRE AND PRESENTED METHOD.

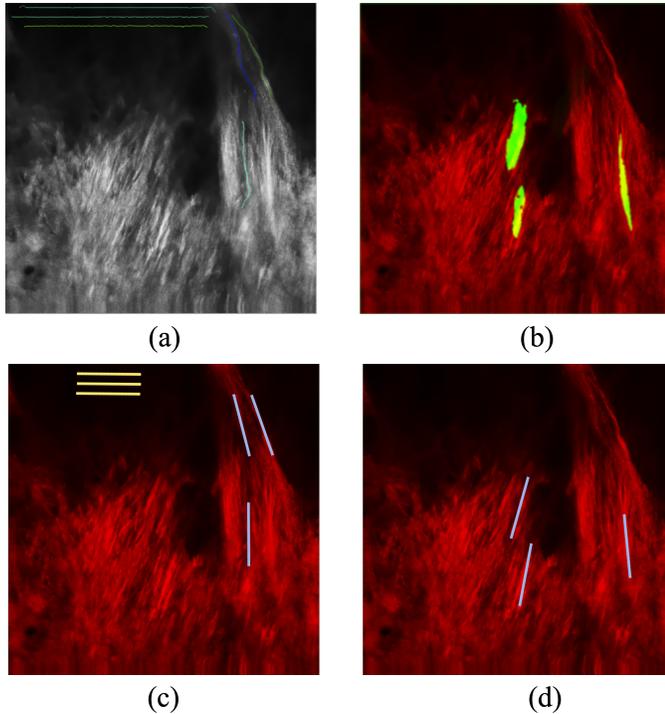


Fig. 16. Comparison of outputs of CT-FIRE and presented algorithm: a-b outputs of CT-FIRE and presented method respectively; c-d a better visualization of detections results with representative rectangular markers. Correct detections are represented with blue and wrong detections represented with orange rectangular marker.

that the method can also detect curvatures on a large scale (Fig. 17). This leads to the idea that by adjusting the filter scale the presented method would be able to determine curvy collagen bundles which are seen as a sign of multiple abnormalities [45], [46]. Hence, even if the method is only validated with healthy collagen bundles, it still has the ability to detect abnormalities in morphology by tuning the parameters. As a result, the presented method can be used for classification purposes to distinguish healthy samples from diseased samples.

V. CONCLUSION

Due to the advantages of SHG and the critical importance of collagen bundles, several studies in the literature focus on SHG images of collagen bundles. However, most of these studies are designed to detect non-local statistical properties of collagen distribution. This led to the loss of local morphological information collagen bundles which are known to be highly significant indicators of multiple abnormalities. The introduced method uses the orientation information to detect individual collagen bundles in 3-dimensional SHG images even if they intersect each other. Hence, the presented method gives deeper

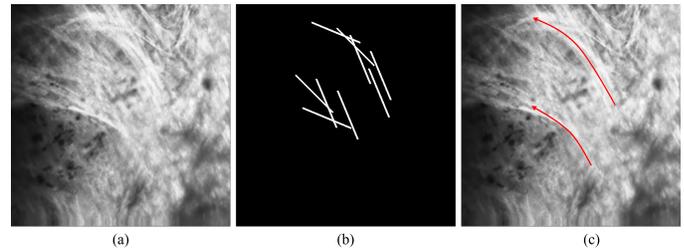


Fig. 17. Curvature detection: (a) raw image, (b) corresponding detected bundles given in white lines, (c) curvy regions path is shown with red arrow. As it is seen through multiple detections of bundles the large curvatures can be determined by the method. This led to the idea that by using small-scale filters it would be possible to detect curvy collagen bundles' structures.

and more detailed information about the distribution and structure of collagen bundles in SHG images. That would lead to more accurate data analysis methods for collagen analysis, which finally lead to better diagnosis methods for collagen-related abnormalities such as cornea diseases and some cancer types.

VI. ACKNOWLEDGMENT

The author thanks Dr. Cafer Tanrıverdi, Mehmet Şerif Aydın, and Olgu Enis Tok for their expertise and valuable assistance for validating the proposed method.

REFERENCES

- [1] K. M. Meek and C. Knupp, "Corneal structure and transparency," *Progress in retinal and eye research*, vol. 49, pp. 1–16, 2015.
- [2] C. Raoux, M. Schmeltz, M. Bied, M. Alnawaiseh, U. Hansen, G. Latour, and M.-C. Schanne-Klein, "Quantitative structural imaging of keratoconic corneas using polarization-resolved shg microscopy," *Biomedical optics express*, vol. 12, no. 7, pp. 4163–4178, 2021.
- [3] M. S. Sridhar, "Anatomy of cornea and ocular surface," *Indian journal of ophthalmology*, vol. 66, no. 2, p. 190, 2018.
- [4] H.-Y. Zhou, Y. Cao, J. Wu, and W.-S. Zhang, "Role of corneal collagen fibrils in corneal disorders and related pathological conditions," *International journal of ophthalmology*, vol. 10, no. 5, p. 803, 2017.
- [5] G. A. Di Lullo, S. M. Sweeney, J. Korkko, L. Ala-Kokko, and J. D. San Antonio, "Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type i collagen," *Journal of Biological Chemistry*, vol. 277, no. 6, pp. 4223–4231, 2002.
- [6] S. Xu, H. Xu, W. Wang, S. Li, H. Li, T. Li, W. Zhang, X. Yu, and L. Liu, "The role of collagen in cancer: from bench to bedside," *Journal of translational medicine*, vol. 17, pp. 1–22, 2019.
- [7] R. M. Martínez-Ojeda, M. D. Pérez-Cárceles, L. C. Ardelean, S. G. Stanciu, and J. M. Bueno, "Multiphoton microscopy of oral tissues," *Frontiers in Physics*, vol. 8, p. 128, 2020.
- [8] E. A. Gibson, O. Masihzadeh, T. C. Lei, D. A. Ammar, and M. Y. Kahook, "Multiphoton microscopy for ophthalmic imaging," *Journal of ophthalmology*, vol. 2011, 2011.
- [9] R. LaComb, O. Nadiarnykh, and P. J. Campagnola, "Quantitative second harmonic generation imaging of the diseased state osteogenesis imperfecta: experiment and simulation," *Biophysical journal*, vol. 94, no. 11, pp. 4504–4514, 2008.
- [10] S. V. Plotnikov, A. C. Millard, P. J. Campagnola, and W. A. Mohler, "Characterization of the myosin-based source for second-harmonic generation from muscle sarcomeres," *Biophysical journal*, vol. 90, no. 2, pp. 693–703, 2006.
- [11] S.-W. Chu, S.-Y. Chen, G.-W. Chern, T.-H. Tsai, Y.-C. Chen, B.-L. Lin, and C.-K. Sun, "Studies of $\chi(2)/\chi(3)$ tensors in submicron-scaled bio-tissues by polarization harmonics optical microscopy," *Biophysical journal*, vol. 86, no. 6, pp. 3914–3922, 2004.
- [12] Z. Liu, K. P. Quinn, L. Speroni, L. Arendt, C. Kuperwasser, C. Sonnenschein, A. M. Soto, and I. Georgakoudi, "Rapid three-dimensional quantification of voxel-wise collagen fiber orientation," *Biomedical optics express*, vol. 6, no. 7, pp. 2294–2310, 2015.
- [13] Z. Liu, D. Pouli, D. Sood, A. Sundarakrishnan, C. K. H. Mingalone, L. M. Arendt, C. Alonzo, K. P. Quinn, C. Kuperwasser, L. Zeng *et al.*, "Automated quantification of three-dimensional organization of fiber-like structures in biological tissues," *Biomaterials*, vol. 116, pp. 34–47, 2017.
- [14] E. C. Rentchler, K. L. Gant, R. Drapkin, M. Patankar, and P. J. Campagnola, "Imaging collagen alterations in stics and high grade ovarian cancers in the fallopian tubes by second harmonic generation microscopy," *Cancers*, vol. 11, no. 11, p. 1805, 2019.
- [15] J. M. Watson, P. F. Rice, S. L. Marion, M. A. Brewer, J. R. Davis, J. J. Rodriguez, U. Utzinger, P. B. Hoyer, and J. K. Barton, "Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma," *Journal of Biomedical Optics*, vol. 17, no. 7, pp. 076002–076002, 2012.
- [16] D. S. James and P. J. Campagnola, "Recent advancements in optical harmonic generation microscopy: Applications and perspectives," *BME Frontiers*, vol. 2021, 2021.
- [17] Y. Ogura, Y. Tanaka, E. Hase, T. Yamashita, and T. Yasui, "Texture analysis of second-harmonic-generation images for quantitative analysis of reticular dermal collagen fibre in vivo in human facial cheek skin," *Experimental Dermatology*, vol. 28, no. 8, pp. 899–905, 2019.
- [18] T. Hsu, A. Calway, and R. Wilson, "Texture analysis using the multiresolution fourier transform," *Bristol, UK, Tech. Rep.*, 1993.
- [19] W. Hu, H. Li, C. Wang, S. Gou, and L. Fu, "Characterization of collagen fibers by means of texture analysis of second harmonic generation images using orientation-dependent gray level co-occurrence matrix method," *Journal of biomedical optics*, vol. 17, no. 2, pp. 026007–026007, 2012.
- [20] C. C. Gotlieb and H. E. Kreytzig, "Texture descriptors based on co-occurrence matrices," *Computer vision, graphics, and image processing*, vol. 51, no. 1, pp. 70–86, 1990.
- [21] C. Y. Park, J. K. Lee, and R. S. Chuck, "Second harmonic generation imaging analysis of collagen arrangement in human cornea," *Investigative ophthalmology & visual science*, vol. 56, no. 9, pp. 5622–5629, 2015.
- [22] M. Sendín-Martín, J. Posner, U. Harris, M. Moronta, J. Conejo-Mir Sánchez, S. Mukherjee, M. Rajadhyaksha, K. Kose, and M. Jain, "Quantitative collagen analysis using second harmonic generation images for the detection of basal cell carcinoma with ex vivo multiphoton microscopy," *Experimental Dermatology*, vol. 32, no. 4, pp. 392–402, 2023.
- [23] Y. Liu, A. Keikhosravi, G. S. Mehta, C. R. Drifka, and K. W. Eliceiri, "Methods for quantifying fibrillar collagen alignment," *Fibrosis: methods and protocols*, pp. 429–451, 2017.
- [24] J. S. Bredfeldt, Y. Liu, C. A. Pehlke, M. W. Conklin, J. M. Szulcowski, D. R. Inman, P. J. Keely, R. D. Nowak, T. R. Mackie, and K. W. Eliceiri, "Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer," *Journal of biomedical optics*, vol. 19, no. 1, pp. 016007–016007, 2014.
- [25] Y. Liu, A. Keikhosravi, C. A. Pehlke, J. S. Bredfeldt, M. Dutson, H. Liu, G. S. Mehta, R. Claus, A. J. Patel, M. W. Conklin *et al.*, "Fibrillar collagen quantification with curvelet transform based computational methods," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 198, 2020.
- [26] J. Liu, M.-y. Xu, J. Wu, H. Zhang, L. Yang, D.-x. Lun, Y.-c. Hu, and B. Liu, "Picosirius-polarization method for collagen fiber detection in tendons: A mini-review," *Orthopaedic Surgery*, vol. 13, no. 3, pp. 701–707, 2021.
- [27] Y. Zhang, Y. Chen, B. Zhao, J. Gao, L. Xia, F. Xing, Y. Kong, Y. Li, and G. Zhang, "Detection of type i and iii collagen in porcine acellular matrix using hplc-ms," *Regenerative biomaterials*, vol. 7, no. 6, pp. 577–582, 2020.
- [28] X. Wang, B. Le, N. Zhang, K. H. Bak, Y. Zhang, and Y. Fu, "Off-flavour compounds in collagen peptides from fish: Formation, detection and removal," *International Journal of Food Science & Technology*, vol. 58, no. 3, pp. 1543–1563, 2023.
- [29] J. Zhang, Y. Ning, H. Zhu, N. J. Rotile, H. Wei, H. Diyabalanage, E. C. Hansen, I. Y. Zhou, S. C. Barrett, M. Sojoodi *et al.*, "Fast detection of liver fibrosis with collagen-binding single-nanometer iron oxide nanoparticles via t 1-weighted mri," *Proceedings of the National Academy of Sciences*, vol. 120, no. 18, p. e2220036120, 2023.
- [30] M. Salarian, H. Yang, R. C. Turaga, S. Tan, J. Qiao, S. Xue, Z. Gui, G. Peng, H. Han, P. Mittal *et al.*, "Precision detection of liver metastasis by collagen-targeted protein mri contrast agent," *Biomaterials*, vol. 224, p. 119478, 2019.
- [31] O. Y. Ibhagui, D. Li, H. Han, G. Peng, M. L. Meister, Z. Gui, J. Qiao, M. Salarian, B. Dong, Y. Yuan *et al.*, "Early detection and staging of lung fibrosis enabled by collagen-targeted mri protein contrast agent," *Chemical & Biomedical Imaging*, 2023.
- [32] C. Roa, V. N. Du Le, M. Mahendroo, I. Saytashev, and J. C. Ramella-Roman, "Auto-detection of cervical collagen and elastin in mueller matrix polarimetry microscopic images using k-nn and semantic segmentation classification," *Biomedical Optics Express*, vol. 12, no. 4, pp. 2236–2249, 2021.
- [33] M. Zaffar and A. Pradhan, "Assessment of anisotropy of collagen structures through spatial frequencies of mueller matrix images for cervical pre-cancer detection," *Applied Optics*, vol. 59, no. 4, pp. 1237–1248, 2020.
- [34] J. Li, Y. Chen, W. Zhi, and Q. Cheng, "Photoacoustics spectral analysis for in vivo detection of collagen contents in cancers," in *2022 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2022, pp. 1–4.
- [35] J.-E. Cota, A. Spadigam, and A. Dhupar, "Detection of type vii collagen in odontogenic keratocyst: An immunohistochemical study," *Journal of clinical and experimental dentistry*, vol. 11, no. 4, p. e310, 2019.
- [36] J. T. Stefano, L. V. Guedes, A. A. A. de Souza, D. S. Vanni, V. A. F. Alves, F. J. Carrilho, A. Largura, M. Arrese, and C. P. Oliveira, "Usefulness of collagen type iv in the detection of significant liver fibrosis in nonalcoholic fatty liver disease," *Annals of hepatology*, vol. 20, p. 100253, 2021.
- [37] R. C. Gonzalez, *Digital image processing*. Pearson education india, 2009.
- [38] C. Kayasandik, P. Negi, F. Laezza, M. Papadakis, and D. Labate, "Automated sorting of neuronal trees in fluorescent images of neuronal networks using neurotree-tracer," *Scientific reports*, vol. 8, no. 1, p. 6450, 2018.
- [39] D. Labate, F. Laezza, P. Negi, B. Ozcan, and M. Papadakis, "Efficient processing of fluorescence images using directional multiscale representations," *Mathematical modelling of natural phenomena*, vol. 9, no. 5, pp. 177–193, 2014.
- [40] C. B. Kayasandik and D. Labate, "Improved detection of soma location and morphology in fluorescence microscopy images of neurons," *Journal of neuroscience methods*, vol. 274, pp. 61–70, 2016.

- [41] B. Ozcan, P. Negi, F. Laezza, M. Papadakis, and D. Labate, "Automated detection of soma location and morphology in neuronal network cultures," *PLoS one*, vol. 10, no. 4, p. e0121886, 2015.
- [42] C. Kayasandik, K. Guo, and D. Labate, "Directional multiscale representations and applications in digital neuron reconstruction," *Journal of computational and applied mathematics*, vol. 349, pp. 482–493, 2019.
- [43] L. Schmarje, C. Zelenka, U. Geisen, C.-C. Glüer, and R. Koch, "2d and 3d segmentation of uncertain local collagen fiber orientations in shg microscopy," in *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41*. Springer, 2019, pp. 374–386.
- [44] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.
- [45] J. N. Ouellette, C. R. Drifka, K. B. Pointer, Y. Liu, T. J. Lieberthal, W. J. Kao, J. S. Kuo, A. G. Loeffler, and K. W. Eliceiri, "Navigating the collagen jungle: the biomedical potential of fiber organization in cancer," *Bioengineering*, vol. 8, no. 2, p. 17, 2021.
- [46] E. Brown, T. McKee, E. DiTomaso, A. Pluen, B. Seed, Y. Boucher, and R. K. Jain, "Dynamic imaging of collagen and its modulation in tumors in vivo using second-harmonic generation," *Nature medicine*, vol. 9, no. 6, pp. 796–800, 2003.



Cihan Bilge Kayasandik is an assistant professor in the Department of Computer Engineering at Istanbul Medipol University, where she has been a faculty member since 2019. She graduated with B.Sc. degree from the Department of Mathematics at Bilkent University in 2012. Then she completed her Ph.D. degree at the Department of Mathematics at University of Houston in 2017. She worked as a researcher in the same university for one more year. Her current research areas include medical image and data analysis.