IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 15, 2021, accepted April 1, 2021, date of publication April 20, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074514

Investigation of Bilateral Similarity in Central Retinal Blood Vessels

SANGEETA BISWAS[1,2], JOHAN ROHDIN[1], ANDRII KAVETSKYI[1], GABRIEL SARAIVA[1,3], ANGKAN BISWAS[1,4], AND MARTIN DRAHANSKY[1], (Senior Member, IEEE)

[1]Faculty of Information Technology, Brno University of Technology, 61200 Brno, Czech Republic
[2]Faculty of Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh
[3]Institute of Computing, Federal University of Amazonas, Manaus 69067-005, Brazil
[4]CAPM Company Ltd., Dhaka 1213, Bangladesh

Corresponding author: Sangeeta Biswas (biswas@fit.vutbr.cz)

This work was supported by the Project "Reliable, Secure, and Efficient Computer Systems", BUT, under Grant FIT-S-20-6427.

ABSTRACT While our left and right eyes clearly have a high degree of external bilateral similarity, it is less obvious to what degree they have internal bilateral similarity. This is especially true for the central retinal blood vessels (CRBVs), which are responsible for supplying blood to retinas and also can be used as a strong biometric. In this paper, we investigate whether the CRBVs of the left and right retinas possess strong enough bilateral similarity so that we reliably tell whether a pair of the left and right retinas belong to a single subject. We evaluate and analyse the performance of both human- and deep neural network-based bilateral verification by experimenting on two publicly available data sets.

INDEX TERMS Retina, symmetry, central retinal blood vessels, deep neural network, biometric system.

I. INTRODUCTION

Symmetry can be defined as *uniformity*, *equivalence*, or *exact similarity* of two parts divided along an axis, whereas *asymmetry* indicates lack of equivalence or similarity. Paired organs, such as eyes, ears, hands, and legs, give a bilateral symmetrical look to the exterior of our human body by dividing it into two parts through an imaginary left-right axis. The bilateral symmetry of paired organs gives a sense of beauty. In medical science, it helps doctors to detect and monitor the progress of unilateral diseases as well as to use one side of a paired organ as a proxy of the other side for post-treatment analysis. In biometric, it opens a possibility of developing side independent biometric systems in which a biometric extracted from one side of a paired organ, such as iris, retina, and palmprint, can be used to access a system developed for the opposite side. The study of bilateral symmetry may also help us to understand how strong a biometric system would be if both sides of a paired organ are used jointly. If the left and right organ based biometrics of a subject are completely different, then an authentication system using both sides would, loosely speaking, be *two times stronger* than an authentication system using one side.

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

None of our paired body organs have identical left and right forms. That means our human body shows *approximate-bilateral*, *pseudo-bilateral*, or *nearly-bilateral* symmetry instead of *perfect-bilateral* symmetry. Furthermore, for many paired organs that seem approximately symmetric from the outside, it is not evident whether they are symmetric internally. Most of the time, the internal sides of paired organs look *asymmetric* instead of *symmetric*. It is, in particular, true for the internal parts of our eyes, e.g., the retina. Especially when 2D color fundus photographs are used for the left and right retinas, the unique *tree* or *venation*-like structure of the central retinal blood vessels (CRBVs) spreading over the retina gives a bilateral asymmetrical look to color fundus photographs at first glance. As shown in Fig. 1, the outside of our eyes has a noticeable approximate-bilateral symmetrical look and the colored retinas show a tiny bilateral symmetrical look. In contrast, the structures of CRBVs of the left and right retinas seem almost unique.

Although the CRBVs of a subject's left and right retinas look asymmetric at first glance, they may be similar in a way that can be detected by humans if they look more carefully as well as by machine learning algorithms. In this paper, using human volunteers and deep neural networks (NNs), we investigate bilateral similarity in the CRBVs of the left and right retinas. Instead of measuring specific features of

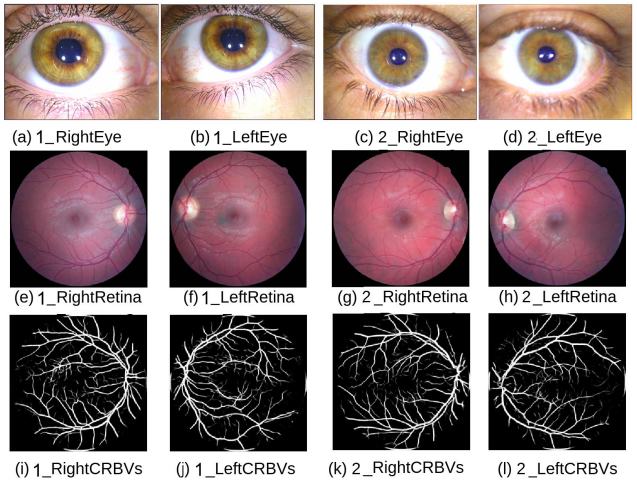


FIGURE 1. Comparison between two pairs of eyes, retina, and segmented CRBVs by a neural network of two subjects. It is easy to see bilateral symmetry in our eyes from the outside. However, seeing bilateral symmetry in retinas is difficult, especially in CRBVs.

the CRBVs such as length or caliber of CRBVs, as is typical in the research of medical science, in this paper, we turn the problem of finding bilateral similarity in CRBVs into a binary classification problem with the segmented CRBVs as input. The basic assumption is that when there is substantial bilateral similarity in a pair of CRBVs, there is a high probability that the pair belong to a single subject. We investigate if it is possible for both untrained and trained human volunteers as well as for deep NNs to tell whether a pair of CRBVs of the left and right retinas belong to a single subject or to two different subjects. From the perspective of biometric, this is a side independent verification task. In the usual CRBVs based biometric system, CRBVs extracted from the retina of the same side are compared in the verification, whereas in this paper, we use CRBVs segmented from both sides of retinas.

In principle, without observing any examples of positive and negative pairs, it is not possible to know in which sense CRBVs from the same subject are similar. For example, if the caliber of the CRBVs depends on the subject, how much can it vary between the two retinas of one subject and how much can it vary from subject to subject? By observing labeled training data, both humans and deep NNs may be able to learn what type of similarities indicates that two retinas are from the same subject. Untrained volunteers may also have a chance to do better than random, either by making educated guesses about in which sense CRBVs from the same subject are similar or by assuming that the pairs with the largest bilateral similarity are positive if they know that there are both positive and negative pairs in the test set.

Our work in this paper is an extension of our previous work reported in [1]. Contrary to [1], here we report the performances of the trained volunteers along with the untrained volunteers. We also analyse which parts of CRBVs are used by humans as well as by the deep NNs to measure the similarity in the CRBVs of the left and right retinas. We investigate different setups of a Y-shaped neural network (we name it YNN). We also analyze how deep NN-based approaches differ from the perception of human volunteers.

This paper is organized as follows. In Section II, we briefly describe the retina's anatomical structures seen in a fundus photograph, previous works on bilateral similarity in CRBVs as well as CRBVs based biometric. In Section III, we explain our approaches for manual and automatic verification. In Section IV, we describe our experimental setup. In Section V, we analyse our results. Finally, we draw our conclusions in Section VI.

II. BACKGROUND

A. BRIEF DESCRIPTION OF RETINA

The retina is a thin, semi-transparent, multi-layered neural tissue that covers two-thirds of each eye's interior. It is anatomically and physiologically considered as an extension of our brain. It is mainly responsible for converting incoming electromagnetic signals from the world outside of our eye into neural signals and then handing the neural signals to the optic nerves. The neural signals, relaying through the optic nerves, form images into the visual cortex of our brain, and therefore, we can have a sense of vision.

The approximately 0.5 mm thick retina is sandwiched between the avascular vitreous and the highly vascular choroid. It is one of the most metabolically active tissues in the human body. It consumes a high level of oxygen and nutrients to ensure our visual functionality. Two kinds of well-organized blood vessels, named CRBVs and choroidal blood vessels, are responsible for supplying oxygen and nutrients to the retina, as well as transporting away waste from the retina. Branching out from the ophthalmic blood vessels, CRBVs pierce the optic nerve and enter the internal side of the retina through the optic disc (as shown in Fig. 2).

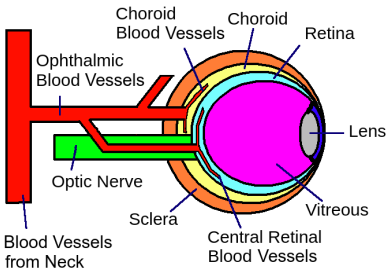


FIGURE 2. A simple schematic diagram of the human eye is drawn to understand the blood supply to-and-from the retina.

As shown in Fig. 3, in a color fundus photograph, we can see only one layered, circular, colored foreground of the multi-layered retina on a dark background. Most of the foreground of a retina is covered by CRBVs which are composed of arteries (i.e., central retinal arteries) and veins (i.e., central retinal veins). Other anatomical structures, such as the macula and the optic disc (OD), are also visible in a color fundus photograph. Depending on the fundus camera, we may see a side indicator (i.e., triangle or oval-shaped bump) always at

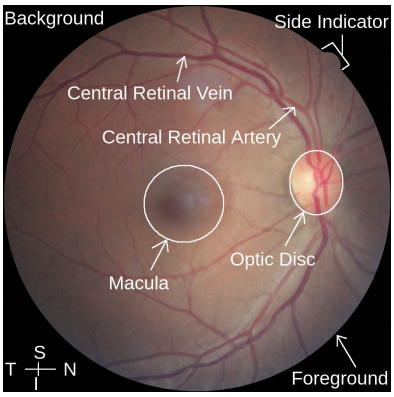


FIGURE 3. Visibility of main anatomical structures in a color fundus photograph of a healthy right retina. T: Temporal, N: Nasal, S: Superior, I: Inferior side of the retina.

the right side, which helps us to determine whether it is a left or right side retina.

Coming out from the OD on the nasal side, the CRBVs form four branches: inferior, superior, nasal, and temporal. The temporal CRBVs are close to the juncture named temple, whereas the other three types of CRBVs are close to the nose, forehead, and cheek, respectively. The first bifurcation of the CRBVs can occur inside the optic nerve, or at the head of the optic nerve, or a bit upper of the OD. The arteries are, in general, bright red, whereas veins are purple-red. Generally, the arteries have a smaller caliber than the veins inside the OD area as well as surrounding the OD area. Near to the boundary of the foreground of the retina, it is hard to distinguish veins from arteries either by color or by caliber. The small branches of arteries and veins are known as arterioles and venules. See [2]–[4] for further details about the retina and blood supply.

B. IMPORTANCE OF CRBVS

The CRBVs play an essential role in our vision. Any kind of disturbance in the blood supply to the inner layers of the retina through the CRBVs can have a negative effect on the retina and, therefore, on our vision. Severe pathology in CRBVs (e.g., diabetic retinopathy such as neovascularization) can even cause irreversible partial or complete vision loss. The CRBVs have similar anatomic, physiological, and embryological characteristics as cerebral vasculature. Many microvascular changes in the brain and other organs are reflected in the changes of the CRBVs. There are associations between different signs of the CRBVs and non-ocular diseases. For example, retinal arteriovenous nicking, focal retinal arteriolar narrowing, an increased arteriolar wall reflex, and alteration of retinal vessel fractal dimensions are associated with stroke [5], [6], whereas arteriolar narrowing and venular dilation in CRBVs are likely associated with increased risk of coronary heart disease in women [7]. Since the CRBVs can be directly and noninvasively visualized in vivo, therefore, the CRBVs draw great research interest in medical science for understanding, collecting information about, and assessing risks of many diseases at an early stage. Moreover, because of the unique and almost lifetime permanent patterns, they also draw huge research interest of biometric researchers.

C. PREVIOUS WORKS ON BILATERAL SIMILARITY IN CRBVS

The bilateral similarity in the retina helps ophthalmologists to use one retina as a proxy of the other retina, especially for post-surgery analysis and for detecting the development of pathology in the retina. For example, it is difficult to measure the vascular density of the retina suffering from epi-retinal membranes and retinal detachment. By using the measurements of bilateral similarity of macular vascular density in normal subjects as a reference value, whereas healthy retina of the patient as pre-operative retina and the retina received medical treatment as post-operative retina [8], ophthalmologists can get an idea about the effect of the medical treatment.

Using different kinds of tomography such as optical coherence tomography (OCT), Heidelberg retina tomography (HRT), Spectral domain OCT (SD-OCT), etc., bilateral symmetry between different anatomical structures and layers of the left and right retinas was reported under the term *interocular symmetry* [8]–[13]. However, we did not find any literature where symmetry in CRBVs of the left and right retina was reported using tomography images. Instead, using fundus photographs, bilateral symmetry in CRBVs of the left and right retina was reported in [14]. Measuring diameters of all retinal arterioles and venules located $0.5 - 1.0$ disc diameters from the OD margin of pairs of fundus photographs taken from 1546 subjects, substantial Pearson correlation, r, was found between the left and right eyes: for arterioles $r = 0.70$ and for venules $r = 0.77$. Moderate Pearson correlation was found for the arteriole-to-venule ratio, which was $r = 0.54$. Similar kinds of correlation were reported in [15]–[17]. Measuring the fractal dimension of the retinal vascular network as a means of quantifying the branching pattern, Taylor *et al.*, reported moderate correlation (i.e., $r \in [0.403, 0.582]$) between the left and right retina in [18]. See [19] for a review on retinal symmetry.

D. CRBVS AS BIOMETRIC

For the last 122 years, the CRBVs have been playing an important role in retina-based biometric, which is considered one of the high-performing biometrics. In 1899, Dr. Levinsohn first mentioned in one of his German articles that retinal images can be used for human identification. Along with the size and shape of the OD and its border, he mainly considered the arrangement of the CRBVs to distinguish humans from each other. In 1918, Haber showed that by placing a fine screen in front of the fundus photographic plate, CRBVs could be represented as a sketch on a grid paper. Later Dr. Blascheck wrote a formula in his unpublished work to distinguish individuals considering the external appearances of the OD and macula, visibility of the choroidal blood vessels, atrophy of the choroid, certain permanent abnormalities

(such as coloboma), the location of the first bifurcation of the arteries, and the number of the veins and the arteries intersecting two parallel lines in the superior and inferior parts and the crossing patterns of the arteries and veins through those lines from left to right. Details of these works can be found in Türkel's book [20] written in German. Furthermore, a short note in English can be found in [21]. In 1935, Dr. Simon and Dr. Goldstein claimed in [22] that even though the patterns of both arteries and veins can be jointly used for human classification, only veins are good enough for human classification since they have a more distinctive appearance than the arteries, being larger and their lumen photographing darker. After studying six pairs of monozygotic twins (i.e., single egg twins who originate from one chorion and a single placenta, and who are always of the same gender and belong to the same blood group) in 1955 [23], Dr. Tower reported that the OD shows a high degree of correlation, whereas the macula and the perimacular region exhibit a fair degree of correlation among pairs of twins. However, dissimilarity prevails in all portions of the CRBVs except those that are at or close to the OD, macula, and perimacular regions of twins. Since even for monozygotic twins, no other anatomic structures of the retina (such as OD and macula) are as unique as CRBVs, these gradually became the main research interest of retina-based biometric researchers instead of the whole fundus photographs.

In the patent US 4,109,237 [24], the patterns of CRBVs of subjects were captured by illuminating the retina with visible monochromatic green light since this light is absorbed by the dark red blood vessels, while substantially reflected by the retinal tissue, which results in high contrast between tissue and vessels. Later it was found that the brightness of the visible illuminating light needed to get a recognizable pattern caused discomfort to the subjects being identified. It also caused the eye's pupil to constrict, making it more difficult to get CRBVs pattern. Therefore, in the later patents, such as US Patent 4393366 [25], US Patent 4620318 [26], and US Patent 5532771 [27], near-infrared (NIR) light was used. The wavelength of NIR light is invisible to the human eye. Therefore, subjects do not feel any discomfort. Moreover, it has a cost-saving advantage as well. It should be noted that scanning with NIR light provides reflections from the CRBVs as well as the choroidal blood vessels. In fact, the choroidal blood vessels reflect most of the useful information needed to identify subjects. Therefore, in the NIR-based retina identification system, the contribution of CRBVs is very small. Since the choroid is not a part of the retina but is located underneath the retina, it is a bit misleading to consider the choroidal blood vessels as a retina-based biometric. However, *retina identification* is a familiar term. Therefore, the choroidal blood vessels based identification is also termed retina identification [28].

Many research works [29]–[37] extracted features such as ridge endings, ridge bifurcations, crossovers, vessels' diameter size, vessels' position and orientation, the skeleton of the whole structure of CRBVs, and so on from the CRBVs, for authentication. In the literature, few works can be found where whole retinal images were used for authentication, such as [38], [39]. However, none of these works considered the scenario where one retina is used for enrollment (registration) and the other is used for authentication. One reason could be the asymmetric look of CRBVs of left and right retinas. At a quick glance, the CRBV patterns of the left and right retinas of an individual do not appear to be more similar than those of different individuals.

III. OUR APPROACH FOR FINDING BILATERAL SIMILARITY IN CRBVS

Contrary to previous works in medical science, such as [15]–[18], [40], in this work, we turn the problem of finding similarity in CRBVs of the left and right retina into a person verification/authentication problem. We perform the person verification task on the masks of CRBVs, i.e., images showing CRBVs in white on the black background. The masks are produced by a convolutional neural network (CNN) taking color (RGB) fundus photographs as input (see Fig. 4 (a)). We choose to use the masks of CRBVs instead of the colored CRBVs segmented from RGB fundus photographs to reduce the impact of intersession variability on our experiments. For example, the light conditions on a particular day may make two images from the same session more similar so that a pair of colored CRBVs from the same subject is correctly recognized because of the light conditions rather than because of the similarity between the two retinas. Ideally, we should use only pairs of fundus photographs from different sessions. However, among the databases we have

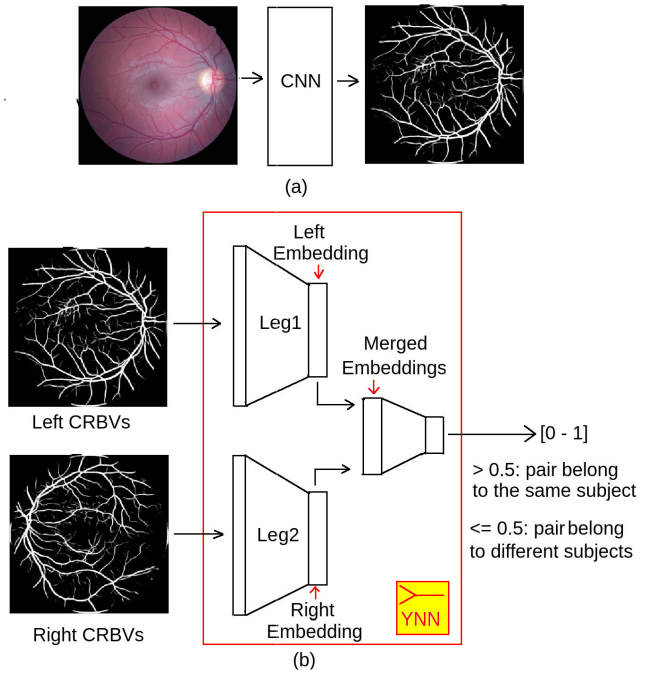


FIGURE 4. (a) Generating masks of CRBVs from color fundus photographs by a deep CNN. (b) Automatic verification of a pair of masks of CRBVs by YNN.

access to, this is only possible in one database which is too small for a detailed experimental analysis. We, therefore, use that database to confirm that for the masks of CRBVs, the intersession variability has no impact on the results, whereas we use a larger database for the main experiments. In the remainder of this article, we will refer to *masks of CRBVs* simply as *CRBVs*.

In order to take human volunteers' opinions, we use a computer interface to show four pairs of CRBVs of the left and right retina in a single frame at once. In total, we show 100 pairs in a random order, among which 50 pairs are positive (i.e., pairs are from the same subject) and 50 pairs are negative (i.e., pairs are from different subjects). Volunteers are asked to make a binary decision about pair of CRBVs. Bowyer *et al.*, [41] performed similar experiments with human subjects for pairs of irises with the aim to find texture similarity in the left and right irises (see Subsection IV-D for details).

In order to do the verification automatically, we develop a Y-shaped deep NN (we name it YNN) by connecting three NNs, as shown in Fig. 4 (b). In this architecture, each CBRVs image in the pair to be verified is sent to an NN (we refer to this NN as a *leg* of the YNN). The outputs of the legs are the compact versions of the inputs and are referred to as *embeddings*. The embeddings are *merged* and further processed by another NN that outputs an estimation of the probability that the two CRBV images belong to the same person. Similar architectures were previously proposed for networks used for fingerprint verification [42] and signature verification [43]. The network proposed for signature verification was referred to as *siamese neural network*. In [42], [43], the images that are compared are the same type of image, i.e., a fingerprint is compared with another fingerprint, or a signature is compared with another signature. In such cases, it is reasonable to share the parameters of the two legs of the NN, hence the name *siamese*. However, we are comparing different types of images, namely an image from the left retina and an image from the right retina. In this case, it is not obvious that sharing parameters is optimal. If, for example, the merging strategy is to subtract one embedding from the other and then take the absolute value, then it is desirable that a specific element in both the left and the right embedding comes from the same region of the corresponding retina, e.g., OD. Otherwise, the OD from one retina is compared with another region in the other retina. Of course, a neural network may, in principle, find ways to solve this problem. For example, if it can detect which side the image is from, it could learn side-specific filters so that some channels contain information about the left retina (while being unused for the right retina) and some channels contain information about the right retina (while being unused for the left retina). Information from the two retinas can then be compared by the NN that comes after the merging. Nevertheless, it may be easier for the YNN to compare two retinas if we either flip one of the images (e.g., the right) and share the parameters of the YNN legs or use different parameters for the two legs. We explore all these approaches as well as several different merging techniques, namely *concatenation*, *absolute subtraction*, *averaging*, or *concatenation* of *averaging* and *absolute subtraction*. We also explore a well-known pre-trained model, named VGG16 [44], as the leg of the YNN.

We also investigate whether a YNN can also be used as a *side independent* verification system, by which we mean a system that can compare two retinas no matter whether they are from the same side or two different sides. In order to achieve this, we need to use a slightly different approach during training a YNN. We need to include not only *left-right* pairs after flipping a specific side image but also *right-right* and *left-left* pairs while training the YNN.

IV. EXPERIMENTAL SETUP

A. HARDWARE & SOFTWARE TOOLS

We did all experiments using two machines with TensorFlow's Keras API 2.0.0, OpenCV 4.2.0, and Python 3.6.9. The first machine is a standard PC with 32 GB memory, Intel(R) Core(TM) i9-9900K CPU having eight cores per socket, and two NVIDIA GeForce GTX 1080 GPUs having 8 GB memory per GPU. The second machine is a standard PC with 32 GB memory, AMD Ryzen Threadripper 2950X CPU having 16 cores per socket, and one GeForce RTX 2080 Super GPU having 8 GB memory.

B. DATA SETS

In order to segment CRBVs from the color fundus photographs, we trained a U-Net with the architecture shown in Fig. 7 (a). The U-Net is well-known for its good performance in medical image segmentation tasks. Even using very few images in the training phase, the U-Net can segment images in great detail. For example, in [45], only 30 images were used to train a U-Net, which outperformed a sliding window CNN for the ISBI neuronal structures in the EM stacks challenge 2012. In order to train and test our U-Net, we used four publicly available data sets: the DRIVE (Digital Retinal Images for Vessel Extraction) [46] data set, the CHASE_DB1 [47], a subset of the CHASE (Child Heart and Health Study in England) data set, the HRF (High-Resolution Fundus) [48] data set, and the STARE (STructured Analysis of the Retina) [49] data set(see Table 1 for details).

TABLE 1. Data sets used for training and testing U-Net for segmenting CRBVs from the color fundus photographs. Note that all these data sets have manually segmented CRBVs as ground truths.

Data Set	Height × Width	# Images	Purpose
DRIVE [46]	584 × 768	40	Training set of U-Net
CHASE_DB1 [47]	960 × 999	28	Development set
HRF [48]	2336 × 3504	45	Test set
STARE [49]	605 × 700	20	

To do the verification, we used RGB retinal images of two publicly available data sets: the Kaggle data set [50] and the Longitudinal diabetic retinopathy screening data set [51].

TABLE 2. Data sets used for verification. [# Subj: Number of subjects, # Pairs: Number of pairs, PPs: Positive pairs, NPs: Negative pairs, LR-SS: pairs of left-right CRBVs from the same session, LR-DS: pairs of left-right CRBVs from different sessions, LL-DS: pairs of left-left CRBVs from different sessions, RR-DS: pairs of right-right CRBVs from different sessions].

Data Set	Height × Width	# Subjs.	# Pairs		Purpose
			PPs	NPs	
Kaggle_SetA.1	3264 × 4928	150	50	50 (Different 50 NPs randomly chosen from 2,500 NPs for different volunteers)	Test set for manual verification
Kaggle_SetA.2				50 (Same NPs for all volunteers)	Test set for manual and automatic verification
Kaggle_SetB.1	3168 × 4752	6834	6834	41,229,522	Training set of YNN
Kaggle_SetB.2		200	200	200	Validation set of YNN
Kaggle_SetC	3264 × 4928	1752	1752	1752	Test set for automatic verification
RODREP_SetA.1	1312 × 2000	69	138 (LR-SS)	138	
RODREP_SetA.2			138 (LR-DS)	138	
RODREP_SetA.3			69 (LL-DS)	69	
RODREP_SetA.4			69 (RR-DS)	69	

1) KAGGLE DATA SET

This data set is prepared for the competition of diabetic retinopathy detection. It is provided by EyePACS and publicly available via the Kaggle online community of data scientists and machine learners. It has 44, 351 pairs of images. In each pair, there is a left and right retinal image belonging to a single subject ID number. Therefore, there are in total 88, 702 RGB retinal images belonging to 44, 351 subject IDs. Images were captured under a variety of conditions. There is no information about whether images from the same subject were captured during the same session, but this is likely the case. Therefore, this database cannot be used to study the effect of session variability. There are images having 27 types of resolutions. We chose images with resolutions 3264×4928 and 3168×4752 because the foreground of these two resolutions has a complete circular shape. We prepared three sets, i.e., Kaggle_A, Kaggle_B, and Kaggle_C, from our chosen images for three purposes (see Table 2 for details).

We prepared two test sets (i.e., Kaggle_SetA.1 and Kaggle_SetA.2) using the images of Kaggle_SetA and one test set (i.e., Kaggle_SetC) using the images of Kaggle_SetC. In principle, it is possible to build 150 positive pairs (i.e., the left and right retinal images of a pair belonging to a single subject ID) and $150 \times 149 = 22,350$ negative pairs (i.e., the left and right retinal images of a pair belonged to two different subject IDs) using the 150 pairs of Kaggle_SetA. However, for human volunteers, it is difficult and time-consuming to give a decision about $150 + 22,350 = 22,500$ pairs. Therefore, we decided to reduce the number of pairs while keeping the variability among pairs as much as possible. For fulfilling that, we divided 150 subjects into three groups: the first 50 subjects were for the positive pairs (PPs), the second 50 were for the left side of negative pairs (NPs), and the third 50 were for the right side of NPs. In this way, we kept only 50 PPs and $50 \times 50 = 2,500$ NPs in Kaggle_SetA.1, and 50 PPs and 50 NPs in Kaggle_SetA.2. The PPs were the same in both test sets, whereas the NPs of Kaggle_SetA.2 were a subset of the NPs of Kaggle_SetA.1. In Kaggle_SetC, there were 1, 752 PPs and 1, 752 NPs. Even though it was possible to make $1,752 \times 1,751 = 3,067,752$ NPs from 1, 752 pairs, we chose only 1, 752 NPs in order to keep a balance between the PPs and NPs. Contrary to Kaggle_SetA.1 and Kaggle_SetA.2, there was a subject overlap between the PPs and NPs, as well as between the left and right sets of NPs in Kaggle_SetC.

2) LONGITUDINAL DIABETIC RETINOPATHY SCREENING DATA SET

This data set was prepared for fundus image registration methods. It has 1120 in total color fundus photographs of 70 patients in the diabetic retinopathy screening program of the Rotterdam Eye Hospital (Rotterdam, The Netherlands). For each patient, there are four types of color fundus photos: macula-centered, optic nerve-centered, superior, and temporal fundus images for both left and right retinas. The images were captured in two sessions and there is a 1-week period gap between the two sessions. We prepared one set named RODREP_SetA by taking only macula-centered images (i.e., two images for each retina of each subject) from this data. Since there is only one session with a macula-centered retinal image for the right side retina for the subject having Patient ID 62, we excluded images of this subject. Therefore, we had 276 images (138 images from the left side and 138 images from the right side) from 69 subjects. We were able to prepare 138 PPs of left-right CRBVs from the same session (i.e., RODREP_SetA.1), 138 PPs of left-right CRBVS from different sessions (i.e., RODREP_SetA.2), 69 PPs of left-left CRBVs from different sessions (i.e., RODREP_SetA.3) and 69 pairs of right-right CRBVs from different sessions (i.e., RODREP_SetA.4). We chose macula-centered images mainly because YNN's training set, i.e., Kaggle_SetB, has macula-centered images.

C. IMAGE PRE-PROCESSING

Since the dark pixels of the background do not provide any information about the retina, at first, the extra dark background was cropped so that the foreground touched the boundary without losing any important pixels of the foreground. For cropping the background of images of the Kaggle and RODREP data sets, we used a simple background cropping algorithm. As shown in Fig. 5, after blurring an image by a 5×5 Gaussian kernel, we detected edges using *Canny's edge detection* algorithm [52]. After that, we found the contour points belonging to each edge. Then we found the contour which had the maximum area. After that, we estimated the radius of the circle that minimally encloses that contour. Using that radius, we cropped the background. We used the functions of the OpenCV library for this part. After cropping the background, since the different data sets have different resolutions, we re-sized all images to 256×256 by bicubic interpolation. Then we re-scaled pixel values to $[0, 1]$ for each channel of each image separately. Except that, no pre-processing was applied to any images.

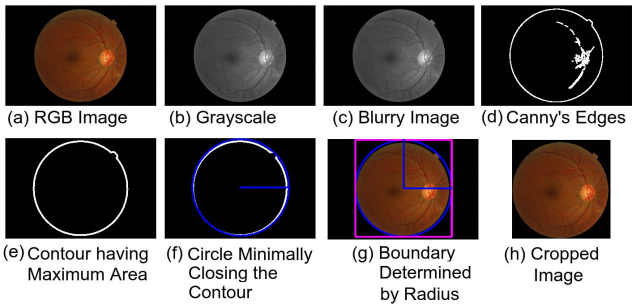


FIGURE 5. Steps for cropping background.

D. SETUP FOR MANUAL VERIFICATION

For manual verification, we asked 20 untrained volunteers who did not know where to find symmetrical properties in retinas to participate in a test. In this test, 25 frames were shown to each volunteer, where each frame contained four pairs of CRBVs side-by-side (as shown in Fig. 6). The right side CRBVs were flipped to make the comparison task easier for human volunteers. All volunteers (i.e., ID 1-20) saw the same 50 PPs but in random orders. Volunteers with ID 1-10 saw 50 different NPs which were randomly chosen from the $2,500$ NPs of Kaggle_SetA.1 so that they were not exhausted by seeing too many NPs. Volunteers with ID 11-20 saw the same 50 NPs but in random orders.

The task of the untrained human volunteers was to click on a pair when they thought that the pair belongs to a single person. Volunteers were allowed to select/deselect any pair as many times as they wanted and could spend as much time on the verification task as they wanted. However, after closing any frame, they were not allowed to see it again. After closing the last frame, each volunteer was asked to write about the factors they considered to make a decision. Twenty volunteers participated in 20 separate sessions. None of them were aware

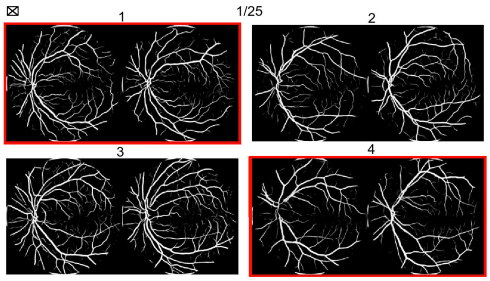


FIGURE 6. An example frame for collecting volunteers' opinions. When a volunteer clicked on any pair, its boundary turned into red color and it meant that the volunteer considered that the pair belonged to a single subject. Numbers 1, 2, 3, 4 were the pair numbers, 1/25 was the frame number, and the *cross* sign was for closing the frame.

of the true answers. All volunteers were requested not to share their assumptions with other volunteers. When writing their points, they were informed of retina-related terms to make their writing easier. After summarizing the features reported by the untrained volunteers, we gave them to four volunteers who work in retina based research field. We prepared a similar interface for them to train themselves. The only difference between this interface and the one used in the test phase was that after clicking on a NP, they saw a blue colored boundary, whereas after clicking on a PP, they saw a red colored boundary so that they could know which pairs are PPs and which pairs are NPs and analyze their decisions. They had to make the decision about a pair before they obtained the answer. They were given three weeks to train themselves. In total, 100 pairs (50 PPs and 50 NPs) were seen by the volunteers in random order during training. These pairs were taken from the Kaggle_SetB. Only when all of them felt they were ready to participate in the final test, they were shown the pairs of the Kaggle_SetA.2.

Contrary to our setup, the setup used in [41] for finding texture similarity in pairs of left and right irises had one pair of irises per frame; humans had to choose an option from a list of five degrees of certainty after seeing a pair only for three seconds instead of taking a binary decision; there were 150 PPs and 150 NPs; humans got a chance to train themselves a bit by seeing three PPs and two NPs with labels at the beginning of the test.

E. SETUP FOR DEEP NEURAL NETWORKS

We trained six kinds of deep neural networks (NNs): U-Net, YNN, Model1: NN having almost the same architecture as YNN except three layers less than YNN after concatenation of the two legs, Model2: NN having the same architecture for the legs as the YNN but then followed by a Cosine distance layer, Model3: NN using the three first layers from a pre-trained VGG16 as legs. Fig. 7 shows the model architectures of our deep NNs. We trained YNN, Model1, Model2, and Model3, in four different ways:

- without sharing/tying parameters of the legs and without flipping the right-side CRBVs.
- without sharing/tying parameters of the legs, but flipping the right-side CRBVs.

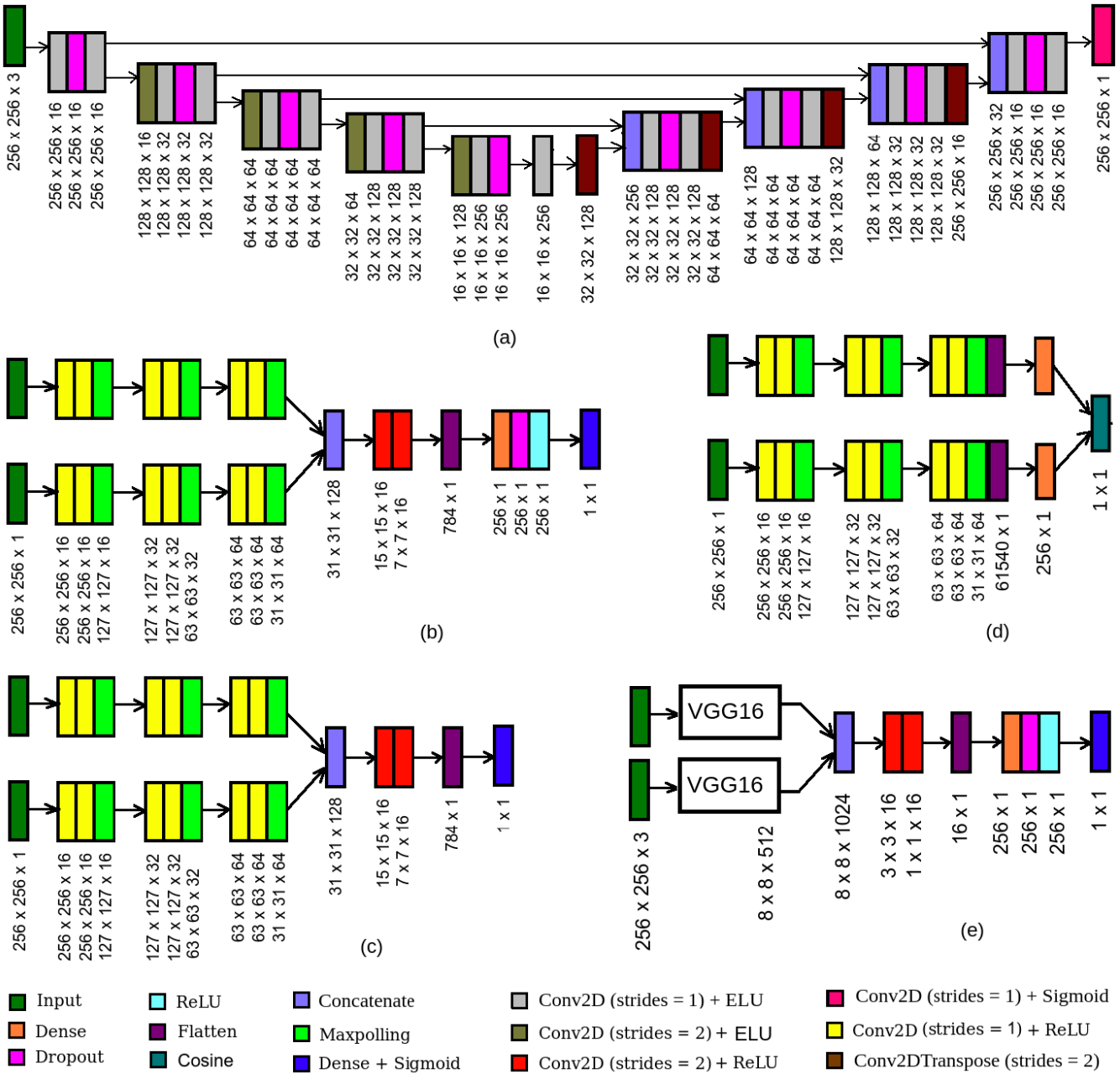


FIGURE 7. Architecture of deep neural networks: (a) U-Net, (b) YNN, (c) Model1, (d) Model2, and (e) Model3. Vertical text shows the output shape of the corresponding layer.

- by sharing/tying parameters of the legs and without flipping the right-side CRBVs.
- by sharing/tying parameters of the legs but flipping the right-side CRBVs.

To train our U-Net, we set *mean-squared-error (mse)* as the loss function; RMSProp with a learning rate of 0.001 as the optimizer and $mini\_batch\_size = 8$. To train all other deep NNs, we set *mean squared error* as the loss function; RMSProp with a learning rate of 0.0001 as the optimizer and $epoch\_no = 50$. We reduced the learning rate if there was no change in the *validation_accuracy* for more than three consecutive epochs. We stopped training if *validation_accuracy* did not change in 15 consecutive epochs. We set $batch\_size = 64$ when training a model except Model3 (i.e., the VGG16 based model). Due to memory limitations, we set $batch\_size = 32$ when training Model3. For all other settings, we used the default values of TensorFlow's Keras API 2.6.1-tf.

The VGG16 model has 13 convolutional layers, five pooling layers, and three dense layers. We formed Model3 by removing the three dense layers from the pre-trained VGG16 model provided by Keras. This model was trained

mainly for classifying images of 1000 classes. Even though the images of CRBVs are 2D, the VGG16 model requires 3D images. Therefore, we turned 2D CRBVs into 3D by repeating the 2D image three times.

For training U-Net, we used unpaired images. For training YNN, Model1, Model2, and Model3, we used left-right paired images. We also trained a variation of YNN using left-right pairs, left-left pairs, and right-right pairs. Since in the Kaggle_SetB, there is only one image per side for a subject, we used data augmentation to create extra left-left and right-right pairs. Specifically, we did this by rotating the image (minimum 5 degrees and maximum 90 degrees), shifting it at most 10% both along the horizontal and vertical direction (using spline interpolation of order three and filling all values beyond the edge with 0.0) and turning non-black pixels into the black in random order. All these three transformations (i.e., rotation, shifting, darkening) were decided automatically and randomly, i.e., for some images, any one, two, or all three transformations were applied, whereas for some images, no transformation was applied. Some training pairs are shown in Fig. 8 as examples.

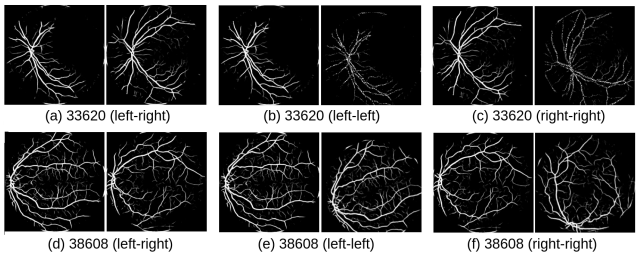


FIGURE 8. Six example pairs used for training YNN. Images of the right CRBVs in (a), (c), (d), and (f) were flipped. Image of the left CRBVs in the right side in (b) was at first shifted 19 pixels along the y-axis, 1 pixel along the x-axis, and then 11.60% of its non-black pixels were turned into black, whereas in (d) was at first rotated 20° and then shifted 20 pixels along the y-axis and 8 pixels along the x-axis. Image of the right CRBVs in the right side in (c) was at first rotated 30°, then shifted 6 pixels along the y-axis and 21 pixels along the x-axis, and then its 22.88% non-black pixels were turned into black, whereas in (f) was rotated only 76°.

For YNN, we tried four different operations to merge the embeddings of the left and right CRBVs: concatenation, average, absolute subtraction, and concatenation of absolute subtraction and average. In order to avoid the effect of randomness caused by different factors, including weight initialization and dropout, on the estimation of performance, we trained all models five times and estimated the performance. That means, in total, we trained YNN 80 (i.e., $4 \times 4 \times 5$) times, whereas we trained Model1, Model2, and Model3 five times. We trained U-Net and the variation of YNN using left-left, right-right, left-right pairs one time.

V. RESULTS & ANALYSIS

In this section, we present and analyse the results of several experiments. In Subsection V-A, we discuss the performance of the U-Net for segmenting the CRBVs from color fundus photographs. All the following human and automatic verification experiments used this segmentation. In Subsection V-B and V-C, respectively, we present the results of manual and automatic *different-side* verification. In Subsection V-D, we present the results of experiments comparing different-side verification with same-side verification as well as different-session trials. In Subsection V-E, we analyse the agreement of human and automatic verification. Finally, we present an analysis of what patterns humans and deep NNs are looking at in Subsection V-F.

A. SEGMENTING CRBVS

The performance of our U-Net trained for segmenting CRBVs from color fundus photographs is shown in Table 3. Comparing to the approaches reported in [53]–[58], the performance of our U-Net may look worse. However, in those works, specific systems for each database were developed by using training, validation/development, and test data from the same database and by optimizing the hyperparameters specifically for each database. In our experiments, the training set, development set, and test set were taken from different data sets as described in Table 1. (We have not investigated whether this was the reason for the different performance since the performance of CRBV segmentation was not the main focus of our work.) In our experiments, the hyperparameters (e.g., number of epochs and learning rate) were selected based on visual inspection of the resulting CRBV images for the development set.

TABLE 3. U-Net's performance measures for segmenting CRBVs from RGB fundus photographs of CHASE_DB1, HRF, and STARE database. The evaluation is done per pixel. [Acc: Accuracy, Sens: Sensitivity, Spec: Specificity, AUC: Area under the receiver operating characteristic (ROC) curve via a Riemann sum].

Data sets	Accuracy	Sens.	Spec.	F1	AUC
CHASE_DB1 (Dev. set)	0.8771	0.4784	0.9911	0.6231	0.9198
HRF	0.7773	0.4128	0.9924	0.5738	0.8594
STARE	0.8637	0.4299	0.9862	0.5689	0.8743

It should be noted that the data sets used in the verification experiments in the later sections, i.e., the Kaggle and the RODREP, do not have manually segmented CRBVs as ground truths, so we cannot train or evaluate CRBV segmentation models on these sets. However, visual inspection of the resulting images on these sets suggests that the performance of automatic CRBV segmentation is similar to the performance on the HRF and the STARE data sets.

As shown in Fig. 9, our U-Net fails to segment CRBV when image quality is very poor. Many factors such as experience level of the operator, operator's finger movement or shaking, different settings of fundus cameras, subject's eye movement or blinking, different amounts of light reflection by different parts of the retina because of its natural curved structure, inadequate illumination, the variation of pupil dilation, poor focus, lossy compression-decompression techniques, noisy transmission channels and so on can result in low-quality retinal images.

TABLE 4. Results of manual verification. [V. ID: Volunteer ID, Acc.: Accuracy, Sens.: Sensitivity, Spec.: Specificity].

	Kaggle_SetA.1										Kaggle_SetA.2													
	Untrained Volunteers										Untrained Volunteers										Trained Volunteers			
V. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Acc.	0.64	0.66	0.66	0.71	0.77	0.74	0.62	0.66	0.72	0.65	0.66	0.68	0.69	0.76	0.69	0.72	0.76	0.66	0.63	0.69	0.70	0.68	0.72	0.75
Sens.	0.74	0.82	0.36	0.68	0.80	0.56	0.92	0.54	0.50	0.70	0.52	0.64	0.60	0.82	0.70	0.64	0.74	0.52	0.34	0.72	0.62	0.76	0.64	0.76
Spec.	0.54	0.50	0.96	0.74	0.74	0.92	0.32	0.78	0.94	0.60	0.80	0.72	0.78	0.70	0.68	0.80	0.78	0.80	0.92	0.66	0.78	0.60	0.80	0.74
F1	0.67	0.71	0.51	0.70	0.78	0.68	0.71	0.61	0.64	0.67	0.60	0.67	0.66	0.77	0.69	0.70	0.76	0.60	0.48	0.70	0.67	0.70	0.70	0.75

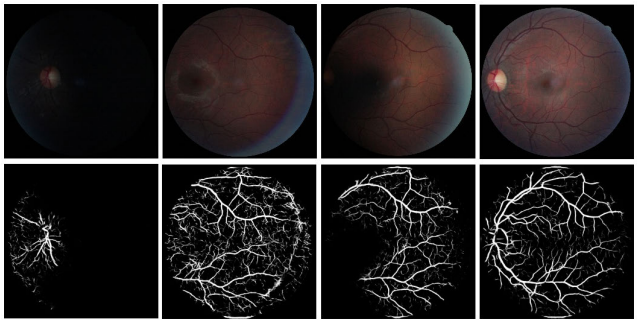


FIGURE 9. Examples of retina images of different qualities from the same subject. Our trained U-Net fails segmenting CRBVs from images of low-quality (e.g.,1st, 2nd, and 3rd column).

Dealing with low-quality images is an important issue since when the image quality is low, the image may not be correctly segmented. For example, CRBVs may not be detected in some region of the image, or vessels may be incorrectly detected. Undetected CRBVs may mislead trained volunteers to make a wrong decision about PPs. For example, untrained volunteers may consider white pixels or lines that are incorrectly segmented as vessels in both sides of a pair due to session effects (e.g., light) as a similarity and incorrectly classify an NP as a PP. Improving the performance of U-Net by adding more low-quality images and their corresponding manually segmented CRBVs in the training set or improving image quality before CRBV segmentation in the pre-processing stage, or developing a neural network, which can handle low-quality images better than the U-Net, is kept as future work.

B. MANUAL VERIFICATION

The results of human verification are shown in Table 4. As can be seen, even the lowest performance among the human volunteers was better than what would be obtained by random guesses. Therefore, we can conclude that there are some similar properties in the CRBVs of our two eyes.

The untrained volunteers reported several different factors as being important for their decision. Some of them were the thickness of the clearly visible CRBVs; the overall structure of CRBVs looking from afar; angles of vessels leaving the root (i.e., OD area), branching pattern; grouping tendency of vessels; vessel pattern close to the empty space (i.e., macula); curvature of the thickest/thicker vessel(s) considering straightness and tortuosity (or waviness) of the vessel(s); the density of vessels; alignment of the root (i.e., OD) with the empty space (i.e., macula); how vessels are spreading, and so on. Some factors (such as the density of vessels) were reported by multiple volunteers.

On average, the trained volunteers performed slightly better than the untrained volunteers, although the two best performing volunteers were untrained. One reason for the relatively small difference between the trained and untrained volunteers could be that there were four pairs of left and right retina in a frame and the volunteers knew beforehand that there would be both PPs and NPs in the test (but not necessarily in every frame), which helped them to gradually make rules about the similarity, some of which were correct. Another reason is that the information reported by the untrained volunteers influenced the trained volunteers a bit. Most probably, some features reported by the untrained volunteers (e.g., the area near to the macula) misled them. Since the untrained volunteers neither marked the area exactly where they saw the similarity nor reported their findings in detail, it was difficult for trained volunteers to get precise information. Another reason could be that the trained volunteers were not as focused as some untrained volunteers during the final test. Since they could not figure out any concrete rule to find similarity/dissimilarity in pairs of the CRBVs during their training, they may be somewhat bored and uninterested when taking the final test.

Fig. 10 shows for each retina pair the number of volunteers that thought the two retinas belong to the same subject. From these graphs, it is clear that there were some easily recognizable positive pairs in Kaggle_SetA.1 and Kaggle_SetA.2, which were recognized by all (or many) volunteers, and some difficult pairs which made almost all volunteers confused. For example, all volunteers correctly recognized eight NPs since no volunteer selected them as PPs (see Fig. 11 for some example pairs).

By visually inspecting some of the images that many volunteers failed to classify correctly, we noticed several reasons why an image pair can be challenging to classify. Different alignment can make two images of CRBVs taken from the same retina look different. Also, the quality of the color fundus photographs and the existence of pathology play important roles in the symmetrical/asymmetrical look of the pairs of CRBVs. The U-Net may end up segmenting very few or discontinuous CRBVs from low-quality fundus photographs. Some pathology (such as retinal hemorrhages) can cause some parts of the blood vessels to be invisible,

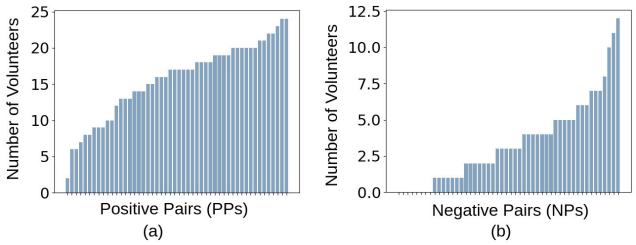


FIGURE 10. The number of pairs considered to be from the same subject for: (a) positive pairs (PPs) by 24 volunteers and (b) negative pairs (NPs) by 14 volunteers. Note that the same 50 PPs were seen by all 24 volunteers. However, only 14 (i.e., 10 untrained and 4 trained) volunteers saw the same 50 NPs. The other 10 untrained volunteers saw different NPs.

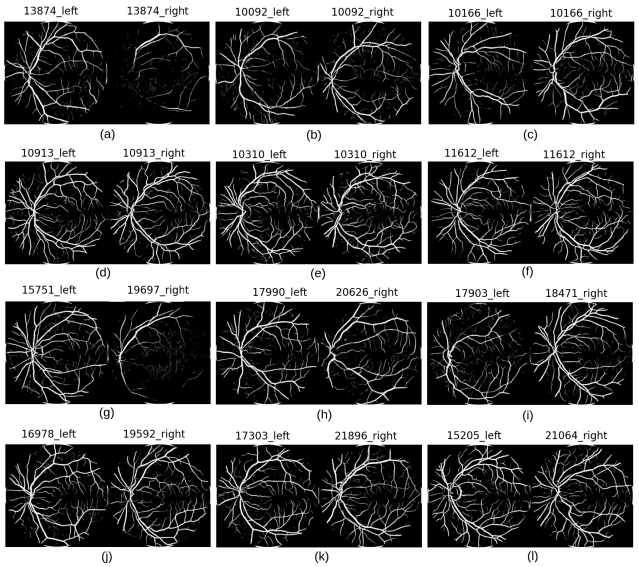


FIGURE 11. 1st row: three difficult PPs in both Kaggle_SetA.1 and in Kaggle_SetA.2. 2nd row: three easily recognized PPs in both Kaggle_SetA.1 and in Kaggle_SetA.2. 3rd Row: three comparatively easily recognized NPs out of eight NPs in Kaggle_SetA.2, which were not mistakenly selected as PPs by any of 14 volunteers. 4th Row: three difficult NPs in Kaggle_SetA.2, which were mistakenly selected as PPs. Note that, in Kaggle_SetA.1, different volunteers have different NPs and most of those NPs were also different from the NPs in Kaggle_SetA.2.

which can make a pair of CRBVs from the same person look different.

C. AUTOMATIC VERIFICATION

Table 5 shows the result for YNN for various combinations of parameter tying (tied/untied), embedding comparison (concatenation/subtraction/average) and input pre-processing (no flipping / flipping right side). From the results of the Kaggle_SetA.2, it is also clear that YNN performs substantially better than the human volunteers. The variations in the results due to initialization and dropout are relatively large, typically around 2%. The results for the Kaggle_SetA.2 and the Kaggle_SetC are almost the same which is expected since these sets have the same properties. However, the Kaggle_SetC is much larger than the Kaggle_SetA.2. Therefore, the Kaggle_SetC provides more reliable results than the Kaggle_SetA.2. The most noticeable result is that the *Average* operation performs badly when the YNN legs are tied. This is not surprising because this operation removes information about the distance between the two embeddings (i.e., the outputs of the last layer of each leg). Especially in the untied case, the YNN should, in principle, overcome this by putting information about the left side image in some elements of the embedding and information about the right side in some other elements of the embedding. The subsequent network can then compare the retina information extracted by the two legs. However, it seems the network could not successfully learn to do this. On the other hand, it seems the YNN successfully learned how to process the data for subtraction merging, even when the parameters of the legs are shared and none of the sides are flipped, which, as discussed in Section III, should be somewhat challenging for it. We have not managed to understand exactly what the YNN did in this case. Other observations are that

- In the case of untied legs, there is no difference between flipped images and unflipped images.
- *Absolute subtraction* and *concatenation of absolute subtraction* and *average* perform nearly the same, i.e., the average does not add any complementary information to the (absolute value of) subtraction.
- Concatenation is marginally the best method for combining the embeddings.

D. SIDE INDEPENDENT VERIFICATION

From the experiments in Section V-C, two important questions remain to be answered. First, how much worse is retina verification based on the left and right retina pairs compared to the verification based on the pairs of two right retinas or two left retinas? Second, how much does the performance reduce if the two images in a pair are from different sessions (e.g., the images are captured on different days)? These questions cannot be answered using the Kaggle data sets used in the previous experiments because these data sets have only one left and one right retinal image for each subject, captured at the same session. To answer these questions, we instead used the RODREP_SetA data set. This data set has two sessions per subject, and in each session, there is a left and a right retinal image. Accordingly, we can compare

- Same-session, left-right verification,
- Different-session, left-right verification.
- Different-session, left-left verification
- Different-session, right-right verification

We used the YNN, which was trained using the left-left, right-right, and left-right pairs after flipping the right side CRBVs. The results are given in Table 6. From the results, it is clear that same-side verification is substantially better than different-side verification and that verification based on images from different sessions is not worse than verification based on images from the same session. The fact that same-side verification is better than different-side verification suggests that the CRBVs of the left and the right retinas do not contain the same information, which means that using

TABLE 5. Effect of randomization introduced during training YNNs for different reasons such as random weight initialization, dropout, etc. [EER: Equal error rate, NF: NoFlip, i.e., images of right side CRBVs were not flipped. RF: RightFlip, i.e., images of right side CRBVs were flipped. Ctt, Avg, Sub, and Ctt(Sub,Avg) are the approaches applied for merging the left and right side legs of YNN. Ctt: Concatenation, Avg: Average, Sub: Absolute subtraction, Ctt(Sub,Avg): Concatenation of absolute subtraction and average].

	Kaggle_SetA.2							
	Untied							
	NF				RF			
	Ctt	Avg	Sub	Ctt(Sub,Avg)	Ctt	Avg.	Sub	Ctt(Sub,Avg)
Accuracy	0.892 ± 0.025	0.884 ± 0.024	0.874 ± 0.021	0.872 ± 0.020	0.870 ± 0.009	0.846 ± 0.034	0.844 ± 0.055	0.844 ± 0.014
Sensitivity	0.976 ± 0.008	0.960 ± 0.022	0.968 ± 0.032	0.960 ± 0.025	0.976 ± 0.015	0.968 ± 0.037	0.976 ± 0.008	0.944 ± 0.034
Specificity	0.808 ± 0.045	0.808 ± 0.047	0.780 ± 0.038	0.784 ± 0.034	0.764 ± 0.029	0.724 ± 0.081	0.712 ± 0.111	0.744 ± 0.046
F1	0.901 ± 0.021	0.892 ± 0.020	0.885 ± 0.019	0.882 ± 0.018	0.883 ± 0.006	0.863 ± 0.026	0.864 ± 0.039	0.858 ± 0.011
AUC	0.919 ± 0.015	0.930 ± 0.023	0.922 ± 0.006	0.933 ± 0.014	0.927 ± 0.019	0.916 ± 0.012	0.898 ± 0.024	0.909 ± 0.013
EER	0.132 ± 0.016	0.140 ± 0.022	0.148 ± 0.020	0.148 ± 0.030	0.156 ± 0.029	0.160 ± 0.022	0.164 ± 0.015	0.168 ± 0.016
	Tied							
	NF				RF			
	Ctt	Avg.	Sub	Ctt(Sub,Avg)	Ctt	Avg.	Sub	Ctt(Sub,Avg)
Accuracy	0.886 ± 0.026	0.738 ± 0.051	0.864 ± 0.019	0.868 ± 0.013	0.882 ± 0.032	0.822 ± 0.029	0.852 ± 0.017	0.858 ± 0.013
Sensitivity	0.952 ± 0.010	0.952 ± 0.048	0.980 ± 0.013	0.976 ± 0.015	0.976 ± 0.015	0.980 ± 0.025	0.976 ± 0.015	0.964 ± 0.034
Specificity	0.820 ± 0.049	0.524 ± 0.105	0.748 ± 0.032	0.760 ± 0.033	0.788 ± 0.055	0.664 ± 0.061	0.728 ± 0.027	0.752 ± 0.020
F1	0.894 ± 0.022	0.785 ± 0.034	0.878 ± 0.015	0.881 ± 0.010	0.893 ± 0.026	0.847 ± 0.022	0.868 ± 0.014	0.871 ± 0.014
AUC	0.938 ± 0.018	0.863 ± 0.007	0.903 ± 0.013	0.923 ± 0.012	0.936 ± 0.014	0.894 ± 0.010	0.889 ± 0.011	0.906 ± 0.003
EER	0.112 ± 0.030	0.188 ± 0.020	0.152 ± 0.016	0.152 ± 0.024	0.148 ± 0.020	0.172 ± 0.024	0.180 ± 0.013	0.168 ± 0.027

	Kaggle_SetC							
	Untied							
	NF				RF			
	Ctt	Avg.	Sub	Ctt(Sub, Avg)	Ctt	Avg.	Sub	Ctt(Sub, Avg)
Accuracy	0.881 ± 0.002	0.865 ± 0.032	0.850 ± 0.024	0.844 ± 0.021	0.850 ± 0.017	0.805 ± 0.033	0.801 ± 0.056	0.819 ± 0.017
Sensitivity	0.989 ± 0.004	0.978 ± 0.008	0.977 ± 0.017	0.978 ± 0.016	0.978 ± 0.004	0.969 ± 0.026	0.961 ± 0.024	0.959 ± 0.039
Specificity	0.772 ± 0.005	0.753 ± 0.058	0.722 ± 0.059	0.710 ± 0.040	0.723 ± 0.035	0.642 ± 0.066	0.641 ± 0.130	0.678 ± 0.022
F1	0.892 ± 0.002	0.880 ± 0.026	0.867 ± 0.017	0.863 ± 0.016	0.867 ± 0.013	0.833 ± 0.025	0.831 ± 0.036	0.841 ± 0.018
AUC	0.932 ± 0.003	0.926 ± 0.019	0.914 ± 0.012	0.912 ± 0.015	0.925 ± 0.007	0.904 ± 0.020	0.896 ± 0.026	0.900 ± 0.016
EER	0.129 ± 0.007	0.140 ± 0.024	0.158 ± 0.013	0.163 ± 0.023	0.149 ± 0.008	0.179 ± 0.025	0.182 ± 0.021	0.178 ± 0.021
	Tied							
	NF				RF			
	Ctt	Avg.	Sub	Ctt(Sub, Avg)	Ctt	Avg.	Sub	Ctt(Sub, Avg)
Accuracy	0.880 ± 0.013	0.679 ± 0.024	0.851 ± 0.016	0.857 ± 0.015	0.875 ± 0.029	0.729 ± 0.033	0.818 ± 0.019	0.820 ± 0.008
Sensitivity	0.984 ± 0.005	0.957 ± 0.057	0.988 ± 0.006	0.986 ± 0.006	0.968 ± 0.010	0.969 ± 0.025	0.982 ± 0.012	0.973 ± 0.022
Specificity	0.776 ± 0.025	0.402 ± 0.046	0.713 ± 0.035	0.728 ± 0.028	0.782 ± 0.054	0.489 ± 0.075	0.655 ± 0.051	0.667 ± 0.028
F1	0.891 ± 0.010	0.748 ± 0.024	0.869 ± 0.012	0.874 ± 0.012	0.886 ± 0.023	0.782 ± 0.019	0.844 ± 0.013	0.844 ± 0.006
AUC	0.934 ± 0.008	0.818 ± 0.033	0.919 ± 0.007	0.921 ± 0.011	0.936 ± 0.017	0.836 ± 0.021	0.909 ± 0.004	0.903 ± 0.007
EER	0.132 ± 0.012	0.257 ± 0.022	0.149 ± 0.010	0.148 ± 0.012	0.135 ± 0.025	0.246 ± 0.026	0.168 ± 0.007	0.170 ± 0.013

both of them jointly should improve the results of biometric systems further. This will be investigated in future work. Another fact is that in the case of different-side verification, the performance of YNN, trained by three kinds of pairs (i.e., left-left, left-right, and right-right), is worse than the YNN, trained by only left-right pairs (see results of Kaggle_SetC in Table 5 and Table 6). Overall, these results show that it is possible to develop CRBVs based side-independent subject authentication systems.

E. HUMAN VS YNN

To compare the agreement between human and automatic classification, we applied majority voting to the decisions of the human volunteers (i.e., 24 volunteers for PPs, 14 for NPs) for the Kaggle_SetA.2 (see Table 7 for the scores of the 12 pairs shown in Fig. 11 as examples). We then compared how well the resulting classification and the different YNNs agreed. The result is shown in Table 8. Interestingly, there is a high agreement between human and automatic classification (and a very high agreement between the different YNN architectures for automatic classification). Whenever there was a disagreement about PPs between the human

TABLE 6. Results of automatic verification when Left-Right, Left-Left, and Right-Right pairs of CRBVs were used to train a tied YNN after flipping the right side retina. [LR-SS: pairs of the left-right CRBVs from the same session, LR-DS: pairs of left-right CRBVs from different sessions, LL-DS: pairs of left-left CRBVs from different sessions, RR-DS: pairs of right-right CRBVs from different sessions].

	Kaggle_SetC	RODREP_SetA			
	LR-SS	LR-SS	LR-DS	LL-DS	RR-DS
Accuracy	0.8154	0.5797	0.6051	0.8551	0.8261
Sensitivity	0.8944	0.1812	0.2391	0.7391	0.7101
Specificity	0.7363	0.9783	0.9710	0.9710	0.9420
F1	0.8289	0.3008	0.3745	0.8361	0.8033
AUC	0.8845	0.7841	0.7702	0.9213	0.9303
EER	0.1832	0.2609	0.2971	0.1449	0.1884

TABLE 7. Scores of the six PPs and six NPs shown in Fig. 11, decided by human volunteers and YNN. The human scores were decided by averaging the decision of 24 volunteers for PPs and 14 volunteers for NPs. In these 12 pairs, there were three easily recognized PPs, three easily recognized NPs as well as three difficult PPs, and three difficult NPs. Note that scores near 1.0 are better for PPs, while scores near 0.0 are better for NPs. [Notation Vlt: Volunteers, Ctt(S, A): Concatenation of absolute subtraction and average].

	Pair ID	Vlt.	YNN_Untied_NoFlip			
			Ctt	Avg	Sub	C(S,A)
PPs	13874_left-13874_right	0.08	0.60	0.75	0.26	0.00
	10092_left-10092_right	0.25	0.99	0.97	0.98	0.92
	10166_left-10166_right	0.25	0.96	0.92	0.93	0.87
	10913_left-10913_right	0.96	0.86	0.98	0.97	0.99
	10310_left-10310_right	1.00	1.00	0.99	0.99	0.99
	11612_left-11612_right	1.00	0.96	0.98	0.93	0.94
NPs	15751_left-19697_right	0.00	0.00	0.00	0.00	0.00
	17990_left-20626_right	0.00	0.00	0.00	0.00	0.00
	17903_left-18471_right	0.00	0.00	0.00	0.00	0.02
	16978_left-19592_right	0.71	0.85	0.97	0.90	0.69
	17303_left-21896_right	0.79	0.33	0.16	0.83	0.99
	15205_left-21064_right	0.86	0.00	0.00	0.01	0.00

TABLE 8. The number of times volunteers and YNN_Untied_NoFlip agreed with each other for 50 PPs and 50 NPs of the Kaggle_SetA.2. [Notation Vlt: Volunteers].

		Vlt.	YNN_Untied_NoFlip			
			Ctt	Avg	Sub	Ctt(Sub,Avg)
PPs	Vlt	-	37	38	38	38
	Ctt	37	-	49	47	47
	Avg	38	49	-	48	48
	Sub	38	47	48	-	50
	Ctt(Sub,Avg)	38	47	48	50	-
NPs	Vlt	-	32	35	36	36
	Ctt	32	-	47	46	46
	Avg	35	47	-	43	47
	Sub	36	46	43	-	46
	Ctt(Sub,Avg)	36	46	47	46	-

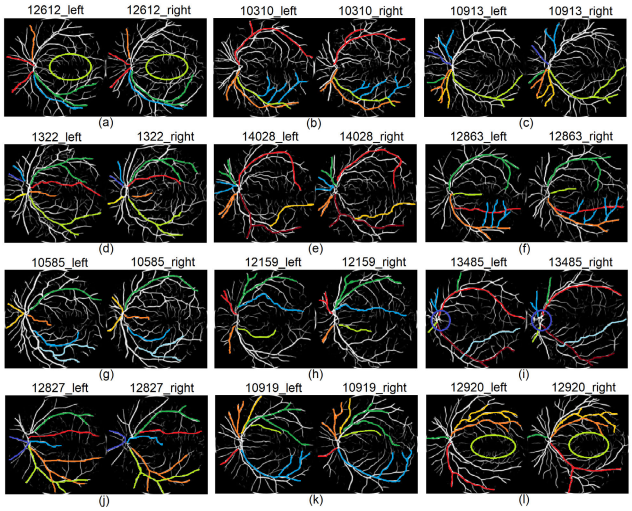


FIGURE 12. Outlined similarities in 12 positive pairs (PPs) out of 50 PPs in Kaggle_SetA.1 and Kaggle_SetA.2. The total number of volunteers selected those pairs [i.e., in (a), (b), (c),.... (l)] are 24, 24, 23, 22, 22, 21, 21, 20, 20, 20, 20, 20, respectively. Note that the vessels that have the same color were classified as PPs by volunteers.

majority vote and the YNNs, the YNNs were correct most of the time, while for NPs, human volunteers were correct most

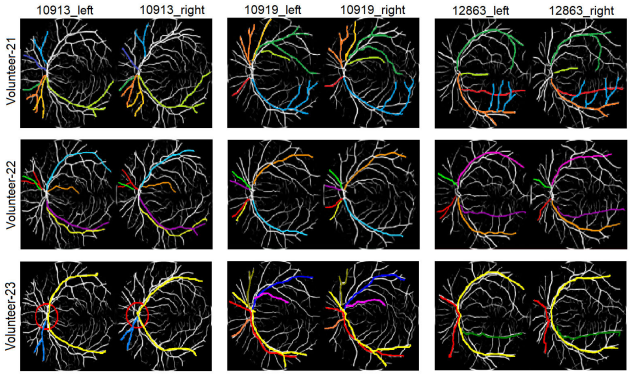


FIGURE 13. Similarities in three PPs marked by three trained volunteers. Vessels marked with the same color in the left and right retina were considered similar by the volunteer. It seems perception about similarity between CRBVs of the left and right retinas is quite subjective. However, all of us can notice some similarities if we look carefully.

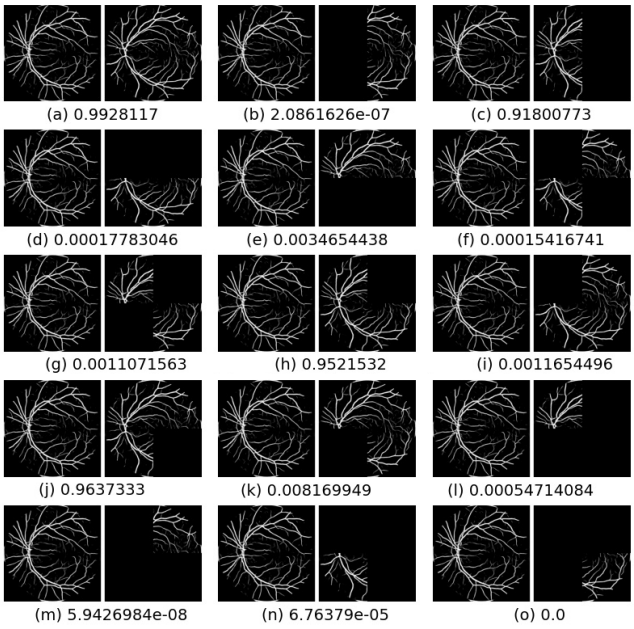


FIGURE 14. Occlusion sensitivity of a PP of Kaggle_SetA.2 having ID 10919. The caption of each subplot indicates the YNN output of the pair (i.e., the probability that the pair belong to a *single subject*). The YNN output dropped drastically when the OD side part was blocked by dark squares. For example, in (b) OD side of the right CRBVs was blocked by two 64 × 64 black squares and YNN output dropped from 0.992817 to 2.0861626e − 07. Note that the right retina was flipped and scores were generated by a YNN whose legs were tied.

of the time, as shown in Table 9. For example, the human majority vote and the YNN whose embeddings were merged by *concatenation*, disagreed about 18 out of the 50 NPs. For these 18 pairs, the human vote was correct for 12 pairs, and the YNN was correct for 6 pairs.

F. WHAT HUMAN VOLUNTEERS AND NEURAL NETWORK LOOK AT FOR COMPARING CRBVS

In Subsection V-B, we have already mentioned several aspects of the CRBVs that the volunteers reported as crucial

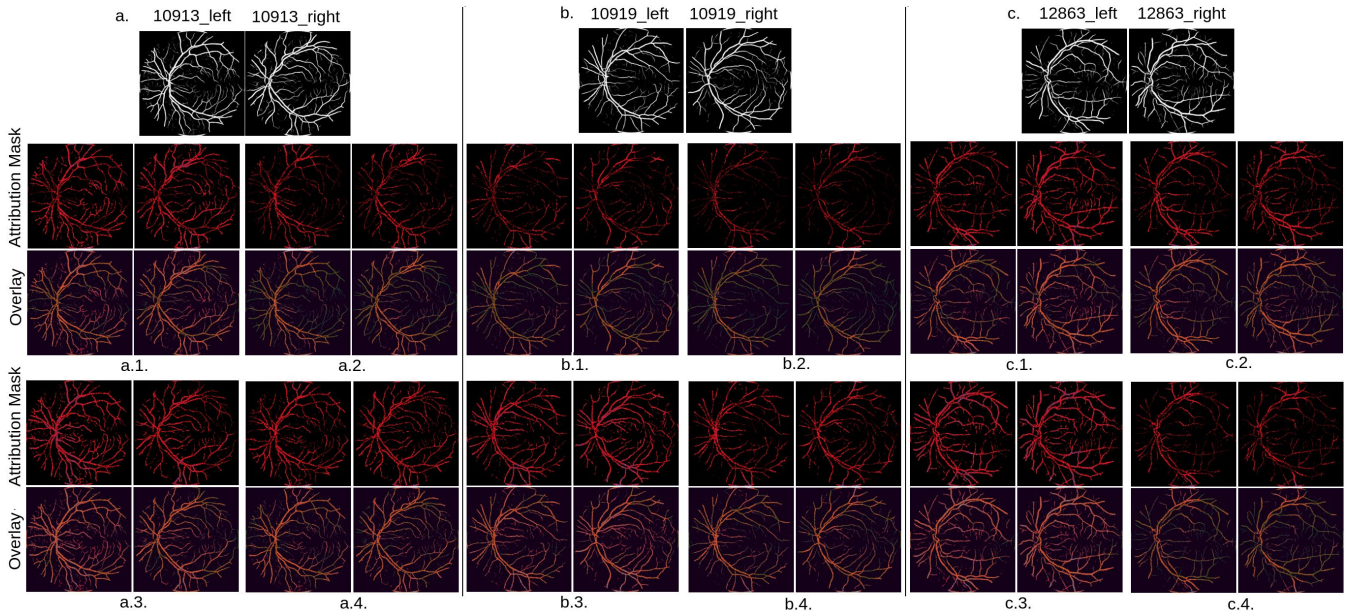


FIGURE 15. **Visualization of attributions for three PPs of CRBVs of the left and right retinas when right retinas were flipped and the YNN had tied legs. Fig. a.1., b.1., & c.1., for *concatenation*; Fig. a.2., b.2. & c.2., for average; Fig. a.3., b.3., & c.3. for absolute subtraction and Fig. a.4., b.4., & c.4., for *concatenation* of *absolute subtraction* and *average*.**

TABLE 9. **The number of times volunteers and YNN_Untied_NoFlip were correct when disagreeing with each other about 50 PPs and 50 NPs of the Kaggle_SetA.2. A cell value x/y means that the system in the row and column header disagreed y times and that the system in the row header was correct x times.**

		Vlt.	YNN_Untied_NoFlip			
			Ctt	Avg	Sub	Ctt(Sub,Avg)
PPs	Vlt	-	1/13	0/12	1/12	1/12
	Ctt	12/13	-	0/1	2/3	2/3
	Avg	12/12	1/1	-	2/2	2/2
	Sub	11/12	1/3	0/2	-	0/0
	Ctt(Sub,Avg)	11/12	1/3	0/2	0/0	-
NPs	Vlt	-	12/18	9/15	10/14	9/14
	Ctt	6/18	-	0/3	2/4	1/4
	Avg	6/15	3/3	-	5/7	2/3
	Sub	4/14	2/4	2/7	-	1/4
	Ctt(Sub,Avg)	5/14	3/4	1/3	3/4	-

for human verification. In this section, we analyse this in more detail and further what features in the CRBVs are important for automatic verification with YNNs.

To better understand what humans are looking at, we asked three of the four trained volunteers to draw the similarities they could find in the PPs that were correctly verified by the majority of the volunteers. The similarities found by one volunteer in twelve of these PPs are shown in Fig. 12. From the highlighted branches and places, we can see that humans can detect similarities in the CRBVs of the left and right retina if they look carefully. The similarities found by three volunteers in three of the PPs are shown in Fig. 13. It can be seen that the perceived similarities in the CRBVs of the left and right retina can be quite subjective, i.e., different volunteers found similarities in different places. However, the similarities in some branches of CRBVs were seen by everyone.

To analyse what features in the CRBVs are important for the decision of the YNN and to answer questions such as ''which parts of CRBVs contribute significantly to the output of the YNN'', ''Does our YNNs find similarity in the same places of CRBVs as human volunteers do'', and ''Do the different YNN architectures look at different things?'' we applied *attribution methods*. The task of an attribution method is to estimate how much each input feature contributes to the decisions of the network [59]. First, we checked the occlusion sensitivity following a simple perturbation-based attribution method applied in [60]. We systematically occluded, i.e., masked different portions of the input image with 64×64 black squares and monitored the output of the YNN. The result is shown in Fig. 14. It can be seen that the YNN output probability dropped significantly when the OD side of the retina was occluded by black squares. This means that the output of the YNN is mostly affected by this side. The macula side, on the other hand, does not contribute significantly to the YNN output. One reason for this could be that the CRBVs on the OD side are thick and easily noticeable, whereas the CRBVs on the macula side are tiny and clearly not visible, which makes our U_Net less accurate at detecting them. Another possible reason is that the OD region of the retinas exhibits larger bilateral symmetry than the macula region.

To analyse what the YNNs look at more systematically than the simple occlusion sensitivity-based approach, we used the integrated gradients method [61] that generates a heatmap representing how important a region of the input image is for the decision. According to this method, for each pair of CRBVs, we generated a path of 64 image pairs, where each image pair was interpolated between a pair of black baseline images and the original pair of CRBVs.

TABLE 10. Performance of varieties of YNN. [NF: NoFlip, i.e., images of right side CRBVs were not flipped. RF: RightFlip, i.e., images of right side CRBVs were flipped. Model1: NN having almost the same architecture as YNN except three layers less than YNN after concatenation of two legs, Model2: NN having the same architecture as YNN and a Cosine distance layer, Model3: NN using top thrown pre-trained VGG16 as legs].

	Kaggle_SetA.2							
	Untied							
	NF				RF			
	YNN	Model1	Model2	Model3	YNN	Model1	Model2	Model3
Accuracy	0.892 ± 0.025	0.858 ± 0.026	0.608 ± 0.063	0.660 ± 0.063	0.870 ± 0.009	0.828 ± 0.027	0.564 ± 0.054	0.672 ± 0.041
Sensitivity	0.976 ± 0.008	0.944 ± 0.020	0.956 ± 0.039	0.728 ± 0.086	0.976 ± 0.015	0.948 ± 0.016	0.932 ± 0.047	0.756 ± 0.062
Specificity	0.808 ± 0.045	0.772 ± 0.048	0.260 ± 0.126	0.592 ± 0.078	0.764 ± 0.029	0.708 ± 0.065	0.196 ± 0.105	0.588 ± 0.047
F1	0.901 ± 0.021	0.870 ± 0.022	0.710 ± 0.034	0.681 ± 0.061	0.883 ± 0.006	0.847 ± 0.020	0.682 ± 0.032	0.697 ± 0.043
AUC	0.919 ± 0.015	0.910 ± 0.027	0.528 ± 0.056	0.703 ± 0.055	0.927 ± 0.019	0.887 ± 0.027	0.498 ± 0.004	0.727 ± 0.039
EER	0.132 ± 0.016	0.148 ± 0.030	0.260 ± 0.070	0.344 ± 0.057	0.156 ± 0.029	0.180 ± 0.033	0.252 ± 0.053	0.344 ± 0.048
	Tied							
	NF				RF			
	YNN	Model1	Model2	Model3	YNN	Model1	Model2	Model3
Accuracy	0.886 ± 0.026	0.864 ± 0.029	0.662 ± 0.071	0.742 ± 0.020	0.882 ± 0.032	0.862 ± 0.012	0.652 ± 0.025	0.678 ± 0.028
Sensitivity	0.952 ± 0.010	0.932 ± 0.037	0.956 ± 0.060	0.892 ± 0.043	0.976 ± 0.015	0.940 ± 0.044	0.832 ± 0.127	0.952 ± 0.020
Specificity	0.820 ± 0.049	0.796 ± 0.051	0.368 ± 0.122	0.592 ± 0.063	0.788 ± 0.055	0.784 ± 0.037	0.472 ± 0.165	0.404 ± 0.056
F1	0.894 ± 0.022	0.873 ± 0.027	0.740 ± 0.051	0.776 ± 0.016	0.893 ± 0.026	0.872 ± 0.013	0.703 ± 0.024	0.748 ± 0.018
AUC	0.938 ± 0.018	0.934 ± 0.016	0.508 ± 0.010	0.795 ± 0.029	0.936 ± 0.014	0.913 ± 0.026	0.500 ± 0.000	0.761 ± 0.022
EER	0.112 ± 0.030	0.132 ± 0.037	0.240 ± 0.066	0.280 ± 0.031	0.148 ± 0.020	0.164 ± 0.020	0.276 ± 0.056	0.320 ± 0.068

	Kaggle_SetC							
	Untied							
	NF				RF			
	YNN	Model1	Model2	Model3	YNN	Model1	Model2	Model3
Accuracy	0.881 ± 0.002	0.850 ± 0.034	0.622 ± 0.048	0.678 ± 0.019	0.850 ± 0.017	0.812 ± 0.025	0.600 ± 0.047	0.684 ± 0.019
Sensitivity	0.989 ± 0.004	0.960 ± 0.019	0.969 ± 0.049	0.692 ± 0.046	0.978 ± 0.004	0.958 ± 0.029	0.957 ± 0.039	0.694 ± 0.067
Specificity	0.772 ± 0.005	0.740 ± 0.069	0.275 ± 0.092	0.664 ± 0.051	0.723 ± 0.035	0.665 ± 0.070	0.244 ± 0.083	0.674 ± 0.051
F1	0.892 ± 0.002	0.866 ± 0.025	0.720 ± 0.031	0.682 ± 0.021	0.867 ± 0.013	0.836 ± 0.016	0.706 ± 0.030	0.686 ± 0.033
AUC	0.932 ± 0.003	0.913 ± 0.020	0.526 ± 0.051	0.742 ± 0.020	0.925 ± 0.007	0.893 ± 0.013	0.509 ± 0.018	0.753 ± 0.017
EER	0.129 ± 0.007	0.163 ± 0.024	0.298 ± 0.063	0.321 ± 0.018	0.149 ± 0.008	0.185 ± 0.015	0.299 ± 0.029	0.313 ± 0.020
	Tied							
	NF				RF			
	YNN	Model1	Model2	Model3	YNN	Model1	Model2	Model3
Accuracy	0.880 ± 0.013	0.864 ± 0.022	0.663 ± 0.040	0.730 ± 0.007	0.875 ± 0.029	0.826 ± 0.022	0.669 ± 0.017	0.711 ± 0.021
Sensitivity	0.984 ± 0.005	0.963 ± 0.027	0.947 ± 0.096	0.843 ± 0.037	0.968 ± 0.010	0.924 ± 0.024	0.838 ± 0.128	0.907 ± 0.023
Specificity	0.776 ± 0.025	0.765 ± 0.046	0.379 ± 0.087	0.616 ± 0.034	0.782 ± 0.054	0.729 ± 0.042	0.500 ± 0.150	0.516 ± 0.062
F1	0.891 ± 0.010	0.876 ± 0.019	0.736 ± 0.041	0.757 ± 0.011	0.886 ± 0.023	0.842 ± 0.019	0.714 ± 0.026	0.759 ± 0.010
AUC	0.934 ± 0.008	0.926 ± 0.014	0.512 ± 0.017	0.796 ± 0.014	0.936 ± 0.017	0.895 ± 0.014	0.505 ± 0.006	0.792 ± 0.017
EER	0.132 ± 0.012	0.146 ± 0.019	0.257 ± 0.062	0.269 ± 0.010	0.135 ± 0.025	0.189 ± 0.020	0.277 ± 0.053	0.277 ± 0.022

We calculated the gradient of the YNN output with respect to each pixel of the pairs of CRBVs, summed the gradients over each interpolated image, and took the absolute value. As shown in Fig. 15, different merging approaches emphasized on slightly different places to find similarities. From Fig. 13 and Fig. 15, we can see that thick blood vessels contribute more to the decisions of both human volunteers and YNNs than the thin blood vessels.

G. VARIETIES OF YNN

To understand the importance of the YNN architecture design, we explored several varieties of YNN. As shown in Table 10, these varieties did not perform better than YNN (i.e., the model used in the previous experiments). Model1 performed slightly worse than YNN. However, it has 54.82% and 68.23% less trainable parameters for untied and tied cases, respectively, comparing to YNN. Even though cosine distance based models performed worse than the other models, they still performed better than random guesses. Transfer learning by a model pre-trained on RGB color images is not very beneficial.

VI. CONCLUSION

In this paper, we have shown that there is a high degree of bilateral similarity in central retinal blood vessels (CRBVs) and, therefore, it is possible to do *side-independent* retina verification. We conducted experiments where both human volunteers and computers by means of neural networks (NN) had to decide whether CRBV images from a left and a right retina are from the same subject or not. For human volunteers, the accuracy in this task was typically in the range of 65%-70%, which is well above the result of random guesses. For most NN architectures, the accuracy was 85%-88%. We further trained a side-independent retina verification system, i.e., a system that can do the verification either by comparing left and right retinas or by comparing two retinas from the same side. In experiments with that system, *same-side* verification had 23% better accuracy (relatively) than *different-side* verification. Finally, we visualized and analysed which features in the CRBVs are useful for verification. Future work shall include improved automatic CRBV segmentation and explore systems where both left and right retina are used jointly for verification.

REFERENCES

[1] S. Biswas, J. Rohdin, and M. Drahansky, "Bilateral symmetry in central retinal blood vessels," in *Proc. 8th Int. Workshop Biometrics Forensics (IWBF)*, Apr. 2020, pp. 1–6.

[2] H. Kolb, "Simple Anatomy of the Retina," in *The Organization of the Retina and Visual System*, H. K. H, E. Fernandez, and R. Nelson, Eds. Salt Lake City, UT, USA: Webvision, 1995.

[3] S. C. Nemeth, C. Shea, M. DiSclafani, and M. Schluter, "The Posterior Segment," in *Ocular Anatomy and Physiology*, 2nd ed. Thorofare, NJ, USA: Slack Incorporated, 2008, ch. 9, pp. 88–99.

[4] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, Dec. 2010.

[5] M. L. Baker, P. J. Hand, J. J. Wang, and T. Y. Wong, "Retinal signs and stroke: Revisiting the link between the eye and brain," *Stroke*, vol. 39, no. 4, pp. 1371–1379, 2008.

[6] F. N. Doubal, T. J. MacGillivray, N. Patton, B. Dhillon, M. S. Dennis, and J. M. Wardlaw, "Fractal analysis of retinal vessels suggests that a distinct vasculopathy causes lacunar stroke," *Neurology*, vol. 74, no. 14, pp. 1102–1107, Apr. 2010.

[7] G. Liew and J. J. Wang, "Retinal vascular signs: A window to the heart?" *Revista Española de Cardiología (English Edition)*, vol. 64, no. 6, pp. 515–521, Jun. 2011.

[8] G. Liu, K. Keyal, and F. Wang, "Interocular symmetry of vascular density and association with central macular thickness of healthy adults by optical coherence tomography angiography," *Sci. Rep.*, vol. 7, no. 1, p. 16297, Dec. 2017.

[9] D. L. Budnez, "Symmetry between the right and left eyes of the normal retinal nerve fiber layer measured with optical coherence tomography (an AOS thesis)," *Trans. Amer. Ophthalmol. Soc.*, vol. 106, pp. 252–275, 2008.

[10] H. Li, P. R. Healey, Y. M. Tariq, E. Teber, and P. Mitchell, "Symmetry of optic nerve head parameters measured by the Heidelberg retina tomograph 3 in healthy eyes: The Blue Mountains Eye study," *Amer. J. Ophthalmol.*, vol. 155, no. 3, pp. 518–523, 2013.

[11] M. Yang, W. Wang, Q. Xu, S. Tan, and S. Wei, "Interocular symmetry of the peripapillary choroidal thickness and retinal nerve fibre layer thickness in healthy adults with isometropia," *BMC Ophthalmol.*, vol. 16, no. 1, p. 182, Dec. 2016.

[12] M. Zhou, B. Lu, J. Zhao, Q. Wang, P. Zhang, and X. Sun, "Interocular symmetry of macular ganglion cell complex thickness in young chinese subjects," *PLoS ONE*, vol. 11, no. 7, Jul. 2016, Art. no. e0159583.

[13] R. R. Mastey, M. Gaffney, K. M. Litts, C. S. Langlo, E. J. Patterson, M. R. Strampe, A. Kalitzeos, M. Michaelides, and J. Carroll, "Assessing the interocular symmetry of foveal outer nuclear layer thickness in achromatopsia," *Transl. Vis. Sci. Technol.*, vol. 8, no. 5, p. 21, Sep. 2019.

[14] H. Leung, J. J. Wang, E. Rochtchina, A. G. Tan, T. Y. Wong, L. D. Hubbard, R. Klein, and P. Mitchell, "Computer-assisted retinal vessel measurement in an older population: Correlation between right and left eyes," *Clin. Exp. Ophthalmol.*, vol. 31, no. 4, pp. 326–330, Aug. 2003.

[15] D. J. Couper, R. Klein, L. D. Hubbard, T. Y. Wong, P. D. Sorlie, L. S. Cooper, R. J. Brothers, and F. J. Nieto, "Reliability of retinal photography in the assessment of retinal microvascular characteristics: The atherosclerosis risk in communities study," *Amer. J. Ophthalmol.*, vol. 133, no. 1, pp. 78–88, Jan. 2002.

[16] T. Y. Wong, R. Klein, F. J. Nieto, B. E. K. Klein, A. R. Sharrett, S. M. Meuer, L. D. Hubbard, and J. M. Tielsch, "Retinal microvascular abnormalities and 10-year cardiovascular mortality: A population-based case-control study," *Ophthalmology*, vol. 110, no. 5, pp. 933–940, 2003.

[17] T. Y. Wong, M. D. Knudtson, R. Klein, B. E. K. Klein, S. M. Meuer, and L. D. Hubbard, "Computer-assisted measurement of retinal vessel diameters in the Beaver Dam Eye Study: Methodology, correlation between eyes, and effect of refractive errors," *Ophthalmology*, vol. 111, no. 6, pp. 1183–1190, 2004.

[18] A. M. Taylor, T. J. MacGillivray, R. D. Henderson, L. Ilzina, B. Dhillon, J. M. Starr, and I. J. Deary, "Retinal vascular fractal dimension, childhood IQ, and cognitive ability in old age: The lothian birth cohort study 1936," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0121119.

[19] J. R. Cameron, R. D. Megaw, A. J. Tatham, S. McGrory, T. J. MacGillivray, F. N. Doubal, J. M. Wardlaw, E. Trucco, S. Chandran, and B. Dhillon, "Lateral thinking–interocular symmetry and asymmetry in neurovascular patterning, in health and disease," *Prog. Retinal Eye Res.*, vol. 59, pp. 131–157, Jul. 2017.

[20] S. Türkel, *Das Auge als Identifizierungsgrundlage: Unter Berücksichtigung von Blascheks Photofundoskopie*. Vienna, Austria: GRAZ, 1927, ch. III, pp. 22–41.

[21] M. E. O'Neill, "A new method of identification," *J. Criminal Law Criminol.*, vol. 26, no. 4, pp. 608–610, 1935.

[22] C. Simon and I. Golstein, "A new scientific method of identification," *New York State J. Med.*, vol. 35, no. 18, pp. 901–906, 1935.

[23] P. Tower, "The fundus oculi in monozygotic twins: Report of six pairs of identical twins," *AMA Arch. Ophthalmol.*, vol. 54, no. 2, pp. 225–239, Aug. 1955.

[24] R. B. Hill, "Apparatus and method for identifying individuals through their retinal vasculature patterns," U.S. Patent 4 109 237, Aug. 22, 1978.

[25] R. B. Hill, "Rotating beam ocular identification apparatus and method," U.S. Patent 4 393 366, Jul. 12, 1983.

[26] R. B. Hill, "Fovea-centered eye fundus scanner," U.S. Patent 4 620 318, Oct. 28, 1986.

[27] J. C. Johnson and R. B. Hill, "Eye fundus optical scanner system and method," U.S. Patent 5 532 771, Jul. 2, 1996.

[28] R. B. Hill, *Retina Identification*. New York, NY, USA: Springer, 1996, pp. 123–141.

[29] Z.-W. Xu, X.-X. Guo, X.-Y. Hu, and X. Cheng, "The blood vessel recognition of ocular fundus," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2005, pp. 4493–4498.

[30] C. Mariño, M. G. Penedo, M. Penas, M. J. Carreira, and F. Gonzalez, "Personal authentication using digital retinal images," *Pattern Anal. Appl.*, vol. 9, no. 1, pp. 21–33, May 2006.

[31] A. Arakala, J. S. Culpepper, J. Jeffers, A. Turpin, S. Boztaş, K. J. Horadam, and A. M. McKendrick, "Entropy of the retina template," in *Advances in Biometrics*, M. Tistarelli and M. S. Nixon, Eds. Berlin, Germany: Springer, 2009, pp. 1250–1259.

[32] M. Ortega, M. G. Penedo, J. Rouco, N. Barreira, and M. J. Carreira, "Personal verification based on extraction and characterisation of retinal feature points," *J. Vis. Lang. Comput.*, vol. 20, no. 2, pp. 80–90, Apr. 2009.

[33] H. Farzin, H. Abrishami-Moghaddam, and M.-S. Moin, "A novel retinal identification system," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, Dec. 2008, Art. no. 280635.

[34] M. Agopov, "Retinal identification," in *IntechOpen Biometrics*, J. Yang, Ed. London, U.K.: InTech, 2011, ch. 5.

[35] W. Barkhoda, F. Akhlaqian, M. D. Amiri, and M. S. Nouroozzadeh, "Retina identification based on the pattern of blood vessels using fuzzy logic," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, p. 113, Dec. 2011.

[36] S. M. Lajevardi, A. Arakala, S. A. Davis, and K. J. Horadam, "Retina verification system based on biometric graph matching," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3625–3635, Sep. 2013.

[37] M. Frucci, D. Riccio, G. Sanniti di Baja, and L. Serino, "Using direction and score information for retina based person verification," *Expert Syst. Appl.*, vol. 94, pp. 1–10, Mar. 2018.

[38] U. Suripon, "Retina recognition using compression-based joint transform correlator," *Opt. Eng.*, vol. 50, no. 9, Sep. 2011, Art. no. 098201.

[39] M. Sabaghi, S. R. Hadianamrei, M. Fattahi, M. R. Kouchaki, and A. Zahedi, "Retinal identification system based on the combination of Fourier and wavelet transform," *J. Signal Inf. Process.*, vol. 03, no. 01, pp. 35–38, 2012.

[40] A. Caramoy, K. Droege, B. Kirchhof, and S. Fauser, "Retinal layers measurements in healthy eyes and in eyes receiving silicone oil-based endotamponade," *Acta Ophthalmol.*, vol. 92, no. 4, pp. e292–e297, 2014.

[41] K. W. Bowyer, S. Lagree, and S. Fenker, "Human versus biometric detection of texture similarity in left and right irises," in *Proc. 44th Annu. IEEE Int. Carnahan Conf. Secur. Technol.*, Oct. 2010, pp. 323–329.

[42] P. Baldi and Y. Chauvin, "Neural networks for fingerprint recognition," *Neural Comput.*, vol. 5, no. 3, pp. 402–418, May 1993.

[43] J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Sackinger, and R. Shah, "Signature verification using a 'siamese' time delay neural network," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 7, no. 4, p. 25, 1993.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[46] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004. [Online]. Available: https://www.isi.uu.nl/Research/Databases/DRIVE/download.php

[47] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009. [Online]. Available: https://staffnet.kingston.ac.uk/~ku15565/CHASE_DB1/assets/CHASEDB1.zip
[48] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *Int. J. Biomed. Imag.*, vol. 2013, Dec. 2013, Art. no. 154860. [Online]. Available: https://www5.cs.fau.de/fileadmin/research/datasets/fundus-images/all.zip
[49] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000. [Online]. Available: http://cecas.clemson.edu/~ahoover/stare/images/all-images.zip
[50] J. Cuadros and G. Bresnick, "EyePACS: An adaptable telemedicine system for diabetic retinopathy screening," *J. Diabetes Sci. Technol.*, vol. 3, no. 3, pp. 509–516, May 2009.
[51] K. M. Adal, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, "Accuracy assessment of Intra- and intervisit fundus image registration for diabetic retinopathy screening," *Investigative Ophthalmol. Vis. Sci.*, vol. 56, no. 3, pp. 1805–1812, Mar. 2015.
[52] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[53] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. V. Gool, "Deep retinal image understanding," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 140–148.
[54] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018.
[55] Y. Zhang and A. C. S. Chung, "Deep supervision with additional labels for retinal vessel segmentation task," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 83–91.
[56] T. Laibacher, T. Weyde, and S. Jalali, "M2U-net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 115–124.
[57] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101556.
[58] N. Tamim, M. Elshrkawey, G. Abdel Azim, and H. Nassar, "Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks," *Symmetry*, vol. 12, no. 6, p. 894, Jun. 2020.
[59] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for Deep Neural Networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–16.
[60] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.
[61] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 3319–3328.

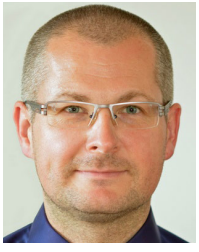
SANGEETA BISWAS received the master's degree in computer science and engineering from the University of Rajshahi (RU), Bangladesh (BD), in 2007, and the master's degree in computer science and the Ph.D. degree from the Tokyo Institute of Technology (TokyoTech), Japan, in 2011 and 2016, respectively. From 2018 to 2020, she worked as a Postdoctoral Researcher with the Faculty of Information Technology (FIT), Brno University of Technology (BUT), Czech Republic (CZ). She is currently working as an Assistant Professor with the Faculty of Engineering, RU, BD. She is also collaborating with the STRaDe Team of FIT, BUT, as a Remote Researcher. Her current research interests include biometrics, medical image processing, and crowd counting.

JOHAN ROHDIN received the M.Sc. degree in engineering physics and mathematics from the Chalmers University of Technology, in 2008, and the Ph.D. degree in computer science from the Tokyo Institute of Technology, in 2015. In 2016, he joined the Faculty of Information Technology, Brno University of Technology, Czech Republic, as a Postdoctoral Researcher, sponsored by the South Moravian Programme for Distinguished Researchers (SoMoPro) for working on neural network approaches to speaker recognition. He currently works with the Faculty of Information Technology, Brno University of Technology, as an Assistant Professor. His primary research interests include statistical pattern recognition and speech processing.

ANDRII KAVETSKYI received the M.Sc. degree in applied mathematics from the Ivan Franko National University of Lviv and in mathematical engineering from the University of L'Aquila. He is currently pursuing the Ph.D. degree in computer science and engineering with the Faculty of Information Technology, Brno University of Technology, Czech Republic. His research interests include biometrics, medical image processing, and mathematical modeling.

GABRIEL SARAIVA is currently pursuing the bachelor's degree in computer science with the Federal University of Amazonas, Brazil. Since January 2020, he has been collaborating with the STRaDe Team of the Faculty of Information Technology, Brno University of Technology, Czech Republic, as a Remote Researcher. His research interests include biometric, image processing, and machine learning.

ANGKAN BISWAS received the MBS degree in accounting from the National University, Bangladesh, in 2011. Since January 2020, he has been collaborating with the STRaDe Team of the Faculty of Information Technology, Brno University of Technology, Czech Republic, as a Remote Researcher. His research interests include image processing and artificial intelligence.



MARTIN DRAHANSKY (Senior Member, IEEE) graduated from the Faculty of Electrotechnics and Computer Science, Brno University of Technology (BUT), Czech Republic, and from the Faculty of Electrotechnics, FernUniversität in Hagen, Germany, in 2001. He received the Ph.D. degree from the Faculty of Information Technology, Brno University of Technology, in 2005. In 2009, he became an Associate Professor with the Faculty of Information Technology, Brno University of Technology, where he has been working as a Full Professor, since 2017. His research interests include biometrics, security and cryptography, artificial intelligence, and sensory systems.