

Chapter 13

Deep Sclera Segmentation and Recognition



Peter Rot, Matej Vitek, Klemen Grm, Žiga Emeršič, Peter Peer and Vitomir Štruc

Abstract In this chapter, we address the problem of biometric identity recognition from the vasculature of the human sclera. Specifically, we focus on the challenging task of multi-view sclera recognition, where the visible part of the sclera vasculature changes from image to image due to varying gaze (or view) directions. We propose a complete solution for this task built around Convolutional Neural Networks (CNNs) and make several contributions that result in state-of-the-art recognition performance, i.e.: (i) we develop a cascaded CNN assembly that is able to robustly segment the sclera vasculature from the input images regardless of gaze direction, and (ii) we present ScleraNET, a CNN model trained in a multi-task manner (combining losses pertaining to identity and view-direction recognition) that allows for the extraction of discriminative vasculature descriptors that can be used for identity inference. To evaluate the proposed contributions, we also introduce a new dataset of ocular images, called the *Sclera Blood Vessels, Periocular and Iris* (SBVPI) dataset, which represents one of the few publicly available datasets suitable for research in multi-view sclera segmentation and recognition. The datasets come with a rich

P. Rot and M. Vitek are first authors with equal contributions.

P. Rot · M. Vitek (✉) · Ž. Emeršič · P. Peer

Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113,
1000 Ljubljana, Slovenia

e-mail: matej.vitek@fri.uni-lj.si

P. Rot

e-mail: peter.rot93@gmail.com

Ž. Emeršič

e-mail: ziga.emersic@fri.uni-lj.si

P. Peer

e-mail: peter.peer@fri.uni-lj.si

K. Grm · V. Štruc

Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25,
1000 Ljubljana, Slovenia

e-mail: klemen.grm@fe.uni-lj.si

V. Štruc

e-mail: vitomir.struc@fe.uni-lj.si

© The Author(s) 2020

A. Uhl et al. (eds.), *Handbook of Vascular Biometrics*, Advances in Computer Vision and Pattern Recognition, https://doi.org/10.1007/978-3-030-27731-4_13

set of annotations, such as a per-pixel markup of various eye parts (including the sclera vasculature), identity, gaze-direction and gender labels. We conduct rigorous experiments on SBVPI with competing techniques from the literature and show that the combination of the proposed segmentation and descriptor-computation models results in highly competitive recognition performance.

Keywords Ocular biometrics · Vascular biometrics · Deep learning · Sclera segmentation · Sclera recognition · Dataset · Eye recognition

13.1 Introduction

With the growing need for secure authentication systems, forensic applications and surveillance software, biometric recognition techniques are attracting interest from research groups and private companies trying to improve the current state of the technology and exploit its immense market potential. Among the existing biometric characteristics used in automated recognition systems, ocular traits offer a number of advantages over other modalities such as contactless data acquisition, high recognition accuracy and considerable user acceptance. While iris recognition is the predominant technology in this area, recent research [1, 2] is looking increasingly at additional ocular characteristics that can complement iris-based features and contribute towards more secure and less-spoofable authentication schemes within this branch of biometrics [3].

One trait that presents itself as a particularly viable option in this context is the vasculature of the sclera. The eye's sclera region contains a rich vascular structure that is considered unique for each individual, is relatively stable over time [4] and can hence be exploited for recognition and authentication purposes, as also evidenced by recent research efforts [1, 5]. As suggested in [6], the vascular patterns also exhibit other desirable properties that make them appealing for recognition systems, e.g. the patterns are discernible despite potential eye redness and also in the presence of contact lenses that may adversely affect iris recognition systems. Despite the potential of the sclera vasculature for biometric recognition, research on this particular trait is still in its infancy and several research problems need to be addressed before the technology can be deployed in commercial systems, e.g.:

- The sclera vasculature contains distinct, but also finer blood vessels that need to be segmented from the input ocular images to ensure competitive recognition performance. As emphasised in the introductory chapter of the handbook, these vessels feature very different border types and have a complex texture that is difficult to model, which makes vasculature segmentation highly challenging. To approach this problem, existing solutions typically adopt a two-stage procedure, where the sclera region is first identified in the ocular images and the vasculature structure is then extracted using established (typically unsupervised) algorithms based, for example, on Gabor filters, wavelets, gradient operators and alike [1, 7–9]. While these approaches have shown promise, recent research suggests that supervised techniques result in much better segmentation performance [5, 10], especially if challenging off-angle ocular images need to be segmented reliably. However,

next to the difficulty of sclera vasculature segmentation task itself, the lack of dedicated and suitably annotated datasets for developing supervised techniques has so far represented one of the major roadblocks in the design of competitive sclera recognition systems.

- Due to the particularities (and potentially unconstrained nature) of the image acquisition procedure, ocular images are in general not aligned well with respect to a reference position. Additionally, as the gaze direction may vary from image to image, not all parts of the sclera vasculature are necessarily visible in every captured image. To efficiently compare sclera images and facilitate recognition, discriminative features need to be extracted from the segmented vasculature. These features have to be robust with respect to variations in position, scale and rotation and need to allow for comparisons with only parts of the located vascular structure. Existing solutions, therefore, commonly rely on hand-crafted image descriptors, such as Scale-Invariant Feature Transforms (SIFTs), Histograms of Oriented Gradients (HOGs), Local Binary Patterns (LBPs) and related descriptors from the literature [5, 8, 9]. These local descriptor-based approaches have dominated the field for some time, but, as indicated by recent trends in biometrics [11–14], are typically inferior to learned image descriptors based, for example, on Convolutional Neural Networks (CNNs).

In this chapter, we try to address some of the challenges outlined above and present a novel solution to the problem of sclera recognition built around deep learning and Convolutional Neural Networks (CNNs). Specifically, we first present a new technique for segmentation of the vascular structure of the sclera based on a cascaded SegNet [15] assembly. The proposed technique follows the established two-stage approach to sclera vasculature segmentation and first segments the sclera region from the input images using a discriminatively trained SegNet model and then applies a second SegNet to extract the final vascular structure. As we show in the experimental section, the technique allows for accurate segmentation of the sclera vasculature from the input images even under different gaze directions, thus facilitating feature extraction and sclera comparisons in the later stages.

Next, we present a deep-learning-based model, called ScleraNET, that is able to extract discriminative image descriptors from the segmented sclera vasculature. To ensure that a single (learned) image descriptor is extracted for every input image regardless of the gaze direction and amount of visible sclera vasculature, we train ScleraNET within a multi-task learning framework, where view-direction recognition is treated as a side task for identity recognition. Finally, we incorporate the segmentation and descriptor-computation approaches into a coherent sclera recognition pipeline.

To evaluate the proposed segmentation and descriptor-computation approaches, we also introduce a novel dataset of ocular images, called *Sclera Blood Vessels, Periocular and Iris (SBVPI)* and make it publicly available to the research community. The dataset represents one of the few existing datasets suitable for research in (multi-view) sclera segmentation and recognition problems and ships with a rich set of annotations, such as a pixel-level markup of different eye parts (including the sclera vasculature) or identity, gaze-direction and gender labels. Using the SBVPI dataset, we evaluate the proposed segmentation and descriptor-computation techniques in

rigorous experiments with competing state-of-the-art models from the literature. Our experimental results show that the cascaded SegNet assembly achieves competitive segmentation performance and that the ScleraNET model generates image descriptors that yield state-of-the-art recognition results.

In summary, we make the following contributions in this chapter:

- We propose a novel model for sclera vasculature segmentation based on a cascaded SegNet assembly. To the best of our knowledge, the model represents the first attempt to perform sclera vasculature segmentation in a supervised manner and is shown to perform well compared to competing solutions from the literature.
- We present ScleraNET, a CNN-based model able to extract descriptive image representations from ocular images with different gaze directions. Different from existing techniques, the model allows for the description of the vascular structure of the sclera using a single high-dimensional image descriptor even if the characteristics (position, scale, translation, visibility, etc.) of the vascular patterns vary from image to image.
- We introduce the *Sclera Blood Vessels, Periocular and Iris* (SBVPI) dataset—a dataset of ocular images with a distinct focus on research into sclera recognition. We make the dataset publicly available: <http://sclera.fri.uni-lj.si/>.

The rest of the chapter is structured as follows: In Sect. 13.2, we survey the relevant literature and discuss competing methods. In Sect. 13.3, we introduce our sclera recognition pipeline and elaborate on the segmentation procedure and ScleraNET models. We describe the novel dataset and its characteristics in Sect. 13.4. All parts of our pipeline are evaluated and discussed in rigorous experiments in Sect. 13.5. The chapter concludes with a brief summary and directions for future work in Sect. 13.6.

13.2 Related Work

In this section, we survey the existing research work relevant to the proposed segmentation and descriptor-computation approaches. The goal of this section is to provide the necessary context for our contributions and motivate our work. The reader is referred to some of the existing surveys on ocular biometrics for a more complete coverage of the field [8, 16–18].

13.2.1 Ocular Biometrics

Research in ocular biometrics dates back to the pioneering work of Daugman [19–21], who was the first to show that the texture of the human iris can be used for identity recognition. Daugman developed an iris recognition system that used Gabor filters to encode the iris texture and to construct a discriminative template that could be used for recognition. Following the success of Daugman’s work, many other hand-crafted feature descriptors were proposed [22–25] to encode the texture of the iris.

With recent research on iris recognition moving towards unconstrained image acquisition settings and away from the Near-Infrared (NIR) spectrum towards visible light (VIS) imaging, more powerful image features are needed that can better model the complex non-linear deformations of the iris typically seen under non-ideal lightning conditions and with off-angle ocular images. Researchers are, therefore, actively trying to solve the problem of iris recognition using deep learning methods, most notably, with Convolution Neural Networks (CNNs). The main advantage of using CNNs for representing the iris texture (compared to the more traditional hand-crafted image descriptors) is that features can be learned automatically from training data typically resulting in much better recognition performance for difficult input samples. Several CNN-based approaches have been described in the literature over the last few years with highly promising results, e.g. [26–30].

Despite the progress in this area and the introduction of powerful (learned) image descriptors, there are still many open research question related mostly to unconstrained image acquisition conditions (e.g. the person is not looking straight into the camera, eyelashes cover the iris, reflections appear in the images, etc.). To improve robustness of ocular biometric systems in such settings, additional ocular traits can be integrated into the recognition process, such as the sclera vasculature [1] or information from the periocular region [31, 32]. These additional modalities have received significant attention from the research community and are at the core of many ongoing research projects—see, for example, [1, 16, 33–40].

The work presented in this chapter adds to the research outlined above and introduces a complete solution to the problem of multi-view sclera recognition with distinct contributions for vasculature segmentation and descriptor computation from the segmented vascular structure.

13.2.2 Sclera Recognition

Recognition systems based on the vasculature of the sclera typically consist of multiple stages, which in the broadest sense can be categorised into a (i) a vasculature segmentation stage that extracts the vascular structure of the sclera from the image, and (ii) a recognition stage, where the vascular structure is represented using suitable image descriptors and the descriptors are then used for comparisons and subsequent identity inference.

The first stage (aimed at vasculature segmentation) is commonly subdivided into two separate steps, where the first step locates the sclera in the image and the second extracts the vasculature needed for recognition. To promote the development of automated segmentation techniques for sclera segmentation (the first step), several competitions were organised in the scope of major biometric conferences [5, 10, 41, 42]. The results of these competitions suggest that supervised segmentation techniques, based on CNN-based models represent the state of the art in this area and significantly outperform competing unsupervised techniques. Particularly successful here are Convolutional Encoder–Decoder (CED) networks (such as Seg-

Net [15]), which represent the winning techniques from the 2017 and 2018 sclera segmentation competitions—see [5, 10] for details. In this chapter, we build on these results and incorporate multiple CED models into a cascaded assembly that is shown in the experimental section to achieve competitive performance for both sclera and vasculature segmentation.

To extract the vascular structure from the segmented sclera region, image operators capable of emphasising gradients and contrast changes are typically used. Solutions to this problem, therefore, include standard techniques based, for example, on Gabor filters, wavelets, maximum curvature, gradient operators (e.g. Sobel) and others [1, 7–9]. As suggested in the sclera recognition survey in [8], a common aspect of these techniques is that they are unsupervised and heuristic in nature. In contrast to the outlined techniques, our approach uses (typically better performing) supervised segmentation models, which are possible due to the manual markup of the sclera vasculature that comes with the SBVPI dataset (introduced later in this chapter) and, to the best of our knowledge, is not available with any of the existing datasets of ocular images.

For the recognition stage, existing techniques usually use a combination of image enhancement (e.g. histogram equalisation, Contrast-Limited Adaptive Histogram Equalization (CLAHE) or Gabor filtering [1, 43]) and feature extraction techniques, with a distinct preference towards local image descriptors, e.g. SIFT, LBP, HOG, Gray-level Co-occurrence Matrices, wavelet features or other hand-crafted representations [6, 8, 44–46]. Both dense and sparse (keypoint) image descriptors have already been considered in the literature. With ScleraNET, we introduce a model for the computation of the first learned image descriptor for sclera recognition. We also make the model publicly available to facilitate reproducibility and provide the community with a strong baseline for future research in this area.

13.2.3 Existing Datasets

A variety of datasets is currently available for research in ocular biometrics [16] with the majority of existing datasets clearly focusing on the most dominant of the ocular modalities—the iris [5, 9, 47, 48, 48–55]. While these datasets are sometimes used for research into sclera recognition as well, a major problem with the listed datasets is that they are commonly captured in the Near-Infrared (NIR) spectrum, where most of the discriminative information contained in the sclera vasculature is not easily discernible. Furthermore, existing datasets are not captured with research on vascular biometrics in mind and, therefore, often contain images of insufficient resolution or images, where the Region-Of-Interest (ROI) needed for sclera recognition purposes is not well visible. While some datasets with characteristics suitable for sclera recognition research have been introduced recently (e.g. MASD [5]), these are, to the best of our knowledge, not publicly available.

Table 13.1 shows a summary of some of the most popular datasets of ocular images and also lists the main characteristics of the SBVPI dataset introduced in this

Table 13.1 Comparison of the main characteristics of existing datasets for ocular biometrics. Note that most of the datasets have been captured with research in iris recognition in mind, but have also been used for experiments with periocular (PO) and sclera recognition techniques. The dataset introduced in Sect. 13.4 of this chapter is the first publicly available dataset dedicated to sclera recognition research

| Dataset | Modality | Public | NIR/VIS | Image size | # Subjects | # Images | SC-M [‡] | VS-M* | Gaze |
|--------------------|--------------------------------|--------|---------|-------------|------------|----------|-------------------|---------|----------|
| CASIA Iris v1 [47] | Iris | Yes | NIR | 320 × 280 | 54 | 756 | No | No | Static |
| CASIA Iris v2 [47] | Iris | Yes | NIR | 640 × 480 | 60 | 2400 | No | No | Static |
| CASIA Iris v3 [47] | Iris | Yes | NIR | 640 × 480 | > 700 | 22034 | No | No | Static |
| CASIA Iris v4 [47] | Iris | Yes | NIR | 640 × 480 | > 2800 | 54601 | No | No | Static |
| ND-IRIS-0405 [49] | Iris | Yes | NIR | 640 × 480 | 356 | 64980 | No | No | Static |
| UTIRIS [50] | Sclera, iris | Yes | Both | 2048 × 1360 | 79 | 1540 | No | No | Static |
| UBIRIS v1 [48] | Sclera, iris | Yes | VIS | 800 × 600 | 241 | 1877 | No | No | Static |
| UBIRIS v2 [52] | Sclera, PO [†] , iris | Yes | VIS | 400 × 300 | 261 | 11102 | No | No | Variable |
| IITD [51] | Iris | Yes | NIR | 320 × 240 | 224 | 1120 | No | No | Static |
| MICHE-I [53] | Sclera, PO, iris | Yes | VIS | 2048 × 1536 | 92 | 3732 | No | No | Static |
| UBIPr [54] | PO | Yes | VIS | 500 × 400 | 261 | 10950 | No | No | Variable |
| IMP [55] | PO | Yes | Both | 260 × 270 | 62 | 930 | No | No | Static |
| IUPUI [9] | Sclera, PO, iris | No | Both | n/a | 44 | 352 | No | No | Variable |
| MASD [5] | Sclera | No | VIS | 7500 × 5000 | 82 | 2624 | Partial | No | Variable |
| SBVPI (ours) | Sclera, PO, iris | Yes | VIS | 3000 × 1700 | 55 | 1858 | Full | Partial | Variable |

[†]PO—periocular, [‡]SC-M—sclera markup, *VS-M—vasculature markup

chapter. While researchers commonly resort to the UBIRISv1 [48], UBIRISv2 [52], UTIRIS [56], or MICHE-I [53] datasets when conducting experiments on sclera recognition, their utility is limited, as virtually no sclera-specific metadata (e.g. sclera markup, vasculature markup, etc.) is available with any of these datasets. SBVPI tries to address this gap and comes with a rich set of annotations that allow for the development of competitive segmentation and descriptor-computation models.

13.3 Methods

In this section, we present our approach to sclera recognition. We start with a high-level overview of our pipeline and then describe all of the individual components.

13.3.1 Overview

A high-level overview of the sclera recognition pipeline proposed in this chapter is presented in Fig. 13.1. The pipeline consists of two main parts: (i) a cascaded SegNet assembly used for Region-Of-Interest (ROI) extraction and (ii) a CNN model (called ScleraNET) for image-representation (or descriptor) computation.

The cascaded SegNet assembly takes an eye image as input and generates a probability map of the vascular structure of the sclera using a two-step segmentation procedure. This two-step procedure first segments the sclera from the input image and then identifies the blood vessels within the sclera region using a second segmentation step.

The CNN model of the second part of the pipeline, ScleraNET, takes a probability map describing the vascular patterns of the sclera as input and produces a discriminative representation that can be used for matching purposes. We describe both parts of our pipeline in detail in the next sections.

13.3.2 Region-Of-Interest (ROI) Extraction

One of the key steps of every biometric system is the extraction of the Region-Of-Interest (ROI). For sclera-based recognition systems, this step amounts to segmenting the vascular structure from the input image. This structure is highly discriminative for every individual and can, hence, be exploited for recognition. As indicated in the previous section, we find the vasculature of the sclera in our approach using a two-step procedure built around a cascaded SegNet assembly. In the remainder of this section, we first describe the main idea behind the two-step segmentation procedure, then briefly review the main characteristics of the SegNet model and finally describe

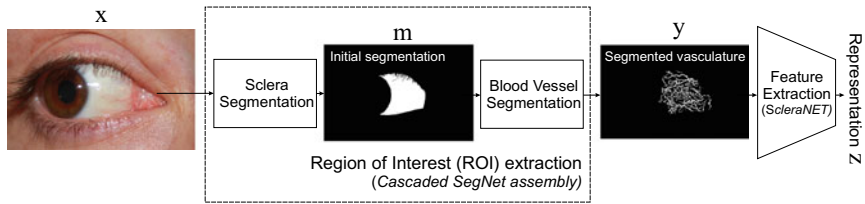


Fig. 13.1 Block diagram of the proposed sclera recognition approach. The vascular structure of the sclera is first segmented from the input image x using a two-step procedure. A probability map of the vascular structure y is then fed to a CNN model (called ScleraNET) to extract a discriminative feature representation that can be used for sclera comparisons and ultimately recognition. Note that m denotes the intermediate sclera region (or masks) generated by the first segmentation step and z represent the learned vasculature descriptor extracted by ScleraNET

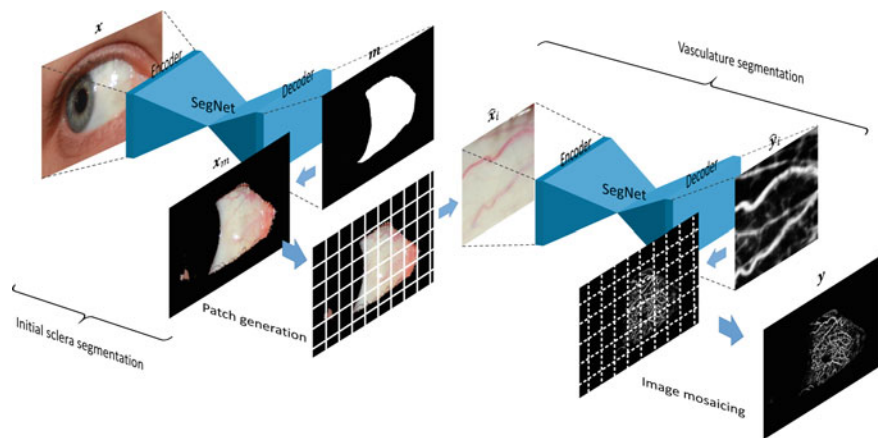


Fig. 13.2 Illustration of the two-step segmentation procedure. In the initial segmentation step, a binary mask of the sclera region is generated by a SegNet model. The mask is used to conceal irrelevant parts of the input image for the second step of the segmentation procedure, where the goal is to identify the vascular structure of the sclera by a second SegNet model. To be able to capture fine details in the vascular structure the second step is implemented in a patch-wise manner followed by image mosaicing. Please refer to the text for an explanation of the symbols used in the image

the training procedure used to learn the parameters of the cascaded segmentation assembly.

13.3.2.1 The Two-Step Segmentation Procedure

The cascaded SegNet assembly used for ROI extraction in our pipeline is illustrated in Fig. 13.2. It consists of two CNN-based segmentation models, where the first tries to generate a binary mask of the sclera region from the input image and the second aims to extract the vascular structure from within the located sclera. The segmenta-

tion models for both steps are based on the recently introduced SegNet architecture from [15]. SegNet was chosen as the backbone model for our segmentation assembly, because of its state-of-the-art performance for various segmentation tasks, competitive results achieved in the recent sclera segmentation competitions [5, 10] and the fact that an open- source implementation is publicly available.¹

Note that our two-step procedure follows existing unsupervised approaches to sclera vasculature segmentation, where an initial sclera segmentation stage is used to simplify the segmentation problem and constrain the segmentation space for the second step, during which the vasculature is extracted. Our segmentation procedure is motivated by the fact that CNN-based processing does not scale well with image size. Thus, to be able to process high-resolution input images, we initially locate the sclera region from down-sampled images in the first segmentation step and then process image patches at the original resolution in the second segmentation step with the goal of capturing the fine-grained information on the vascular structure of the sclera. Note that this information would otherwise get lost if the images were down-sampled to a size manageable for CNN-based segmentation.

If we denote the input RGB ocular image as \mathbf{x} and the binary mask of the sclera region generated by the first SegNet model as \mathbf{m} , then the first (initial) segmentation step can formally be described as follows:

$$\mathbf{m} = f_{\theta_1}(\mathbf{x}), \quad (13.1)$$

where f_{θ_1} denotes the mapping from the input \mathbf{x} to the segmentation result \mathbf{m} by the first CNN model and θ_1 stands for the model parameters that need to be learned during training.

Once the sclera is segmented, we mask the input image \mathbf{x} with the generated segmentation output \mathbf{m} and, hence, exclude all image pixels that do not belong to the sclera from further processing, i.e.:

$$\mathbf{x}_m = \mathbf{x} \odot \mathbf{m}, \quad (13.2)$$

where \odot denotes the Hadamard product. The masked input image \mathbf{x}_m is then used as the basis for the second segmentation step.

Because the vasculature of the sclera comprises large, but also smaller (finer) blood vessels, we use a patch-wise approach in the second segmentation step. This patch-wise approach allows us to also locate large blood vessels within the sclera region, but also the finer ones that would get lost (or overseen) within a holistic segmentation approach due to poor contrast and small spatial area these vessels occupy. Towards this end, we split the masked input image \mathbf{x}_m into M non-overlapping patches $\{\hat{\mathbf{x}}_i\}_{i=1}^M$ and subject them to a second segmentation model f_{θ_2} that locates the vascular structure $\hat{\mathbf{y}}_i$ within each patch:

$$\hat{\mathbf{y}}_i = f_{\theta_2}(\hat{\mathbf{x}}_i), \quad \text{for } i = 1, \dots, M. \quad (13.3)$$

¹SegNet on GitHub: <https://github.com/alexgkendall/caffe-segnet>.

Here, θ_2 denotes the model parameters of the second SegNet model that again need to be learned on some training data.

The final map of the vascular structure \mathbf{y} is generated by re-assembling all generated patches $\hat{\mathbf{y}}_i$ using image mosaicing. Note that different from the first segmentation step, where a binary segmentation mask \mathbf{m} is generated by the segmentation model, \mathbf{y} represents a probability map, which was found to be better suited for recognition purposes than a binary mask of the vasculature (details on possible segmentation outputs are given in Sects. 13.3.2.2 and 13.3.2.3).

To ensure robust segmentation results when looking for the vascular structure of the sclera in the second segmentation step, we use a data augmentation procedure at run-time. Thus, the masked image \mathbf{x}_m is randomly rotated, cropped and shifted to produce multiple versions of the masked sclera. Here, the run-time augmentation procedure selects all image operations with a probability of 0.5 and uses rotations in the range of $\pm 8^\circ$, crops that reduce the image size by up to 1% of the spatial dimensions, and shifts up to ± 20 pixels in the horizontal and up to ± 10 pixels in the vertical direction. Each of the generated images is then split into M patches which are fed independently to the segmentation procedure. The output patches $\hat{\mathbf{y}}_i$ are then reassembled and all generated maps of the vascular structure are averaged to produce the final segmentation result.

As indicated above, the basis for the ROI extraction procedure is the SegNet architecture, which is used in the first, but also the second segmentation step. We, therefore, briefly describe the main SegNet characteristics in the next section.

13.3.2.2 The SegNet Architecture

SegNet [15] represents a recent convolutional encoder–decoder architecture proposed specifically for the task of semantic image segmentation. The architecture consists of two high-level building blocks: an encoder and a decoder. The goal of the encoder is to compress the semantic content of the input and generate a descriptive representation that is fed to the decoder to produce a segmentation output [57, 58].

SegNet’s encoder is inspired by the VGG-16 [59] architecture, but unlike VGG-16, the encoder uses only convolutional and no fully connected layers. The encoder consists of 13 convolutional layers (followed by batch normalisation and ReLU activations) and 5 pooling layers. The decoder is another (inverted) VGG-16 model again without fully connected layers, but with a pixel-wise softmax layer at the top. The softmax layer generates a probability distribution for each image location that can be used to classify pixels into one of the predefined semantic target classes. During training, the encoder learns to produce low-resolution semantically meaningful feature maps, whereas the decoder learns filters capable of generating high-resolution segmentation maps from the low-resolution feature maps produced by the encoder [57].

A unique aspect of SegNet are so-called skip-connections that connect the pooling layers of the encoder with the corresponding up-sampling layers of the decoder. These skip-connections propagate spatial information (pooling indices) from one part of the model to the other and help avoid information loss throughout the network.

Consequently, SegNet’s output probability maps have the same dimensions (i.e. width and height) as the input images, which allows for relatively precise segmentation. The number of output probability maps is typically equal to the number of semantic target classes—one probability map per semantic class [57]. The reader is referred to [15] for more information on the SegNet model.

13.3.2.3 Model Training and Output Generation

To train the two SegNet models, f_{θ_1} and f_{θ_2} , and learn the model parameters θ_1 and θ_2 needed by our segmentation procedure, we use categorical cross-entropy as our training objective. Once the models are trained, they return a probability distribution over the $C = 2$ target classes (i.e. *sclera vs. non-sclera* for the first SegNet and *blood vessels vs. other* for the second SegNet in the cascaded assembly) for each pixel location. This is, for every location $s = [x, y]^T$ in the input image, the model outputs a distribution $\mathbf{p}_s = [p_{sC_1}, p_{sC_2}]^T \in \mathbb{R}^{C \times 1}$, where p_{sC_i} denotes the probability that the pixel at location s belongs to the i th target class C_i and $\sum_{i=1}^C p_{sC_i} = 1$ [57]. In other words, for each input image the model returns two probably maps, which, however are only inverted versions of each other, because $p_{sC_1} = 1 - p_{sC_2}$.

When binary segmentation results are needed, such as in the case of our sclera region \mathbf{m} , the generated probability maps are thresholded by comparing them to a predefined segmentation threshold Δ .

13.3.3 ScleraNET for Recognition

For the second part of our pipeline, we rely on a CNN model (called ScleraNET) that serves as a feature extractor for the vasculature probability maps. It needs to be noted that recognition techniques based on the vascular structure of the sclera are sensitive to view (or gaze) direction changes, which affect the amount of visible vasculature and consequently the performance of the final recognition approach. As a consequence, the vasculature is typically encoded using local image descriptors that allow for parts-based comparisons and are to some extent robust towards changes in the appearance of the vascular structure. Our goal with ScleraNET is to learn a single discriminative representation of the sclera that can directly be used for comparison purposes regardless of the given gaze direction. We, therefore, use a Multi-Task Learning (MTL) objective that takes both identity, but also gaze direction into account when learning the model parameters. As suggested in [60], the idea of MTL is to improve learning efficiency and prediction accuracy by considering multiple objectives when learning a shared representation. Because domain information is shared during learning due to the different objectives (pertaining to different tasks), the representations learned by the model offer better generalization ability than representations that rely only on a single objective during training. Since

we try to jointly learn to recognise gaze direction and identity from the vascular structure of the sclera with ScleraNET, the intermediate layers of the model need to encode information on both tasks in the generated representations.

In the following sections, we elaborate on ScleraNET and discuss its architecture, training procedure and deployment as a feature (or descriptor) extractor.

13.3.3.1 ScleraNET Architecture

The ScleraNET model architecture builds on the success of recent CNN models for various recognition tasks and incorporates design choices from the AlexNet [61] and VGG models [59]. We design the model as a (relatively) shallow network with a limited number of trainable parameters that can be learned using a modest amount of training data [11], but at the same time aim for a network topology that is able to generate powerful image representations for recognition. Consequently, we built on established architectural design choices that have proven to work well for a variety of computer vision tasks.

As illustrated in Fig. 13.3 and summarised in Table 13.2, the architecture consists of 7 convolutional layers (with ReLU activations) with multiple max-pooling layers in between followed by a global average pooling layer, one dense layer and two softmax classifiers at the top.

The first convolutional layer uses 128 reasonably large 7×7 filters with a stride of 2 to capture sufficient spatial context and reduce the dimensionality of the generated feature maps. The layer is followed by a max-pooling layer that further reduces the size of the feature maps by $2 \times$ along each dimension. Next, three blocks consisting of two convolutional and one max-pooling layer are utilised in the ScleraNET model. Due to the max-pooling layers, the spatial dimensions of the feature maps are halved after each block. To ensure a sufficient representational power of the feature maps, we double the number filters in the convolutional layers after each max-pooling operation. The output of the last of the three blocks is fed to a global average pooling layer and subsequently to a 512-dimensional Fully Connected (FC) layer. Finally, the FC layer is connected to two softmax layers, upon which an identity-oriented and a view-direction-oriented loss is defined for the MTL training procedure. The softmax layers are not used during run-time.

13.3.3.2 Learning Objective and Model Training

We define a cross-entropy loss over each of the two softmax classifiers at the top of ScleraNET for training. The first cross-entropy loss L_1 penalises errors when classifying subjects based on the segmented vasculature, and the second L_2 penalises errors when classifying different gaze directions. The overall training loss is a Multi-Task Learning (MTL) objective:

$$L_{total} = L_1 + \lambda L_2. \quad (13.4)$$

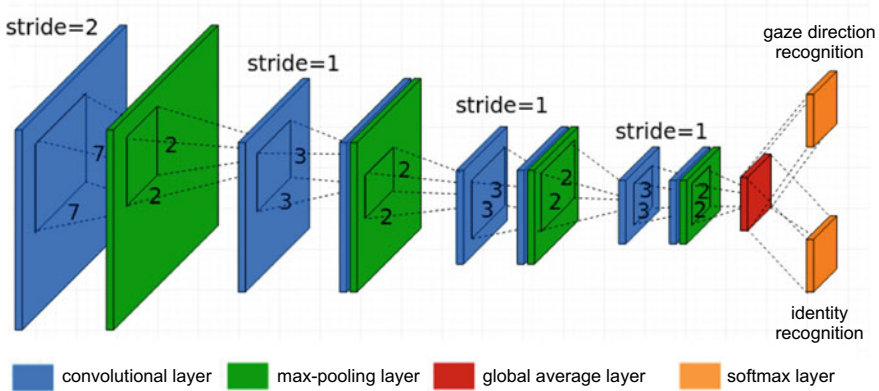


Fig. 13.3 Overview of the ScleraNET model architecture. The model incorporates design choices from the AlexNet [61] and VGG [59] models and relies on a Multi-Task Learning (MTL) objective that combines an identity and gaze-direction-related loss to learn discriminative vasculature representations for recognition

Table 13.2 Summary of the ScleraNET model architecture

| No. | Layer type | # Filters | Description |
|-----|------------------------|-----------|----------------------------|
| 1. | conv | 128 | 7×7 (stride of 2) |
| 2. | max-pooling | | 2×2 |
| 3. | conv | 128 | 3×3 (stride of 1) |
| 4. | conv | 128 | 3×3 (stride of 1) |
| 5. | max-pooling | | 2×2 |
| 6. | conv | 256 | 3×3 (stride of 1) |
| 7. | conv | 256 | 3×3 (stride of 1) |
| 8. | max-pooling | | 2×2 |
| 9. | conv | 512 | 3×3 (stride of 1) |
| 10. | conv | 512 | 3×3 (stride of 1) |
| 11. | max-pooling | | 2×2 |
| 12. | global average pooling | | |
| 13. | dense | 512 | |
| 14. | softmax (2 \times) | | Multi-task objective |

To learn the parameters θ of ScleraNET, we minimise the combined loss over some training data and when doing so give equal weights to both loss terms, i.e. $\lambda = 1$.

As suggested earlier, the intuition behind the MTL objective is to learn feature representations that are useful for both tasks and, thus, contribute to (identity) recognition performance as well as to the accuracy of gaze-direction classification. Alternatively, one can interpret the loss related to gaze-direction classification as a regularizer for the identity recognition process [62]. Hence, the additional term helps to learn (to a certain extent) view-invariant representations of the vasculature, or to put it dif-

ferently, it contributes towards more discriminative feature representations across different views.

13.3.3.3 Identity Inference with ScleraNET

Once the ScleraNET model is trained, we make it applicable to unseen identities by performing network surgery on the model and removing both softmax layers. We then use the 512-dimensional output from the fully connected layer as the feature representation of the vascular structure fed as input to the model.

If we again denote the probability map of the vascular structure produced by our two-step segmentation procedure as \mathbf{y} then the feature representation calculation procedure implemented by ScleraNET can be described as follows:

$$\mathbf{z} = g_{\theta}(\mathbf{y}), \quad (13.5)$$

where g_{θ} again denotes the mapping from the vascular structure \mathbf{y} to the feature representation \mathbf{z} by the ScleraNET model and θ stands for the model's parameters. The feature representation can ultimately be used with standard similarity measures to generate comparison scores for recognition purposes.

13.4 The Sclera Blood Vessels, Periocular and Iris (SBVPI) Dataset

In this section, we describe a novel dataset for research on sclera segmentation and recognition called *Sclera Blood Vessels, Periocular and Iris* (SBVPI), which we make publicly available for research purposes from <http://sclera.fri.uni-lj.si/>. While images of the dataset contain complete eyes, including the iris and periocular region, the focus is clearly on the sclera vasculature, which makes SBVPI the first publicly available dataset dedicated specifically to sclera (segmentation and) recognition research. As emphasised in the introductory chapter of the handbook, currently there exists no dataset designed specifically for sclera recognition, thus, SBVPI aims to fill this gap.

In the remainder of this section, we describe the main characteristics of the introduced dataset, discuss the acquisition procedure and finally elaborate on the available annotations.

13.4.1 Dataset Description

The SBVPI (Sclera Blood Vessels, Periocular and Iris) dataset consists of two separate parts. The first part is a dataset of periocular images dedicated to research in periocular biometrics and the second part is a dataset of sclera images intended for

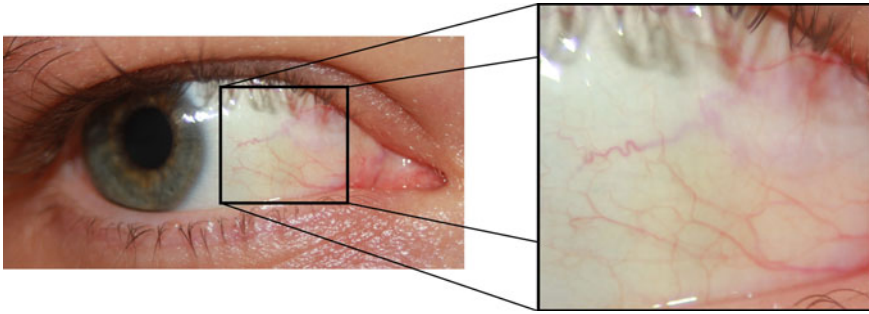


Fig. 13.4 An example image from the SVBPI dataset with a zoomed in region that shows the vascular patterns of the sclera

research into vascular biometrics. We focus in this chapter on the second part only, but a complete description of the data is available from the webpage of SBVPI.

The sclera-related part of SBVPI contains 1858 RGB images of 55 subjects. Images for the dataset were captured during a single recording session using a Digital Single-Lens Reflex camera (DSLR) (Canon EOS 60D) at the highest resolution and quality setting. Macro lenses were also used to capitalise on the quality and details visible in the captured images. The outlined capturing setup was chosen to ensure high-quality images, on which the vascular patterns of the sclera are clearly visible, as shown in Fig. 13.4.

During the image capturing process, the camera was positioned at a variable distance between 20 and 40 centimetres from the subjects. Before acquiring a sclera sample, the camera was always randomly displaced from the previous position by moving it approximately 0–30 cm left/right/up/down. During the camera-position change, the subjects also slightly changed the eyelid position and direction of view. With this acquisition setup, we ensured that the individual samples of the same eye looking at the same direction is always different from all other samples of the same eye looking in the same direction. It is known that the small changes in view direction cause complex non-linear deformations in the appearance of the vascular structure of the sclera [7] and we wanted our database to be suitable for the development of algorithms robust to such kind of changes.

The captured samples sometimes contained unwanted facial parts (e.g. eyebrows, parts of the nose, etc.). We, therefore, manually inspected and cropped (using a fixed aspect ratio) the captured images to ensure that only a relatively narrow periocular region was included in the final images as shown in the samples in Fig. 13.5. The average size of the extracted Region-Of-Interest (ROI) was around 1700×3000 pixels, which is sufficient to also capture the finer blood vessels of the sclera in addition to the more expressed vasculature. Thus, 1700×3000 px was selected as the target size of the dataset and all samples were rescaled (using bicubic interpolation) to this target size to make the data uniform in size.

The image capturing process was inspired by the MASD dataset [5]. Each subject was asked to look in one of four directions at the time, i.e. straight, left, right and up. For each view direction, one image was captured and stored for the dataset. This

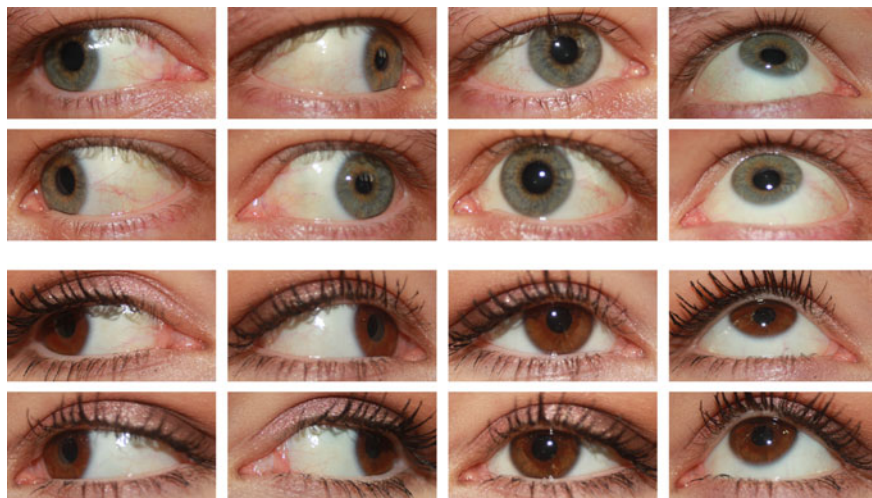


Fig. 13.5 Sample images from the SVBP dataset. The dataset contains high-quality samples with a clearly visible sclera vasculature. Each subject has at least 32 images covering both eyes and 4 view directions, i.e. up, left, right and straight. The top two rows show 8 sample images of a male subject and the bottom two rows show 8 sample images of a female subject from the dataset

process was repeated four times, separately for the left and right eye, and resulted in a minimum of 32 images per subject (i.e. 4 repetitions \times 4 view directions \times 2 eyes)—some subjects were captured more than four times. The images were manually inspected for blur and focus and images not meeting subjective quality criteria were excluded during the recording sessions. A replacement image was taken if an image was excluded. Subjects with sight problems were asked to remove prescription glasses, while contact lenses, on the other hand, were allowed. Care was also taken that no (or minimal) reflections caused by the camera's flash were visible in the images.

The final dataset is gender balanced and contains images of 29 female and 26 male subjects all of Caucasian origin. The age of the subjects varies from 18 to 80 with the majority of subjects being below 35-year old. SBVP contains eyes of different colours, which represents another source of variability in the dataset. A summary of the main characteristics of SBVP is presented in Table 13.3. For a high-level comparison with other datasets of ocular images, including those used for research in sclera recognition, please refer to Table 13.1.

13.4.2 Available Annotations

The dataset is annotated with identity (one of 55 identities), gender (male or female), eye class (left eye or right eye) and view/gaze-direction labels (straight, left, right, up), which are available for each of the 1858 SVBPI sclera images. Additionally,

Table 13.3 Main characteristics of the SVBP dataset

| Characteristic | Description |
|------------------------------|---|
| Acquisition device | DSLR camera, Canon EOS 60D + macro lenses |
| Number of images | 1858 |
| Number of subjects | 55 |
| Number of images per subject | 32 minimum, but variable |
| Image size | 1700 × 3000 px |
| Available annotations | Identity, gender, view direction, eye markup (segmentation masks) |



Fig. 13.6 Examples of the markups available with the SBVPI dataset. All images contain manually annotated irises and sclera regions and a subset of images has a pixel-level markup of the sclera vasculature. The images show (from left to right): a sample image from SBVPI, the iris markup, the sclera markup and the markup of the vascular structure

ground truth information about the location of certain eye parts is available for images in the dataset. In particular, all 1858 images contain a pixel-level markup of the sclera and iris regions, as illustrated in Fig. 13.6. The vascular structure and pupil area are annotated for a subset of the dataset i.e. 130 images. The segmentation masks were generated manually using the GNU Image Manipulation Program (GIMP) and stored as separate layers for all annotated images. The markups are included in SBVPI in the form of metadata.

The available annotations make our dataset suitable for research work on sclera recognition, but also segmentation techniques, which is not the case with competing datasets. Especially the manual pixel-level markup of the sclera vasculature is a unique aspect of the sclera-related part of SBVPI.

13.5 Experiments and Results

In this section, we evaluate our sclera recognition pipeline. We start the section with a description of the experimental protocol and performance metrics used, then discuss the training procedure for all parts of our pipeline and finally proceed to the presentation of the results and corresponding discussions. To allow for reproducibility of our results, we make all models, data, annotations and experimental scripts publicly available through <http://sclera.fri.uni-lj.si/>.

13.5.1 Performance Metrics

The overall performance of our recognition pipeline depends on the performance of the segmentation part used to extract the vascular structure from the input images and on the discriminative power of the feature representation extracted from the segmented vasculature. In the experimental section we, therefore, conduct separate experiments for the segmentation and feature extraction parts of our pipeline. Next, we describe the performance metrics used to report results for these two parts.

Performance metrics for the segmentation experiments: We measure the performance of the segmentation models using standard performance metrics, such as *precision*, *recall* and the *F1-score*, which are defined as follows [57, 58, 63]:

$$precision = \frac{TP}{TP + FP}, \quad (13.6)$$

$$recall = \frac{TP}{TP + FN}, \quad (13.7)$$

$$F1\text{-score} = 2 \cdot \frac{precision \cdot recall}{precision + recall}, \quad (13.8)$$

where TP denotes the number of true positive pixels, FP stands for the number of false positive pixels and FN represents the number of false negative pixels.

Among the above measures, precision measures the proportion of correctly segmented pixels with respect to the overall number of true pixels of the target class (e.g. the sclera region) and, hence, provides information about how many segmented pixels are in fact relevant. Recall measures the proportion of correctly segmented pixels with respect to the overall number of pixels assigned to the target class and, hence, provides information about how many relevant pixels are found/segmented. Precision and recall values are typically dependent—it is possible to increase one at the expense of the other and vice versa by changing segmentation thresholds. If a simple way to compare two segmentation models is required, it is, therefore, convenient to combine precision and recall into a single metric called F1-score, which is also used as an additional performance metric in this work [57].

Note that when using a fixed segmentation threshold Δ , we obtain fixed precision and recall values for the segmentation outputs, while the complete trade-off between precision and recall can be visualised in the form of precision–recall curves by varying the segmentation threshold Δ over all possible values. This trade-off shows a more complete picture of the performance of the segmentation models and is also used in the experimental section [57].

Performance metrics for the recognition experiments: We measure the performance of the feature extraction (and recognition) part of our pipeline in verification experiments and report performance using standard False Acceptance (FAR) and False Rejection error Rates (FRR). FAR measures the error over the illegitimate verification attempts and FRR measures the error over the legitimate verification

attempts. Both error rates, FAR and FRR, depend on the value of a decision threshold (similar to the precision and recall values from the previous section) and selecting a threshold that produces low FAR values contributes towards high FRR scores and vice versa, selecting a threshold that produced low FRR values generates high FAR scores. Both error rates are bounded between 0 and 1. A common practice in biometric research is to report Verification Rates (VER) instead of FRR scores, where VER is defined as $1 - \text{FRR}$ [11, 64–66]. We also adopt this practice in our experiments.

To show the complete trade-off between FAR and FRR (or VER), we generate Receiver Operating Characteristic (ROC) curves by sweeping over all possible values of the decision threshold. We then report on several operating points from the ROC curve in the experiments, i.e. the verification performance at a false accept rate of 0.1% (VER@0.1FAR), the verification performance at a false accept rate of 1% (VER@1FAR) and the so-called Equal Error Rate (EER), which corresponds to the ROC operating point, where FAR and FRR are equal. Additionally, we provide Area Under the ROC Curve (AUC) scores for all recognition experiments, which is a common measure of the accuracy of binary classification tasks, such as biometric identity verification.

13.5.2 *Experimental Protocol and Training Details*

We conduct experiments on the SBVPI dataset introduced in Sect. 13.4 and use separate experimental protocols for the segmentation and recognition parts of our pipeline. The protocols and details on the training procedures are presented below.

13.5.2.1 **Segmentation Experiments**

The segmentation part of our pipeline consists of two components. The first generates an initial segmentation result and locates the sclera region in the input image, whereas the second segments the vasculature from the located sclera.

Sclera segmentation: To train and test the segmentation model for the first component of our pipeline, we split the sclera-related SBVPI data into two (image and subject) disjoint sets:

- A *training set* consisting of 1160 sclera images. These images are further partitioned into two subsets. The first, comprising 985 images, is used to learn the model parameters and the second, comprising 175 images, is employed as the validation set and used to observe the generalization abilities of the model during training and stop the learning stage if the model starts to over-fit.
- A *test set* consisting of 698 sclera images. This set is used to test the final performance of the trained segmentation model and compute performance metrics for the experiments.

To avoid over-fitting, the training data (i.e. 985 images) is augmented by a factor of 40 by left–right flipping, cropping, Gaussian blurring, changing the image brightness and application of affine transformations such as scale changes, rotations (up to $\pm 35^\circ$) and shearing.

Training of the SegNet model for the initial segmentation step (for sclera segmentation) is conducted on a GTX 1080 Ti with 11GB of RAM. We use the Caffe implementation of SegNet made available by the authors² for the experiments. The input images are rescaled to fixed size of 360×480 pixels for the training procedure. The model weights are learned using Stochastic Gradient Descent (SGD) and Xavier initialization [67]. The learning rate is set to 0.001, the weight decay to 0.0005, the momentum to 0.9 and the batch size to 4. The model converges after 26, 000 iterations.

Vasculature segmentation: The second component of our pipeline requires a pixel-level markup of the vascular structure of the sclera for both the training and the testing procedure. The SBVP dataset contains a total of 130 such images, which are used to learn the SegNet model for this part and assess its performance. We again partition the data into two (image and subject) disjoint sets:

- A *training set* of 98 images, which we split into patches of manageable size, i.e. 360×480 pixels. We generate a total of 788 patches by sampling from the set of 98 training images and randomly select 630 of these patches for learning the model parameters and use the remaining 158 patches as our validation set during training. To avoid over-fitting, we again augment the training patches 40-fold using random rotations, cropping and colour manipulations.
- A *test set* consisting of 32 images. While the test images are again processed patch-wise, we report results over the complete images and not the intermediate patch representations.

To train the segmentation model for the vascular structure of the sclera, we use the same setup as described above for the sclera segmentation model.

13.5.2.2 Recognition Experiments

The vascular structure of the sclera is an epigenetic biometric characteristic with high discriminative power that is known to differ between the eyes of the same subject. We, therefore, treat the left and right eye of each subject in the SBVPI dataset as a unique identity and conduct recognition experiments with 110 identities. Note that such a methodology is common for epigenetic biometric traits and has been used regularly in the literature, e.g. [68, 69].

For the recognition experiments, we split the dataset into subject disjoint training and test sets, where the term subject now refers to one of the artificially generated 110 identities. The training set that is used for the model learning procedure consists of 1043 images belonging to 60 different identities. These images are divided between

²Available from: <https://github.com/alexkendall/caffe-segnet>.

the actual training data (needed for the learning model parameters) and the validation data (needed for the early stopping criterion) in a ratio of 70% versus 30%. The remaining 815 images belonging to 50 subjects are used for testing purposes.

For the training procedure, we again use a GTX 1080 Ti GPU. We implement our ScleraNET model in Keras and initialize its weights in accordance with the method from [67]. We use the Adam optimizer with a learning rate of 0.001, beta1 equal to 0.9 and beta2 equal to 0.999 to learn the model parameters. We augment the available training data on the fly to avoid over-fitting and to ensure sufficient training material. We use random shifts (± 20 pixels in each direction) and rotations ($\pm 20^\circ$) for the augmentation procedure. The model reaches stable loss values after 70 epochs. As indicated in Sect. 13.3.3.3, once trained, the model takes 400×400 px images as input and returns a 512-dimensional feature representation at the output (after network surgery). The input images to the model are complete probability maps of the sclera vasculature down-sampled to the target size expected by ScleraNET. Note that because the down-sampling is performed after segmentation of the vasculature, information on the smaller veins is not completely lost when adjusting for the input size of the descriptor-computation model.

13.5.3 Evaluation of Sclera Segmentation Models

We start our experiments with an evaluation of the first component of the sclera recognition pipeline, which produces the initial segmentation of the sclera region. The goal in this series of experiments is to show how the trained SegNet architecture performs for this task and how it compares to competing deep models and existing sclera segmentation techniques. We need to note that while the error from this stage is propagated throughout the entire pipeline to some extent, these errors are not as critical as long as the majority of the sclera region is segmented from the input images. Whether the segmentation is precise (and able to find the exact border between the sclera region and fine details such as the eyelashes, eyelids, etc.) is not of paramount importance at this stage.

To provide a frame of reference for the performance of SegNet, we implement 4 additional segmentation techniques and apply them to our test data. Specifically, we implement 3 state-of-the-art CNN-based segmentation models and one segmentation approach designed specifically for sclera segmentation. Note that these techniques were chosen, because they represent the top performing techniques from the sclera segmentation competitions of 2017 and 2018. Details on the techniques are given below:

- *RefineNet-50* and *RefineNet-101*: RefineNet [70] is recent deep segmentation model built around the concept of residual learning [71]. The main idea of RefineNet is to exploit features from multiple levels (i.e. from different layers) to produce high-resolution semantic feature maps in a coarse-to-fine manner. Depending on the depth of the model, different variants of the model can be trained.

In this work, we use two variants, one with 50 model layers (i.e. RefineNet-50) and one with 101 layers (i.e. RefineNet-101). We train the models on the same data and with the same protocol as SegNet (see Sect. 13.5.2.1) and use a publicly available implementation for the experiments.³ Note that RefineNet was the top performer of the sclera 2018 segmentation competition held in conjunction with the 2018 International Conference on Biometrics (ICB) [10].

- *UNet*: The UNet [72] model represents a popular CNN architecture particularly suited for data-scarce image translation tasks such as sclera segmentation. Similarly to SegNet, the model uses an encoder–decoder architecture but ensures information flow from the encoder to the decoder by concatenating feature maps from the encoder with the corresponding outputs of the decoder. We train the models on the same data and with the same protocol as SegNet. For the experiments we use our own Keras (with TensorFlow backend) implementation of UNet and make it publicly available to the research community.⁴
- *Unsupervised Sclera Segmentation (USS)* [73]: Different from the models above, USS represents an unsupervised segmentation technique, which does not rely on any prior knowledge. The technique operates on greyscale images and is based on an adaptive histogram normalisation procedure followed by clustering and adaptive thresholding. Details on the method can be found in [73]. The technique was ranked second in the 2017 sclera segmentation competition. Code provided by the author of USS was used for the experiments to ensure a fair comparison with our segmentation models.

Note that the three CNN-based models produce probability maps for the sclera vasculature, whereas the USS approach returns only binary masks. In accordance with these characteristics we report precision, recall and F1-scores for all tested methods (the CNN models are thresholded with a value of Δ that ensures the highest possible F1-score) in Table 13.4 and complete precision–recall curves only for the CNN-based methods in Fig. 13.7. For both the quantitative results and the performance graphs, we also report standard deviations to have a measure of dispersion across the test set.

The results show that the CNN-based models perform very similarly (there is no statistical difference in performance between the models). The unsupervised approach USS, on the other hand, performs somewhat worse, but the results are consistent with the ranking reported in [5]. Overall, the CNN models all achieve near-perfect performance and are able to ensure F1-scores of around 0.95. Note that such high results suggest that performance for this task is saturated and further improvements would likely be a consequence of over-fitting to the dataset and corresponding manual annotations.

The average processing time per image (calculated over a test set of 100 images) is 1.2s for UNet, 0.6s for RefineNet-50, 0.8s for RefineNet-101, 0.15s for SegNet and 0.34s for USS. In our experiments, SegNet is the fastest of the tested models.

We show some examples of the segmentation results produced by the tested segmentation models in Fig. 13.8. Here, the first column shows the original RGB ocular

³Available from <https://github.com/guosheng/refinenet>.

⁴Available from: <http://sclera.fri.uni-lj.si/>.

Table 13.4 Segmentation results generated based on binary segmentation masks. For the CNN-based models, the masks are produced by thresholding the generated probability maps with a value of Δ that ensures the highest possible F1-score, whereas the USS approach is designed to return a binary mask of the sclera region only. Note that all CNN perform very similarly with no statistical difference in segmentation performance, while the unsupervised USS approach performs somewhat worse. The reported performance scores are shown in the form $\mu \pm \sigma$, computed over all test images

| Algorithm | Precision | Recall | F1-score |
|-------------------------------------|-------------------|-------------------|-------------------|
| UNet [72] (ours) | 0.936 ± 0.044 | 0.930 ± 0.037 | 0.933 ± 0.037 |
| RefineNet-50 [70] (ours) | 0.959 ± 0.020 | 0.959 ± 0.020 | 0.959 ± 0.018 |
| RefineNet-101 [70] (ours) | 0.953 ± 0.025 | 0.951 ± 0.023 | 0.952 ± 0.021 |
| SegNet [5, 57] (ours, this chapter) | 0.949 ± 0.024 | 0.949 ± 0.022 | 0.949 ± 0.021 |
| USS [5, 73] | 0.729 ± 0.041 | 0.718 ± 0.039 | 0.723 ± 0.036 |

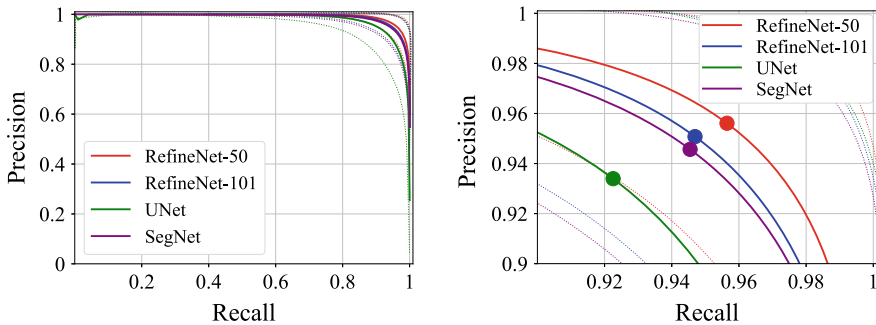


Fig. 13.7 Precision–recall curves for the tested CNN models. USS is not included here, as it returns only binary masks of the sclera region. The left graph shows the complete plot generated by varying the segmentation threshold Δ over all possible values, whereas the right graph shows a zoomed in region to highlight the minute differences between the techniques. The marked points stand for the operating points with the highest F1-Score. The dotted lines show the dispersion (σ) of the precision and recall scores over the test images

images, the second shows the manually annotated ground truth and the remaining columns show results generated by (from left to right): USS, RefineNet-50, RefineNet-101, SegNet and UNet. These results again confirm that all CNN-based models ensure similar segmentation performance. All models segment the sclera region well and differ only in some finer details, such as eyelashes, which are not really important for the second segmentation step, where the vasculature needs to be extracted from the ocular images.

Consequently, any of the tested CNN-based segmentation models could be used in our sclera recognition pipeline for the initial segmentation step, but we favour

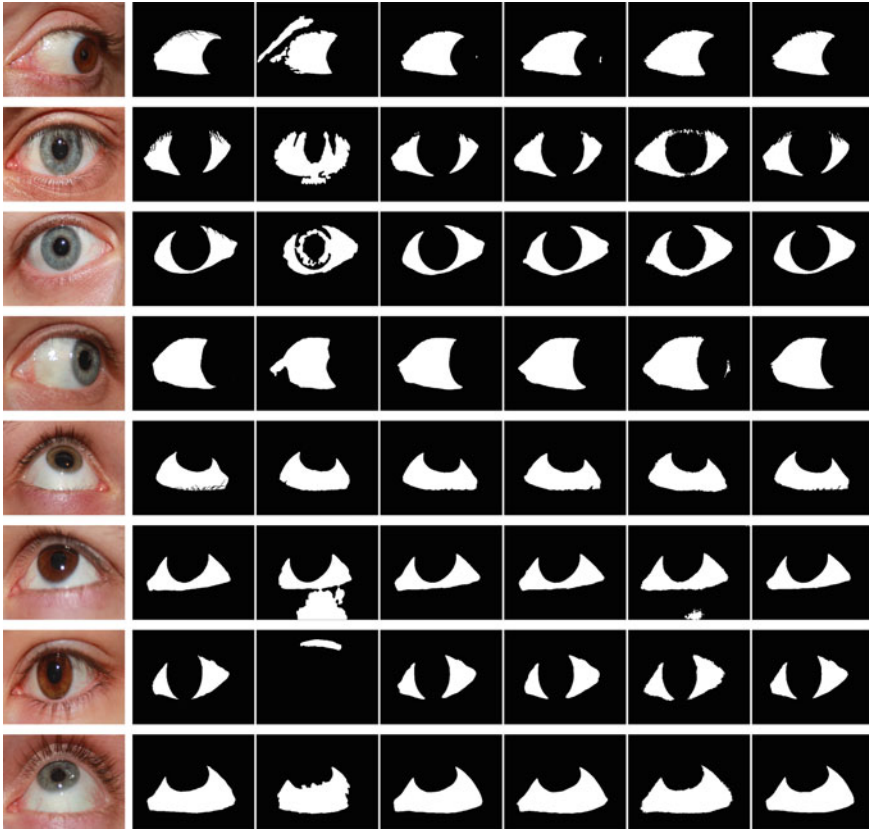


Fig. 13.8 Visual examples of the segmentation results produced by the tested segmentation models. The first column shows the input RGB ocular images, the second the manually annotated ground truth and the remaining columns show the results generated by (from left to right): USS, RefineNet-50, RefineNet-101, SegNet and UNet. Note that the CNN models (last four columns) produce visually similar segmentation results and differ only in certain fine details

SegNet because of the fast prediction time, which is 4 times faster the second fastest CNN model, i.e. RefineNet-50.

13.5.4 Evaluation of Vasculature Segmentation Models

In the next series of experiments, we evaluate the performance of the second segmentation step of our pipeline, which aims to locate and segment the vascular structure of the sclera from the input image. The input to this step is again an RGB ocular image (see Fig. 13.9), but masked with the segmentation output produced by the SegNet model evaluated in the previous section.

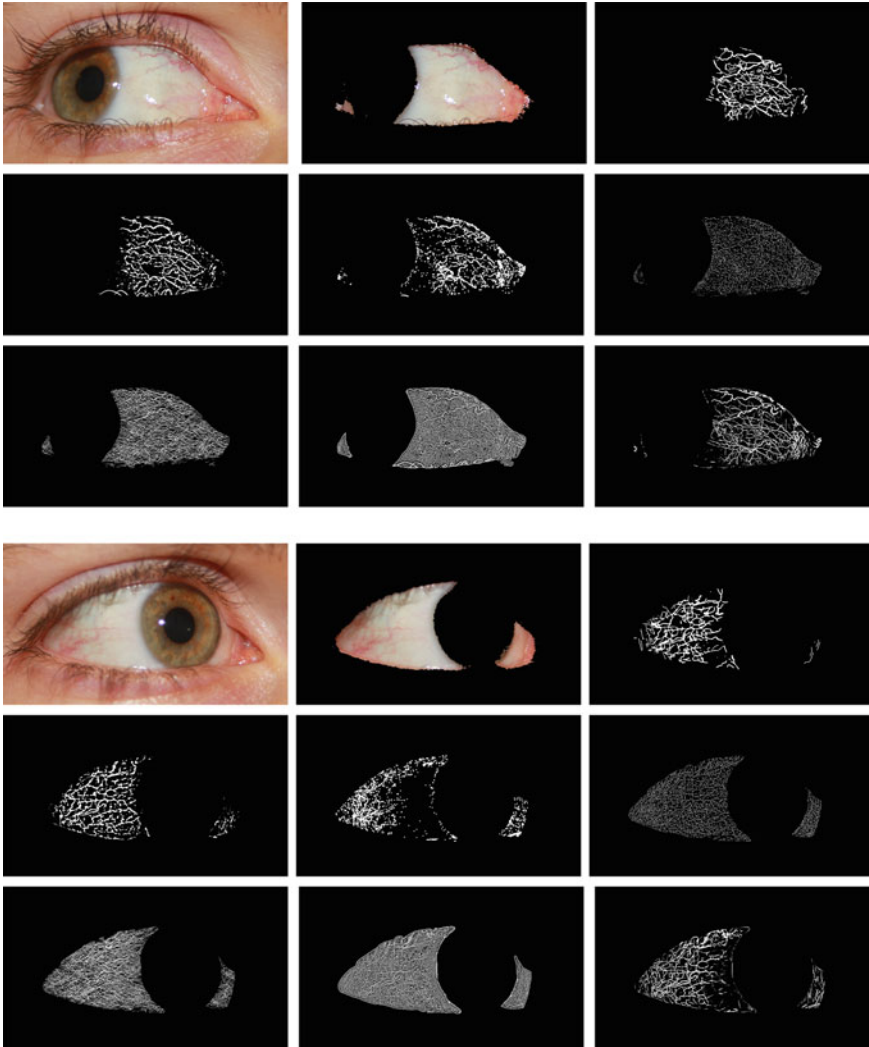


Fig. 13.9 Examples of vasculature segmentation results. Each of the two image blocks shows (from left to right and top to bottom): the input RGB ocular image, the input image masked with the sclera region produced by the initial segmentation step, the ground truth markup, results for the proposed cascaded SegNet assembly, and results for the Adaptive Gaussian Thresholding (AGT), and the NMC, NRLT, Coye and B-COSFIRE approaches. The results show the generated binary masks corresponding to the operating point used in Table 13.5. Note that the proposed approach most convincingly captures the characteristics of the manual vasculature markup. Best viewed electronically and zoomed in

As emphasised earlier, we conduct segmentation with our approach in a patch-wise manner to ensure that information about the finer details of the sclera vasculature is not lost. Because the second SegNet model of the cascaded assembly outputs probability maps, we use adaptive Gaussian thresholding [74] to generate binary masks to compare with the manually annotated ground truth. To assess performance, we compute results over the binary masks and again report fixed precision, recall and F1-score values in this series of experiments. The performance scores are computed for the operating point on the precision–recall curve that corresponds to the maximum possible F1-score. We again report standard deviations in addition to the average scores to have a measure of dispersion for the results of the test data.

For comparison purposes, we implement a number of competing techniques from the literature that are regularly used for vessel segmentation in the field of vascular biometrics, i.e. (i) Adaptive Gaussian Thresholding (AGT) [74], (ii) Normalized Maximum Curvature (NMC) [75], (iii) Normalized Repeated Line Tracking (NRLT) [76], (iv) Coye filtering [77] and (v) the B-COSFIRE approach from [78, 79]. The NMC and NRLT approaches represent a modified version of the original segmentation techniques and are normalised to return continuous probability maps rather than binarized segmentation results. The hyper-parameters of all baseline techniques (if any) are selected to maximise performance. The techniques are implemented using publicly available source code.⁵ We note again that no supervised approach to sclera vasculature segmentation has been presented in the literature so far. We focus, therefore, exclusively on unsupervised segmentation techniques in our comparative assessment.

The results of the experiments are presented in Table 13.5. As can be seen, SegNet ensures the best overall results by a large margin with an average F1-score of 0.727. The B-COSFIRE techniques, regularly used for vasculature segmentation in retina images, is the runner-up with an average F1-score of 0.393, followed closely by AGT thresholding with an F1-score of 0.306. The NMC, NRLT and Coye filter approaches result in worse performance with F1-scores below 0.25. While the performance difference between the SegNet model and the competing techniques is considerable, it is also expected, as SegNet is trained on the manually annotated vasculature, while the remaining approaches rely only on local image characteristics to identify the vascular structure of the sclera. As a result, the vasculature extracted by the unsupervised techniques (NMC, NRLT, Coye filter and B-COSFIRE) does not necessarily correspond to the markup generated by a human annotator. However, the low-performance scores of the unsupervised techniques do not indicate that the extracted vasculature is useless for recognition, but only that there is low correspondence with the manual markup. To investigate the usefulness of the extracted vascular patterns of these

⁵Code for the techniques is available from: AGT from OpenCV: <https://opencv.org/>, NMC and NRLT from Mathworks: <https://www.mathworks.com/matlabcentral/fileexchange/35716-miura-et-al-vein-extraction-methods>

Coye filter from Mathworks: <https://www.mathworks.com/matlabcentral/fileexchange/50839-novel-retinal-vessel-segmentation-algorithm-fundus-images>

B-COSFIRE from Mathworks: <https://www.mathworks.com/matlabcentral/fileexchange/49172-trainable-cosfire-filters-for-curvilinear-structure-delineation-in-images>.

Table 13.5 Comparison of vasculature segmentation techniques. Results are presented for the proposed cascaded SegNet assembly, as well as for five competing unsupervised segmentation approaches from the literature. The probability maps generated by the techniques have been thresholded to allow for comparisons with the annotated binary vasculature markup. Note that the proposed approach achieves the best overall performance by a large margin

| Algorithm | Precision | Recall | F1-score |
|---|-------------------|-------------------|-------------------|
| SegNet [15] + AGT (ours, this chapter) | 0.806 ± 0.155 | 0.675 ± 0.131 | 0.727 ± 0.120 |
| Adaptive Gaussian thresholding (AGT) | 0.308 ± 0.119 | 0.372 ± 0.201 | 0.306 ± 0.120 |
| Normalized maximum curvature (NMC) | 0.240 ± 0.097 | 0.247 ± 0.044 | 0.232 ± 0.062 |
| Normalized repeated line tracking (NRLT) | 0.145 ± 0.055 | 0.314 ± 0.114 | 0.191 ± 0.066 |
| Coye filter | 0.143 ± 0.070 | 0.376 ± 0.085 | 0.198 ± 0.078 |
| B-COSFIRE | 0.351 ± 0.142 | 0.480 ± 0.083 | 0.393 ± 0.116 |

techniques for recognition, we conduct a series of recognition experiments in the next section.

To put the reported results into perspective and show what the scores mean visually, we present in Fig. 13.9 some qualitative segmentation results. Here, each of the two image blocks shows (from left to right and top to bottom): the input ocular image, the masked sclera region, the ground truth annotation and results for the proposed cascaded SegNet assembly, the Adaptive Gaussian Thresholding (AGT), and the NMC, NRLT, Coye and B-COSFIRE techniques. It is interesting to see what level of detail the SegNet-based model is able to recover from the input image. Despite the relatively poor contrast of some of the finer veins, the model still successfully segments the sclera vasculature from the input images. The B-COSFIRE results are also convincing when examined visually, but as emphasised earlier do not result in high-performance scores when compared to the manual markup. Other competing models are less successful and generate less precise segmentation results. However, as suggested above, the competing models use no supervision to learn to segment the vascular structures and therefore generate segmentation results that do not correspond well to the manual markup.

To further highlight the quality of the segmentation ensured by the SegNet-based model, we show a close up of the vascular structure of an eye and the corresponding segmentation output in Fig. 13.10. We see that the model successfully segments most of the vascular structure, but also picks up on the eyelashes, which very much resemble the vein patterns of the sclera even from a human perspective. In the area where reflections are visible, the model is not able to recover the vascular structure from the input image. Furthermore, despite the patch-wise processing used with the cascaded SegNet segmentation approach, we observe no visible artifacts caused

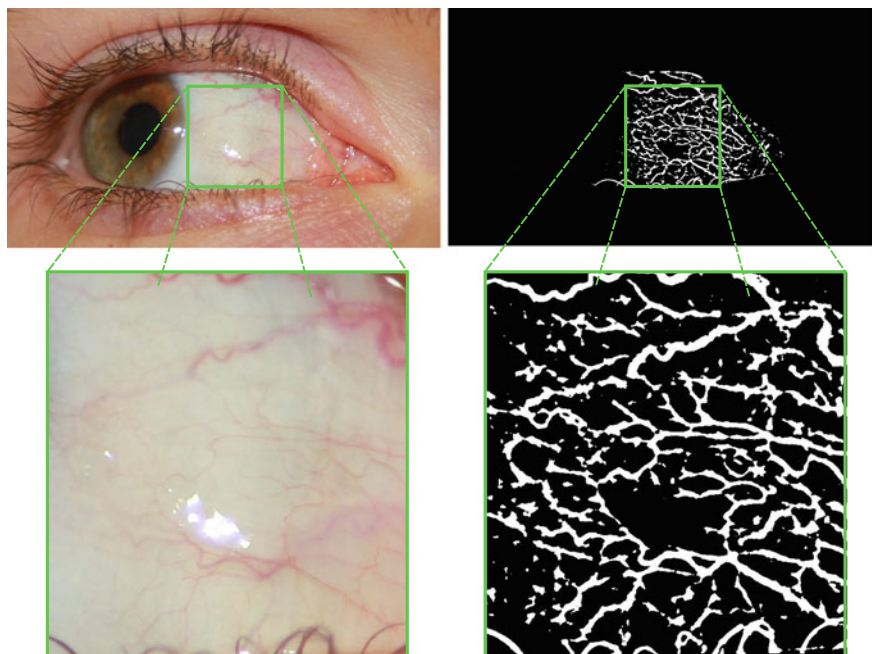


Fig. 13.10 Visualisation of the fine vascular structure recovered by our segmentation model. The image shows a zoomed in region of the vascular structure of the eye (on the left) and the corresponding binarized output of our model (on the right)

by the re-assembly procedure. We assume this is a consequence of the run-time augmentation step that smooths out such artifacts.

Because the segmentation is performed in a patch-wise manner, the average time needed to process one input image with the proposed model in this part is 5.6 seconds when using a single GPU (please note that this step can be parallelised using multiple GPUs, because patch predictions can be calculated independently). For comparison, the average processing time for AGT is 1.2s, for NMC it is 32.5s, for NRLT the processing time is 7.9s, for Coye it is 1.2s and for the B-COSFIRE the processing time is 13.9s. However, note that different programming languages were used for the implementation of the segmentation methods, so the processing times need to be interpreted accordingly. For the proposed cascaded SegNet assembly, the entire region-of-interest extraction step (which comprises the initial sclera segmentation and vascular structure segmentation steps), takes around 6s using a single GPU for one input image on average.

Overall, these results suggest that the trained segmentation model is able to produce good quality segmentation results that can be used for recognition purposes. We evaluate the performance of our recognition approach with the generated segmentation outputs next.

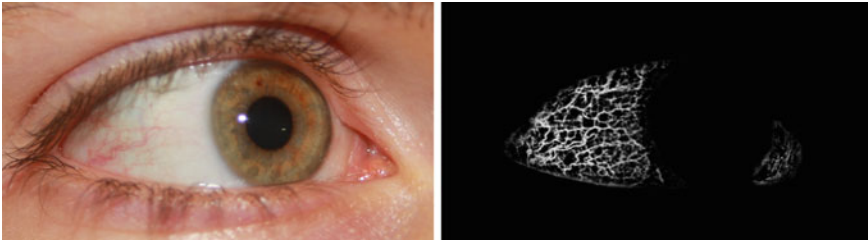


Fig. 13.11 Example of an input image and the corresponding probability map generated by the SegNet model. The probability mask on the left is used as input to the ScleraNET model

13.5.5 Recognition Experiments

In the last series of experiments, we assess the performance of the entire recognition pipeline and feed the segmented sclera vasculature into our ScleraNET model for feature extraction. Note again that we use the probability output of the segmentation models as input to ScleraNET (marked y in Fig. 13.2) and not the generated binary masks of the vasculature. An example of the probability map generated with the SegNet model is shown in Fig. 13.11. Once a feature representation is computed from the input image, it is used with the cosine similarity to compute similarity scores and to ultimately conduct identity inference. The feature computation procedure takes 0.1 s per image on average.

To evaluate the recognition performance of ScleraNET, we conduct verification experiments using the following experimental setup:

- We first generate user templates by randomly selecting four images of each subject in the test set. We sample the test set in a way that ensures that each template contains all four gaze directions (i.e. up, down, left and right). Since each subject has at least 4 images of each gaze direction, we are able to generate multiple templates for each subject in the test set.
- Next, we use all images in the test set and compare them to the generated user templates. The comparison is conducted by comparing (using the cosine similarity) the query vasculature descriptor to the descriptors of each image in the template. The highest similarity score is kept as the score for the query-to-template comparison. If the query image is also present in the template, we exclude the score from the evaluation.
- We repeat the entire process 5-times to estimate average performance scores as well as standard deviations. The outlined setup results in a total of 1228 legitimate and 121572 illegitimate verification attempts in each of the 5 repetitions.

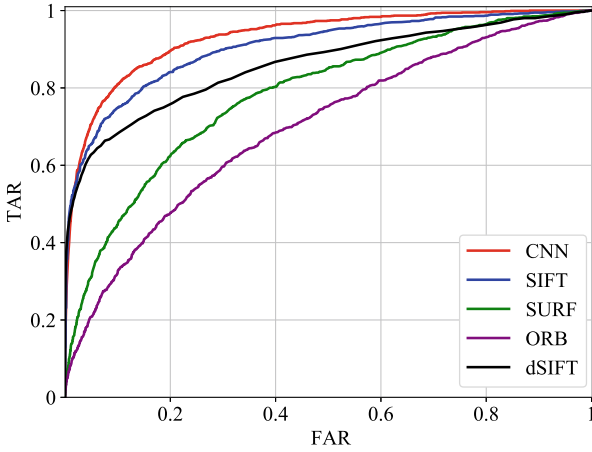
Because the ocular images are not aligned, we implement multiple descriptor-based approaches for comparison. Specifically, we implement the dense SIFT (dSIFT hereafter) approach from [8] and several keypoint based techniques. For the latter, we compute SIFT [80], SURF [81] and ORB [82] descriptors using their corresponding keypoint detectors. For each image-pair comparison, we use the average Euclidean

Table 13.6 Results of the recognition experiments. The table shows performance scores for five different descriptor-computation strategies and five approaches to vasculature segmentation. For each performance metric, the best overall result is coloured red and the best results for a given segmentation approach is coloured blue. The proposed ScleraNET model ensures competitive performance significantly outperforming the competing models when applied on the segmentation results generated by the proposed cascaded SegNet assembly

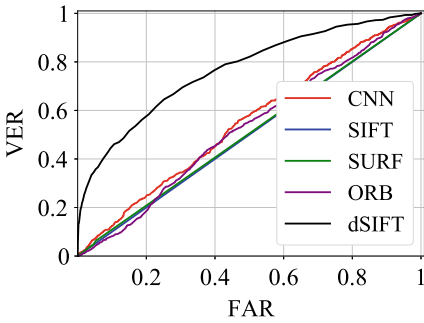
| Segment. | Algorithm | VER@0.1FAR | VER@1FAR | EER | AUC |
|------------------|------------------|----------------------|----------------------|----------------------|----------------------|
| C. SegNet (ours) | ScleraNET (ours) | 0.181 ± 0.009 | 0.459 ± 0.009 | 0.145 ± 0.002 | 0.933 ± 0.002 |
| | SIFT | 0.184 ± 0.076 | 0.452 ± 0.040 | 0.176 ± 0.005 | 0.903 ± 0.005 |
| | SURF | 0.023 ± 0.007 | 0.126 ± 0.010 | 0.286 ± 0.004 | 0.782 ± 0.005 |
| | ORB | 0.017 ± 0.004 | 0.080 ± 0.011 | 0.351 ± 0.003 | 0.704 ± 0.005 |
| | Dense SIFT | 0.326 ± 0.016 | 0.507 ± 0.010 | 0.221 ± 0.004 | 0.865 ± 0.002 |
| NMC | ScleraNET | 0.002 ± 0.001 | 0.023 ± 0.003 | 0.425 ± 0.004 | 0.596 ± 0.004 |
| | SIFT | 0.000 ± 0.000 | 0.000 ± 0.000 | 0.500 ± 0.000 | 0.500 ± 0.000 |
| | SURF | 0.017 ± 0.024 | 0.031 ± 0.016 | 0.488 ± 0.013 | 0.535 ± 0.010 |
| | ORB | 0.000 ± 0.000 | 0.005 ± 0.005 | 0.504 ± 0.006 | 0.497 ± 0.006 |
| | Dense SIFT | 0.063 ± 0.014 | 0.184 ± 0.028 | 0.371 ± 0.012 | 0.683 ± 0.010 |
| NRLT | ScleraNET | 0.112 ± 0.011 | 0.311 ± 0.006 | 0.196 ± 0.008 | 0.888 ± 0.004 |
| | SIFT | 0.000 ± 0.000 | 0.014 ± 0.005 | 0.500 ± 0.001 | 0.500 ± 0.002 |
| | SURF | 0.000 ± 0.000 | 0.021 ± 0.013 | 0.492 ± 0.008 | 0.509 ± 0.005 |
| | ORB | 0.000 ± 0.000 | 0.021 ± 0.010 | 0.502 ± 0.005 | 0.499 ± 0.007 |
| | Dense SIFT | 0.047 ± 0.004 | 0.153 ± 0.010 | 0.362 ± 0.008 | 0.701 ± 0.004 |
| Coye | ScleraNET | 0.067 ± 0.008 | 0.215 ± 0.007 | 0.267 ± 0.006 | 0.812 ± 0.004 |
| | SIFT | 0.000 ± 0.000 | 0.036 ± 0.014 | 0.496 ± 0.001 | 0.507 ± 0.002 |
| | SURF | 0.000 ± 0.000 | 0.000 ± 0.000 | 0.500 ± 0.005 | 0.497 ± 0.005 |
| | ORB | 0.002 ± 0.001 | 0.023 ± 0.005 | 0.451 ± 0.005 | 0.568 ± 0.006 |
| | Dense SIFT | 0.091 ± 0.005 | 0.234 ± 0.018 | 0.300 ± 0.004 | 0.772 ± 0.004 |
| B-COSFIRE | ScleraNET | 0.042 ± 0.004 | 0.140 ± 0.008 | 0.337 ± 0.005 | 0.723 ± 0.006 |
| | SIFT | 0.000 ± 0.000 | 0.012 ± 0.005 | 0.488 ± 0.002 | 0.522 ± 0.003 |
| | SURF | 0.000 ± 0.000 | 0.000 ± 0.000 | 0.494 ± 0.005 | 0.513 ± 0.003 |
| | ORB | 0.000 ± 0.000 | 0.008 ± 0.002 | 0.467 ± 0.003 | 0.539 ± 0.004 |
| | Dense SIFT | 0.110 ± 0.011 | 0.242 ± 0.011 | 0.325 ± 0.006 | 0.748 ± 0.005 |

distance between matching descriptors as the similarity score for recognition. Since the descriptor-based approaches are local and rely on keypoint correspondences, they are particularly suitable for problems such as sclera recognition, where (partially visible) unaligned vascular structures under different views need to be matched against each other. We conduct experiments with the vasculature extracted with the proposed cascaded SegNet assembly, so we are able to evaluate our complete processing pipeline, but also with the segmentation results produced by the competing segmentation approaches evaluated in the previous section, i.e. NMC, NRLT, Coye and B-COSFIRE.

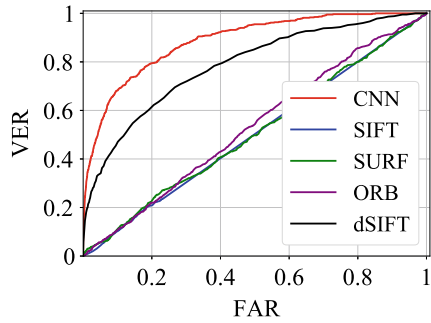
From the results in Table 13.6 and Fig. 13.12 (results for ScleraNET in the figures are marked as CNN), we see that the proposed pipeline (cascaded SegNet assembly +



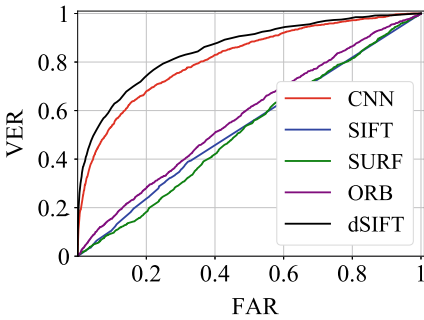
(a) Recognition results based on vasculature extracted with the SegNet assembly.



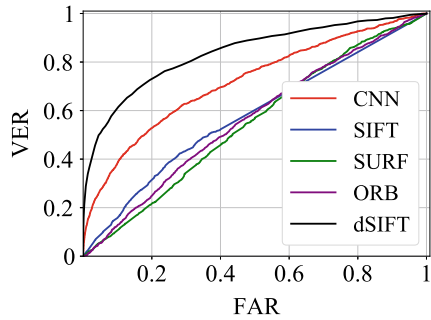
(b) NMC-based ROC curves.



(c) NRLT-based ROC curves.



(d) Coye-filter-based ROC curves.



(e) B-COSFIRE-based ROC curves.

Fig. 13.12 Results of the verification experiments. The graphs show recognition results for several feature extraction techniques and multiple approaches to vasculature segmentation. The pipeline proposed in this chapter results in the best overall performance

ScleraNET) ensures an average AUC of 0.933 for the verification experiments compared to the average AUC of 0.903 for the runner-up, the SIFT-based approach. Interestingly, the dSIFT approach is very competitive at the lower FAR values, but becomes less competitive at the higher values of FAR—see Fig. 13.12a. This behaviour can likely be ascribed to the dense nature of the descriptor, which makes it difficult to reliably compare images when there is scale and position variability present in the samples. The remaining three descriptors, SIFT, SURF and ORB, are less competitive and result in lower performance scores.

The segmentation results generated by the proposed cascaded SegNet assembly appear to be the most suitable for recognition purposes, as can be seen by comparing the ROC curves from Fig. 13.12b–e, to the results in Fig. 13.12a, or examining the lower part of Table 13.6. While the NMC, NRLT, Coye and B-COSFIRE segmentation results (in the form of probability maps) result in above-random verification performance with the ScleraNET and dSIFT descriptors, the performance is at chance for the keypoint-descriptor-based methods—SIFT, SURF and ORB. The reason for this is the difficulty of finding matching descriptors in the images, which leads to poor performance. The ScleraNET model, on the other hand, seems to generalise reasonably well to segmentation outputs with characteristics different from those produced by the cascaded SegNet assembly. It achieves the best performance with the NRLT and Coye segmentation techniques, it is comparable in performance to dSIFT on B-COSFIRE segmented vasculature and is second only to dSIFT with the NMC approach. This is surprising, as it was not trained on vascular images produced by these methods. Nonetheless, it seems to be able to extract useful descriptors for recognition from these images as well.

Overall, the results achieved with the proposed pipeline are very encouraging and present a good foundation for further research, also in the context of multi-modal biometric systems built around (peri-)ocular information.

13.6 Conclusion

We have presented a novel approach to sclera recognition built around convolutional neural networks. Our approach uses a two-step procedure that first locates the vascular structure of the sclera from the input image and then extracts a discriminative representation from the segmented vasculature that can be used for image comparisons and ultimately recognition. The two-step segmentation procedure is based on cascaded SegNet assembly, the first supervised approach to sclera vasculature segmentation presented in the literature, while the descriptor-computation procedure is based on a novel CNN-based model, called ScleraNET, trained in a multi-task manner. We evaluated our approach on a newly introduced and publicly available dataset of annotated sclera images and presented encouraging comparative results with competing methods. As part of our future work, we plan to integrate the presented pipeline with other ocular traits into a multi-modal recognition system.

Acknowledgements This research was supported in parts by ARRS (Slovenian Research Agency) Research Program P2-0250 (B) Metrology and Biometric Systems, ARRS Research Program P2-0214 (A) Computer Vision, and the RS-MIZŠ and EU-ESRR funded GOSTOP. The GPU used for this research was donated by the NVIDIA Corporation.

References

1. Alkassar S, Woo WL, Dlay SS, Chambers JA (2017) Robust sclera recognition system with novel sclera segmentation and validation techniques. *IEEE Trans Syst Man Cybern Syst (TSMC)* 47(3):474–486
2. Ali Z, Park U, Nang J, Park JS, Hong T, Park S (2017) Periocular recognition using uMLBP and attribute features. *KSII Trans Internet Inf Syst (TIIS)* 11(12):6133–6151
3. Jain AK, Ross A, Prabhakar S et al (2004) An introduction to biometric recognition. *IEEE Trans Circuits Syst Video Technol* 14(1)
4. Patil V, Patil AM (2017) Human identification method: sclera recognition. *Int J Comput Sci Netw (IJCSN)* 6(1):24–29
5. Das A, Pal U, Ferrer MA, Blumenstein M, Štepec D, Rot P, Emeršič Ž, Peer P, Štruc V, Kumar SA et al (2017) SSERBC 2017: sclera segmentation and eye recognition benchmarking competition. In: *International joint conference on biometrics (IJCB)*, pp 742–747
6. Derakhshani R, Ross A (2007) A texture-based neural network classifier for biometric identification using ocular surface vasculature. In: *International joint conference on neural networks 2007 (IJCNN 2007)*, pp 2982–2987. *IEEE*
7. Zhou Z, Du EY, Thomas NL, Delp EJ (2012) A new human identification method: sclera recognition. *IEEE Trans Syst Man Cybern (TSMC) Part A Syst Hum* 42(3):571–583
8. Das A, Pal U, Ferrer Ballester MA, Blumenstein M (2013) Sclera recognition using dense-sift. In: *2013 13th international conference on intelligent systems design and applications (ISDA)*, pp 74–79. *IEEE*
9. Zhou Z, Du EY, Thomas NL, Delp EJ (2011) Multi-angle sclera recognition system. In: *2011 IEEE workshop on computational intelligence in biometrics and identity management (CIBIM)*, pp 103–108. *IEEE*
10. Das A, Pal U, Ferrer MA, Blumenstein M, Štepec D, Rot P, Peer P, Štruc V (2018) SSBC 2018: sclera segmentation benchmarking competition. In: *International conference on biometrics (ICB)*, pp 303–308
11. Emeršič Ž, Štepec D, Štruc V, Peer P (2017) Training convolutional neural networks with limited training data for ear recognition in the wild. [arXiv:1711.09952](https://arxiv.org/abs/1711.09952)
12. Emeršič Ž, Meden B, Peer P, Štruc V (2018) Evaluation and analysis of ear recognition models: performance, complexity and resource requirements. *Neural computing and applications*, pp 1–16
13. Emeršič Ž, Štepec D, Štruc V, Peer P, George A, Ahmad A, Omar E, Boulton TE, Safdaii R, Zhou Y et al (2017) The unconstrained ear recognition challenge. In: *2017 IEEE international joint conference on biometrics (IJCB)*, pp 715–724. *IEEE*
14. Grm K, Štruc V, Artiges A, Caron M, Ekenel HK (2017) Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biom* 7(1):81–89
15. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 39(12):2481–2495
16. Nigam I, Vatsa M, Singh R (2015) Ocular biometrics: a survey of modalities and fusion approaches. *Inf Fusion* 26:1–35
17. De Marsico M, Petrosino A, Ricciardi S (2016) Iris recognition through machine learning techniques: a survey. *PRL* 82:106–115

18. Nguyen K, Fookes C, Jillela R, Sridharan S, Ross A (2017) Long range iris recognition: a survey. *PR* 72:123–143
19. Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 15(11):1148–1161
20. Daugman J (2009) How iris recognition works. In: *The essential guide to image processing*, pp 715–739. Elsevier
21. Daugman J (2007) New methods in iris recognition. *IEEE Trans Syst Man Cybern (TSMC) Part B* 37(5):1167–1175
22. Monro DM, Rakshit S, Zhang D (2007) DCT-based iris recognition. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 29(4):586–595
23. Miyazawa K, Ito K, Aoki T, Kobayashi K, Nakajima H (2008) An effective approach for iris recognition using phase-based image matching. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 30(10):1741–1756
24. Sun Z, Tan T (2009) Ordinal measures for iris recognition. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 31(12):2211–2226
25. Sun Z, Zhang H, Tan T, Wang J (2014) Iris image classification based on hierarchical visual codebook. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 36(6):1120–1133
26. Gangwar A, Joshi A (2016) Deepirisnet: deep iris representation with applications in iris recognition and cross-sensor iris recognition. In: *2016 IEEE international conference on image processing (ICIP)*, pp 2301–2305, Sept 2016
27. Liu N, Zhang M, Li H, Sun Z, Tan T (2016) Deepiris: learning pairwise filter bank for heterogeneous iris verification. *Pattern Recogn Lett* 82:154–161
28. Xingqiang T, Jiangtao X, Peihu L (2017) Deep convolutional features for iris recognition. In: *Proceedings of the 12th Chinese conference biometric recognition CCBR 2017*, pp 391–400, Cham, 2017. Springer International Publishing
29. Zhao Z, Kumar A (2017) Towards more accurate iris recognition using deeply learned spatially corresponding features. In: *International conference on computer vision, ICCV 2017*, pp 1–10
30. Nguyen K, Fookes C, Ross A, Sridharan S (2018) Iris recognition with off-the-shelf CNN features: a deep learning perspective. *IEEE Access* 6:18848–18855
31. Park U, Ross A, Jain AK (2009) Periocular biometrics in the visible spectrum: a feasibility study. In: *2009 IEEE 3rd international conference on biometrics: theory, applications, and systems*, pp 1–6. IEEE
32. Uzair M, Mahmood A, Mian A, McDonald C (2013) Periocular biometric recognition using image sets. In: *IEEE winter conference on applications of computer vision (WACV)*, pp 246–251
33. Sequeira A, Chen L, Ferryman J, Wild P, Alonso-Fernandez F, Bigun J, Raja K, Raghavendra R, Busch C, Freitas Pereira T et al (2017) Cross-eyed 2017: cross-spectral iris/periocular recognition database and competition. In: *IEEE international joint conference on biometrics (IJCB)*
34. Proença H, Neves J (2018) Deep-prwis: periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Trans Inf Forensics Secur (TIFS)* 13(4):888–896
35. Zhao Z, Kumar A (2017) Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Trans Inf Forensics Secur (TIFS)* 12(5):1017–1030
36. Alonso-Fernandez F, Bigun J (2015) Near-infrared and visible-light periocular recognition with gabor features using frequency-adaptive automatic eye detection. *IET Biom* 4(2):74–89
37. Das A, Mondal P, Pal U, Blumenstein M, Ferrer M (2017) Sclera vessel pattern synthesis based on a non-parametric texture synthesis technique. In: *Proceedings of computer vision and image processing (CVIP)*, pp 241–250. Springer
38. Alkassar S, Woo WL, Dlay S, Chambers J (2016) Sclera recognition: on the quality measure and segmentation of degraded images captured under relaxed imaging conditions. *IET Biom* 6(4):266–275
39. Yadav D, Kohli N, Doyle J, Singh R, Vatsa M, Bowyer K (2014) Unraveling the effect of textured contact lenses on iris recognition. *IEEE Trans Inf Forensics Secur (TIFS)* 9(5):851–862

40. Raja KB, Raghavendra R, Vemuri VK, Busch C (2015) Smartphone based visible iris recognition using deep sparse filtering. *PRL* 57:33–42
41. Das A, Pal U, Ferrer MA, Blumenstein M (2015) SSBC 2015: sclera segmentation benchmarking competition. In: International conference on biometrics: theory, applications, and systems (BTAS), pp 1–6
42. Das A, Pal U, Ferrer MA, Blumenstein M (2013) SSRBC 2016: sclera segmentation and recognition benchmarking competition. In: International conference on biometrics (ICB), pp 1–6
43. Maxwell EG, Tripti C (2013) A comparison between contrast limited adaptive histogram equalization and gabor filter sclera blood vessel enhancement techniques. *Int J Soft Comput Eng (IJSCE)* 3(4):22–5
44. Tankasala SP, Doynov P, Derakhshani RR, Ross A, Crihalmeanu S (2011) Biometric recognition of conjunctival vasculature using GLCM features. In: 2011 international conference on image information processing (ICIIP), pp 1–6, Nov 2011
45. Das A, Pal U, Ferrer Ballester MA, Blumenstein M (2014) Fuzzy logic based sclera recognition. In: 2014 IEEE international conference on fuzzy systems, pp 561–568
46. Das A, Pal U, Ferrer Ballester MA, Blumenstein M (2013) A new method for sclera vessel recognition using OLBP. In: Sun Z, Shan S, Yang G, Zhou J, Wang Y, Yin Y (eds) *Biometric Recognition*. Springer International Publishing, Cham, pp 370–377
47. CASIA Iris Image Database. <http://biometrics.idealtest.org/>
48. Proença H, Alexandre L (2005) Ubiris: a noisy iris image database. In: International conference on image analysis and processing (ICIAP), pp 970–977. Springer
49. Bowyer KW, Flynn PJ (2016) The ND-IRIS-0405 iris image dataset. Technical report, Notre Dame University
50. Hosseini MS, Araabi BN, Soltanian-Zadeh H (2010) Pigment melanin: pattern for iris recognition. *IEEE Trans Instrum Meas* 59(4):792–804
51. Kumar A, Passi A (2010) Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recogn* 43(3):1016–1026
52. Proença H, Filipe S, Santos R, Oliveira J, Alexandre LA (2010) The UBIRIS.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 32(8):1529–1535
53. De Marsico M, Nappi M, Riccio D, Wechsler H (2015) Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recogn Letters* 57:17–23
54. Padole CN, Proença H (2012) Periocular recognition: analysis of performance degradation factors. In: 5th IAPR international conference on biometrics (ICB), pp 439–445
55. Sharma A, Verma S, Vatsa M, Singh R (2014) On cross spectral periocular recognition. In: IEEE international conference on image processing (ICIP), pp 5007–5011. IEEE
56. Hosseini MS, Araabi BN, Soltanian-Zadeh H (2010) Pigment melanin: pattern for iris recognition. *IEEE Trans Instrum Meas (TIM)* 59(4):792–804
57. Rot P, Štruc V, Peer P (2018) Deep multi-class eye segmentation for ocular biometrics. In: IEEE international work conference on bioinspired intelligence (IWOBi)
58. Emeršič Ž, Gabriel L, Štruc V, Peer P (2018) Convolutional encoder-decoder networks for pixel-wise ear detection and segmentation. *IET Biom* 7(3):175–184
59. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition
60. Kendall A, Gal Y, Cipolla R (2018) Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7482–7491
61. Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. In: Conference on neural information processing systems (NIPS), pp 1097–1105
62. Yin X, Liu X (2017) Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans Image Process (TIP)*

63. Lozej J, Meden B, Štruc V, Peer P (2018) End-to-end iris segmentation using u-net. In: IEEE international work conference on bioinspired intelligence (IWOBI)
64. Križaj J, Štruc V, Dobrišek S (2013) Combining 3d face representations using region covariance descriptors and statistical models. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp 1–7. IEEE
65. Štruc V (2012) The PhD face recognition toolbox: toolbox description and user manual. Faculty of Electrical Engineering Ljubljana
66. Vesnicer B, Žganec Gros J, Pavešić N, Štruc V (2012) Face recognition using simplified probabilistic linear discriminant analysis. *Int J Adv Robot Syst (IJARS)* 9(5):180
67. Glorot X, Bengio Y (2012) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics (AISTATS), pp 249–256
68. Štruc V, Pavešić N (2009) Phase congruency features for palm-print verification. *IET Signal Process* 3(4):258–268
69. Savič T, Pavešić N (2007) Personal recognition based on an image of the palmar surface of the hand. *Pattern Recogn* 40(11):3152–3163
70. Lin G, Milan A, Shen C, Reid I (2017) Refinenet: multi-path refinement networks for high-resolution semantic segmentation. In: IEEE conference on computer vision and pattern recognition (CVPR)
71. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X (2017) Residual attention network for image classification. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 6450–6458
72. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: 18th international conference, proceedings, part III, medical image computing and computer-assisted intervention (MICCAI), pp 234–241. Springer International Publishing, Cham
73. Riccio D, Brancati N, Frucci M, Gragnaniello D (2017) An unsupervised approach for eye sclera segmentation. In: Iberoamerican congress on pattern recognition (CIARP), pp 550–557. Springer
74. Gonzalez RC, Woods RE (2007) Image processing. *Digit Image Process* 2
75. Miura N, Nagasaka A, Miyatake T (2007) Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE Trans Inf Syst* 90(8):1185–1194
76. Miura N, Nagasaka A, Miyatake T (2004) Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach Vis Appl* 15(4):194–203
77. Coye T (2015) A novel retinal blood vessel segmentation algorithm for fundus images. <http://www.mathworks.com/matlabcentral/fileexchange/50839>. matlab central file exchange. Accessed 4 Mar 2019
78. Azzopardi George, Strisciuglio Nicola, Vento Mario, Petkov Nicolai (2015) Trainable cosfire filters for vessel delineation with application to retinal images. *Med Image Anal* 19(1):46–57
79. Strisciuglio Nicola, Azzopardi George, Vento Mario, Petkov Nicolai (2016) Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters. *Mach Vis Appl* 27(8):1137–1149
80. Lowe DG (1999) Object recognition from local scale-invariant features. *Int J Comput Vis (ICCV)* 2:1150–1157. IEEE
81. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: European conference on computer vision, pp 404–417. Springer
82. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to sift or surf

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

