ASSIGNMENT FLOW FOR ORDER-CONSTRAINED OCT SEGMENTATION

DMITRIJ SITENKO, BASTIAN BOLL, CHRISTOPH SCHNÖRR

ABSTRACT. At the present time Optical Coherence Tomography (OCT) is among the most commonly used non-invasive imaging methods for the acquisition of large volumetric scans of human retinal tissues and vasculature. The substantial increase of accessible highly resolved 3D samples at the optic nerve head and the macula is directly linked to medical advancements in early detection of eye diseases. To resolve decisive information from extracted OCT volumes and to make it applicable for further diagnostic analysis, the exact identification of retinal layer thicknesses serves as an essential task be done for each patient separately. However, the manual examination of multiple OCT scans in a row is a demanding and time consuming task, which results in a lengthy qualification process and is frequently confounded in the presence of tissue-dependent speckle noise. Therefore, the elaboration of automated segmentation models has become an important task in the field of medical image processing.

We propose a novel, purely data driven *geometric approach to order-constrained 3D OCT retinal cell layer segmentation* which takes as input data in any metric space and comes along with basic operations that can be effectively computed in parallel. As opposed to many established retina detection methods, our presented formulation avoids the use of any shape prior and accomplishes the natural order of the retina in a purely geometric way, while maintaining the high level of accuracy. This makes the approach unbiased and hence suited for the detection of local anatomical changes of retinal tissue structure. To demonstrate robustness of the proposed approach, we compare two different choices of features on a data set of manually annotated 3D OCT volumes of healthy human retina. The quality of computed segmentations is compared to the state of the art in terms of mean absolute error and the Dice similarity coefficient. Visualizations of segmented volumes are also provided. The results indicate a great potential for applying our method to the classification of diseased retina and opens a new research direction regarding the joint segmentation of retinal cell layers and blood vessel structures.

CONTENTS

1. INTRODUCTION

1.1. **Overview, Motivation.** Optical Coherence Tomography (OCT) is a non-invasive imaging technique which measures the intensity response of back scattered light from millimeter penetration depth and provides information about retinal tissue structure in vivo to understand human eye functionalities, see Figure 1.1 for a more descriptive anatomical explanation. OCT devices record multiple two-dimensional B-scans in rapid succession and combine them to a single volume in a subsequent alignment step. We focus specifically on the application of OCT in ophthalmology for the aquisition of high-resolution volume scans of the human retina. Taking an OCT scan only takes minutes and can help detect symptoms of pathological conditions such as glaucoma, diabetes, multiple sclerosis or age-related macular degeneration. The relative ease of data acquisition also enables to use multiple OCT volume scans of a single patient over time to track the progression of a pathology or quantify the success of therapeutic treatment. As a consequence of the technological progress in OCT imaging which was made over past few decades since its invention in 1991 [HSL$^+$91], more expertise for extraction of manual annotations is required which in the presence of big volumetric data sets is difficult to access.

To better leverage the availability of retinal OCT data in both clinical settings and empirical studies, much work is focused on the analysis of appropriate automatic feature extraction techniques. In particular, the access to such methods is especially crucial for achieving enhanced effectiveness of existing quantitative retinal multi cell layer segmentation approaches, and for increasing their clinical potential in real life applications, such as detection of fluid regions and reconstruction of vascular structures. The difficulty of these tasks lies in the challenging signal-to-noise ratio which is influenced by multiple factors including mechanical eye movement during registration and the presence of speckle noise.

In this paper, we extend the approach [ÅPSS17] for labeling data on graphs to automatic cell layer segmentation in OCT data. After a feature extraction step, each voxel is labeled by smoothing local layer
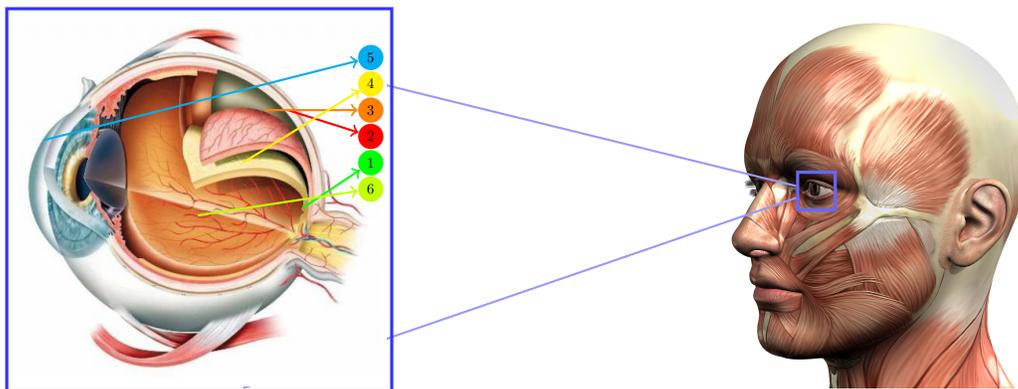
FIGURE 1.1. Schematic illustration designed by [Kjp] of human eye functionality: The light enters the Cornea ⑤ though the vitreous humour ⑥ towards retina ④ and choroid ③ which are located around the fovea ①.

decisions and jointly leveraging a global geometric invariant – the natural order of cell layers along the vertical axis of each B-scan, as shown in the third row of Figure 1.2. We are able to produce high-quality segmentations of OCT volumes by using *local* features as input for a purpose-built assignment flow variant which serves to incorporate global context in a controlled way. This is in contrast to common machine learning approaches which use essentially full B-scans as input. By incorporating global context into the feature extraction process, the latter methods are at increased risk of overfitting training data and potentially missinterpreting unseen pathologies.

Our segmentation approach is a smooth image labeling algorithm based on geometric numerical integration on an elementary statistical manifold. It can work with input data from any metric space, making it agnostic to the choice of feature extraction and suitable as plug-in replacement in diverse pipelines. In addition to respecting the natural order of cell layers, the proposed segmentation process has a high amount of built-in parallelism such that modern graphics acceleration hardware can easily be leveraged. We compare the effectiveness of our novel approach between a selection of input features ranging from traditional covariance descriptors to convolutional neural networks. Figure 1.2 shows one specific example of an segmented OCT-volume with clearly visible difficulties of speckle noise and jagged patterns followed by an labeled (2D) cutout which is a typical result our novel approach.

1.2. **Related Work.** Effective segmentation of OCT volumes is a very active area of research. Here, we briefly review the current state of the art approaches originating from the broad research fields of graphical models, variational methods and machine learning.

1.2.1. *Graphical Models.* The first mathematical access to the problem is provided by the theory of graphical models which transforms the segmentation task into an optimization problem with hard pairwise interaction constraints between voxels. Starting with Li et. al. [KXCS06] and Haeker [HAW+07], simultaneous retina layer detection attempts were made by finding an s-t minimum graph cut. Garvin et. al. [GAW+09] further extended this approach with a shape prior modeling layer boundaries. The methods benefit from low computational complexity, but are lacking of robustness in the presence of speckle and therefore require additional preprocessing steps. Along this line of reasoning, B.J. Anthony et. al [AAL+10] used a two stage segmentation process by applying anisotropic diffusion in a preprocessing step and consequently segmenting outer retina layers using graphical models. Similarly, Kafieh et. al. [KRAS13] proposed to use specific distances based on diffusion maps which are computed by coarse graining the original graph. However, increased performance for noisy OCT data gained by regularizing in this way comes at the cost of introducing bias in the preprocessing step which in turn inpairs robustness in settings with medical pathologies.

Motivated by [SBG$^+$13], Dufour et. al. [DCA$^+$13] comes up with a circular shape prior for segmentation of 6 retinal layers by incorporating soft constraints which are more suitable for the robust detection of pathological retina structures. Chiu. et al. [CAM$^+$15] relies on a graphical model approach as a post-processing step after applying a supervised kernel regression classification with features extracted according to [QLD$^+$10]. Rathke et. al. [RSS14] reduced the overall complexity by a parallelizable segmentation approach based on probabilistic graphical models with global low-rank shape prior representing interacting retina tissues surfaces. While the *global* shape prior works well for non-pathological OCT data, it cannot be adapted to the broad range of variations caused by *local pathological* structure resulting in a inherent limitation of this approach.
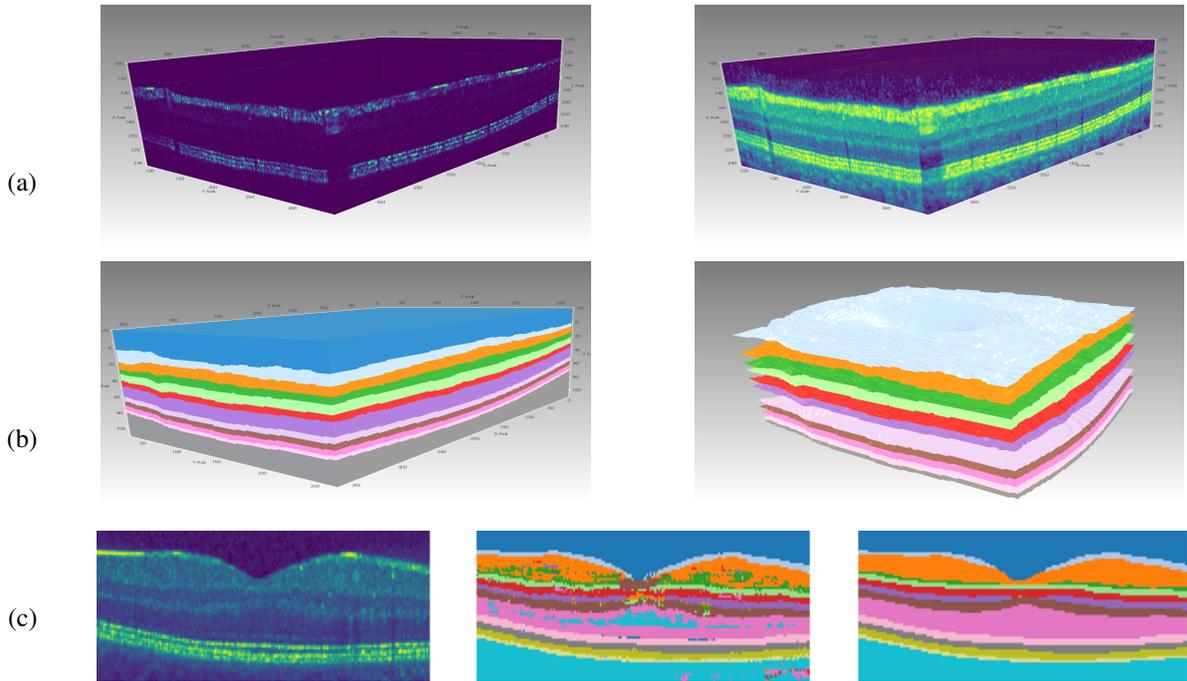

FIGURE 1.2. **(a)**: 3D OCT volume scan dimension $512 \times 512 \times 256$ of healthy human retina with ambiguous locations of layer boundaries with normalized view on the right. **(b)**: The resulting segmentation of 11 layers displaying the order preserving labeling of the proposed approach. Boundary surfaces between different segmented cell layers are illustrated. **(c)**: Typical result of the proposed segmentation approach for a single B-scan of healthy retina. *Left*: raw OCT input data. *Middle*: segmentation by locally selecting the label with maximum score for each voxel after feature extraction. *Right*: segmentation by the proposed assignment flow approach using the same extracted features.

1.2.2. *Variational Methods.* Another category of layer detection methods focus on minimizing of energy functional to express the quantity of interest as the solution to an optimization problem. To this class of methods for retina detection level set approaches have proven to be particularly suitable by encoding each retina layer as the zero level sets of a certain functional. Yazdanpanah et. al. [YHSS11] introduces a level set method for minimizing an active contour functional supported by a multiphase Chan Vese model [CV01] as circular shape prior, to avoid limitations of hard constraints as opposed to graphical model proposed by [GAW$^+$09]. Duan et. al. [DTG$^+$15] suggests the approach to model layer boundaries with a mixture of Mumford Shah and Vese and Osher functionals by first preprocessing the data in the Fourier domain. A capable level set approach for joint segmentation of pathological retina tissues was reported in [NVd$^+$17]. However, due to the involved hierarchical optimization, their method is computationally expensive. One

common downside of the above algorithms are their inherent limitations to only include local notions of layer ordering, making their extension to cases with pathologically caused retina degeneracy a difficult task.

1.2.3. *Machine Learning.* Much recent work has focused on the use of deep learning to address the task of cell layer segmentation in a purely data driven way. The U-net architecture [RFB15] has proven influential in this domain because of its good predictive performance in settings with limited availability of training data. Multiple modifications of U-net have been proposed to specifically increase its performance in OCT applications [RCK+17, LCF+19]. The common methods largely rely on convolutional neural networks to predict layer segmentations for individual B-scans which are subsequently combined to full volumes. These methods have also been used as part of a two-stage pipeline where additional prior knowledge such as local regularity and global order of cell layers along a spatial axis is incorporated through graph-based methods [FCW+17] or a second machine learning component [HCL+19]. However, because global context is already used in feature extraction, the risk of overfitting remains and unseen pathologies may result in unpredictable behavior.

1.3. **Contribution, Organization.** We propose a geometric assignment approach to retinal layer segmentation. By leveraging a continuous characterization of layer ordering, our method is able to simultaneously perform local regularization and incorporate the global topological ordering constraint in a *single smooth* labeling process. The segmentation is computed from a distance matrix containing pairwise distances between data for each voxel and prototypical data for each layer in some feature space. This highlights the ability to extract features from raw OCT data in a variety of different ways and to use the proposed segmentation as a plug-in replacement for other graph-based methods.

As a result of the proposed method, it becomes possible to compute high-quality cell layer segmentations of OCT volumes by using only local features for each voxel. This is in contrast to competing deep learning approaches which commonly use information from an entire B-scan as input. In addition, the exclusive use of local features combats bias introduced through limited data availability in training and makes incorporation of three-dimensional information easily possible without limiting runtime scalability. To demonstrate this, we implement two feature extraction approaches. The first is based on identifying each datum with a covariance descriptor and finding prototypical descriptors as cluster centers. For each voxel, Riemannian distances to the prototypical descriptors are used as input for subsequent segmentation. The second is based on training a convolutional neural network to classify small voxel patches of raw OCT data. Predicted class scores for each voxel are subsequently used as input for the proposed segmentation method.

The final pipeline enables robust cell layer segmentation for raw OCT volumes at scale, labeling an entire OCT volume in the time frame between 30 seconds and several minutes on a single GPU and in general leads to increased performance in the case of more informative features. This is without using any prior knowledge other than local regularity and order of cell layers. In particular, no global shape prior is used thus making our proposed approach suited for retina detection in OCT volumes with observable pathological patterns.

Our paper considerably elaborates the conference version [SBS20] and is organized as follows. The assignment flow approach is summarized in Section 2 and extended in Section 4 in order to take into account the order of layers as a global constraint. In Section 3, we consider the Riemannian manifold $\mathcal{P}_d$ of positive definite matrices as a suitable feature space for local OCT data descriptors. Various Riemannian metrics are discussed with regard to computational efficiency of clustering. The resulting features are subsequently compared to local features extracted by a convolutional network in Section 5. Performance measures for OCT segmentation will be reported for our novel approach and for two other state-of-the-art methods with available standalone software, that were evaluated in detail as summarized in Section 5. In Section 6, we shortly discuss the access to appropriate ground truth data and the impact of feature locality underlying our approach.

2. ASSIGNMENT FLOW

We summarize the assignment flow approach introduced by [ÅPSS17] and refer to the recent survey [Sch20] for more background and a review of recent related work.

2.1. Assignment Manifold.

Let $(\mathcal{F}, d_{\mathcal{F}})$ be a metric space and

$$\mathcal{F}_n = \{f_i \in \mathcal{F} \colon i \in \mathcal{I}\}, \qquad |\mathcal{I}| = n \tag{2.1a}$$

given data. Assume that a predefined set of prototypes

$$\mathcal{F}_* = \{f_j^* \in \mathcal{F} \colon j \in \mathcal{J}\}, \qquad |\mathcal{J}| = c \tag{2.1b}$$

is given. *Data labeling* denotes the assignments

$$j \to i, \qquad f_j^* \to f_i \tag{2.2}$$

of a single prototype $f_j^* \in \mathcal{F}_*$ to each data point $f_i \in \mathcal{F}_n$. The set $\mathcal{I}$ is assumed to form the vertex set of an undirected graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ which defines a relation $\mathcal{E} \subset \mathcal{I} \times \mathcal{I}$ and neighborhoods

$$\mathcal{N}_i = \{k \in \mathcal{I} \colon ik \in \mathcal{E}\} \cup \{i\}, \tag{2.3}$$

where ik is a shorthand for the unordered pair (edge) $(i, k) = (k, i)$. We require these neighborhoods to satisfy the symmetry relation

$$k \in \mathcal{N}_i \quad \Leftrightarrow \quad i \in \mathcal{N}_k, \qquad \forall i, k \in \mathcal{I}. \tag{2.4}$$

The assignments (labeling) (2.2) are represented by matrices in the set

$$\mathcal{W}_* = \{W \in \{0, 1\}^{n \times c} \colon W \mathbb{1}_c = \mathbb{1}_n\} \tag{2.5}$$

with unit vectors W_i, $i \in \mathcal{I}$, called *assignment vectors*, as row vectors. These assignment vectors are computed by numerically integrating the assignment flow below (2.24) in the following geometric setting. The integrality constraint of (2.5) is relaxed and vectors

$$W_i = (W_{i1}, \dots, W_{ic})^\top \in \mathcal{S}, \quad i \in \mathcal{I}, \tag{2.6}$$

that we still call *assignment vectors*, are considered on the elementary Riemannian manifold

$$(\mathcal{S}, g), \qquad \mathcal{S} = \{p \in \Delta_c \colon p > 0\} \tag{2.7}$$

with the probability simplex $\Delta_c = \left\{p \in \mathbb{R}_+^c \colon \sum_{i=1}^c = \langle \mathbb{1}, p \rangle = 1\right\}$, the barycenter

$$\mathbb{1}_{\mathcal{S}} = \frac{1}{c} \mathbb{1}_c \in \mathcal{S}, \qquad \text{(barycenter)} \tag{2.8}$$

tangent space

$$T_0 = \{v \in \mathbb{R}^c \colon \langle \mathbb{1}_c, v \rangle = 0\} \tag{2.9}$$

and tangent bundle $T\mathcal{S} = \mathcal{S} \times T_0$, the orthogonal projection

$$\Pi_0 \colon \mathbb{R}^c \to T_0, \qquad \Pi_0 = I - \mathbb{1}_{\mathcal{S}} \mathbb{1}^\top \tag{2.10}$$

and the Fisher-Rao metric

$$g_p(u, v) = \sum_{j \in \mathcal{J}} \frac{u^j v^j}{p^j}, \quad p \in \mathcal{S}, \quad u, v \in T_0. \tag{2.11}$$

Based on the linear map

$$R_p \colon \mathbb{R}^c \to T_0, \qquad R_p = \mathrm{Diag}(p) - pp^\top, \qquad p \in \mathcal{S} \tag{2.12}$$

that satisfies

$$R_p = R_p \Pi_0 = \Pi_0 R_p, \tag{2.13}$$

exponential maps and their inverses are defined by

$$\mathrm{Exp}\colon \mathcal{S} \times T_0 \to \mathcal{S}, \qquad\qquad (p, v) \mapsto \mathrm{Exp}_p(v) = \frac{pe^{\frac{v}{p}}}{\langle p, e^{\frac{v}{p}} \rangle}, \qquad (2.14\mathrm{a})$$

$$\mathrm{Exp}_p^{-1}\colon \mathcal{S} \to T_0, \qquad\qquad q \mapsto \mathrm{Exp}_p^{-1}(q) = R_p \log \frac{q}{p}, \qquad (2.14\mathrm{b})$$

$$\exp_p\colon T_0 \to \mathcal{S}, \qquad\qquad \exp_p = \mathrm{Exp}_p \circ R_p, \qquad (2.14\mathrm{c})$$

$$\exp_p^{-1}\colon \mathcal{S} \to T_0, \qquad\qquad \exp_p^{-1}(q) = \Pi_0 \log \frac{q}{p}. \qquad (2.14\mathrm{d})$$

Applying the map $\exp_p$ to a vector in $\mathbb{R}^c = T_0 \oplus \mathbb{R}\mathbb{1}$ does not depend on the constant component of the argument, due to (2.13).

Remark 2.1. The map Exp corresponds to the e-connection of information geometry [AN00], rather than to the exponential map of the Riemannian connection. Accordingly, the affine geodesics (2.14a) are not length-minimizing. But they provide a close approximation [ÅPSS17, Prop. 3] and are more convenient for numerical computations.

The *assignment manifold* is defined as

$$(\mathcal{W}, g), \qquad \mathcal{W} = \mathcal{S} \times \cdots \times \mathcal{S}. \qquad (n = |\mathcal{I}| \text{ factors}) \qquad (2.15)$$

We identify $\mathcal{W}$ with the embedding into $\mathbb{R}^{n \times c}$

$$\mathcal{W} = \{W \in \mathbb{R}^{n \times c}\colon W\mathbb{1}_c = \mathbb{1}_n \text{ and } W_{ij} > 0 \text{ for all } i \in [n], j \in [c]\}. \qquad (2.16)$$

Thus, points $W \in \mathcal{W}$ are row-stochastic matrices $W \in \mathbb{R}^{n \times c}$ with row vectors $W_i \in \mathcal{S}$, $i \in \mathcal{I}$ that represent the assignments (2.2) for every $i \in \mathcal{I}$. We set

$$\mathcal{T}_0 := T_0 \times \cdots \times T_0 \qquad (n = |\mathcal{I}| \text{ factors}). \qquad (2.17)$$

Due to (2.16), the tangent space $\mathcal{T}_0$ can be identified with

$$\mathcal{T}_0 = \{V \in \mathbb{R}^{n \times c}\colon V\mathbb{1}_c = 0\}. \qquad (2.18)$$

Thus, $V_i \in T_0$ for all row vectors of $V \in \mathbb{R}^{n \times c}$ and $i \in \mathcal{I}$. All mappings defined above factorize in a natural way and apply row-wise, e.g. $\mathrm{Exp}_W = (\mathrm{Exp}_{W_1}, \ldots, \mathrm{Exp}_{W_n})$ etc.

2.2. **Assignment Flow.** Based on (2.1a) and (2.1b), the distance vector field

$$D_{\mathcal{F};i} = \big(d_{\mathcal{F}}(f_i, f_1^*), \ldots, d_{\mathcal{F}}(f_i, f_c^*)\big)^\top, \qquad i \in \mathcal{I} \qquad (2.19)$$

is well-defined. These vectors are collected as row vectors of the *distance matrix*

$$D_{\mathcal{F}} \in S_+^n, \qquad (2.20)$$

where S_+^n denotes the set of symmetric and entrywise nonnegative matrices.

Remark 2.2. In this paper, we build upon two different types of features to determine vectors (2.19) which are serving as input before mapping the assembled matrix (2.20) onto the assignment manifold as explained below. Hereby, the first class of features access our model by calculating distance to prototypes (2.1) with metric introduced in section (3.2) while the second feature class directly possess the form of (2.20) as argued in section (5.2.3).

The *likelihood map* and the *likelihood vectors*, respectively, are defined as

$$L_i\colon \mathcal{S} \to \mathcal{S}, \qquad L_i(W_i) = \exp_{W_i}\Big(-\frac{1}{\rho}D_{\mathcal{F};i}\Big) = \frac{W_i e^{-\frac{1}{\rho}D_{\mathcal{F};i}}}{\langle W_i, e^{-\frac{1}{\rho}D_{\mathcal{F};i}}\rangle}, \qquad i \in \mathcal{I}, \qquad (2.21)$$

where the scaling parameter $\rho > 0$ is used for normalizing the a-prior unknown scale of the components of $D_{\mathcal{F};i}$ that depends on the specific application at hand.

A key component of the assignment flow is the interaction of the likelihood vectors through *geometric* averaging within the local neighborhoods (2.3). Specifically, using weights

$$\omega_{ik} > 0 \quad \text{for all } k \in \mathcal{N}_i,\ i \in \mathcal{I} \quad \text{with} \quad \sum_{k \in \mathcal{N}_i} w_{ik} = 1, \tag{2.22}$$

the *similarity map* and the *similarity vectors*, respectively, are defined as

$$S_i \colon \mathcal{W} \to \mathcal{S}, \qquad S_i(W) = \mathrm{Exp}_{W_i} \Big(\sum_{k \in \mathcal{N}_i} w_{ik} \, \mathrm{Exp}_{W_i}^{-1} \big(L_k(W_k) \big) \Big), \qquad i \in \mathcal{I}. \tag{2.23}$$

If Exp_{W_i} were the exponential map of the Riemannian (Levi-Civita) connection, then the argument inside the brackets of the right-hand side would just be the negative Riemannian gradient with respect to W_i of center of mass objective function comprising the points L_k, $k \in \mathcal{N}_i$, i.e. the weighted sum of the squared Riemannian distances between W_i and L_k [Jos17, Lemma 6.9.4]. In view of Remark 2.1, this interpretation is only approximately true mathematically, but still correct informally: $S_i(W)$ moves W_i towards the geometric mean of the likelihood vectors L_k, $k \in \mathcal{N}_i$. Since $\mathrm{Exp}_{W_i}(0) = W_i$, this mean precisely is W_i if the aforementioned gradient vanishes.

The *assignment flow* is induced on the assignment manifold $\mathcal{W}$ by the locally coupled system of nonlinear ODEs

$$\dot{W} = R_W S(W), \qquad W(0) = \mathbb{1}_{\mathcal{W}}, \tag{2.24a}$$

$$\dot{W}_i = R_{W_i} S_i(W), \qquad W_i(0) = \mathbb{1}_{\mathcal{S}}, \quad i \in \mathcal{I}, \tag{2.24b}$$

where $\mathbb{1}_{\mathcal{W}} \in \mathcal{W}$ denotes the barycenter of the assignment manifold (2.15). The solution $W(t) \in \mathcal{W}$ is numerically computed by geometric integration [ZSPS20] and determines a labeling $W(T) \in \mathcal{W}_*$ for sufficiently large T after a trivial rounding operation. Convergence and stability of the assignment flow have been studied by [ZZS20].

3. OCT DATA REPRESENTATION BY COVARIANCE DESCIPTORS

In this section, we work out the basic geometric notation for representation of OCT data by means of covariance descriptors [TPM06]. Specifically, the metric data space $(\mathcal{F}, d_{\mathcal{F}})$ underlying (2.1) will be identified with the Riemannian manifold $(\mathcal{P}_d, d_g)$ of positive definite matrices of dimension $d \times d$, with Riemannian metric g and Riemannian distance d_g as specified in section 5. In particular regarding the computation of corresponding prototypes (2.1b), an important aspect concerns the trade-off between respecting the Riemannian distance d_g of the matrix manifold $\mathcal{P}_d$ and approximating surrogate distance functions, that enable to compute more efficiently Riemannian means of covariance descriptors while adopting their natural geometry. We review and discuss various choices in Section 3.2 after reviewing few required concepts of Riemannian geometry in Section 3.1.

3.1. **The Manifold** $\mathcal{P}_d$. We collect few concepts related to data $p \in \mathcal{M}$ taking values on a general Riemannian manifold $(\mathcal{M}, g)$ with Riemannian metric g; see, e.g., [Lee13, Jos17] for background reading. Then we apply these concepts to the specific manifold $(\mathcal{P}_d, g)$ and the corresponding distance d_g, keeping the symbol g for the metric for simplicity. We refer to, e.g., [Bha07, Bha13, PFAE06, MB06] for further reading and to the references in Section 3.2.

Let $\gamma \colon [0, 1] \to \mathcal{M}$ a smooth curve connecting two points $p = \gamma(0)$ and $q = \gamma(1)$. The *Riemannian distance* between p and q is given by

$$d_g(p, q) = \min_{\gamma \colon \gamma(0) = p, \gamma(1) = q} L(\gamma) \tag{3.1a}$$

with

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)} \, dt = \int_0^1 \sqrt{g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big)} \, dt \,. \tag{3.1b}$$

Assume the minimum of the right-hand side of (3.1a) is attained at $\overline{\gamma}$. Then the *exponential map* at p is defined on some neighborhood $V_p \subseteq T_p\mathcal{M}$ of 0 in the tangent space to $\mathcal{M}$ at p by

$$\exp_p \colon V_p \supseteq T_p\mathcal{M} \;\to\; U_p \subseteq \mathcal{M}, \qquad v \mapsto \exp_p(v) := \overline{\gamma}(1). \tag{3.2}$$

This mapping is a diffeomorphism of V_p and its inverse $\exp_p^{-1} \colon U_p \to V_p$ exists on a corresponding open neighborhood U_p. Let $\mathcal{X}(\mathcal{M})$ denote the set of all smooth vector fields on $\mathcal{M}$, i.e. $X \in \mathcal{X}(\mathcal{M})$ evaluates to a tangent vector $X_p \in T_p\mathcal{M}$ smoothly depending on p. The set of all smooth covector fields (one-forms) is denoted by $\mathcal{X}^*(\mathcal{M})$, and $df(X)$ denotes the action of the differential $df \in \mathcal{X}^*(\mathcal{M})$ of a smooth function $f \colon \mathcal{M} \to \mathbb{R}$ on a vector field X. The *Riemannian gradient* of f is the vector field $\operatorname{grad} f \in \mathcal{X}(\mathcal{M})$ defined by

$$g(\operatorname{grad} f, X) = df(X) = Xf, \qquad \forall X \in \mathcal{X}(\mathcal{M}). \tag{3.3}$$

We now focus on the following problem: Given a set of points $\{p_i\}_{i \in [N]} \subset \mathcal{M}$, compute the *weighted Riemannian mean* as minimizer of the objective function

$$\overline{p} = \arg\min_{q \in \mathcal{M}} J(q), \qquad J(q) = \sum_{i \in [N]} \omega_i d_g^2(q, p_i), \qquad \omega_i > 0, \ \forall i, \qquad \sum_{i \in [N]} \omega_i = 1. \tag{3.4}$$

The Riemannian gradient of this objective function is given by [Jos17, Lemma 6.9.4]

$$\operatorname{grad} J(p) = -\sum_{i \in [N]} \omega_i \exp_p^{-1}(p_i). \tag{3.5}$$

Hence the Riemannian mean $\overline{p}$ is determined by the optimality condition

$$\sum_{i \in [N]} \omega_i \exp_{\overline{p}}^{-1}(p_i) = 0. \tag{3.6}$$

A basic numerical method for computing $\overline{p}$ is the fixed point iteration

$$q_{(t+1)} = \exp_{q_{(t)}} \Big(\sum_{i \in [N]} \omega_i \exp_{q_{(t)}}^{-1}(p_i) \Big), \qquad t = 1, 2, \dots \tag{3.7}$$

that may converge for a suitable initialization $q_{(0)}$ to $\overline{p}$.

We now focus on the specific manifold $(\mathcal{P}_d, g)$

$$\mathcal{P}_d = \{S \in \mathbb{R}^{d \times d} \colon S = S^\top, \ S \text{ is positive definite}\} \tag{3.8}$$

equipped with the Riemannian metric

$$g_S(U, V) = \operatorname{tr}(S^{-1} U S^{-1} V), \qquad U, V \in T_S\mathcal{P}_d = \{S \in \mathbb{R}^{d \times d} \colon S^\top = S\}. \tag{3.9}$$

The Riemannian distance (3.1a) is given by

$$d_{\mathcal{P}_d}(S, T) = \Big(\sum_{i \in [d]} \big(\log \lambda_i(S, T) \big)^2 \Big)^{1/2}, \tag{3.10}$$

whereas the exponential map (3.2) reads

$$\exp_S(U) = S^{\frac{1}{2}} \operatorname{expm}(S^{-\frac{1}{2}} U S^{-\frac{1}{2}}) S^{\frac{1}{2}}, \tag{3.11}$$

and $\operatorname{expm}(\cdot)$ denotes the matrix exponential. Finally, given a smooth objective function $J \colon \mathcal{P}_d \to \mathbb{R}$, the Riemannian gradient is given by

$$\operatorname{grad} J(S) = S\big(\partial J(S)\big)S \in T_S\mathcal{P}_d, \tag{3.12}$$

where the symmetric matrix $\partial J(S)$ denotes the Euclidean gradient of J at S. Since $\mathcal{P}_d$ is a simply connected, complete and nonpositively curved Riemannian manifold [BH99, Section 10], the exponential map (3.11) is globally defined and bijective, and the Riemannian mean always exists and is uniquely defined as minimizer of the objective function (3.4), after substituting the Riemannian distance (3.10).

3.2. **Computing Prototypical Covariance Descriptors.** In this section, we focus on the computational differential geometric framework required for extraction of prototypes (2.1b) as Riemannian means from a set of covariance descriptors assembled from OCT data. Application details are reported in Section 5. Particularly with regard to more efficient handling present volumetric data and to reduce the computational costs, a surrogate metrics and distances are reviewed in Sections 3.2.2 and 3.2.3. Their qualitative comparison is reported in Section 5.

3.2.1. *Computing Riemannian Means.* Given a set of covariance descriptors

$$\mathcal{S}_N = \{(S_1, \omega_1), \ldots, (S_N, \omega_N)\} \subset \mathcal{P}_d \tag{3.13}$$

together with positive weights ω_i, we next focus on the solution of the problem (3.4) for specific geometry (3.8),

$$\overline{S} = \arg\min_{S \in \mathcal{P}_d} J(S; \mathcal{S}_N), \qquad J(S; \mathcal{S}_N) = \sum_{i \in [N]} \omega_i d_{\mathcal{P}_d}^2(S, S_i), \tag{3.14}$$

with the distance $d_{\mathcal{P}_d}$ given by (3.10). From (3.11), we deduce

$$U = \exp_S^{-1} \circ \exp_S(U) = S^{\frac{1}{2}} \operatorname{logm}\left(S^{-\frac{1}{2}} \exp_S(U) S^{-\frac{1}{2}}\right) S^{\frac{1}{2}} \tag{3.15}$$

with the matrix logarithm $\operatorname{logm} = \operatorname{expm}^{-1}$ [Hig08, Section 11]. As a result, optimality condition (3.6) reads

$$\sum_{i \in [N]} \omega_i \overline{S}^{\frac{1}{2}} \operatorname{logm}\left(\overline{S}^{-\frac{1}{2}} S_i \overline{S}^{-\frac{1}{2}}\right) \overline{S}^{\frac{1}{2}} = 0. \tag{3.16}$$

Applying the corresponding basic fixed iteration (3.7) has two drawbacks, however [CABM15]: Convergence is not theoretically guaranteed and if the iteration converges, than at a linear rate only. Since each iterative step requires nontrivial numerical matrix decomposition that has to be applied multiple times to every voxel (vertex) of a 3D gridgraph, this results in an overall quite expensive approach, in particular when larger data sets are involved as is the case for highly resolved 3D OCT volumetric scans.

The following variant proposed by [BI13] is guaranteed to converge at a *quadratic* rate assuming the matrices $\{S_1, \ldots, S_N\}$ to pairwise commute. Using the parametrization

$$S = LL^\top \tag{3.17}$$

corresponding to the Cholesky decomposition replacing the map of fixed point iteration (3.7) with its linearization leads to the following fixed point iteration

$$F_\tau(L; \mathcal{S}_N) = LL^\top - \tau \sum_{i \in [N]} \omega_i L^\top \operatorname{logm}(L^{-\top} S_i^{-1} L^{-1}) L, \qquad \tau > 0, \tag{3.18}$$

with damping parameter τ. Comparing to (3.16) shows that the basic idea is to compute the Riemannian mean $\overline{S}$ as fixed point of the iteration

$$\overline{S} = \lim_{t \to \infty} S_{(t)}, \qquad S_{(t+1)} = F(S_{(t)}; \mathcal{S}_N). \tag{3.19}$$

Algorithm 1 provides a refined variant of this iteration including adaptive stepsize selection. See [CABM15] for alternative algorithms that determine the Riemannian mean.

Algorithm 1: Fixed Point Iteration for Computing the Riemannian Matrix Mean.

Initialization

ϵ (termination threshold)

$t = 0, \quad S_{(0)} = LL^\top$, with $S_{(0)}$ solving (3.21).

$c_0 = \frac{\lambda_{\max}(S_{(0)})}{\lambda_{\min}(S_{(0)})}, \{\alpha_0, \beta_0\} = \left[\frac{\log(c_0)}{c_0-1}, c_0\frac{\log(c_0)}{c_0-1}\right]$ (condition number and step size selection parameters)

$\tau_0 = \frac{2}{\alpha_0+\beta_0}$

$S_{(1)} = F_\tau(L; \mathcal{S}_N)$ (iterative step)

$\epsilon_1 = \left\| \sum_{i\in[N]} \omega_i \operatorname{logm}(S_{(1)}^{\frac{1}{2}} S_i^{-1} S_{(1)}^{\frac{1}{2}}\right\|_F, \quad t = 1$

while $\epsilon_t > \epsilon$ **do**

    $S_{(t)} = LL^\top$

    $c_t = \frac{\lambda_{\max}(S_{(t)})}{\lambda_{\min}(S_{(t)})}$

    if $c_t = 1$ **then**

        stop

    $\{\alpha_t, \beta_t\} = \{\sum_{k=0}^t \frac{\log(c_k)}{c_k-1}, c_k\frac{\log(c_k)}{c_k-1}\}$

    $\tau_t = \frac{2}{\alpha_t+\beta_t}$

    $S_{(t+1)} = F_{\tau_t}(L; \mathcal{S}_N)$

    $\epsilon_{t+1} := \left\| \sum_{i\in[N]} \omega_i \operatorname{logm}(S_{(t+1)}^{\frac{1}{2}} S_i^{-1} S_{(t+1)}^{\frac{1}{2}})\right\|_F, \quad t \leftarrow t+1$

3.2.2. *Log-Euclidean Distance and Means.* A computationally cheap approach was proposed by [AFPA07] (among several other ones). Based on the operations

$$S_1 \odot S_2 = \operatorname{expm}\big(\operatorname{logm}(S_1 + \operatorname{logm}(S_2))\big), \tag{3.20a}$$

$$\lambda \cdot S = \operatorname{expm}\big(\lambda \operatorname{logm}(S)\big), \tag{3.20b}$$

the set $(\mathcal{P}_s, \odot, \cdot)$ becomes isomorphic to the vector space where $\odot$ plays the role of addition. Consequently, the mean of the data $\mathcal{S}_N$ given by (3.13) is defined analogous to the arithmetic mean by

$$\overline{S} = \operatorname{expm}\Big(\sum_{i\in[N]} \omega_i \operatorname{logm}(S_i)\Big). \tag{3.21}$$

While computing the mean is considerably cheaper than integrating the flow (3.12) using approximation Algorithm 1, the critical drawback of relying on (3.21) is not taking into account the (curved structure) of the manifold $\mathcal{P}_d$. Therefore, in the next section, we additionally consider another approximation of the Riemannian mean that better respects the underlying geometry but can still be evaluated more efficiently than the Riemannian mean of Section 3.2.1.

3.2.3. *S-Divergence and Means.* A general approach to the approximation of the objective function (3.4) is to replace the squared Riemannian $d_g^2(p, q)$ distance by a divergence function

$$D(p, q) \approx \frac{1}{2}d_g^2(p, q) \tag{3.22}$$

that satisfies

$$D(p, q) \geq 0 \quad \text{and} \quad D(p, q) = 0 \Leftrightarrow p = q, \tag{3.23a}$$

$$\partial_1^2 D(p, q) \succ 0, \quad \forall p \in \operatorname{dom} D(\cdot, q). \tag{3.23b}$$

We refer to, e.g., [CZ97, BB97] for a complete definition. Property (3.23b) says that, for any feasible p, the Hessian with respect to the first argument is positive definite. In fact, suitable divergence functions D recover in this way locally the metric tensor of the underlying manifold $\mathcal{M}$, in order to qualify as a surrogate for the squared Riemannian distance (3.22).

For the present case $\mathcal{M} = \mathcal{P}_d$ of interest, Sra [Sra16] proposed the divergence function, called *Stein divergence*

$$D_s(S_1, S_2) = \log \det \left(\frac{S_1 + S_2}{2} \right) - \frac{1}{2} \log \det(S_1 S_2), \qquad S, S_1, S_2 \in \mathcal{P}_d. \tag{3.24}$$

Regarding the task of evaluating the Riemannian distance (3.10), which is required for the second term of problem (3.14) for subsequential extraction of prototypes (2.1b) in Section (5), while avoiding to solve the numerically involved numerical generalized eigenvalue problem, we replace (3.14) by

$$\overline{S} = \arg \min_{S \in \mathcal{P}_d} J_s(S; \mathcal{S}_N), \qquad J_s(S; \mathcal{S}_N) = \sum_{i \in [N]} \omega_i D_s(S, S_i). \tag{3.25}$$

The resulting Riemannian gradient flow reads

$$\dot{S} = -\operatorname{grad} J_s(S; \mathcal{S}_N) \overset{(3.12)}{=} -S \partial J(S; \mathcal{S}_N) S \tag{3.26a}$$

$$= -\frac{1}{2} \big(S R(S; \mathcal{S}_N) S - S \big), \qquad R(S; \mathcal{S}_N) = \sum_{i \in [N]} \omega_i \Big(\frac{S + S_i}{2} \Big)^{-1}. \tag{3.26b}$$

Discretizing the flow using the geometric explicit Euler scheme with step size h,

$$S_{(t+1)} = \exp_{S_{(t)}} \big(- h \operatorname{grad} J_s(S_{(t)}; \mathcal{S}_N) \big) \tag{3.27a}$$

$$\overset{(3.11)}{=} S_{(t)}^{\frac{1}{2}} \operatorname{expm} \Big(\frac{h}{2} \big(I - S_{(t)}^{\frac{1}{2}} R(S_{(t)}; \mathcal{S}_N) S_{(t)}^{\frac{1}{2}} \big) \Big) S_{(t)}^{\frac{1}{2}} \tag{3.27b}$$

and using the log-Euclidean mean (3.21) as initial point $S_{(0)}$, defines Algorithm 2 as listed below.

Algorithm 2: Computing the Geometric Matrix Mean Based on the S-divergence.

Initialization
ϵ (termination threshold)
$t = 0, \quad S_{(0)}$ solves (3.21)
$\epsilon_0 > \epsilon$ (any value ϵ_0)
while $\epsilon_t > \epsilon$ **do**

$\quad \Big| \quad LL^\top = S_{(t)}$

$\quad \Big| \quad L_i L_i^\top = \frac{S_{(t)} + S_i}{2}$ for $i \in [N]$

$\quad \Big| \quad U = I - S_{(t)}^{\frac{1}{2}} \big(\sum_{i \in [N]} \omega_i (L_i L_i^\top)^{-1} \big) S_{(t)}^{\frac{1}{2}}$

$\quad \Big| \quad S_{(t+1)} = S_{(t)}^{\frac{1}{2}} \operatorname{expm}(\frac{h}{2} U) S_{(t)}^{\frac{1}{2}}$

$\quad \Big| \quad \epsilon_{t+1} := \|U\|_F, \quad t \leftarrow t + 1$

---

4. ORDERED LAYER SEGMENTATION

In this section, we work out an extension of the assignment flow (Section 2) which is able to respect the order of cell layers as a global constraint while remaining in the same smooth geometric setting. In particular, existing schemes for numerical integration still apply to the novel variant.

4.1. **Ordering Constraint.** With regard to segmenting OCT data volumes, the order of cell layers is crucial prior knowledge. In this paper we focus on segmentation of the following 11 retina layers: Retinal Nerve Fiber Layer (RNFL), Ganglion Cell Layer (GCL), Inner Nuclear Layer (INL), Outer Plexiform Layer (OPL), Outer Nuclear Layer (ONL), two photoreceptor layers (PR1, PR2) separated by the External Limiting Membrane (ELM) and the Retinal Pigment Epithelium (RPE) together with the Choroid Section (CS). Figure 4.1 also contains positions for the Internal Limiting Membrane (ILM) and Brunch Membrane (BM).

FIGURE 4.1. OCT volume acquisition: ① is the A-scan axis (single A-scan is marked yellow). Multiple A-scans taken in rapid succession along axis ② form a two-dimensional B-scan (single B-scan is marked blue). The complete OCT volume is formed by repeating this procedure along axis ③. A list of retina layers that we expect to find in every A-scan is shown on the left.

To incorporate this knowledge into the geometric setting of Section 2, we require a smooth notion of ordering which allows to compare two probability distributions. In the following, we assume prototypes $f_j^* \in \mathcal{F}$, $j \in [n]$ in some feature space $\mathcal{F}$ to be indexed such that ascending label indices reflect the physiological order of cell layers.

Definition 4.1 (Ordered Assignment Vectors). A pair of voxel assignments $(w_i, w_j) \in \mathcal{S}^2$, $i < j$ within a single A-scan is called *ordered*, if $w_j - w_i \in K = \{By : y \in \mathbb{R}_+^c\}$ with the matrix

$$B = \begin{pmatrix} -1 & & & & \\ 1 & -1 & & & \\ & 1 & \ddots & & \\ & & \ddots & -1 & \\ & & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{c \times c} . \tag{4.1}$$

This new continuous ordering of probability distributions is consistent with discrete ordering of layer indices in the following way.

Lemma 4.2. *Let $w_i = e_{l_1}$, $w_j = e_{l_2}$, $l_1, l_2 \in [c]$ denote two integral voxel assignments. Then $w_j - w_i \in K$ if and only if $l_1 \leq l_2$.*

Proof. B is regular with inverse

$$B^{-1} = -Q, \qquad Q_{i,j} = \begin{cases} 1 & \text{if } i \geq j \\ 0 & \text{else} \end{cases} \tag{4.2}$$

and $w_j - w_i \in K \Leftrightarrow B^{-1}(w_j - w_i) \in \mathbb{R}_+^c$. It holds

$$B^{-1}(w_j - w_i) = Qe_{l_1} - Qe_{l_2} = \sum_{k=l_1}^{c} e_k - \sum_{k=l_2}^{c} e_k \tag{4.3}$$

such that $B^{-1}(w_j - w_i)$ has nonnegative entries exactly if $l_1 \leq l_2$. $\qquad\square$

The continuous notion of order preservation put forward in Definition 4.1 can be interpreted in terms of a related discrete graphical model. Consider a graph consisting of two nodes connected by a single edge. The order constrained image labeling problem on this graph can be written as the integer linear program

$$\min_{W \in \{0,1\}^{2 \times c}, M \in \Pi(w_i, w_j)} \langle W, D \rangle + \theta \langle Q - \mathbb{I}, M \rangle \tag{4.4}$$

where $\Pi(w_i, w_j)$ denotes the set of coupling measures for marginals w_i, w_j and $\theta \gg 0$ is a penalty associated with violation of the ordering constraint. By taking the limit $\theta \to \infty$ we find the more tightly constrained problem

$$\min_{W \in \{0,1\}^{2 \times c}, M \in \Pi(w_i, w_j)} \langle W, D \rangle \qquad \text{s.t. } \langle Q - \mathbb{I}, M \rangle = 0 . \tag{4.5}$$

Its feasible set has an informative relation to Definition 4.1 examined in Proposition 4.4.

Lemma 4.3. *Let $M \in \mathbb{R}^{c \times c}$ be an upper triangular matrix with non-negative entries above the diagonal and non-negative marginals*

$$M\mathbb{1}_c \geq 0, \qquad M^\top \mathbb{1}_c \geq 0 . \tag{4.6}$$

Then there exists a modified matrix M^1 with the same properties such that $M^1 \geq 0$.

Proof. (4.6) directly implies $M_{11} \geq 0$ and $M_{cc} \geq 0$ because M is upper triangular. For row indices $l \neq m$ and column indices $q \neq r$, define the matrix $O^{lm,qr}$ with

$$O_{ij}^{lm,qr} = \begin{cases} -1 & \text{if } (i,j) = (l,q) \vee (i,j) = (m,r) \\ 1 & \text{if } (i,j) = (l,r) \vee (i,j) = (m,q) \\ 0 & \text{else} \end{cases} . \tag{4.7}$$

Then $O^{lm,qr}\mathbb{1} = (O^{lm,qr})^\top \mathbb{1} = 0$. Adding a matrix $O^{lm,qr}$ to M does therefore not change its marginals, but it redistributes mass from the positions (l,q) and (m,r) to the positions (l,r) and (m,q). Due to (4.6), it is possible to choose scalars $\alpha_{lr}^k \geq 0$ such that

$$M + \sum_{2 \leq k \leq c-1} \sum_{\substack{l < k \\ r > k}} \alpha_{lr}^k O^{lk,kr} \geq 0 . \tag{4.8}$$

$\qquad\square$

Proposition 4.4. *A pair of voxel assignments $(w_i, w_j) \in \mathcal{S}^2$ within an single A-scan is ordered if and only if the set*

$$\Pi(w_i, w_j) \cap \{M \in \mathbb{R}^{c \times c} \colon \langle Q - \mathbb{I}, M \rangle = 0\} \tag{4.9}$$

is not empty.

Proof. " $\Leftarrow$ " Suppose there exists a measure $M \in \mathbb{R}^{c \times c}$ with marginals w_i, w_j and $\langle Q - \mathbb{1}, M \rangle = 0$. Then

$$w_j - w_i = By \iff Q(M - M^\top)\mathbb{1} = y. \tag{4.10}$$

It suffices to show that no entry of y is negative. Define the shorthand $\zeta = (M - M^\top)\mathbb{1}$. Further, let $M_{.,k}$ denote the k-th column of M and let $M_{k,.}$ denote the k-th row of M. ζ has entries

$$\zeta_l = (M - M^\top)\mathbb{1}|_l = \langle M_{l,.} - M_{.,l}, \mathbb{1} \rangle = \sum_{k=l}^{c} M_{l,k} - \sum_{k=1}^{l} M_{k,l}, \qquad l \in [c]. \tag{4.11}$$

By (4.10), the entries of y read

$$y_r = \sum_{q=1}^{r} \zeta_q. \tag{4.12}$$

We can now inductively show that $y_r \geq 0$ for all $r \in [c]$. The cases $r = 1$ and $r = c$ are immediate:

$$y_1 = \zeta_1 = \sum_{k=1}^{c} M_{1,k} - M_{1,1} = \sum_{k=2}^{c} M_{1,k} \geq 0 \tag{4.13}$$

$$y_c = \langle \zeta, \mathbb{1} \rangle = \langle M - M^\top, \mathbb{1}\mathbb{1}^\top \rangle = \sum_{i,j \in [c]} M_{i,j} - \sum_{i,j \in [c]} M_{i,j}^\top = 0. \tag{4.14}$$

For $r \in \{2, \ldots, c-1\}$ we make the hypothesis that

$$y_r = \sum_{q=1}^{r} \zeta_q = \sum_{k=r+1}^{c} (M_{1,k} + \ldots + M_{r,k}) \geq 0 \tag{4.15}$$

which is consistent with the result for $r = 1$ in (4.13). It follows

$$y_{r+1} = \sum_{q=1}^{r+1} \zeta_q \tag{4.16}$$

$$= \zeta_{r+1} + \sum_{k=r+1}^{c} (M_{1,k} + \ldots + M_{r,k}) \tag{4.17}$$

$$= \sum_{k=r+1}^{c} M_{r+1,k} - \sum_{k=1}^{r+1} M_{k,r+1} + \sum_{k=r+1}^{c} (M_{1,k} + \ldots + M_{r,k}) \tag{4.18}$$

$$= \sum_{k=r+2}^{c} M_{r+1,k} + \sum_{k=r+2}^{c} (M_{1,k} + \ldots + M_{r,k}) \tag{4.19}$$

$$= \sum_{k=r+2}^{c} (M_{1,k} + \ldots + M_{r,k} + M_{r+1,k}) \tag{4.20}$$

where we used (4.15) in (4.17). This completes the inductive step and thus shows $y \geq 0$.

" $\Rightarrow$ " Let (w_i, w_j) be ordered. Following Definition (4.1), it holds

$$B^{-1}(w_j - w_i) = Q(w_i - w_j) \in \mathbb{R}_+^c. \tag{4.21}$$

We show the existence of a transport plan $M \geq 0$ satisfying

$$M\mathbb{1} = w_i, \quad M^\top \mathbb{1} = w_j \tag{4.22}$$

as well as the ordering constraint $\langle Q - \mathbb{1}, M \rangle = 0$ by direct construction. For $c = 2$,

$$M = \begin{pmatrix} (w_j)_1 & (w_i)_1 - (w_j)_1 \\ 0 & 1 - (w_i)_1 \end{pmatrix} \tag{4.23}$$

satisfies these requirements. Now, let $c > 2$ and define the mapping

$$C_1^{c-1} : \Delta_c \to \Delta_{c-1} \tag{4.24}$$

$$w \mapsto \tilde{w} = (w_2, \ldots, w_c) + \frac{w_1}{c-1} \mathbb{1}_{c-1}. \tag{4.25}$$

If $(w_i, w_j) \in \Delta_c^2$ is ordered, then $(\tilde{w}_i, \tilde{w}_j) := (C_1^{c-1}(w_i), C_1^{c-1}(w_j)) \in \Delta_{c-1}^2$ is ordered as well because

$$Q(\tilde{w}_i - \tilde{w}_j) = Q(\bar{w}_i - \bar{w}_j) + \frac{(w_i)_1 - (w_j)_1}{c-1} Q\mathbb{1} \geq 0 \tag{4.26}$$

where $\bar{w}_i$ denotes the vector $((w_i)_2, \ldots, (w_i)_c)$. Suppose a transport plan $\tilde{M} \in \mathbb{R}^{(c-1)\times(c-1)}$ exists such that

$$\tilde{M}\mathbb{1}_{c-1} = \tilde{w}_i \quad \tilde{M}^\top \mathbb{1}_{c-1} = \tilde{w}_j, \quad \tilde{M} \geq 0. \tag{4.27}$$

To complete the inductive step, we consider the matrix

$$M^0 := \begin{pmatrix} (w_j)_1 & s^\top \\ 0 & \tilde{M} - \frac{(w_i)_1}{c-1} I \end{pmatrix}, \quad s = \frac{(w_i)_1 - (w_j)_1}{c-1} \mathbb{1}_{c-1} \tag{4.28}$$

which satisfies (4.22) as well as $\langle Q - \mathbb{1}, M^0 \rangle = 0$. By Lemma 4.3, M^0 can be modified to yield a transport plan with the desired properties. $\square$

Proposition 4.4 shows that transportation plans between ordered voxel assignments w_i and w_j exist which do not move mass from w_{i,l_1} to w_{j,l_2} if $l_1 > l_2$. This characterizes order preservation for non-integral assignments as put forward in Definition 4.1.

4.2. Ordered Assignment Flow.
Likelihoods as defined in (2.21) emerge by lifting $-\frac{1}{\rho} D_{\mathcal{F}}$ regarded as Euclidean gradient of $-\frac{1}{\rho}\langle D_{\mathcal{F}}, W \rangle$ to the assignment manifold. It is our goal to encode order preservation into a generalized likelihood matrix $L_{\text{ord}}(W)$. To this end, consider the assignment matrix $W \in \mathcal{S}^N$ for a single A-scan consisting of N voxels. We define the related matrix $Y(W) \in \mathbb{R}^{N(N-1)\times c}$ with rows indexed by pairs $(i,j) \in [N]^2$, $i \neq j$ in fixed but arbitrary order. Let the rows of Y be given by

$$Y_{(i,j)}(W) = \begin{cases} Q(w_j - w_i) & \text{if } i > j \\ Q(w_i - w_j) & \text{if } i < j \end{cases}. \tag{4.29}$$

By construction, an A-scan assignment W is ordered exactly if all entries of the corresponding $Y(W)$ are nonnegative. This enables to express the ordering constraint on a single A-scan in terms of the energy objective

$$E_{\text{ord}}(W) = \sum_{(i,j)\in[N]^2,\ i\neq j} \phi(Y_{(i,j)}(W)). \tag{4.30}$$

where $\phi \colon \mathbb{R}^c \to \mathbb{R}$ denotes a smooth approximation of $\delta_{\mathbb{R}_+^c}$. In our numerical experiments, we choose

$$\phi(y) = \left\langle \gamma \exp\left(-\frac{1}{\gamma} y\right), \mathbb{1} \right\rangle \tag{4.31}$$

with a constant $\gamma > 0$. Suppose a full OCT volume assignment matrix $W \in \mathcal{W}$ is given and denote the set of submatrices for each A-scan by $C(W)$. Then order preserving assignments consistent with given distance data $D_{\mathcal{F}}$ in the feature space $\mathcal{F}$ are found by minimizing the energy objective

$$E(W) = \langle D_{\mathcal{F}}, W \rangle + \sum_{W_A \in C(W)} E_{\text{ord}}(W_A) . \tag{4.32}$$

We consequently define the generalized likelihood map

$$L_{\text{ord}}(W) = \exp_W \left(-\nabla E(W) \right) = \exp_W \left(-\frac{1}{\rho} D_{\mathcal{F}} - \sum_{W_A \in C(W)} \nabla E_{\text{ord}}(W_A) \right) \tag{4.33}$$

and specify a corresponding assignment flow variant.

Definition 4.5 (Ordered Assignment Flow)**.** The dynamical system

$$\dot{W} = R_W S(L_{\text{ord}}(W)), \qquad W(0) = \mathbb{1}_{\mathcal{W}} \tag{4.34}$$

evolving on $\mathcal{W}$ is called the *ordered assignment flow*.

By applying known numerical schemes [ZSPS20] for approximately integrating the flow (4.34), we find a class of discrete-time image labeling algorithms which respect the physiological cell layer ordering in OCT data. In chapter 5, we benchmark the simplest instance of this class, emerging from the choice of geometric Euler integration.

5. EXPERIMENTAL RESULTS

5.1. **Data, Competing Approaches, Performance Measures.**

5.1.1. *OCT-Data.* In the following sections, after introducing key terminology in volumetric OCT data we describe experiments performed on a set of OCT volumes depicting the intensity of light reflection in chorioretinal tissues centered around the fovea. The scans were obtained using a spectral domain OCT device (Heidelberg Engineering, Germany) for multiple patients at a variety of resolutions by averaging various registered B-scan which share same location to reduce the speckle noise. This is representative of the fact that different resolutions may be desirable in clinical settings at the preference of medical practitioners. In the following, we always assume an OCT volume in question to consist of N_B B-scans, each comprising N_A A-scans with N voxels and use the term surface to represent the set of voxels located in the interface of two retina layers. See Figure 4.1 for a schematic acquisition illustration of retina layers and separating membranes.

In this publication we use OCT volumes of size $(N \times N_A \times N_B) = (498 \times 768 \times 61)$ with a approximate resolution of $3.87 \,\mu\text{m/voxel}$ along N, N_A direction and with a resolution of $11.11 \,\mu\text{m/voxel}$ on N_B axis. The volume set was divided into training set and the testing set where the latter consists of 8 volumes extracted from different patients without any observable pathological retina changes. Figure (5.1) provides a view on a Bscan located in the OCT volume on which the ordered assignment flow is validated. The right plot depicts the noisy signal along an A-scan indicated by a yellow vertical line which underpins the difficulty of segmenting the underlying data sets.

5.1.2. *Reference Methods.* To assess the segmentation performance of our proposed approach we compare ourselves to state of the art retina segmentation methods presented in Rathke [RSS14] and Li [KXCS06] which are applicable for both healthy and pathological patient data. In particular, we prefer these reference methods over [DCA+13], [SBG+13] and [GAW+09] because available implementations of the latter are limited to the segmentation of up to 9 retina layers.
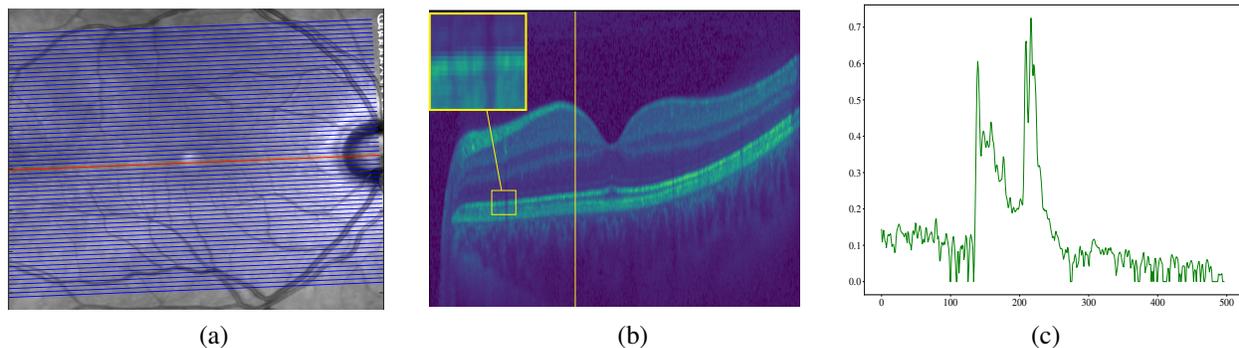
FIGURE 5.1. **Left**: En-face view on the volumetric OCTA data with red line indicating the location of the B-scan shown in the center image. **Center**: The enlarged view depicts typical artifacts such as shadow regions and speckle noise. **Right**: The gray value intensity of one vertical A-scan within the B-Scan represented by yellow line in the center image. The noisy intensity plot indicates the difficulty of application to segment the OCT-volume accurately while respecting layer ordering.

IOWA Reference Algorithm: A well-known graph-based approach to segmentation of macular volume data was developed by the Retinal Image Analysis Laboratory at the Iowa Institute for Biomedical Imaging [KXCS06, AGS10, GAW$^+$09]. The problem of localizing cell layer boundaries in 3D OCT volumes is posed and ultimately transformed into a minimum st-cut problem on a non-trivially constructed graph G. To this end, a distance tensor $D_k \in \mathbb{R}^{N_B \times N_A \times N}$ is formed in a feature extraction step for each boundary $k \in [c-1]$. This encodes $c-1$ separate binary segmentation problems on a geometric graph G_k spanning the volume. In each instance, voxels are to be classified as either belonging to boundary k or not belonging to boundary k. By utilizing a (directed) neighborhood structure on each G_k, smoothness constraints are introduced and regulated via user-specified stiffness parameters. To model interactions between different boundaries, the graphs G_k are combined to a global graph G, introducing additional edges between them. The latter set up constraints on the distance between consecutive boundaries within each A-scan which can be used to enforce physiological ordering of cell layers. On G, the problem of optimal boundary localization takes the form of minimal closed set construction which is in turn transformed into a minimum st-cut problem for which standard methods exist. Their standalone software is freely available for research purposes[1].

Probabilistic Model: Rathke et al. [RSS14] proposed a graph-based probabilistic approach for segmenting OCT volumes for given data y by leveraging the Bayesian ansatz

$$p(y, s, b) = p(y|s)p(s|b)p(b) . \tag{5.1}$$

Here, the tensor $b \in \mathbb{R}^{N_B \times N_A \times (c-1)}$ contains real-valued boundary positions between retina layers and s denotes discrete (voxel-wise) segmentation. The appearance terms $p(y|s)$, $p(s|b)$ and $p(b)$ represent data likelihood, Markov random field regularizer and global shape prior respectively. In order to approximate the desired posterior

$$p(b, s|y) = \frac{p(y|s)p(s|b)p(b)}{p(y)} , \tag{5.2}$$

a variational inference strategy is employed. This aims to find a tractable distribution q decoupled into

$$q(b, s) = q_b(b)q_s(s) \tag{5.3}$$

---

[1]see https://www.iibi.uiowa.edu/oct-reference

which is close to $p(b, s|y)$ in terms of the relative entropy $\mathrm{KL}(q\,|\,p)$. The shape prior $p(b)$ is learned offline by maximum likelihood estimation in the space of normal distributions using a low-rank approximation of the involved covariance matrix. Ordering constraints

$$1 \leq s_{1,ij} \leq s_{2,ij} \leq \cdots \leq s_{c-1,ij}, \quad ij \in [N_B] \times [N_A] \tag{5.4}$$

are enforced for the discrete segmentation s and are not enforced for the continuous boundaries b. This is in contrast to the proposed model which integrates the ordering of retina layers by adding a cost function (4.9) penalizing the overall deviation of soft assignments by integrating (2.24) from the subspace of probability distributions satisfying (4.1). The method comes along with a standalone software free available under [2].

5.1.3. *Performance Measures.* We will evaluate the computed segmentations by their direct comparison with manual annotations regarded as gold standard which were realized by an medical expert. Respective metrics are suitable for segmentation tasks that involve multiple tissue types [CCH06]. Specifically, we report the mean DICE similarity coefficient [Dic45] for each segmented cell layer.

Definition 5.1. *(DICE)* Given two sets A, B the *DICE similarity coefficient* is defined as

$$\mathrm{DSC}(A, B) := \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \in [0, 1], \tag{5.5}$$

where $\{TP, FN, FP\}$ denotes the number of *true positives*, *false negatives* and *false positives* respectively.

The DICE similarity coefficient quantifies the region agreement between computed segmentation results and manually labeled OCT volumes which serve as ground truth. High similarity index $\mathrm{DSC}(A, B) \approx 1$ indicates large relative overlap between the sets A and B. This metric is well suited for average performance evaluation and appears frequently in the literature (e.g. [CAM$^+$15], [YHSS11] and [NVd$^+$17]). It is closely related to the positively correlated Jaccard similarity measure [Jac08] which in contrast to (5.5) is more strongly influenced by worst case performance.

In addition, we report the mean absolute error (MAE) of computed layer boundaries used in [RSS14] and [GAW$^+$09] to make our results more directly comparable to these references.

Definition 5.2. *(Mean Absolute Error)* For a single A-scan indexed by $ij \in [N_B] \times [N_A]$, let $e_{ij} := |g_{ij} - p_{ij}|$ denote the absolute difference between a layer boundary position g_{ij} in the gold standard segmentation and a predicted layer boundary p_{ij}. The mean absolute error (MAE) is defined as the mean value

$$\mathrm{MAE}(g, p) = \frac{1}{N_B N_A} \sum_{ij \in [N_B] \times [N_A]} e_i . \tag{5.6}$$

5.2. **Feature Extraction.**

5.2.1. *Region Covariance Descriptors.* To apply the geometric framework proposed in Section 3 we next introduce the *region covariance descriptors* [TPM06] which have been widely applied in computer vision and medical imaging, see e.g. [CS16, TS16, DFVDVM14, SSR15]. We model the raw intensity data for a given OCT volume by a mapping $I : \mathcal{D} \to \mathbb{R}_+$ where $\mathcal{D} \subset \mathbb{R}^3$ is the underlying spatial domain. To each voxel $v \in \mathcal{D}$, we associate the local feature vector $f : \mathcal{D} \to \mathbb{R}^{10}$,

$$f : \mathcal{D} \to \mathbb{R}^{10} \tag{5.7}$$

$$v \mapsto (I(v), \nabla_x I(v), \nabla_y I(v), \nabla_z I(v), \sqrt{2}\nabla_{xy} I(v), \sqrt{2}\nabla_{yz} I(v), \nabla_{xx} I(v), \nabla_{yy} I(v), \nabla_{zz} I(v))^\top . \tag{5.8}$$

assembled from the intensity $I(v)$ as well as first- and second-order responses of derivative filters capturing information from larger scales following [HS87]. To improve the segmentation accuracy we combine the

---

[2] https://github.com/FabianRathke/octSegmentation

derivative filter responses from various scales in an computationally efficient way we first normalize the derivatives of the input volume $I(v)$ at every scale σ_s by convolution each dimension with a 1D window:

$$\nabla_x \tilde{I}_{\sigma_s}(v) = \sigma_s^2 \frac{\partial}{\partial x} \tilde{G}(v, \sigma_s) \tag{5.9}$$

where $\tilde{G}(v, \sigma_s)$ is an approximation to Gaussian window $\big(G(v, \sigma_s) * I\big)(v)$ at scale σ_s as in detail described in [HS87]. Subsequently we follow the idea presented by [Lin04] by taking local maxima over scales

$$\nabla_x \tilde{I}(v) = \max_{\sigma_s} \nabla_x \tilde{I}_{\sigma_s}(v), \tag{5.10}$$

which are serving for the mapping (5.7).

By introducing a suitable geometric graph spanning $\mathcal{D}$, we can associate a neighborhood $\mathcal{N}_i$ of fixed size with each voxel $i \in [n]$ as in (2.23). For each neighborhood, we define the regularized *region covariance descriptor*

$$S_i := \sum_{j \in \mathcal{N}_i} \theta_{ij}(f_j - \overline{f_i})(f_j - \overline{f_i})^T + \epsilon I, \quad \overline{f_i} = \sum_{k \in \mathcal{N}_i} \theta_{ik} f_k, \tag{5.11}$$

as a weighted empirical covariance matrix with respect to feature vectors f_j. The small value $1 \gg \epsilon > 0$ acts as a regularization parameter enforcing positive definiteness of S_i. Diagonal entries of each covariance matrix C_i are empirical variances of feature channels in (5.7) while the off-diagonal entries represent empirical correlations within the region $\mathcal{N}_i$.

5.2.2. *Prototypes on* $\mathcal{P}^d$. In view of the assignment flow framework introduced in Section 2, we interpret region covariance descriptors (5.11) as data points in the metric space $\mathcal{P}^d$ of symmetric positive definite matrices and model each retina tissue indexed by $l \in [c]$ with a random variable S_l taking values in $\mathcal{P}^d$. Suppose we draw N_l samples $\{S_l^k\}_{k=1}^{N_l}$ from the distribution of S_l. The most basic way to apply assignment flows to data in $\mathcal{P}^d$ is based on computing a prototypical element of $\mathcal{P}^d$ for each tissue layer, e.g. the Riemannian center of mass of $\{S_l^k\}_{k=1}^{N_l}$. This corresponds to directly choosing $\mathcal{P}^d$ as feature space $\mathcal{F}$ in (2.1a). We find that superior empirical results are achieved by considering a dictionary of $K_l > 1$ prototypical elements for each layer $l \in [c]$. This entails partitioning the samples $\{S_l^k\}_{k=1}^{N_l}$ into K_l disjoint subsets $\hat{S}_l^j \subseteq \{S_l^k\}_{k=1}^{N_l}$, $j \in [K_l]$ with representatives $\tilde{S}_l^j$ determined offline.

To find a set of representatives which captures the structure of the data, we minimize expected loss measured by the Stein divergence (3.24) leading to the K-means like functional

$$\mathbb{E}_{p_l}(\tilde{S}_l) = \sum_{j=1}^{K_l} p(j) \sum_{S_l^i \in \hat{S}_l^j} \frac{p(i|j)}{p(j)} D_S(S_l^i, \tilde{S}_l^j), \quad p(i,j) = \frac{1}{N_l}, p_l(j) = \frac{N_j}{N_l}. \tag{5.12}$$

A hard partitioning is achieved by applying Lloyd's algorithm in conjunction with algorithm 2 for mean retrieval. We additionally employ the more common soft K-means like approach for determining prototypes by employing the mixture exponential family model based on Stein divergence to given data

$$p(S_l^i, \Gamma_l) = \sum_{j=1}^K \pi_l^j p(S_l^i, \tilde{S}_l^j)), \tag{5.13}$$

where the parameters

$$\Gamma_l = \{(\pi_l^j\}_{j=1}^K, \{\tilde{S}_l^j\}_{j=1}^K), \quad (\pi_l^1, \cdots, \pi_l^{|J|}) \in S \tag{5.14}$$

have to be adjusted to given data. The prototypes are recovered as mean parameters $S_l^{j,T}$ though an iterative process commonly refered to as *expectation maximation* (EM) defined by alternation of the following iterations

$$p_l(j|S_l^i, \Gamma_l^t) = \frac{\pi_l^{(j,t)} e^{-D_S(S_l^i, \tilde{S}_l^{(j,t)})}}{\sum_{k=1} \pi_l^{(k,t)} e^{-D_S(S_l^i, \tilde{S}_l^{(k,t)})}}, \qquad \textbf{\textit{(Expectation step)}} \qquad (5.15)$$

followed by updating the marginals at each time step up to final time T

$$\pi_l^{(j,t+1)} = \sum_{i=1}^{N_j} p_l(j|S_l^i, \Gamma_l^t) \tilde{S}^{j,t} \qquad\qquad (5.16)$$

$$\tilde{S}_l^{j,t+1} = \mathrm{argmin}_{S \in \mathcal{P}_d} \Big(\sum_{i=1}^{n} p(j|\Gamma_i^t) D_S(S_l^i, S) \Big), \qquad \textbf{\textit{(Maximization step)}}. \qquad (5.17)$$

The decision to approximate the Riemannian metric on $\mathcal{P}_d$ by the Stein divergence (3.24) can be backed up empirically. To this end, we randomly select descriptors (5.11) representing the nerve fibre layer in real-world OCT data and compute their Riemannian mean as well as their mean w.r.t. the Log-Euclidian metric (3.20) and Stein divergence (3.24). Figure 5.3 illustrates that Stein divergence approximates the full Riemannian metric more precisely than the Log-Euclidian metric while still achieving a significant reduction in computational effort. Furthermore to evaluate the classification we extracted a dictionary of 200 prototypes for representing each retina tissue for different choice of metric and subsequently evaluated the resulting segmentation accuracy by assigning each voxel to a class containing the prototype with smallest distance using a cropped OCT Volume of size $138 \times 100 \times 40$ taken from the testing set.

Figure 5.2 visualizes the correct classification matches for retina layers ordered by color according to Figure 4.1. In particular, we inspect a notable gain of correct matches while respecting the Riemannian geometry (first column) as opposed to Log-Euclidean setting (third column). Regarding the approximation of (3.10) by (3.24), we are observing more effective detection of outer Photoreceptor Layer (PR1), Inner Nuclear Layer (INL) and Retinal Pigment Epithelium (RPE). Furthermore, taking a closer look at (OPL) and (ONL) we note a typical tradeoff between the number of prototypes and detection performance indicating superior retina to voxel allocation by applying (3.20), whereas the surrogate divergence metric (3.24) has the tendency to improve the accuracy while increasing the size of evaluated prototypes in contrast to flattening curves when relying on (3.21).

This illustrates a tradeoff between computational effort and labeling performance, cf. Figure 5.3. Note that prototypes are computed offline, making runtime performance less relevant to medical practitioners. However, building a distance matrix involves computing $n \sum_{l \in [c]} K_l$ Riemannian distances resp. Stein divergences to prototypes. This still leads to a large difference in (online) runtime since evaluation of the Riemannian distance (3.10) involves generalized eigendecomposition while less costly Cholesky decomposition suffices to evaluate the Stein divergence (3.24).

Summarizing the discussed results concerning the application of Algorithm 1 and Algorithm 2, we point out that respecting the Riemannian geometry leads to superior labeling results providing more descriptive prototypes.

5.2.3. *CNN Features.* In addition to the covariance features in Section 5.2.1, we compare a second approach to local feature extraction based on a convolutional neural network architecture. For each node $i \in [n]$, we trained the network to directly predict the correct class in $[c]$ using raw intensity values in $\mathcal{N}_i$ as input. As output, we find a score for each layer which can directly be transformed into a distance vector suitable as input to the ordered assignment flow (4.34) via (4.33). The specific network used in our experiments has a ResNet architecture comprising four residually connected blocks of 3D convolutions and ReLU activation.
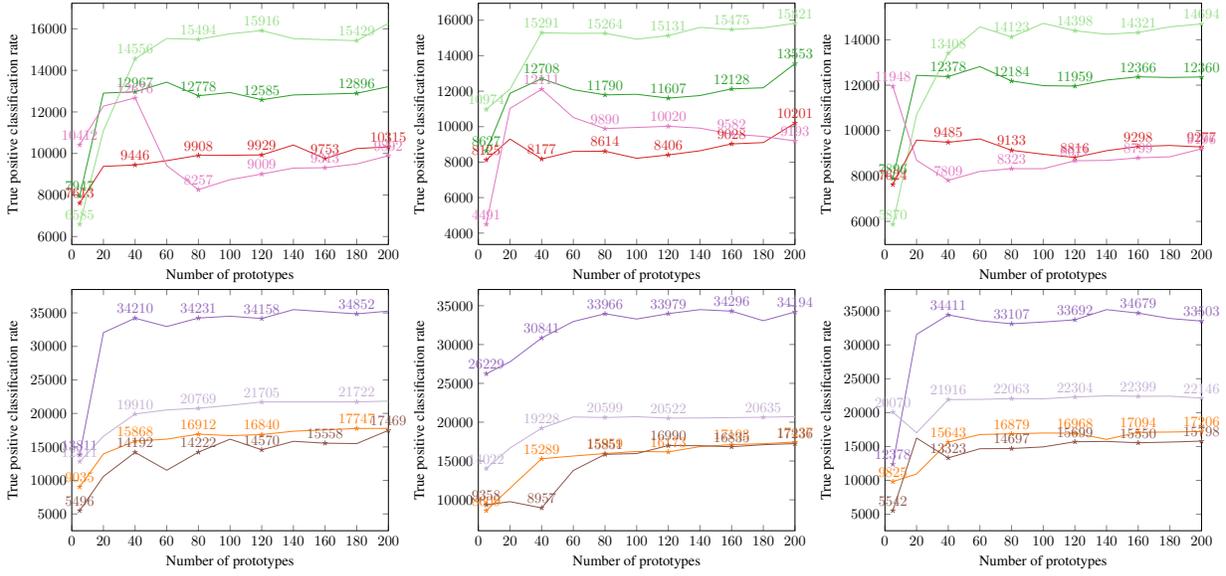
FIGURE 5.2. **Top**: Metric classification evaluated on thin layers (IPL,INL,OPL,PR2). **Bottom**: Analogous metric evaluation for (GCL,ONL,PR1,RPE). **From left to right**: The number of true outcomes after direct comparison with ground truth, for the choice of the exact Riemannian geometry of $\mathcal{P}_d$, Stein divergence and Log-Euclidean distance for geometric mean computation. The results of first two columns indicate higher detection performance while respecting the Riemannian geometry of a curved manifold. Enlarging the set of prototypical covariance descriptors leads to increased matching accuracy which is in contrast to the observed flattening of matching curves when using the Log-Euclidean distance.
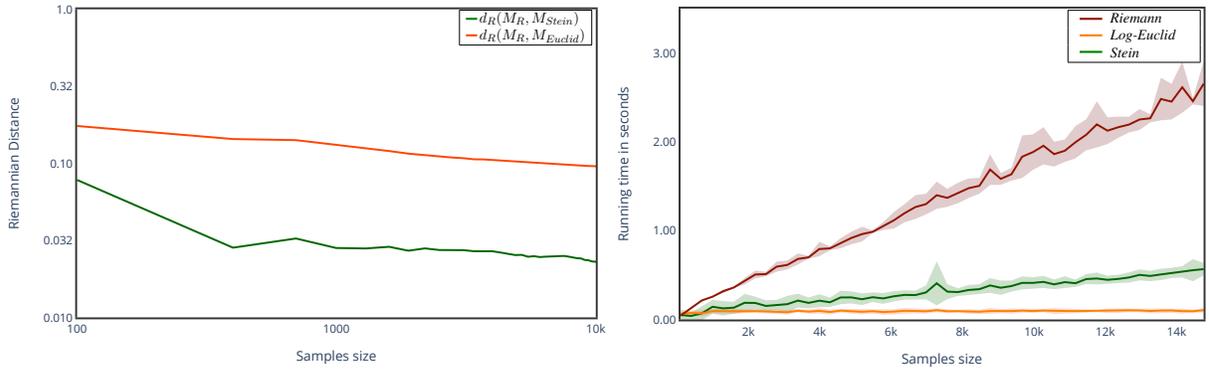
FIGURE 5.3. **Left**: Deviation of the geometric means computed using the Log-Euclidian metric and Stein divergence, respectively, from the true Riemannian mean. **Right**: Runtime for geometric mean computation using the different metrics. All evaluations were performed on a randomly chosen subset of covariance descriptors representing the retinal nerve fibre layer in a real-world OCT scan. Both graphics clearly highlight the advantages of using Stein the divergence in terms of approximation accuracy and efficient numerical computation.

Model size was hand-tuned for different sizes of input neighborhoods, adjusting the number of convolutions per block as well as corresponding channel dimensions. In particular, labeling accuracy is increased for the detection of RPE and PR2 layers, as illustrated in the last row of Figure 5.7.

5.3. **Segmentation via Ordered Assignment.** By numerically integrating the ordered assignment flow (4.5) parametrized by the distance matrix D, an assignment state W is evolved on $\mathcal{W}$ until mean entropy
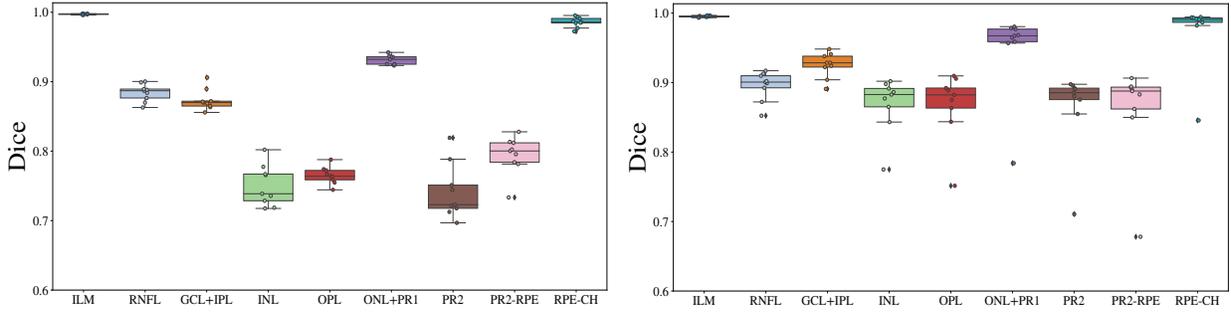
FIGURE 5.4. Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. **Left**: Probabilistic approach (5.1) proposed in [RSS14]. **Right**: OAF based on CNN features. See Table 2 for mean and standard deviations. Direct comparison shows a notably higher detection performance for segmenting the intraretinal layers using OAF (B).
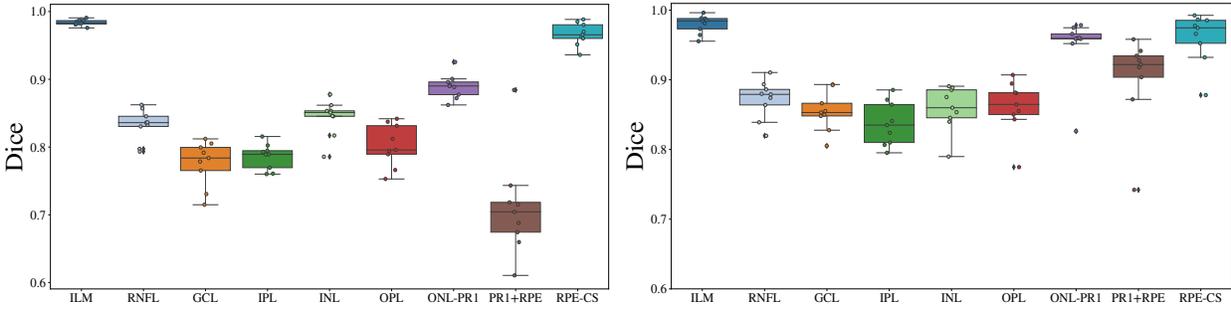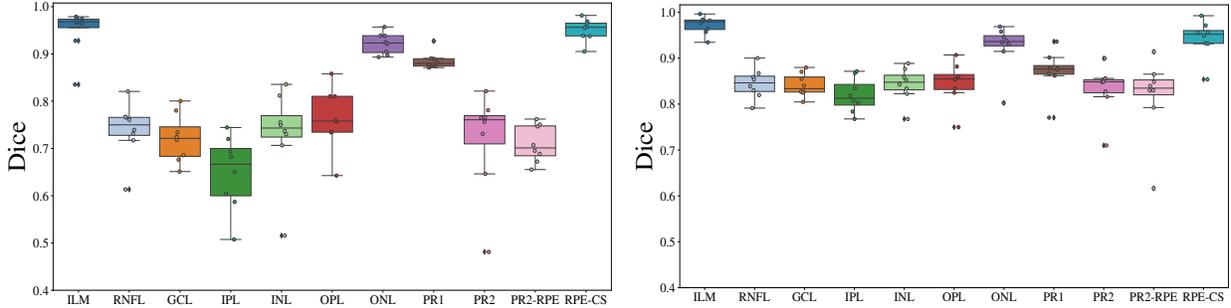

FIGURE 5.5. Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. **Left**: IOWA reference algorithm [GAS]. **Right**: OAF based on CNN features. See Table 3 for mean and standard deviations. Exploiting OAF (B) for retina tissue classification results in improved overall layer detection performance, especially for the PR2-RPE region.

FIGURE 5.6. Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. **Left**: OAF (A). **Right**: OAF (B). The OAF based on CNN features yields improved segmentations for all retina layers.

of pixel assignments is low. We specifically use geometric Euler integration steps on $T\mathcal{W}$ with a constant step-length of $h = 0.1$ (see [ZSPS20] for details of this process). Geometric averaging with uniform weights leads to local regularization of assignments which smooths regions in which the features do not conclusively point to any label. More global knowledge about the ordering of cell layers is incorporated into E_{ord} which

addresses more severe inconsistencies between local features and global ordering. In all experiments, the neighborhood of each voxel $i \in [n]$ is choosen as the voxel patch of size 5 x 5 x 3 centered at i.

5.4. **Evaluation.** To benchmark our novel segmentation approach, we first extract local features for each voxel from a raw OCT volume. As described above, either region covariance descriptors (Section 5.2.1) or class scores predicted by a CNN (Section 5.2.3) are computed for segmenting the retina layers with ordered assignment flow which we in the following abbreviate as **OAF (A)** and **OAF (B)** respectively. In the former case, a set of $k = 400$ prototypical cluster centers on the positive definite cone (3.8) has been determined offline for each cell layer. These are compared to descriptors extracted from the unseen volume by computing the pair-wise distance with respect to the metric induced by Stein divergence (3.2.3). The minimum value corresponding to the lowest of all the distances for each pair of voxel $i \in [n]$ and cell layer $j \in [c]$ is noted as entry d_{ij} of the distance matrix D_{cov}, i.e. for every voxel i the distance to optimal fitting representative to layer j is given by

$$(D_{\mathrm{cov}})_{ij} := \min_{k \in [400]} D_S(S_i, \tilde{S}_j^k). \tag{5.18}$$

In the latter case, class scores $C \in \mathbb{R}^{n \times c}$ predicted by the neuronal network (5.2.3) are transformed into a distance matrix $D_{\mathrm{cnn}} = -C$ simply by switching their sign followed by adjusting the parameter (2.21) to weight the data relevance in the likelihood matrix.

A naive way to segment the volume in accordance to the data is by choosing $\arg\min_{j \in [c]} D_{ij}$ for each voxel i. However, due to the challenging signal-to-noise ratio in real-world OCT data, classes will not usually be well-separated in the feature space at hand. The resulting uncertainty pertaining to the assignment of classes using exclusively local features is encoded into each distance matrix. To assess the segmentation performance of our proposed approach, we first compared to the state of the art graph-based retina segmentation method of 10 intra-retinal layers developed by the Retinal Image Analysis Laboratory at the Iowa Institute for Biomedical Imaging [KXCS06, AGS10, GAW+09], also referred to as the IOWA Reference Algorithm. We quantify the region agreement with manual segmentation regarded as gold standard. Since both the augmented volumes and the compared reference methods determine boundary locations of retina layers intersections, we first transfer the retina surfaces to a layer mask by rounding to the voxel size and assign to voxels within each A-scan the associated layer label, starting from the observed boundary to the location of the next detected intersection surface of two neighboring layers.

Specifically, we calculate the DICE similarity coefficient [Dic45] and the mean absolute error for segmented cell layers within the pixel size of $3.87 \, \mu\mathrm{m}$ compared to human grader by segmenting 8 OCT volumes consisting of 61 B-scans. We first directly compare the performance accuracy of using the local features given by the covariance descriptor (5.2.1) by constructing a dictionary with 400 prototypes for each retina layer using the iterative clustering with (5.13) against the features extracted from an CNN net (5.2.3).

The experimental results discussed next illustrate the relative influence of the covariance descriptors (5.11) and regularization property of the ordered assignment flow, respectively. Throughout, we fixed the grid connectivity $\mathcal{N}_i$ for each voxel $i \in \mathcal{I}$ to $3 \times 5 \times 5$. Figure 5.7 illustrates real-world labeling performance based on extracting a dictionary of 400 prototypes per layer by minimizing (5.12) and employing Algorithm 2 for mean retrieval. Second row in Figure 5.7, illustrates a typical result of nearest neighbor assignment and the volume segmentation *without* ordering constraints. As inspected, the high texture similarity between the choroid and GCL layer yields wrong predictions resulting in violation of biological retina ordering through the whole volume which cannot be resolved with the based assignment flow approach given in Section 2. On the other side using pairwise correlations captured by covariance matrices provides accurate detection of large signal intensity internal limiting membrane (ILM) with its characteristic highly reflective boundary as well to meaningful segmentation of light rejecting fiber layers RNFL, PR1 and RPE. For the particularly challenging inner layers such as GCL, INL and ONL that are mainly comprised of weakly reflective neuronal
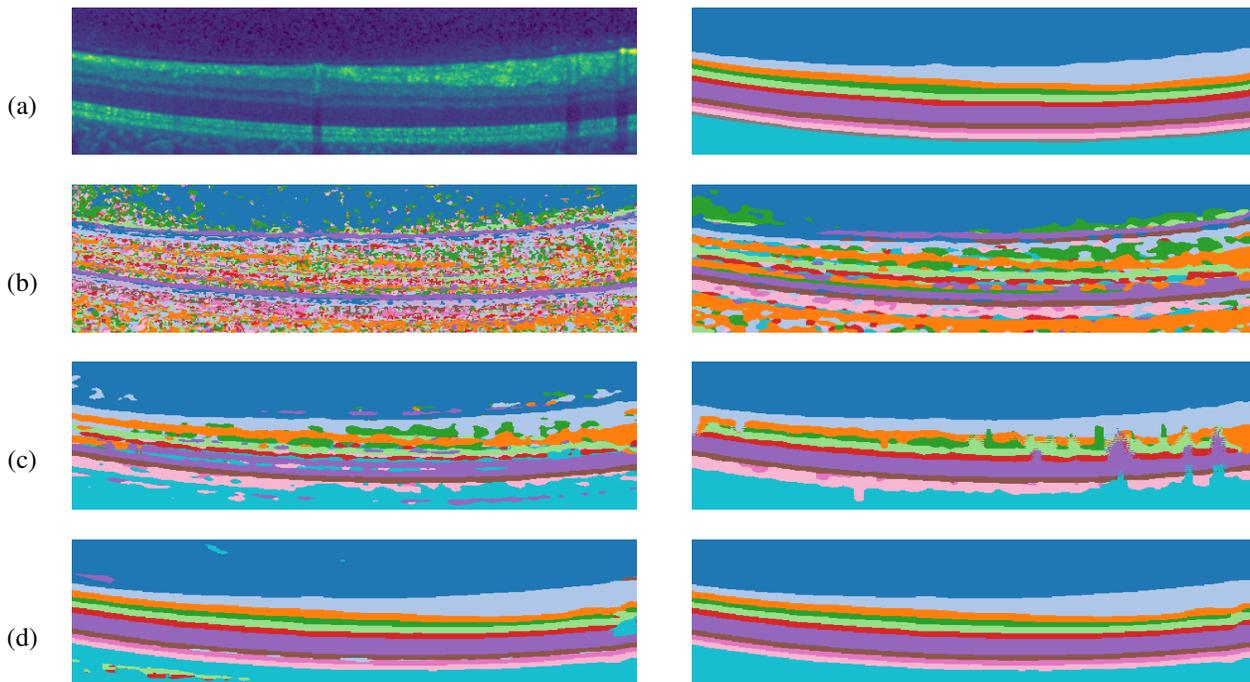
FIGURE 5.7. **From top to bottom: Row (a)**: One B-scan from a OCT-volume showing the shadow effect, with ground truth plot on the right. **Row (b)**: Local nearest neighbor assignments based on prototypes by minimizing (5.12) computed with Stein divergence, with the result of the segmentation returned by the basic assignment flow (Section 2) on the right. **Row (c)**: Proposed *layer-ordered* volume segmentation based on covariance descriptors. From left to right: ordered volume segmentation for different $\gamma = 0.5, \gamma = 0.1$ (cf. Eq. (4.31)). **Row (d)**: Local rounding result extracted from Res-Net on the left and the result of the ordered assignment flow on the right.

cell bodies, regularization by imposing (4.30) is required. In the third row of Figure 5.7, we plot the *ordered* volume segmentation by stepwise increasing the parameter γ defined in (4.31), which controls the ordering regularization by means of the novel generalized likelihood matrix (4.33). The direct comparison with the ground truth remarkably shows how the ordered labelings evolve on the assignment manifold while simultaneously giving accurate data-driven detection of RNFL, OPL, INL and the ONL layer. For the remaining critical inner layers, the local prototypes extracted by (5.12) fail to segment the retina layers properly and are still revealing artifacts due to the presence of vertical shadow regions caused by existing blood vessels, which contribute to a loss of the interference signal during the scanning process of the OCT-data, as shown in Figure 5.7.

After segmentation of the test data set, the mean and standard deviation were calculated for better assessment of the retina layer detection accuracy of the proposed segmentation method, according to the performance measures (5.6) and (5.5). The evaluation results for each retina tissue as depicted in Figure 4.1, are detailed in Table 1. The first row of Figure 5.12 clearly shows the superior detection accuracy of utilizing the Ordered Assignment Flow for the first outer retina layers (RNFL, GCL, IPL, INL) and the (PR2-RPE) region in connection with local features extracted from an CNN net (5.2.3). Nonetheless, the covariance descriptor achieves comparable results for characterization of the outer plexiform layer (OPL) and exhibits increased retina detection regarding the photoreceptor region (PR1,PR2) and outer nuclear region (ONL). Additionally, Table 1 includes the evaluation based on DICE similarity which is less sensitive to outliers and serves as an appropriate metric for calculating the performance measures across large 3D volumes. To obtain
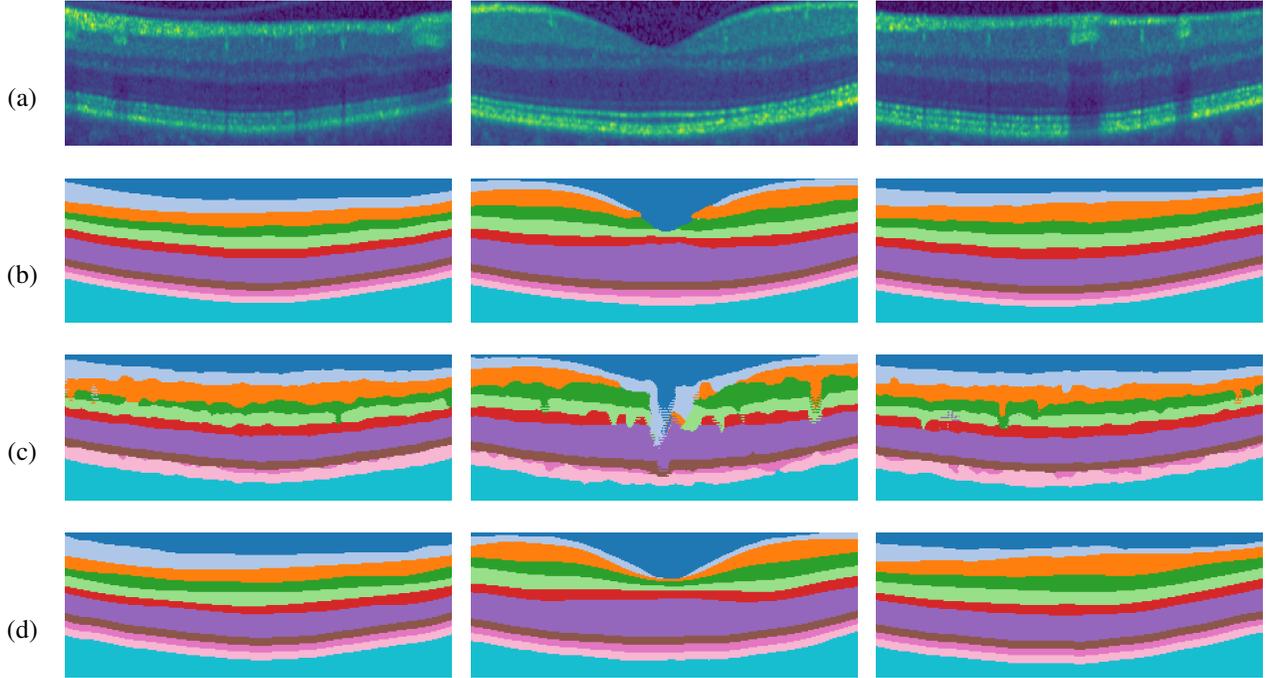
FIGURE 5.8. **From top to bottom: Row (a)**: Three sample B-Scans extracted for different locations from a healthy OCT volume with 61 scans, with the fovea centered OCT scan visualized in the middle column. **Row (b)**: The associated augmented labeling. **Row (c)**: OAF (A) segmentation using a dictionary of covariance descriptors determined by (5.13). **Row (d)**: OAF (B) segmentation using features determined the CNN network. In contrast to to results achieved by OAF (A), the above visualization indicates more accurate detection of retina boundaries using OAF (B), in particular near the fovea region (middle column).

a consistent and clear comparability between the involved features on which we rely to tackle the specific problem of retina layer segmentation, the corresponding results are visualized in Figure 5.6. The graphic illustrates higher Dice similarity and relative small standard deviation when incorporating features (5.2.3) as input to our model, which characterizes their superior informative content. According to the left plot, the covariance descriptor performs well for detecting the prototypical textures of the internal limiting membrane (ILM), the (ONL) and (PR1) layers as well as the RPE boundary to the choroid section. Especially this highlights the ability of using gradient based features for accurate detection of retina tissues indicating sharp contrast between the neighboring layers, as is the case for ONL and PR1.

On the contrary, the decrease of effective detection regarding the remaining layers reflects the present high texture variability, which can by tackled by increasing the size of prototypes on the cost of efficient computation of the generalized likelihood (4.33). In general, the more robust retina detection features from an CNN net can be attributed to the underlying manifold geometry of symmetric positive definite matrices where the data partition is performed linearly by hyperplanes. This further indicates the nonlinear structure of the acquired volumetric OCT data. Figure 5.8 presents typical labelings of a B-scan for different locations in the segmented healthy OCT-volume obtained with the proposed approach. Direct comparison with the ground truth, as depicted in row (b), demonstrate higher accuracy and smoother boundary transitions by using CNN features instead of covariance descriptors. In particular, for the challenging segmentation of the ganglion cell layer (GCL) with a typical thinning near the macular region (middle scan), we report a Dice
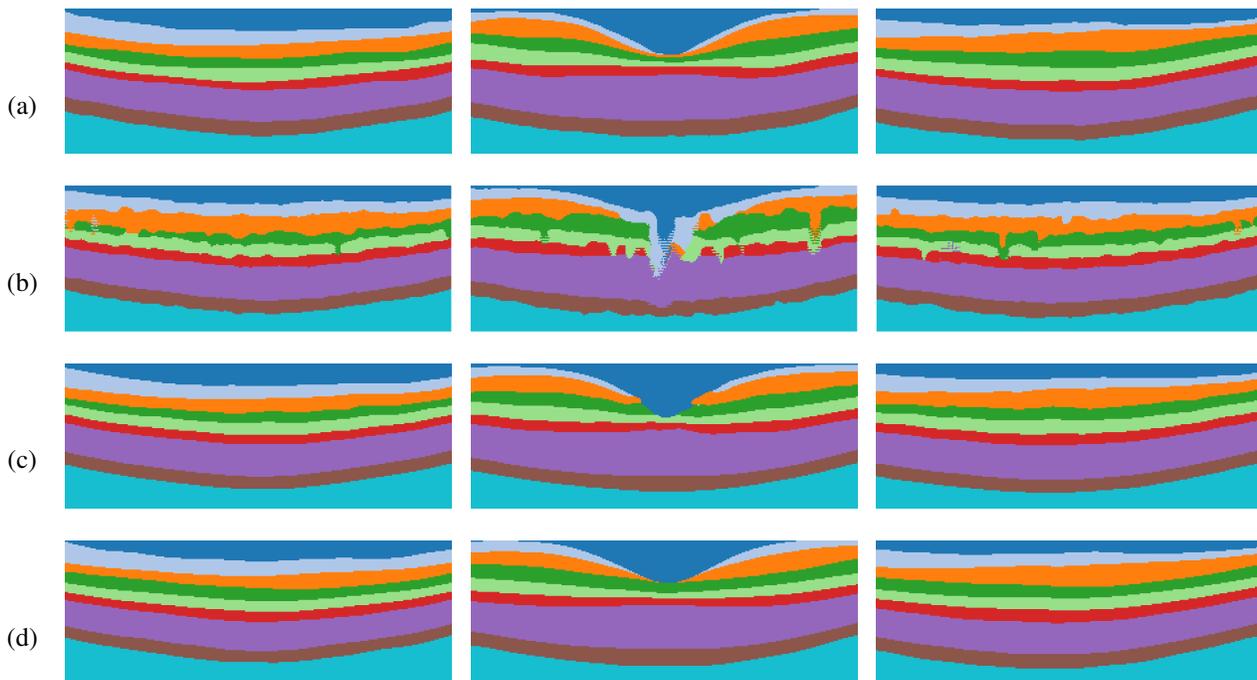
FIGURE 5.9. Illustration of retina layer segmentation results listed in Table (3). **Row (a)**: Ground truth labeling. **Row (b-c)**: Labeled retina tissues using the proposed approach based on covariance descriptors and CNN features, respectively. **Row (d):** The resulting segmentation obtained using the IOWA reference algorithm. Visual comparison with manually annotated retina layers (a) shows that segmentation with OAF (B) leads to more reliable layer thickness of regions concentrated near the fovea region (middle column), as opposed to IOWA reference method.

index of $0.8373 \pm 0.0263\,\mu m$ as opposed to the result $0.6657 \pm 0.1909\,\mu m$. The remaining numerical experiments are focusing on the validation of OAF against the retina segmentation methods serving as reference, as summarized in Section 5.1.2.

To access a quantitative direct comparison with the IOWA reference algorithm, the tested OCT volumes were imported into OCTExplorer 3.8.0 and segmented using the predefined Macular-OCT IOWA software after properly adjusting the resolution parameters. Additionally, we preprocessed each volume by removing 2 B-scans from each side to get rid of boundary artifacts. We calculated and compared the segmentation results for layers consistent with available OCT volumes which were augmented by a medical expert. As before, we use the mean average error and the Dice index after segmenting the 8 volumes with no observable intraretinal diseases which are reported in Table (3). Figure 5.5 provides a statistical illustration of the Dice index which reveals the high performance accuracy for methods which is in accordance with the mean average error shown in the last row of Figure 5.12. In particular, we observe a notable increase of performance using the OAF for detection of the ganglion cell layer with overall accuracy of $0.8546 \pm 0.0281\,\mu m$, see Figure 5.9 for visualized segmentations of 3 B-scans.

Next, we provide a visual and statistical comparison of the proposed approach and the probabilistic state of the art retina segmentation approach [RSS14] underlying Eq. (5.1). As before, to achieve a direct comparison with the proposed approach, we first adopted the OCT volumes to match the shape and parameters given in [RSS14]. Subsequently, we removed the boundary between the GCL and IPL layers to obtain one characteristic layer which has to be detected. Figure 5.10 displays the labeling accuracy. Both methods perform well by accurately segmenting flat shaped retina tissues, as shown in the first and last columns.
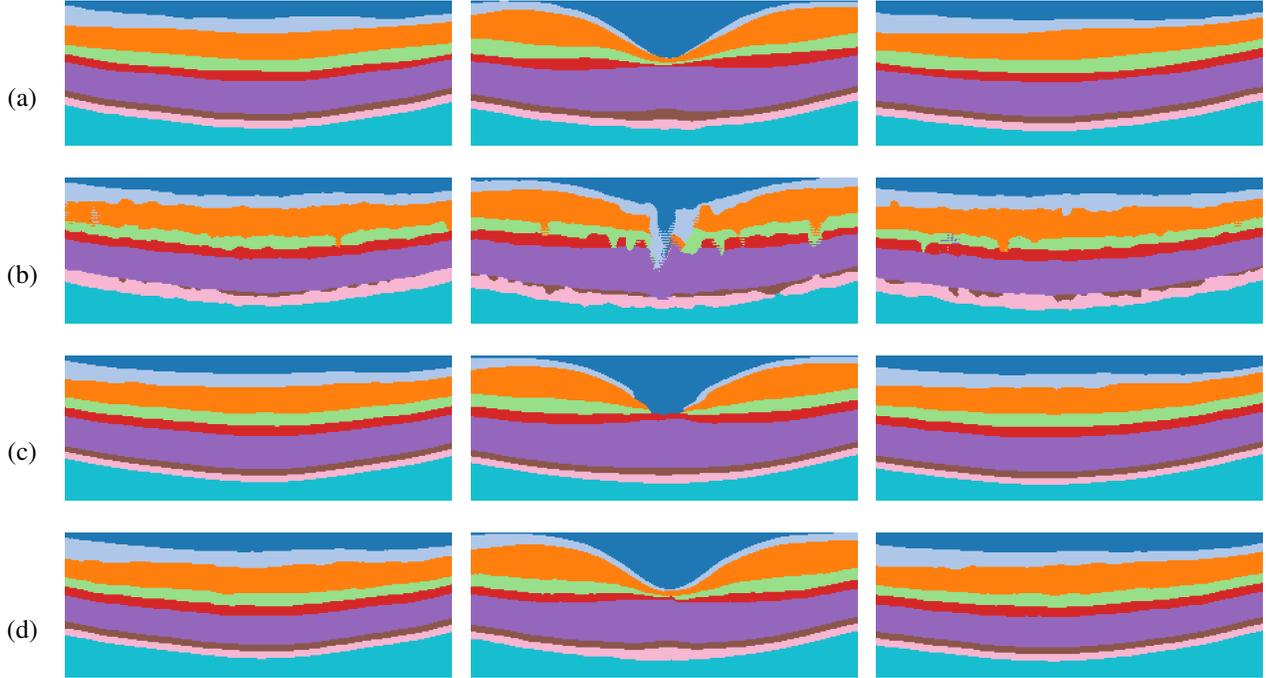
FIGURE 5.10. **From top to bottom: Row (a)**: Ground truth for the augmented retina layer corresponding to Table 2. **Row (b) and (c)**: Segmentation results of the OAF based on manifold valued features and on CNN features, respectively. **Row (d)**: Segmentation results achieved by the probabilistic graphical model approach [RSS14]. Both methods provide extraordinary performance for flat retina detection whereas our method is more accurate regarding the photoreceptor layers (below PR1).

However, closer inspection of the second column reveals a more accurate detection of layer thickness for the (PR2-RPE) and (INL) regions below the concave curved fovea region by using OAF(B). This is mainly due to the connectivity constraints imposed on boundary detection in [RSS14]. By contrast, our method is capable to deal with rapidly decreasing layer thickness near the fovea region, as observed for GCL and IPL layers in the middle column of Figure 5.10 after visual comparison against the manual delineations (first row). This observation is supported by missing connectivity constraints for retina boundaries of the proposed method, as opposed to the Gaussian shape prior model in [RSS14], and only relying on layer ordering entirely included within the Fisher Rao geometry that underlies the assignment manifold. Therefore, the OAF is amenable for extension to layer detection on pathological volumes with vanishing retina boundaries, as for vitreomacular traction or diabetic macular edema. Figure 5.11 additionally provides a 3D view on detected retina surfaces for each evaluated reference method used in this publication. The corresponding performance measures given in Table 2 underpin the notably higher Dice similarity for (PR2-RPE) and (INL) layers with overall accuracy (0.8606 ± 0.0706) µm and (0.8690 ± 0.0396) µm, respectively. The statistical plots for the mean average error and the Dice similarity index are given in Figures (5.4) and (5.12), clearly showing the superiority of OAF (B) with both the higher Dice index and the mean average errors for all layers. In particular, following Table 2, small error rates are observed among all the segmented layers, except for the (ILM) boundary which is detected by all methods with high accuracy. We point out that in general our method is not limited to any number of segmented layers, if ground truth is available.

Concluding the validation, both methods accurately detect the RNFL layer width whereas for the remaining retina tissues the layer extraction with the ordered assignment flow indicates the smallest mean absolute

error supported by the highest Dice similarity index. This demonstrates the superior performance of order preserving labeling regarding accuracy and robustness, in view of segmenting the retina layer for classifying volumetric OCT data.
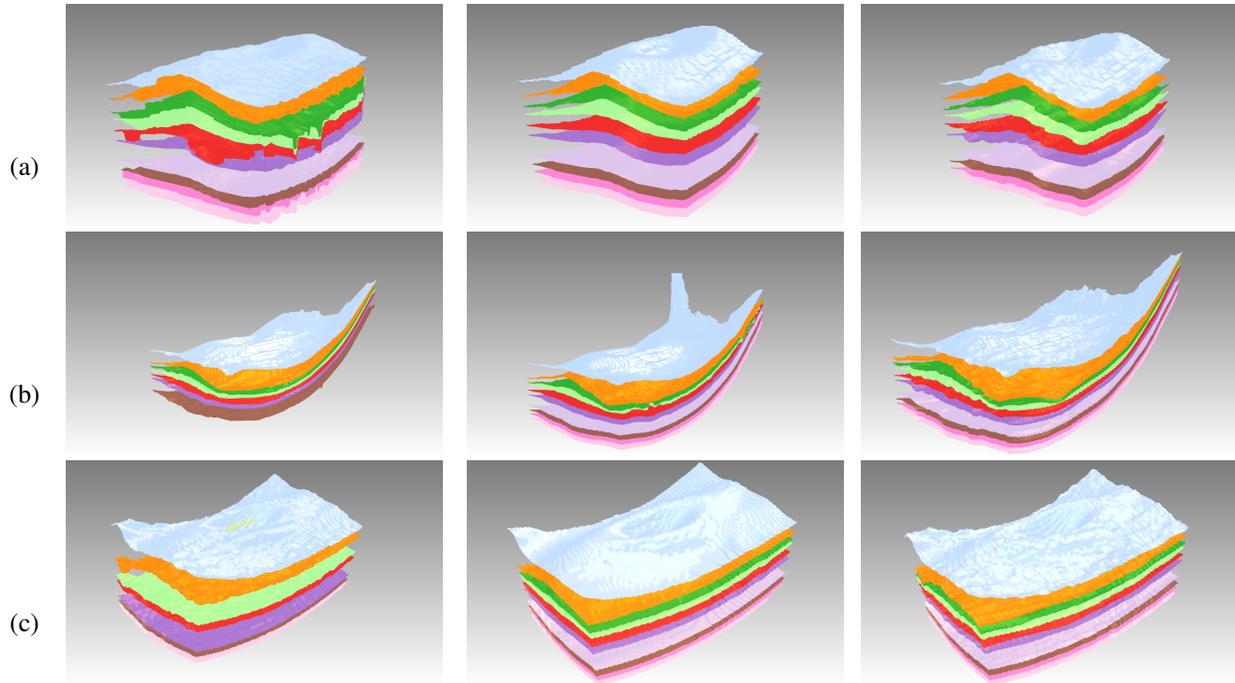

(a)

(b)

(c)

FIGURE 5.11. **Row (a)**: **From left to right:** 3D retinal surfaces determined using OAF (A) (left column) and OAF (B) (middle column). The last column depicts ground truth. **Row (b)**: **From left to right:** Segmentation of retinal tissues with the IOWA reference algorithm (left column) with the proposed approach (middle column). **Row (c):** Visual comparison of the probabilistic method [RSS14] (left column) left and the OAF (B) (middle column). Our approach OAF (B) leads to accurate retina layer segmentation with smooth layer boundaries, as observed in the middle column.
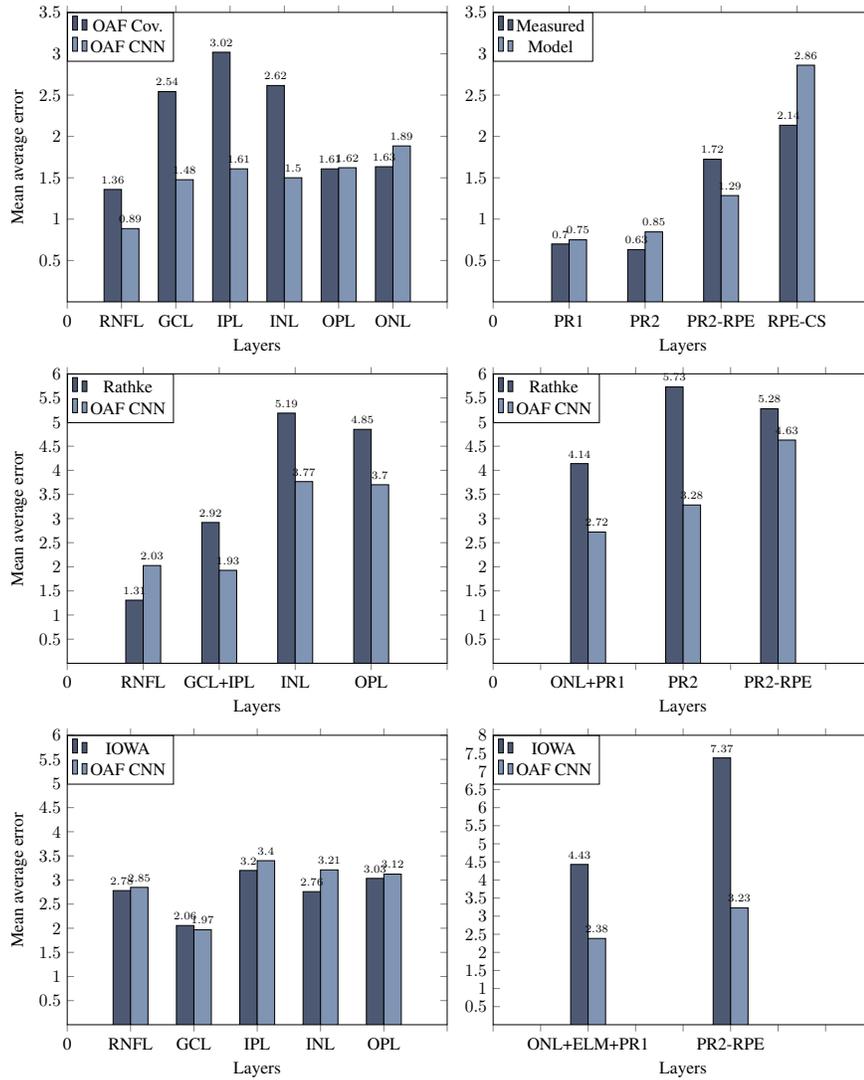
FIGURE 5.12. Performance measures per layer in terms of the mean average error based on the segmentation of 10 healthy OCT volumes. **Top row**: Error bars for retina layers separated by the external limiting membrane (ELM) corresponding to OAF (A) and OAF (B). **Middle row**: Comparison of the mean errors of OAF (B) and the probabilistic method [RSS14]. **Bottom row**: Comparison of mean average errors of OAF (B) and the the IOWA reference algorithm.

OAF Cov	DICE index	Mean absolute error	OAF CNN	DICE index	Mean absolute error
ILM	0.8837 ± 0.2564	–	ILM	0.9739 ± 0.0189	–
RNFL	0.6963 ± 0.1998	1.3590 ± 0.4114	RNFL	0.8842 ± 0.0313	0.8856 ± 0.3513
GCL	0.6657 ± 0.1909	2.5426 ± 0.7819	GCL	0.8373 ± 0.0263	1.4767 ± 0.5589
IPL	0.5853 ± 0.1773	3.0183 ± 1.0682	IPL	0.8151 ± 0.0367	1.6082 ± 1.5291
INL	0.6671 ± 0.1773	2.6160 ± 1.1294	INL	0.8414 ± 0.0035	1.5004 ± 0.8652
OPL	0.7018 ± 0.2013	1.6080 ± 0.5120	OPL	0.8442 ± 0.0437	1.6220 ± 1.0786
ONL	0.8575 ± 0.2523	1.6342 ± 0.7174	ONL	0.9254 ± 0.0486	1.8853 ± 1.3951
PR1	0.8199 ± 0.2407	0.6995 ± 0.2467	PR1	0.8717 ± 0.0441	0.7500 ± 0.3216
PR2	0.6787 ± 0.1976	0.6320 ± 0.2442	PR2	0.8330 ± 0.0516	0.8458 ± 0.4914
PR2-RPE	0.6313 ± 0.1821	1.7244 ± 0.6038	PR2-RPE	0.8213 ± 0.0835	1.2850 ± 1.3660
RPE-CS	0.8606 ± 0.2469	2.1354 ± 1.0836	RPE-CS	0.9445 ± 0.0488	2.8613 ± 2.5612

TABLE 1. Mean and standard deviations of the Dice index and mean absolute errors in pixels (1 pixel = 3.87 μm). **Left**: Errors of the proposed approach based on covariance descriptors (OAF (A)). **Right**: Errors of the proposed approach based on CNN features (OAF (B)).

[RSS14]	DICE index	Mean absolute error	OAF CNN	DICE index	Mean absolute error
ILM	0.9972 ± 0.0006	–	ILM	0.9953 ± 0.0011	–
RNFL	0.8841 ± 0.0125	1.3080 ± 0.6039	RNFL	0.8954 ± 0.0208	2.0256 ± 0.7660
GCL+IPL	0.8735 ± 0.0152	2.9180 ± 1.0303	GCL+IPL	0.9250 ± 0.0180	1.9267 ± 0.7975
INL	0.7501 ± 0.0292	5.1853 ± 1.3642	INL	0.8690 ± 0.0396	3.7660 ± 2.3101
OPL	0.7651 ± 0.0124	4.8489 ± 1.5898	OPL	0.8680 ± 0.0048	3.7010 ± 2.3561
ONL+PR1	0.9312 ± 0.0068	4.1490 ± 1.2310	ONL+PR1	0.9485 ± 0.0622	2.7209 ± 2.3594
PR2	0.7416 ± 0.0395	5.7281 ± 1.5411	PR2	0.8647 ± 0.0592	3.2810 ± 2.0854
PR2-RPE	0.7945 ± 0.0271	5.2757 ± 1.6452	PR2+RPE	0.8606 ± 0.0706	4.6270 ± 3.1891
RPE-CS	0.9858 ± 0.0073	–	RPE-CS	0.9743 ± 0.0484	–

TABLE 2. Mean and standard deviations of the Dice index and mean absolute errors in pixels (1 pixel = 3.87 μm). **Left**: Errors of the probabilistic approach [RSS14]. **Right**: Errors of the proposed approach OAF (B). These numbers demonstrate the superior performance of our novel order-preserving labeling approach.

IOWA	DICE index	MAE	OAF CNN	DICE index	MAE
ILM	0.9837 ± 0.0043	–	ILM	0.9795 ± 0.0130	–
RNFL	0.8323 ± 0.0236	2.7799 ± 0.9485	RNFL	0.8717 ± 0.0277	2.8470 ± 1.0758
GCL	0.7757 ± 0.0334	2.0561 ± 0.4978	GCL	0.8546 ± 0.0281	1.9683 ± 0.6678
IPL	0.7860 ± 0.0189	3.1970 ± 1.1408	IPL	0.8370 ± 0.0313	3.4000 ± 1.5535
INL	0.8434 ± 0.0269	2.7583 ± 1.3776	INL	0.8587 ± 0.0320	3.2094 ± 1.6950
OPL	0.8024 ± 0.0311	3.0330 ± 1.2837	OPL	0.8613 ± 0.0387	3.1217 ± 1.7875
ONL+ELM+PR1	0.8893 ± 0.0182	4.4292 ± 1.5052	ONL+ELM+PR1	0.9482 ± 0.0465	2.3842 ± 1.6869
PR2-RPE	0.7120 ± 0.0756	7.3738 ± 3.2031	PR2+RPE	0.9021 ± 0.0648	3.2296 ± 1.9627
RPE-CS	0.9667 ± 0.0167	–	RPE-CS	0.9605 ± 0.0362	–

TABLE 3. Mean and standard deviations of the Dice index and mean absolute errors in pixels (1 pixel = 3.87 μm). **Left**: Errors of the IOWA reference algorithm [KXCS06]. **Right**: Errors of the proposed approach OAF (B). These numbers demonstrate the superior performance of our novel order-preserving labeling approach.

6. Discussion

We discuss additional aspects pertaining to the data used for training feature extractors as well as the locality of extracted features.

6.1. **Ground Truth Generation.** The training and evaluation of supervised models for feature extraction requires a sizeable amount of high-quality labeled ground truth data. This presents a commonly encountered challenge in 3D OCT segmentation [DCA$^+$13, KXCS06], because the process of manually labeling every voxel of a 3D volume is extremely laborious. The desire to account for inter-observer variability in manual segmentations further compounds this problem. OCT volumes used for testing purposes in the present paper were initially segmented by an automatic procedure based on hand-crafted features. In a subsequent step, each B-scan segmentation was manually corrected by a medical practitioner. The automatic method used for initial segmentation only explicitly regularizes on each individual B-scan, leading to irregularity between consecutive B-scans (see Figure 6.1). Manual correction of initial automatic segmentations leads to
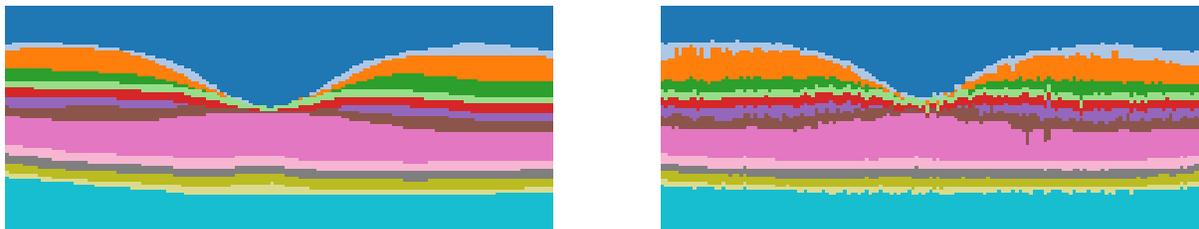

FIGURE 6.1. **Left**: Initial automatic segmentation of individual B-scan based on hand-crafted features. **Right**: Section of the same automatically segmented volume orthogonal to each B-scan.

a noticeable reduction of irregularity but does not completely remove it. We therefore cannot rule out that a small bias towards the initial automatic segmentation based on hand-crafted features may still be present in the ground truth segmentations that we used to quantify segmentation performance of novel methods as well as baseline methods in this paper. During feature extraction, deep learning models may be capable of discovering the specific hand-crafted features used for initial automated segmentation which may in turn lead to exploitation of any bias towards them. In contrast, because the reference methods are not trained on the same data, they can not exploit any such bias, putting them at a possible disadvantage.

6.2. **Feature Locality.** The ordered assignment flow segmentation approach can work with data from any metric space and is hence completely agnostic to the choice of preliminary feature extraction method. In this paper, we chose to limit the field of view of deep networks such that features with local discriminative information are extracted. This makes empirical results directly comparable between features based on covariance descriptors and features extracted by these networks. In addition, we conjecture that local features may generalize better to unseen pathologies. Specifically, if a pathological change in retinal appearance pertains to the global shape of cell layers, local features are largely uneffected and therefore significant for detecting irregularly shaped retina boundaries on unhealthy OCT data. In this way, we expect segmentation performance to be relatively consistent on real-world data. Conversely, widening the field of view in feature extraction should be accompanied by a well-considered training procedure in order to achieve similar generalization behavior, by employing e.g. extensive data augmentation. While raw OCT volume data has become relatively plentiful in clinical settings, large volume datasets with high-quality gold-standard segmentation are not widely available at the time of writing. Therefore, by representing a given OCT scan *locally* as opposed to incorporating global context at every stage, it is our next hypothesis that superior generalization can be achieved in the face of limited data availability. Similarly, although based on local features, the method

proposed by [RSS14] combines local knowledge in accordance with a *global* shape prior. This again limits the methods ability to generalize to unseen data if large deviation from the expected global shape seen in training is present.

7. Conclusion

In this paper we presented a novel, fully automated and purely data driven approach for retina segmentation in OCT-volumes. Compared to methods [KXCS06] [DCA$^+$13] and [RSS14] that have proven to be particularly effective on tissue classification with a priory known retina shape orientation, our ansatz merely relies on local features and yields ordered labelings which are directly enforced through the underlying geometry of statistical manifold (2.15). To address the task of leveraging 3D-texture information, we proposed two different feature selection processes by means of region covariance descriptors (5.11) and the output obtained by training a CNN network (5.2.3), which are both based on the interaction between local feature responses.

As opposed to other machine learning methods developed for segmenting human retina from volumetric OCT data, the proposed method only takes the pairwise distance between voxels and prototypes (2.1b) as input. As a direct consequence our approach can be applied in connection with broader range of features living in any metric space and additionally provides the incorporation of outputs from trained neuronal convolution networks interpreted as image features, where a particular instance of such type was demonstrated in Section (5.2.3). Even in view of the moderate result achieved after segmentation using OAF (A) in connection with covariance descriptors, we observe the importance of our automatic algorithm by its high level of regularization. Compared to the approach presented in [CAM$^+$15] which employs a higher number of input features but still requires postprocessing steps to yield order preserving labeling, our approach provides a way to perform this tasks simultaneously.

Using locally adapted features for handling volumetric OCT data sets from patients with observable pathological retina changes is in particular valuable to suppress wrong layer boundaries predictions caused by prior assumptions on retinal layer thicknesses typically made by graphical model approaches as in [DCA$^+$13] and [SBG$^+$13]. Our method overcomes this limitation by mainly avoiding any bias towards using priors to global retina shape and instead only relies on the natural biological layer ordering, which is accomplished by restricting the assignment manifold to probabilities that satisfy the ordering constraint presented in Section (4). The experimental results reported in Section 5, and the direct comparison to the state of the art segmentation techniques [GAS] and [RSS14] by using common validation metrics, underpin a notable performance and robustness of the geometric segmentation approach introduced in Section 2, that we extended to order-preserving labeling in Section 4. Furthermore, the results indicate that the ordered assignment flow successfully tackles problems in the field of retinal tissue classification on 3D-OCT data which are typically corrupted by speckle noise, with achieved performance comparable to manual graders which makes it to a method of choice for medical image applications and extensions therein. We point out that our approach consequently differs from common deep learning methods which explicitly aim to incorporate global context into the feature extraction process. In particular, throughout the experiments we observed higher regularization resulting in smoother transitions of layer boundaries along the B-scan acquisition axis similar to the effect in [RSS14] where the used smooth global Gaussian prior leads to limitations for pathological applications.

To reduce the reliance of manually segmented ground truth for extracting dictionaries of prototypes, our method can easily be extended to unsupervised scenarios in the context of [ZZPS20]. To deal with highly variable layer boundaries another possible extension of our method is to predict weights for geometric averaging (2.22) in an optimal control theoretic way, to cope with the linearized dynamics of the assignment flow [ZSPS20] as in detail elaborated in [HSPS20]. Consequently, by building on the feasible concept of spatially regularized assignments [Sch20], the ordered flow (4.5) possesses the potential to be extended towards the

detection of pathological retina changes and vascular vessel structure. We expect that the joint interaction of retina tissues and blood vessels during the segmentation with the assignment flow will lead to more effective layer detection, which is the objective of our current research.

ORCID IDS

Dmitrij Sitenko ⓘ 0000-0002-0022-3891
Bastian Boll ⓘ 0000-0002-3490-3350
Christoph Schnörr ⓘ 0000-0002-8999-2338

REFERENCES

[AAL+10]   B. Antony, M. Abramoff, K. Lee, P. Sonkova, P. Gupta, Y. Kwon, M. Niemeijer, Z. Hu, and M. Garvin, *Automated 3-D Segmentation of Intraretinal Layers from Optic Nerve Head Optical Coherence Tomography Images*, Progress in Biomedical Optics and Imaging - Proc.SPIE **7626** (2010), 249 – 260.

[AFPA07]   V.. Arsigny, P.. Fillard, X. Pennec, and N. Ayache, *Geometric Means in a Novel Vector Space Structure on Symmetric Positive Definite Matrices*, SIAM Journal on Matrix Analysis and Applications **29** (2007), no. 1, 328–347.

[AGS10]   M. D. Abràmoff, M. K. Garvin, and M. Sonka, *Retinal Imaging and Image Analysis*, IEEE Reviews in Biomedical Engineering **3** (2010), 169–208.

[AN00]   S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, Amer. Math. Soc. and Oxford Univ. Press, 2000.

[ÅPSS17]   F. Åström, S. Petra, B. Schmitzer, and C. Schnörr, *Image Labeling by Assignment*, Journal of Mathematical Imaging and Vision **58** (2017), no. 2, 211–238.

[BB97]   H. H. Bauschke and J. M. Borwein, *Legendre Functions and the Method of Random Bregman Projections*, J. Convex Analysis **4** (1997), no. 1, 27–67.

[BH99]   M. R. Bridson and A. Häflinger, *Metric Spaces of Non-Positive Curvature*, Springer, 1999.

[Bha07]   R. Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[Bha13]   R. Bhatia, *The Riemannian Mean of Positive Matrices*, pp. 35–51, Springer Berlin Heidelberg, 2013.

[BI13]   D. A. Bini and B. Iannazzo, *Computing the Karcher Mean of Symmetric Positive Definite Matrices*, Linear Alg. Appl. **438** (2013), 1700–1710.

[CABM15]   M. Congedo, B. Afsari, A. Barachant, and M. Moakher, *Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices*, PLoS ONE **10** (2015), no. 4, e0121423.

[CAM+15]   S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, *Kernel Regression Based Segmentation of Optical Coherence Tomography Images with Diabetic Macular Edema*, Biomed. Opt. Express **6** (2015), no. 4, 1172–1194.

[CCH06]   W. R. Crum, O. Camara, and D. L. G. Hill, *Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis*, IEEE Transactions on Medical Imaging **25** (2006), no. 11, 1451–1461.

[CS16]   A. Cherian and S. Sra, *Positive Definite Matrices: Data Representation and Applications to Computer Vision*, Algorithmic Advances in Riemannian Geometry and Applications (H. Minh and V. Murino, eds.), Springer, 2016, pp. 93–114.

[CV01]   T. F. Chan and L. A. Vese, *Active Contours without Edges*, IEEE Transactions on Image Processing **10** (2001), no. 2, 266–277.

[CZ97]   Y. A. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford Univ. Press, New York, 1997.

[DCA+13]   P. A. Dufour, L. Ceklic, H. Abdillahi, S. Schroder, S. De Dzanet, U. Wolf-Schnurrbusch, and J. Kowal, *Graph-Based Multi-Surface Segmentation of OCT Data Using Trained Hard and Soft Constraints*, IEEE Transactions on Medical Imaging **32** (2013), no. 3, 531–543.

[DFVDVM14]   A. Depeursinge, A. R. Foncubierta, D. Van De Ville, and H. Müller, *Three-Dimensional Solid Texture Analysis in Biomedical Imaging: Review and Opportunities*, Medical Image Analysis **18** (2014), no. 1, 176 – 196.

[Dic45]   L. R. Dice, *Measures of the Amount of Ecologic Association Between Species*, Ecology **26** (1945), no. 3, 297–302.

[DTG⁺15]   J. Duan, C. Tench, I. Gottlob, F. Proudlock, and L. Bai, *New Variational Image Decomposition Model for Simultaneously Denoising and Segmenting Optical Coherence Tomography Images*, Physics in Medicine and Biology **60** (2015), 8901–8922.

[FCW⁺17]   L. Fang, D. Cunefare, C. Wang, R.H. Guymer, S. Li, and S. Farsiu, *Automatic Segmentation of Nine Retinal Layer Boundaries in OCT Images of Non-Exudative AMD Patients Using Deep Learning and Graph Search*, Biomed. Opt. Express **8** (2017), no. 5, 2732–2744.

[GAS]   M. D. Garvin, M. K. Abràmoff, and M. Sonka, *The Iowa Reference Algorithms (Retinal Image Analysis Lab, Iowa Institute for Biomedical Imaging, Iowa City, IA)*, https://www.iibi.uiowa.edu/oct-reference.

[GAW⁺09]   M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, *Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images*, IEEE Transactions on Medical Imaging (2009), no. 9, 1436–1447.

[HAW⁺07]   M. Haeker, M. Abramoff, X. Wu, R. Kardon, and M. Sonka, *Use of Varying Constraints in Optimal 3-D Graph Search for Segmentation of Macular Optical Coherence Tomography Images*, MICCAI **10** (2007), 244–51.

[HCL⁺19]   Y. He, A. Carass, Y. Liu, B.M. Jedynak, S.D. Solomon, S. Saidha, P.A. Calabresi, and J.L. Prince, *Deep Learning Based Topology Guaranteed Surface and MME Segmentation of Multiple Sclerosis Subjects from Retinal OCT*, Biomed. Opt. Express **10** (2019), no. 10, 5042–5058.

[Hig08]   N.J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, 2008.

[HS87]   M. Hashimoto and J. Sklansky, *Multiple-Order Derivatives for Detecting Local Image Characteristics*, Computer Vision, Graphics, and Image Processing **39** (1987), no. 1, 28 – 55.

[HSL⁺91]   D Huang, EA Swanson, CP Lin, JS Schuman, WG Stinson, W Chang, MR Hee, T Flotte, K Gregory, CA Puliafito, and al. et, *Optical Coherence Tomography*, Science **254** (1991), no. 5035, 1178–1181.

[HSPS20]   R. Hühnerbein, F. Savarino, S. Petra, and C. Schnörr, *Learning Adaptive Regularization for Image Labeling Using Geometric Assignment*, J. Math. Imaging Vision (2020), https://doi.org/10.1007/s10851-020-00977-2.

[Jac08]   Paul Jaccard, *Nouvelles recherches sur la distribution florale*, Bulletin de la Societe Vaudoise des Sciences Naturelles **44** (1908), 223–70.

[Jos17]   J. Jost, *Riemannian Geometry and Geometric Analysis*, 7th ed., Springer-Verlag, 2017.

[Kjp]   Kjpargeter,Freepik, *The muscles of the head*, http://www.freepik.com, Accessed: 2020-09-09.

[KRAS13]   R. Kafieh, H. Rabbani, M. Abramoff, and M. Sonka, *Intra-Retinal Layer Segmentation of 3D Optical Coherence Tomography Using Coarse Grained Diffusion Map*, Medical Image Analysis **17** (2013), 907–928.

[KXCS06]   Kang, L., Xiaodong, W., D. Z. Chen, and M. Sonka, *Optimal Surface Segmentation in Volumetric Images-A Graph-Theoretic Approach*, IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006), no. 1, 119–134.

[LCF⁺19]   X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, *Semi-Supervised Automatic Segmentation of Layer and Fluid Region in Retinal Optical Coherence Tomography Images Using Adversarial Learning*, IEEE Access **7** (2019), 3046–3061.

[Lee13]   J. M. Lee, *Introduction to Smooth Manifolds*, Springer, 2013.

[Lin04]   T. Lindeberg, *Feature Detection with Automatic Scale Selection*, International Journal of Computer Vision **30** (2004), 79–116.

[MB06]   M. Moakher and P. G. Batchelor, *Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization*, Visualization and Processing of Tensor Fields, Springer, Berlin, Heidelberg, 2006, pp. 285–298.

[NVd⁺17]   J. Novosel, K. A. Vermeer, J. H. de Jong, Z. Wang, and L. J. van Vliet, *Joint Segmentation of Retinal Layers and Focal Lesions in 3-D OCT Data of Topologically Disrupted Retinas*, IEEE Transactions on Medical Imaging **36** (2017), no. 6, 1276–1286.

[PFAE06]   X. Pennec, P. Fillard, N. Ayache, and P. Epidaure, *A Riemannian Framework for Tensor Computing*, International Journal of Computer Vision **66** (2006), 41–66.

[QLD⁺10]   G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka, *Three-Dimensional Analysis of Retinal Layer Texture: Identification of Fluid-Filled Regions in SD-OCT of the Macula*, IEEE Transactions on Medical Imaging **29** (2010), no. 6, 1321–1330.

[RCK⁺17]   A.G. Roy, S. Conjeti, S.P.K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, *ReLayNet: Retinal Layer and Fluid Segmentation of Macular Optical Coherence Tomography Using Fully Convolutional Networks*, Biomed. Opt. Express **8** (2017), no. 8, 3627–3642.

[RFB15]   O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, MICCAI (Cham), Springer International Publishing, 2015, pp. 234–241.

[RSS14]   F. Rathke, S. Schmidt, and C. Schnörr, *Probabilistic Intra-Retinal Layer Segmentation in 3-D OCT Images Using Global Shape Regularization*, Medical Image Analysis **18** (2014), no. 5, 781–794.

[SBG+13] Q. Song, J. Bai, M. K. Garvin, M. Sonka, J. M. Buatti, and X. Wu, *Optimal Multiple Surface Segmentation With Shape and Context Priors*, IEEE Transactions on Medical Imaging **32** (2013), no. 2, 376–386.

[SBS20]   D. Sitenko, B. Boll, and C. Schnörr, *Assignment Flow for Order-Constrained OCT Segmentation*, GCPR, 2020.

[Sch20]   C. Schnörr, *Assignment Flows*, Variational Methods for Nonlinear Geometric Data and Applications (P. Grohs, M. Holler, and A. Weinmann, eds.), Springer, 2020, pp. 235–260.

[Sra16]   Sra, S., *Positive Definite Matrices and the S-Divergence*, Proceedings of the American Mathematical Society **144** (2016), no. 7, 2787–2797.

[SSR15]   K. Sirinukunwattana, D. R.J. Snead, and N. M. Rajpoot, *A Novel Texture Descriptor for Detection of Glandular Structures in Colon Histology Images*, Medical Imaging: Digital Pathology, vol. 9420, SPIE, 2015, pp. 186 – 194.

[TPM06]   O. Tuzel, F. Porikli, and P.: Meer, *Region Covariance: A Fast Descriptor for Detection and Classification*, Proc. ECCV, 2006, pp. 589–600.

[TS16]   P. Turaga and A. Srivastava (eds.), *Riemannian Computing in Computer Vision*, Springer, 2016.

[YHSS11] A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic, *Segmentation of Intra-Retinal Layers From Optical Coherence Tomography Images Using an Active Contour Approach*, IEEE Transactions on Medical Imaging **30** (2011), no. 2, 484–496.

[ZSPS20] A. Zeilmann, F. Savarino, S. Petra, and C. Schnörr, *Geometric Numerical Integration of the Assignment Flow*, Inverse Problems **36** (2020), no. 3, 034004 (33pp).

[ZZPS20] M. Zisler, A. Zern, S. Petra, and C. Schnörr, *Self-Assignment Flows for Unsupervised Data Labeling on Graphs*, SIAM Journal on Imaging Sciences **13** (2020), no. 3, 1113–1156.

[ZZS20]   A. Zern, A. Zeilmann, and C. Schnörr, *Assignment Flows for Data Labeling on Graphs: Convergence and Stability*, CoRR abs/2002.11571 (2020).

(D. Sitenko) IMAGE AND PATTERN ANALYSIS GROUP (IPA) AND HEIDELBERG COLLABORATORY FOR IMAGE PROCESSING (HCI), HEIDELBERG UNIVERSITY, GERMANY
  E-mail address: dmitrij.sitenko@iwr.uni-heidelberg.de

(B. Boll) IMAGE AND PATTERN ANALYSIS GROUP (IPA) AND HEIDELBERG COLLABORATORY FOR IMAGE PROCESSING (HCI), HEIDELBERG UNIVERSITY, GERMANY
  E-mail address: bastian.boll@iwr.uni-heidelberg.de

(C. Schnörr) IMAGE AND PATTERN ANALYSIS GROUP, HEIDELBERG UNIVERSITY, GERMANY
  E-mail address: schnoerr@math.uni-heidelberg.de
  URL: https://ipa.math.uni-heidelberg.de