Multi-view Learning over Retinal Thickness and Visual Sensitivity on Glaucomatous Eyes

Toshimitsu Uesaka
Grad. Sch. of Inf. Sci. & Tech., The
University of Tokyo,
Tokyo, Japan 113-8656
uesaka.toshimitu@ci.i.u-tokyo.ac.jp

Kai Morino
Institute of Industrial Science,
The University of Tokyo,
Tokyo, Japan 153-8505
morino@sat.t.u-tokyo.ac.jp

Hiroki Sugiura
Grad. Sch. of Inf. Sci. & Tech., The
University of Tokyo,
Tokyo, Japan 113-8656
sugiura.hiroki@ci.i.u-tokyo.ac.jp

Taichi Kiwaki
Grad. Sch. of Inf. Sci. & Tech., The
University of Tokyo,
Tokyo, Japan 113-8656
kiwaki@mist.i.u-tokyo.ac.jp

Hiroshi Murata
Department of Ophthalmology, The
University of Tokyo,
Tokyo, Japan 113-8655
hmurata-tky@umin.net

Ryo Asaoka
Department of Ophthalmology, The
University of Tokyo,
Tokyo, Japan 113-8655
rasaoka-tky@umin.ac.jp

Kenji Yamanishi
Grad. Sch. of Inf. Sci. & Tech., The
University of Tokyo,
Tokyo, Japan 113-8656
yamanishi@mist.i.u-tokyo.ac.jp

ABSTRACT

Dense measurements of visual-field, which is necessary to detect glaucoma, is known as very costly and labor intensive. Recently, measurement of retinal-thickness can be less costly than measurement of visual-field. Thus, it is sincerely desired that the retinal-thickness could be transformed into visual-sensitivity data somehow. In this paper, we propose two novel methods to estimate the sensitivity of the visual-field with SITA-Standard mode 10-2 resolution using retinal-thickness data measured with *optical coherence tomography* (*OCT*). The first method called *Affine-Structured Non-negative Matrix Factorization* (*ASNMF*) which is able to cope with both the estimation of visual-field and the discovery of deep glaucoma knowledge. While, the second is based on *Convolutional Neural Networks* (*CNNs*) which demonstrates very high estimation performance. These methods are kinds of multi-view learning methods because they utilize visual-field and retinal thickness data simultaneously. We experimentally tested the performance of our methods from several perspectives. We found that ASNMF worked better for relatively small data size while CNNs did for relatively large data size. In addition, some clinical knowledge are discovered via ASNMF. To the best of our knowledge, this is the first paper to address the dense estimation of the visual-field based on the retinal-thickness data.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Computing methodologies** → *Non-negative matrix factorization*; Neural networks;

KEYWORDS

Glaucoma; Retinal thickness; Non-negative matrix factorization (NMF); Convolutional neural networks (CNNs)

1 INTRODUCTION

1.1 Background of This Study

Glaucoma is an eye disease that is the second leading cause of blindness globally [14]. As the disease progresses, the thickness of the retinal layers is reduced, which eventually causes irreversible damage to the visual-field. Early detection of glaucoma requires the ability to precisely measure the visual-field. Studies have shown that dense measurements within the central 10-degree visual-field are effective at detecting glaucoma because the sensitivity of this region is closely related to eyesight quality [2, 12]. However, such dense measurements are typically conducted using a *Humphrey Field Analyzer* (*HFA*), which is very costly and labor intensive.

Meanwhile, thickness of retinal layer can be measured with *Optical Coherence Tomography* (*OCT*) [11]. This measurement is much easier and less costly than that of visual-field using a HFA. Hence it is sincerely desired that the retinal thickness data could be transformed into visual sensitivity data somehow. In the area of ophthalmology, it has turned out that retinal thickness and sensitivity of the visual-field are related each other [5, 10, 13, 35], but their quantitative relation has not yet been clarified. To the best of our knowledge, it has been an open problem how we can estimate visual sensitivity from retinal thickness. If this problem were solved, it would give a strong impact to ophthalmology.

In this paper, we present a methodology for estimating the central 10-degree visual-field from the retinal-thickness data obtained using OCT. Our proposed methodology is based on two data-mining techniques: *Affine Structured Non-negative Matrix Factorization* (*AS-NMF*), and *Convolutional Neural Networks* (*CNNs*), which are described in detail later in this paper. We offer that the ASNMF or CNNs should be employed depending on the purpose and data size. That is, ASNMF works better for relatively small data size while CNNs do for relatively large data size. Further, ASNMF is suitable for extracting features of visual patterns on glaucomatous eyes while CNNs are suitable for predicting visual sensitivity as accurately as possible.

1.2 Novelty and Significance

The novelty and significance of this paper are as follows:

(1) A novel methodology for the dense estimation of the visual-field from OCT data. The proposed methods estimate the visual-field with the resolution of the SITA-Standard mode 10-2 from OCT data, which is important for estimates that will be used at actual clinical sites. To the best of our knowledge, this is the first work that proposes a method to estimate visual filed with the resolution SITA-Standard mode 10-2 from OCT-data. Thus, our proposed framework will contribute to the improvement of visual-field prediction.

(2) The high estimation accuracy enables use in practice. To achieve high accuracy for a single patient, both the methods leverage the data from other patients to construct a model of the relationship between the OCT data and visual field. For CNNs, this relationship can be automatically detected through multilayers which are initialized with pre-learning. For larger data sets, the accuracy of the CNNs is such that it can be used in practice.

(3) The reliability of estimations is controllable. In actual clinical applications, performing an estimation with poor accuracy is much worse than doing nothing at all. In the proposed methodology, we established a criterion for determining whether the computed estimation is reliable based on classifying the data as predictable or unpredictable using a support vector machine (SVM) model.

(4) The OCT-visual field patterns for glaucomatous eyes can be successfully extracted. The ASNMF extracts the latent relationships between the visual-field and retinal-layer data from the decomposition of the ASNMF matrices which agree with the clinical knowledge.

1.3 Related Works

Many studies have been conducted to improve the prediction accuracy of glaucoma progression [9, 26]; however, most suffered from the shortage of relevant patient data due to the high cost of clinical measurements. Recent analyses have overcome this problem by leveraging similar data from other patients [8, 20, 23, 24, 32]. We also utilize similarities between the data to improve the accuracy.

Studies have also been conducted to determine if an eye was glaucomatous based on eye photographs and/or retinal-layer thickness measurements [35, 36]. Very recently, the estimation of visual-fields from the retina-layer thickness has been intensively studied. Hood et al. [10] and Ajtony et al. [1] showed that there is a statistically-significant correlation between the sensitivity of the visual-field and the retinal-layer thickness, as measured by OCT. Eura et al. [5]
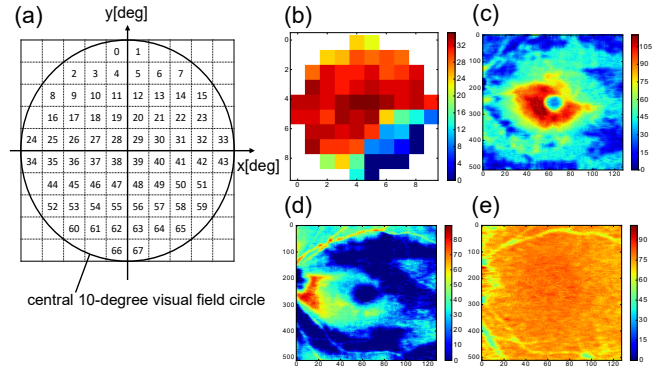

Figure 1: (a) Visual-field distribution. Examples of (b) TH, (c) GCIPL, (d) RNFL, and (e) RCL data.

employed multiple linear regressions to demonstrate that the averaged thickness of the *ganglion cell and inner plexiform layer* (GCIPL) on the macula is related to the average sensitivity of the visual field. However, these previous works could not estimate the sensitivity at a specific local visual-field point nor infer the interaction between the visual-field and retinal-layer thickness. Our proposed method overcomes these difficulties. To the best of our knowledge, this is the first paper that shows how to estimate the visual-field with full resolution of SITA-Standard mode 10-2. In developing our methods, we simultaneously analyzed the data in different domains. For this reason, our methodology can be considered as a type of multi-view learning technique [34].

1.4 Organization of This Paper

The remainder of this paper is organized as follows. Section 2 introduces the glaucoma dataset. We outline our proposed methods in Sections 3 to 5. Section 6 discusses control of the prediction accuracy. Section 7 provides an overview of how to obtain knowledge of glaucoma. In Section 8, the results of four numerical experiments are presented. Finally, the paper is concluded in Section 9.

2 GLAUCOMA DATA

2.1 Data of Visual Field

In our dataset, the state of the visual-field is represented by an integer value *TH*, which was measured using a Humphrey Field Analyzer (HFA) and SITA-Standard mode 10-2. The range of TH is within [0, 40]. When the photosensitivity of a local point in the visual-field increases, the value of TH also increases. See [28] for the details of the SITA-Standard.

Our TH dataset was measured at 68 points within the central 10-degree visual-field region, and the points were located at 2-degree intervals. Its distribution for the left eye is illustrated in Fig. 1(a). Images of the right eye were horizontally flipped for analysis using the same method as that for the left. This measurement is much denser than that for the usual visual-field test settings, such as 24-2 (central 24-degree) and 30-2 (central 30-degree). Our dataset is therefore a better representation of the fields that are critical to eyesight. An example of real TH data is shown in Fig. 1(b):

Raw Data	Method	Prediction		Knowledge Discovery
		Small Data	Large Data	
TH GCIPL RNFL RCL	ASNMF	Very good	Good	Very good
	CNN	Good	Very good	X

Figure 2: Comparison ASNMF with CNNs for prediction accuracy and knowledge-discovery ability.

temporal and nasal area is located on the left- and right-hand sides, respectively. In this figure, defects are present at the lower-right visual field.

2.2 Data of the Retinal Layers

Our retina-thickness data references the *Retinal Nerve Fiber Layer* (*RNFL*), *Ganglion Cell-inner Plexiform Layer* (*GCIPL*), and *Rod and Cone Layer* (*RCL*). This data was measured using *Optical Coherence Tomography* (*OCT*) [11] as pictures with 512×128 pixels. Our data was focused around the macula because it is the central region and dominates the function of the eyes. Examples of the GCIPL, RNFL, and RCL figures are shown in Fig. 1. In glaucomatous eyes, death of retinal cells causes the retinal-layers to be thinner. The relationship between the retinal-layers thickness and the visual-field photosensitivity was studied in [2, 10, 13].

3 OVERVIEW OF NOVEL METHODS

The proposed two methods are introduced in the following sections. The comparison with ASNMF and CNNs is summarized in Fig. 2. As in Sec. 4, the ASNMF has an elaborated structure consisting of both common and private features and of parts to exclude biases. Because of this elaborated framework, ASNMF can provide good performance in cases where there is a relatively small amount of data. As in Sec. 5, CNNs have a huge number of parameters to be learned, which a huge number of data is required to learn. Thus, pre-learning is essential for our glaucoma dataset. This is why CNNs demonstrate good performance for cases in which there is a relatively large dataset.

4 NMF-BASED PROPOSED METHOD

4.1 Background of Our Proposed Method

The first proposed method is a type of *Non-negative Matrix Factorization* (*NMF*) in which a matrix $X \in \mathbb{R}^{n \times d}$ is decomposed into the product of two matrices Y and Z by solving the following optimization problem:

$$\min_{Y,Z} ||X - YZ||_2^2 \quad \text{s.t.} \quad Y \geq O \quad \text{and} \quad Z \geq O, \quad (1)$$

where $Y \in \mathbb{R}_{\geq 0}^{n \times k}$ and $Z \in \mathbb{R}_{\geq 0}^{k \times d}$, and the dimension k is a hyper parameter which should be appropriately determined. The elements of the matrix O are all 0. This technique has been often used for extracting knowledge from data because we can simultaneously find the latent components of the data in Z and their contribution ratio in Y. We can find knowledge from Y and Z because a vector

of X can be seen as the weighted sum of vectors of Z. In addition, the non-negative restriction means that the matrices are more realistic, which enables users to easily understand the meaning of the decomposition matrices. To more effectively extract knowledge, many variants of NMF have been proposed [33].

One variant is called *Structured NMF* (*SNMF*). SNMF is a kind of NMF with structured decomposition, e.g., [18]. This framework is well suited to data composed of multiple domains, i.e., common features among domains and private features of each domain are described by embedding a structure into the matrices of Eq. (1), as follows:

$$X := (X_1 \ X_2), Y := (Y_{\text{cmm}} \ Y_{\text{pri}}), Z := \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & O \end{pmatrix}. \quad (2)$$

We can see that Y_{pri} affects X_1 but not X_2 because O is embedded in Z. Thus, SNMF can detect the common and private relationships among multiple domains.

Another variant is *Affine* NMF (*ANMF*) [17], which is suited for the analysis of data that includes biases. In ANMF, biases can be excluded by modifying the matrices of Eq. (1) as follows:

$$Y := \begin{pmatrix} 1 & Y_1 \end{pmatrix}, Z := \begin{pmatrix} \mathbf{b}^\top \\ Z_1 \end{pmatrix}, \quad (3)$$

where $\mathbf{1}$ is a vector whose components are all one. We find that X can be approximated as $\mathbf{1}\mathbf{b}^\top + Y_1 Z_1$. Thus, ANMF can be used to remove the bias effects from the decomposed matrices Y and Z.

Our proposed method is based on the above characteristics of both SNMF and ANMF and is named the *Affine Structured Non-negative Matrix Factorization* (*ASNMF*) in this paper. Our glaucoma dataset is constructed of TH, RNFL, and GCIPL data, and is considered to be multi-domain data. In addition, the structures of the eyes are similar between humans, so this dataset is considered to include biases. To discover deep knowledge from the glaucoma dataset, we embedded the SNMF and ANMF architectures directly into the prediction model. However, from the viewpoint of prediction accuracy, strongly-structural models can sometimes reduce their power of expression. In the case of ASNMF, non-negativity restricts the representability of the reconstruction. For example, the method often predicts visual-field points with zero TH values as positive TH values, because all the corresponding components of decomposed matrices must be zero to estimate the zero TH value correctly. So, the estimated TH values tend to be larger than the real TH values in this case. The difficulty of estimation also happens in the case of large TH values. These tendency can be observed from training data, thus, we alleviate such defects by utilizing the residual of the training data in this paper. See Fig. 3(a) for the flow of ASNMF.

4.2 Generating Low-dimensional Features

When the NMF-family of methods is applied to raw glaucoma data, the decomposed matrices are difficult to understand because the raw data is very high-dimensional. Before analyzing data, we must convert the data into low-dimensional features in the three domains (see Sec. 2).

As shown in Fig. 3(b), for each domain, we separately obtain low-dimensional features using the normal NMF defined in Eq. (1). A raw data matrix of TH-data is defined as $X^{(\text{TH})} \in \mathbb{R}^{N_{\text{TH}} \times d_{\text{TH}}}$ whose
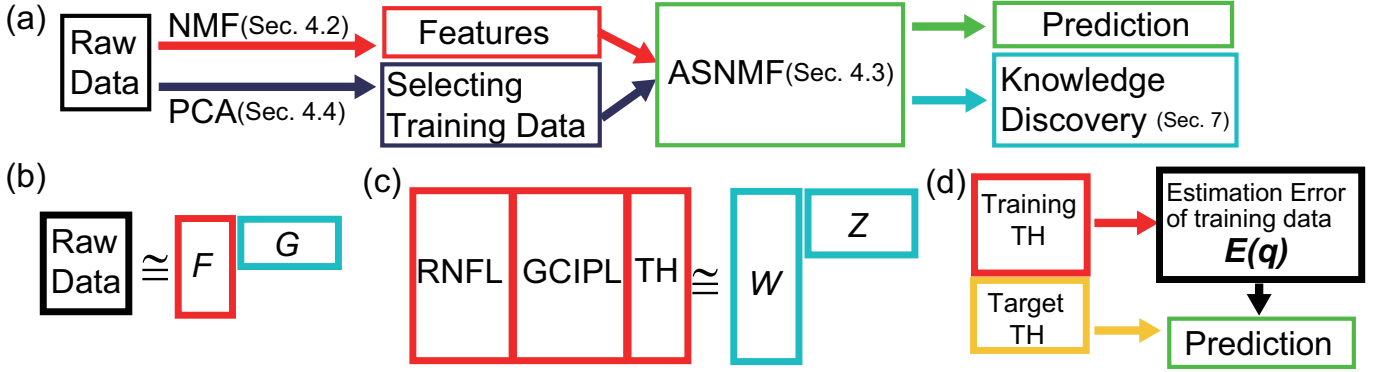
Figure 3: (a) The flow of ASNMF. Clinical knowledges are discovered in light-blue rectangles of (b) and (c). (b) Features F and bases G are generated with NMF. (c) Matrices W and Z are generated with ASNMF. (d) Estimation error of training data $E(q)$ enhances the prediction accuracy of the target TH.

ith row vector $\mathbf{x}_i^\top$ is the TH-data of the ith eye. Note that N_{TH} is the number of eyes and d_{TH} is the dimension of the data. We decompose $X^{(\text{TH})}$ into a product $F^{(\text{TH})}G^{(\text{TH})}$, where $F^{(\text{TH})} \in \mathbb{R}_{\geq 0}^{N_{\text{TH}} \times k_{\text{TH}}}$ and $G \in \mathbb{R}_{\geq 0}^{k_{\text{TH}} \times d_{\text{TH}}}$. A row vector of $G^{(\text{TH})}$ provides the characteristic base of the data. The ith row vector of $F^{(\text{TH})}$ denoted by $\mathbf{f}_i^{(\text{TH})\top}$ indicates the contribution ratio of the bases for the ith eye. Because $F^{(\text{TH})}$ is low-dimensional compared to the raw data, we analyze $F^{(\text{TH})}$ with ASNMF instead of the raw data. We can obtain the original-dimensional data from the product $\mathbf{f}_i^{(\text{TH})\top} G$. For RNFL and GCIPL, we regard these 2-dim. images as 1-dim. vectors. Therefore, we can obtain the features of RNFL and GCIPL in the same manner. Note that the index TH of the above variables is replaced by RNFL and GCIPL, respectively. We define k_{sum} as $k_{\text{TH}} + k_{\text{RNFL}} + k_{\text{GCIPL}}$.

Finally, we obtain a feature matrix $F \in \mathbb{R}^{N \times k_{\text{sum}}}$ into which the ASNMF is applied. The ith row vector of F is denoted by $\mathbf{f}_i^\top$, which is defined as $\mathbf{f}_i^\top := (\mathbf{f}_i^{(\text{RNFL})\top}\ \mathbf{f}_i^{(\text{GCIPL})\top}\ \mathbf{f}_i^{(\text{TH})\top})$ corresponding to the ith eye-data. The ith row ($i = 1, \ldots, N - 1$) corresponds to the training eye-data and the Nth row to the target eye. Because the TH of the target eye is unknown, the corresponding part $\mathbf{f}_N^{(\text{TH})\top}$ is left blank.

4.3 Proposed ASNMF

Our ASNMF decomposes the feature matrix F into a product of W and Z by solving the following optimization problem:

$$\min_{W,Z} ||(F - WZ) \odot \bar{M}||_2^2 \ \text{ s.t. } W \geq O \text{ and } Z \geq O, \quad (4)$$

where $\bar{M}$ is a mask matrix, which is defined as $\bar{M}_{ij} = 0$ if F_{ij} is blank and $\bar{M}_{ij} = 1$ otherwise. Note that $(A \odot B)_{ij} := A_{ij}B_{ij}$. See Fig. 3(c) about this decomposition.

Based on the architecture of SNMF and ANMF, we define the matrix $W \in \mathbb{R}_{\geq 0}^{N \times (l_{\text{sum}}+1)}$ as:

$$W := [\mathbf{1}\ W_{\text{cmn}}\ W_{\text{RNFL}}\ W_{\text{GCIPL}}], \quad (5)$$

where $W_{\text{cmn}} \in \mathbb{R}_{\geq 0}^{N \times l_{\text{cmn}}}, W_{\text{RNFL}} \in \mathbb{R}_{\geq 0}^{N \times l_{\text{RNFL}}}$, and $W_{\text{GCIPL}} \in \mathbb{R}_{\geq 0}^{N \times l_{\text{GCIPL}}}$. The ith row vector of W, denoted by $\mathbf{w}_i^\top$, indicates the contribution ratio for the ith eye. Note that l_{sum} is defined as $l_{\text{cmn}} + l_{\text{RNFL}} + l_{\text{GCIPL}}$.

The matrix $Z \in \mathbb{R}_{\geq 0}^{(l_{\text{sum}}+1) \times k_{\text{sum}}}$ is also defined as:

$$Z := \begin{pmatrix} \mathbf{b}_{\text{RNFL}}^\top & \mathbf{b}_{\text{GCIPL}}^\top & \mathbf{b}_{\text{TH}}^\top \\ Z_{\text{RNFL}} & Z_{\text{GCIPL}} & Z_{\text{TH}} \\ V_{\text{RNFL}} & O & O \\ O & V_{\text{GCIPL}} & O \end{pmatrix}, \quad (6)$$

where $\mathbf{b}_{\text{RNFL}} \in \mathbb{R}_{\geq 0}^{k_{\text{RNFL}}}, \mathbf{b}_{\text{GCIPL}} \in \mathbb{R}_{\geq 0}^{k_{\text{GCIPL}}}, \mathbf{b}_{\text{TH}} \in \mathbb{R}_{\geq 0}^{k_{\text{TH}}}, Z_{\text{RNFL}} \in \mathbb{R}_{\geq 0}^{l_{\text{cmn}} \times k_{\text{RNFL}}}, Z_{\text{GCIPL}} \in \mathbb{R}_{\geq 0}^{l_{\text{cmn}} \times k_{\text{GCIPL}}}, Z_{\text{TH}} \in \mathbb{R}_{\geq 0}^{l_{\text{cmn}} \times k_{\text{TH}}}, V_{\text{RNFL}} \in \mathbb{R}_{\geq 0}^{l_{\text{RNFL}} \times k_{\text{RNFL}}}$, and $V_{\text{GCIPL}} \in \mathbb{R}_{\geq 0}^{l_{\text{GCIPL}} \times k_{\text{GCIPL}}}$. After W and Z are obtained, we reconstruct the feature matrix $\hat{F} = WZ$ to obtain the reconstructed data $\hat{X} = \hat{F}G$.

However, as was mentioned in the last paragraph of Sec. 4.1, the reconstructed matrix $\hat{X}^{(\text{TH})}$ is not sufficient for prediction. We now demonstrate a way for alleviating the defects of strong structures coming from the above procedure by utilizing the residual of this estimation for the training data (see Fig. 3(d)). For the reconstructed matrix $\hat{X}^{(\text{TH})}$ and original data matrix $X^{(\text{TH})}$, we obtain the following function:

$$E(q) := \frac{\sum_{i=1}^{N-1} \sum_{j=1}^{d_{\text{TH}}} \{\hat{X}_{(i,j)}^{(\text{TH})} - X_{(i,j)}^{(\text{TH})}\} \delta(\hat{X}_{(i,j)}^{(\text{TH})} - q)}{\sum_{i=1}^{N-1} \sum_{j=1}^{d_{\text{TH}}} \delta(\hat{X}_{(i,j)}^{(\text{TH})} - q)}, \quad (7)$$

where $\delta(\cdot)$ is the Kronecker delta function. The value of $E(q)$ provides the averaged error of the matrix-reconstruction when the estimated value is q. Thus, the prediction target vector $\hat{\mathbf{x}}_N^{(\text{TH})\top} = \hat{\mathbf{f}}_N^{(\text{TH})\top} G^{(\text{TH})}$ can be modified by $E(q)$. The hth element of $\hat{\mathbf{x}}_N^{(\text{TH})\top}$ is described as follows:

$$\left[\hat{\mathbf{f}}_N^{(\text{TH})\top} G^{(\text{TH})}\right]_h = \left[\mathbf{w}_N^\top \begin{pmatrix} \mathbf{b}_{\text{TH}} \\ Z_{\text{TH}} \\ O \\ O \end{pmatrix} G^{(\text{TH})}\right]_h, \quad (8)$$

where $[\mathbf{a}]_i$ is the ith element of a vector $\mathbf{a}$. The prediction of the hth local visual-field point can then be written as:

$$\left[\mathbf{w}_N^\top \begin{pmatrix} \mathbf{b}_{\text{TH}} \\ Z_{\text{TH}} \\ O \\ O \end{pmatrix} G^{(\text{TH})} \right]_h - E\left(\left[\hat{\mathbf{x}}_N^{(\text{TH})\top} \right]_h \right). \tag{9}$$

We note that the basis matrix G was obtained by NMF described in Sec. 4.1. For this NMF procedure, we use TH-data except for the target eye. Our proposed ASNMF gives prediction of TH based on the above procedure using the EM algorithm.

4.4 Selecting Training Data with PCA

Disease data, including glaucoma, is often widely heterogeneous. This means that some training data is similar to the prediction-target data, while other data is completely different. Thus, we selected appropriate training data when applying ASNMF that would enhance the prediction accuracy. This approach was used in glaucoma studies [20, 23, 24, 32].

Our approach is based on a *principal component analysis* (*PCA*) of the RNFL and GCIPL raw data, as shown in Fig. 3(a). By applying PCA to the training and target data, we obtain M_{R} principal components for the RNFL and M_{G} principal components for the GCIPL. Note that these principal components are obtained from the top. From the data for each eye, we obtain the sign of these principal components and generate a label $\mathbf{s} \in \{+, -\}^{M_{\text{R}} + M_{\text{G}}}$ whose element is the sign of each principal component. This label indicates the similarity of the data, and allows us to only use the training data whose labels agree with those of the target data in subsequent procedures. When $M_{\text{R}} = M_{\text{G}} = 0$, we use all the data without selection.

4.5 Combined Features

Instead of using the features proposed in Sec. 4.2, we can use other features called *combined-features*. We first assemble a data vector $\mathbf{x}_i^\top$ in which the left part is from the RNFL, and the right part is from the GCIPL. Then, we apply the NMF method to the resulting matrix X and generate the features denoted by $\mathbf{f}_i^{(\text{cmb})}$, which is referred to as the *combined-feature* in this paper. Instead of separately using $\mathbf{f}_i^{(\text{RNFL})}$ and $\mathbf{f}_i^{(\text{GCIPL})}$, we can use $\mathbf{f}_i^{(\text{cmb})}$ as a feature of the OCT data. When the combined-feature is used, we also apply the ASNMF in a similar manner. The difference in Eq. (6) is that we use $Z_{\text{cmb}} \in \mathbb{R}^{l_{\text{cmn}} \times k_{(\text{cmb})}}$ instead of $\begin{pmatrix} Z_{\text{RNFL}} & Z_{\text{GCIPL}} \end{pmatrix}$, and $V_{\text{cmb}} \in \mathbb{R}^{l_{\text{cmb}} \times k_{(\text{cmb})}}$ instead of $\begin{pmatrix} V_{\text{RNFL}} & O \\ O & Z_{\text{GCIPL}} \end{pmatrix}$.

5 CNNS-BASED PROPOSED METHOD

In this section, we describe the second proposed method based on a *Convolutional Neural Networks* (*CNNs*) [16, 29, 31]. A CNN is a type of deep neural network with a distinct structure of alternating convolution and pooling layers. The convolution layers extract features by convolving a set of filters over the input, while the pooling layers spatially coarsen the input to make the output robust against translation or distortion in the input. This method predicts TH using CNNs that process RNFL, GCIPL, and RCL images as its input. Figure 4 outlines the proposed method, which has two parts:

The first extracts the key features of the images, while the second predicts TH from the cross-channel mixing.

5.1 Extraction of Key Features

First, 512 images of 14×14 pixels were generated from the three RNFL, GCIPL, and RCL images of 512×128 pixels. These images were flipped vertically because we need to adopt the visual-functional relationships of these data to the physical-position relationships. We then employed VGG16 [29], which has already been trained on 1.3 million images of the LSVRC14 dataset [27] for 1,000 categories of object classification. Our glaucoma dataset was too small for proper training; however, we used transfer learning [6], and initialized the network parameters based on a network that had already been trained with LSVRC14 dataset, a large dataset unrelated to the glaucoma dataset. The first part of the proposed method had an identical structure to VGG16, except that the last pooling layer, three fully connected layers, and the softmax layer have been removed. Thus, our network had 13 convolution layers, 4 max pooling layers, and 1 output layer (see Fig. 4(c)). We also used the initial network parameters from the corresponding part of VGG16. Note that the three input-images were horizontally up-sampled and vertically downsampled from 512×128 to 224×224 so that they fit the design of VGG16. Through the first convolution layer, 64 images of 224×224 pixels were generated from the simultaneously-inputted three images of 224×224 pixels (RNFL, GCIPL, and RCL). Such convolution and pooling procedures were then proceeded. In addition, the value of all image pixels that were greater than 255 were replaced with a constant 255.

5.2 Prediction of Eyesight

Second, the TH is predicted by linearly mixing the outputs of the first part $(d_{lmn}) = D \in \mathbb{R}^{512 \times 14 \times 14}$. This prediction bases on the spatial correspondence of D because clinical knowledge suggests a location-wise correlation between the retinal thickness and visual field [10]. In other words, the predicted TH $\hat{x}_{mn}$ at retinal location (m, n) is computed in a linear manner for mixing different channels as:

$$\hat{x}_{mn} = \sum_{l=1}^{512} a_{lmn} d_{lmn} + b_{mn}, \tag{10}$$

where $(a_{lmn}) = A \in \mathbb{R}^{512 \times 14 \times 14}$ and $(b_{mn}) = B \in \mathbb{R}^{14 \times 14}$. The initial parameters were determined according to Xavier's rule [7]. This architecture allows the number of parameters to be reduced compared to those of a prevalent fully connected layer, and thus inhibits the method from overfitting the small glaucomatous dataset. Note that we only made a prediction based on the inner 68 locations (see Fig. 4 for detail) out of 14×14 possible locations because the target values of TH were only available at these locations, and the values of d_{lmn} at the other locations were discarded. The proposed model can be trained on the glaucoma dataset by minimizing the square regression loss $\sum_{(m,n) \in \text{inner}} \| \hat{x}_{mn} - x_{mn}^{\text{TRUE}} \|^2$ through the back propagation technique. Note that the network was trained on centered target signals where the mean of the target signals in the training set was subtracted. The prediction from the model was acquired by adding the mean to the output of the model. The network was trained using the momentum stochastic gradient descent
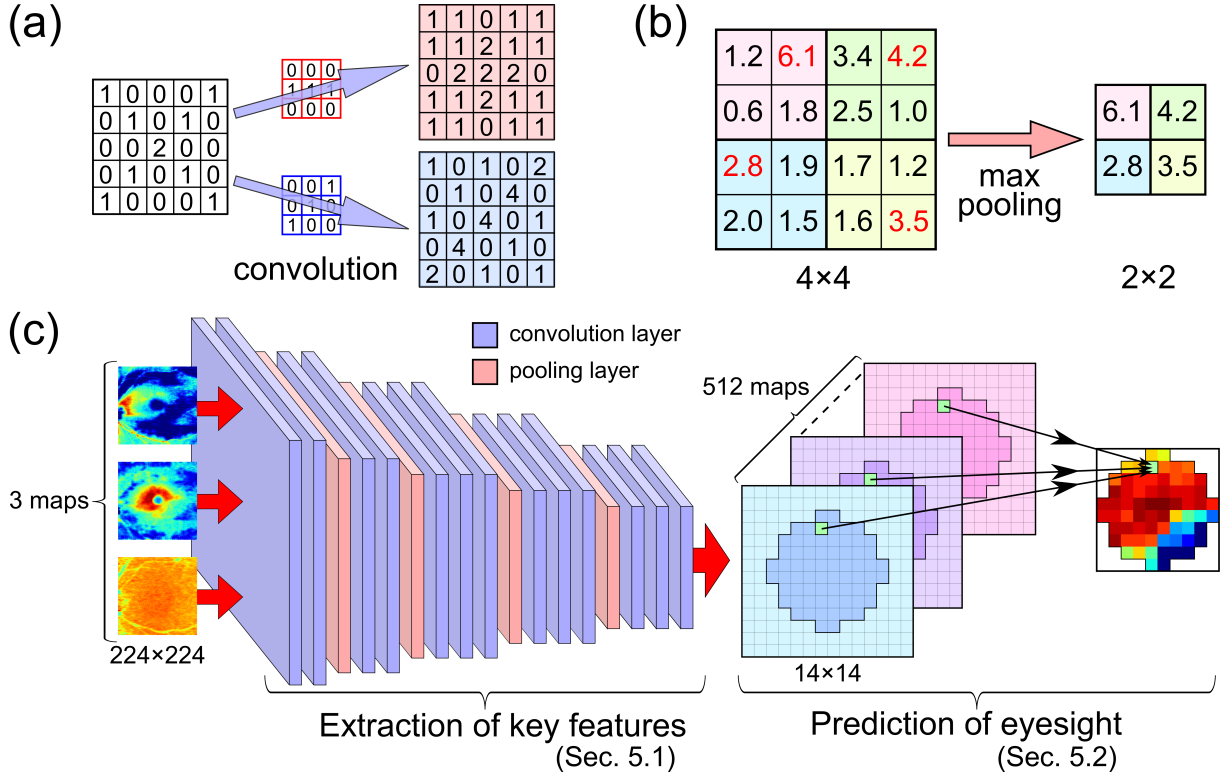
Figure 4: (a) A convolution layer extracts multiple feature maps. (b) A max pooling layer downsizes images. (c) The flow of our CNNs.

(SGD) and L2 regularization. We stopped the training process early to address overfitting, where the number of training epochs was determined via cross validation.

6 GUARANTEE ON ACCURACY

In this paper, we introduced methods to predict eyesight quality based on OCT data. It is important to consider the reliability of the prediction because, if the prediction reliability is low or the prediction error varies significantly, then the eyesight must be re-measured.

In our framework, we can guarantee the prediction accuracy by employing an *SVM*, which is a well-known classification model [4]. The features of an eye are the principal components of the RNFL and GCIPL data. The label of the eye indicates whether its prediction error is higher or lower than θ. By using an SVM with this data, we can determine whether the prediction error of the target eye is above or below θ. Thus, we are able to control the acceptable prediction error by changing a single parameter θ. To enhance the prediction accuracy, we should not make a prediction for eyes that have been deemed unpredictable. Instead, medical doctors should manually perform these measurements.

7 DISCOVERING GLAUCOMA KNOWLEDGE

Here, we show how to discover deep knowledge from the glaucoma data using ASNMF. The knowledge comes from the two matrices, namely G and Z (see Fig. 3).

The first matrix G gives bases of the raw glaucoma data (see Fig. 3(b)). These bases are obtained by NMF and shows the essential information of the raw glaucoma dataset. However, this result lacks information which describes the relationship among these pieces of glaucoma knowledge.

We overcome this problem with Z, which reconstructs the original matrix via ASNMF (see Fig. 3(c)). Our proposed model learns Z about a latent relationships among three domains: TH, RNFL, and GCIPL. Except biases $\mathbf{b}$, each row vector in Z provides a latent base of the feature matrix F. Z_{RNFL} and V_{RNFL} correspond to the reconstruction of RNFL features; Z_{GCIPL} and V_{GCIPL} correspond to the reconstruction of GCIPL-features; and Z_{TH} corresponds to the reconstruction of TH features. There is no private part of TH (so to speak, V_{TH}) in Eq. (6) to avoid the case where the solution to the optimization problem (4) does not converge. Therefore, we can find deep knowledge by mining the matrix Z.

We note that the remaining matrix W also contains rich information of individual eyes, although we did not utilize W in this paper. The matrix W shows the contribution ratio of the latent-bases of Z. The left part of $\mathbf{w}_i^\top$ corresponding to W_{cmn} is commonly used for the reconstruction of RNFL, GCIPL, and TH. The center and right

parts of those corresponding to W_{RNFL} and W_{GCIPL} are only used for the reconstruction of RNFL and GCIPL, respectively. For the ith eye, $\mathbf{w}_i^\top$ provides a connection between the view-field and OCT data. Therefore, in future, it should be discussed the role of W in discovering knowledge of glaucoma.

8 NUMERICAL EXPERIMENTS

In this section, we describe the experiments conducted to evaluate the proposed methods.

8.1 Details of Dataset

The dataset used in this paper was provided by the Department of Ophthalmology at The University of Tokyo. The dataset consisted of (1) 591 examples of eye-data that included both TH and OCT data; (2) 1,127 examples of eye-data that included only OCT data; and (3) 644 examples of eye-data that included only TH data. Note that the examples of eye-data that included either TH or OCT were only used to obtain the features listed in Sec. 4.2. The TH data was measured with a HFA using SITA-Standard mode 10-2. There TH data had $d = 68$ dimensions, each datum was within the range $[0, 40]$, the average was 18.9, and the standard deviation was 14.2. The RNFL, GCIPL, and RC data were measured using OCT within a 9 mm × 9 mm region on the macula. These examples were in $\mathbb{R}^{512\times128}$. In our dataset, for the GCIPL data, the value of each cell was within the range $[17.5, 68.8](\mu m)$ and its average and standard deviation were 41.9 μm and 8.72 μm, respectively. For the RNFL data, the range was $[5.0, 63](\mu m)$, the average was 32.8 μm, and the standard deviation was 8.5 μm. For the RCL data, the range was $[47.85, 75.52](\mu m)$, the average was 67.12 (μm), and the standard deviation was 3.757 (μm).

8.2 Experimental Setting

Here, we summarize the experimental settings. When the predicted TH was not within the defined range of TH, we modified it to be within the range: If the predicted TH was greater than 40, we set it to 40. If it was smaller than 0, we set it to 0. The prediction error of TH was evaluated using the *root mean squared error* (*RMSE*), which was defined as $\sqrt{||\mathbf{x}_{\text{true}} - \hat{\mathbf{x}}||_2^2/68}$, where $\mathbf{x}_{\text{true}}$ and $\hat{\mathbf{x}}$ were the actual and predicted TH values for the target eye, respectively. Our source codes are available at https://www.dropbox.com/sh/j twzrns8t6l6fgx/AAAbhC6iecMJvMHqFaRfjWEoa?dl=0.

The hyper parameter we used in the following experiments were the same for CNNs: We employ L2 penalty multiplier for weight decay with $5 \cdot 10^{-3}$. We also set the learning rate, the batch size, and the momentum as $3 \cdot 10^{-7}$, 40, and 0.9, respectively. See [29] for details of these parameters.

8.3 Experiment 1: Visual-field Prediction

First, we evaluated the prediction accuracy of TH using RNFL, GCIPL, and RCL data. We tested four cases: (1) ASNMF with the low-dimensional features; (2) ASNMF with combined-features; (3) CNNs; and the (4) *Multiple regression* (*MR*) method whose multiple regression coefficients are estimated between features $\mathbf{f}^{(\text{TH})}$ and features ($\mathbf{f}^{(\text{GCIPL})}$ $\mathbf{f}^{(\text{RNFL})}$).

Table 1: Predicted RMSE with whole data.

Method	Median	Mean	Max.
ASNMF with low-dim-features	7.864	7.320	22.32
ASNMF with combined-features	7.620	7.269	22.04
CNNs	6.749	6.786	23.22
MR with features	8.063	7.490	22.45

Table 2: Predicted RMSE without estimating training error.

Method	Median	Mean	Max.
ASNMF with low-dim-features	8.462	7.758	26.00
ASNMF with combined-features	8.288	7.690	28.55

We performed leave-one-out cross validation for the ASNMF and MR method, and performed a 5-fold cross validation for the CNNs. We evaluated the prediction performance by varying hyper-parameters of our methods to select best parameters. The best parameters of the ASNMF were: For low-dimensional features, $k_{\text{TH}} = k_{\text{RNFL}} = k_{\text{GCIPL}} = 8$, $l_{\text{cmn}} = 14$, $l_{\text{RNFL}} = l_{\text{GCIPL}} = 0$, $M_{\text{R}} = 2$, and $M_{\text{G}} = 0$; for combined features, $k_{\text{cmb}} = 14, k_{\text{TH}} = 6, l_{\text{cmn}} = 13, l_{\text{cmb}} = 0, M_{\text{R}} = 2, M_{\text{G}} = 0$. Those of MR were $k_{\text{TH}} = 7, k_{\text{RNFL}} = 6, k_{\text{GCIPL}} = 8$, and $M_{\text{R}} = 1$, and $M_{\text{G}} = 0$.

In Table 1, we summarize the results of prediction. We show the median, mean, and maximum values of the predicted TH values. We found that the CNNs were the best. The ASNMF using combined-features demonstrated better performance than the ASNMF with simple features; thus, the usefulness of the combined-features is evident. These results indicate the superiority of the CNNs from the viewpoint of the prediction accuracy. In the next experiment, we decreased the size of the dataset to better understand the applicability of the proposed methods.

Estimating training error based on $E(q)$ (see Eq. (7)) improved the prediction accuracy. Without applying $E(q)$, the performance of ASNMF with low-dimensional features and that of ASNMF with combined features decreased as shown in Table 2. One can find that the median of prediction error were largely decreased by applying $E(q)$. For combined-features, the best parameters used in the Table 2 experiments were $k_{\text{cmb}} = 14$, $k_{\text{TH}} = 6$, $l_{\text{cmn}} = 14$, $l_{\text{cmb}} = 0$, $M_{\text{R}} = 1$, and $M_{\text{G}} = 0$. The best parameters of this ASNMF with low-dimensional features were the same as those selected in the result of Table 1.

8.4 Experiment 2: Data Size Dependency

This experiment was to evaluate the robustness of the methods against the size of the glaucoma dataset. We generated random subsets from the glaucoma dataset and then determined the prediction accuracy for the subsets. The dataset sizes N_s were 450, 300, 200, 150, 75, and 50. The number of subsets was 12 for $N_s = 450, 300$, and 200; and is 15 for $N_s = 150, 75$, and 50. Note that the parameters for each method, e.g. the bases' dimension, intrinsic bases' dimension, M_{R}, and M_{G}, were those which provided the best performance. Appropriate parameter values may depend on individual
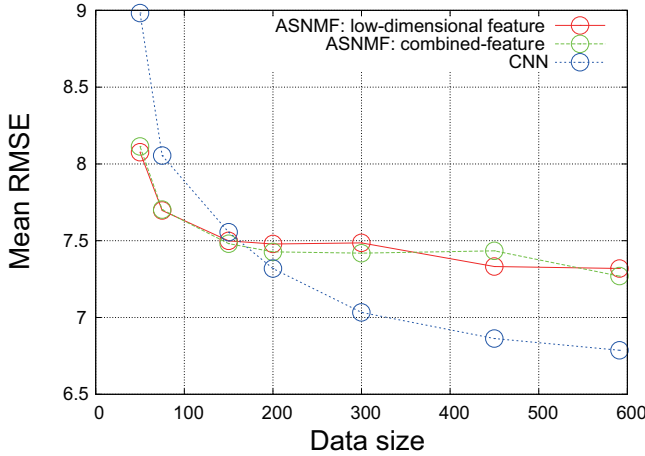
Figure 5: Dependence of the prediction accuracy, mean RMSE, on the data size.

data, therefore, it is challenging to investigate how we can choose the parameters.

The averaged RMSE was shown in Fig. 5. One can see that the performance of the CNNs was highly dependent on the data size. As the size of the dataset decreased, the averaged RMSE increased drastically. This is because CNNs often require large datasets to learn effectively. In contrast, the performance of the ASNMF did not depend on the data size so much. In the ASNMF procedure, we generated low-dimensional features and then made a prediction. Because these features can be obtained from small datasets, the performance of the ASNMF was moderate.

8.5 Experiment 3: Guarantee of Accuracy

We discuss a method for controlling the prediction accuracy using an SVM with a hyper parameter θ. In the SVM framework, the data was composed of *input-data* and *label-data*. From the input-data of our glaucoma dataset, we selected 200 of the top principal components from each the GCIPL and RNFL data. For the label-data of our glaucoma dataset, we determined whether the prediction error was under or over the threshold θ. Initially, we generated label-data by varying θ using the final results of Experiment 1 for both the ASNMF and CNNs for all 591 eye examples at once. Then, the leave-one-out procedure was done: We selected eye-data for the test data of the SVM; The rest of the data was used as training data for the SVM. After the SVM learning process was complete, we evaluated whether the test data was predictable. If it was judged to be predictable, we applied the ASNMF or CNNs to obtain the RMSE. We employed the 3rd-order polynomial kernel defined as $\langle \mathbf{x}, \mathbf{y} \rangle := (\gamma \mathbf{x}^{\top} \mathbf{y} + r)^3$ with parameters being $C = 1, \gamma = 0.001$, and $r = 0$. Experiments with these parameters showed the best performance among those we tested.

Figure 6(a) shows the average RMSE versus θ. Both the averaged RMSE of the ASNMF and CNNs are below the diagonal purple line RMSE $= \theta$. This indicates that the averaged accuracy of the predictions can be controlled within the desired θ. Figure 6(b) shows the averaged RMSE versus the size of the predicted dataset. Note
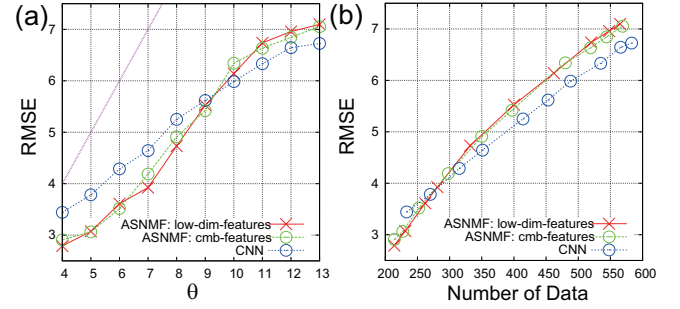

Figure 6: Prediction accuracy dependency for SVM-selection on (a) θ and on (b) selected data size.
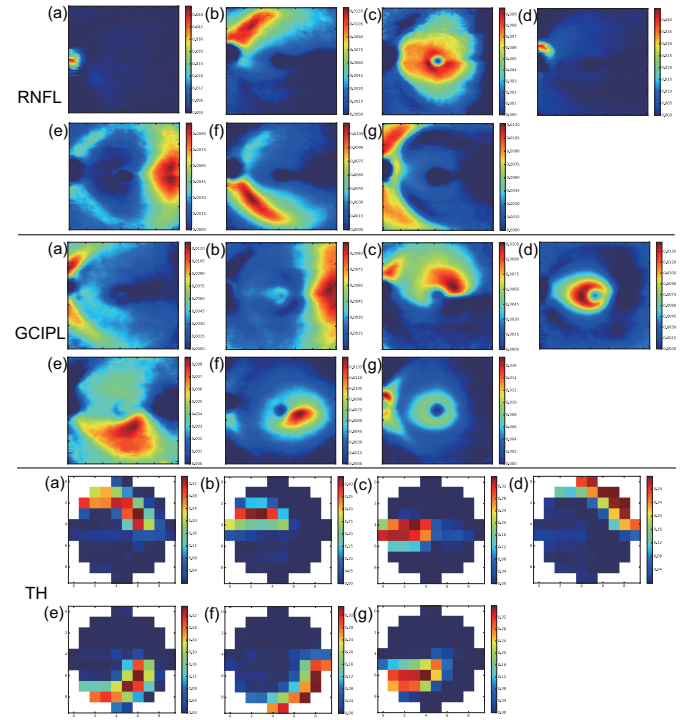

Figure 7: Bases of RNFL, GCIPL, and TH.

that the symbols were obtained by varying the hyper parameter θ. The CNNs were shown to have better performance for large data sizes, and the ASNMF had better performance for smaller data sizes. This result is similar to that obtained in Experiment 2. These results suggest possibility to detect the prediction reliability.

8.6 Experiment 4: Knowledge Discovery

Finally, we show the two types of deep knowledge obtained by mining the glaucoma dataset with ASNMF (see Sec. 7). Note that all of the data was used for ASNMF in this section, because the purpose was to mine knowledge from the complete dataset.

(1) Bases of glaucoma data. We show the figures of the normalized bases of G obtained by NMF shown in Sec. 4.2 in Fig. 7. The dimensions of the bases were: $k_{TH} = k_{NT} = k_{GT} = 7$. For the RNFL and GCIPL bases, red indicates thick and blue indicates thin. For the TH bases, red indicates good sensitivity of visual-field and blue indicates bad one. Original images were described as weighted average of these bases. These bases showed the characteristic damage patterns on visual-field and retinal-layers because defects were expressed by small coefficients of the bases. We describe examples of clinical correspondence of these patterns. For the RNFL bases, the arcuate scotoma in Bjerrum area were found in (b), (f), and (g); damages on nasal step in (e); the arcuate scotoma in Bjerrum area and in the nasal step in (c). For the GCIPL bases, the arcuate scotoma in Bjerrum area were found in (a) and (e); damages on nasal step in (b); the arcuate scotoma in Bjerrum area and in the nasal step in (c), (d), (f), and (g). For the TH bases, damages were found in superior hemifield in (a), (b), and (d); those in inferior hemifield in (c), (e), (f), and (g); the arcuate scotoma in Bjerrum area were found in (a), (b), (c), (e), and (g), whereas those in Bjerrum area and nasal step were connected in (d) and (f).

(2) Relationship between domains. Then, we focused on the biases and the latent bases of Z obtained by ASNMF in Sec. 4.3. The bases indicated the co-occurrence relationships among three domains RNFL, GCIPL, and TH. That is, the bases demonstrate how the damages of retinal-layers affected the visual-field. We show some characteristic examples whose clinical suggestion was present in Fig. 8. In (a), we show biases of ASNMF, which reflected the terminal stage of glaucomatous eyes. They were obtained by $\mathbf{b}_{RNFL}^\top G^{(RNFL)}$, $\mathbf{b}_{GCIPL}^\top G^{(GCIPL)}$, and $\mathbf{b}_{TH}^\top G^{(TH)}$, respectively. As described above, the biases was always included in reconstruction for every data. In other words, even if weights of the latent bases were all zero, the biases still remained in the reconstructed images. Thus, it is reasonable that the biases corresponded to the terminal stage. From (b) to (d), the figures show clinical knowledge that damages in lower-half (upper-half) of the retinal-layers affected the superior (inferior) hemifield of the visual-field. The figures corresponding to RNFL were obtained by $Z_{RNFL}^{(i)\top} G^{(RNFL)}$, where $Z_{RNFL}^{(i)\top}$ is the transposed ith row vector of Z_{RNFL}. The figures corresponding to GCIPL and TH were also obtained by $Z_{GCIPL}^{(i)\top} G^{(GCIPL)}$ and $Z_{TH}^{(i)\top} G^{(TH)}$, respectively. These examples suggest that the latent bases of our proposed AS-NMF includes deep knowledge of glaucoma. In addition, the matrix W contains characteristics of individual eyes. Therefore, to discover deeper knowledge of glaucoma, it should be focused on the matrix W in future.

9 CONCLUSION

In this paper, we proposed two methods for estimating the state of the visual-field based on the thickness of the retinal-layers: *Affine Structured Non-negative Matrix Factorization*, which is able to cope with both the estimation of the visual-field and detection of deep knowledge of glaucoma, a method based on a *Convolutional Neural Network*, which demonstrated very high estimation performance for large datasets. During the four experiments that were conducted, we evaluated the prediction performance of the proposed methods, the dependence of the prediction accuracy on the data size,
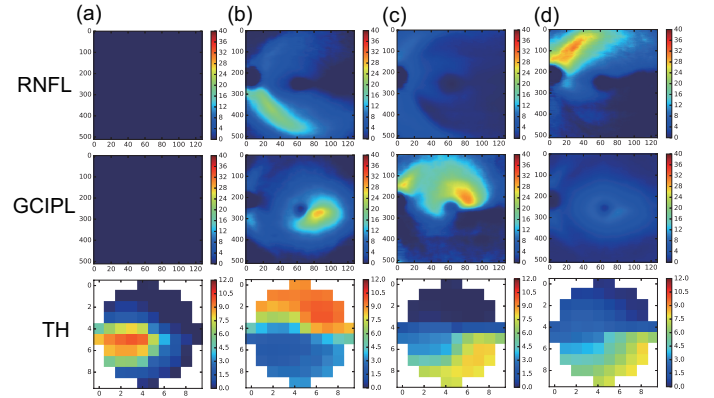

Figure 8: Relationship between the domains.

the possibility of controlling the prediction errors, and how deep knowledge of glaucoma can be found.

To the best of our knowledge, this paper is the first approach to estimate eye-sight damage with the full resolution of SITA-Standard mode 10-2. Because the measurements of the visual-field is a heavy burden on clinical sites, a method, such as the one proposed in this paper, is strongly desired to estimate the state of the visual-field from retinal-layer data measured with OCT. Based on our analysis, our proposed methods have the potential to positively impact the field of ophthalmology.

REFERENCES

[1] C. Ajtony, Z. Balla, S. Somoskeoy, and B. Kovacs. "Relationship between Visual Field Sensitivity and Retinal Nerve Fiber Layer Thickness as Measured by Optical Coherence Tomography," *Invest. Ophth. Vis. Sci.* **48**(1), p. 258, 2007.

[2] R. Asaoka. "The relationship between visual acuity and central visual field sensitivity in advanced glaucoma." *Brit. J. Ophthalmol.* **97**(10), pp. 1355-1356, 2013.

[3] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations*, John Wiley & Son, 2009.

[4] C. Cortes and V. Vapnik. "Support vector networks," *Machine Learning* **20**(3):273-297, 1995.

[5] M. Eura, "Correspondence Between Visual Field Test Results and GCL+IPL Thickness in the Maculae of Glaucomatous Eyes," Ph.D. Thesis, Kinki University, 2014 (written in Japanese).

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." in CVPR, pp. 580-587, 2014.

[7] X. Glorot and Y. Bengio. "Understanding the difficulty of training deep feedforward neural networks." In AISTATS Vol. **9**, pp. 249-256, 2010.

[8] M. Higaki, K. Morino, H. Murata, R. Asaoka, and K. Yamanishi, "Predicting Glaucoma Visual Field Loss by Hierarchically Aggregating Clustering-based Predictors," arXiv:1603.07094, 2016.

[9] C. Holmin and C. E. T. Krakau, "Regression analysis of the central visual field in chronic glaucoma cases," *Acta Ophthalmologica* **60**(2), pp. 267 – 274, 1982.

[10] D. C. Hood and R. H. Kardon. "A framework for comparing structural and functional measures of glaucomatous damage," *Progress in Retinal and Eye Research* **26**(6), pp. 688-710, 2007.

[11] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto. "Optical coherence tomography," *Science* **254**(5035), pp. 1178-1181, 1991.

[12] S. Izumi, S. Shiroaki, S. Matsumoto, and M. Araie. "The relationship between visual disability and visual field in patients with glaucoma." *Ophthalmology* **110**(2), pp. 332-339, 2003.

[13] A. Kanamori, M. Nakamura, M. F.T Escano, R. Seya, H. Maeda, and A. Negi. "Evaluation of the Glaucomatous Damage on Retinal Nerve Fiber Layer Thickness Measured by Optical Coherence Tomography." *Am. J. Ophthalmol.* **135**(4), pp. 513-520, 2003.

[14] S. Kingman. "Glaucoma is second leading cause of blindness globally," *B. World health Organ.* **82**(11), pp. 887-888, 2004.

[15] Y. Kita, R. Kita, A. Takeyama, A. Anraku, G. Tomita, and I. Goldberg. "Relationship between macular ganglion cell complex thickness and macular outer retinal thickness: a spectral-domain optical coherence tomography study," *Clinical & Experimental Ophthalmology* **41**(7), pp. 674-682, 2013.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks". In NIPS, pp. 1097-1105, 2012.

[17] H. Laurberg and L. K. Hansen. "On Affine Non-Negative Matrix Factorization," in ICASSP2007, Vol. **II**, pp. 653 - 656, 2007.

[18] H. Laurberg, M. N. Schmidt, M. G. Christensen, and S. H. Jensen, "Structured non-negative matrix factorization with sparsity patterns," in ASILOMAR2008, pp.1693-1697, 2008.

[19] D. D. Lee and H. S. Seung. "Algorithms for Non-negative Matrix Factorization," in NIPS, pp. 556-562, 2000.

[20] Z. Liang, R. Tomioka, H. Murata, R. Asaoka, and K. Yamanishi. "Quantitative Prediction of Glaucomatous Visual Field Loss from Few Measurements," in ICDM2013, pp. 1121-1126, 2013.

[21] M. Lin, Q. Chen, and S. Yan. "Network in Network." arXiv:1312.4400, 2013.

[22] J. Liu, C. Wang, J. Gao, and J. Han. "Multi-View Clustering via Joint Non-negative Matrix Factorization," in SDM2013, pp. 252-260, 2013.

[23] S. Maya, K. Morino, H. Murata, R. Asaoka, and K. Yamanishi. "Discovery of Glaucoma Progressive Patterns Using Hierarchical MDL-Based Clustering." in KDD2015, pp. 1979-1988, 2015.

[24] S. Maya, K. Morino, and K. Yamanishi, "Predicting Glaucoma Progression using Multi-task Learning with Heterogeneous Features," in IEEE BigData2014, pp. 261-270, 2014.

[25] J. Nayak, R. Acharya U., P. S. Bhat, N. Shetty, and T. C. Lim. "Automated Diagnosis of Glaucoma Using Digital Fundus Images," *Journal of Medical Systems* **33**(5), pp. 337-346, 2009.

[26] B. N. Noureddin, D. Poinoosawmy, F. W. Fietzke, and R. A. Hitchings. "Regression Analysis of Visual Field Progression in Low Tension Glaucoma". *Brit. J. Ophthalmol.* **75**(8), pp. 493-495, 1991.

[27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge." *Int. J. Comput. Vision* **115**(3), pp. 211-252, 2015.

[28] T. M. Shaarawy, M. B. Sherwood, R. A. Hitchings, and J. G. Crowston. *GLAUCOMA MEDICAL DIAGNOSIS & THERAPY* VOLUME ONE, Elsevier, 2009.

[29] K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2015.

[30] W. H. Swanson, J. Felius, and F. Pan. "Perimetric Defects and Ganglion Cell Damage: Interpreting Linear Relations Using a Two-Stage Neural Model," *Invest. Ophth. Vis. Sci.* **45**(2), p. 466, 2004.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in CVPR, pp. 1-9, 2015.

[32] K. Tomoda, K. Morino, H. Murata, R. Asaoka, and K. Yamanishi. "Predicting Glaucomatous Progression with Piecewise Regression Model from Heterogeneous Medical Data." in HEALTHINF2016, pp. 93-104, 2016.

[33] Y.-X. Wang and Y.-J. Zhang, "Nonnegative Matrix Factorization: A Comprehensive Review," *IEEE Trans. Knowl. Data Eng.* **25**(6), pp. 1336-1353, 2013.

[34] C. Xu, D. Tao, and C. Xu. "A Survey on Multi-view Learning." arXiv:1304.5634, 2013.

[35] S. Yousefi, M. H. Goldbaum, M. Balasubramanian, T. P. Jung, R. N. Weinreb, F. A. Medeiros, L. M. Zangwill, J. M. Liebmann, C. A. Girkin, and C. Bowd. "Glaucoma Progression Detection Using Structural Retinal Nerve Fiber Layer Measurements and Functional Visual Field Points." *IEEE T. Bio-med. Eng.* **61**(4), pp. 1143-1154, 2014.

[36] S. Yousefi, M. H. Goldbaum, M. Balasubramanian, F. A. Medeiros, L. M. Zangwill, J. M. Liebmann, C. A. Girkin, R. N. Weinreb, and C. Bowd, "Learning From Data: Recognizing Glaucomatous Defect Patterns and Detecting Progression From Visual Field Measurements," *IEEE T. Bio-med Eng.* **61**(7), pp. 2112- 2124, 2014.