

# Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization

Hiroki Sugiura

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

Hiroshi Murata

Department of Ophthalmology, The University of Tokyo, Tokyo, Japan  
hmurata-tky@umin.net

Taichi Kiwaki

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

Ryo Asaoka

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

Siamak Yousefi

Depart. of Ophth., University of Tennessee Health Science Center Memphis, Texas, USA  
siamak.yousefi@uthsc.edu

Kenji Yamanishi

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

## ABSTRACT

Conventionally, glaucoma is diagnosed on the basis of visual field sensitivity (VF). However, the VF test is time-consuming, costly, and noisy. Using retinal thickness (RT) for glaucoma diagnosis is currently desirable. Thus, we propose a new methodology for estimating VF from RT in glaucomatous eyes. The key ideas are to use our new methods of *pattern-based regularization* (PBR) and *pattern-based visualization* (PBV) with convolutional neural networks (CNNs). PBR effectively conducts supervised learning of RT–VF relations in combination with unsupervised learning from non-paired VF data. We can thereby avoid overfitting of a CNN to small sized data. PBV visualizes functional correspondence between RT and VF with its nonlinearity preserved. We empirically demonstrate with real datasets that a CNN with PBR achieves the highest estimation accuracy to date and that a CNN with PBV is effective for knowledge discovery in an ophthalmological context.

## CCS CONCEPTS

- Applied computing → Health informatics; • Computing methodologies → Neural networks; Dimensionality reduction and manifold learning; Regularization; Non-negative matrix factorization; Principal component analysis;

## KEYWORDS

Glaucoma, Convolutional Neural Networks, Non-negative Matrix Factorization, Regularization, Visualization

### ACM Reference Format:

Hiroki Sugiura, Taichi Kiwaki, Siamak Yousefi, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2018. Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '18, August 19–23, 2018, London, United Kingdom

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5552-0/18/08...\$15.00

<https://doi.org/10.1145/3219819.3219866>

In KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3219819.3219866>

## 1 INTRODUCTION

### 1.1 Motivation

Glaucoma is a progressive eye disease that brings about the death of retinal nerves and disorder in visual perception. It can eventually cause the permanent loss of sight. The progression is irreversible but can be slowed down with medical and/or surgical treatment. Therefore, detecting glaucoma progression in the early stages is important. VF is a principal measure for diagnosing glaucoma and monitoring glaucoma progression. VF is commonly measured with a Humphrey field analyzer (HFA, Carl Zeiss Meditec, Dublin, California, USA). However, the VF test is very labor-intensive, very time-consuming, and very noisy [25]. Reports have shown that retinal thickness (RT) is closely correlated to VF [6, 13, 14]. RT is measured using optical coherence tomography (OCT) [15]. This measurement is much less expensive, less time-consuming, and less noisy than the VF test. Thus, developing technology for accurately estimating VF from RT measured by OCT is important. This would drastically reduce the current costs for glaucoma diagnoses and eventually have a large impact on ophthalmology.

Uesaka et al. [30] proposed two methods to realize this estimation task based on affine structured non-negative matrix factorization (ASNMF) and a convolutional neural network (CNN). They showed that the ASNMF-based method works better for relatively small data sizes, while the CNN-based method works better for larger data sizes. They also showed that the ASNMF-based method was useful for ophthalmological knowledge discovery. To the best of our knowledge, [30] was the first paper on estimating the VF from RT data.

The main drawbacks of the previous methods are that 1) the estimation accuracy is not high enough for practical use, and 2) the acquired knowledge fails to reflect nonlinear relations between the VF and RT [14].

Regarding 1), the root mean squared error (RMSE) for estimation with these methods was 6.8–7.3 in [30], which is insufficient to be used at the real clinical practices. The reason for this inaccuracy is

that significant amounts of paired data of VF and RT are required for training, but only imbalanced data are available. That is, a small amount of paired data are available, while a large amount of non-paired VF data without the corresponding RT data are available. This is the *imbalanced non-paired data problem*.

Regarding 2), Uesaka et al. [30] acquired knowledge about the functional relations between RT and VF using a linear model. The true functional relations, however, can be intrinsically nonlinear [14]. That is, VF values can suddenly change as RT values smoothly change. This nonlinearity is a key to early detection of glaucoma. CNNs can learn such nonlinear relations. However, it has remained open how we can visualize the nonlinear relations. This is the *RT-VF nonlinearity problem*.

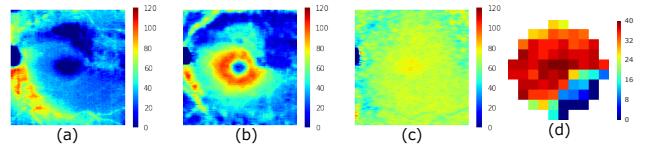
The purpose of this paper is twofold. First we propose a novel universal methodology for estimating VF from RT data by resolving the imbalanced non-paired data problem and RT-VF nonlinearity problem. Second, we empirically demonstrate the effectiveness and usefulness of our proposed methodology to obtain the best performance recorded in this field.

## 1.2 Contribution and Significance

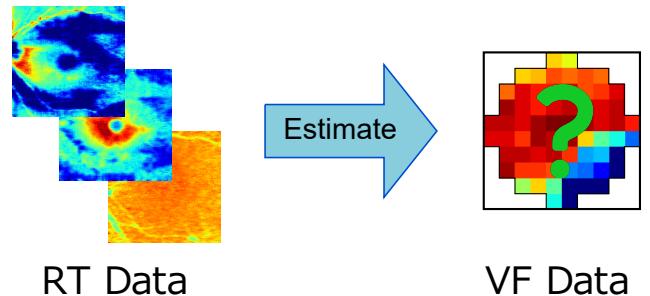
The contributions of our paper can be summarized as follows:

A) *Development of pattern-based regularization learning*: The key idea to resolve the imbalanced non-paired data problem is to first learn the patterns from non-paired data (VF data) using an unsupervised method and then to regularize the supervised learning from paired data (RT+VF data) with the obtained patterns. We call this *pattern-based regularization* (PBR). Note that PBR is completely different from the conventional semi-supervised learning (SSL) method [4]. Let  $x$  be RT and  $y$  be VF. In SSL, the unsupervised learning for generating  $x$  is utilized in order to learn the conditional probability model  $p(y|x)$ . In our method, the unsupervised learning for generating  $y$  is utilized for the same purpose. Here, we employ CNNs for supervised learning, while we may employ non-negative matrix factorization (NMF), principal component analysis (PCA), or auto-encoders (AEs) for unsupervised learning. Thus, we can make the best use of both non-paired and paired data with PBR. We applied PBR to real datasets of glaucoma patients and showed that our method achieved an RSME of 6.0, which is the highest accuracy for estimating VF from RT to date.

B) *Development of a novel methodology of “pattern-based visualization”*: The key idea for resolving the RT-VF nonlinearity problem is to first extract the characteristic patterns in the VF and RT domains using NMF and then to represent the gradient information of CNNs with respect to these patterns. This method is called *pattern-based visualization* (PBV). PBV is able to represent *nonlinear* relations between RT and VF. That is, edges of a PBV graph suddenly appear as the glaucoma starts to damage the VF. Hence edge detection in PBV helps us understanding the progression of glaucoma. This is in contrast with ASNMF, which is only capable of visualizing the linear dependency between the domains. We empirically demonstrate that PBV can detect potential VF defects which are suspected to take place in the future.



**Figure 1: RT and VF data.** (a) RNFL, (b) GCIPL, (c) RCL, (d) VF.



**Figure 2: Problem overview.** Models are trained to estimate the VF (on the right) from the RT data (on the left)

## 1.3 Related Work

There have been many studies on predicting glaucoma progression from VF data (e.g. [12, 22]). However, most suffered from the problem of a shortage of relevant patient data. A number of recent studies overcame this problem by leveraging similar data from other patients [10, 19–21]. Other studies on glaucoma progression prediction have used not only VF but also other features such as intraocular pressure [29] and retinal nerve fiber layer (RNFL) data around the optic disk of the eye [32].

Recently, Hood et al. [14] and Ajtorny et al. [1] showed a statistically-significant correlation between VF and RT. Eura et al. [6] demonstrated that the averaged thickness of the ganglion cell and inner plexiform layer (GCIPL) is related to VF. However, these previous studies did not indicate any method for estimating VF from RT.

Uesaka et al. [30] did the first work on estimating VF measured with HFA (SITA-Standard mode 10-2) from RT data. This methodology was based on multi-view learning over two different domains: VF and RT. As already discussed, however, this methodology suffered from the imbalanced non-paired data problem and RT-VF nonlinearity problem. Our proposed methodology in this paper is designed to overcome these difficulties.

Our proposed method of PBR is closely related to SSL in that both combine supervised learning with unsupervised learning to achieve better results [4]. However, the problem settings differ from each other as already discussed. The methodology for SSL cannot straightforwardly be applied to our problem.

The rest of this paper is organized as follows: Section 2 introduces basic notions of VF, RT, CNN, and NMF as preliminaries. Sections 3 and 4 introduce PBR and PBV, respectively. Section 5 presents experimental results on real glaucoma datasets. Section 6 concludes the paper.

## 2 PRELIMINARIES

### 2.1 Visual Field Data

VF is expressed by an integer value of the visual sensitivity threshold belonging to  $\{0, \dots, 40\}$ . The threshold is measured with HFA using SITA-Standard mode and 10-2 programme. A higher photosensitivity at a VF location means a higher threshold value. See [25] for details on SITA-Standard.

In our dataset, the VF was measured at 68 locations within the region of the central 10-deg visual field. These locations are distributed over the visual field every 2 deg. An example of real VF data is shown in Fig. 1 (d). It indicates that the lower half of the VF has defects. The problem with the VF test is that it is very time-consuming, and the measurement is very noisy. The test generally takes more than 20 min for both eyes.

### 2.2 Retinal Thickness Data

Researchers have shown that VF is very much correlated with RT [2, 6, 14, 16]. In glaucomatous eyes, there are retinal layers become thin as glaucoma progresses.

The RT data in our dataset consisted of the RNFL, GCIPL, and rod and cone layer (*RCL*). The data were measured using OCT [15], and obtained as images with a resolution of  $512 \times 128$  pixels. An example of RT data with structural abnormality is shown in Fig. 1 ((a)-(c)).

In general, RT tests using OCT are much less time-consuming than VF tests using HFA. The former takes only a few seconds and is much less sensitive to measurement noise. Hence glaucoma diagnosis based on RT tests is desired in order to reduce the costs. The aim of our study was to develop an effective method for estimating VF from RT (See Fig. 2).

### 2.3 Estimation with Convolutional Neural Networks

We employ a convolutional neural network (CNN) [17, 27, 28] as a basic model for estimating VF from RT following [30]. A CNN is a deep neural network that consists of alternating convolution and pooling layers. The CNN takes RNFL, GCIPL, and RCL images as its input and estimates VF.

The size of clinically available glaucoma dataset is usually not large enough for training a CNN. Hence we utilize the framework of transfer learning [31]. Within this framework, network parameters are initialized with those of a network that has already been trained with another large dataset. Specifically, we employ VGG16 [27] trained for natural image classification [24].

On the top of the last convolutional layer of the pre-trained architecture, a location-wise linear regression layer was stacked to estimate VF in [30]. We use this architecture as one of the baselines. Let  $\mathbf{h}_l(\mathbf{x}) \in \mathbb{R}^M$  be the  $l$ th output channel (i.e. the  $l$ th feature map) of the convolutional layer. Then, VF is estimated as follows:

$$\mathbf{y}_{\text{pred}}(\mathbf{x}) = \sum_{m=1}^M \left( \sum_{l=1}^L a_{lm} \mathbf{h}_l(\mathbf{x}) + b_m \right) \odot \delta_m, \quad (1)$$

where  $L$  is the number of filters in the last convolutional layer (for VGG16,  $L = 512$ ),  $(a_{lm}) \in \mathbb{R}^{L \times M}$  and  $(b_m) \in \mathbb{R}^M$  are regression parameters,  $\odot$  is an operator of taking component-wise products,

and  $\delta_m \in \mathbb{R}^M$  is an indicator vector for the  $m$ th location i.e., its  $m$ th element is 1 and others are 0. These regression parameters are randomly initialized and optimized from scratch, unlike the other parameters of the network.  $M$  is the number of VF locations; for all of the datasets dealt with in this paper,  $M = 68$ .

The model is trained on the glaucoma dataset by minimizing the squared regression error

$$\frac{1}{N} \sum_{n=1}^N \|\mathbf{y}_{\text{pred}}(\mathbf{x}_n) - \mathbf{y}_n\|^2 \quad (2)$$

with  $\{\mathbf{y}_n\}$  being the ground truth,  $N$  being the number of paired data, and  $\|\cdot\|^2$  being the  $L_2$  norm. Standard techniques like back-propagation can be used here.

### 2.4 Extracting Characteristic Patterns with Non-negative Matrix Factorization

We employ non-negative matrix factorization (NMF) for extracting characteristic patterns in PBR and PBV. NMF [18] is a method which approximates a matrix of non-negative data points  $Y = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times M}$  with a product of two non-negative matrices. Namely, with a given rank  $K$ , NMF approximately factorizes  $Y$  as  $Y \approx W\Theta$  where  $W \in \mathbb{R}^{N \times K}$  and  $\Theta \in \mathbb{R}^{K \times M}$ . Rows of  $\Theta$ , which are denoted by  $\{\theta_k\}$ , are *characteristic patterns* of data. Data points are approximately made up from linear combinations of these patterns. This factorization can be obtained by solving the following optimization problem:

$$\min_{W \geq 0, V \geq 0} \|Y - W\Theta\|_F^2, \quad (3)$$

where  $\|\cdot\|_F$  is the Frobenius norm.

Since both VF and RT data are non-negative, NMF is suitable for extracting characteristic patterns in these domains. It has actually been demonstrated that NMF can extract useful and medically plausible patterns of glaucoma data [30].

## 3 PATTERN-BASED REGULARIZATION

When we apply the CNN in Sec. 2.3 to a real glaucoma dataset, it often suffers from the problem that insufficient paired data (VF and RT) are available for it to be trained. However, abundant non-paired data (i.e., VF only) are available. The shortage of paired data may cause overfitting to the training data if we simply train a CNN with many parameters only from the paired data. This is the imbalanced non-paired data problem. To resolve this problem, we introduce a novel learning method called PBR. The key idea is to regularize CNN learning by making use of characteristic patterns obtained from non-paired VF data. Note that non-paired VF data were not utilized for CNNs in the previous study [30].

We propose two techniques for PBR:

- (A) *regularizing objective functions*,
- (B) *regularizing network structures*.

Both regularize CNN learning using VF patterns. We give their details below.

### 3.1 Regularizing Objective Functions

Regularizing objective functions is conducted in two steps. First, we extract characteristic VF patterns with an unsupervised learning

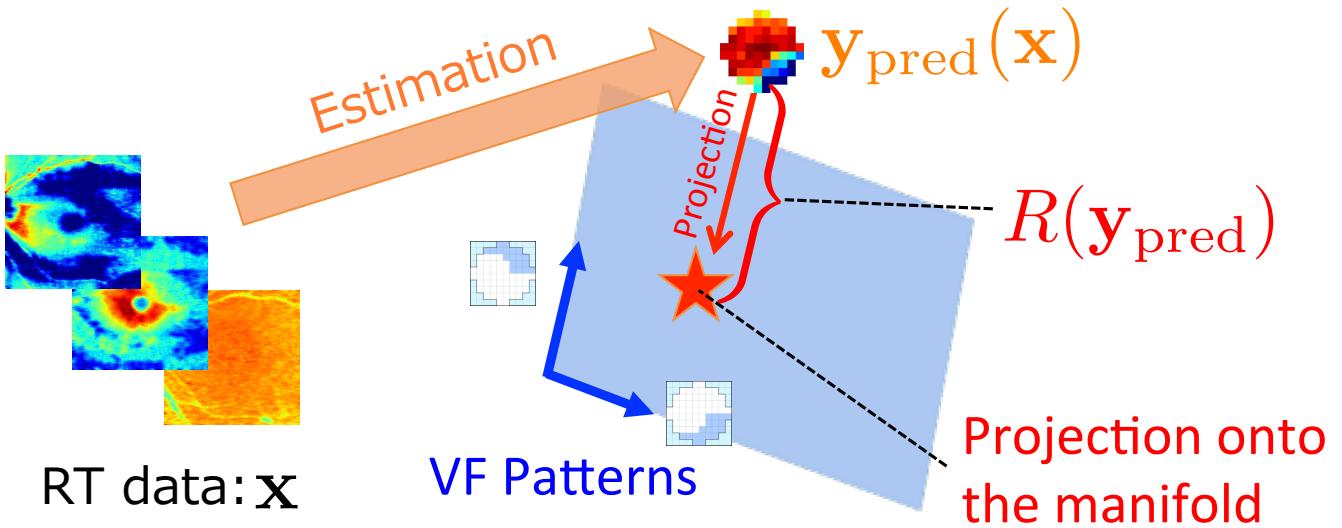


Figure 3: Pattern based regularization

method. Most VF data lie in a low-dimensional manifold formed by these patterns. Second, we regularize a model so that the estimation falls near the manifold as illustrated in Fig. 3. We can then regularize models without diminishing their representation capacity.

We realize these two steps by regularizing the objective function as follows. Suppose that a low-dimensional manifold has already been obtained by an unsupervised learning method. We define the regularized objective function as:

$$\frac{1}{N} \sum_{n=1}^N \|\mathbf{y}_{\text{pred}}(\mathbf{x}_n) - \mathbf{y}_n\|^2 + \frac{\beta}{2} R(\mathbf{y}_{\text{pred}}(\mathbf{x}_n)), \quad (4)$$

where  $R(\mathbf{y}_{\text{pred}})$  is a novel regularization term that penalizes the model if the estimation result  $\mathbf{y}_{\text{pred}}$  is far from the manifold.  $\beta (> 0)$  is a hyper parameter that tunes the strength of the regularization.

Note that PBR is similar to SSL [33] in that both utilize unsupervised learning results for supervised learning. However, they differ from each other because PBR utilizes the unsupervised learning for the objective variables (i.e., non-paired VF data) for supervised learning of the RT-VF relationship, while SSL utilizes the unsupervised learning for the explanatory variables (i.e., non-paired RT data) for the supervised learning of the relationship.

The concrete form of the regularization term  $R(\mathbf{y}_{\text{pred}})$  may depend on the choice of an unsupervised learning algorithm. Below we give examples of  $R(\mathbf{y}_{\text{pred}})$  corresponding to the three methods of unsupervised learning: NMF, PCA, and AEs.

**3.1.1 Pattern-Based Regularization with Non-negative Matrix Factorization and Principal Component Analysis.** Since both NMF and PCA are linear dimension reduction algorithms, we can calculate the regularization term  $R(\mathbf{y}_{\text{pred}})$  in a general form. Suppose that a  $K$ -dimensional affine subspace embedded in  $\mathbb{R}^M$  with  $K \leq M$ :

$$\{\Theta^\top \mathbf{w} + \mathbf{b} | \mathbf{w} \in \mathbb{R}^K\}, \quad (5)$$

is obtained by NMF or PCA, where  $\Theta \in \mathbb{R}^{K \times M}$  and  $\mathbf{b} \in \mathbb{R}^M$ . Then we calculate the regularization term  $R(\mathbf{y}_{\text{pred}})$  as follows:

$$R(\mathbf{y}_{\text{pred}}) \stackrel{\text{def}}{=} \frac{1}{M} \|\mathbf{y}_{\text{pred}} - \mathbf{y}_{\text{proj}}\|^2, \quad (6)$$

where  $\mathbf{y}_{\text{proj}}$  is the projection of  $\mathbf{y}_{\text{pred}}$  onto the low-dimensional subspace:

$$\mathbf{y}_{\text{proj}} \stackrel{\text{def}}{=} \Theta^\top (\Theta \Theta^\top)^{-1} \Theta (\mathbf{y}_{\text{pred}} - \mathbf{b}) + \mathbf{b}. \quad (7)$$

Once NMF is performed on non-paired VF data, we obtain characteristic VF patterns  $\Theta$  in Eq. (3). PBR can then be realized with NMF by plugging the resulting  $\Theta$  and  $\mathbf{b} = \mathbf{0}$  into Eq. (7).

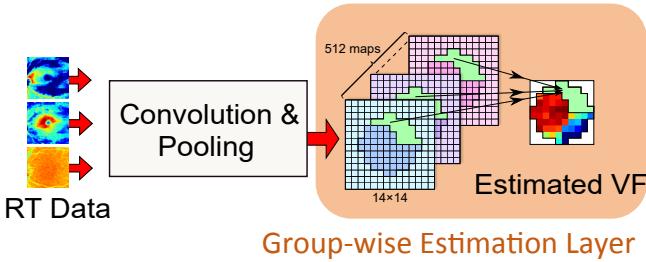
PCA is one of most typical dimensional reduction methods. It calculates the variance-covariance matrix from  $Y$  and then calculates its eigenvalues and the orthogonal normalized bases. The bases corresponding to larger eigenvalues are principal components. Let  $\Theta$  be a set of bases corresponding to the top  $K$  largest eigenvalues. Let  $\mathbf{b}$  be the sample mean. Then we can realize PBR with PCA by plugging such  $\Theta$  and  $\mathbf{b}$  into Eq. (7).

**3.1.2 Auto Encoder-based Regularization.** We can employ a nonlinear unsupervised learning algorithm in PBR. Below we give an example of PBR with AEs [5, 11]. An AE is a neural network that computes  $f : \mathbb{R}^M \rightarrow \mathbb{R}^M$  as

$$f(\mathbf{y}; \Theta) \stackrel{\text{def}}{=} f_2(W_2^\top f_1(W_1^\top \mathbf{y} + \mathbf{b}_1) + \mathbf{b}_2), \quad (8)$$

where  $\Theta = \{W_1, \mathbf{b}_1, W_2, \mathbf{b}_2\}$  with  $W_1 \in \mathbb{R}^{M \times K}$ ,  $W_2 \in \mathbb{R}^{K \times M}$ ,  $\mathbf{b}_1 \in \mathbb{R}^K$ , and  $\mathbf{b}_2 \in \mathbb{R}^M$  are parameters, and  $f_1(\cdot)$ ,  $f_2(\cdot)$  are respectively nonlinear functions for the encoder and decoder. Let  $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  be non-paired VF dataset. An AE can be trained on this dataset by solving the following optimization problem:

$$\Theta = \arg \min_{\Theta} \frac{1}{N} \sum_{n=1}^N \|\mathbf{y}_n - f(\mathbf{y}_n; \Theta)\|^2. \quad (9)$$



**Figure 4: PBR of network structures**

An AE can learn a low-dimensional manifold where the input data lies [9]. Mathematically, the manifold is defined as the invariant set of a recurrence map  $\mathbf{y}_{t+1} \leftarrow f(\mathbf{y}_t)$ . The closer  $\mathbf{y}_t$  approaches the manifold, the smaller  $\mathbf{y}_{t+1} - f(\mathbf{y}_t)$  gets. This observation motivates us to define  $R(\mathbf{y}_{\text{pred}})$  for PBR with an AE as follows:

$$R(\mathbf{y}_{\text{pred}}) \stackrel{\text{def}}{=} \|\mathbf{y}_{\text{pred}} - f(\mathbf{y}_{\text{pred}}; \Theta)\|^2. \quad (10)$$

### 3.2 Regularizing Network Structures

Regularizing network structures is a technique for reducing the number of parameters via group-wise estimation. Without this regularization, we would have to make a location-wise estimation, i.e., estimate VF location by location. This would require a huge number of parameters:  $ML = 68 \times 512 = 34816$ . Thus, location-wise estimation may cause an overfitting problem because the data size will be 600 at most, which is much smaller than the number of these parameters.

In order to avoid this problem, we propose drastically reducing the number of parameters by conducting group-wise estimation rather than location-wise one. The former means that we first categorize VF locations into a number of groups depending on functional similarity of the locations; then an estimation model is constructed with these groups where the parameters are shared within each group as in Fig. 4. Letting the total number of groups be  $K$ , then the total number of parameters required by group-wise estimation is  $K \times 512$ , which can be much smaller than that required for location-wise estimation. Note that grouping visual field locations is reasonable from a medical viewpoint [7, 14].

NMF can be used to group functionally similar VF locations. Let  $\theta_k = [\theta_{k1}, \dots, \theta_{kM}]^\top$  be the characteristic VF patterns extracted from non-paired VF data with NMF. Large values of  $\theta_{km}$  tend to be assigned to similar VF locations in each pattern. A pattern thus indicates a group of VF locations in a soft way (i.e., with continuous values). Regularizing network structures can be achieved by using  $\theta_k$  instead of the location indicator  $\delta_m$  in Eq. (1) as:

$$\mathbf{y}_{\text{pred}}(\mathbf{x}) = \sum_{k=1}^K \left( \sum_{l=1}^L \tilde{a}_{kl} \mathbf{h}_l(\mathbf{x}) + \tilde{b}_k \right) \odot \theta_k, \quad (11)$$

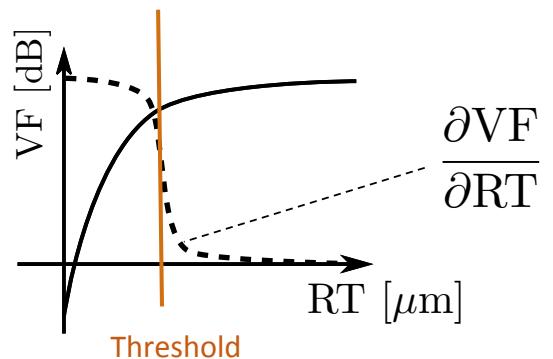
where  $\tilde{a}_{kl}$  and  $\tilde{b}_k$  are parameters assigned to the  $k$ th pattern.

## 4 PATTERN-BASED VISUALIZATION

In this section, we introduce PBV. Characteristic patterns extracted by CNNs play key roles in PBV. We first give the motivation for using the differential information of CNNs. Then, we summarize the problems of conventional CNN-based visualization techniques. Finally, we present the proposed PBV to resolve the problems.

### 4.1 Motivation

There is known to be a significant correlation between RT at a location and VF at the functionally corresponding location [6, 14]. This relation is further known to be nonlinear [14]. In other words, VF is not much influenced by RT unless the latter exceeds a certain threshold. Once it exceeds the threshold, VF suddenly changes. Fig. 5 shows this relation.



**Figure 5: Early glaucoma detection and RT–VF gradient**

This nonlinear relation suggests that the derivative of VF with respect to RT gives much more information than VF itself for the early detection of glaucoma (see Fig. 5). This is because, in the early stages of glaucoma, the absolute derivative of VF may be large while the VF value is still near normal levels. Such early stages may be overlooked if only VF values are examined.

Our proposed PBV is motivated by this observation. We can extract derivative information from trained CNNs, which may be utilized to visualize the correspondence between RT and VF.

### 4.2 Visualization Technique of CNNs

Observing the backpropagated gradients of a CNN is a widely used technique in deep learning to interpret functionalities of networks [26].

The function  $f(\bullet)$ , which is represented by a CNN, can be linearized around a data point  $\mathbf{x}$  with respect to local changes in the data domains  $\Delta\mathbf{x}$  and  $\Delta\mathbf{y}$  as follows:

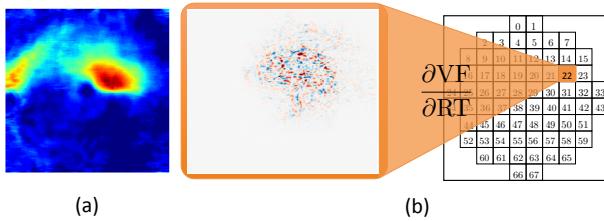
$$\Delta\mathbf{y} = A\Delta\mathbf{x}, \quad (12)$$

where  $A = (a_{ml}) = [\mathbf{a}_1, \dots, \mathbf{a}_M]^\top$  is the Jacobian matrix:

$$a_{ml} \stackrel{\text{def}}{=} \left. \frac{\partial f_m(\mathbf{x}; \Theta)}{\partial x_l} \right|_{\mathbf{x}}. \quad (13)$$

Row vectors  $\mathbf{a}_m$  can be visualized as images. This indicates which part of the input  $\mathbf{x}$  is responsible for an estimation outcome.

There are two problems in the application of this visualization technique to our case. First, RT-VF gradients of a CNN largely disagree with established medical knowledge on RT progression [14]. Fig. 6 compares  $a_{22}$  of an eye and a medically-understandable RT pattern extracted with NMF. They apparently differ; RT thinning patterns like  $a_{22}$  are not medically plausible [14]. Therefore, a straightforward application of the existing technique cannot be a reliable predictor for glaucomatous progression. Second, the granularity of the visualization in terms of the VF domain is too fine for our purpose. With a straightforward application of the existing technique, we have to check  $a_m$  for each of VF locations. From the ophthalmological viewpoint, however, we are less interested in progression at individual VF locations. We are rather interested in progression patterns that span over entire VF [7, 13, 30]. This suggests that we should consider VF patterns rather than individual VF locations.



**Figure 6: Comparison between a typical RT deterioration pattern and RT-VF gradients:** (a) a RT progression pattern, (b)  $a_{22}$  visualized as an image.

To resolve these two problems, we need to develop a new method that can deal with patterns in low-dimensional manifolds both in the VF and RT domains. Thus, we propose PBV. With this method, we extract characteristic damage patterns in both of the domains. We employ these patterns to restrict the translation vectors  $\Delta\mathbf{x}$  and  $\Delta\mathbf{y}$  to be within the subspaces. With PBV, we can represent RT-VF relations in a concise fashion, which can be easily referred to medical knowledge.

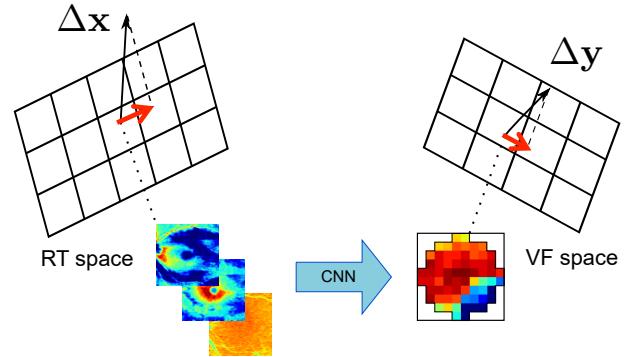
### 4.3 Proposed Method: Pattern-Based Visualization

It has been demonstrated [30] that medically plausible patterns of the VF and RT of glaucomatous eyes can be extracted by using NMF. Here, we also employ NMF for extracting the patterns. In PBV, NMF is conducted under the assumption that VF and RT can be represented as linear combinations of such patterns. This assumption mathematically implies that VF or RT data lie on a linear subspace embedded in a high-dimensional Euclidian space.

The PBV process is given as follows. First, we conduct NMF on non-paired RT and VF data to extract glaucomatous damage patterns in the respective domains. We denote these patterns as  $\{\mathbf{b}_1^{\text{RT}}, \dots, \mathbf{b}_{L'}^{\text{RT}}\}$  and  $\{\mathbf{b}_1^{\text{VF}}, \dots, \mathbf{b}_{M'}^{\text{VF}}\}$ . We then restrict the local translations  $\Delta\mathbf{x}$  and  $\Delta\mathbf{y}$  as:

$$\Delta\mathbf{x} = B_{\text{RT}}z_{\text{RT}}, \Delta\mathbf{y} = B_{\text{VF}}z_{\text{VF}}, \quad (14)$$

where  $z_{\text{RT}}$  and  $z_{\text{VF}}$  are local change vectors with respect to the NMF bases (i.e., glaucomatous patterns).  $B_{\text{RT}}$  and  $B_{\text{VF}}$  are matrices



**Figure 7: Pattern based visualization**

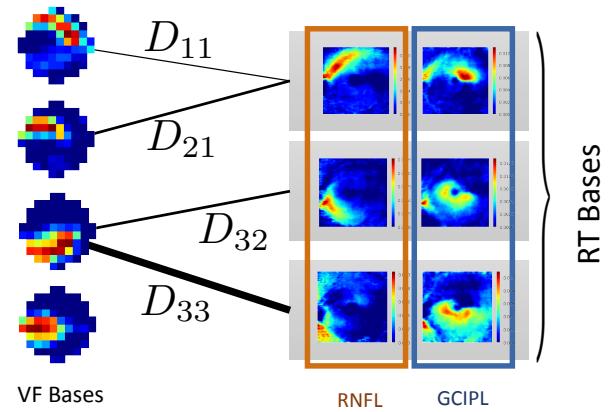
that consist of the RT and VF bases:  $B_{\text{RT}} = [\mathbf{b}_1^{\text{RT}}, \dots, \mathbf{b}_{L'}^{\text{RT}}]^{\top}$  and  $B_{\text{VF}} = [\mathbf{b}_1^{\text{VF}}, \dots, \mathbf{b}_{M'}^{\text{VF}}]^{\top}$ . Fig. 7 illustrates this idea. Eq. (14) is plugged into Eq. (12) and solved for  $z_{\text{VF}}$  with the Moore-Penrose inverse to yield

$$z_{\text{VF}} = Dz_{\text{RT}}, \quad (15)$$

where

$$D \stackrel{\text{def}}{=} (B_{\text{VF}}B_{\text{VF}}^{\top})^{-1}B_{\text{VF}}AB_{\text{RT}}^{\top}. \quad (16)$$

Note that local translation vectors are restricted to be within low-dimensional linear subspaces. Thus, the granularity of visualization is no longer too fine. In addition, the resulting correspondence can easily be checked whether it matches existing medical knowledge. Furthermore, PBV is able to visualize nonlinear relations between the VF and RT patterns. By the nonlinearity, an edge may suddenly emerge in the PBV graph as the glaucoma starts to damage the VF, even if the VF is nearly at the normal levels. This implies that edge detection in PBV helps us detect the early damage of glaucoma.



**Figure 8: Bipartite graph drawn by PBV.**

Because of the non-negativity of the NMF bases and positive correlation between RT and VF, most elements of  $D$  are positive. Therefore we ignore negative values in  $D$ . This allows us to visualize the relations in Eq. (16) as a bipartite graph as in Fig. 8. The thickness

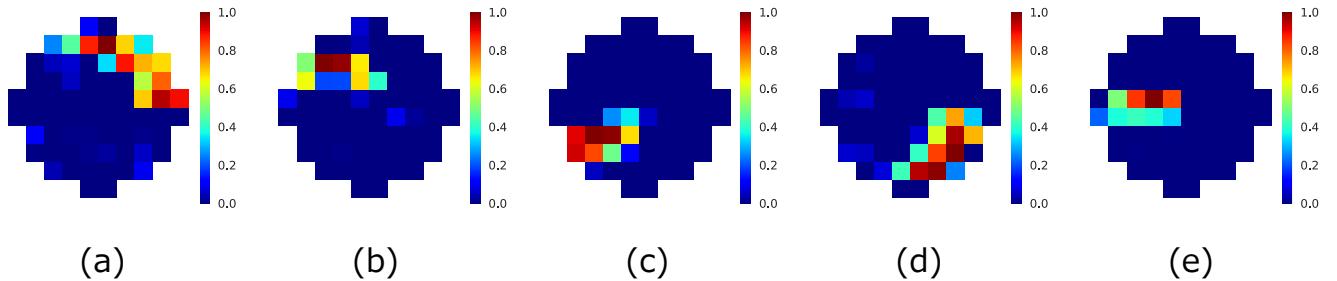


Figure 9: VF patterns extracted with NMF.

of an edge is proportional to the corresponding element of  $D$ . Hence a thicker edge means a stronger correlation between the VT and RT bases. The connectivity of the bases shows which part of the retina is related to the defect or potential defect detected with PBV. The graph appearance with regard to edge connections thus indicates the eye’s health condition.

## 5 EXPERIMENTS

### 5.1 Data Specifications

The dataset used in our experiments was provided by the Department of Ophthalmology, The University of Tokyo Hospital, Inouye Eye Hospital, or JR Tokyo General Hospital. The number of paired data was 591, the number of non-paired RT data was 1488, and the number of non-paired VF data was 7715. Each data point represented a mutually different eye. For patients with multiple RT or VF measurements, we only used the latest measurements.

VF data were measured with HFA using SITA-Standard mode and 10-2 programme. The dimension of the VF data was  $M = 68$ . The range of the VF data was  $\{0, \dots, 40\}$ . RT data were measured with OCT (RS3000, Nidek Co Ltd., Gamagori, Aichi, Japan). All of the measurements were taken around the macula ( $9.0 \times 9.0$ mm, corrected with axial length). The measurements had three channels corresponding to the three different retinal layers: RNFL, GCIPL, and RCL. Each measurement channel was a two-dimensional array with the size  $512 \times 128$ . The range of the RNFL data was  $[0, 1848]$ , and the average was 32.3. The range of the GCIPL data was  $[0, 1673]$ , and the average was 41.5. The range of the RCL data was  $[0, 394.3]$ , and the average was 67.0. All measurements were in micrometers ( $\mu m$ ). Following the manufacturer’s recommendation, imaging data with quality factor  $< 7$  were also excluded.

### 5.2 Data Preprocessing

The dataset consisted of both right and left eyes. We horizontally flipped the data of the right eyes to reconcile the horizontal symmetry. Furthermore, the RT data were vertically and horizontally flipped to reconcile the optical image flip in the eyes. All of the RT channels were resized to  $224 \times 224$  to fit the design of the pre-trained model [27]. The resized RNFL, GCIPL, and RCL data were fed to the red, green, and blue channels of the model. The estimated VF values were rectified to the range  $[0, 40]$  by clipping values that were outside this range.

**Table 1: Comparison of estimation accuracies. Mean, median, and worst RMSEs are reported.**

	Median	Mean	Worst
CNN [30]	6.68	6.76	21.7
CNN + DA	6.70	6.52	21.7
CNN + DA + PBR1(NMF)	6.41	6.30	23.5
CNN + DA + PBR1(PCA)	6.31	6.26	24.1
CNN + DA + PBR1(AE)	6.39	6.26	24.2
CNN + DA + PBR1(PCA) + PBR2	<b>6.00</b>	<b>6.16</b>	21.4
CNN + DA + PBR1(AE) + PBR2	6.02	<b>6.16</b>	21.5

We considered the following five types of data augmentation: 1) vertical flip, 2) horizontal flip, 3) small rotation (within 5 deg), 4) small magnification (within 10 pixels), and 5) small transition (within 10 pixels). We tested all and found that only the vertical flip improved the estimation accuracy. In the following sections, we report on the data augmented via the vertical flip.

### 5.3 Experimental Setting

The parameters of the CNNs in the last layer were randomly initialized according to [8]. All other parameters were inherited from VGG16 [27]. The training data were randomly reordered every epoch. We trained models with stochastic gradient descent with momentum [23] and weight decay. Our source code is available at <https://goo.gl/MzfMY6>.

To make the best use of the limited amount of data, we validated the effect of the proposed methods via nested cross validation (CV) [3]. In the outer CV, the generalization error was estimated. Here, the paired data were split into training and validation sets in  $k$  ways with  $k$  being an integer. In the inner CV, the hyper parameters were optimized for a training set generated in the outer CV. The training set was further split into training and validation sets in  $k'$  ways with  $k'$  being another integer. In our experiments, both the inner and outer CVs were five-fold, i.e.,  $k = k' = 5$ .

We measured the estimation accuracy for an eye in terms of the RMSE defined as

$$\text{RMSE} \stackrel{\text{def}}{=} \sqrt{\sum_{m=1}^M (y_{\text{pred } m} - y_{\text{true } m})^2 / M},$$

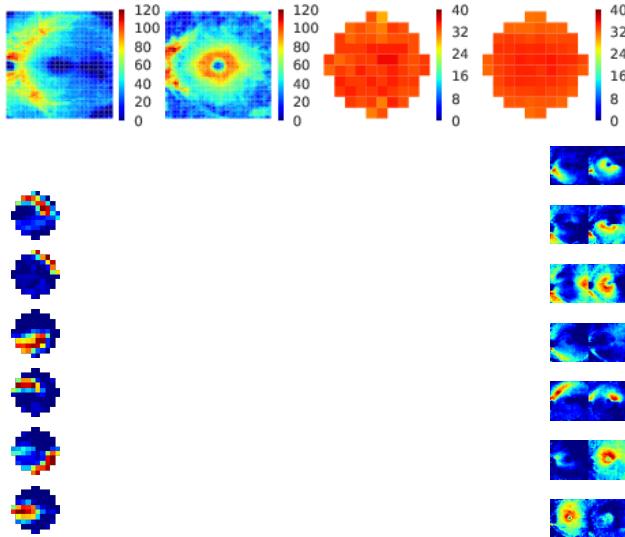


Figure 10: Patient with normal VF and RT

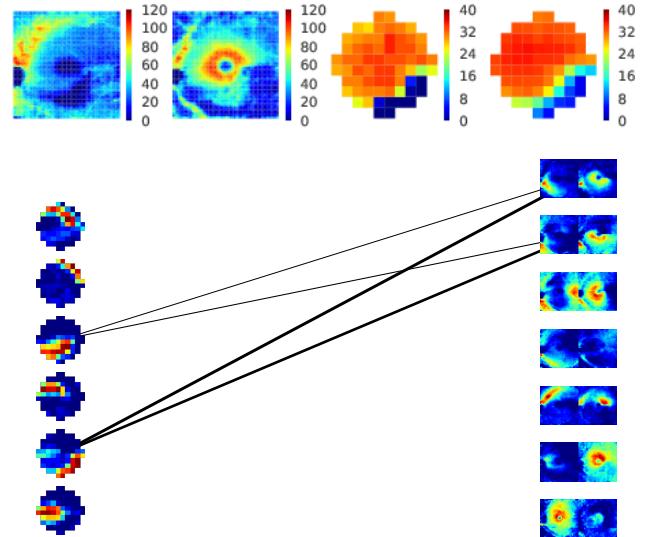


Figure 12: Patient with VF defect in inferior area

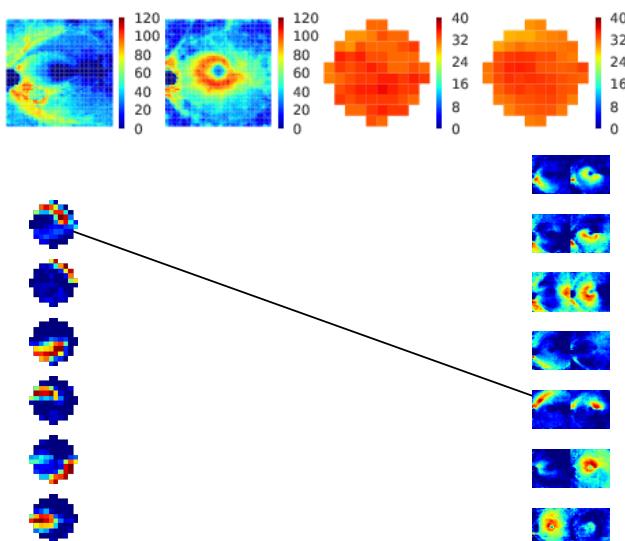


Figure 11: Patient with defected RT and normal VF

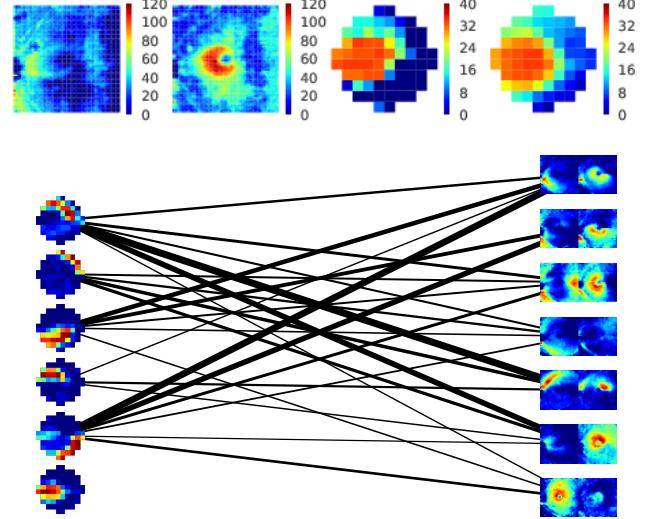


Figure 13: Patient with extensively defected VF and RT

where  $y_{\text{pred}}^m$  and  $y_{\text{true}}^m$  are the estimated value and ground truth, respectively, of the VF at the  $m$ th location. Models were validated based on the mean RMSE, which was the average RMSE taken over all of the eyes in the validation set.

The hyper parameters were tuned via inner CV. An early stopping criterion was determined in all of our experiments. For PBR of objective functions, the rank or dimension of the latent space for NMF, PCA, or AE was selected from  $\{7, 10, 13\}$ . The regularization coefficients were selected from  $\{1.0 \times 10^{-4}, 3.0 \times 10^{-4}, 1.0 \times 10^{-3}\}$ . For PBR of network structures, the rank for NMF was selected from  $\{35, 40, 45\}$ .

#### 5.4 Experiment 1: Estimation of VF from RT

Below, PBR1 denotes the technique of regularizing objective functions in PBR, while PBR2 denotes that of regularizing network structures in PBR. Table 1 presents the mean RMSEs for a number of combinations of PBR1 (with various unsupervised learning algorithms: NMF, PCA, AE), PBR2, and DA (data augmentation as given in Sec. 5.2). It also presents the results with the straightforward CNN from the previous work [30] and with DA as baselines. Both PBR1 and PBR2 outperformed the baselines. The statistical significance of the results was verified with the t-test at  $p = 0.05$  for the mean RMSEs. We observed that the mean RMSE was significantly reduced when using PBR1 and RBM2 together. Specifically, PBR1

with PCA/AE achieved the *highest accuracy* (6.00/6.02 in median, 6.16 in mean), to the best of our knowledge. This demonstrates that PBR can significantly improve the estimation accuracy for CNNs.

Fig. 9 shows examples of NMF bases for PBR2. Red indicates a high VF value while blue indicates a low VF value. In Fig. 9, VF remains only in the superior area (i.e., upper part) in (a), VF remains only in the superior juxta-central area (i.e., upper middle part) in (b), VF remains only in the inferior juxta-central area (i.e., lower middle part) in (c), VF remains only in the inferior area (i.e., lower part) in (d), and VF remains only in the central area in (e). They show that the resulting groupings of VF locations are reasonable from a medical viewpoint.

## 5.5 Experiment 2: Pattern-Based Visualization

We conducted PBV as described in Sec. 4.3 to demonstrate that early glaucomatous damage is reflected in the visualization.

PBV was performed on a CNN with PBR1(AE) and PBR2, which performed best as described in Sec. 5.4. Characteristic patterns in the VF and RT domains were extracted by NMF with data of 119 eyes in the validation set. There were seven RT bases and six VF bases. The extracted bases were normalized with the  $L_1$  norm. For the RT bases, we ignored elements in the RCL channel because glaucomatous progression is usually not observed in this layer in general [25].

We computed  $D$  according to Eq. (16) for each eye in the validation set. The thickness of an edge was proportional to the corresponding value in  $D$ . A thicker edge meant a stronger correlation between the RT and VF bases. For values in  $D$  below  $2.0 \times 10^{-3}$ , edges were not displayed for clarity of the plot.

Figs 10–13 show bipartite graphs corresponding to  $D$  for a number of patients. In each graph, we assign the VF bases to the left side and name them as VF basis1, VF basis2, ..., VF basis6, etc. in descending order. We assign the RT bases to the right side and name them as RT basis1, RT basis2, ..., RT basis6, etc. in descending order. In VF and RT bases, red indicates a high VF (or RT) value while blue indicates a low VF (or RT) value. Fig. 10 shows the graph of a patient for which the VF and RT are normal. Fig. 11 shows the graph of a patient for which the VF is normal but RT is defective. Fig. 12 shows the graph of a patient with VF and RT defects in the inferior area. Fig. 13 shows the graph of a patient at the terminal stage glaucoma. In the top of each figure, the RNFL, GCIPL, VF, and estimated VF are shown from left to right.

We see from Fig. 10 that the graph for the patient with normal eyes has no edge. Meanwhile, we see from Fig. 13 that the graph for the patient with defective RT has many edges. This implies that there is nonlinear relations between RT and VF, where VF defect progresses nonlinearly as RT value decreases. This agrees with established medical knowledge [14].

Fig. 12 demonstrates that PBV successfully captured functional correspondence between RT and VF. The patient has defect in the inferior area. The edges are drawn between VF bases responsible for defects in the inferior area and RT bases responsible for retinal deterioration in the inferior area. This result matches the medical knowledge on functional correspondence between VF and RT [14, 25]. In other words, the medical knowledge was confirmed from the view of data science.

In Fig. 11, the VF and estimated VF may not seem abnormal at first glance. However, PBV shows that there is an edge between VF basis1 and RT basis5. This implies that a mild decrease in the RT on the superior area can greatly affect the corresponding part of the VF. Therefore, this eye is suspected to show glaucomatous damage in this area. Such RT–VF relations represented by PBV will not be obtained by linear methods that discover static correspondence between RT and VF [30]. Our PBV was able to discover such relations because the trained CNN contained the useful nonlinear information between RT and VF, and was thus able to detect a sudden appearance of RT–VF correlation. We expect that PBV can also be used for identifying retinal regions related to the potential VF defect progression.

## 6 CONCLUSION

We proposed a novel methodology for estimating VF from RT in glaucomatous eyes. This technology is designed for the purpose of reducing the cost of current glaucoma diagnoses using HFA. The key ideas in our methodology are PBR and PBV, which are implemented with CNNs. PBR conducts unsupervised learning of the non-paired VF data and utilizes the result for supervised learning of RT–VF relations. PBR resolves the imbalanced non-paired data problem to prevent a CNN from overfitting to small datasets. PBV visualizes functional correspondences between RT and VF preserving their nonlinearity attained by CNNs. PBV help us understand the stages of glaucoma and identify retinal regions that are related to VF defects. We used real datasets to demonstrate that a CNN with PBR has achieved the highest estimation accuracy to date, and that PBV is effective in ophthalmological knowledge discovery. Future work may include; A) extension to RT time series prediction, B) extension to a framework using non-paired RT data, and C) further validation of the practical usefulness of our methodology on the basis of an external dataset.

## ACKNOWLEDGMENTS

This work was partially supported by JST CREST Grant Number JPMJCR1304, Japan, and by JSPS KAKENHI Grant Number JP18K18121. We thank Yuri Fujino and Masato Matsuura, Department of Ophthalmology, The University of Tokyo, for providing us useful comments. We thank Kai Morino, Shin Matsushima, Kohei Miyaguchi, and Atsushi Suzuki, The University of Tokyo, for helpful discussion. We thank Kenji Inoue, Inouye Eye Hospital, and Junkichi Yamagami, JR Tokyo General Hospital, for providing us the data.

## REFERENCES

- [1] Csilla Ajtony, Zsolt Balla, Szabolcs Somoskóy, and Balint Kovacs. 2007. Relationship between visual field sensitivity and retinal nerve fiber layer thickness as measured by optical coherence tomography. *Investigative ophthalmology & visual science* 48, 1 (2007), 258–263.
- [2] Ryo Asaoka. 2013. The relationship between visual acuity and central visual field sensitivity in advanced glaucoma. *British Journal of Ophthalmology* (2013), bjophthalmol–2013.
- [3] Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* 11, Jul (2010), 2079–2107.
- [4] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien (Eds.). 2006. *Semi-Supervised Learning*. MIT Press.
- [5] David DeMers and Garrison W Cottrell. 1993. Non-linear dimensionality reduction. In *Advances in neural information processing systems*. 580–587.

- [6] Mariko Eura. 2014. Correspondence Between Visual Field Test Results and GCL+IPL Thickness in the Maculae of Glaucomatous Eyes. *Ph.D. Thesis* (2014).
- [7] Stuart K Gardiner, Steven L Mansberger, and Shaban Demirel. 2017. Detection of Functional Change Using Cluster Trend Analysis in Glaucoma. *Investigative ophthalmology & visual science* 58, 6 (2017), BIO180–BIO190.
- [8] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Yee Whye Teh and Mike Titterington (Eds.), Vol. 9. 249–256.
- [9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. The MIT Press.
- [10] Motohide Higaki, Kai Morino, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2016. Predicting Glaucoma Visual Field Loss by Hierarchically Aggregating Clustering-based Predictors. *arXiv preprint arXiv:1603.07094* (2016).
- [11] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *science* 313, 5786 (2006), 504–507.
- [12] CATHARINA HOLMIN and CET Krakau. 1982. Regression analysis of the central visual field in chronic glaucoma cases. *Acta ophthalmologica* 60, 2 (1982), 267–274.
- [13] Donald C Hood. 2016. Improving our understanding, and detection, of glaucomatous damage: An approach based upon optical coherence tomography (OCT). *Progress in retinal and eye research* (2016).
- [14] Donald C Hood and Randy H Kardon. 2007. A framework for comparing structural and functional measures of glaucomatous damage. *Progress in retinal and eye research* 26, 6 (2007), 688–710.
- [15] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. 1991. Optical coherence tomography. *science* 254, 5035 (1991), 1178–1181.
- [16] Akiyasu Kanamori, Makoto Nakamura, Michael FT Escano, Ryu Seya, Hidetaka Maeda, and Akira Negi. 2003. Evaluation of the glaucomatous damage on retinal nerve fiber layer thickness measured by optical coherence tomography. *American journal of ophthalmology* 135, 4 (2003), 513–520.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*. Curran Associates Inc., USA, 1097–1105.
- [18] Daniel D Lee and H Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*. 556–562.
- [19] Zenghan Liang, Ryota Tomioka, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2013. Quantitative prediction of glaucomatous visual field loss from few measurements. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 1121–1126.
- [20] Shigeru Maya, Kai Morino, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2015. Discovery of Glaucoma Progressive Patterns Using Hierarchical MDL-Based Clustering. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1979–1988.
- [21] Shigeru Maya, Kai Morino, and Kenji Yamanishi. 2014. Predicting glaucoma progression using multi-task learning with heterogeneous features. In *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 261–270.
- [22] BN Noureddin, D Poinoosawmy, FW Fietzke, and RA Hitchings. 1991. Regression analysis of visual field progression in low tension glaucoma. *British journal of ophthalmology* 75, 8 (1991), 493–495.
- [23] Ning Qian. 1999. On the momentum term in gradient descent learning algorithms. *Neural networks* 12, 1 (1999), 145–151.
- [24] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.
- [25] TM Shaarawy, MB Sherwood, RA Hitchings, and JG Crowston. 2009. *Glaucoma: Medical Diagnosis and Therapy*. Vol. 1. Vol. 1. Elsevier.
- [26] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034* (2013).
- [27] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [28] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, et al. 2015. Going Deeper with Convolutions. In *CVPR*.
- [29] Kyosuke Tomoda, Kai Morino, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2016. Predicting Glaucomatous Progression with Piecewise Regression Model from Heterogeneous Medical Data. *HEALTHINF 2016* (2016).
- [30] Toshimitsu Uesaka, Kai Morino, Hiroki Suguri, Taichi Kiwaki, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2017. Multi-view Learning over Retinal Thickness and Visual Sensitivity on Glaucomatous Eyes. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2041–2050.
- [31] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.
- [32] Siamak Yousefi, Michael H Goldbaum, Madhusudhanan Balasubramanian, Tzyy-Ping Jung, Robert N Weinreb, Felipe A Medeiros, Linda M Zangwill, Jeffrey M Liebmann, Christopher A Girkin, and Christopher Bowd. 2014. Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Transactions on Biomedical Engineering* 61, 4 (2014), 1143–1154.
- [33] Xiaojin Zhu. 2006. Semi-supervised learning literature survey. *Computer Science, University of Wisconsin-Madison* 2, 3 (2006), 4.