

Development of a β -Variational Autoencoder for Disentangled Latent Space Representation of Anterior Segment Optical Coherence Tomography Images

Kilhwan Shon^{1,2}, Kyung Rim Sung³, Jiehoon Kwak³, Joong Won Shin³, and Joo Yeon Lee^{2,4}

¹ Department of Ophthalmology, Gangneung Asan Hospital, Gangneung, Korea

² Asan Artificial Intelligence Institute, Hwaseong-si, Gyeonggi-do, Korea

³ Department of Ophthalmology, College of Medicine, University of Ulsan, Asan Medical Center, Seoul, Korea

⁴ Camp 9 Orthopedic Clinic, Hwaseong-si, Gyeonggi-do, Korea

Correspondence: Kyung Rim Sung, Department of Ophthalmology, College of Medicine, University of Ulsan, Asan Medical Center, 388-1 Pungnap-2-dong, Songpa-gu, Seoul 138-736, Korea.
e-mail: sungeye@gmail.com

Received: October 8, 2021

Accepted: January 17, 2022

Published: February 8, 2022

Keywords: anterior segment OCT; deep learning; artificial intelligence; β -variational autoencoder

Citation: Shon K, Sung KR, Kwak J, Shin JW, Lee JY. Development of a β -variational autoencoder for disentangled latent space representation of anterior segment optical coherence tomography images. *Transl Vis Sci Technol.* 2022;11(2):11, <https://doi.org/10.1167/tvst.11.2.11>

Purpose: To investigate the feasibility of extracting a low-dimensional latent structure of anterior segment optical coherence tomography (AS-OCT) images by use of a β -variational autoencoder (β -VAE).

Methods: We retrospectively collected 2111 AS-OCT images from 2111 eyes of 1261 participants from the ongoing Asan Glaucoma Progression Study. After hyperparameter optimization, the images were analyzed with β -VAE.

Results: The mean participant age was 64.4 years, with mean values of visual field index and mean deviation of 86.4% and -5.33 dB, respectively. After experiments, a latent space size of 6 and β value of 5^3 were selected for latent space analysis with β -VAE. Latent variables were successfully disentangled, showing readily interpretable distinct characteristics, such as the overall depth and area of the anterior chamber (η_1), pupil diameter (η_2), iris profile (η_3 and η_4), and corneal curvature (η_5).

Conclusions: β -VAE can successfully be applied for disentangled latent space representation of AS-OCT images, revealing the high possibility of applying unsupervised learning in the medical image analysis.

Translational Relevance: This study demonstrates that a deep learning-based latent space model can be applied for the analysis of AS-OCT images.

Introduction

Anterior segment optical coherence tomography (AS-OCT) has become a crucial tool for assessing the iridocorneal angle in recent decades.¹ Assessment of the iridocorneal angle is of paramount importance in the diagnosis of primary angle closure (PAC). Parameters derived from AS-OCT have been used in various tasks, including subclassification, monitoring pre- and post-laser peripheral iridotomy changes, and the natural course of long-term structural changes in

eyes with PAC.²⁻⁵ Such studies have relied on manual measurements necessitated by the technical difficulties of developing satisfactory machine learning models. In the 2010s, deep neural networks, more specifically convolutional neural networks (CNNs), were applied for various computer vision tasks with great success.⁶

Several deep neural networks have been successfully trained to achieve accuracy comparable to human measurements of AS-OCT images, including in the detection of scleral spur, prediction of plateau iris, detection and grading of angle closure, and assessment of angle parameters.⁷⁻¹⁰ However, labeling is

vital in the training of the supervised model, and every parameter has to be manually defined, marked, and measured in these studies. Additionally, parameter definitions need to be simple and clear enough to measure manually because overly complex parameters are not only laborious to measure but also difficult to reproduce in a deep learning model. Even though the accuracy of reproducing human labels is comparable to human measurement in supervised models, it suffers from high correlation among the parameters themselves, especially when the number of parameters increases, which limits the interpretability of the outcome.¹¹

A good example of a conventional method that analyzes many intercorrelated parameters is principal components analysis (PCA). PCA transforms original variables into a set of new orthogonal variables (or principal components).¹² By choosing the first few principal components, PCA can be used to reduce the data dimensions and could be applied for medical image analysis.^{13,14} Such a procedure could be applied to a large set of parameters defined and measured by experts; however, while PCA effectively accounts for intercorrelation and reduces parameter dimensions, it is not intuitively interpreted and does not guarantee a representation of the full image, including geometric and physical features. To overcome such limitations, a hybrid approach combining CNN-based autoencoder with PCA has been developed for optical coherence tomography images demonstrating good accuracy in discrimination of glaucomatous versus nonglaucomatous optic nerve head as well as excellent visualization of latent structure.¹⁵

Lately, a family of models called deep generative models has been developed with recent advances in the field of deep learning.¹⁶ Such models aim at learning a disentangled latent space, a nonlinear low-dimensional representation of the data space, which can be used to generate new images. “Disentangled” is analogous to “independent” in conventional statistics. One of the widely used generative models is the variational autoencoder (VAE).¹⁷ When combined with CNN architecture, a combination sometimes called a convolutional VAE, a VAE framework becomes a relatively simple but powerful tool for unsupervised learning of disentangled latent space from images. Therefore, VAE architecture can be applied to extract the essence of the entire image in a low-dimensional form via latent space projection, making it suitable for deep learning analysis.

A strong advantage of CNNs over traditional human-measured parameters is visualization. Conventionally, every parameter has been defined by experienced specialists to capture an important aspect of the image. Hence, an interpretation of a single param-

eter, which is a scalar quantity, is straightforward per se, but it becomes increasingly difficult with a larger number of parameters. In contrast, while also being a scalar quantity, every latent variable in a VAE represents certain aspects of the whole image. Also, a latent variable or combination of latent variables can be directly visualized, which is invaluable in terms of interpretation. Once modeled, latent variables can be applied to tasks, such as diagnosis, classification, or longitudinal analysis.

This was a pilot study of unsupervised deep learning aiming to investigate the capacity of VAEs in ophthalmologic image analysis. Using AS-OCT images, we first optimized hyperparameters and then explored latent variables in detail to ensure that the latent variables represent interpretable distinct anterior segment characteristics.

Methods

Participants

We retrospectively reviewed electronic medical records of all participants who had undergone an AS-OCT examination (Visante OCT, version 3.0; Carl Zeiss Meditec, Dublin, CA, USA) under controlled room lighting conditions (0.5 cd/m^2) from the ongoing Asan Glaucoma Progression Study. All participants initially underwent a complete ophthalmic examination, including a review of medical history, measurement of best-corrected visual acuity, measurement of manifest refraction, slit-lamp biomicroscopy, Goldmann applanation tonometry, gonioscopy, fundoscopic examination, stereoscopic optic disc photography, retinal nerve fiber layer photography, and a visual field test (Humphrey field analyzer; Swedish Interactive Threshold Algorithm 24-2; Carl Zeiss Meditec, Jena, Germany). We excluded the following patients: those with acute angle closure, phacomorphic glaucoma, or phagocytic glaucoma; patients with a history of ophthalmic surgery, including laser peripheral iridotomy, implantable Collamer lens insertion, anterior chamber lens implantation, penetrating keratoplasty, cataract surgery, or vitrectomy; patients with secondary glaucoma, including neovascular or uveitic glaucoma; and those with anterior chamber or corneal abnormalities, including iridocorneal synechiae and high iris insertion, iridocorneal syndrome, or keratoconus. If a participant had undergone multiple AS-OCT examinations, the highest-quality image, defined as showing good visibility of the scleral spur, was selected for analysis. Both eyes of the same patient were included if eligible. A 0° to 180° scan acquired in anterior segment quad mode

with a size of 1200×1500 pixels (height by width) was center cropped to create a $512\text{-pixel} \times 1024\text{-pixel}$ image, which was then resized to 256×512 pixels and a grayscale color mode. As a result, 2111 eyes of 1261 patients were included in this analysis.

All procedures conformed to the Declaration of Helsinki, and this study was approved by the institutional review board of the Asan Medical Center, University of Ulsan, Seoul, Korea. The requirement for informed consent was waived due to the study's retrospective design.

Image Preparation With Automated Segmentation

We followed the methods described by Pham et al. to construct a deep CNN for segmenting AS-OCT images into three classes: (1) 130 randomly chosen AS-OCT images from the training set were manually segmented by an experienced glaucoma specialist (KHS) using Fiji¹⁸ software into three segments—the iris, the corneal shell, and the anterior chamber; (2) a modified U-net was trained with manually drawn segmented images; and (3) all images were segmented using a modified U-net (Supplementary Figs. S1 and S2).^{18–20} The resulting segmented images had three channels corresponding to the iris, corneal shell, and anterior chamber, each coded with 1 on a 0 background. Segmented images were aligned with rotation and translation with a spatial-transformer network.²¹

Construction of a Deep Neural Network

Obtaining an interpretable disentangled latent space was key to our study. Ideally, each latent variable would represent a distinct but interpretable aspect of anterior segment configuration, which, in combination, would describe important aspects of an AS-OCT image completely. In reality, an image cannot be perfectly represented by latent space, nor can the latent space be completely disentangled and interpretable. There are trade-offs: (1) increasing the number of latent variables will result in a better overall representation of an AS-OCT image but with less interpretable latent space, and (2) forcing a higher degree of disentanglement might result in a posterior collapse and blurred images with a β -VAE framework. Hence, despite the model being unsupervised, the optimization of hyperparameters was a critical part of the study.

To explore the capability of deep generative models for representing AS-OCT images, we analyzed four steps: (1) construction of a β -VAE model, (2) determination of the optimal size of the latent space,

(3) determination of the optimal value of β by comparing reconstruction accuracy and exploring the latent space, and (4) visualization and interpretation of the latent space in the final model. Actually, steps 2 and 3 were performed concurrently, but we have presented them sequentially for practical reasons.

Constructing a β -VAE Model

A VAE is a probabilistic deep generative model in which the distribution of the latent variables, posterior $p_\theta(z|x)$, is approximated to prior $p_\phi(z)$, most commonly assumed to be a Gaussian distribution $N(0, I)$.¹⁷ A loss function of the VAE is given as

$$L(\theta, \phi, \beta; x) = -\mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] + \beta D_{KL}(q_\phi(z|x)||p(z)) \quad (1)$$

where in our case, x is an image, z is a latent space, $q_\phi(z|x)$ is the estimated probability function of latent factors (probabilistic encoder parametrized with Φ), $p_\theta(x|z)$ is the likelihood of generating a true image given latent factors (probabilistic decoder parametrized with θ), \mathbb{E} is the expected value, and D_{KL} is the Kullback–Leibler divergence (KLD).

The practical problem behind every VAE is a weighting of the reconstruction loss (the first term on the right side of Equation (1)) and KLD. Too much weight on the reconstruction loss will result in a poor disentanglement of the latent space, while too much weight on the KLD will result in a posterior collapse. As our segmented images are essentially one-hot encoded, we have binary cross-entropy for the calculation of reconstruction loss, which is a negative log-likelihood of the Bernoulli distribution. If each pixel in the image were independent, the total log-likelihood of the image would be equal to the sum of the log-likelihoods of the individual pixels. However, this is not applicable to all cases.

To address this problem, we have used a β -VAE framework,²² which has been proposed to enhance disentanglement in VAE, where the loss function is defined as follows:

$$L(\theta, \phi, \beta; x) = -\mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] + \beta D_{KL}(q_\phi(z|x)||p(z)) \quad (2)$$

To specifically address the problem of correlation between image pixels, a weight could also be added to the reconstruction term. But because there is no universal solution for the determination of appropriate weight for reconstruction loss of image data, we decided to manipulate β rather than further complicating the model by the addition of a hyperparameter.

The encoder and decoder components of the β -VAE were constructed according to the model described by Han et al.²³ with rescaling of parameters: (1) input

and output image sizes were scaled up from 128×128 to $192 \times 448 \times 3$, (2) the number of filters in each step was reduced (multiplied by a filter multiplier), and (3) latent space was set to values explained in the following section. To determine the optimal number of filters, we compared models with the number of filters reduced by filter multipliers of 1/16, 1/8, and 1/4. There was no significant improvement in reconstruction accuracy, while some instability issues arose with filter multipliers of 1/8 and 1/4 in certain situations. We decided to use a filter multiplier of 1/16 in all models presented in this article.

Determination of the Optimal Size of the Latent Space

Two additional manipulations were done to segmented images before training the β -VAE: (1) a corresponding horizontally flipped image was created for every image, and (2) margins were cropped out, resulting in images with a size of 192×448 pixels. For training and validation of β -VAE models, we randomly assigned 80% of the data to the training set (1692 eyes of 1007 patients) and 20% to the validation set (419 eyes of 254 patients). Both eyes of the same patient were guaranteed to be assigned to the same group using stratified sampling. After a preliminary explorative step to roughly approximate reasonable values of β , we constructed eight models with β set to 5^3 and another eight models with β set to 5^4 , with latent space dimensions from 4 to 11. The resulting reconstruction accuracy and KLD were compared to determine the optimal size of the latent space.

Determination of the Optimal Value of β

Using a latent space size determined in the previous step, we constructed eight otherwise identical models with differing β values ranging from 5^{-1} to 5^6 , with each step multiplied by a factor of 5. We compared the reconstruction accuracy and KLD of the models, and we inspected selected samples. We explored full latent space for all models, but we present only the most prominent latent variable from each model in the Results section for practical reasons.

Visualization and Interpretation of the Latent Space

Using the hyperparameters determined in previous steps, we trained a final model. Full visualization of latent space for z values of -2 to 2 is presented with manual rearrangement of latent variables from η_1 to η_6 . We also generated a dynamic animated image to aid interpretation.

Table. Baseline Demographics

Characteristic	Training Set	Validation Set	P Value
Age, y	63.2 ± 12.2	64.1 ± 11.9	0.335
Sex, female	675 (67.0)	169 (66.8)	0.960
Laterality, right	852 (50.3)	213 (51.0)	0.817
VFI, %	86.3 ± 23.7	86.9 ± 23.7	0.667
MD, decibels	-5.38 ± 7.69	-5.12 ± 7.83	0.527

Values are presented as mean \pm SD or number (%). MD, mean deviation; VFI, visual field index.

Statistical Analysis

Pytorch (version 1.9.1; Facebook's AI Research lab (FAIR), Menlo Park, CA, USA) running in a Python (version 3.7; Python Software Foundation, Wilmington, DE, USA) environment was used for neural network training. The mean binary cross-entropy multiplied by the number of pixels (258,048) was used to compare original and reconstructed images, and a sum was used to compare KLD. All models were trained with Adam optimizer for 3000 epochs. Analyses were performed using SAS 9.4 (SAS Institute, Inc., Cary, NC, USA). The χ^2 test was used for comparison of categorical variables and the independent Student's t -test for continuous variables.

Results

Demographics

The mean participant age was 64.4 years, with more females (66.9%) than males (33.1%) and all patients being of East Asian ethnicity (Korean). The proportion of left and right eyes was roughly equal (49.5% vs. 50.5%, respectively), while the mean values of visual field index and mean deviation were 86.4% and -5.33 dB, respectively. There were no significant differences in any of the parameters between the training set and validation set (Table).

Determination of the Optimal Latent Space Size

In models with β set to 5^3 , increasing the size of the latent space resulted in improved reconstruction accuracy and KLD (Fig. 1A). In contrast, with β set to 5^4 , reconstruction accuracy did not improve significantly with a latent space size of 6, which could have implied the presence of a posterior collapse (Fig. 1B).

While numerically more accurate, excessively large latent space is likely to generate latent variables that are indiscernible from one another by the human

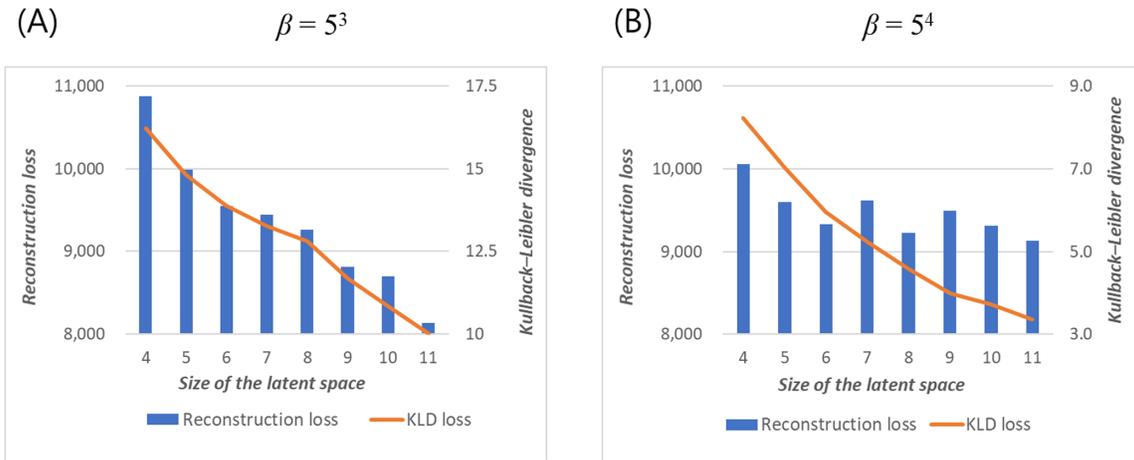


Figure 1. Reconstruction loss (left axis, same in both figures) and KLD (right axis, different between figures) of the test set by latent space dimensions at (A) $\beta = 5^3$ and (B) $\beta = 5^4$.

eye, decreasing the model's interpretability. Hence, we decided to use 6 for the size of the latent space in the following analyses, whereafter some posterior collapse might have occurred at higher β values.

Reconstruction Accuracy and KLD for Different β Values

Reconstruction accuracy was relatively stable until β reached 5^4 , whereafter the reconstructed image became significantly blurry. At $\beta = 5^6$, a complete posterior collapse occurred. While comparing $\beta = 5^3$ and $\beta = 5^4$, reconstruction accuracy was better at $\beta = 5^3$ because of some visually noticeable blurring at $\beta = 5^4$. In sample image 1, the reconstructed image is less accurate at $\beta \geq 5^4$ than $\beta \leq 5^3$. In sample image 4, the thickness of the iris and diameter of the pupil keep decreasing until $\beta = 5^4$ (Fig. 2).

Exploration of Latent Space for Different β Values

We chose one of the most prominent latent variables—that which generated the deepest anterior chamber—from each model for a visual comparison of latent space between different β values. At $\beta = 5^0$, the latent variable did not seem to be properly centered around 0, and the generated images were noisy, but this gradually improved with increasing β . At β values of 5^3 through 5^5 , the main feature represented by the latent variable became clear—the depth of the anterior chamber. A posterior collapse happened at a β value of 5^6 , meaning that manipulating z did not introduce any noticeable change (Fig. 3).

Visualization and Interpretation of the Latent Space in the Final Model

The most prominent latent variable was η_1 , which accounted for the largest variability of anterior chamber dimensions. On detailed exploration, we noticed that with decreasing anterior chamber dimensions, not only depth but width of the anterior chamber also decreased. Additionally, the iris seemed to become more curved, with decreasing anterior chamber dimensions, while cornea seemed to be slightly thicker with the deeper anterior chamber. The second latent variable, η_2 , seemed to be mainly related to pupil diameter. Also, the anterior surface of the lens seemed to shift forward with a larger pupil. The next two latent variables, η_3 and η_4 , seemed to represent differences in iris profiles—both variables generated thinner irides at $z = -2$ and thicker irides at $z = 2$. However, η_4 generated more curved irides than η_3 at $z = -2$, whereas η_3 generated more curved irides than η_4 at $z = 2$. η_5 seemed to be mainly related to corneal curvature and lens vault; at $z = -2$, the corneosclera was flatter while the lens vault was smaller than at $z = 2$. Differences induced by changing z values of η_6 were only subtle but seemed to be mainly related to asymmetry (Fig. 4 and Supplementary Movie S1).

Despite the model being unsupervised, all latent variables were interpretable, suggesting good disentanglement, which can be confirmed by a visual inspection. To summarize, each latent variable seemed to represent mainly the following aspects: the overall depth and area of the anterior chamber (η_1), pupil diameter (η_2), iris profile (η_3 and η_4), corneal curvature and lens vault (η_5), and overall asymmetry (η_6).

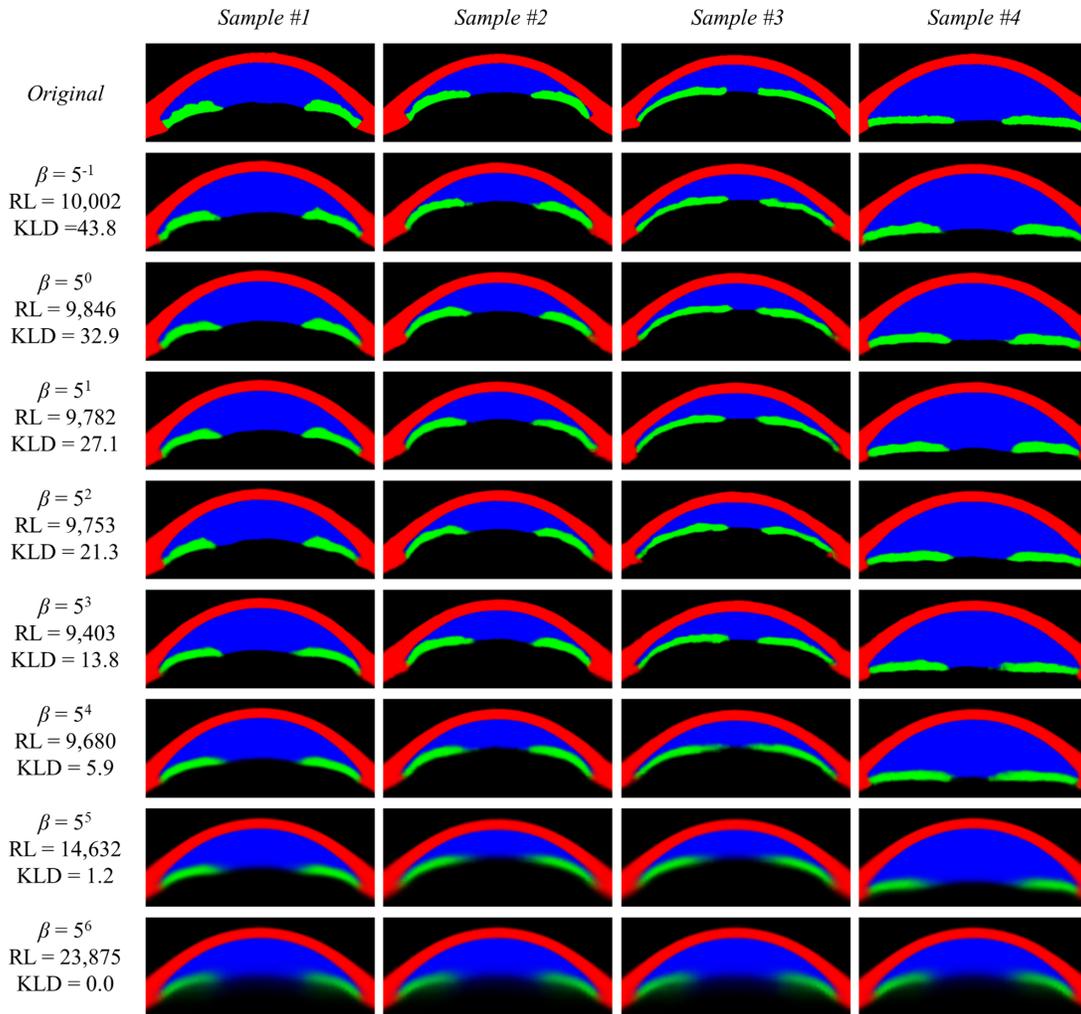


Figure 2. Reconstructed sample images from models with different β values with reconstruction loss calculated on the test set. RL, reconstruction loss.

Discussion

With conventional methods, morphologic analysis of the AS-OCT has been largely dependent on the parameters deduced from clinical experience. In the context of angle assessment, AS-OCT images have been extensively analyzed using various parameters that can be roughly categorized into the following:

1. Anterior chamber dimensions: anterior chamber depth, anterior chamber width, anterior chamber area, anterior chamber volume.
2. Angle width: angle opening distance, angle recess area, trabecular iris angle, trabecular iris space area, scleral spur angle.
3. Iris profile: pupil diameter (PD), iris area (IA), iris curvature, iris thickness.
4. Lens position: lens vault.

The above list can be expanded if we include measurements taken at different locations (e.g., 500 vs. 750 μm from scleral spur), secondary parameters (e.g., relative lens position), or dynamic parameters (e.g., iris area change by different lighting conditions). Given that most anterior chamber parameters are strongly correlated, combining those parameters into a single model is not only technically challenging but also difficult to interpret.¹¹ The problem gets worse when accounting for the dynamic nature of the eye—the pupil diameter changes with lighting conditions, which affect anterior chamber parameters.^{24,25} Also, despite the complexity, while effective in capturing clinically important and prominent characteristics, such parameters have limited ability to represent the AS-OCT image as a whole.

To address this limitation, we have applied a nonlinear latent space model called convolutional β -VAE and examined hyperparameters to find one that fits the best

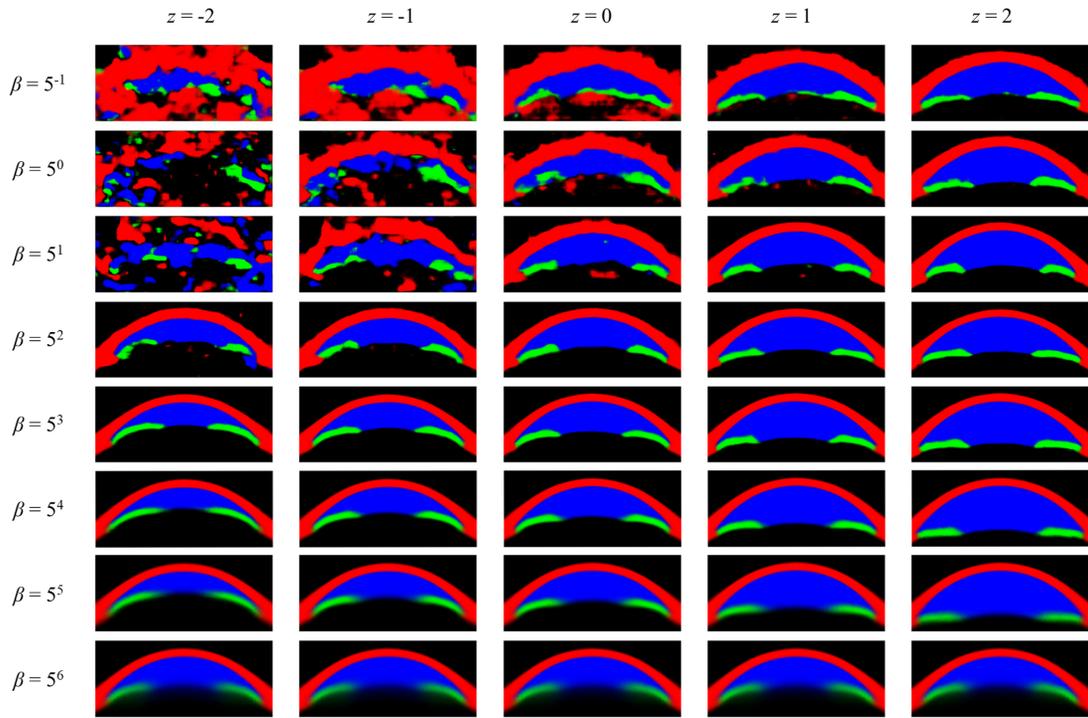


Figure 3. Latent variable generating the deepest anterior chamber for each β value. Sign of z values was adjusted to show a generated image with a deeper anterior chamber on the *right side*.

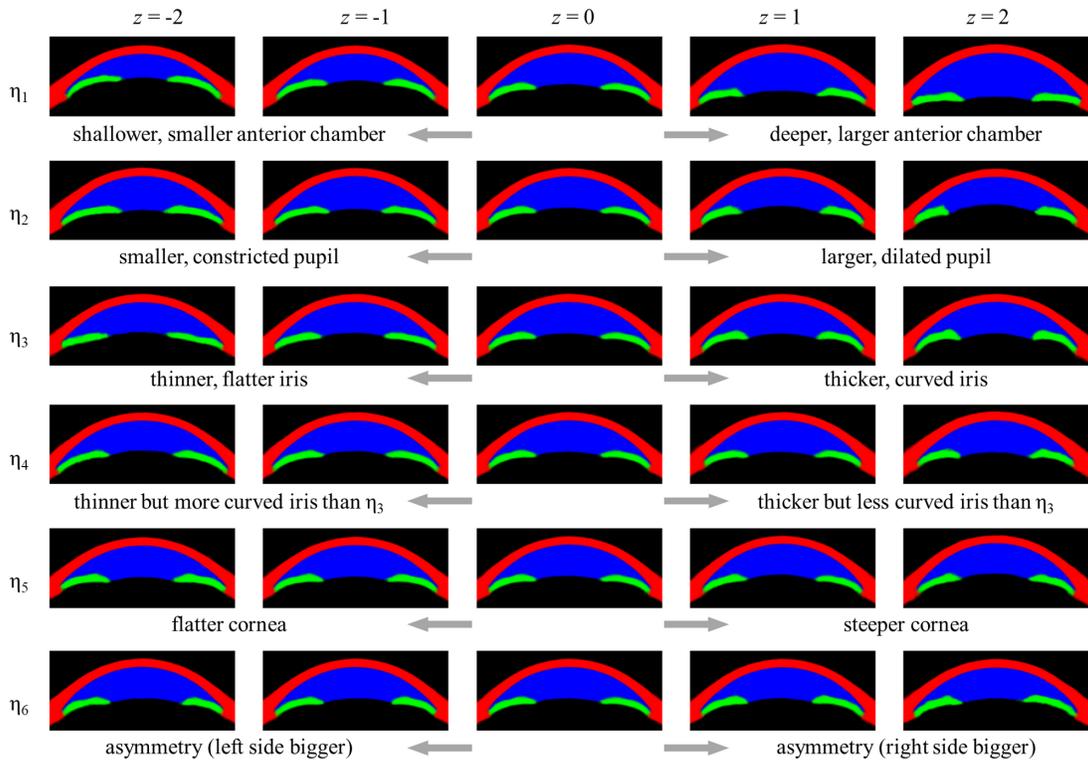


Figure 4. Generated images from latent space of $\beta = 5^3$, with z values from -2 to 2 . Each latent variable that has been altered is denoted with a subscript η , while other latent variables are kept at 0 . The sign of z values was adjusted so that for every latent variable, a generated image with a deeper anterior chamber or larger pupil is displayed on the *right side*.

scenario. Our model demonstrated encouraging results despite being unsupervised, successfully disentangling latent space so that it captures both intuitive and clinically meaningful features of AS-OCT: the overall depth and area of the anterior chamber (η_1), pupil diameter (η_2), iris profile (η_3 and η_4), and corneal curvature (η_5).

Zhang et al.^{26,27} have developed several algorithms for detecting PAC using AS-OCT by mechanisms of angle closure: pupillary block, plateau iris configuration, thick peripheral iris roll, and all three mechanisms. The authors measured 12 anterior segment parameters twice under different lighting conditions. In backward logistic regression models, with the exception of model 1 for plateau iris configuration, all eight remaining models included no more than one parameter from each category: anterior chamber dimensions (anterior chamber depth or volume), angle width (angle recess area), iris profile (iris thickness or curvature), lens vault, and dynamic changes (IA change or IA change/PD change). Therefore, only four to five parameters remained in the final models.

Compared with conventional parameters, latent variables from a β -VAE model have some characteristics more analogous to the actual physiology:

1. η_1 could be a better single representation of the anterior chamber dimension than any single conventional parameter.
2. η_1 could be a good single indicator of angle closure: a smaller anterior chamber is associated with a smaller anterior chamber width (analogous to the phenomenon known as a crowded anterior chamber) as well as a more curved iris (which is hypothesized to happen in a pupillary block).
3. Angle width is determined from a combination of latent variables rather than from a separate entity; the angle in the actual eye is likely to be a result of multiple anatomic and physiologic features.
4. Half of all latent variables (η_2 , η_3 , and η_4) seem to be primarily associated with the iris, with features difficult to parametrize. The iris is a highly diverse component in the anterior chamber.
5. Lens vault seems to increase with a larger pupil, which also seems to happen in the real eye but with a lesser magnitude.^{24,26}

While every latent variable is unique to the specific model, and thus the reproducibility and comparability are limited, once established, the latent space model can provide advantages and possibilities. As the model is mathematically designed to promote disentanglement of the latent space, we can expect the amount of structural information contained to be maximized. Hence,

the latent space model could be better suited for modeling the dynamic patterns of the anterior chamber from differing lighting conditions or degrees of accommodation, which is difficult with conventional parameters. Such models would greatly facilitate multicenter and longitudinal research. Also, latent variables could be used for unsupervised diagnosis of PAC, classification of PAC depending on the mechanism of the disease, or quantitative and qualitative analysis of underlying mechanisms of the disease.

Our model shares many similarities to the model developed by Panda et al.¹⁵ for the analysis of optical coherence tomography images of the optic nerve head—both models involve segmentation, use a convolutional autoencoder, and extract a low-dimensional latent representation of target structure. However, there are also differences—Pandas et al.¹⁵ have combined segmentation and generation of the latent space in one network followed by PCA for further reduction of dimension of the latent space. In contrast, we have built a dedicated network for segmentation while the β -VAE solely focuses on obtaining disentangled low-dimensional latent representation of already segmented images. As a result, we expect the former approach to be more efficient at extracting most informative latent vectors for discrimination of glaucomatous versus nonglaucomatous optic discs, while our approach is better suited for general structural exploration of the anterior segment.

We have shown that careful tailoring of hyperparameters is necessary to achieve a meaningful representation of latent space. Our results suggest that even with a latent space size of 6, the meaning of the last latent variable is less intuitive. While increasing the size of latent space might improve reconstruction accuracy, derived latent space is likely to have worse interpretability. In addition, manual optimization of hyperparameter β is the key to constructing a β -VAE model. While increasing the value of β is expected to promote disentanglement of the latent space, it also might lead to poorer reconstruction. Currently, there is no universally accepted measure to find the optimal value of β , especially for unlabeled data as in our example; hence, visual inspection plays an important role.²² We have found that while the reconstructed image might look similar, the latent space is drastically different depending on the value of β . At lower β values, reconstructed images from a single latent variable looked noisy, somewhat reminiscent of an impressionist or abstract painting. With increasing β , latent space representation improved until β reached 5^3 , whereafter reconstructed images became blurry, and a complete posterior collapse happened at a β of 5^6 . Our example underlines the importance of adjusting hyperparam-

ter β while designing a β -VAE. Also, because the degree of correlation between pixels will be different for every data set, β must be carefully tailored for every model to achieve a meaningful representation of the latent space.

Nonetheless, our model also has demonstrated shortfalls and limitations. While reconstructed images are surprisingly good given the small latent dimension of 6, there is still room for improvement. Also, as can be expected from a VAE framework, increasing β resulted in better latent space disentanglement; however, at higher values, the reconstruction accuracy deteriorated rapidly. Hence, hyperparameters have to be optimized manually, which can be subjective and laborious, a known problem for a VAE framework.²⁸ Another important potential issue includes the balance of labels: the model will be less effective at representing rare instances. Simple categorical labels are relatively straightforward to balance, but balancing images—which are all unique themselves—is more difficult. Manual balancing by the researcher might introduce subjective bias, while there seems to be no universal standard for unsupervised balancing of unlabeled images. Also, while true latent structure of the anterior segment does not change, results from the generative model are dependent on the data set, meaning that the results may differ depending on the subjects and imaging modality. Newer devices would capture more details of the anterior segment, which could be informative for generative models also.

Conclusions

We have shown that a generative model can be applied for disentangled low-dimensional latent space representation of AS-OCT images. Further optimization of the latent space model and analyses using the latent space are warranted.

Acknowledgments

Disclosure: **K. Shon**, None; **K.R. Sung**, None; **J. Kwak**, None; **J.W. Shin**, None; **J.Y. Lee**, None

References

1. Triolo G, Barboni P, Savini G, et al. The use of anterior-segment optical-coherence tomography for the assessment of the iridocorneal angle and its alterations: update and current evidence. *J Clin Med*. 2021;10(2):231.

2. Kwon J, Sung KR, Han S, et al. Subclassification of primary angle closure using anterior segment optical coherence tomography and ultrasound biomicroscopic parameters. *Ophthalmology*. 2017;124(7):1039–1047.
3. Baek S, Sung KR, Sun JH, et al. A hierarchical cluster analysis of primary angle closure classification using anterior segment optical coherence tomography parameters. *Invest Ophthalmol Vis Sci*. 2013;54(1):848–853.
4. Lee KS, Sung KR, Kang SY, et al. Residual anterior chamber angle closure in narrow-angle eyes following laser peripheral iridotomy: anterior segment optical coherence tomography quantitative study. *Jpn J Ophthalmol*. 2011;55(3):213–219.
5. Kwon J, Sung KR, Han S. Long-term changes in anterior segment characteristics of eyes with different primary angle-closure mechanisms. *Am J Ophthalmol*. 2018;191:54–63.
6. Ciresan DC, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*. Menlo Park, CA, USA: AAAI Press/International Joint Conferences on Artificial Intelligence; 2011:1237–1242.
7. Xu BY, Chiang M, Pardeshi AA, et al. Deep neural network for scleral spur detection in anterior segment OCT images: the Chinese American eye study. *Transl Vis Sci Technol*. 2020;9(2):1–10.
8. Wanichwecharungruang B, Kaothanthong N, Pattanapongpaiboon W, et al. Deep learning for anterior segment optical coherence tomography to predict the presence of plateau iris. *Transl Vis Sci Technol*. 2021;10(1):1–10.
9. Fu H, Baskaran M, Xu Y, et al. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. *Am J Ophthalmol*. 2019;203:37–45.
10. Hao H, Zhao Y, Yan Q, et al. Angle-closure assessment in anterior segment OCT images via deep learning. *Med Image Anal*. 2021;69:101956.
11. Shan J, Pardeshi AA, Varma R, et al. Correlations between anterior segment optical coherence tomography parameters in different stages of primary angle closure disease. *Invest Ophthalmol Vis Sci*. 2018;59(9):5896.
12. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdisciplinary Rev Comp Stat*. 2010;2(4):433–459.
13. Nandi D, Ashour AS, Samanta S, et al. Principal component analysis in medical image processing: a study. *Int J Image Mining*. 2015;1(1):65.

14. Gong S, Boddeti VN, Jain AK. On the intrinsic dimensionality of image representations. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, D.C., USA: IEEE Computer Society; 2019:3982–3991.
15. Panda SK, Cheong H, Tun TA, et al. Describing the structural phenotype of the glaucomatous optic nerve head using artificial intelligence. *Am J Ophthalmol*. 2021;236:172–182.
16. Ruthotto L, Haber E. An introduction to deep generative modeling. *GAMM Mitteilungen*. 2021;44(2):1–26.
17. Kingma DP, Welling M. Auto-encoding variational Bayes. In: *2nd International Conference on Learning Representations, ICLR 2014—Conference Track Proceedings*. 2014:1–14, open access via openreview.net.
18. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–682.
19. Pham TH, Devalla SK, Ang A, et al. Deep learning algorithms to isolate and quantify the structures of the anterior segment in optical coherence tomography images. *Br J Ophthalmol*. 2021;105(9):1231–1237.
20. Ronneberger O, Fischer P, U-Net Brox T. Convolutional networks for biomedical image segmentation. *IEEE Access*. 2015;9:16591–16603.
21. Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks. In: *Advances in Neural Information Processing Systems*. Massachusetts, USA: Morgan Kaufmann Publishers Inc.; 2015:2017–2025.
22. Higgins I, Matthey L, Pal A, et al. β -VAE: Learning Basic Visual Concepts With a Constrained Variational Framework, *ICLR 2017*, open access via openreview.net.
23. Han K, Wen H, Shi J, et al. Variational autoencoder: an unsupervised model for modeling and decoding fMRI activity in visual cortex. *Neuroimage*. 2019;198:125–136.
24. Lin J, Wang Z, Chung C, et al. Dynamic changes of anterior segment in patients with different stages of primary angle-closure in both eyes and normal subjects. *PLoS ONE*. 2017;12(5):1–15.
25. Nakamine S, Sakai H, Arakaki Y, et al. The effect of internal fixation lamp on anterior chamber angle width measured by anterior segment optical coherence tomography. *Jpn J Ophthalmol*. 2018;62(1):48–53.
26. Zhang Y, Zhang Q, Li L, et al. Establishment and comparison of algorithms for detection of primary angle closure suspect based on static and dynamic anterior segment parameters. *Transl Vis Sci Technol*. 2020;9(5):1–10.
27. Zhang Y, Dong Z, Zhang Q, et al. Detection of primary angle closure suspect with different mechanisms of angle closure using multivariate prediction models. *Acta Ophthalmologica*. 2021;99(4):e576–e586.
28. Mathieu E, Rainforth T, Siddharth N, Teh YW. Disentangling disentanglement in variational autoencoders. In *International Conference on Machine Learning*. 2019:4402–4412. PMLR.

Supplementary Material

Supplementary Movie S1. An animated movie showing the dynamic changes of generated images. Each image represents a single latent variable each (range $z = -2$ to $z = 2$), while other latent variables are kept at $z = 0$.