2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897646

DISENTANGLED SEQUENTIAL AUTOENCODER WITH LOCAL CONSISTENCY FOR INFECTIOUS KERATITIS DIAGNOSIS

Yuxuan Si[1,2,3], *Zhengqing Fang*[1,2,3] , *Kun Kuang*[1,3], *Zhengxing Huang*[1] , *Yu-Feng Yao*[2,3] *and Fei Wu*[1,3,4,5]

[1] College of Computer Science and Technology Zhejiang University, Hangzhou 310027, China
[2] Zhejiang University School of Medicine Sir Run Run Shaw Hospital, Hangzhou 310027, China
[3] Key Laboratory for Corneal Diseases Research of Zhejiang Province, Hangzhou 310027, China
[4] Shanghai Institute for Advanced Study of Zhejiang University, Shanghai 201203, China
[5] Shanghai AI Laboratory, Shanghai 201203, China

ABSTRACT

Infectious keratitis is a major cause of visual impairment and a common blinding eye disease. Deep learning based prior researches mainly regard infectious keratitis diagnosis as a classification task on the slit-lamp images of single-visit. However, in real clinical applications, it is critical to analyze the lesion evolution characteristics represented by time-varying features over multiple-visits. To bridge this gap, in this paper, we focus on the problem with sequential clinical images of patients, and propose a novel disentangled sequential auto-encoder (DSLC-VAE) algorithm to separate the time-varying pathological features from the time-invariant ones for infectious keratitis diagnosis. Specifically, a inference model is exploited to generate time series of the shape and appearance of corneal lesions to represent keratitis progression, which are combined with location-related features to identify keratitis pathogen. Moreover, we construct a local consistent regularizer with a self-supervised task to enhance the consistency of the time-varying features across different infectious keratitis. Extensive experiments on real world dataset demonstrate superiority of our DSLC-VAE on both representation disentanglement and diagnosis accuracy.

Index Terms— Disentangled Representation Learning, Sequential Learning, Variational Autoencoder, Local Consistency Constraint, Infectious Keratitis

1. INTRODUCTION

Infectious keratitis [1, 2] is a common blinding eye disease worldwide and a major cause of visual impairment. As shown in Fig.1, Bacterial Keratitis (BK), [3], Fungal Keratitis (FK) [4], Herpes Simplex Virus Stromal Keratitis (HSK) [5] and Acanthamoeba Keratitis (AK) [6] are the most common keratitis. Early detection and prompt medical intervention are required to halt the disease progression[3, 7]. Otherwise keratitis can rapidly worsen over time, potentially leading to permanent vision loss or even corneal perforation [8, 9, 10]. In current clinical applications, the average diagnosis accuracy of keratitis is about 50% [11]. Therefore, It is vital to explore further ways for higher diagnostic accuracy.

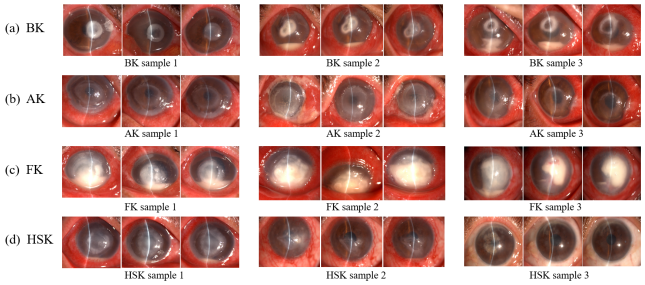


Fig. 1. Examples of four categories of corneal diseases, include bacterial keratitis (BK), acanthamoeba keratitis (AK), fungal keratitis (FK), and herpes simplex viral stromal keratitis (HSK), among which the manifestations of the diseases are subtle for identification.

Recently, many deep learning solutions treating infectious keratitis diagnosis as a image classification problem [11, 12] have been proposed by inputting the slit-lamp image of patient's single visit and predicting its category (e.g., BK, AK, FK, HSK as shown in Fig.1). The SOS [11] method imitated the way ophthalmologists making diagnosis intuitively, they extracted the image features through a sequential learning mechanism from the center of the lesion to the surrounding area. Fang et al. proposed the VCEC [12] focus on the interpretability of infectious keratitis diagnosis by mining the interpretable visual concept of the clinical image. However, these methods ignored the clinical practice of ophthalmologists: do detection and treatment on multiple-visit record. [9, 13]. It would be more practical to develop a model to analyze patient's multiple-visit record(i.e., sequential clinical images).

This work was supported in part by the Program of Zhejiang Province Science and Technology (2022C01044) and the Fundamental Research Funds for the Central Universities (226-2022-00142, 226-2022-00051)

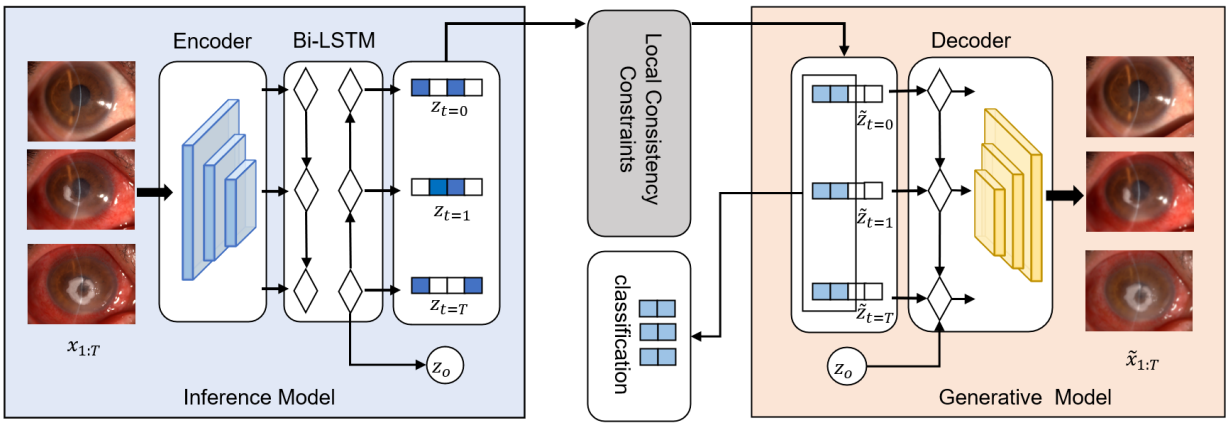


Fig. 2. The framework of our proposed model. Feed the sequential images into the inference model to obtain the manifold posterior of a time-varying latent variable $\{q(z_t \mid \boldsymbol{x}_{\leq t})\}_{t=1}^{T}$ and the posterior of a static latent variable $q(\boldsymbol{z}_o \mid \boldsymbol{x}_{1:T})$. Then, from the corresponding posteriors resample variables z_o and $z_{1:T}$ and concatenate them to feed into the decoder to generate reconstructed sequential images. Local consistency regularizer is imposed on time-varying latent variable to encourage the representation disentanglement.

In this paper, we focus on the problem of infectious keratitis diagnosis by inputting the sequential clinical images of patients and predicting its category. From the experience of ophthalmologists [8, 9, 10], we know the key to diagnose infectious keratitis is the time-varying pathological characteristics cross the sequential clinical images. Inspired by this, we propose a disentangled sequential auto-encoder with local consistency (DSLC-VAE) to learn the disentangled representations of time-varying and time-invariant pathological characteristics for infectious keratitis diagnosis. Specifically, we learn the time-varying and time-invariant features by combining the consensus pattern of sequential images and specific texture of single frame, which leads to the generation of time series of the shape and visual appearance of lesions to represent keratitis progression. Furthermore, they are combined with location-related features to classify keratitis pathogen.

Considering the consistency of the time-varying features across different infectious keratitis, we utilize local consistency regularizer as a self-supervised task to further disentangle stable pathological factors from the time-varying features. Experimental results on real world dataset demonstrate superiority on both representation disentanglement and diagnosis accuracy.

The contributions of the paper include: 1) We study a sequential disentangled representation problem on the real medical multiple-visit application of infectious keratitis classification. 2) We propose a novel disentangled sequential auto-encoder with local consistency (DSLC-VAE) on sequential clinical images. Our approach achieve state-of-the-art diagnosis accuracy. 3) Extensive experiments demonstrate that our extracted time-varying pathological characteristics are practical and precise on keratitis diagnosis.

2. THE PROPOSED APPROACH

2.1. Notation Definition

Let $\mathcal{D} = \{X^i\}_{i=1}^{N}$ denotes the dataset of N independent samples X, where $X \equiv x_{1:T} = (x_1, x_2, ..., x_T)$ indexes a sequence of clinical images collected from patient' T follow-up visits. A variational autoencoder is applied to extract the low-dimensional latent variable Z from X. In order to remove the confounding information irrelevant to diagnosis, the latent variable Z can be further disentangled into a time-invariant factor z_o and a time-varying factor $z_{1:T}$ where z_o refers to the the ocular surface structure and the complications of keratitis and $z_{1:T}$ refers to the pathological features varied over time. pathological features at time $t \in \{1 : T\}$ named z_t helps to improve the accuracy of downstream tasks such as classification of infectious keratitis.

2.2. Disentangled Sequential Autoencoder

2.2.1. *generative model*

The generative model generates high-dimensional sequence image data close to the original sequence data when the two low-dimensional latent variables $z_{1:T}$ and z_o are known. According to whether the generated results are similar to the original high-dimensional data, we can effectively test whether the construction of the low-dimensional manifold space is reasonable. To achieve this goal, we define the following probabilistic generative model by assuming that $z_{1:T}$ and z_o are independent:

$$p(X, Z) = p(z_o) \prod_{t=1}^{T} p(x_t \mid z_o, z_t)\, p(z_t \mid z_{<t}) \qquad (1)$$

For the prior distributions, we choose $z_o \sim \mathcal{N}(0,1)$, and $z_t \mid z_{<t} \sim \mathcal{N}(\mu(z_{<t}), \sigma^2(z_{<t}))$ where $\mu(\cdot)$ and $\sigma(\cdot)$ are modeled by Bi-LSTM.

2.2.2. *inference model*

In order to obtain a low-dimensional manifold space by reducing the high-dimensional data space from the original sequential dataset, we build an inference model. To make full use of the contextual information of sequential data, we choose sequential variational autoencoder for feature extractor model, the posterior distribution inferred by the model is:

$$q(Z \mid X) = q(z_o \mid x_{1:T}) \prod_{t=1}^{T} q(z_t \mid x_{\leq t}) \tag{2}$$

$z_o \sim \mathcal{N}\left(\mu_s(z_o), \sigma_s^2(z_o)\right)$, $z_t \sim \mathcal{N}\left(\mu_d(z_t), \sigma_d^2(z_t)\right)$ where $\mu_s(\cdot)$ and $\sigma_s(\cdot)$ are modeled by static variable encoder ψ_s^{Encoder} take the whole sequence as input, $\mu_d(\cdot)$ and $\sigma_d(\cdot)$ are modeled by time-varying recursive encoder ψ_d^{Encoder} take only the previous frame as input.

The variational inference method is applied to fit approximate posterior distribution $q(z|x)$ from inference model with our prior hypothesis. Since the $q(z|x)$ is parameterized by weight set θ in the network, it can be written as $q_\theta(z|x)$. Similarly, the likelihood distribution $p(x|z)$ from generative model is parameterized by weight set ψ, hence it can be written as $p_\psi(x|z)$. The Kullback-Leibler divergence (KL divergence) between the approximate posterior distribution q_θ and the real posterior distribution is calculated by

$$\begin{aligned} &D_{KL}\left(q_\theta(z \mid x_i) \| p(z \mid x_i)\right) \\ &= -\int q_\theta(z \mid x_i) \log\left(\frac{p(z \mid x_i)}{q_\theta(z \mid x_i)}\right) dz \geq 0 \end{aligned} \tag{3}$$

Through mathematical recurrence, it can be transformed into

$$\begin{aligned} \log p(x_i) \geq &- D_{KL}\left(q_\theta(z \mid x_i) \mid p(z)\right) \\ &+ E_{\sim q_\theta}(z \mid x_i)\left[\log p_\psi(x_i \mid z)\right] \end{aligned} \tag{4}$$

Therefore, objective function of sequential VAE is calculated as the time-varying negative variaional lower bound by accepting the assumptions of the model, as shown in Eq 5.

$$\begin{aligned} \mathcal{L}_{VAE} = &E_{q(z_{1:T}, z_o \mid x_{1:T})}\left[-\sum_{t=1}^{T} \log p(x_t \mid \mathrm{z}_o, z_t)\right] \\ &+ KL\left(q(z_o \mid x_{1:T}) \mid p(z_o)\right) \\ &+ \sum_{t=1}^{T} KL\left(q(z_t \mid x_{\leq t}) \mid p(z_t \mid z_{<t})\right) \end{aligned} \tag{5}$$

2.3. Local Consistency Constraint

Considering that Etiology remains unchanged on patients' follow-up visits, the extent of the pathogen infiltration determines the morphology and manifestation of lesions at different progression. Therefore, the pathological factor is the inherent part of the time-varying factors. Assuming that the dimension of the pathological factor is k , the time-varying factor $z_{1:T}$ can further be disentangled into invariant pathological factor $z_{1:T}^k$. By controlling the value of low-dimensional pathological factors, we can generate corresponding pathological characteristics and manifestations of infectious keratitis at different stages of disease progression. Therefore, we use weakly-supervised disentangled method [14] to constrain $z_{1:T}$ for different disease stages.

We utilize different disease progression sequential data of the same patient as a set of data pairs (X^m, X^n). The time-varying factors extracted from this pair are $(z_{1:T}^m, z_{1:T}^n)$. Suppose S is a set of low-dimensional variables that do not cause changes in the pathological characteristics, and $\bar{S}$ on the contrary. Therefore, the true posterior distribution of (X^m, X^n) should satisfy the following constraints:

$$\begin{aligned} &p(z_i \mid X^m) = p(z_i \mid X^n) \quad \forall i \in S \\ &p(z_i \mid X^m) \neq p(z_i \mid X^n) \quad \forall i \in \bar{S} \end{aligned} \tag{6}$$

The approximate posterior distribution $q_\psi(\hat{z} \mid X)$ we generate should also satisfy this alignment relationship. Since the factor set S is unknown, in order to obtain an estimate of S $\hat{S}$, We select $d-k$ coordinates with the closest KL divergence in the approximate posterior distribution produced by the inference network for (X^m, X^n), where KL divergence is defined as $D_{\mathrm{KL}}\left(q_\psi(\widehat{z_t} \mid \mathrm{X}^m) \mid\mid q_\psi(\widehat{z_t} \mid \mathrm{X}^n)\right)$, d is the dimension of (z_t^m, z_t^n), and k is the dimension of factors that controls the pathological characteristics.

To enforce the constraint, we replace each shared coordinate with the mean value τ of the two posterior probabilities $\left(q_\psi(\widehat{z_t} \mid \mathrm{X}^m), q_\psi(\widehat{z_t} \mid \mathrm{X}^n)\right)$:

$$\widetilde{q_\psi}(\widehat{z_t} \mid X^m) = \begin{cases} \tau & \text{for } \forall i \in \hat{S} \\ q_\psi(\widehat{z_t} \mid X^m) & \text{otherwise} \end{cases} \tag{7}$$

and $\widetilde{q_\psi}(z_t \mid X^n)$ is obtained in the same way.

3. EXPERIMENTATION AND RESULTS

3.1. Dataset

Our dataset contains 2284 images from 867 patients. We took the patients' images of three consecutive visits as a data sample. Specifically, the training set is consisted of 255 randomly selected image sequences of AK, 222 image sequences of BK, 439 image sequences of FK, and 429 image sequences of HSK. The testing set is consisted of 64 randomly selected image sequences of AK, 48 image sequences of BK, 103 image sequences of FK, and 106 image sequences of HSK.

3.2. Implementation details

In the experiments, we set the input sequence length to 3 sheets, the size of image channels to $3 \times 128 \times 128$, the batch-size to 32. We use adaptive moment estimation (Adam) to

Table 1. The Balanced Accuracy and Macro F1 Score comparison of the classification results

Method	Balanced Acc	Macro F1 Score
Resnet50	0.625	0.619
Resnet101	0.643	0.625
Resnet152	0.633	0.630
Densenet121	0.703	0.696
SOS [11]	0.663	0.659
DS-VAE [15]	0.642	0.623
R-WAE [16]	0.649	0.643
Ours without constraint	0.653	0.650
DSLC-VAE	**0.727**	**0.718**

Table 2. The ARI and AMI scores comparison of the clustering results

Method	ARI		AMI	
	z_o	$z_{1:T}$	z_o	$z_{1:T}$
DS-VAE [15]	0.092	0.006	0.123	0.018
R-WAE [16]	0.097	0.011	0.133	0.058
DSLC-VAE	**0.102**	**0.270**	**0.157**	**0.359**

optimize the learning rate of the optimizer and set it as 0.001. For the time-invariant latent variable z_o, we set its dimension to 512, and the time-varying latent variable $z_{1:T}$ dimension to 256. Dimension size of inherent factor of the time-varying latent variable $z^k_{1:T}$ is 128.

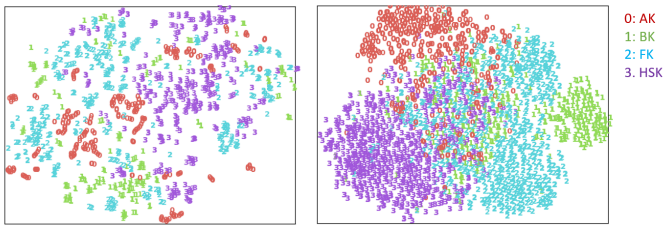
3.3. Results and Discussions

3.3.1. Classification Analysis

We adopt Balanced Accuracy and Macro F1 Score as evaluating metrics. To validate the benefits of local consistency constraint, we conduct ablation experiment. The comparison results are shown in table 1. We can observe that our model outperforms baselines by at least 2% on balanced accuracy, showing the effectiveness and robustness of our method. The performance of our model without local consistency constraint drops slightly, which means representation disentanglement can be encouraged through the self-supervised task in our model.

3.3.2. Clustering Analysis

The self-supervised task constructed by local consistency constraint is an important part in our model. To estimate whether the time-varying features $z_{1:T}$ play a key role in discriminating pathogenic characteristics of different types of infectious keratitis, we utilize t-distributed stochastic neighbor embedding (t-SNE) for visualizing the time-invariant factors z_o and the time-varying factors $z_{1:T}$ by giving each data point a location in a two or three-dimensional map.



(a) clustering result of z_o (b) clustering result of $z_{1:T}$

Fig. 3. Clustering results of time-invariant factor z_o and time-varying factors $z_{1:T}$

According to the cluster visualization results shown as Fig.3, we can find that the time-invariant factor z_o, which represents the ocular surface structure and the complications of keratitis, does not have clustering characteristics. z_o tends to be the same of one sequential data sample, and tends to be different between different samples. For $z_{1:T}$, it represents the intrinsic pathogenic characteristics, the differences in the pathogen of infectious keratitis lead to different manifestations. Thus, $z_{1:T}$ have clustering characteristics.

For a more precise elucidation, we utilize the adjusted rand coefficient (ARI) [17] and adjusted mutual information (AMI) [18] metrics to quantify the clustering results. Compared with other sequential disentangled algorithms, The observation as shown in table 2 agrees with our model design that the representation disentanglement between time-invariant factor z_o and time-varying factors $z_{1:T}$ enables the discriminator to distinguish between ocular surface structure and the complications of keratitis with both lesion shape and disease manifestations, thus in turn improving the classification performance of infectious keratitis.

4. CONCLUSION

In this paper, we study the diagnosis of infectious keratitis with inputting the sequential clinical images of patients and predicting its category. We utilize the time-varying features critical in clinical practice to eliminate the confusion leading to misdiagnosis and propose a disentangled sequential autoencoder with local consistency to achieve a better performance on the representation disentanglement. The proposed method disentangles time-varying features (pathological factors) from time-invariant (confounding factors) in a joint frame-work of image reconstruction and local consistency constraint to intensify the learning of time-varying features. Experiments have verified the superiority of our method and provide a pathological characteristics disentangled representation results on the diagnosis of infectious keratitis for future study.

5. REFERENCES

[1] Lamprini Papaioannou, Michael Miligkos, and Miltiadis Papathanassiou, "Corneal collagen cross-linking for infectious keratitis: a systematic review and meta-analysis," *Cornea*, vol. 35, no. 1, pp. 62–71, 2016.

[2] Ariana Austin, Tom Lietman, and Jennifer Rose-Nussbaumer, "Update on the management of infectious keratitis," *Ophthalmology*, vol. 124, no. 11, pp. 1678–1689, 2017.

[3] T Bourcier, F Thomas, V Borderie, C Chaumeil, and L Laroche, "Bacterial keratitis: predisposing factors, clinical and microbiological review of 300 cases," *British Journal of Ophthalmology*, vol. 87, no. 7, pp. 834–838, 2003.

[4] M Srinivasan, "Fungal keratitis," *Current opinion in ophthalmology*, vol. 15, no. 4, pp. 321–327, 2004.

[5] Jared E Knickelbein, Robert L Hendricks, and Puwat Charukamnoetkanok, "Management of herpes simplex virus stromal keratitis: an evidence-based review," *Survey of ophthalmology*, vol. 54, no. 2, pp. 226–234, 2009.

[6] Christopher D Illingworth and Stuart D Cook, "Acanthamoeba keratitis," *Survey of ophthalmology*, vol. 42, no. 6, pp. 493–508, 1998.

[7] Usha Gopinathan, Savitri Sharma, Prashant Garg, and Gullapalli N Rao, "Review of epidemiological features, microbiological diagnosis and treatment outcome of microbial keratitis: experience of over a decade," *Indian journal of ophthalmology*, vol. 57, no. 4, pp. 273, 2009.

[8] Vinay Agrawal, Jyotirmay Biswas, HN Madhavan, Gurmeet Mangat, Madhukar K Reddy, Jagjit S Saini, Savitri Sharma, and M Srinivasan, "Current perspectives in infectious keratitis," *Indian journal of ophthalmology*, vol. 42, no. 4, pp. 171, 1994.

[9] Nikhil S Gokhale, "Medical management approach to infectious keratitis," *Indian journal of ophthalmology*, vol. 56, no. 3, pp. 215, 2008.

[10] Lei Wang, Kuan Chen, Han Wen, Qinxiang Zheng, Yang Chen, Jiantao Pu, and Wei Chen, "Feasibility assessment of infectious keratitis depicted on slit-lamp and smartphone photographs using deep learning," *International journal of medical informatics*, vol. 155, pp. 104583, 2021.

[11] Yesheng Xu, Ming Kong, Wenjia Xie, Runping Duan, Zhengqing Fang, Yuxiao Lin, Qiang Zhu, Siliang Tang, Fei Wu, and Yu-Feng Yao, "Deep sequential feature learning in clinical image classification of infectious keratitis," *Engineering*, vol. 7, no. 7, pp. 1002–1010, 2021.

[12] Zhengqing Fang, Kun Kuang, Yuxiao Lin, Fei Wu, and Yu-Feng Yao, "Concept-based explanation for fine-grained images and its application in infectious keratitis classification," in *Proceedings of the 28th ACM international conference on Multimedia*, 2020, pp. 700–708.

[13] Noopur Gupta, Radhika Tandon, Sanjeev K Gupta, V Sreenivas, and Praveen Vashist, "Burden of corneal blindness in india," *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, vol. 38, no. 4, pp. 198, 2013.

[14] Francesco Locatello, Ben Poole, Gunnar Rätsch, Bernhard Schölkopf, Olivier Bachem, and Michael Tschannen, "Weakly-supervised disentanglement without compromises," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6348–6359.

[15] Yingzhen Li and Stephan Mandt, "Disentangled sequential autoencoder," *arXiv preprint arXiv:1803.02991*, 2018.

[16] Jun Han, Martin Renqiang Min, Ligong Han, Li Erran Li, and Xuan Zhang, "Disentangled recurrent wasserstein autoencoder," *arXiv preprint arXiv:2101.07496*, 2021.

[17] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis, "Cluster validity methods: part i," *ACM Sigmod Record*, vol. 31, no. 2, pp. 40–45, 2002.

[18] Nguyen Xuan Vinh, Julien Epps, and James Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *The Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, 2010.