Enhancing Diagnostic Accuracy in Rare and Common Fundus Diseases with a Knowledge-Rich Vision-Language Model

Meng Wang[a,b*], Tian Lin[c*], Aidi Lin[c], Yih Chung Tham[a,b], Dianbo Liu[a,b], Wendy Wong[a,b], Sahil Thakur[d], Beau Fenner[d,e], Haoyu Chen[c✉], Huazhu Fu[f✉], Ching-Yu Cheng[a,b✉]

[a] *Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.*

[b] *Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.*

[c] *Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, 515041 Shantou, Guangdong, China.*

[d] *Singapore Eye Research Institute, Singapore National Eye Centre, Republic of Singapore.*

[e] *Ophthalmology & Visual Sciences Academic Clinical Program (EYE ACP), Duke-NUS Medical School, Singapore.*

[f] *Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore.*

** Co-first authors.*

✉*Co-corresponding authors and contributed equally.*

1. Introduction

Recent advances in artificial intelligence (AI) have led to promising fundus disease screening systems for disease detection and patient referral, but most are tailored to specific diseases—such as diabetic retinopathy[1,2], glaucoma[3,4], and retinopathy of prematurity[5,6]—and are trained on task-specific datasets. This specialization results in misdiagnosis when encountering new data (e.g., images from different cameras) or adapting to new or rare disease categories. Collecting comprehensive datasets covering all fundus abnormalities is challenging due to limited healthcare resources and varying disease prevalence, restricting AI models' feature representation, and necessitating extensive retraining for different real-world applications. While large foundation models (LFMs) have excelled in computer vision tasks by providing rich feature support for downstream applications[7,8], current ophthalmic LFMs are pre-trained on extensive yet categorically limited datasets. To address these challenges, we collected 341,896 fundus image–text pairs encompassing over 400 retinal and optic nerve diseases from diverse sources across multiple countries, regions, and ethnicities. This study developed RetiZero, an LFM based on a contrastive vision–language pretraining framework that integrates masked autoencoder-based pretraining knowledge and low-rank training methods. Additionally, we introduced an uncertainty vision–language feature calibration method using Dirichlet reparameterization to further align vision and language features in high-dimensional embedding space. Consequently, RetiZero achieved superior performance across various downstream tasks, marking a significant advancement in ophthalmic artificial intelligence.

2. Methods

To address these problems and challenges, we collected 341,896 fundus images paired with text descriptions from publicly available datasets, ophthalmology literatures, and online resources, encompassing over 400 retinal and optic nerve diseases. As shown in Fig. 1, RetiZero is based on a contrastive vision-language pretraining framework that integrates MAE-based pretraining knowledge and low-rank training methods. Moreover, we introduced an uncertainty vision-language feature calibration method using Dirichlet reparameterization within the contrastive vision-language pretraining framework, to further better align vision-language features in the high-dimensional embedding space. Consequently, RetiZero achieved superior performance in various downstream tasks.

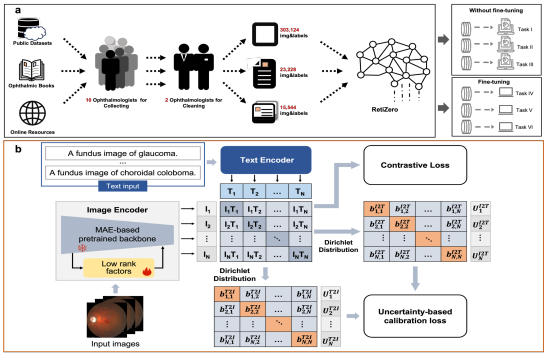


Fig. 1: Overview of the framework. a, Datasets for RetiZero pretraining. b, RetiZero, which combines the strengths of self-supervised learning based on the MAE architecture and contrastive learning from the CLIP architecture.

3. Results

The biggest advantage of RetiZero is the capability of zero-shot learning, which enables RetiZero to recognize fundus diseases using only textual prompts, without needing to retrain or fine-tune the model with labelled fundus images. As shown in Fig. 2a, RetiZero achieved overall Top-1, Top-3, and Top-5 scores of 0.442, 0.702, 0.840, respectively, for recognizing 15 common fundus diseases and normal condition of 30,089 fundus images (Fig. 2e). To further validate RetiZero's zero-shot capability in more challenging clinical scenarios, we assembled a more demanding dataset named EYE-52 (including 7,007 fundus images from various ophthalmology clinics, covering 52 fundus diseases, Fig. 2f.). As depicted in Fig. 2b, RetiZero achieved overall Top-1, Top-3, and Top-5 scores of 0.360, 0.626, and 0.756, respectively, for recognizing these 52 types of fundus diseases in a zero-shot manner. Fig. 2c further illustrates the excellent performance of RetiZero in identifying 15 fundus diseases through image-to-image retrieval. The overall scores for Top-1, Top-3,

Enhancing Diagnostic Accuracy in Rare and Common Fundus Diseases with a Knowledge-Rich Vision-Language Model

Meng Wang[a,b*], Tian Lin[c*], Aidi Lin[c], Yih Chung Tham[a,b], Dianbo Liu[a,b], Wendy Wong[a,b], Sahil Thakur[d], Beau Fenner[d,e], Haoyu Chen[c✉], Huazhu Fu[f✉], Ching-Yu Cheng[a,b✉]

[a] *Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.*

[b] *Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.*

[c] *Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, 515041 Shantou, Guangdong, China.*

[d] *Singapore Eye Research Institute, Singapore National Eye Centre, Republic of Singapore.*

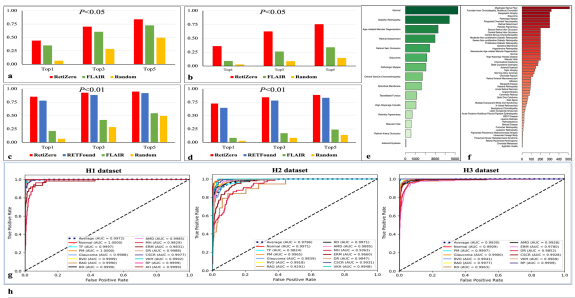
[e] *Ophthalmology & Visual Sciences Academic Clinical Program (EYE ACP), Duke-NUS Medical School, Singapore.*

[f] *Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore.*

** Co-first authors.*

✉*Co-corresponding authors and contributed equally.*

and Top-5 are 0.854, 0.928, and 0.950, respectively. In the more challenging Eye-52 dataset, RetiZero achieved overall Top-1, Top-3, and Top-5 scores of 0.726, 0.843, and 0.886, respectively (Fig. 2d). For fundus disease identification, RetiZero achieved average AUCs of 0.9972, 0.9796, and 0.9930 on the three datasets (Fig. 2h), respectively, each encompassing 15, 13, and 12 different categories of retinal diseases/normal condition, respectively.



Category	H1 dataset	H2 dataset	H3 dataset	Total
AH	60	/	/	60
AMD	1,443	837	1,011	3,291
CSCR	429	534	752	1,715
DR	1,661	1,263	1,498	4,422
ERM	301	500	819	1,620
Glaucoma	1,026	682	452	2,160
MH	265	322	/	587
Normal	2,125	1,463	1,693	5,281
PM	863	231	978	2,072
RAO	183	94	283	560
RD	1,359	677	944	2,980
RP	307	/	/	685
RVO	650	402	378	2,446
TF	478	661	/	1,139
VKH	264	146	1,394	1,071
Total	11,414	7,812	10,863	30,089

AH: Asteroid Hyalosis, AMD: Age-related Macular Degeneration, CSCR: Central Serous Chorioretinopathy, DR: Diabetic Retinopathy, ERM: Epiretinal Membrane, MH: Macular Hole, PM: Pathologic Myopia, RAO: Retinal Artery Occlusion, RD: Retinal Detachment, RP: Retinitis Pigmentosa, RVO: Retinal Vein Occlusion, TF: Tessellated Fundus, VKH: Vogt-Koyanagi-Harada disease.

Fig. 2: a, The overall Top-1, Top-3, and Top-5 scores for zero-shot performance on EYE-15 dataset. b, The overall Top-1, Top-3, and Top-5 scores image-to-image retrieval performance on EYE-15 dataset. c, The zero-shot performance on the EYE-52 dataset. d, Image-to-image retrieval performance on EYE-52 dataset. e, Data distribution of EYE-15. f, Data distribution of EYE-52. g, ROC curves for fundus disease identification. h, Data distribution of different datasets.

4. Conclusion

In this study, we trained a vision-language-foundation model, RetiZero, using a vast amount of image-text pairs. Comprehensive experimental results demonstrated that RetiZero has strong capability in representing fundus disease features across a wide range of downstream tasks of fundus disease identification, including zero-shot recognition, image-to-image retrieval, and fundus disease identification. The performance of RetiZero is superior to two state-of-the-art ophthalmic LFMs, RETFound and FLAIR. These results collectively demonstrated the superior performance of RetiZero in both common and rare fundus disease identification.

References

[1]. Bellemo, V., et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. The Lancet Digital Health. 1, e35-e44 (2019).

[2]. Xie, Y., et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. The Lancet Digital Health, 2, e240-e249 (2020).

[3]. Liu, H., et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA ophthalmology, 137, 1353-1360 (2019).

[4]. Wang, M., et al. Characterization of central visual field loss in end-stage glaucoma by unsupervised artificial intelligence. JAMA ophthalmology, 138, 190-198, (2020).

[5]. Peng, Y., et al. Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. IEEE Transactions on Medical Imaging, 40, 1750-1762 (2021).

[6]. Taylor, S., et al. Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning. JAMA ophthalmology, 137, 1022-1028 (2019).

[7]. Zhou, Y., et al. A foundation model for generalizable disease detection from retinal images. Nature, 622, 156-163 (2023).

[8].Silva-Rodriguez J, et al. A foundation language-image model of the retina (flair): Encoding expert knowledge in text supervision. Medical Image Analysis, 2025, 99: 10335.