

ReDIVE: A Two-Stage Hybrid Network for Keratitis Diagnosis via Retinex-based Enhancement and Depth-Informed Vertical Embedding

Anxu Liao^{1,2}, Ling-Xiao Zhang¹, Fang-Lue Zhang³, Jie Zhou¹, Mengying Qi⁴,
Fen Chen⁴, Qingyan Zeng⁴, Lin Gao^{1,2}, and Yiqiang Chen^{1,*}

¹ Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
liaoanxu20@mails.ucas.ac.cn, {zhanglingxiao, zhoujie, gaolin, yqchen}@ict.ac.cn

² University of Chinese Academy of Sciences, Beijing, China
³ the University of New South Wales, Sydney, Australia
fanglue.zhang@unsw.edu.au

⁴ Wuhan Aier Hankou Eye Hospital, Wuhan, China
{qimengying2014@163.com, 1402796997@qq.com, zengqingyan@aierchina.com}

Abstract. Infectious keratitis is a leading cause of corneal blindness worldwide, and in vivo confocal microscopy (IVCM) has become a key non-invasive diagnostic tool for this condition. However, existing methods predominantly rely on single-frame classification, ignoring pathogen-specific infection depths and frequently causing misdiagnosis. Furthermore, inherent imaging degradations like low signal-to-noise ratios and poor contrast severely compromise classification accuracy. To address these challenges, we propose ReDIVE, a Retinex-enhanced Depth-Informed Vertical Embedding framework for sequence-level infectious keratitis classification. ReDIVE integrates a customized, differentiable multi-scale Retinex preprocessing module, a robust encoder-decoder network, and a specialized depth embedding module. This approach leverages clinical prior knowledge to enhance sequence analysis while significantly improving image quality. Experiments on a large-scale dataset demonstrate that ReDIVE outperforms state-of-the-art methods in accuracy, F1-score, and AUROC, establishing a new paradigm for IVCM-based diagnosis.

Keywords: In Vivo Confocal Microscopy · Infectious Keratitis · Depth-Aware Classification · Retinex-Based Image Enhancement · Sequence-Level Diagnosis.

1 Introduction

Infectious keratitis is a leading cause of corneal blindness, necessitating timely diagnosis to prevent irreversible visual impairment [1]. In vivo confocal microscopy

* Corresponding author.

(IVCM) has become a key diagnostic tool, allowing non-invasive, high-resolution visualization of corneal microstructural abnormalities at the cellular level [2, 3]. Although numerous methods [4] have been developed to classify infectious keratitis using IVCM images, the majority of state-of-the-art methods treat IVCM analysis as a generic single-image classification task [5, 6]. These methods fail to account for the spatial correlations between images. Consequently, leveraging the inherently volumetric nature of IVCM data for robust diagnosis remains a critical task.

In contrast to these single-image approaches, Li et al. [7] attempt to utilize sequence-level information for diagnosing fungal keratitis. However, this formulation assumes that simple sequential continuity is sufficient for diagnosis, ignoring the explicit anatomical depth relationships between images. In clinical practice, pathogen infection is intrinsically linked to corneal depth; for instance, fungal and *Acanthamoeba* keratitis present characteristic signatures within distinct corneal layers, while viral keratitis often triggers varying inflammatory patterns across the corneal stroma [8–11]. Without accounting for these depth-level clinical priors, this method struggles to differentiate keratitis types that may exhibit similar cellular-level features but diverge significantly in their depth distributions.

Furthermore, image preprocessing is essential to mitigate inherent IVCM degradations like low signal-to-noise ratios and non-uniform illumination [12]. For instance, Liu et al. [13] employed adaptive histogram equalization, while Liang et al. [14] emphasized pathological textures via structure-aware approaches. Among these, Retinex models excel in simultaneous noise suppression and illumination correction [15]. However, traditional non-differentiable variational Retinex methods require intensive iterative computations and hand-crafted hyperparameters, incurring prohibitive costs and restricting generalization across clinical samples.

To overcome these limitations, we propose ReDIVE, a two-stage hybrid network for sequence-level infectious keratitis classification integrating widely validated image preprocessing and clinical prior-guided depth-aware learning. A Customized Differentiable Multi-Scale Retinex (CDMSR) module first rapidly and effectively enhances IVCM image texture features. Stage one employs a ConvNeXt encoder to extract frame-level visual features, fusing them with high-frequency absolute and relative depth embeddings. Stage two uses a Vision Transformer (ViT) decoder to model inter-frame correlations from these tokens for final diagnosis.

Overall, our contributions are summarized as follows:

- We propose ReDIVE, the first depth-informed framework for IVCM sequence classification, enabling rapid and accurate infectious keratitis diagnosis.
- We introduce a depth-informed feature embedding mechanism that encodes high-frequency depth cues into a transformer architecture. In addition, we design a differentiable Retinex-based enhancement network to correct non-uniform illumination and suppress noise, thereby improving classification performance.
- Extensive experiments on a newly collected large-scale IVCM dataset demonstrate that ReDIVE consistently outperforms state-of-the-art methods.

2 Methodology

2.1 Preliminaries

Let $\mathcal{X} = \{x_1, x_2, \dots, x_L\} \in \mathbb{R}^{L \times H \times W}$ denote an input IVC sequence consisting of L grayscale frames, and $\mathcal{D} = \{d_1, d_2, \dots, d_L\} \in \mathbb{R}^L$ represent the corresponding depth metadata. Our goal is to learn a mapping function $\mathcal{F} : (\mathcal{X}, \mathcal{D}) \rightarrow y$, where $y \in \{1, \dots, C\}$ is the pathogen category. As illustrated in Fig. 1, the proposed ReDIVE (Retinex-enhanced Depth-Informed Vertical Embedding) framework decomposes the mapping function $\mathcal{F}$ into three tightly coupled components: an adaptive enhancement module, a hybrid depth-informed embedding mechanism, and a progressive optimization scheme.

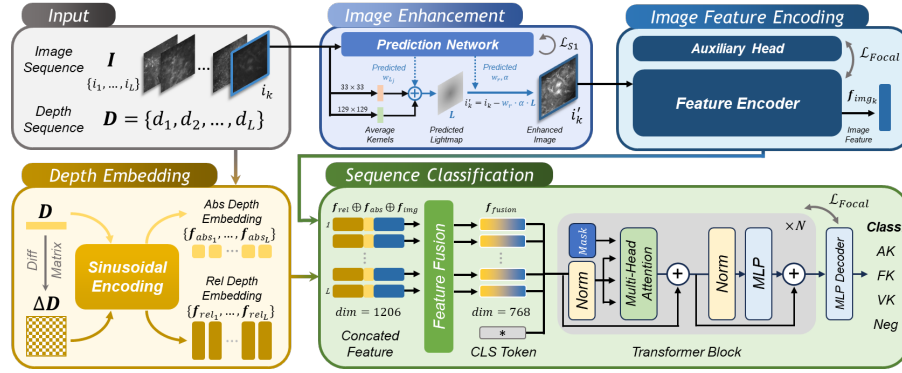


Fig. 1. Overview of the ReDIVE framework. Given length- L image and depth sequences (I_L, D_L) , the image enhancement module (CDMSR) generates enhanced images $I'_k \in I'_L$, which a fine-tuned encoder converts into visual features $f_{vis_k} \in f_{img}$. Simultaneously, sinusoidal functions encode depth scalars $d_k \in D_L$ into absolute depth vectors $f_{abs_k} \in D_{abs}$, while D_L forms a relative depth matrix $\Delta D \in \mathbb{R}^{L \times L}$ yielding relative depth vectors $f_{rel_k} \in D_{rel}$. These components are projected into fused features $f_{fusion_k} \in F_{fusion}$ as ViT-like input tokens. Finally, ViT self-attention processes the concatenated sequence $[B, L, 768]$ to capture correlations for sequence classification.

2.2 Customized Differentiable Multi-Scale Retinex

Classic Retinex theory models an image I as the composition of reflectance R and illumination L [16]. Inspired by this, our proposed Customized Differentiable Multi-Scale Retinex (CDMSR) module learns L from I to decouple R , fusing R with the original I to yield the enhanced output. This fully differentiable pipeline dynamically adapts to infectious keratitis textures, emphasizing pathogen-related semantics. The core processing logic proceeds as follows:

Following an initial min-max stretching of the input image I , we employ a lightweight MobileNetV3 [17] hyper-network to predict pixel-wise weighting

maps directly from the normalized image. The hyper-network extracts multi-scale features and decodes them to generate spatially adaptive parameters: a modulation map α and an illumination discount weight w_r . These parameters are constrained via sigmoid and softmax functions to ensure numerical stability and bound the enhancement extent.

To capture illumination variations across different spatial frequencies, we apply a bank of convolutional kernels G_{σ_j} with varying spatial scales. By evaluating combinations of varying numbers and sizes of averaging kernels, we empirically set the number of scales to $m = 2$ with kernel sizes $\sigma \in \{33, 129\}$. The multi-scale illumination component L is computed as a weighted average of these responses:

$$L = \sum_{j=1}^m w_{L,j} (I * G_{\sigma_j}) \quad (1)$$

where $w_{L,j}$ represents learnable, softmax-normalized scale weights.

We first estimate the spatial-map modulated reflectance R :

$$R = I - \alpha \odot L \quad (2)$$

To preserve dark-region details, we fuse the original image I with R using the predicted weight w_r :

$$I' = (1 - w_r) \odot I + w_r \odot R \quad (3)$$

Substituting R simplifies the entire pipeline to a spatially adaptive illumination discounting operation:

$$I' = I - (\alpha \odot w_r) \odot L \quad (4)$$

Empirically, computing these variables in logarithmic space yields better enhancement. Finally, we restore global contrast via robust linear normalization, clipping at the 99.8th percentile intensity.

2.3 Hybrid Depth-Informed Feature Embedding

To leverage both detailed spatial textures and volumetric context, we construct a hybrid architecture combining CNNs and Transformers (Fig. 1).

We employ *ConvNeXt V2 Base*[18] as the image feature encoder. This selection is grounded in two factors: first, it demonstrated superior classification performance in our empirical experiments; second, the inherent inductive bias of CNNs is crucial for modeling local pathological features in medical images. The enhanced frame I'_{raw} is transformed into a 1024-dimensional feature vector f_{img} .

Depth metadata is explicitly encoded via high-frequency sinusoidal functions to provide pathogen observation depth context. With embedding dimensions set empirically, absolute depth (f_{abs}) maps the slice depth d_{abs} into a 128-dimensional vector, while relative depth (f_{rel}) encodes the inter-frame difference ΔD into a 54-dimensional representation ($L \times 6, L = 9$) to capture inter-slice correlations.

To bridge the semantic gap between visual textures and geometric depth signals, we project the heterogeneous features into a unified latent space. Let $\oplus$ denote channel-wise concatenation. The fused feature $f_{fusion}^{(i)}$ for the i -th frame is formulated as:

$$f_{fusion_k} = \mathbf{W}_{proj} \cdot (f_{img_k} \oplus f_{abs_k} \oplus f_{rel_k}) + \mathbf{b}_{proj} \quad (5)$$

where $\mathbf{W}_{proj} \in \mathbb{R}^{768 \times (1024+128+54)}$, $\mathbf{b}_{proj} \in \mathbb{R}^{768}$ is the learnable projection matrix and bias vector.

2.4 Sequence Classification with Vision Transformer

A Vision Transformer [19] processes the fused feature sequence $F_{seq} \in \mathbb{R}^{B \times L \times 768}$. Each frame-level feature acts as a distinct token, analogous to a spatial patch in standard ViT. Given vectorized inputs, we disable standard positional encoding, as empirical results confirm its inclusion has no impact on performance; instead, fused depth embeddings natively correlate inter-frame spatial relationships. We employ this architecture as self-attention naturally models sequential dependencies and global context. To handle variable sequence lengths, a padding mask M ignores zero-padded frames within the self-attention mechanism:

$$\text{Attention}(Q, K, V, M) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} + M \right) V \quad (6)$$

The final class token predicts four categories: Fungal Keratitis, Viral Keratitis, Acanthamoeba Keratitis, and Negative Controls.

2.5 Progressive Training Strategy

To ensure stability and optimal convergence, we adopt a three-stage progressive training strategy (for specific hyperparameters and training settings, please refer to Section 3.2 Implementation Details):

Stage 1: Unsupervised Image Enhancement. Inspired by traditional Retinex optimization method [20], we optimize the CDMSR module using a compound loss $\mathcal{L}_{S1}$ that balances three complementary objectives: parameter smoothness, content fidelity, and texture amplification.

$$\mathcal{L}_{S1} = \lambda_{tv} \mathcal{L}_{TV}(\Theta) + \lambda_{con} \|I'_{raw} - I_{in}\|_1 + \lambda_{ent} \mathcal{L}_{ent}(I'_{raw}) \quad (7)$$

The total variation loss $\mathcal{L}_{TV}$ acts on the predicted parameter maps Θ (including $w_{L,j}, w_r, \alpha$), ensuring spatially continuous enhancement and preventing abrupt artifacts. This translates the illumination smoothness prior (e.g., $\|\nabla L\|$ in [20]) into direct regularization on modulation parameters. The normalized pixel-wise L_1 term preserves global structural consistency between the enhanced output and the min-max stretched input I_{in} , thereby avoiding distortion caused by extreme image enhancement. Finally, the entropy maximization loss $\mathcal{L}_{ent}$ acts as reflectance gradient regularization, encouraging locally diverse gradients to

amplify fine-grained pathological features. All loss weights are determined empirically to ensure effective and stable enhancement.

Stage 2: Feature Encoder Pre-training. The CDMSR module is frozen, and ConvNeXt is trained to enhance discriminative feature learning at the single-frame level.

Stage 3: Sequence Classifier Training. The ViT module is introduced to aggregate temporal information across frames.

To address the inherent class imbalance within the dataset, the classification networks in both Stage 2 and Stage 3 are optimized using the Focal Loss, formulated as:

$$\mathcal{L}_{Focal} = - \sum_{c=1}^C (1 - p_c)^\gamma \log(p_c) \quad (8)$$

where p_c represents the predicted probability for class c , and γ is the focusing parameter. In Stage 3, after the initial temporal aggregation, the ConvNeXt backbone is unfrozen for full end-to-end optimization.

3 Experiment

3.1 Setup

We evaluate our approach on a large-scale, in-house in vivo confocal microscopy (IVCM) dataset collected between 2017 and 2024. The dataset contains 15,572 images with depth meta-information, categorized into four classes: Acanthamoeba Keratitis (AK), Viral Keratitis (VK), Fungal Keratitis (FK), and Negative Controls (Neg). The training set consists of 13,206 images from 86 patients (AK: 1,278/17, VK: 4,334/16, FK: 1,707/16, Neg: 5,887/37 [images/cases]). The test set comprises 2,366 images from 15 patients (AK: 637/4, VK: 556/3, FK: 561/4, Neg: 612/4). Ground truth diagnoses were confirmed via gold-standard microbiological culture, with image-level annotations provided by at least two ophthalmologists in a double-blind manner. All images were acquired using the Heidelberg Retina Tomograph III system at a resolution of 384×384 , with depth meta-information automatically recorded during acquisition.

3.2 Implementation Details

Implemented in PyTorch 2.5.1 on a single NVIDIA RTX A6000 GPU, our method adopts a three-stage linear probing-then-fine-tuning strategy to ensure stable convergence and prevent catastrophic forgetting.

Stage 1: Unsupervised Image Enhancement. We independently train the CDMSR module for 4 epochs to align visual domains. Optimization utilizes AdamW (learning rate 1×10^{-4} , weight decay 1×10^{-2}) and a dynamically weighted objective combining parametric total variation, content preservation, and entropy maximization losses.

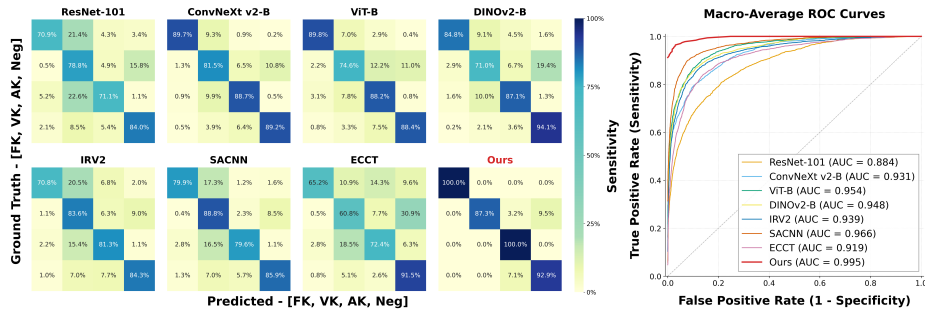
Stage 2: Feature Encoder Training. We train the ImageNet-21k initialized ConvNeXt V2 Base encoder for 50 epochs. This comprises 10 linear probing

Table 1. Quantitative comparison with state-of-the-art methods and ablation study on the IVCN dataset. The best results are highlighted in **bold**.

Method	Params (M)	FPS $\uparrow$	Acc (%) $\uparrow$	F1 (%) $\uparrow$	AUROC $\uparrow$
ResNet-101 [21]	42.51	117.75	76.20	76.66	0.8836
ConvNeXt v2-B [18]	87.51	83.73	87.36	87.48	0.9303
ViT-B [19]	85.61	106.43	85.46	85.44	0.9533
DINOv2-B [22]	85.47	85.79	84.62	84.54	0.9481
IRV2 [5]	54.67	44.37	80.14	80.44	0.9383
SACNN [14]	141.17	96.43	83.47	83.84	0.9661
ECCT [6]	54.28	21.67	72.91	72.82	0.9188
Ours w/o (CDMSR & D_{rel} & D_{abs})	-	-	93.75	93.78	0.9924
Ours w/o (D_{rel} & D_{abs})	-	-	94.59	94.59	0.9903
Ours w/o (D_{abs})	-	-	95.00	94.96	0.9905
Ours (Full)	174.67	73.26	95.42	95.40	0.9948

epochs training only the auxiliary head with a frozen backbone (learning rate 1×10^{-3}), followed by 40 joint fine-tuning epochs (backbone learning rate 1×10^{-5} , head learning rate 1×10^{-4}). We employ Focal loss with class weights [1.0, 1.3, 1.3, 1.0] alongside a cosine annealing scheduler featuring a 5-epoch linear warmup and 0.05 weight decay.

Stage 3: Sequence Classifier Training. We train the ImageNet-21k initialized ViT-Base classifier (patch size 16) for 15 epochs with a frozen ConvNeXt backbone (projection learning rate 2×10^{-4} , classifier learning rate 1×10^{-4}). Finally, we unfreeze the backbone for 25 joint fine-tuning epochs at reduced learning rates (backbone learning rate 1×10^{-5} , classifier learning rate 5×10^{-5}).

**Fig. 2.** Left: Normalized confusion matrices for all methods, adjusted for sample size disparities between single-frame baselines and our sequence-level approach. Right: Macro-averaged AUC curves.

3.3 Evaluation

We evaluate our method against classic architectures (ResNet-101, ConvNeXt V2 Base, ViT-Base, DINOv2 Base) and recent IVCN models (IRV2, SACNN, ECCT). Table 1 demonstrates that our method significantly outperforms these baselines. Utilizing a larger parameter space to ensure robust generalization, our multi-stage framework maintains excellent diagnostic efficiency and accuracy. Prioritizing clinical precision, our approach directly yields sequence-level diagnoses. This circumvents the complex post-aggregation required by single-frame models, fully justifying the slightly lower FPS. Furthermore, Fig. 2 confirms high accuracy on the imbalanced test set.

Ablation studies demonstrate that the CDMSR module and depth embedding consistently improve performance. Specifically, the CDMSR module converges in just 4 epochs and supports large-scale pre-training, highlighting its potential as a universal, plug-and-play component for clinical microscopy analysis.

Regarding the depth embedding, we analyze depth meta-information to provide empirical motivation. Fig. 3 (showing 10 μm -bin histograms, normalized kernel density estimation, and cumulative distribution functions) reveals distinct depth preferences across the four pathogen categories. Quantitatively, a Kruskal-Wallis H-test confirms significant depth distribution differences ($H = 1889.43$, $p < 0.001$), while post-hoc Mann-Whitney U tests reveal significant pairwise differences between all groups (FDR-corrected $p < 0.05$). These results statistically validate pathogen-specific imaging depths. Consequently, compared to raw sequences, embedded depth helps the attention mechanism better capture pathogen features and structural relationships across depths.

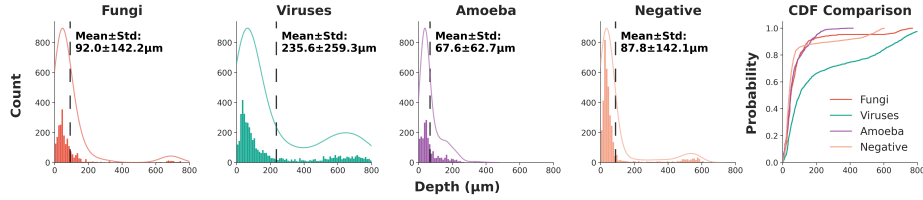


Fig. 3. Depth distributions across four pathogen categories. 10 μm -interval histograms, normalized kernel density curves, and CDFs reveal distinct depth preferences.

4 Discussion and Conclusion

We propose ReDIVE, a Retinex-enhanced Depth-Informed Vertical Embedding framework for infectious keratitis diagnosis via IVCN. By processing sequences instead of single frames, it integrates volumetric data and clinical priors.

To mitigate IVCN degradations, our Customized Differentiable Multi-Scale Retinex (CDMSR) module amplifies pathological features without traditional

variational latency. A hybrid mechanism encodes absolute and relative depth meta-information based on statistical depth preferences, fusing it with spatial textures. A Vision Transformer then aggregates these features to capture long-range correlations for comprehensive sequence classification.

Evaluated on a 15,572-image dataset, ReDIVE outperforms state-of-the-art baselines with 95.42% Accuracy, 95.40% F1-Score, and 0.9948 AUROC, balancing diagnostic efficiency and precision. Ablation studies confirm the necessity of CDMSR and depth embedding.

ReDIVE establishes a paradigm for sequence-level keratitis diagnosis. However, the lack of public IVCN datasets with depth meta-information restricts validation to single-center data. Additionally, our current data covers only four categories. Future work will focus on multi-center validation and expanding the framework to a broader spectrum of infectious keratitis pathologies.

Acknowledgments. This work was supported by the HNXJ Philanthropy Foundation (No. KY25008, KY24014), in part by the Natural Science Foundation of Hunan Province (No. 2024JJ9002), in part by the The Science and Technology Innovation Program of Hunan Province and in part by the National Natural Science Foundation of China (No. 62322210).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Koganti, R., Yadavalli, T., Naqvi, R.A., Shukla, D., Naqvi, A.R.: Pathobiology and treatment of viral keratitis. *Experimental Eye Research* **205**, 108483 (2021). <https://doi.org/10.1016/j.exer.2021.108483>
2. Chiang, J.C.B., Roy, M., Kim, J., Markoulli, M., Krishnan, A.V.: In-vivo corneal confocal microscopy: Imaging analysis, biological insights and future directions. *Commun. Biol.* **6**, 652 (2023). <https://doi.org/10.1038/s42003-023-05005-8>
3. Labbé, A., Khammari, C., Dupas, B., Gabison, E., Brasnu, E., Labetoulle, M., Baudouin, C.: Contribution of In Vivo Confocal Microscopy to the Diagnosis and Management of Infectious Keratitis. *The Ocular Surface* **7**(1), 41–52 (2009). [https://doi.org/10.1016/S1542-0124\(12\)70291-4](https://doi.org/10.1016/S1542-0124(12)70291-4)
4. Assaf, J.F., Ahuja, A.S., Kannan, V., Yazbeck, H., Krivit, J., Redd, T.K.: Applications of Computer Vision for Infectious Keratitis: A Systematic Review. *Ophthalmology Science* **5**, 100861 (2025). <https://doi.org/10.1016/j.xops.2025.100861>
5. Tang, N., Huang, G., Lei, D., Jiang, L., Chen, Q., He, W., Tang, F., Hong, Y., Lv, J., Qin, Y., Lin, Y., Lan, Q., Qin, Y., Lan, R., Pan, X., Li, M., Xu, F., Lu, P.: An artificial intelligence approach to classify pathogenic fungal genera of fungal keratitis using corneal confocal microscopy images. *Int. Ophthalmol.* **43**, 2203–2214 (2023). <https://doi.org/10.1007/s10792-022-02616-8>
6. Qu, J., Qin, X., Xie, Z., Qian, J., Zhang, Y., Sun, X., Sun, Y., Peng, R., Xiao, G., Lin, J., Bian, X., Chen, T., Cheng, Y., Gu, S., Wang, H., Hong, J.: Establishment of an automatic diagnosis system for corneal endothelium diseases using artificial intelligence. *J. Big Data* **11**, 67 (2024). <https://doi.org/10.1186/s40537-024-00913-w>

7. Li, C.-P., Dai, W., Xiao, Y.-P., Qi, M., Zhang, L.-X., Gao, L., Zhang, F.-L., Lai, Y.-K., Liu, C., Lu, J., Chen, F., Chen, D., Shi, S., Li, S., Zeng, Q., Chen, Y.: Two-stage deep neural network for diagnosing fungal keratitis via in vivo confocal microscopy images. *Sci. Rep.* **14**, 18432 (2024). <https://doi.org/10.1038/s41598-024-68768-y>
8. Chopra, R., Mulholland, P.J., Hau, S.C.: In Vivo Confocal Microscopy Morphologic Features and Cyst Density in Acanthamoeba Keratitis. *Am. J. Ophthalmol.* **217**, 38–48 (2020). <https://doi.org/10.1016/j.ajo.2020.03.048>
9. Peng, Y., Chen, Q., Wei, Y., Wang, L., Zhang, Z., Wei, Z., Pang, J., Peng, B., Shi, Q., Wang, Z., Zhang, Y., Chen, K., Xu, X., Liang, Q.: Clinical Characteristics and In Vivo Confocal Microscopic Study in Candida Keratitis. *Trans. Vis. Sci. Tech.* **14**(1), 23 (2025). <https://doi.org/10.1167/tvst.14.1.23>
10. Huang, P., Tepelus, T., Vickers, L.A., Baghdasaryan, E., Huang, J., Irvine, J.A., Hsu, H.Y., Sadda, S., Lee, O.L.: Quantitative Analysis of Depth, Distribution, and Density of Cysts in Acanthamoeba Keratitis Using Confocal Microscopy. *Cornea* **36**(8), 927–932 (2017). <https://doi.org/10.1097/ICO.0000000000001239>
11. Gu, S.-F., Gao, S., Wang, H.-K., He, L.-H., Peng, R.-M., Xiao, G.-G., Hong, J.: Three-dimensional in vivo evaluation of the cornea in patients with unilateral posterior interstitial keratitis. *Front. Med.* **10**, 1180208 (2023). <https://doi.org/10.3389/fmed.2023.1180208>
12. Kempe, M., Genack, A.Z., Rudolph, W., Dorn, P.: Ballistic and diffuse light detection in confocal and heterodyne imaging systems. *J. Opt. Soc. Am. A* **14**, 216 (1997). <https://doi.org/10.1364/JOSAA.14.000216>
13. Liu, Z., Cao, Y., Li, Y., Xiao, X., Qiu, Q., Yang, M., Zhao, Y., Cui, L.: Automatic diagnosis of fungal keratitis using data augmentation and image fusion with deep convolutional neural network. *Comput. Methods Programs Biomed.* **187**, 105019 (2020). <https://doi.org/10.1016/j.cmpb.2019.105019>
14. Liang, S., Zhong, J., Zeng, H., Zhong, P., Li, S., Liu, H., Yuan, J.: A Structure-Aware Convolutional Neural Network for Automatic Diagnosis of Fungal Keratitis with In Vivo Confocal Microscopy Images. *J. Digit. Imaging* **36**, 1624–1632 (2023). <https://doi.org/10.1007/s10278-021-00549-9>
15. Han, R., Tang, C., Xu, M., Lei, Z.: A Retinex-based variational model for noise suppression and nonuniform illumination correction in corneal confocal microscopy images. *Phys. Med. Biol.* **68**, 025025 (2023). <https://doi.org/10.1088/1361-6560/acaef>
16. Jobson, D.J., Rahman, Z., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **6**(7), 965–976 (1997). <https://doi.org/10.1109/83.597272>
17. Howard, A., Sandler, M., Chu, G., et al.: Searching for MobileNetV3. In: *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324 (2019)
18. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. In: *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16133–16144 (2023)
19. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: *International Conference on Learning Representations (ICLR)* (2021)
20. Zhao, Y., Zhang, J., Pereira, E., Zheng, Y., Su, P., Xie, J., Zhao, Y., Shi, Y., Qi, H., Liu, J., Liu, Y.: Automated Tortuosity Analysis of Nerve Fibers in

- Corneal Confocal Microscopy. *IEEE Trans. Med. Imaging.* **39**, 2725–2737 (2020).
<https://doi.org/https://doi.org/10.1109/TMI.2020.2974499>
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)
 22. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research* (2024)