SABPI-Net: A Novel Structure-Aware Network for Accurate and Domain-Invariant Retinopathy of Prematurity Diagnosis

Shaobin Chen[1], Xinyu Zhao[2], Huazhu Fu[3], Tao Tan[1], Jiaju Huang[1], Xiangyu Xiong[1], Zhenquan Wu[2], Behdad Dashtbozorg[4], Baiying Lei[5], Guoming Zhang[2], and Yue Sun[1⋆(✉)]

[1] Faculty of Applied Sciences, Macao Polytechnic University, Macao, China
yuesun@mpu.edu.mo
[2] Shenzhen Eye Hospital, Jinan University, Shenzhen Eye Institute, Shenzhen, China
[3] Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore
[4] Netherlands Cancer Institute
[5] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen, China

Abstract. Delayed treatment of retinopathy of prematurity (ROP) can diminish therapeutic efficacy and may lead to severe, potentially irreversible damage. Automated diagnosis of ROP presents significant challenges, including the detection of subtle early lesions, the variability of clinical phenotypes, and inconsistencies in imaging quality. To address these, which cannot be well addressed by existing general foundation models, we propose structure-aware proxy interaction network (SABPI-Net) within a universal learning framewrok. SABPI-Net incorporates a high-frequency mapping branch, and introduces a proxy interaction attention module to enable effective interaction between its trunk feature encoding branch and the high-frequency mapping branch. This enhances the model's ability to perceive fine retinal detail structures. Domain-agnostic embedding space self-matching, guided by a memory-bank low-frequency component replacement strategy, facilitates domain-invariant learning and ensures consistent model performance across diverse image styles. In this study, classification task for ROP is conducted on the largest clinical color fundus photography dataset to date, achieving an accuracy of 95.32%. Extensive experiments further validate the effectiveness and superiority of SABPI-Net in diagnosing ROP diseases.

Keywords: Retinopathy of prematurity · Structure-aware · Interaction · Ophthalmology.

⋆ Corresponding author

1 Introduction

Retinopathy of prematurity (ROP), a vasoproliferative disease, is the leading cause of visual impairment and irreversible blindness in children worldwide. It primarily affects preterm infants with very low birth weight or those small for gestational age [27]. Currently, ROP diagnosis relies on binocular indirect ophthalmoscopy or wide-field digital fundus image interpretation by experienced ophthalmologists [6]. However, ROP diagnosis remains subjective, with significant diagnostic inconsistency even among specialists [7]. Recent studies have shown some encouraging results in using artificial intelligent (AI) systems for automated ROP screening using color fundus photography (CFP). Various studies, such as those by Huang et al. [12] and Zhang et al. [31], have employed different convolutional neural networks to evaluate their effectiveness in detecting ROP [10, 32]. Additionally, several studies have developed deep learning systems to recognise the presence of plus disease, with some extending their work to categorise pre-plus disease [26, 25, 23, 24].

In recent years, research has increasingly focused on other clinical features of ROP [1]. For instance, studies on ROP severity grading seek to subclassify images that are identified as having ROP features according to the severity of the disease [13, 10, 25]. Xie et al. [28] introduced an adversarial learning-based multi-level dense knowledge distillation approach for ROP detection. Subsequent studies incorporated aggressive posterior ROP into the grading system [31, 28]. Peng et al. [20] proposed a network for five-level staging of ROP that combines a multistream parallel feature extractor.

Despite significant advancements, these studies exhibit two critical limitations. First, current automated ROP classification systems are not comprehensive, as they primarily focus on identifying ROP occurrence, plus disease, and disease severity. There remains a lack of a complete diagnostic framework capable of addressing the complexity of clinical requirements. For example, such a system should encompass routine ROP, aggressive ROP (A-ROP), laser-treated ROP and other infantile retinal diseases. Second, existing models, often trained on small datasets, still have room for improvement in terms of performance and disease coverage. In addition to these limitations, automated ROP diagnosis faces several inherent challenges: A) Early-stage ROP are marked by subtle and often difficult-to-detect characteristic changes [4, 22]. B) The diversity and complexity of disease phenotypes in infants require a model capable of extracting more distinct and comprehensive features. C) The imaging heterogeneity in CFP further complicates effective feature extraction.

To extract more discriminative features, it is essential to accurately perceive the detailed structure of retinal images. According to the retinex theory [14], the detailed structure of the retina is concentrated in the high-frequency domain of the retinal image, whereas the style texture of the image is concentrated in the low-frequency component. In recent years, spectral knowledge has garnered increasing attention in the diagnosis of retinal diseases [16]. Li et al. [15] proposed GFE-Net, which utilizes image frequency information to extract robust structural perception representations, particularly from degraded images. Fur-

thermore, studies [29, 30] have shown that domain shift [11] between the source domain and target domain can be mitigated by integrating or exchanging low-frequency spectrum components. To address domain shift in optic disc and cup segmentation, Chen et al. introduced VPTTA [3], a visual prompt-based test-time adaptation method. VPTTA adapts to each test image by training specific low-frequency prompts that effectively capture image style and texture.

To address the challenges of automatic diagnosis of ROP, we propose a structure-aware bidirectional proxy interaction network (SABPI-Net), which explicitly models the structural information of CFP images and leverages proxy interaction mechanisms to enhance feature representation capabilities, thereby achieving accurate recognition of ROP. Our main contributions include: 1) To address pathological diversity (challenges A and B), we design a high-frequency mapping branch. This branch is integrated with the trunk encoding branch via a proxy interactive attention module, improving the model's ability to capture detailed retinal structures. 2) To mitigate imaging heterogeneity (challenge C), we propose a domain-agnostic embedding space self-matching strategy, ensuring consistent model performance across different imaging styles. 3) We construct a dataset of over 170,000 CFP images, enabling the classification task related to ROP. This dataset surpasses existing studies in both scale and disease coverage in infants. Code is available at https://github.com/SB-Chen/SABPI-Net.

2 Method

The proposed SABPI-Net is shown in Fig.1. The high-frequency mapping (HFM) branch focuses on the high-frequency components of the CFP images, and the trunk feature encoding (TFE) branch of the CFP images interacts with the HFM branch through the proxy interaction attention module (PIAM) to enhance the model's perception of the detailed structures of the retinal. The domain-agnostic embedding space self-matching (DAESSM) learning facilitates the model to maintain consistent performance under various retinal image styles.

2.1 High frequency mapping

Based on the Retinex theory [14] , the high frequency region of the fundus image contains the most important details of the retina. In clinical diagnosis, the characteristic changes of retinal diseases are usually identified by observing the detailed structure of the retina (such as retinal blood vessels, macular area and optic disc area). Therefore, the ability of the model to finely perceive the detailed structure of retinal images is critical to the extraction of discriminative features. To effectively extract the high-frequency components of the retinal region and eliminate interference from irrelevant background areas, the invalid regions are filled with the median value of the image to prevent their impact on subsequent calculations, which is defined as follows:

$$x_p = M(x) \times x + (1 - M(x)) \times (median(x) + \alpha), \tag{1}$$

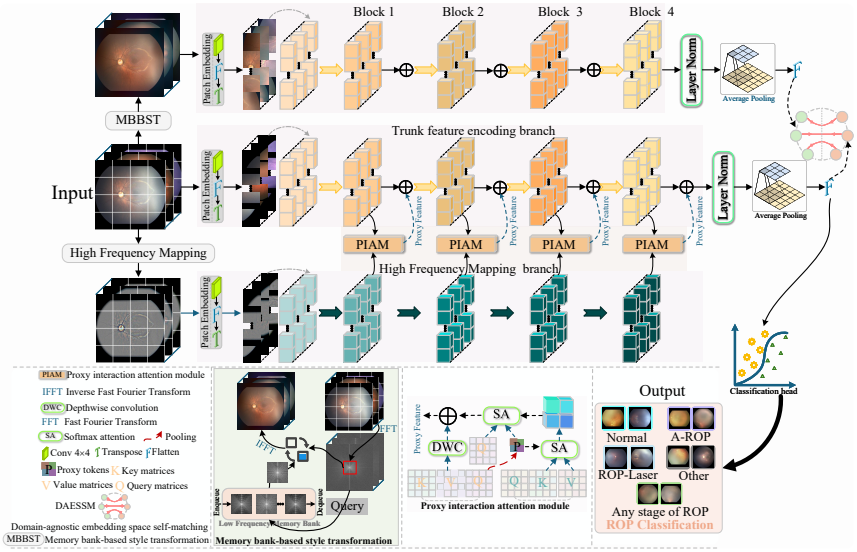

Fig. 1. Overall framework of the proposed SABPI-Net. The architecture incorporates standard hierarchical transformer blocks in Blocks 1 to 4.

where x represents the CFP image, *M(x)* represents the corresponding fundus region mask. On the basis of the median, increasing the offset α can avoid the median filling being too smooth and enhance the contrast of high-frequency features. In this study, α is set to 0.2. Subsequently, a Gaussian convolution kernel is employed for the convolution operation to extract low-frequency components. The high frequency components of the input image are calculated as follows:

$$\mathcal{F}(x_p) = (clamp(x_p - x_p * g(r, \sigma), -1, 1) + 1) \cdot M(x) - 1, \tag{2}$$

where $\mathcal{F}$ is a high-frequency mapping function, *clamp* is a value clipping function, and g is a Gaussian filter with a kernel of (r, σ).

2.2 Proxy interaction attention module

For an input with N tokens that can be represented as $x \in \mathbb{R}^{N \times C}$, in each head, the self-attention query, key, and value matrices can be formulated as follows:

$$Q = xW_Q, \quad K = xW_K, \quad \text{and} \quad V = xW_V, \tag{3}$$

where $W_{Q/K/V} \in \mathbb{R}^{C \times d}$ are projection matrices, $Q, K, V \in \mathbb{R}^{N \times C}$, and C and d are the embedding dimension of modules and each head, respectively. In order to facilitate subsequent analysis, the superscripts m and h are used to denote the variables of the TFE branch and the HFM branch, respectively. A straightforward pooling strategy is applied to the query Q^m, thereby yielding the high-frequency proxy tokens that interact with the HFM branch: $P = \text{pooling}(Q^m)$,

where $P \in \mathbb{R}^{N \times C_P}$, C_P denotes the number of high-frequency proxy tokens. The PIAM uses high-frequency proxy tokens to aggregate high-frequency detail information from the HFM branch and distribute it to individual image tokens in the TFE branch. Specifically, P is taken as the query and the Softmax attention computation is performed between P, the key K^h, and the value V^h to aggregate the high frequency feature F_P. Subsequently, utilising P as the key and F_P as the value, in conjunction with the query Q^m, a secondary Softmax attention computation is performed. This process disseminates the high frequency detail information from the proxy high frequency feature to each query token in the TFE branch. We set the C_P to a small hyperparameter, which can be significantly smaller than the number of query tokens. Based on these designs, PIAM can be formalized as:

$$O_P = \sigma(Q^m P^T + B_2)\,[\sigma(P(K^h)^T) + B_1)\,V^h] + DWC(V^m), \tag{4}$$

where $B_1 \in \mathbb{R}^{C_P \times N}$, $B_2 \in \mathbb{R}^{N \times C_P}$ are the bias terms of high frequency proxy, $\sigma(\cdot)$ represents the softmax function, and DWC is the depthwise convolutional module. The integration of high-frequency proxy bias terms can make better use of the location information, thereby enabling different high-frequency proxy tokens to concentrate on distinct high-frequency regions through the incorporation of spatial information into Softmax attention. In order to alleviate the effect of insufficient feature diversity [8], the DWC is utilised to introduce the V^m, thereby enhancing the feature diversity. Ultimately, the features at each block of the TFE branch are fused with the high-frequency proxy features from the output of PIAM before proceeding to the next block of feature extraction.

2.3 Self-matching in a domain-agnostic embedding space

A memory bank-based low-frequency component replacement strategy is proposed to simulate the low-frequency components of the transformed training images. The memory bank is constructed on the low-frequency components of preceding training samples and updated using a first-in-first-out strategy. The memory bank, designated as M, is responsible for storing the low-frequency components $\{f_{low}^s\}_{s=1}^S$ of S training images. The Fast Fourier Transform and Inverse Fast Fourier Transform operations are denoted by $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ respectively, while the amplitude and phase components are denoted by $\mathcal{F}^A(\cdot)$ and $\mathcal{F}^P(\cdot)$. To attain a transformed image that aligns with the current training image, it is necessary to initially extract the low-frequency component $\mathcal{F}_{low}^A(X_i)$ of the current training image. It is noteworthy that the amplitude component in $\mathcal{F}^A(\cdot)$ has been shifted towards the centre. We use the Euclidean paradigm to compute the similarity between $\mathcal{F}_{low}^A(X_i)$ and each low-frequency component f_{low}^s in the memory bank. The similarity scores are then ranked, and the $\mathcal{F}_{low}^A(X_i)$ of the original image is replaced with the lowest similarity low-frequency component. This process forms a matching alternative stylistic texture image, X_i^*:

$$X_i^* = \mathcal{F}^{-1}([swap(\mathcal{F}^A(X_i), min(sim(\mathcal{F}_{low}^A(X_i), \{f_{low}^s\}_{s=1}^S))), \mathcal{F}^P(X_i)]) \tag{5}$$

In order to achieve stable performance of retinal disease diagnosis across imaging styles, we used cosine similarity to calculate cross-stylistic texture correlations:

$$\mathbb{R}_{X_i,X_i^*} = \frac{Z_i \cdot (Z_i^*)^T}{\|Z_i\|_2 \|Z_i^*\|_2}, \tag{6}$$

where Z_i,$Z_i^* \in \mathbb{R}^{b\times 8C}$ denote the matched X_i and X_i^* output features after image encoder respectively, b denotes the batch size and C is the feature embedding dimension. The objective is to ensure the consistency of the semantic relationship between the image features of X_i and X_i^*. The optimisation objective of self-matching in the domain-agnostic embedding space can be formulated as follows:

$$\mathcal{L}_{match} = 1 - mean(\mathbb{R}_{X_i,X_i^*}) \tag{7}$$

We adopt the cross-entropy loss function as the training objective for the classification task. Therefore, the final training objective function is the weighted sum of all loss functions.

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}) + \lambda\mathcal{L}_{match}, \tag{8}$$

where λ is the weight of the loss $\mathcal{L}_{match}$, which is set to 0.01 in this paper.

3 Experiments

Dataset and Implementation Details. This study collate data from 13,314 patients undergoing ROP screening at Shenzhen Eye Hospital, resulting in a dataset of 174,540 CFPs categorized as: normal (112,770), laser treated ROP (29,442), any stage of ROP (19,513, encompassing Stages 1-5 of ROP), A-ROP (2,363), and other (10,452, representing infant retinal diseases other than ROP). It is divided into training, validation, and test sets at a ratio of approximately 7:2:1 at the patient level. All training is performed on NVIDIA 4090 GPUs. The initial learning rate is set to 0.001, and a cosine annealing schedule is employed. The default optimiser employed is SGD, with a momentum setting of 0.9. The weight decay parameter is set to 1e-4. The number of training epochs is 100. The image data is uniformly resized to 224×224 pixels before being fed into the model for training. During the training process, data enhancement techniques such as random cropping, scaling, horizontal flipping and vertical flipping are sequentially applied. We also validate in the public dataset OIA-ODIR [17].

Comparision with State-of-the-Arts. Table 1 reports the classification performance of the proposed SABPI-Net with seven SOTA models in the in-house ROP and OIA-ODIR datasets. The proposed SABPI-Net outperforms other methods on all metrics for both the in-house and OIA-ODIR datasets. As shown in Fig.2, the majority of models demonstrate high sensitivities within the Normal, Laser-treated ROP and Any Stage of ROP. However, sensitivities within

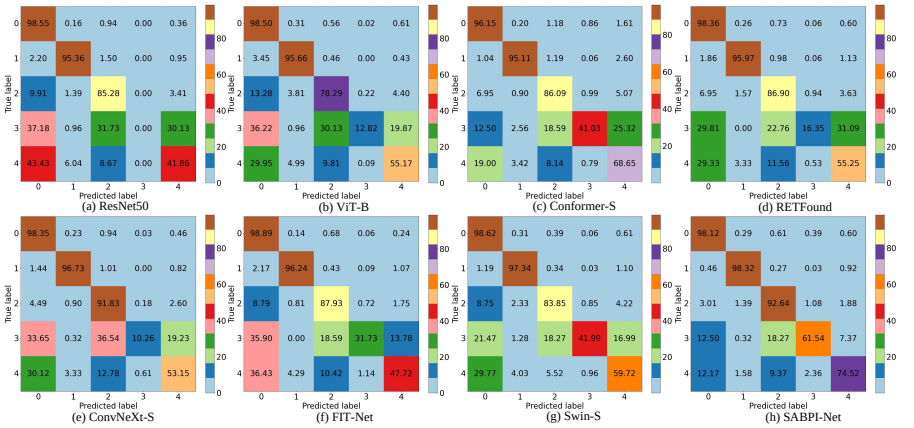

Fig. 2. Comparison of confusion matrix diagram of different methods.

the A-ROP and Other are significantly lower. For instance, ConvNeXt-S demonstrates a sensitivity of 10.26% on A-ROP, while the highest sensitivities observed in the A-ROP and Other within the SOTA methods are 41.99% and 68.65%. It is noteworthy that the SABPI-Net demonstrates significantly higher sensitivity in the A-ROP and Other, with 61.54% and 74.52%, respectively. With the exception of the Normal, SABPI-Net attains the highest level of sensitivity among all other categories.

Table 1. Comparison results on different methods

	In-house dataset							OIA-ODIR[17]			
Method	Acc	AUC	Pre	Sen	F1	Spec	Kappa	Acc	AUC	Sen	Spec
ResNet50[9]	90.96	96.85	67.92	64.21	65.48	96.50	83.58	58.52	81.89	49.62	92.54
ViT-B[5]	91.19	97.34	85.80	68.09	71.48	96.61	84.05	58.76	85.90	49.22	92.63
Conformer-S[21]	91.98	97.42	78.50	77.41	77.84	97.50	86.01	58.10	84.06	48.65	92.30
RETFound[33]	92.30	97.74	81.03	70.57	73.25	97.23	86.23	61.40	86.03	49.10	92.75
ConvNeXt-S[19]	92.81	98.05	84.24	70.06	72.31	97.42	87.16	55.28	81.70	45.81	91.72
FIT-Net[2]	92.59	97.57	86.03	72.50	76.89	97.05	86.60	60.14	84.76	49.37	92.76
Swin-S[18]	93.07	98.15	86.51	76.30	80.17	97.41	87.59	60.62	85.21	50.15	92.87
Our	**95.32**	**99.23**	**87.22**	**85.03**	**86.03**	**98.54**	**91.82**	**63.57**	**88.71**	**50.28**	**93.23**

Ablation Study. Table 2 shows the results of the ablation studies of different modules in SABPI-Net. We first examine the impact of DAESSM learning on the model. It is found that the introduction of DAESSM improves the overall Acc by 1.1%. Next, we introduce the HFM branch and implement the interaction between the HFM branch and the TFE branch through the PIAM. And the introduction of HFM and PIAM can improve the overall Acc by 1.23%. Finally, when DAESSM, HFM and PIAM are combined, it is found that the overall Acc can be further improved, with Acc improvements of 2.18%. To investigate the

Table 2. Ablation study experiments

Module ablation							Low-frequency sizes in DAESSM learning				
HFM	PIAM	DAESSM	Acc	AUC	F1	Spec	Method	Acc	AUC	F1	Spec
			93.14	98.35	78.88	97.32	Baseline	93.14	98.35	78.88	97.32
		✓	94.24	99.11	84.70	97.78	S12	93.45	98.69	77.67	97.76
✓	✓		94.37	98.84	85.15	98.01	S6	93.60	98.64	82.10	97.69
✓	✓	✓	**95.32**	**99.23**	**86.03**	**98.54**	S3	**94.24**	**99.11**	**84.70**	**97.78**

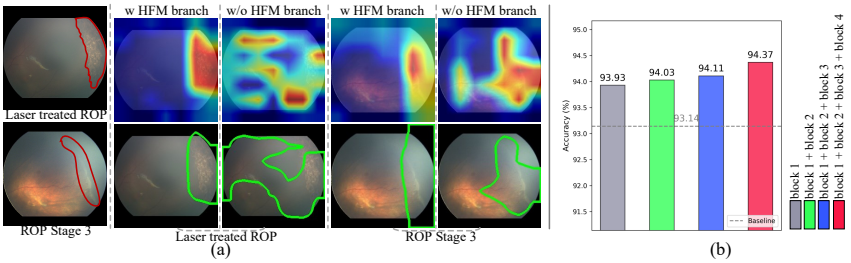

Fig. 3. (a) Comparison of Grad-Cam visualisation results with and without HFM branch. (b) Applying PIAM at different blocks of the Baseline.

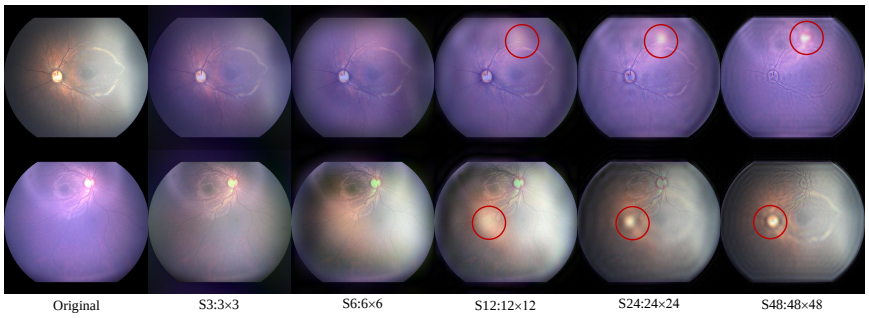

Fig. 4. Transformation of the image style texture after replacing the low-frequency components of different sizes.

impact of PIAM at different blocks, an ablation study is conducted. As demonstrated in Fig. 3(b), the model's Acc undergoes a progressive enhancement as the number of blocks employing PIAM increases. The application of PIAM to facilitate interaction between the two branches at all four blocks results in the model achieving optimal performance. To assess the impact of low-frequency component replacement sizes, we conduct ablation experiments. As shown in Table 2, the model's Acc slightly decreases as the replacement size increases. Furthermore, as illustrated in Fig. 4, when the replacement size is set to S3, the style texture transformation achieves the best effect. However, as the size increases, the image's detailed information gradually becomes blurred, potentially altering

the original structure (e.g., shifting the optic disc area), which is an undesired outcome. Consequently, S3 is selected as the optimal replacement size. To further elucidate the model's decision-making process, we employed Grad-CAM to generate heatmaps highlighting the regions in the input images that contribute most significantly to a specific class prediction. As shown in Fig. 3(a), the addition of the HFM branch enables the model to better perceive detailed structures in the retinal region and focus more precisely on disease lesion locations. Conversely, without the HFM branch, the model's sensitivity to subtle details diminishes, leading to an inability to focus on the actual lesion location for certain diseases.

4 Conclusion

This paper proposes SABPI-Net, a diagnostic model for ROP trained on a large-scale dataset of over 170,000 color fundus photographs, representing the most extensive ROP dataset to date. Comprehensive experiments demonstrate that SABPI-Net achieves exceptional performance in ROP classification tasks, highlighting the significant potential of AI-based models utilizing CFP for early screening and diagnosis of retinal diseases in infants.

Acknowledgments. This work is supported in part by Shenzhen Medical Research Funds under Grant C2301005, in part by Macao Polytechnic University under Grant RP/FCA-08/2024, in part by Science and Technology Development Fund of Macao under Grant 0088/2023/ITP2, in part by STCSM under Grant 22DZ2229005, in part by the H.Fu's Agency for Science, Technology and Research (A*STAR) Central Research Fund ("Robust and Trustworthy AI system for Multi-modality Healthcare").

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

References

1. Chen, S., Zhao, X., , et al.: Multi-risk factors joint prediction model for risk prediction of retinopathy of prematurity. EPMA Journal **15**(2), 261–274 (Jun 2024)
2. Chen, S., et al.: Fit-net: Feature interaction transformer network for pathologic myopia diagnosis. IEEE Transactions on Medical Imaging **42**(9), 2524–2538 (2023)
3. Chen, Z., Pan, Y., Ye, Y., Lu, M., Xia, Y.: Each test image deserves a specific prompt: Continual test-time adaptation for 2d medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11184–11193 (June 2024)
4. Chiang, M.F., Quinn, G.E., et al.: International classification of retinopathy of prematurity, third edition. Ophthalmology **128**(10), e51–e68 (2021)
5. Dosovitskiy, A., Beyer, L., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
6. Fierson, W.M., Chiang, M.F., et al.: Screening examination of premature infants for retinopathy of prematurity. Pediatrics **142**(6), e20183061 (12 2018)

7. Gschließer, A., Stifter, E., et al.: Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. American Journal of Ophthalmology **160**(3), 553–560.e3 (2015)
8. Han, D., Pan, X., Han, Y., et al.: Flatten transformer: Vision transformer using focused linear attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5961–5971 (October 2023)
9. He, K., Zhang, X., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
10. Hu, J., Chen, Y., et al.: Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Transactions on Medical Imaging **38**(1), 269–279 (2019)
11. Huang, J., Chen, S., et al.: All-in-one multi-organ segmentation in 3d ct images via self-supervised and cross-dataset learning. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2025)
12. Huang, Y.P., Vadloori, S., et al.: Deep learning models for automated diagnosis of retinopathy of prematurity in preterm infants. Electronics **9**(9) (2020). https://doi.org/10.3390/electronics9091444, https://www.mdpi.com/2079-9292/9/9/1444
13. Huang, Y.P., Wu, W.C., et al.: Automated detection of early-stage rop using a deep convolutional neural network. British Journal of Ophthalmology **105**(8), 1099–1103 (2021). https://doi.org/10.1136/bjophthalmol-2020-316526
14. Land, E.H.: The retinex theory of color vision. Scientific American **237**(6), 108–129 (1977), http://www.jstor.org/stable/24953876
15. Li, H., Liu, H., et al.: A generic fundus image enhancement network boosted by frequency self-supervised representation learning. Medical Image Analysis **90**, 102945 (2023). https://doi.org/https://doi.org/10.1016/j.media.2023.102945
16. Li, H., Liu, J., et al.: Frequency-mixed single-source domain generalization for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 127–136. Springer Nature Switzerland, Cham (2023)
17. Li, N., Li, T., et al.: A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In: Wolf, F., Gao, W. (eds.) Benchmarking, Measuring, and Optimizing. pp. 177–193. Springer International Publishing, Cham (2021)
18. Liu, Z., Lin, Y., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10012–10022 (October 2021)
19. Liu, Z., Mao, H., et al.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11976–11986 (June 2022)
20. Peng, Y., Zhu, W., et al.: Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. IEEE Transactions on Medical Imaging **40**(7), 1750–1762 (2021). https://doi.org/10.1109/TMI.2021.3065753
21. Peng, Z., Huang, W., Gu, S., et al.: Conformer: Local features coupling global representations for visual recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 367–376 (October 2021)
22. of Retinopathy of Prematurity*, C.: The International Classification of Retinopathy of Prematurity Revisited. Archives of Ophthalmology **123**(7), 991–999 (07 2005). https://doi.org/10.1001/archopht.123.7.991
23. Ramachandran, S., Niyas, P., et al.: A deep learning framework for the detection of plus disease in retinal fundus images of preterm

infants. Biocybernetics and Biomedical Engineering **41**(2), 362–375 (2021). https://doi.org/https://doi.org/10.1016/j.bbe.2021.02.005, https://www.sciencedirect.com/science/article/pii/S0208521621000139
24. Ramachandran, S., Strisciuglio, N., et al.: U-cosfire filters for vessel tortuosity quantification with application to automated diagnosis of retinopathy of prematurity. Neural Computing and Applications **32**(16), 12453–12468 (Aug 2020). https://doi.org/10.1007/s00521-019-04697-6
25. Tong, Y., Lu, W., et al.: Automated identification of retinopathy of prematurity by image-based deep learning. Eye and Vision **7**(1), 40 (Aug 2020). https://doi.org/10.1186/s40662-020-00206-2
26. Wang, J., Ji, J., et al.: Automated Explainable Multidimensional Deep Learning Platform of Retinal Images for Retinopathy of Prematurity Screening. JAMA Network Open **4**(5), e218758–e218758 (05 2021). https://doi.org/10.1001/jamanetworkopen.2021.8758
27. Wu, Q., Hu, Y., et al.: Development and Validation of a Deep Learning Model to Predict the Occurrence and Severity of Retinopathy of Prematurity. JAMA Network Open **5**(6), e2217447–e2217447 (06 2022). https://doi.org/10.1001/jamanetworkopen.2022.17447
28. Xie, H., Lei, B., et al.: Adversarial learning-based multi-level dense-transmission knowledge distillation for ap-rop detection. Medical Image Analysis **84**, 102725 (2023). https://doi.org/https://doi.org/10.1016/j.media.2022.102725
29. Xu, Q., Zhang, R., et al.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14383–14392 (June 2021)
30. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
31. Zhang, R., Zhao, J., et al.: Automatic diagnosis for aggressive posterior retinopathy of prematurity via deep attentive convolutional neural network. Expert Systems with Applications **187**, 115843 (2022). https://doi.org/https://doi.org/10.1016/j.eswa.2021.115843
32. Zhao, X., Chen, S., et al.: A fundus image dataset for intelligent retinopathy of prematurity system. Scientific Data **11**(1), 543 (May 2024)
33. Zhou, Y., Chia, M.A., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (Oct 2023)