

Generalized Robust Fundus Photography-based Vision Loss Estimation for High Myopia

Zipei Yan¹, Zhile Liang², Zhengji Liu³, Shuai Wang⁴, Rachel Ka-Man Chun³, Jizhou Li¹, Chea-su Kee^{3,4}, and Dong Liang^{3,4,✉}

¹ School of Data Science, City University of Hong Kong, Kowloon, Hong Kong

² School of Engineering and Technology, The University of Newcastle, Australia

³ School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong
dong1.liang@connect.polyu.hk

⁴ Centre for Eye and Vision Research, 17W Hong Kong Science Park, Hong Kong

⁵ Department of Biomedical Engineering, Tsinghua University, Beijing, China

Abstract. High myopia significantly increases the risk of irreversible vision loss. Traditional perimetry-based visual field (VF) assessment provides systematic quantification of visual loss but it is subjective and time-consuming. Consequently, machine learning models utilizing fundus photographs to estimate VF have emerged as promising alternatives. However, due to the high variability and the limited availability of VF data, existing VF estimation models fail to generalize well, particularly when facing out-of-distribution data across diverse centers and populations. To tackle this challenge, we propose a novel, parameter-efficient framework to enhance the generalized robustness of VF estimation on both in- and out-of-distribution data. Specifically, we design a Refinement-by-Denoising (RED) module for feature refinement and adaptation from pretrained vision models, aiming to learn high-entropy feature representations and to mitigate the domain gap effectively and efficiently. Through independent validation on two distinct real-world datasets from separate centers, our method significantly outperforms existing approaches in RMSE, MAE and correlation coefficient for both internal and external validation. Our proposed framework benefits both in- and out-of-distribution VF estimation, offering significant clinical implications and potential utility in real-world ophthalmic practices.

Keywords: Vision loss estimation · Visual field · Fundus photograph · Feature learning · Denoising · Generalization.

1 Introduction

High myopia (HM) is a significant risk factor for irreversible vision impairment, primarily due to its association with several ocular conditions, such as myopic maculopathy and retinal detachment [19]. The assessment of visual field (VF) sensitivity is crucial for evaluating the risk and extent of vision loss, as it provides a systematic quantification of visual function [21]. However, traditional

VF assessment using perimetry is notably subjective and time-consuming, as it highly requires patients’ compliance throughout the test [14].

Unlike VF assessment, fundus photography as an imaging technique for retinal morphology, not only offers an objective measurement of the retinal structure, but also emerged as a valuable alternative for evaluating the retinal function, based on the theory of “structure-function relationship” [27]. Recent advances deep learning-based methods [13,7,29] have shown potential in leveraging fundus photos for accurate vision loss estimation, even point-wise VF estimation [29]. These methods provide more objective and efficient alternatives to assess the risk and extent of vision loss.

Despite the potential of deep learning-based VF estimations, their practical implementation in clinical settings is compromised by two primary challenges associated with VF data. Firstly, the inherent high variability within VF data—largely caused by its subjective assessment [26]—and the variation in imaging devices and protocols across diverse centers and populations, will introduce unforeseen shifts in data distribution. This phenomenon, known as out-of-distribution (OOD) data, hinders the models’ ability to generalize across datasets, presenting a significant challenge to the robustness and reliability of VF estimation. Secondly, the limited available VF data—mainly due to the cost and difficulty of acquisition [9]—further poses a significant barrier to the development of a robust deep learning model for VF estimation.

To tackle these challenges, our study proposes a novel parameter-efficient framework to enhance the generalized robustness of VF estimation on both in-distribution and OOD data. Specifically, we design a Refinement-by-Denoising (RED) module for feature refinement and adaptation from pretrained vision models, aiming to mitigate noises originating from the domain gap and learn high-entropy feature representations for VF estimation.

In brief, our method utilizes pretrained vision models for initial feature extraction from fundus photos. While these models are originally trained on natural images, their application to fundus photos introduces a domain gap and leads to corrupted raw features, mainly due to the differences between the natural image and fundus image domains. Therefore, to mitigate this domain gap, RED refines the raw features by unsupervised denoising, which not only removes the noise but also encourages high-entropy representations, facilitating the downstream VF estimation. As a result, our method significantly outperforms existing approaches in the validation of two distinct real-world datasets.

Our contributions are summarized as follows:

- We propose a novel parameter-efficiency framework, RED, for robust VF estimation from fundus photos. RED is designed for feature refinement and adaptation from pretrained vision models, and could learn high-entropy feature representations and mitigate the domain gap effectively.
- Our proposed RED outperforms existing models in RMSE, MAE and correlation coefficient for both internal and external validation on two distinct real-world datasets, offering significant clinical implications and potential utility in future ophthalmic practices.

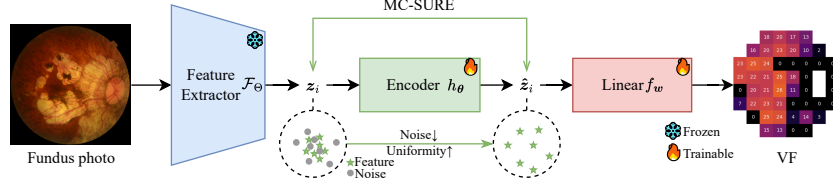


Fig. 1: Overview of the proposed method.

- This is the first work to systematically assess the robustness of VF estimation models across datasets from different centers and populations.

2 Problem Formulation

Given a training set $\mathcal{D} = \{(\mathbf{X}^i, \mathbf{y}^i)\}_{i=1}^N$, where $\mathbf{X}^i \in \mathbb{R}^{H \times W \times C}$ denotes the fundus photo, $\mathbf{y}^i \in \mathbb{R}^M$ denotes its corresponded vectorized VF. The objective is to train a model on the given $\mathcal{D}$, which learns a mapping from the fundus photo to the target VF. The challenges mainly come from the limited training set. The training set is in a limited size and the fundus photo is in a high-dimensional space, therefore learning the model is challenging. Besides, conventional regression fails to predict VF accurately [29], due to its inability to learn high entropy feature representations [31]. Furthermore, we aim to learn a robust model that is generalized to both in- and out-of-distribution data.

3 Proposed Method

3.1 Overview

As illustrated in Fig. 1, we present an overview of the proposed method. Specifically, we address this problem from a feature refinement perspective. Our methods consist of three components: feature extraction, feature denoising and regression. For feature extraction, we utilize pretrained vision models to extract the feature vectors from the fundus photo, thereby significantly reducing its dimensionality. However, the pretrained vision models are generally trained in the natural image domain; therefore, there is a domain gap between the natural image and the fundus photo domain. To address this problem, we propose a refinement by denoising (RED) module, which removes the noise in features and benefits the following regression process. Thereafter, the final regression learns to predict VF based on the refined features.

3.2 Feature Refinement by Denoising (RED)

Given a pretrained vision model $\mathcal{F}_\Theta(\cdot) : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}^K$ where Θ denotes its parameters, for a fundus photo $\mathbf{X}$, its feature vector can be extracted as follows:

$$\mathbf{z} = \mathcal{F}_\Theta(\mathbf{X}), \quad (1)$$

Table 1: Main results. $\dagger$ denotes fine-tuning the feature extractor.

Method	Trainable Params (MB)	Validation set			Test set		
		RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)
Classification	1.04	4.96	3.81	0.32	3.17	2.39	0.19
EMD [12]	1.04	4.94	3.76	0.45	3.19	2.38	0.32
SOFT [1]	1.04	4.95	3.80	0.33	3.16	2.39	0.20
CORAL [2]	1.02	4.78	3.68	0.41	3.14	2.43	0.30
OLL [3]	1.04	4.57	3.45	0.55	3.08	2.37	0.40
VF-HM [29]	11.02 $\dagger$	4.69	3.61	0.49	3.32	2.58	0.22
Regression	0.03	5.09	4.00	0.44	3.99	3.44	-0.08
Regression + OE [31]	0.53	4.36	3.25	0.60	3.08	2.39	0.31
Regression + RED	0.53	4.21	3.13	0.63	2.92	2.19	0.53

where $\mathbf{z} \in \mathbb{R}^K$ denotes the extracted feature vector.

However, the pretrained vision models are generally trained in the natural image domain, and our input modality is the fundus photo, which is captured by the specialized digital camera. Although it is in RGB mode, there is a domain gap between the natural image domain and fundus photo domain. Following [15], we model the domain gap by assuming that each feature vector $\mathbf{z}$ is corrupted by additive white Gaussian noise (AWGN), which is formulated as follows:

$$\mathbf{z} = \mathbf{z}^* + \mathbf{n}, \quad (2)$$

where $\mathbf{z}^* \in \mathbb{R}^K$ denotes the underlying cleaned signal and $\mathbf{n} \in \mathbb{R}^K$ denotes an *i.i.d.* Gaussian noise, i.e., $n_k \sim \mathcal{N}(0, \sigma^2)$.

Thereafter, we propose a Refinement by Denoising (RED) module for mitigating this gap by denoising. Specifically, we parameterize RED by an encoder $h_{\theta}(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}^K$, which denoises the features by $\hat{\mathbf{z}} = h_{\theta}(\mathbf{z})$. However, the ground truth $\mathbf{z}^*$ is not available in practice, therefore optimizing $h_{\theta}(\cdot)$ with the mean squared error (MSE) is impossible. And the MSE is defined as follows:

$$\text{MSE}(h_{\theta}(\mathbf{z})) = \frac{1}{K} \|\mathbf{z}^* - \hat{\mathbf{z}}\|^2 = \frac{1}{K} \|\mathbf{z}^* - h_{\theta}(\mathbf{z})\|^2, \quad (3)$$

where $\|\cdot\|$ denotes ℓ^2 norm.

Instead, following Stein's unbiased risk estimator (SURE) [25], we can derive the unbiased estimation of the above MSE in Eq.(3), which only requires noisy feature $\mathbf{z}$. And the SURE is formulated as follows:

$$\text{SURE}(h_{\theta}(\mathbf{z})) = \frac{1}{K} \|\mathbf{z} - h_{\theta}(\mathbf{z})\|^2 - \sigma^2 + \underbrace{\frac{2\sigma^2}{K} \sum_{k=1}^K \frac{\partial h_{\theta}(\mathbf{z})_k}{\partial z_k}}_{\text{divergence}}. \quad (4)$$

Theorem 1. ([25]) *The random variable $\text{SURE}(h_{\theta}(\mathbf{z}))$ is an unbiased estimator of $\text{MSE}(h_{\theta}(\mathbf{z}))$, that is, $\mathbb{E}_{\mathbf{n}}\{\text{MSE}(h_{\theta}(\mathbf{z}))\} = \mathbb{E}_{\mathbf{n}}\{\text{SURE}(h_{\theta}(\mathbf{z}))\}$,*

Table 2: Results of abnormal cases on Test set.

Category	GHT: Outside Normal Limits			PSD: $P \leq 0.5\%$		
Method	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)
Classification	4.32	3.83	0.28	5.53	3.06	0.13
EMD [12]	4.55	3.57	0.49	5.57	3.05	0.32
SOFT [1]	4.33	3.40	0.27	5.51	3.03	0.15
CORAL [2]	4.11	3.25	0.45	5.45	3.01	0.23
OLL [3]	4.10	3.21	0.58	5.40	2.94	0.34
VF-HM [29]	4.25	3.38	0.17	5.54	3.18	0.12
Regression	4.38	3.67	-0.23	5.52	3.70	0.38
Regression + OE [31]	4.24	3.31	0.58	5.40	2.88	0.31
Regression + RED	3.95	3.05	0.62	5.15	2.72	0.47

where $\mathbb{E}_{\mathbf{n}}\{\cdot\}$ represents the expectation with respect to $\mathbf{n}$.

For fast approximating the last divergence term in Eq.(4), we follow the MC-SURE [22], which utilizes Monte Carlo method [18] in approximation.

Theorem 2. ([22]) Let $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \in \mathbb{R}^K$ be a zero-mean i.i.d. normal vector, then,

$$\sum_{k=1}^K \frac{\partial h_{\theta}(\mathbf{z})_k}{\partial \mathbf{z}_k} = \lim_{\epsilon \rightarrow 0} \mathbb{E}_{\mathbf{b}} \left\{ \mathbf{b}^T \left(\frac{h_{\theta}(\mathbf{z} + \epsilon \mathbf{b}) - h_{\theta}(\mathbf{z})}{\epsilon} \right) \right\}, \quad (5)$$

where ϵ denotes a fixed small positive scalar, and $(\cdot)^T$ denotes the transpose operator.

Based on the above Theorem 2, the last divergence term in Eq.(4) can be approximated as follows [22,24]:

$$\frac{1}{K} \sum_{k=1}^K \frac{\partial h_{\theta}(\mathbf{z})_k}{\partial \mathbf{z}_k} \approx \frac{1}{\epsilon K} \mathbf{b}^T (h_{\theta}(\mathbf{z} + \epsilon \mathbf{b}) - h_{\theta}(\mathbf{z})). \quad (6)$$

Therefore, the training objective function of $h_{\theta}(\cdot)$ is formulated as follows:

$$\text{MC-SURE}(h_{\theta}(\mathbf{z})) = \frac{1}{K} \|\mathbf{z} - h_{\theta}(\mathbf{z})\|^2 - \sigma^2 + \frac{2\sigma^2}{\epsilon K} \mathbf{b}^T (h_{\theta}(\mathbf{z} + \epsilon \mathbf{b}) - h_{\theta}(\mathbf{z})). \quad (7)$$

Finally, we utilize a regression module $f_{\mathbf{w}}(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}^M$ for predicting VF based on the denoised features, that is, $\hat{\mathbf{y}} = f_{\mathbf{w}}(\hat{\mathbf{z}})$. And the overall training objective is summarized as follows:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \|\hat{\mathbf{y}}^i - \mathbf{y}^i\|^2 + \lambda \cdot \text{MC-SURE}(h_{\theta}(\mathbf{z}^i)) \right), \quad (8)$$

where $\lambda \in \mathbb{R}_+$ is a trade-off hyper-parameter.

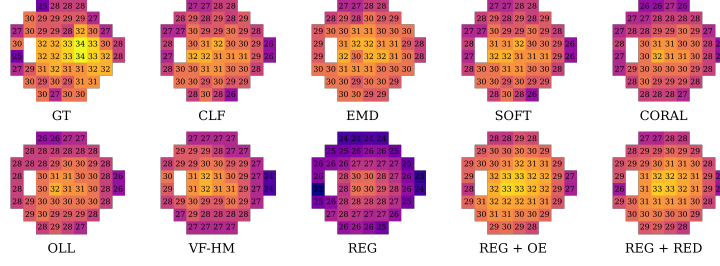


Fig. 2: Visualization of predictions of a case in the Test set.

3.3 Analysis of the role of noise

To understand the role of noise in features, we conduct analyses on both noisy features $\mathbf{z}$ and denoised features $\hat{\mathbf{z}}$, respectively. For simplicity, we consider a linear regression model $f_{\mathbf{w}}(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}$ for predicting a single VF point based on a given feature; and without loss of generality, we ignore its bias term, and it is defined as:

$$f_{\mathbf{w}}(\mathbf{z}) = \mathbf{w}^T \mathbf{z}, \quad (9)$$

where $\mathbf{w} \in \mathbb{R}^K$ denotes its weight parameter.

Recall the noise formulation in Eq.(2) for features $\mathbf{z}$, and consider the training objective of $f_{\mathbf{w}}(\cdot)$ is a squared error, then it can be expressed as follows:

$$\begin{aligned} (f_{\mathbf{w}}(\mathbf{z}) - y)^2 &= (\mathbf{w}^T (\mathbf{z}^* + \mathbf{n}) - y)^2 = ((\mathbf{w}^T \mathbf{z}^* - y) + \mathbf{w}^T \mathbf{n})^2 \\ &= (\mathbf{w}^T \mathbf{z}^* - y)^2 + \|\mathbf{w}^T \mathbf{n}\|^2 \\ &= (\mathbf{w}^T \mathbf{z}^* - y)^2 + \underbrace{K\sigma^2 \|\mathbf{w}\|^2}_{\text{weight decay}}, \end{aligned} \quad (10)$$

therefore, the noise in noisy features $\mathbf{z}$ can be interpreted as a regularization [4], i.e., a constant weight decay.

Furthermore, we investigate the effect of denoised feature $\hat{\mathbf{z}}$. Inspired by [23,5,6,30], we decompose it as follows:

$$\hat{\mathbf{z}} = \mathbf{z}^* + \mathbf{e}, \quad (11)$$

where $\mathbf{e}$ denotes the residual error, which is composed of the lost signals of $\mathbf{z}^*$ and the remaining noise from $\mathbf{n}$.

Similarly, the training objective can be expanded as follows:

$$(f_{\mathbf{w}}(\hat{\mathbf{z}}) - y)^2 = (\mathbf{w}^T \mathbf{z}^* - y)^2 + \underbrace{\cos^2(\mathbf{e}, \mathbf{w}) \|\mathbf{e}\|^2 \|\mathbf{w}\|^2}_{\text{weight decay}}, \quad (12)$$

where the error in denoised features $\hat{\mathbf{z}}$ also performs as a regularization, yet, an adaptive weight decay depending on the residual error $\mathbf{e}$, as well as the quality of the denoising.

Table 3: Alternative denoising kernels. Table 4: Alternative feature extractors.

Method	Validation set			Test set		
	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)
Regression	5.09	4.00	0.44	3.99	3.44	-0.08
+ Mean (k=2)	5.01	3.93	0.14	3.65	3.05	-0.10
+ Mean (k=3)	5.00	3.93	0.00	3.64	3.03	-0.11
+ Mean (k=4)	5.02	3.94	-0.05	3.63	3.03	-0.11
+ Median (k=3)	4.98	3.89	0.15	3.65	3.05	-0.01
+ Median (k=5)	5.03	3.94	-0.04	3.64	3.03	-0.12
+ Median (k=7)	5.07	3.99	-0.20	3.69	3.09	-0.14
+ RED	4.21	3.13	0.63	2.92	2.19	0.53

Method	Validation set			Test set		
	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)	RMSE($\downarrow$)	MAE($\downarrow$)	PCC($\uparrow$)
ResNet-18 [11]	5.09	4.00	0.44	3.99	3.44	-0.08
+ RED	4.21	3.13	0.63	2.92	2.19	0.53
ConvNeXt-T [17]	4.91	3.82	0.53	3.90	3.31	-0.08
+ RED	4.39	3.26	0.57	3.01	2.27	0.49
ViT-B-16 [10]	4.93	3.84	0.31	4.10	3.58	0.06
+ RED	4.42	3.30	0.59	3.05	2.38	0.35
Swin-T [16]	4.85	3.75	0.46	4.18	3.64	-0.12
+ RED	4.12	3.03	0.63	3.56	2.99	0.38

4 Experiments

4.1 Experimental Data

The experimental data (including training, validation, and test set) comes from the HM population. Specifically, the training and validation sets are from the same clinic site, while the test set is from another clinic site, there are no overlapping patients among them. The training set includes 254 eyes, the validation set includes 45 eyes, and the test set consists of 92 eyes. Besides, the fundus photo is captured in RGB colorful mode, the VF is measured in the 24-2 mode with 52 effective points.

4.2 Experimental Setup

Data pre-processing. We choose the left eye pattern as our base [29], while we horizontally flip these data that are not in the left eye pattern.

Baselines. We mainly compare our method to classification and regression baselines, which utilize extracted features. Specifically, classification baselines including conventional and ordinal classifications, including EMD [12], SOFT [1], CORAL [2] and OLL [3]. Besides, we also compare our method to VF-HM [29], which is based on CORAL and auxiliary learning. Additionally, we consider the regression regularized by ordinal entropy (OE) [31].

Evaluation methods. Following [20,32,28,7,8,29], we utilize two metrics: RMSE and MAE for quantitative evaluation. In addition, we utilize the Pearson correlation coefficient (PCC) to measure the linear relationship between the predicted and ground-truth VF. For qualitative evaluation, we visualize some representative predictions from the test set.

Implementation details. We follow the official implementations for all baselines. And for a fair comparison, we utilize the ImageNet-1K pretrained ResNet-18 [11] as the feature extractor. For the encoder $h_\theta(\cdot)$ in RED and OE, we utilize a ReLU-based MLP with only one hidden layer⁶.

4.3 Experimental Results

Main results. Table 1 reports the results of different methods on both validation and test sets. In general, RED outperforms others on both validation and

⁶ Our code is available at https://github.com/yanziwei/VF_RED

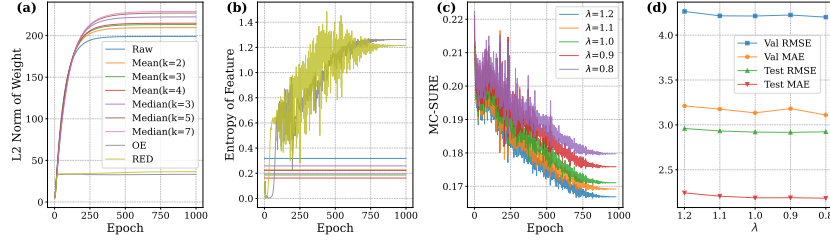


Fig. 3: (a) L2 Norm of the weight of the regression layer (b) Entropy of feature space (c) MC-SURE loss during training, and (d) RMSE and MAE metrics corresponding to different λ .

test sets. Specifically, for classifications, we observe that OLL achieves better performance and CORAL is the runner-up. Besides, VF-HM outperforms OLL on the validation set but is worse on the test set, indicating fine-tuning overfits and poorly generalizes to unseen OOD test data. For regressions, we observe that RED can improve the generalization on both in- and out-of-distribution data, according to the significant improvement from RMSE, MAE and PCC on both validation and test sets; meanwhile, OE mainly improves the generalization of in-distribution data.

Results on abnormal cases. Table 2 reports the results of abnormal cases on the Test set. Abnormal cases are identified based on Glaucoma Hemifield Test (GHT, outside normal limited) and Pattern Standard Deviation (PSD, $P \leq 0.5\%$). Similar to the main results, RED achieves better performance than others in abnormal cases, suggesting the RED’s ability to identify and predict abnormal cases more accurately.

Visualization of predictions. Fig. 2 visualizes predictions from different models for a representative case on test data. Specifically, conventional regression fails to predict vision loss, while RED predicts local vision loss more precisely.

4.4 Ablation Study

Apart from the analysis in Sec 3.3, we conduct experiments to analyze the effectiveness of the proposed RED.

Effectiveness of RED. We compared RED to different denoising kernels including mean and median kernels. In Table 3, we observe that RED is more effective than others. Besides, in Fig. 3(a) and (b) verify our method works as a weight decay regularized to smaller weights, and encourages to learn high entropy feature representations.

Impact of Hyper-parameter. As shown in Fig. 3(c), we explore the impact of different λ for MC-SURE loss; besides, according to the result in Fig. 3(d), we observe that $\lambda = 1.0$ is a good trade-off parameter.

Alternative feature extractors. We also explore alternative feature extractors, such as ConvNeXT [17], ViT [10], and Swin [16]. In Table 4, we observe that RED has broad applicability to different feature extractors.

5 Conclusion

In this work, we propose a parameter-efficiency framework: RED for robust VF estimation from fundus photos and validate it on two distinct real-world datasets from separate centers. RED significantly outperforms existing methods in both internal and external validation, suggesting a more robust and reliable VF estimation approach. Our work benefits not only in-distribution but also out-of-distribution data, highlighting significant clinical implications and potential utility in future ophthalmic practices. A notable limitation of our work is the demographic homogeneity of the datasets used, which consisted exclusively of patients with non-glaucomatous HM from the same ethnicity, although they were collected from diverse sites using diverse imaging devices and protocols. This may affect the applicability of our model to a broader population with diverse demographics (e.g., ethnicity, ocular diseases). Further study is required to evaluate the influence of demographic background on VF estimation.

Acknowledgements This work is supported by Centre for Eye and Vision Research (CEVR), InnoHK CEVR Project 1.5, 17W Hong Kong Science Park, Hong Kong; Research Centre for SHARP Vision (RCSV) and Research Institute for Artificial Intelligence of Things (RIAIoT), The Hong Kong Polytechnic University; and PolyU Grants (P0035514).

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

References

1. Bertinetto, L., Mueller, R., Tertikas, K., Samangoee, S., Lord, N.A.: Making better mistakes: Leveraging class hierarchies with deep networks. In: CVPR. pp. 12506–12515 (2020)
2. Cao, W., Mirjalili, V., Raschka, S.: Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters* **140**, 325–331 (2020)
3. Castagnos, F., Mihelich, M., Dognin, C.: A simple log-based loss function for ordinal text classification. In: COLING. pp. 4604–4609 (2022)
4. Chapelle, O., Weston, J., Bottou, L., Vapnik, V.: Vicinal risk minimization. In: NeurIPS. pp. 416–422 (2000)
5. Chen, C., Xiong, Z., Tian, X., Wu, F.: Deep boosting for image denoising. In: ECCV. pp. 3–19 (2018)
6. Chen, C., Xiong, Z., Tian, X., Zha, Z., Wu, F.: Real-world image denoising with deep boosting. *IEEE Transactions on Pattern Analysis and Machine* **42**(12), 3071–3087 (2020)
7. Christopher, M., Bowd, C., Belghith, A., Goldbaum, M.H., Weinreb, R.N., Fazio, M.A., Girkin, C.A., Liebmann, J.M., Zangwill, L.M.: Deep learning approaches predict glaucomatous visual field damage from oct optic nerve head en face images and retinal nerve fiber layer thickness maps. *Ophthalmology* **127**(3), 346–356 (2020)

8. Datta, S., Mariottoni, E.B., Dov, D., Jammal, A.A., Carin, L., Medeiros, F.A.: Retinervenet: Using recursive deep learning to estimate pointwise 24-2 visual field data based on retinal structure. *Scientific Reports* **11**(1), 1–10 (2021)
9. Delgado, M.F., Nguyen, N.T., Cox, T.A., Singh, K., Lee, D.A., Dueker, D.K., Fechtner, R.D., Juzych, M.S., Lin, S.C., Netland, P.A., Pastor, S.A., Schuman, J.S., Samples, J.R.: Automated perimetry: A report by the american academy of ophthalmology. *Ophthalmology* **109**(12), 2362–2374 (2002)
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: *ICLR* (2021)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. pp. 770–778 (2016)
12. Hou, L., Yu, C.P., Samarasinghe, D.: Squared earth mover’s distance-based loss for training deep neural networks. *arXiv preprint arXiv:1611.05916* (2016)
13. Lee, J., Kim, Y.W., Ha, A., Kim, Y.K., Park, K.H., Choi, H.J., Jeoung, J.W.: Estimating visual field loss from monoscopic optic disc photography using deep learning model. *Scientific reports* **10**(1), 1–10 (2020)
14. Lewis, R.A., Johnson, C.A., Keltner, J.L., Labermeier, P.K.: Variability of quantitative automated perimetry in normal observers. *Ophthalmology* **93**(7), 878–881 (1986)
15. Li, X., Dai, Y., Ge, Y., Liu, J., Shan, Y., Duan, L.Y.: Uncertainty modeling for out-of-distribution generalization. In: *ICLR* (2022)
16. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: *ICCV*. pp. 10012–10022 (2021)
17. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: *CVPR*. pp. 11976–11986 (June 2022)
18. Metropolis, N., Ulam, S.: The monte carlo method. *Journal of the American Statistical Association* **44**(247), 335–341 (1949)
19. Ohno-Matsui, K., Wu, P.C., Yamashiro, K., Vutipongsatorn, K., Fang, Y., Cheung, C.M.G., Lai, T.Y.Y., Ikuno, Y., Cohen, S.Y., Gaudric, A., Jonas, J.B.: Imi pathologic myopia. *Investigative Ophthalmology & Visual Science* **62**(5), 5–5 (05 2021)
20. Park, K., Kim, J., Lee, J.: A deep learning approach to predict visual field using optical coherence tomography. *PLOS ONE* **15**(7), 1–19 (2020)
21. Phu, J., Khuu, S.K., Yapp, M., Assaad, N., Hennessy, M.P., Kalloniatis, M.: The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clinical & Experimental Optometry* **100**(4), 313–332 (Jul 2017)
22. Ramani, S., Blu, T., Unser, M.: Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Transactions on Image Processing* **17**(9), 1540–1554 (2008)
23. Romano, Y., Elad, M.: Boosting of image denoising algorithms. *SIAM Journal on Imaging Sciences* **8**(2), 1187–1219 (2015)
24. Soltanayev, S., Chun, S.Y.: Training deep learning based denoisers without ground truth data. In: *NeurIPS*. pp. 3261–3271 (2018)
25. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* pp. 1135–1151 (1981)
26. Wild, J.M., Pacey, I.E., Hancock, S.A., Cunliffe, I.A.: Between-algorithm, between-individual differences in normal perimetric sensitivity: full threshold, FASTPAC,

- and SITA. Swedish Interactive Threshold algorithm. *Investigative Ophthalmology & Visual Science* **40**(6), 1152–1161 (05 1999)
27. Xie, S., Kamoi, K., Igarashi-Yokoi, T., Uramoto, K., Takahashi, H., Nakao, N., Ohno-Matsui, K.: Structural Abnormalities in the Papillary and Peripapillary Areas and Corresponding Visual Field Defects in Eyes With Pathologic Myopia. *Investigative Ophthalmology & Visual Science* **63**(4), 13–13 (04 2022)
 28. Xu, L., Asaoka, R., Kiwaki, T., Murata, H., Fujino, Y., Yamanishi, K.: Pami: A computational module for joint estimation and progression prediction of glaucoma. In: *KDD*. pp. 3826–3834 (2021)
 29. Yan, Z., Liang, D., Xu, L., Li, J., Liu, Z., Wang, S., Cao, J., Kee, C.: VF-HM: Vision loss estimation using fundus photograph for high myopia. In: *MICCAI*. vol. 14226, pp. 649–659 (2023)
 30. Yan, Z., Liu, Z., Li, J.: Boosting of implicit neural representation-based image denoiser. In: *ICASSP* (2024)
 31. Zhang, S., Yang, L., Mi, M.B., Zheng, X., Yao, A.: Improving deep regression with ordinal entropy. In: *ICLR* (2023)
 32. Zheng, Y., Xu, L., Kiwaki, T., Wang, J., Murata, H., Asaoka, R., Yamanishi, K.: Glaucoma progression prediction using retinal thickness via latent space linear regression. In: *KDD*. pp. 2278–2286 (2019)