Domain Adaptive Diabetic Retinopathy Grading with Model Absence and Flowing Data

Wenxin Su[1], Song Tang*[1,2,3], Xiaofeng Liu[4], Xiaojing Yi[5], Mao Ye[6], Chunxiao Zu[1], Jiahao Li[7], and Xiatian Zhu[8]

[1]University of Shanghai for Science and Technology [2]Universität Hamburg, [3]ComOriginMat Inc. [4]Yale University [5]Sichuan Eye Hospital [6]University of Electronic Science and Technology of China [7]Chinese Academy of Medical Sciences & Peking Union Medical College [8]University of Surrey

steventangsong@gmail.com

1. Proof of Theorem

1.1. A Proof of Theorem 1

Recalling traditional unadversarial learning. Unadversarial learning aims to develop an image perturbation that enhances the performance on a specific class, which can be succinctly described as follows:

$$\hat{\delta} = \arg\min_{\delta} L(f_\theta(x+\delta), y), s.t.\ ||\delta|| \leq \epsilon \tag{1}$$

where $L(\cdot)$ denotes objective function, x and y are input image and its label, f_θ is a pre-trained model with parameters θ, δ is a perturbation, ϵ is a small threshold. Solves this problem in an iterative way formulated as

$$\delta_{k+1} = \delta_k + \alpha \cdot \text{sign}\left(\nabla_x L(f_\theta(x+\delta_k), y)\right), k \in [0, K-1], \tag{2}$$

where α is a trade-off parameter, K is iteration number, δ_0 is an initial random noise. We re-consider the iterative optimization process above and obtain the theorem below.

Restatement of Theorem 1 *Given the unadversarial learning problem defined in Eq. (1), the iterative process featured by Eq. (2) can be expressed as the following generative form.*

$$\delta_k = \delta_0 + V \cdot F_\Phi\left(\frac{\partial \delta_0}{\partial x}\right), \tag{3}$$

where δ_0 is an initial random noise, V is a bound constant, F_Φ is a generative function.

Proof. First, according to the chain principle, we can convert Eq. (2) into

$$\delta_{k+1} = \delta_k + \alpha \cdot \left(\frac{\partial L}{\partial f_\theta} \cdot \frac{\partial f_\theta}{\partial x} \cdot \left(1 + \frac{\partial \delta_k}{\partial x}\right)\right). \tag{4}$$

*Corresponding author

Since that the learning will converge to the unadversarial examples, $\alpha \cdot \frac{\partial L}{\partial f_\theta} \cdot \frac{\partial f_\theta}{\partial x}$ is bounded by a certain constant, denoted by $U_k > 0$, thereby Eq. (4) become

$$\delta_{k+1} \leq \delta_k + U_k\left(1 + \frac{\partial \delta_k}{\partial x}\right). \tag{5}$$

We make a further substitution on δ_k according to the law presented in Eq. (5), leading to

$$\delta_{k+1} \leq \left[\delta_{k-1} + U_{k-1}\left(1 + \frac{\partial \delta_{k-1}}{\partial x}\right)\right] + U_k\left(1 + \frac{\partial \delta_k}{\partial x}\right). \tag{6}$$

By continuing this substitution on $\delta_{k-1}, \cdots, \delta_0$ in order, we have

$$\begin{aligned}\delta_{k+1} &\leq \delta_0 + U_0\left(1 + \frac{\partial \delta_0}{\partial x}\right) + U_1\left(1 + \frac{\partial \delta_1}{\partial x}\right) \\ &\quad + \cdots + U_i\left(1 + \frac{\partial \delta_i}{\partial x}\right) + \cdots + U_k\left(1 + \frac{\partial \delta_k}{\partial x}\right) \\ &\leq \delta_0 + U_m\left[k + \frac{\partial \delta_0}{\partial x} + \frac{\partial \delta_1}{\partial x} + \cdots + \frac{\partial \delta_k}{\partial x}\right],\end{aligned} \tag{7}$$

where $U_m = max\{U_0, U_1, \cdots, U_k\}$.

To obtain generative form, we explore the relationships between $\{\frac{\partial \delta_1}{\partial x}, \frac{\partial \delta_2}{\partial x}, \cdots, \frac{\partial \delta_k}{\partial x}\}$ and $\frac{\partial \delta_0}{\partial x}$, respectively. To this end, we first investigate the relationship between $\frac{\partial \delta_1}{\partial x}$ and $\frac{\partial \delta_0}{\partial x}$, combining Eq. (5).

$$\frac{\partial \delta_1}{\partial x} \leq \frac{\partial \delta_0}{\partial x} + U_1 \cdot \frac{\partial}{\partial x}\left(\frac{\partial \delta_0}{\partial x}\right) = h_1\left(\frac{\partial \delta_0}{\partial x}\right) \tag{8}$$

where, $h_1(\cdot)$ stands for an equivalent function. For $\frac{\partial \delta_2}{\partial x}$, we

have the following equation based on Eq. (5) and Eq. (8).

$$\begin{aligned}\frac{\partial \delta_2}{\partial x} &\leq \frac{\partial \delta_1}{\partial x} + U_2 \cdot \frac{\partial}{\partial x}\left(\frac{\partial \delta_1}{\partial x}\right) \\ &= h_1\left(\frac{\partial \delta_0}{\partial x}\right) + U_2 \cdot \frac{\partial}{\partial x}\left(h_1\left(\frac{\partial \delta_0}{\partial x}\right)\right) \\ &= h_2\left(\frac{\partial \delta_0}{\partial x}\right)\end{aligned} \tag{9}$$

In the recursion way presented by Eq. (8) and Eq. (9), $\{\frac{\partial \delta_3}{\partial x}, \cdots, \frac{\partial \delta_k}{\partial x}\}$ can be expressed as

$$\frac{\partial \delta_3}{\partial x} \leq h_3\left(\frac{\partial \delta_0}{\partial x}\right), \cdots, \frac{\partial \delta_k}{\partial x} \leq h_k\left(\frac{\partial \delta_0}{\partial x}\right) \tag{10}$$

Therefore, substituting Eq. (8), (9) and (10) into Eq. (7), we have

$$\delta_{k+1} \leq \delta_0 + U_m\left[k + \frac{\partial \delta_0}{\partial x} + h_1\left(\frac{\partial \delta_0}{\partial x}\right) + \cdots + h_k\left(\frac{\partial \delta_0}{\partial x}\right)\right]. \tag{11}$$

Let $F_\Phi\left(\frac{\partial \delta_0}{\partial x}\right) = \left[k + \frac{\partial \delta_0}{\partial x} + h_1\left(\frac{\partial \delta_0}{\partial x}\right) + \cdots + h_k\left(\frac{\partial \delta_0}{\partial x}\right)\right]$ and V be a value that makes the equality relationship hold. Eq. (11) becomes the generative form below.

$$\delta_k = \delta_0 + V \cdot F_\Phi\left(\frac{\partial \delta_0}{\partial x}\right). \tag{12}$$

1.2. A Proof of Theorem 2

Recalling the calculation of the fine-grained saliency map. It calculates saliency by measuring central-surround differences within images.

$$\begin{aligned}&\mathrm{G}(h,w) = \sum_{\varsigma} \max\left\{\mathrm{cen}(h,w) - \mathrm{sur}(h,w,\varsigma), 0\right\}, \\ &\mathrm{cen}(h,w) = I(h,w), \\ &\mathrm{sur}(h,w,\varsigma) = \frac{\sum_{h'=-\varsigma}^{h'=\varsigma}\sum_{w'=-\varsigma}^{w'=\varsigma} I(h+h', w+w') - I(h,w)}{(2\varsigma+1)^2 - 1},\end{aligned} \tag{13}$$

where (h, w) is the coordinate of one pixel in grey-scale image (transformed by x_t) with its corresponding value denoted as $I(w, h)$, and $\varsigma \in \{1, 3, 7\}$ denotes surrounding values.

Restatement of Theorem 2 *Given the partial derivatives of the initial random noise δ_0 w.r.t image x is $\frac{\partial \delta_0}{\partial x}$ and x's saliency map is $s = G(x)$ where G is the computation function of saliency map. We have the following relationship:*

$$\frac{\partial \delta_0}{\partial x} \leq U \cdot s, \tag{14}$$

where $U > 0$ is a bound constant.

Proof. we treat s as a middle variable, thus $\frac{\partial \delta_0}{\partial x}$ can be expressed as the following equation by the chain law.

$$\frac{\partial \delta_0}{\partial x} = \frac{\partial \delta_0}{\partial s} \cdot \frac{\partial s}{\partial x} \leq U \cdot \frac{\partial s}{\partial x}, \tag{15}$$

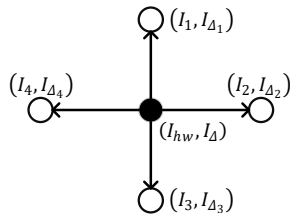


Figure 1. Illustration of $x + \triangle_x$ at coordinate (h, w) as we select the simplest surround case $\varsigma = 1$.

where $U > 0$ is a bound constant. In Eq. (15), the inequality holds because both the initial noise and the specific saliency map are bounded, resulting in the relative changes between them also being restricted. In addition, according to the definition of derivative, we have

$$\frac{\partial s}{\partial x} = \frac{\partial G(x)}{\partial x} \approx \frac{G(x + \triangle_x) - G(x)}{\triangle_x}, \tag{16}$$

where $\triangle_x$ is a tiny variation.

It is known that the saliency map at (h, w) is only related to itself and its surrounding pixels. Without loss of generality, we build the proof based on the simplest surround case $\varsigma = 1$ where $\triangle_x$ at (h, w) is presented by Fig. 1. According to Eq. (13), we have

$$\begin{aligned}&\mathrm{cen}(h,w,\triangle_x) = \mathrm{cen}(h,w) + I_\triangle = I_{hw} + I_\triangle. \\ &\mathrm{sur}(h,w,\varsigma,\triangle_x) \\ &= \frac{\sum_{i=1}^4 (I_i + I_{\triangle i}) - (I_{hw} + I_\triangle)}{8}, \\ &= \frac{\left(\sum_{i=1}^4 I_i - I_{hw}\right) + \left(\sum_{i=1}^4 I_{\triangle i} - I_\triangle\right)}{8}, \\ &= \mathrm{sur}(h,w,\varsigma) + \frac{\mathrm{sur}_\triangle(\varsigma,\triangle_x)}{8}.\end{aligned} \tag{17}$$

Thus, $G(x + \triangle_x)$ at (h, w) can be expressed as

$$\begin{aligned}G_{hw}(x+\triangle_x) = \sum_{\varsigma} &\max\{[\mathrm{cen}(h,w) - \mathrm{sur}(h,w,\varsigma)] \\ &- \left[\frac{1}{8}\mathrm{sur}_\triangle(\varsigma,\triangle_x) - I_\triangle\right], 0\}\end{aligned} \tag{18}$$

Let $A_1 = \mathrm{cen}(h,w)$ and $A_2 = \mathrm{sur}(h,w,\varsigma)$, $B_1 = \mathrm{cen}(h,w) - \mathrm{sur}(h,w,\varsigma)$, $B_2 = \frac{1}{8}\mathrm{sur}_\triangle(\varsigma,\triangle_x) - I_\triangle$. Eq. (16) has two situations as follows.

- S-1. When $A_1 > A_2, B_1 > B_2$,

$$\begin{aligned}&\frac{G(x+\triangle_x) - G(x)}{\triangle_x} \\ &= \frac{I_\triangle - \frac{1}{8}\mathrm{sur}_\triangle(\varsigma,\triangle_x)}{I_\triangle} \\ &= \frac{1}{2} - \sum_{i=1}^4 \frac{I_{\triangle i}}{I_\triangle}\end{aligned} \tag{19}$$

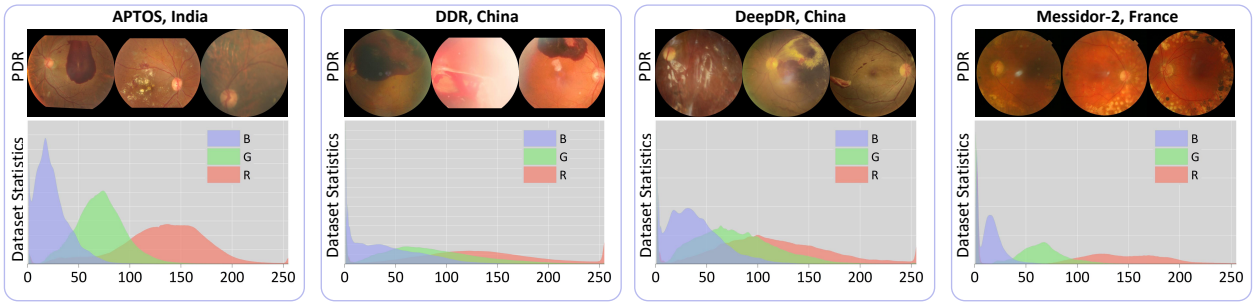


Figure 2. Visualize the styles and characteristics of each dataset by analyzing the RGB statistics of proliferative diabetic retinopathy (PDR) samples across APTOS, DDR, DeepDR, and Messidor-2.

- S-2. When $A_1 < A_2, B_1 < B_2$,

$$\frac{G(x+\triangle_x) - G(x)}{\triangle_x} = 0 \tag{20}$$

Let $d_1 = \text{cen}(h,w) - \text{sur}(h,w,\varsigma)$, $d_2 = \frac{1}{2} - \sum_{i=1}^{4} \frac{I_{\triangle_i}}{I_{\triangle}}$, we have another two cases:

- S-3. When $A_1 < A_2, B_1 > B_2$,

$$\frac{G(x+\triangle_x) - G(x)}{\triangle_x} \leq D_1(1+\frac{d_2}{d_1})d_1, \tag{21}$$

where D_1 is a bound constant.

- S-4. When $A_1 > A_2, B_1 < B_2$,

$$\frac{G(x+\triangle_x) - G(x)}{\triangle_x} \leq D_2 d_1, \tag{22}$$

where D_2 is another bound constant.

The four cases above provide us with an insight that $\frac{\partial s}{\partial x}$ is proportional to the saliency map s, namely

$$\frac{\partial s}{\partial x} \propto s. \tag{23}$$

There are two reasons contributing to this conclusion. First, $\frac{\partial s}{\partial x}$' values confine to a binary situation. More importantly, in S-3 and S-4, $\frac{\partial s}{\partial x}$ is proportional to the central-surround pixel value difference d_1, which are also depicted by saliency maps. Combing Eq. (15) and Eq. (23), we have

$$\frac{\partial \delta_0}{\partial x} \leq U \cdot \frac{\partial s}{\partial x} \propto U \cdot s. \tag{24}$$

2. Implementation Details

2.1. Datasets Details

Dataset description. We evaluate the proposed method on four standard DR benchmarks. Their details are presented as follows.

- **APTOS** [1] The dataset originates from Kaggle's APTOS 2019 Blindness Detection Contest, organized by the Asia Pacific Tele-Ophthalmology Society (APTOS). It comprises a total of 5,590 fundus images provided by Aravind Eye Hospital in India. However, only the annotations for the training set (3,662 images) are publicly accessible, and these are used in this study.
- **DDR** [3] The DDR dataset comprises 13,673 fundus images collected from 9,598 patients across 23 provinces in China. These images are classified by seven graders based on features such as soft exudates, hard exudates, and hemorrhages.
- **DeepDR** [5] The DeepDR dataset comprises 2,000 fundus images of both left and right eyes from 500 patients in Shanghai, China.
- **Messidor-2** [2] The Messidor-2 dataset includes 1,748 macula-centered eye fundus images. This dataset partially originates from the Messidor program partners, with additional images contributed by Brest University Hospital in France.

Table 1. Label distribution of the four evaluation datasets: **APTOS**, **DDR**, **DeepDR**, and **Messidor-2**.

Dataset	No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR	Total
APTOS	1,805	370	999	193	295	3,662
DDR	6,265	630	4,477	236	913	13,673
DeepDR	914	222	398	354	112	2,000
Messidor-2	1,017	270	347	75	35	1,748

The label distribution of datasets. All datasets exhibit imbalanced class distributions, as shown in Table 1. Specifically, in APTOS, the "No DR" class comprises about **49.2**% of all samples. In DDR, "No DR" accounts for approximately **45.8**%, while in DeepDR, it makes up around **45.7**%. In Messidor-2, the "No DR" class represents about **58.2**% of the total data.

The domain shift of datasets. Each dataset is treated as a distinct domain, with significant variations from factors like country of origin, patient demographics, and differences in imaging equipment used for acquisition. Additionally, analysis of the RGB statistics for proliferative DR (PDR) samples across these datasets/domains reveals distinct fluctuations in

Table 2. Performance of test time adaptation methods evaluated in ACC, QWK, and AVG across different batch sizes

Method	ACC							QWK							AVG						
Source	53.9							60.1							57.0						
	Test Time Adaptation Batch Size							Test Time Adaptation Batch Size							Test Time Adaptation Batch Size						
	2	4	8	16	32	64	Avg.	2	4	8	16	32	64	Avg.	2	4	8	16	32	64	Avg.
SHOT-IM [4]	44.9	54.2	58.5	58.0	59.2	59.0	55.6	60.8	60.9	62.0	63.2	64.4	64.7	62.7	52.8	57.5	60.0	60.8	61.8	61.9	59.1
TENT [7]	56.3	57.1	57.8	58.8	59.7	59.3	58.2	25.1	30.2	39.7	47.2	54.1	59.2	42.6	40.7	43.6	48.7	53.0	56.9	59.3	50.4
SHOT-IM+GUES	60.0	60.9	61.4	61.5	61.4	62.0	61.2	64.7	65.2	65.6	65.8	66.1	66.9	65.7	62.4	63.1	63.5	63.6	63.7	64.5	63.5
TENT+GUES	60.6	61.0	61.3	61.2	61.1	61.0	61.0	62.5	62.3	62.2	62.4	62.5	63.3	62.5	61.5	61.7	61.8	61.8	61.8	62.2	61.8

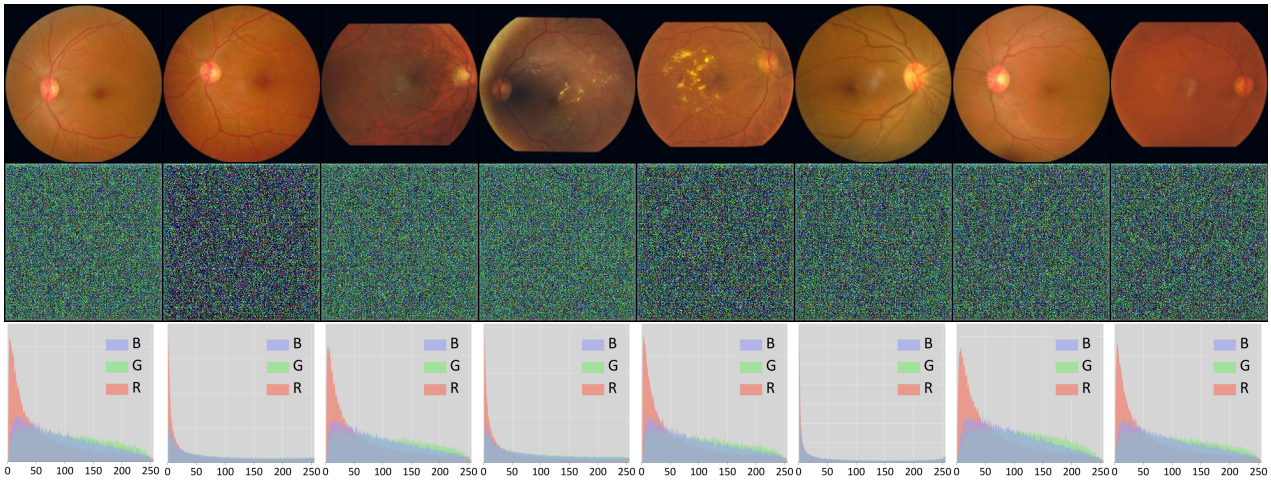


Figure 3. Visualization for input images, generative perturbations, and RGB statistic of the corresponding perturbations on transfer task DDR→APTOS.

each channel (R, G, and B), highlighting the unique visual styles and characteristics of each dataset, as shown in Fig. 2.

3. Evaluation metrics

The computation rules for accuracy (termed ACC), Quadratic Weighted Kappa (termed QWK), and the average of QWK and ACC (termed AVG) are as follows.

$$
\begin{aligned}
ACC &= \frac{TP+TN}{TP+TN+FP+FN}, \\
QWK &= 1-\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W(i,j)\cdot O(i,j)}{\sum_{i=1}^{n}\sum_{j=1}^{n} W(i,j)\cdot E(i,j)},\ W_{i,j}=\frac{(i-j)^2}{(C-1)^2} \\
AVG &= \frac{1}{2}\left(ACC+QWK\right),
\end{aligned}
\tag{25}
$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. i is a true category, j is a predicted category, C is the number of classes, and n is the total number of samples. $O(i,j)$ is the observed frequency, which represents how many times the true category i was predicted as category j, and $E(i,j)$ is the expected frequency, which indicates how many times category i would be predicted as category j under random guessing, $E(i,j) = P(i) \times P(j) \times n$.

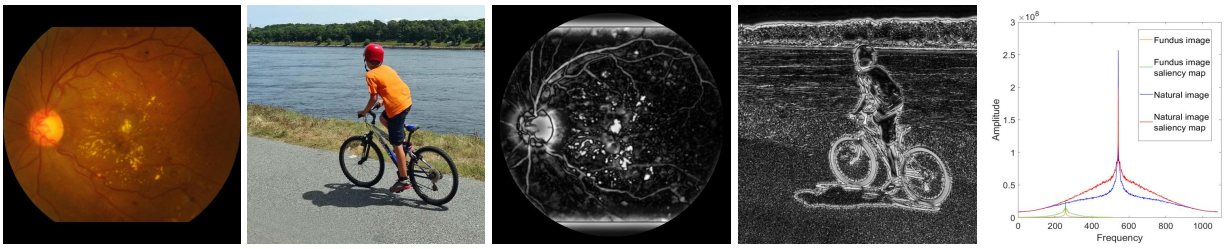
4. Supplementary Experiment Results

4.1. Results with Varying Batch Size

As a supplement to the results with varying batch sizes, Table 2 presents the complete performance of three evaluation metrics across all 12 tasks. TTA methods SHOT-IM and TENT show a performance drop when the batch size is small. Specifically, SHOT-IM decreases by approximately **14.1**% in ACC, **3.9**% in QWK, and **9.1**% when comparing batch sizes of 2 and 64. TENT decreases by approximately **3.0**% in ACC, **34.1**% in QWK, and **18.6**% when comparing batch sizes of 2 and 64. However, when these methods are combined with our proposed method, GUES, the decline is not as significant. In SHOT-IM+GUES, the performance shows a decrease of only **2.0**% in ACC, **2.0**% in QWK, and **2.1**% in AVG. In TENT+GUES, the performance shows a decrease of only **0.4**% in ACC, **0.8**% in QWK, and **0.7**% in AVG. These results indicate that our method can prevent declines when the batch size is small, as it predicts individual perturbations that are robust to batch size variations.

4.2. Visualization for Generative Perturbations

As depicted in Fig. 3, it is evident that different input images exhibit distinct perturbations, as observed directly in the sec-



(a) Fundus image (b) Natural image (c) Fundus image saliency map (d) Natural image saliency map (e) Amplitude spectrum

Figure 4. Visualization of a fundus image, a natural image, and their corresponding saliency maps. The fundus image is sampled from APTOS, and the natural image is sampled from Office-Home [6]. In (e), the amplitude spectrum of these four images is displayed.

ond row. To be more specific, the RGB distribution of the perturbations, illustrated in the third row, further highlights their variability. This analysis demonstrates how GUES dynamically adjusts the perturbations to account for the unique characteristics of each input image, effectively tailoring them to align with the target domain.

4.3. Why are Saliency Maps Unsuitable for Natural Images?

As we early stated, the proposed method cannot tackle the natural image scenarios well. This part executes a further discussion for this issue using two typical images illustrated in Fig. 4 (a) and (b). There are two key observations to note. First, the fundus image has a simpler background and structure compared to the natural image, which features richer semantics, including diverse shapes, complex relative structures, and intricate backgrounds. This difference is reflected in the amplitude spectrum in Fig. 4(e), where the fundus image displays a significantly lower frequency band. Second, the saliency maps effectively highlight variations in both fundus and natural images. This is indicated by the fact that the amplitudes of the saliency maps are much larger than the corresponding amplitudes of the images at similar frequencies.

The effects of this enhancement differ between fundus images and natural images. For simpler fundus images, the noticeable variations are typically related to lesions, making the enhancement useful for highlighting these specific regions (see Fig. 4 (c)). In contrast, complex natural images exhibit variations that span the entire scene, such as areas of forest, grass, shadows, and a person riding a bike. In this case, the enhancement draws attention to all elements in the image, which can obscure the factors that are relevant to the task at hand. Therefore, we believe that refining a proper self-supervised signal for natural images represents a promising research direction for the future.

Table 3. Inference time comparison (per-image) on DDR→APTOS.

Method	SHOT-IM	TENT	SAR	GUES
Time (ms)	11.9	5.6	24.1	14.7

4.4. Comparison of Training Time

As shown in Tab. 3, GUES demonstrates an average speed among the competitors. The primary reason GUES is not the fastest is that the VEA model inherently requires more training time than the ResNet50 model. Specifically, the VEA model's loss function is based on reconstruction loss, which is computationally intensive, whereas ResNet50 handles classification tasks using cross-entropy loss, which is more computationally efficient. Furthermore, the VEA model's decoder typically takes low-dimensional latent representations as input and decodes them into high-resolution images through multiple layers. This process is significantly more computationally demanding compared to ResNet50, which directly processes raw image features. Therefore, the training time of GUES is relatively longer.

References

[1] Aptos: Aptos 2019 blindness detection website. https://www.kaggle.com/c/aptos2019-blindness-detection, accessed February 20, 2022.

[2] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, John-Richard Ordóñez-Varela, Pascale Massin, Ali Erginay, et al. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, pages 231–234, 2014.

[3] Tao Li, Yingqi Gao, Kai Wang, Song Guo, Hanruo Liu, and Hong Kang. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences*, 501: 511–522, 2019.

[4] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*, 2020.

[5] Ruhan Liu, Xiangning Wang, Qiang Wu, Ling Dai, Xi Fang, Tao Yan, Jaemin Son, Shiqi Tang, Jiang Li, Zijian Gao, et al.

Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge. *Patterns*, 3(6), 2022.

[6] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017.

[7] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *ICLR*, 2020.