Leveraging Diffusion Models for Continual Test-Time Adaptation in Fundus Image Classification

Mingsi Liu[1,2][0009−0002−0882−5640], Xiang Li[6][0000−0001−7955−6498], Mengxiang Guo[7][0000−0002−5974−3533], Lixin Duan[1,2⋆][0000−0002−0723−4016], Huihui Fang[4,5⋆⋆][0000−0003−3380−7970], and Yanwu Xu[3,5][0000−0002−1779−931X]

[1] School of Computer Science and Engineering & Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
[2] Sichuan Provincial Key Laboratory for Human Disease Gene Study and the Center for Medical Genetics, Department of Laboratory Medicine, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, UESTC
[3] School of Future Technology, South China University of Technology
[4] College of Computing and Data Science, Nanyang Technological University
[5] Pazhou Lab
[6] School of Intelligent Science and Technology, Nanjing University
[7] Guangzhou Women and Children's Medical Center
lxduan@uestc.edu.cn, huihui.fang@ntu.edu.sg

Abstract. Continual Test-Time Adaptation (CTA) aims to improve model generalization under distribution shifts by adapting to incoming test data. However, conventional CTA methods, such as pseudo-label refinement and entropy minimization, face challenges in fundus image classification due to the limited number of training samples and class categories, which lead to overconfident yet miscalibrated predictions, making traditional adaptation methods ineffective. To address these issues, we propose a novel diffusion-based CTA framework, DiffCTA, which leverages the generative capabilities of diffusion models to refine test samples and align them with the source domain distribution without modifying the source model. DiffCTA enhances test-time adaptation using diffusion guidance while preserving diagnostic features. Specifically, we integrate content guidance to retain anatomical structures, consistency guidance to stabilize predictions via entropy minimization, style guidance for CLIP-based domain alignment, and a sampling optimization module that dynamically adjusts guidance strength across diffusion timesteps. We conducted experiments on glaucoma classification and diabetic retinopathy grading tasks. In the glaucoma classification task, our method outperformed the best existing approach by 2.6%, demonstrating its effectiveness in handling domain shifts without modifying the source model. The code is available at: https://github.com/mingsiliu557/DiffCTA.

Keywords: Fundus Image · Continual Test-Time Adaptation · Diffusion Model · Glaucoma Classification · Diabetic Retinopathy Grading

⋆ Corresponding author: lxduan@uestc.edu.cn
⋆⋆ Corresponding author: huihui.fang@ntu.edu.sg

1 Introduction

Medical image classification is a critical task for accurately diagnosing diseases by identifying pathological and anatomical features, yet models often struggle with real-world deployment due to distribution shifts caused by variations in imaging protocols, operators, and scanners [7]. In fundus disease diagnosis, early detection is vital to prevent vision loss. Color fundus photography provides detailed views of the optic nerve and retinal structures, which are essential for detecting disease progression. However, domain variability across imaging devices and settings degrades the performance of trained models during testing, making consistent and reliable diagnostics challenging in clinical practice.

Test-time adaptation (TTA) allows models to handle distribution shifts using only test data, without requiring a labeled target dataset as in traditional domain adaptation (DA) [24,3,25,29]. Most TTA methods use self-supervised auxiliary losses, such as entropy minimization [22] and self-training [18], to adapt model parameters. However, these methods often struggle in fundus imaging due to dynamic domain shifts caused by variations in imaging protocols and equipment across hospitals. To address evolving distributions, continual test-time adaptation (CTA) has been proposed [23], requiring models to adapt sequentially across multiple domains. Traditional CTA methods mitigate error accumulation and catastrophic forgetting through loss optimization [23,14] or regularization [13], but they rely on self-supervised losses prone to noisy supervision and overfitting. In fundus disease classification, the limited data and category diversity lead to overfitted source models with overconfident, low-entropy predictions, making entropy-based adaptation methods (e.g., EATA [13], CoTTA [23], TENT [22]) ineffective in handling domain shifts. Thus, to improve adaptation in dynamic fundus imaging, VPTTA [4] uses Fourier-based prompt tuning for target-to-source transformation without modifying the source model but requires extra prompt training. In contrast, our diffusion-based approach directly refines target images while preserving structural integrity.

Diffusion models enable continual test-time adaptation by iteratively denoising test samples, aligning them with the source distribution [6,21]. Unlike traditional methods, they adapt without modifying the source model, making them well-suited for fundus image classification.

Inspired by this, we introduce Leveraging Diffusion Models for Continual Test-Time Adaptation in Fundus Image Classification (DiffCTA), a diffusion–driven framework designed specifically for fundus disease classification under continual test-time adaptation. The main contributions of this work are as follows: 1) To the best of our knowledge, this is the first work to address continual test-time adaptation for fundus image classification, introducing a novel diffusion-based adaptation framework to mitigate domain shifts without altering the source model; 2) We design four key components to enhance adaptation: content guidance to preserve anatomical structures, consistency guidance to stabilize predictions via entropy minimization, style guidance to harmonize domain shifts using CLIP-based alignment, and a sampling optimization module that dynamically regulates guidance strength across diffusion timesteps; 3) Extensive

experiments on multiple fundus imaging datasets demonstrate that DiffCTA achieves state-of-the-art performance, highlighting its effectiveness in handling domain shifts and its potential impact in medical imaging applications.

2 Proposed method

Our DiffCTA framework applies diffusion-based TTA by introducing structured guidance at each reverse diffusion step. Specifically, we design three guidance mechanisms: content guidance for anatomical preservation, consistency guidance for prediction stability, and style guidance for semantic information of fundus images. Additionally, we incorporate an anatomy-aware sampling optimization strategy for fundus disease classification. Inspired by [6], our method enhances adaptation through dynamic integration of these components. The overall architecture is shown in Fig. 1.

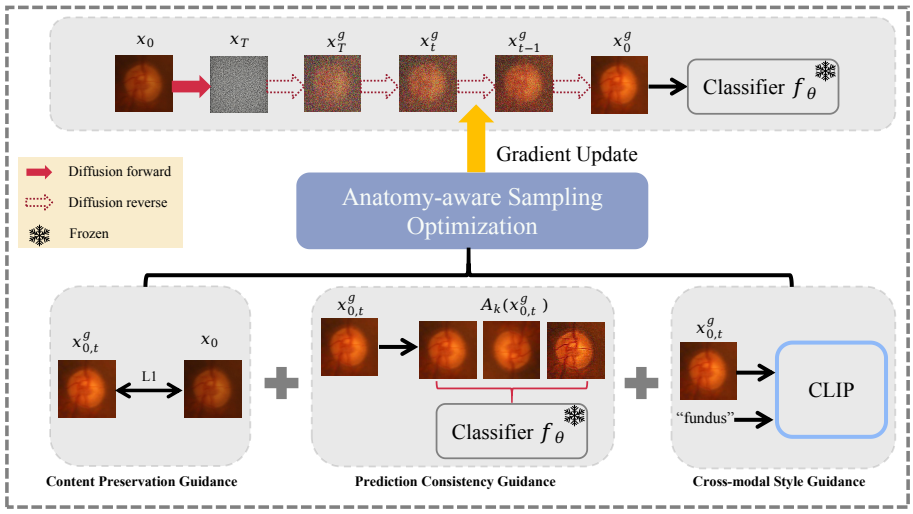

Fig. 1. The DiffCTA architecture applies content, consistency, and style guidance during reverse diffusion to refine target images. An anatomy-aware sampling strategy updates each step, with $x_{0,t}^g$ predicted from x_t^g via the DDIM reverse process.

2.1 DiffCTA Framework and Diffusion Process

Our DiffCTA framework integrates diffusion dynamics with clinical prior guidance for fundus domain adaptation. Given a pre-trained unconditional diffusion model on the source domain $\mathcal{X}_S$ and a target domain input x_0, the model generates samples $\hat{x}_0$ that progressively align with the source domain $\mathcal{X}_S$. We utilize Denoising diffusion probabilistic models (DDPM) for adaptation [10], where the forward process gradually adds Gaussian noise to a target domain image x_0 over

Algorithm 1 DiffCTA Adaptation Algorithm

Input: Target image x_0, timesteps T, noise predictor ϵ_θ, classifier f_ϕ
Output: Adapted prediction $\hat{y}$
$x_T^g \leftarrow q(x_T|x_0), \quad x_T^g \sim \mathcal{N}(1,0)$ ▷ Forward process
Set $x^{\text{blk}} \leftarrow \mathbf{0}$, $x^{\text{whi}} \leftarrow \mathbf{1}$ ▷ Generate full white and full black sample
for $t = T$ **downto** 1 **do**
    Compute x_{t-1}^g via Eq. 2 ▷ Reverse process
    Estimate intermediate clean image:

$$x_{0,t}^g = \frac{x_t^g - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t^g, t)}{\sqrt{\bar{\alpha}_t}}$$

    if $||x_0 - x_{0,t}^g||_1 \leq \max(||x^{\text{blk}} - x_{0,t}^g||_1, ||x^{\text{whi}} - x_{0,t}^g||_1)$ **then**
        $g_{\text{content}} \leftarrow \nabla_{x_t^g} ||x_{0,t}^g - x_0||_1$ ▷ Content guidance
        $g_{\text{style}} \leftarrow \eta \nabla_{x_t^g} \left(\frac{\langle E_{\text{i}}(x_{0,t}^g), E_{\text{t}}(r) \rangle}{||x_{0,t}^g|| \cdot ||r||} \right)$ ▷ Style guidance
        $g_{\text{consist}} \leftarrow \nabla_{x_t^g} \mathcal{H}\Big(\frac{1}{K} \sum_{k=1}^{K} p_\theta(y|\mathcal{A}_k(x_{0,t}^g)) \Big)$ ▷ Consistency guidance
        Update: $x_{t-1}^g \leftarrow x_{t-1}^g - \lambda\,(g_{\text{content}} + g_{\text{style}} + g_{\text{consist}})$
    end if
end for
$\hat{y} \leftarrow f_\phi(x_0^g)$
return $\hat{y}$

T steps using a fixed Markov chain. At each timestep t, noise with variance β_t is added, producing a noisy sequence $[x_0, x_1, ..., x_T]$. This forward process is defined as:

$$q(x_t|x_0) = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0,1)$ is the noise we add, $\alpha_t = 1-\beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. The reverse process iteratively removes noise to generate denoised images $[x_T^g, x_{T-1}^g, ..., x_0^g]$, with each step defined as:

$$x_{t-1}^g = \frac{1}{\sqrt{\alpha_t}} \left(x_t^g - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t^g, t) \right) + \sigma_t \epsilon, \tag{2}$$

where ϵ_θ is the model to predict noise, and σ_t is the noise variance.

2.2 Content Preservation Guidance

To preserve key fundus structures such as the optic disc and vasculature, we enforce content consistency during the denoising process. Specifically, we generate an intermediate clean estimate $x_{0,t}^g$ using the DDIM reverse formulation[20] (Eq. 3):

$$x_{0,t}^g = \frac{x_t^g - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t^g, t)}{\sqrt{\bar{\alpha}_t}}. \tag{3}$$

To ensure that $x_{0,t}^g$ retains structural fidelity with the target image x_0, we apply an L1 loss that enforces pixel-wise consistency and preserves essential anatomical features (Eq. 4):

$$g_{\text{content}} = \nabla_{x_t^g} ||x_{0,t}^g - x_0||_1, \tag{4}$$

where g_{content} represents the gradient that guides the diffusion process to maintain fundus content. This constraint prevents excessive distortion during adaptation, ensuring that the generated images remain diagnostically reliable while aligning with the source domain.

2.3 Prediction Consistency Guidance

To mitigate stochastic noise and accumulated errors during adaptation, we enforce prediction consistency by leveraging augmentation invariance. Specifically, a set of transformations $\{\mathcal{A}_k\}_{k=1}^{K}$ (e.g., slight rotations, intensity shifts) is applied to x_t, and the classifier predictions are averaged to form a smoothed probability distribution (Eq. 5):

$$\bar{p}_\theta(y|x_{0,t}^g) \approx \frac{1}{K}\sum_{k=1}^{K} p_\theta(y|\mathcal{A}_k(x_{0,t}^g)), \tag{5}$$

where $p_\theta(y|x_{0,t}^g)$ is the classifier prediction for $x_{0,t}^g$, and $\bar{p}_\theta(y|x_{0,t}^g)$ is the mean prediction over all augmented samples. To enhance adaptation stability, we minimize the uncertainty of $\bar{p}_\theta(y|x_{0,t}^g)$ by computing the entropy loss gradient (Eq. 6):

$$g_{\text{consist}} = \nabla_{x_t^g}\mathcal{H}\Big(\bar{p}_\theta(y|x_{0,t}^g)\Big), \tag{6}$$

where $\mathcal{H}(\cdot)$ denotes the entropy function that quantifies prediction uncertainty. This consistency constraint steers the diffusion model towards stable and source-aligned predictions, reducing the impact of distributional shifts and ensuring robustness across augmentations.

2.4 Cross-Modal Style Guidance

To better preserve the semantic information of fundus images, we incorporate CLIP-based guidance. A predefined text prompt r (e.g., "fundus") encodes the desired visual characteristics of the source domain. The alignment is enforced by computing the cosine similarity between the image and text embeddings (Eq. 7):

$$\mathcal{S}(x_{0,t}^g, r) = \frac{\langle E_{\text{i}}(x_{0,t}^g), E_{\text{t}}(r)\rangle}{||x_{0,t}^g|| \cdot ||r||}, \tag{7}$$

where $E_{\text{i}}(x_{0,t}^g)$ and $E_{\text{t}}(r)$ represent the CLIP-encoded embeddings of the image and text, respectively. To ensure the generated image inherits the style of the source domain, we optimize this similarity by computing its gradient (Eq. 8):

$$g_{\text{style}} = \eta\nabla_{x_t^g}\mathcal{S}(x_{0,t}^g, r), \tag{8}$$

where η is a scaling factor controlling the strength of style alignment. This guidance refines the appearance of the adapted image, making it visually consistent with the source domain while preserving diagnostic integrity.

2.5 Anatomy-Aware Sampling Optimization

To avoid applying guidance when the image is dominated by noise, inspired by [26], we introduce a structural check to determine whether the generated sample contains meaningful fundus structures. Specifically, guidance is only activated if the L1 distance between $x^g_{0,t}$ and x_0 is smaller than the maximum distance between $x^g_{0,t}$ and reference blank images x^{blk} (fully black) and x^{whi} (fully white) (Eq. 9):

$$||x_0 - x^g_{0,t}||_1 \leq \max(||x^{\text{blk}} - x^g_{0,t}||_1, ||x^{\text{whi}} - x^g_{0,t}||_1), \tag{9}$$

where x^{blk} and x^{whi} are blank and white images of the same size as x_0. This condition ensures that guidance is only applied when $x^g_{0,t}$ exhibits discernible anatomical structures, preventing unnecessary updates in early noisy timesteps.

3 Experiments

3.1 Datasets and Implementation

The glaucomatous classification dataset comprises five public datasets collected from different medical centers, denoted as domain A (RIM-ONE-r3 [5]), B (REFUGE-train [15]), C (ORIGA [28]), D (ACRIMA [16]), and E (Drishti-GS [19]). These datasets contain 159, 400, 650, 705, and 101 images, respectively. We cropped a region of interest (ROI) centered at the optic disc (OD) with a size of 800×800 following [11], resized it to 512×512, and applied min-max normalization. This is a binary classification task where each image is categorized as either glaucomatous or non-glaucomatous, and accuracy is used for evaluation.

The diabetic retinopathy grading dataset consists of four public datasets from different medical centers: domain A (aptos2019-train [1]), B (Messidor2 [2]), C (IDRiD [17]), and D (SUSTech-SYSU [12]), containing 3662, 1744, 516, and 1219 images, respectively. Each image is graded into five levels based on the severity of diabetic retinopathy: No DR, Mild, Moderate, Severe, and Proliferative DR. Accuracy is used as the evaluation metric.

Implementation details. We implemented DiffCTA using torch 1.9.0 and cuda 11.1, and conducted all experiments on a single GeForce RTX 4090. For each task, we trained the source model on a single dataset and evaluated it across the remaining datasets as target domains, reporting the average performance across all target domains. The diffusion model is an unconditional 256×256 model trained on the source dataset. We used the publicly released BioMedCLIP [27] checkpoint as the visual-language backbone for CLIP-related components. ResNet50 [8] was used as the baseline for both tasks. During test-time adaptation, we applied a single-iteration adaptation per batch (batch size = 1) across all experiments for DiffCTA and competing methods. We employed DDPM for forward and reverse diffusion with a total timestep of $T = 50$. For the marginal entropy loss, we utilized AugMix [9], which randomly applies augmentations such as posterization, rotation, and equalization. The style alignment

factor η and gradient scaling factor λ were set to 5 and 8, respectively, to regulate style guidance during reverse diffusion and balance adaptation strength.

3.2 Results

Comparison on the glaucomatous classification task. We evaluate our proposed DiffCTA on the glaucoma classification task, comparing it with the "Source Only" baseline and five state-of-the-art CTA methods. The results, summarized in Table 1, show that all adaptation methods outperform the baseline, confirming their ability to mitigate domain shifts. Notably, DiffCTA achieves the highest accuracy across all domains, demonstrating its superior adaptability and robustness. It not only performs well in domains where most methods achieve reasonable results (e.g., Domain E) but also excels in challenging domains where other methods struggle (e.g., Domain C). Furthermore, as shown in Fig. 2, our adapted fundus images exhibit greater alignment with the source domain, highlighting the effectiveness of diffusion-based adaptation.

Table 1. Glaucoma accuracy (%) on five target domains. Underline = 2nd best; Bold = best.

Method	Domain A	Domain B	Domain C	Domain D	Domain E	AVG
Source Only	68.37	50.68	65.74	<u>34.98</u>	<u>43.43</u>	52.64
TENT [22]	65.84	<u>58.84</u>	60.36	28.42	42.76	51.24
CoTTA [23]	64.01	58.51	61.25	24.02	33.59	48.28
EATA [13]	66.46	58.50	63.34	33.41	40.42	52.43
SAR [14]	66.57	58.81	63.21	32.87	33.52	51.00
DDA [6]	<u>69.71</u>	56.06	<u>67.20</u>	34.22	39.73	<u>53.38</u>
DiffCTA	**70.47**	**59.34**	**68.46**	**35.59**	**45.93**	**55.96**

Comparison on the diabetic retinopathy grading task. For the diabetic retinopathy grading task, we conducted a comparative study under the same experimental setup. As presented in Table 2, DiffCTA consistently outperforms competing methods, benefiting from training on the current test image without modifying the source model. This ability to effectively adapt without parameter updates enables DiffCTA to achieve the highest performance across multiple target domains, further validating its robustness in handling domain shifts in medical image classification.

Ablation study. We conducted an ablation study on the glaucoma classification task using the REFUGE dataset to evaluate the contribution of our proposed components. The introduced components include: 1) **Content guidance**, preserving anatomical structures by enforcing pixel-wise consistency through L1 regularization; 2) **Style guidance**, aligning the target domain with the source distribution using CLIP-based feature matching; 3) **Consistency guidance**, stabilizing predictions via entropy minimization; and 4) **Sampling optimization**, dynamically regulating guidance strength across diffusion timesteps. The

Table 2. Diabetic Retinopathy accuracy (%) across four target domains. Underline = 2nd best; Bold = best.

Method	Domain A	Domain B	Domain C	Domain D	AVG
Source Only	59.81	46.65	<u>49.35</u>	52.18	52.00
TENT [22]	52.19	48.24	35.21	48.76	46.10
CoTTA [23]	36.27	43.45	26.36	35.60	35.42
EATA [13]	56.39	48.13	45.09	43.93	48.39
SAR [14]	45.18	48.24	36.03	43.51	43.24
DDA [6]	<u>60.99</u>	<u>55.97</u>	49.33	<u>52.45</u>	<u>54.68</u>
DiffCTA	**61.64**	**57.32**	**49.89**	**53.35**	**55.55**

results, shown in Table 3, demonstrate the effectiveness of each component in improving adaptation performance.

Table 3. Results of the ablation study on the glaucoma classification task using the REFUGE dataset (source domain).

Components	ACC (%)
Content	56.06
Content + Style	58.10
Content + Style + Consistency	<u>58.23</u>
Content + Style + Consistency + Sampling Optimization	**59.34**

4 Conclusion

In this work, we present the first study on CTA for fundus disease classification and propose a novel diffusion-based framework, DiffCTA. By leveraging the reverse diffusion process, DiffCTA aligns target domain images with the source distribution while preserving key diagnostic features. Unlike traditional CTA methods, which suffer from overfitting and unstable adaptation, DiffCTA addresses these challenges through content, consistency, and style guidance, along with an optimized sampling strategy. Extensive experiments validate the effectiveness of DiffCTA, achieving state-of-the-art performance across multiple datasets. These results highlight the potential of diffusion-based adaptation in clinical medical imaging, especially in scenarios where direct model parameter updates are impractical. In the future, we plan to extend DiffCTA to segmentation and other medical imaging tasks to enhance its applicability.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (No. 82121003), the Guangzhou Science and Technology Project (2024D03J0013), and the Pazhou Lab's basic cloud computing platform.

Disclosure of Interests. The authors declare no competing interests.

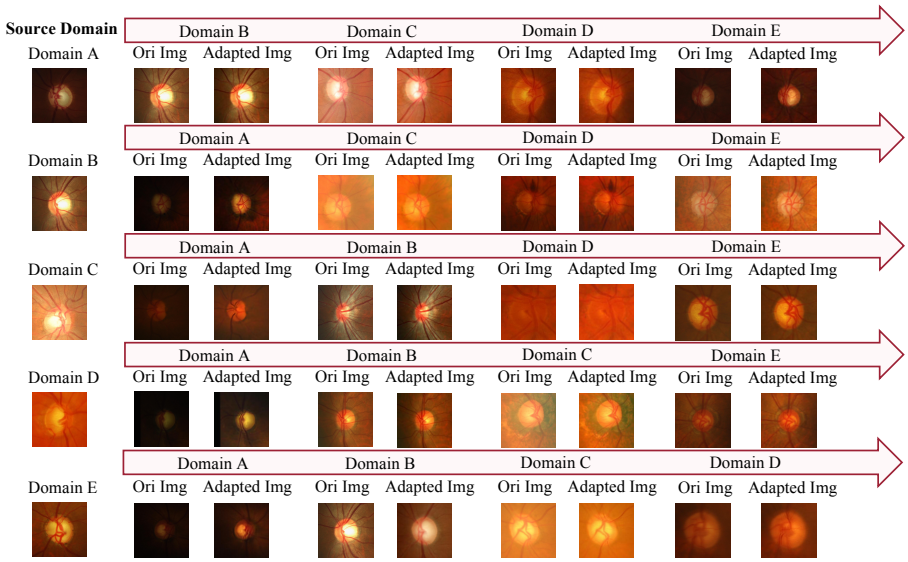

Fig. 2. Visualization results on the glaucoma classification task. The domains under the red arrows represent the target domains.

References

1. APTOS 2019 Blindness Detection. Online (2019), `https://www.kaggle.com/c/aptos2019-blindness-detection/`, accessed: 10-Jun-2019
2. Messidor Fundus Image Dataset. Online (nd), `https://www.adcis.net/en/third-party/messidor/`, kindly provided by the Messidor program partners.
3. Chen, L., Chen, H., Wei, Z., Jin, X., Tan, X., Jin, Y., Chen, E.: Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 7181–7190 (2022)
4. 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. pp. 11184–11193 (2024)
5. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: RIM-ONE: An open retinal image database for optic nerve evaluation. In: Int. Symp. Comput.-based Med. Syst. pp. 1–6. IEEE (2011)
6. Gao, J., Zhang, J., Liu, X., Darrell, T., Shelhamer, E., Wang, D.: Back to the source: Diffusion-driven test-time adaptation. arXiv preprint arXiv:2207.03442 (2022)
7. Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R., de Leeuw, F.E., Tempany, C.M., Van Ginneken, B., et al.: Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In: Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. pp. 516–524. Springer (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 770–778 (2016)

9. Hendrycks, D., Mu, N., Cubuk, E.D., Zoph, B., Gilmer, J., Lakshminarayanan, B.: Augmix: A simple data processing method to improve robustness and uncertainty (2020)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Hu, S., Liao, Z., Xia, Y.: Prosfda: Prompt learning based source-free domain adaptation for medical image segmentation. arXiv preprint arXiv:2211.11514 (2022)
12. Lin, L., Li, M., Huang, Y., Cheng, P., Xia, H., Wang, K., Yuan, J., Tang, X.: The sustech-sysu dataset for automated exudate detection and diabetic retinopathy grading. Scientific Data **7**(1), 409 (2020)
13. Niu, S., Wu, J., Zhang, Y., Chen, Y., Zheng, S., Zhao, P., Tan, M.: Efficient test-time model adaptation without forgetting. In: Int. Conf. Mach. Learn. pp. 16888–16905. PMLR (2022)
14. Niu, S., Wu, J., Zhang, Y., Wen, Z., Chen, Y., Zhao, P., Tan, M.: Towards stable test-time adaptation in dynamic wild world. In: Int. Conf. Learn. Represent. (2023)
15. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med. Image Anal. **59**, 101570 (2020)
16. Ovreiu, S., Paraschiv, E.A., Ovreiu, E.: Deep learning & digital fundus images: Glaucoma detection using densenet. In: 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). pp. 1–4 (2021). https://doi.org/10.1109/ECAI52376.2021.9515188
17. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid) (2018). https://doi.org/10.21227/H25W98, https://dx.doi.org/10.21227/H25W98
18. Sinha, S., Gehler, P., Locatello, F., Schiele, B.: Test: Test-time self-training under distribution shift. In: IEEE Wint. Conf. Appli. Comput. Vis. pp. 2759–2769 (2023)
19. Sivaswamy, J., Krishnadas, S., Joshi, G.D., Jain, M., Tabish, A.U.S.: Drishti-GS: Retinal image dataset for optic nerve head (onh) segmentation. In: IEEE Int. Symp. on Bio. Imaging. pp. 53–56. IEEE (2014)
20. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
21. Tsai, Y.Y., Chen, F.C., Chen, A.Y., Yang, J., Su, C.C., Sun, M., Kuo, C.H.: Gda: Generalized diffusion for robust test-time adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23242–23251 (2024)
22. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: Int. Conf. Learn. Represent. (2021)
23. Wang, Q., Fink, O., Van Gool, L., Dai, D.: Continual test-time domain adaptation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 7201–7211 (2022)
24. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. pp. 102–110. Springer (2019)
25. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 4085–4095 (2020)
26. Yu, J., Wang, Y., Zhao, C., Ghanem, B., Zhang, J.: Freedom: Training-free energy-guided conditional diffusion model. CoRR **abs/2303.09833** (2023). https://doi.org/10.48550/arXiv.2303.09833, https://doi.org/10.48550/arXiv.2303.09833

27. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., et al.: Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv preprint arXiv:2303.00915 (2023)
28. Zhang, Z., Yin, F.S., Liu, J., Wong, W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: ORIGA-light: An online retinal fundus image database for glaucoma analysis and research. In: Int. Conf. IEEE Eng. in Med. and Bio. pp. 3065–3068. IEEE (2010)
29. Zuo, Y., Yao, H., Xu, C.: Attention-based multi-source domain adaptation. IEEE Trans. Image Process. **30**, 3793–3803 (2021)