Computers in Biology and Medicine 174 (2024) 108418

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Source-free active domain adaptation for diabetic retinopathy grading based on ultra-wide-field fundus images

Jinye Ran [a], Guanghua Zhang [b,d], Fan Xia [c], Ximei Zhang [d], Juan Xie [e], Hao Zhang [f,*]

[a] College of Computer and Information Science, Southwest University, Chongqing 400700, China
[b] School of Big Data Intelligent Diagnosis and Treatment Industry, Taiyuan University, Taiyuan 030002, China
[c] Reading Academy, Nanjing University of Information Science and Technology, Nanjing 210044, China
[d] College of Biomedical Engineering, Taiyuan University of Technology, Taiyuan 030600, China
[e] Shanxi Eye hospital, Taiyuan 030002, China
[f] College of Chemistry and Chemical Engineering, Southwest University, Chongqing 400700, China

ARTICLE INFO

Keywords:
Diabetic retinopathy grading
Source-free domain adaptation
Ultra-wide-field fundus images
Active learning

ABSTRACT

Domain adaptation (DA) is commonly employed in diabetic retinopathy (DR) grading using unannotated fundus images, allowing knowledge transfer from labeled color fundus images. Existing DAs often struggle with domain disparities, hindering DR grading performance compared to clinical diagnosis. A source-free active domain adaptation method (SFADA), which generates features of color fundus images by noise, selects valuable ultra-wide-field (UWF) fundus images through local representation matching, and adapts models using DR lesion prototypes, is proposed to upgrade DR diagnostic accuracy. Importantly, SFADA enhances data security and patient privacy by excluding source domain data. It reduces image resolution and boosts model training speed by modeling DR grade relationships directly. Experiments show SFADA significantly improves DR grading performance, increasing accuracy by 20.90% and quadratic weighted kappa by 18.63% over baseline, reaching 85.36% and 92.38%, respectively. This suggests SFADA's promise for real clinical applications.

1. Introduction

Diabetic retinopathy (DR) is one of the most common secondary microvascular complications in diabetic patients, caused by damage to the blood vessels in the light-sensitive tissue at the back of the eye [1,2]. With the increasing number of diabetic patients, the prevalence and blindness rate of DR have also been increasing, and it is currently the leading cause of blindness [3]. According to the Global Diabetes Report released by the World Health Organization in 2017, the number of diabetic patients in China is as high as 116 million, and the prevalence of DR is from 24.70% to 37.50% [4–6]. Since DR has no obvious characteristics in the early stage, finding potential patients and intervening in time through regular fundus screening can reduce the risk of severe blindness faced by 94.40% of patients [7,8]. However, there are fewer than 1,000 doctors in China who specialize in fundus medical services and research, resulting in more than 90% of diabetic patients unable to undergo regular fundus screening [6].

In traditional clinical practice, DR is classified into non-proliferative DR (NPDR) and proliferative DR (PDR) based on the detailed examination of microaneurysms (MA), hemorrhages (HE), soft exudates (SE), and hard exudates (EX) in fundus images [9,10]. Each fundus image can be specifically divided into five DR stages of progressive severity: normal, mild, moderate, severe non-proliferative, and proliferative according to international protocols [11,12]. Although precise and repeatable diagnostic criteria for DR grading is developed, the identification of retinopathy areas and categories is so difficult that it faces problems of inefficiency, subjectivity, and even the risk of misdiagnosis. [10] In recent years, computer-aided diagnosis (CAD) technology based on deep learning has gradually been applied to the task of DR grading based on traditional color fundus images, which can greatly alleviate the anxiety of ophthalmologists [13]. It is reported that these advanced CAD systems achieve DR grading performance equal to or even exceeding that of humans on some real-world DR datasets [14]. The misdiagnosis rate of DR is significantly correlated with the visualization range of the fundus [15]. Traditional fundus photography techniques expand the visualization range to 75° by stitching together seven standard 30° fields of view using montage technology [16]. However, this approach requires extensive patient cooperation after mydriasis and takes more than two hours to complete. In contrast, ultra-wide-field (UWF) retinal imaging devices can capture 80% of the retinal image in 0.4 s with a 200° UWF angle of view [17]. This allows for the

* Corresponding author.
E-mail address: haozhang@swu.edu.cn (H. Zhang).

https://doi.org/10.1016/j.compbiomed.2024.108418
Received 27 October 2023; Received in revised form 19 March 2024; Accepted 4 April 2024
Available online 5 April 2024

observation of a larger area of the retina, which can lead to improved detection of DR lesions such as non-perfusion areas, neovascularization, and panretinal photocoagulation. Thus, UWF fundus angiography has become the mainstream technique for DR diagnosis and management. . However, due to the difficulty of both acquiring and labeling UWF fundus images, there are currently no available UWF fundus datasets comparable in scale to traditional color fundus datasets. As a result, most CAD systems based on supervised learning paradigms hardly benefit from the advances of photography technology.

To remedy this defect and obtain better performance in DR grading, a large number of DR grading studies based on UWF fundus images have been reported in the research community. One line of studies is to annotate the UWF fundus images and to propose a new CAD system, following the previous research paradigm of supervised learning, such as [15,18]. However, the cost of annotation is so expensive that most attempts are based on a few hundred UWF fundus images. While better performance in DR grading can be obtained, there is always a risk of overfitting owing to the scarcity of samples for neural network training. Another line of studies is to switch the research paradigm to transfer learning, such as domain adaptation (DA) [19,20], which enables automatic DR grading of UWF fundus images by transferring the annotated knowledge from the traditional color fundus images domain to the almost unannotated UWF fundus images domain. Traditional domain adaptation methods require access to both source domain and target domain data. However, in the medical domain, source domain data often contains highly sensitive personal privacy information and is stored in a distributed manner across multiple hospitals. Laws and regulations enacted by governments severely restrict the accessibility of data in real-world medical scenarios, rendering domain adaptation methods infeasible. Consequently, source-free domain adaptation (SFDA), which utilizes only the source-trained model and unlabeled target data to adapt to the target domain, has garnered increasing attention in the medical field. However, due to the existence of huge domain gaps and the quality of fundus images, most SFDA methods are hard to achieve an ideal DR grading performance. Generally, increasing the proportion of labeled data in a dataset can enhance model performance to a certain extent. However, given the high cost of medical data labeling, it is desirable to minimize the proportion of labeled data while ensuring the accuracy of DR grading. Active learning selects samples with high information content from the unlabeled database for labeling and incorporates them into the training set via a rationally designed query function. Through optimal experimental design, it can effectively reduce labeling costs while improving the performance of the DR grading model. Therefore, source-free active domain adaptation (SFADA) that integrates active learning strategies with SFDA may serve as an effective approach to improving the performance of ultra-widefield retinal DR grading models.

This paper presents a novel SFADA framework for DR grading based on UWF fundus images. Our approach prioritizes data privacy and achieving superior DR grading performance. We aim to extract DR grading knowledge from a well-trained model on a dataset of traditional color fundus images. Subsequently, we leverage this knowledge to develop a robust and powerful DR grading model for UWF fundus images. Notably, the proposed framework achieves superior DR grading performance in more realistic application scenarios. This advancement significantly improves the potential for practical application of DA in automated DR diagnosis. The main contributions of this paper are as follows:

- To address the challenge of SFADA for fine-grained DR grading, we propose a novel approach which models the evolving relationships between different DR grades in color fundus images and leverages noise to generate features for these images. This strategy not only protects data privacy but also promotes computational efficiency.
- We propose a novel active local representation matching for valuable UWF fundus images selecting. This method can effectively reduce the impact of outliers on sample actively selecting process and is suitable for the source-free DA scenario of real-world fundus images.
- A novel active local representation matching approach for selecting valuable UWF fundus images is introduced in this paper. This method effectively mitigates the impact of outliers during the active sample selection process, making it well-suited for the source-free DA scenario commonly encountered with real-world fundus image datasets.
- Our extensive experimental results demonstrate that the proposed SFADA framework achieves state-of-the-art performance in DR grading, while simultaneously ensuring data privacy and computational efficiency.

The rest of the paper is organized as follows: The related work of this paper is reviewed in Section 2. The details of our proposed SFADA framework are elaborated in Section 3. The experimental particulars and results are presented in Section 4, along with the discussion and the ablation experiments for each part of the method, and the conclusion is shown in Section 5.

2. Related work

Our work is closely related to domain adaptation, source-free domain adaptation, and active learning for DR grading. A brief introduction to these three aspects will be presented in this section.

2.1. Domain adaptation for DR grading

Domain adaptation has been widely studied in the field of medical images, which can significantly reduce the dependence of research on annotated fundus images. There are also a lot of related studies based on domain adaptation methods in terms of automatic DR diagnosis. Ju et al. [21] proposed an adversarial domain adaptation to bridge the domain between UWF fundus images and traditional color fundus images. With the help of pseudo-labeling and adversarial learning, the performance of DR grading is improved. Bai et al. [22] focused on interpretability in DR grading and investigated an unsupervised lesion-aware transfer learning framework for DR grading in UWF fundus images. A lesion external attention module is introduced to transfer fine lesion knowledge, obtaining a significant improvement in DR grading performance. Inspired by the success of fixed-ratio based mixup in unsupervised domain adaptation, Wei et al. [20] designed cross-domain collaborative learning for recognizing multiple retinal diseases. By employing Transformers for producing scale-invariant features and limited annotation of the target domain, more meaningful knowledge transfer is achieved. However, although the methods mentioned above have improved the performance of DR grading compared with their baseline, from the perspective of DR diagnosis, the DR grading performance needs to be further improved.

2.2. Source-free domain adaptation for DR grading

Source-free domain adaptation is a new setting of domain adaptation proposed in recent years, considering the data privacy issues in real scenarios. In the field of DR grading, data privacy is even more involved. Zhang et al. [23] proposed a source-free transfer learning approach for DR grading. The source domain samples with similar target domain samples are generated by a generative adversarial network to realize knowledge transfer with no source domain data. Different from the above methods, Zhou et al. [24] argued to generate source domain samples directly by a generative adversarial network and then begin to adapt based on these forged source domain data with label and target domain data. Pourreza [25] study the source-free domain adaptation in

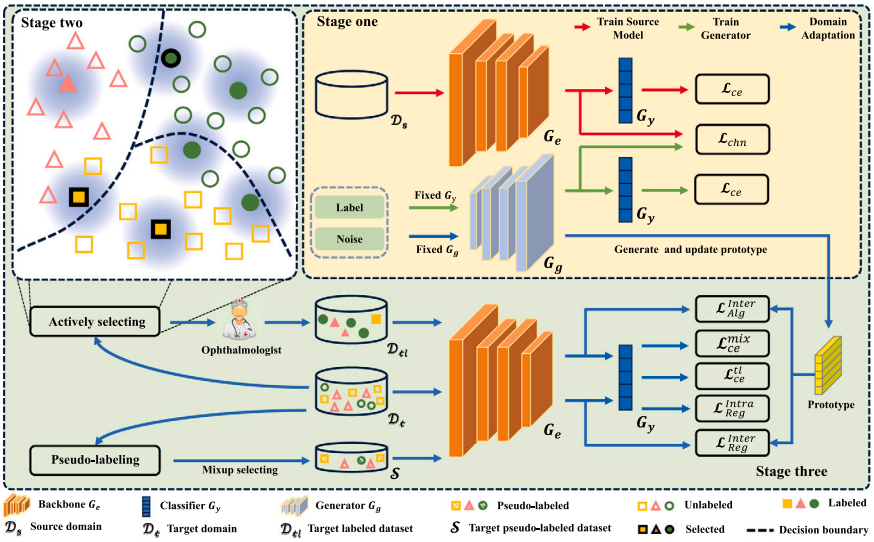


Fig. 1. The flowchart of the proposed source-free active domain adaptation (SFADA). The Stage one is to train a source pre-trained model with annotated color fundus images and a source feature generator. The Stage two aims to actively select a few valuable ultra-wide-field (UWF) fundus images for labeling, and the Stage three adapts the source pre-trained model to the UWF fundus images. Notably, The Stage two and the Stage three run alternately.

open-set. The source classifier is fixed, while fine-tuning is performed on the source backbone using target domain data to achieve source-free domain adaptation. However, these methods are hard to further develop due to problems such as high computational costs and difficulty accounting for source domain knowledge.

2.3. Active learning for DR grading

Active learning is a paradigm that focuses on human–computer interaction and is often used to model complex data in real-world scenarios. Active learning provides a means to leverage expert knowledge to intervene in the neural network modeling process, generally leading to better performance in automatic DR grading. Qureshi et al. [26] proposed an active learning method named expected gradient length to realize DR grading. With the help of clinical specialists, it achieved remarkable DR grading results on the Eyepacs. Ahsan et al. [27] proposed a hybrid model for active learning. By quantifying the uncertainty and selecting the worthiest sample to label, nearly perfect area under the curve performance was obtained. However, active learning is still rare in the DR grading field of UWF fundus images.

Inspired by the work mentioned above, this paper proposes a novel SFADA for DR grading of UWF fundus images, focusing on data privacy, computational efficiency, and superior DR grading performance.

3. Methodology

3.1. Problem formulation

In the SFADA setting, there is a source pre-trained model $G = \{G_e, G_y\}$ trained by annotated color fundus images and a target domain consisting of unannotated UWF fundus images $\mathcal{D}_t = \{x_i^t\}$, where G_e and G_y are, respectively, source classifier and source backbone. Source domain consisting of annotated color fundus images is unavailable during domain adaptation. Meanwhile, a certain amount of unannotated UWF fundus images $\mathcal{B}$ will be actively selected for labeling ($\mathcal{B} \ll \mathcal{D}_t$). Let all of the selected target domain data be $\mathcal{D}_{tl} = \{x_i^{tl}, y_i^{tl}\}$. The primary goal of SFADA is to adapt G to the target domain with $\mathcal{D}_t$ and $\mathcal{D}_{tl}$, thereby improving the DR grading performance of G in UWF fundus images.

3.2. Overall scheme

We exploit to study a superior performance domain adaptation for UWF-based DR grading, which is closer to the real-world scenarios. As shown in Fig. 1, the proposed SFADA consist of three stages: source domain feature generation, active local representation matching, and lesion-based prototype domain adaptation. After stage one, stage two and stage three will run alternately for several rounds until the algorithm ends in accordance with the established hyper-parameter settings. More specific details refer to Algorithm 1.

Algorithm 1: Overall Scheme of SFADA

Input : Source pre-trained model $\{G_e, G_y\}$; Target domain data $\mathcal{D}_t$; Source domain data $\mathcal{D}_s$; Generator G_g; Training epochs E, M, N; Number classes of DR grades K;
Output: Adapted model $\{G_e, G_y\}$

Initialization $\mathcal{D}_{tl} = \emptyset$; $\mathcal{D}_{tu} = \mathcal{D}_t$; Source prototype P_k;
for $e \leftarrow 0$ **to** E **do**
  Randomly sample a mini-batch from $\mathcal{D}_s$;
  Update parameters of $\{G_e, G_y\}$ with Eq. (1) and (2);
end
Freeze parameters of G_y;
for $n \leftarrow 0$ **to** N **do**
  Randomly sample $\xi \sim \mathcal{N}(0,1)$, $\kappa \sim U(0,K)$;
  Update parameters of G_g with Eq. (1) and (2);
end
Freeze parameters of G_g;
for $m \leftarrow 0$ **to** M **do**
  if *need active selection* **then**
    Obtain active selection set $\mathcal{D}_a$ as described in Algorithm 2;
    $\mathcal{D}_{tl} \leftarrow \mathcal{D}_{tl} \cup \mathcal{D}_a$, $\mathcal{D}_{tu} \leftarrow \mathcal{D}_{tu} \setminus \mathcal{D}_a$;
  end
  Initialize $\mathcal{D}_{add} = \emptyset$, $\mathcal{D}_{rev} = \emptyset$, $\mathcal{S} \leftarrow \mathcal{D}_{tl}$, iterations I on $\mathcal{D}_{tu}$;
  for $i \leftarrow 0$ **to** I **do**
    Update source prototype P_k with G_g and Eq. (5);
    Sample mini-batches $\{x_{tl}, y_{tl}\}$ from $\mathcal{S}$, x_t from $\mathcal{D}_{tu}$;
    Calculate $\mathcal{L}_{ce}^{tl}$, $\mathcal{L}_{Alg}^{Inter}$, $\mathcal{L}_{Reg}^{Inter}$, $\mathcal{L}_{Reg}^{Intra}$, $\mathcal{L}_{ce}^{mix}$ with Eq. (6), (7), (8), (9), (11), and Update the parameters of $\{G_e, G_y\}$;
    if *need pseudo label* **then**
      Update $\mathcal{D}_{add}$ and $\mathcal{D}_{rev}$ as described in Algorithm 3;
      if *iteration over* $\mathcal{S}$ *finishes* **then**
        $\mathcal{S} \leftarrow \mathcal{S} \setminus \mathcal{D}_{rev} \cup \mathcal{D}_{add}$, $\mathcal{D}_{add} = \emptyset$, $\mathcal{D}_{rev} = \emptyset$;
      end
    end
  end
end

3.3. Source domain feature generation

Source-free domain adaptation is a successful attempt to respond to the data privacy of DR grading [23,24]. However, most mainstream methods suffer from the computational efficiency, which is not conducive to further development. To alleviate this, inspired by [28], we propose a novel source domain feature generation for domain adaptation without source data. Specifically, our approach is divided into two steps. The first step is to obtain a source pre-trained model $\{G_e, G_y\}$ with color fundus images via a cross-entropy loss function

$$\mathcal{L}_{ce} = \mathbb{E}_{(x,y)\sim D_s} - \sum_{k=0}^{K} y_k log(G_y(G_e(x))) \tag{1}$$

where k is one hot label. The second step is to train a source feature generator G_g using a normal noise $\xi \sim \mathcal{N}(0,1)$ and a uniform label $\kappa \sim U(0,K)$ under the condition of a source classifier G_y, which loss function is still $\mathcal{L}_{ce}$. While ξ is multiplied with the embedded κ as input x to G_g to ensure diversity in the generated source features, the parameters of G_y are simultaneously fixed to maintain a close relationship between the generated source features and their corresponding labels. Notably, instead of generating source-like samples, our method generates source features for source-free domain adaptation, which can improve computational efficiency.

To further control the computational cost of source-free domain adaptation, we re-think the influence of the fundus image resolution on DA. DR grading using lower-resolution fundus images can significantly improve computational efficiency, while some fundus lesion regions (e.g., HE and EX) become so small that fine-grained feature adaptation is required to be considered. To deal with this issue, we propose to model the continuously evolving relationship from normal to proliferative DR. In terms of mutual information, this implies greater mutual information between neighboring DR grades. However, high-dimensional mutual information is difficult to estimate and calculate [29]. Also, as a viable alternative, due to the unboundedness in the maximization process and the suppression of inter-class mutual information by later cross-entropy optimization, directly maximizing the lower bound of mutual information will conflict with the optimization objective of Eq. (1), leading to a potential collapse in certain scenarios. Fortunately, we can implicitly optimize the mutual information via contrastive loss function [30,31]. So, we propose an improved contrastive loss function to model the continuously evolving relationship, which can be formed as

$$\mathcal{L}_{chn} = -log \frac{exp((p \cdot k^{+} - \gamma)/\tau)}{exp((p \cdot k^{+} - \gamma)/\tau) + \sum_{j=1}^{C-1} exp((p \cdot k^{-} - \beta_j)/\tau)} \tag{2}$$

where p, k^{+}, and k^{-} denote anchor samples, positive samples for anchor sample, and negative samples for anchor sample. γ, τ, and β = represent hyper-parameters, respectively. As shown in Fig. 2. the effect of the proposed loss function $\mathcal{L}_{chn}$ on color fundus images distribution in the feature space is visualized onto the hyper-sphere. The left hyper-sphere represents the source pre-trained model trained with $\mathcal{L}_{ce}$, and the right one denotes the source pre-trained model trained with $\mathcal{L}_{ce} + \mathcal{L}_{chn}$. The histogram is the radian between the features of each interval relationship before and after using $\mathcal{L}_{chn}$. Obviously, the $\mathcal{L}_{chn}$ drives the DR features of adjacent interval relationship closer together, establishing a more ordered relationship among different interval relationship, which means $\mathcal{L}_{chn}$ successfully models the relative relationship between different DR grades.

To this end, both the source pre-trained model and the source feature generator are trained with $\mathcal{L}_{ce}$ and $\mathcal{L}_{chn}$. Essentially, $\mathcal{L}_{chn}$ increases the convergence difficulty of $\mathcal{L}_{ce}$ while modeling the continuously evolving relationship from normal to proliferative DR, obtaining a more powerful source classifier G_y, which is conducive to train the source feature generator G_g and adapt the fine-grained knowledge. The flowchart of source domain feature generation is illustrated in stage one of Fig. 1.

3.4. Active local representation matching

Guided by uncertainty and representativeness criteria, active learning selects a few valuable UWF fundus images for labeling and supervised training, thus greatly improving the performance of DR grading. However, due to outliers such as eyelash occlusion and huge domain gaps, the general active learning method is difficult to directly apply to DA in real-world UWF fundus images. To handle this, inspired by the exploration of local information in active domain adaptation [32,33], we propose an active local representation matching (ALRM) for active sample selection of UWF-based DR grading in DA.

Let $p(\cdot) = \sigma(\mathbf{W} \otimes h(\cdot))$ denote different DR probabilities after a softmax layer $\sigma(\cdot)$, where $\mathbf{W}$ and $h(\cdot)$ are the parameters of classifier G_y and feature extracted by source backbone G_e, respectively. To alleviate the effect of source-free setting and make full use of source pre-trained model, $\mathbf{W}$ is regarded as class prototypes to map low-dimensional semantic information $p(\cdot)$ to high-dimensional feature information $p(x_t^i) \otimes \mathbf{W}^{\top}$. Then, we propose a novel criterion to measure the local representation of each UWF fundus image x_t, which can be formed as

$$LR(x_t) = \frac{1}{N} \sum_{i=1}^{N} [h(x_t^i) \odot p(x_t^i) \otimes \mathbf{W}^{\top}] + h(x_t^i) \tag{3}$$

where x_t^i is one of the N nearest neighbors of x_t selected by cosine similarity $<h(\cdot), h(\cdot)^{\top}>$. There is an insight that local representation is essentially an averaging of the most similar sample features, reducing the influence of outliers on data-sensitive active sample selection. Meanwhile, the introduction of class probabilistic information also reduces the risk of collapse.

Based on obtained all of local representations of $\mathcal{D}_t$, we try to actively find a subset $\mathcal{D}_{tl} \subset \mathcal{D}_t$ such that $\mathcal{D}_{tl}$ matches $\mathcal{D}_t$ as well as possible. Although this is a NP-hard problem that cannot be solved in polynomial time, it could be approximated by a greedy selection process with relatively correct results. Inspired by [34], an unlabeled UWF fundus image that minimizes the squared Maximum Mean Discrepancy (MMD) between $\mathcal{D}_t$ and $\mathcal{D}_{tl}$ is selected each time, which can be formulated as

$$\arg\min_{x_t} MMD^2(\mathcal{D}_t, \mathcal{D}_{tl} \leftarrow x_t) \tag{4}$$

The process of local representation matching is shown in stage two of Fig. 1. An interesting finding is that the designed subset matching works only with our proposed local representation. After further analysis of UWF fundus images, we speculate that the intra-class differences of different DR grades are too large, and the limited active samples could not fully match the entire feature space, resulting in the deterioration of active learning performance. Additionally, considering the initial poor performance of G on the UWF fundus images, active learning is performed for $\mathcal{K}$ rounds, and $\mathcal{B}/\mathcal{K}$ UWF fundus images of the target domain are selected in each round. More details refer to Algorithm 2.

3.5. Lesion-based prototype domain adaptation

After actively selecting, domain adaptation is usually implemented using classical DANN [35] or MME [36], while supervised training is performed on the active UWF fundus images. However, the DR grading performance achieved by these methods is far from our expectations. To re-think the advantages of actively selecting samples, we note that these annotated UWF fundus images can be more fully exploited by semi-supervised methods, like consistency regularity [37]. To this end, we propose a lesion-based prototype domain adaptation (LPDA) to adapt G to UWF fundus images better, which is shown in stage three of Fig. 1. The proposed LPDA includes three parts: prototype-based domain contrastive alignment, inter and intra domain consistency regularity, and selecting pseudo label with mixup samples.

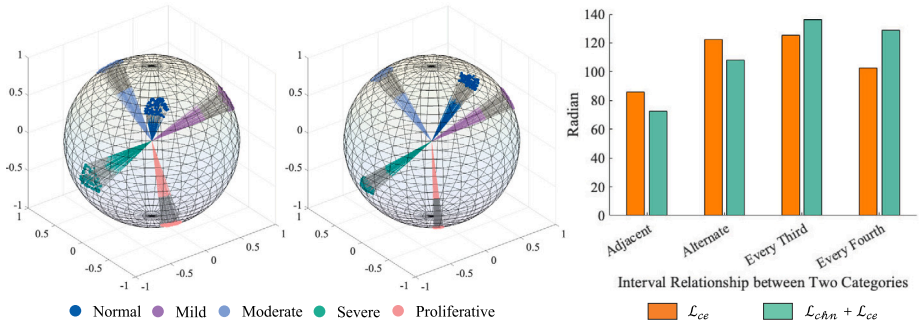


Fig. 2. Visualization of source domain feature in the hyper-sphere. The left hyper-sphere is trained by $\mathcal{L}_{ce}$ and the right hyper-sphere is trained by $\mathcal{L}_{ce} + \mathcal{L}_{chn}$. The histogram is the radian between the features of each interval relationship before and after using $\mathcal{L}_{chn}$.

Algorithm 2: Process of ALRM

Input : Per-round budget $\mathcal{B}/\mathcal{K}$; Target data $\mathcal{D}_t$; Labeled target data $\mathcal{D}_{tl}$
Output: selected set $\mathcal{D}_a$

Initialization selected set $\mathcal{D}_a = \emptyset$;
Obtain local representation LR_t of $\mathcal{D}_t$ with Eq. (3);
for $b \leftarrow 0$ **to** $\mathcal{B}/\mathcal{K}$ **do**
  $\mathcal{M} \leftarrow \mathcal{D}_{tl} \cup \mathcal{D}_a$;
  for $x_t \in \mathcal{D}_t$ **do**
    $\mathcal{M} \leftarrow \mathcal{M} \cup x_t$;
    Obtain local representation $LR_{\mathcal{M}}$ of $\mathcal{M}$ with Eq. (3);
    $\mathcal{Z}_t \leftarrow MMD^2(LR_t, LR_{\mathcal{M}})$;
    $\mathcal{M} \leftarrow \mathcal{M} \setminus x_t$;
  end
  $\mathcal{D}_a \leftarrow \arg\min_{x_t} \mathcal{Z}_t$;
end

3.5.1. Prototype-based domain contrastive alignment

Although supervised training of actively selecting UWF fundus images is an efficient way to improve the DR grading performance, the updating direction of the model gradient may not be conducive to the annotated knowledge transfer of color fundus images or even harmful. To avoid this, considering lack of color fundus images, we argue for a prototype contrastive loss function for explicit domain alignment. Specifically, the source features generated by G_g are to update a memory-bank as follows

$$\mathbf{h}_i \leftarrow \beta \mathbf{h}_i + (1-\beta)\mathbf{o}_i \tag{5}$$

where h_i and o_i denote the source class prototypes in the memory-bank and class features of the current step, and $\beta = 0.99$ is a hyper-parameter to control the ratio of class prototypes updating. The effect of the memory-bank is to ensure the stability of the source class prototype. Therefore, when the cross-entropy loss function

$$\mathcal{L}_{ce}^{tl} = \mathbb{E}_{(x,y)\sim D_{tl}} - \sum_{k=0}^{K} y_k log(C_y(G_e(x))) \tag{6}$$

is used for supervised training, the prototype contrastive loss can be used to align the features of the annotated UWF fundus images with the source class prototype in the memory-bank. The prototype contrastive loss is formed as

$$\mathcal{L}_{Alg}^{Inter} = -log \frac{exp(h_{tl} \cdot h^+/\tau)}{exp(h_{tl} \cdot h^+/\tau) + \sum_{n=1}^{P-1} exp(h_{tl} \cdot h^-/\tau)} \tag{7}$$

where h_{tl}, h^+, h^-, P, and τ denote the feature of x_{tl} extracting by G_e, source class prototype with the same label as x_{tl}, source class prototype with the different labels as x_{tl}, number classes of DR grades, and hyper-parameter.

3.5.2. Inter and intra domain consistency regularity

Inspired by consistency regularity in semi-supervised learning, we can further exploit unannotated UWF fundus images and propose to regularize the inter-domain consistency and intra-domain consistency. Following the previous literature [37], weak and strong data augmentations are applied to the unannotated UWF fundus image, obtaining two features h_{xu}^w and h_{xu}^s after extracting by source backbone G_e. The inter-domain consistency is to transport h_{xu}^w to all of the source class prototypes and get a soft probability. Then, the soft probability can be used as a supervised signal to regularize the model G with h_{xu}^s. This loss function can be formed as

$$\mathcal{L}_{Reg}^{Inter} = \mathcal{L}_{L1}(\sigma(\gamma\langle h_{xu}^w, p\rangle), G_y(h_{xu}^s)) \tag{8}$$

where γ, p, and $\mathcal{L}_{L1}$ denote optimal transport strategy, source class prototypes, and L1-Loss, respectively. The intra-domain consistency is to obtain the class probabilities of h_{xu}^w and h_{xu}^s after the source classifier G_y and a softmax layer, enforcing both class probabilities to have similar probabilistic output. This loss function can be implemented by

$$\mathcal{L}_{Reg}^{Intra} = p_{tu}^w \otimes p_{tu}^{s\top} - \mathrm{tr}(p_{tu}^w \otimes p_{tu}^{s\top}) \tag{9}$$

where p_{tu}^w and p_{tu}^s denote the class probabilities of the same UWF fundus image after weak data augmentation and strong data augmentation. With the help of inter-domain consistency regularity and intra-domain consistency regularity, more details that can be used for DR grading in UWF fundus images are captured by the model G.

3.5.3. Selecting pseudo label with mixup samples

We notice that the fine-grained problem and source-free setting make DA sensitive to the pseudo-labels. It is hard to find a method to deal with this issue after investigating a lot of literature. Fortunately, the annotated UWF fundus images selected by ALRS prompt us to propose a novel method to select pseudo label with mixup samples (S-PMiS). The proposed S-PMiS allows for better DR grading performance after active sample selection. More details, at the beginning of adaptation, the model G is trained with selected UWF fundus images augmented by mixup

$$\begin{cases} \lambda = B(\alpha, \beta) \\ x_{mix}^{tl} = \lambda \cdot x_i^{tl} + (1-\lambda) \cdot x_j^{tl} \\ y_{mix}^{tl} = \lambda \cdot y_i^{tl} + (1-\lambda) \cdot y_j^{tl} \end{cases} \tag{10}$$

where B refers to a *Beta* distribution. The mixup augmentation can make the model better able to distinguish whether a mixup sample is generated by two same label UWF fundus images, which can be optimized by a cross-entropy loss function

$$\mathcal{L}_{ce}^{mix} = \mathbb{E}_{(x,y)\sim D_{tl}} - \sum_{k=0}^{K} y_{mix}^{tl} log(C_y(G_e(x_{mix}^{tl}))) \tag{11}$$

When starting to adapt with a pseudo label, one of the samples x_r with the least similarity to the annotated UWF fundus image x_{tl} is selected to feed to model G, getting its corresponding class probability p_r and pseudo label y_r. The similarity between UWF fundus images is measured by a cosine metric based on local representation. All candidates that can be selected are formed as

$$\{x_r^1, x_r^2, \ldots, x_r^k\} = TopK(-\langle LR(x_{tl}), LR(\omega)\rangle) \tag{12}$$

where k and ω denote a hyper-parameter and a feature memory-bank of all UWF fundus images updated by an exponential moving average (EMA) [38], respectively. Notably, x_r is randomly selected among candidates. If the maximum p_r exceeds the threshold ϕ_a, x_r is pseudo-labeled and mixed up with an annotated UWF fundus image x_{tl} whose label is the same as the pseudo label of x_r to generate a mixup sample x_s, following as

$$x_s = 0.5 \cdot x_r^{pl} + 0.5 \cdot x_{tl}^{lb} \quad \text{s.t. } pl = lb \tag{13}$$

where pl and lb denote the pseudo label of x_r and annotation of x_{tl}. Then, x_s is fed to model G to obtain class probability p_s and pseudo label y_s (MiS PL). If the maximum p_s is exceeds the threshold ϕ_b and y_s is equal to y_r, the pseudo label of x_r is considered to be reliable. At the loop end of the dataloader, x_r with its pseudo label y_r will be added to $\mathcal{D}_{tl}$ as an annotated UWF fundus image (Add PL). After the first Add PL, all the UWF fundus images with pseudo labels will be re-checked multiple times by the MiS PL. At this time, if the predicted label of x_s by model G is different from its corresponding label of x_{tl}, the pseudo label y_r is considered to be unreliable, and x_r will be removed from $\mathcal{D}_{tl}$ before the next loop of the dataloader (Rev PL). More details are shown in Algorithm 3.

Algorithm 3: Process of S-PMiS

Input : Model $G = \{G_e, G_y\}$; Memory-bank ω; Annotated target data $\{x_{tl}, y_{tl}\}$; Unannotated target data $\{x_{tu}\}$; $\mathcal{D}_{rev}$; $\mathcal{D}_{add}$; $\mathcal{D}_{tl}$
Output: $\mathcal{D}_{rev}$; $\mathcal{D}_{add}$

$p_{tu} = softmax(G_y(G_e(x_{tu})))$, $y_{tu} = \arg\max p_{tu}$;
Obtain $LR(x_{tu})$ with Eq. (3) and update ω with EMA;
if $x_{tl} \in \mathcal{D}_{add}$ **then** // MiS PL
  Generate x_s with Eq. (13) and obtain pseudo label y_s by G;
  if $y_s \neq y_{tl}$ **then** // Rev PL
    $\mathcal{D}_{rev} \leftarrow \mathcal{D}_{rev} \cup \{x_{tl}, y_{tl}\}$;
  end
end
Obtain all candidates with Eq. (12) and randomly select x_r;
$p_r = softmax(G_y(G_e(X_r)))$, $y_r = \arg\max p_r$;
if $p_r[y_r] \geq \phi_a$ **then** // MiS PL
  Generate x_s with Eq. (13);
  $p_s = softmax(G_y(G_e(X_s)))$, $y_s = \arg\max p_s$;
  if $p_s[y_s] > \phi_b$ *and* $y_s = y_r$ **then** // Add PL
    $\mathcal{D}_{add} \leftarrow \mathcal{D}_{add} \cup \{x_r, y_r\}$;
  end
end

3.5.4. Overall loss function

Based on the above works, the proposed LPDA is optimized by

$$\mathcal{L} = \mathcal{L}_{ce}^{tl} + \alpha \cdot \mathcal{L}_{Alg}^{Inter} + \beta \cdot \mathcal{L}_{Reg}^{Inter} + \gamma \cdot \mathcal{L}_{Reg}^{Intra} + \kappa \cdot \mathcal{L}_{ce}^{mix} \tag{14}$$

where $\mathcal{L}_{ce}^{tl}$ is the supervised cross entropy loss (Eq. (6)), $\mathcal{L}_{Alg}^{Inter}$ is the prototype contrastive loss (Eq. (7)), $\mathcal{L}_{Reg}^{Inter}$ is the inter-domain consistency regularity loss (Eq. (8)), $\mathcal{L}_{Reg}^{Intra}$ is the intra-domain consistency regularity loss (Eq. (9)) and $\mathcal{L}_{ce}^{mix}$ is the supervised cross entropy loss (Eq. (11)) with mixup, respectively. α, β, γ, and κ denote hyper-parameters for different loss terms.

4. Results and discussion

4.1. Datasets and metrics

The public Eyepacs dataset and a private UWF fundus images dataset are employed as the source domain and target domain of our experiments, respectively. Considering overall efficiency and a realistic setup, 6000 color fundus images are randomly selected from Eyepacs to establish a new subset for the training of a source pre-trained model and other comparative methods while preserving the issue of data imbalance. The private dataset consists of 2816 UWF fundus images and has the same DR grade as Eyepacs (i.e., 906 normal, 553 mild, 720 moderate, 247 severe, and 390 proliferative). Meanwhile, to avoid a potential domain shift between the train dataset and test dataset, we follow previous DA practices [28,32], retaining all of the UWF fundus image for training and testing rather than k-fold cross-validation. For a comprehensive quantitative evaluation, overall and class-specific accuracy (ACC), F1-score, receiver operating characteristic (ROC), area under the curve (AUC), kappa, and quadratic weighted kappa (Q.W.Kappa) are reported in our experimental results.

4.2. Implementation details

The resolution of fundus images is resized to 512×512 for memory efficiency. Empirically, the same data augmentation techniques as used in the VisDA2017 [39] are applied to augment all of the fundus images, obtaining a resolution of 224×224 for computational efficiency. The source pre-trained model and the source feature generator are trained for 100 epochs and 1000 epochs, respectively, using the same mini-batch size of 64. The architecture of source feature generator is based on [28] and $\beta = [0.1, 0, -0.1, 0.2]$, $\gamma = 0.5$. The source pre-trained model adapts to UWF fundus images for 15 epochs with a mini-batch size of 32, including the first 10 epochs with ALRM and the last 5 epochs with S-PMiS. All of the experiments are optimized by Adadelta [40] with default settings except $lr = 0.1$. To obtain better performance, the backbone in the source pre-trained model adopt tiny Swim Transformer [41]. Meanwhile, the backbone of comparative methods is also replaced for fair comparison. The network of the source feature generator is referred to [42] which is initialized by random normal. Our experiments are run on the PyTorch platform with a single RTX3090 GPU. More details refer to our Github(https://github.com/JinyeRAN/source-free_active_domain_adaptation).

4.3. Overall evaluation

Since our SFADA mainly focuses on implementing a more realistic and superior method in DR grading, it is difficult to find an available baseline for improvement. Therefore, we modify LADA [32] to make it our baseline to study the effectiveness of our proposed methodology. In addition, due to the lack of a publicly available UWF benchmark dataset, we cannot directly utilize the reported domain adaptation performance of DR grading between color fundus images and UWF fundus images from existing literature. So, all the experimental results in this paper are reproduced by us on our dataset. We choose some methods that are similar to our research background for side-by-side comparison, including DUC [43], EADA [44], LADA [32], LADMA [34], SDM-AG [45], and ENLP [33]. All experiments are conducted with an active budget of 5%. The quantitative evaluation performance is presented in Table 1. The highest score is shown in bold font, while the second-highest score is underlined. The comparison results indicate that (1) Our SFADA achieves remarkable performance far beyond other methods in all five overall evaluation metrics, as well as highly competitive results in class-specific accuracy. Specifically, compared to baseline, the proposed method achieves performance improvements of 18.63% in Q.W.Kappa and 20.90% in ACC, both of which are crucial for DR diagnosis. (2) Moreover, compared to other DA methods that rely on source data for DR grading, our SFADA can solve the privacy of medical data to some extent. In contrast to other source-free domain adaptation approaches, SFADA obtains better DR grading performance. (3) As for computational efficiency, although it is difficult to give a quantitative evaluation due to fairness, considering that SFADA adopts one of the smallest fundus image resolutions in DR grading, employs the backbone with fewer parameters, and utilizes generating source domain features rather than generating source-like samples to transfer knowledge, we can make a qualitative conclusion that our approach achieves better computational efficiency than most DA-based DR grading methods.

It is worth noting that the DR grading performance of selected comparison methods is far from our expectation. Specifically, while

Table 1
Quantitative results of SFADA compared with other methods. The best performance is bold fonts and the second performance is underlined. All experiments are conducted with an active budget $\mathcal{B} = 5\%$.

Method	Source-free	Normal	Mild	Moderate	Severe	Proliferative	ACC	F1-Score	AUC	Kappa	Q.W.Kappa
DUC	✗	64.05	24.80	80.90	21.35	71.79	58.80	52.45	81.01	45.08	72.00
EADA	✗	69.24	26.91	**89.86**	67.42	83.33	68.60	65.45	91.05	58.72	75.00
LADA	✗	59.11	57.52	76.80	**84.27**	80.45	68.14	68.60	92.36	59.02	81.25
LADMA	✗	80.80	17.94	75.24	74.72	75.96	67.31	63.20	84.13	56.06	79.75
SDM-AG	✗	87.47	28.76	40.93	48.31	75.64	61.33	56.67	84.40	47.27	72.19
ENLP	✔	61.39	47.23	18.91	75.28	82.05	53.00	50.47	84.83	39.66	72.74
Baseline	✔	71.01	38.00	59.84	57.30	**91.67**	64.46	60.90	90.41	53.66	73.75
SFADA	✔	**93.54**	**74.14**	80.50	78.65	90.06	**85.36**	**83.71**	**95.70**	**80.54**	**92.38**

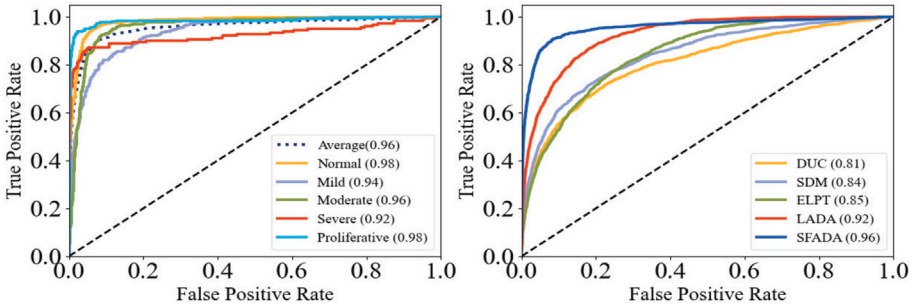


Fig. 3. The receiver operating characteristic curve (ROC) of model trained by SFADA and other methods. The left one is class-specific results of SFADA and the right one is ROC of SFADA and others.

both EADA and LAMDA achieve relatively high overall DR grading performance, they, like DUC and SDM-AG, suffer from issues with specific DR category. Having excluded external influencing factors and summarized the commonalities among these methods, we conclude that these sample-wise active sampling methods fail to address the presence of outliers in real-world fundus images, which is a key reason for the specific DR grading failure. LADA, which obeys a local-wise active sampling method, achieves more average grading performance across all DR categories, providing further evidence for our conclusion. However, local-wise and sample-wise methods each have advantages and disadvantages. The former can address outliers effectively but comes with higher computational costs and more hyperparameters. On the other hand, the latter are computationally efficient and can also produce good results on suitable application scenarios. As for ENLP, although it is also a local-wise active sampling method, a more powerful DA is required to adapt real-world fundus images under the setting of source-free. To further verify the performance of the adapted model on UWF fundus images, we plot the receiver operating characteristic and report the area under the curve, which is shown in Fig. 3. The left one is the class-specific results of SFADA, and the right one is the overall results compared with other methods. It is obvious that (1) the automatic DR grading model trained by SFADA can perform DR diagnosis at a high threshold and (2) is better than those trained by other DA methods.

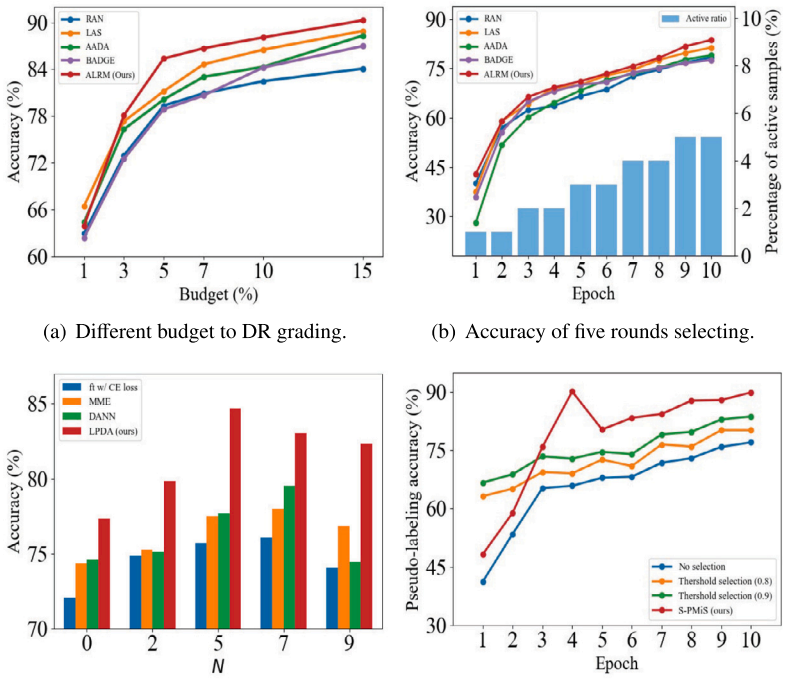
4.4. Specific discussion

Apart from the source feature generation stage, our SFADA performs active sample selection in the target domain and adaptation between source and target domains alternately. We wonder whether the proposed ALRM and LPDA are advanced in the DA tasks of DR grading. Therefore, we select some classical active sample selection methods such as Random, AADA [46], BADGE [47], LAS [32] and some domain adaptation methods such as ft w/ CE, DANN [35], RAA [32] from the related field for side-by-side comparison. The random and ft w/ CE mean randomly selecting some UWF fundus images as active samples and fine-tune the model only using active samples, respectively. LAS and RAA are the methods in our baseline. All the experimental settings are the same as quantitative evaluation. The results are presented in Table 2. The highest score is shown in bold font, while the second-highest score is underlined. Clearly, thanks to the consideration of outliers and the full utilization of active samples, the proposed ALRM and LPDA are more suitable for domain adaptation in DR grading than other methods. In particular, the proposed methods achieve the best performance on all the overall evaluation matrices.

Table 2
Quantitative results of LPDA and ALRM compared with other methods. The best performance is bold fonts and the second performance is underlined. All experiments are conducted with an active budget $\mathcal{B} = 5\%$.

AL Method	DA Method	ACC	F1_Score	AUC	QWK
Random		79.33	76.85	93.80	86.94
AADA		80.11	78.78	93.63	89.70
BADGE	LPDA	78.86	77.49	92.99	87.36
LAS		81.16	79.75	94.72	88.75
ALRM		**85.36**	**83.71**	**95.70**	**92.38**
	ft w/ CE loss	76.20	73.77	93.22	85.89
	DANN	78.78	75.72	94.66	84.40
ALRM	MME	78.55	75.93	93.97	84.97
	RAA	64.87	62.88	89.75	76.61
	LPDA	**85.36**	**83.71**	**95.70**	**92.38**

We conduct further study on the proposed ALRM, and the results refer to Fig. 4(a) - Fig. 4(c). The DR grading performance of different active budgets $\mathcal{B}$ is reported in Fig. 4(a), which shows that (1) our ALRM have a rapid DR grading performance improvement with increasing $\mathcal{B}$ when $\mathcal{B}$ is small. This is because too few active samples cannot effectively represent the entire dataset. (2) Starting from $\mathcal{B}$=3%, our ALRM achieves the best DR grading performance in the remaining experiments. Fig. 4(b) is the performance of five rounds of active selection with $\mathcal{B}$=5% and indicates that our ALRM obtains the best DR grading performance in every training epoch. We also discuss the effect of different local setting N of samples on the final DR grading performance in Fig. 4(c). It demonstrates that (1) a larger N in a certain range can make the proposed ALRM actively select more valuable UWF fundus images for labeling, which can effectively improve DR grading performance. However, (2) too large N will make the proposed ALRM ignore some high-value UWF fundus images, which is harmful to final DR grading performance. This may be caused by the definition of



(a) Different budget to DR grading. (b) Accuracy of five rounds selecting.

(c) The effect of N to DR grading. (d) Accuracy of pseudo label

Fig. 4. (a)–(c) Analysis of the impact of active sample selection on the final DR grading performance, where (a) reports the impact of different budgets, (b) reports the accuracy of train processing with a fixed budget $\mathcal{B}$=5%, and (c) reports the influence of different numbers of nearest neighbors N. (d) The influence of the proposed selecting pseudo label with mixup samples on the accuracy of pseudo label.

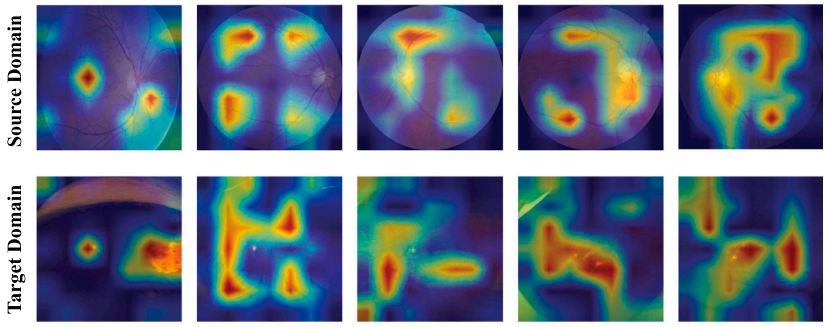


Fig. 5. Visualization of the regions attended by the source pre-trained model in the source domain and the regions attended by the adapted model in the target domain. The technique of visualization is Ablation-CAM.

local representation. In other words, the essence of local representation is the average of neighbors, and too large N will cause the local representation of each sample to converge, thus leading to unsuccessful matching.

To further investigate the reason why the proposed S-PMiS can improve rather than harm the DR grading performance, while making the results more significant, we modify the timing of pseudo-label training to report the accuracy of pseudo-labeling in the first 10 epochs and conduct another four experiments , which is shown in Fig. 4(d). Since our S-PMiS relies on the performance of model G, the pseudo-label accuracy generated by S-PMiS in the early stage increases rapidly with the improvement of model G performance. When the performance of model G reaches a certain level, the pseudo-label accuracy of S-PMiS consistently outperforms others

4.5. Visualization

It is a fact that there are more details in the UWF fundus images that can be employed for DR grading than that in the colors fundus images [21]. DA may focus on common knowledge to grade DR between colors fundus images and UWF fundus images, and ignore some UWF-specific information. Fortunately, active sampling with supervised training can alleviate this issue to some extent. As illustrated in Fig. 5, We visualize the feature map of the last LayerNorm in Transformer with the help of the Ablation-CAM [48]. (1) The interested regions focused by the source pre-trained model in the source domain color fundus images heatmap are similar to those focused by the adapted model in the target domain UWF fundus images heatmap, indicating that the meaningful knowledge is adapted to the DR grading of UWF fundus images. (2) Meanwhile, more interested regions in the heatmap of UWF fundus images than traditional color fundus images show that our SFADA captures more UWF-specific details for superior performance DR grading.

To visually demonstrate the relationship between all target domain samples and actively selected samples, we also visualize the active selection results of each round by t-SNE [49] technology, which is shown in Fig. 6. Due to the poor performance of the source pre-trained model in the UWF fundus images, the selection results of the first round are far from our expectations. With the progress of DA and the increase

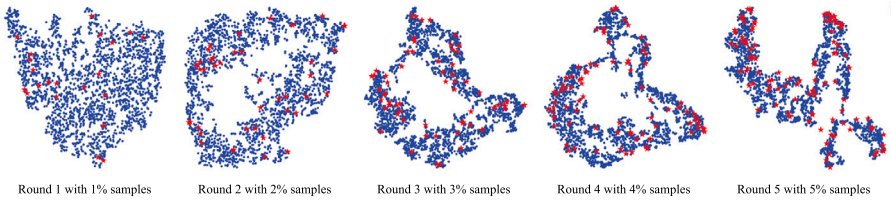


Fig. 6. Visualization of the relationship between the selected UWF fundus images (red) and all UWF fundus images (blue) in five round with 5% budget. Local representations of each sample are reduced the dimensionality by t-SNE.

Table 3
Ablation study on our proposed $\mathcal{L}_{chn}$ in the training stage of both G_e and G_g. The best performance is **highlighted** and the second performance is underlined. All experiments are conducted with an active budget of 5%.

Input resolution	MACs(G)	G_e		G_g		ACC	Kappa
		$\mathcal{L}_{ce}$	$\mathcal{L}_{chn}$	$\mathcal{L}_{ce}$	$\mathcal{L}_{chn}$		
448 × 448	17.96	✓		✓		**84.95**	**80.07**
224 × 224	4.49	✓		✓		76.10	68.80
224 × 224	4.49	✓	✓	✓		81.17	74.75
224 × 224	4.49	✓		✓	✓	79.83	73.06
224 × 224	4.49	✓	✓	✓	✓	84.81	79.91

Table 4
Ablation study on our proposed LPDA. The best performance is **highlighted** and the second performance is underlined. All experiments are conducted with an active budget of 5%.

$\mathcal{L}_{ce}^{tl}$	$\mathcal{L}_{Alg}^{Inter}$	$\mathcal{L}_{Reg}^{Inter}$	$\mathcal{L}_{Reg}^{Intra}$	S-PMiS	ACC	Kappa
✓					76.79	68.74
✓	✓				80.29	73.73
✓	✓	✓			83.98	78.47
✓	✓		✓		83.38	77.80
✓	✓	✓	✓		84.81	79.91
✓	✓	✓	✓	✓	**85.36**	**80.54**

in labeling, the subset composed of active samples can be well matched with the whole target domain in the space of local representation.

4.6. Ablation study

The purpose of $\mathcal{L}_{chn}$ is to mitigate the impact of input resolution reduction on DR grading performance. To further explore the role of $\mathcal{L}_{chn}$, we conducted five ablation experiments in Table 3, where the multiply-accumulate operations (MACs) is calculated in one forward propagation by calflops [50]. The results indicate that (1) reducing the input resolution can significantly reduce the demand for computational resources, but it can also seriously impair the DR grading performance. (2) The proposed $\mathcal{L}_{chn}$ can effectively mitigate this damage, and when used in training of G_e and G_g, $\mathcal{L}_{chn}$ can almost completely overcome this performance loss. However, there may be a discussion here about whether using $\mathcal{L}_{chn}$ in G_e is consistent with the source-free setting. We believe that the purpose of source-free setting is to protect data privacy. As long as the source domain data is not involved in the DA process, the limited requirements for the model are reasonable.

The LPDA is the key to successfully adapt the source pre-training model to UWF fundus images. To investigate the impact of each component in LPDA on the final DR grading performance, six ablation experiments are designed, and the quantitative evaluation results are shown in Table 4. It is clear that each of the components proposed in the LPDA contributes to improve the performance of DR grading. Specifically, the consistency regularity of unannotated UWF fundus images has the greatest improvement on the DR grading performance. Although the proposed S-PMiS improves DR grading performance less significantly than other proposed components in LPDA, it solves the problem that pseudo-labeling is harmful to the domain adaptation of DR grading in UWF fundus images, which can lead to a better performance DR grading model (for more details, refer to the next paragraph).

Table 5
Ablation study on our proposed S-PMiS. The best performance is **highlighted** and the second performance is underlined. All experiments are conducted with an active budget of 5%.

Mixup	Add PL	MiS PL	Rev PL	ACC	Kappa
				84.81	79.91
✓				84.39	79.24
✓	✓			84.43	79.26
✓	✓	✓		85.08	80.21
✓	✓	✓	✓	**85.36**	**80.54**
	✓	✓	✓	83.79	78.46

To further explore the necessity of various designs in S-PMiS, the results of six additional ablation experiments are presented in Table 5. The results indicate that (1) both pseudo-labeling and mixup compromise the DR grading performance. (2) Based on the training of mixup, the proposed Add PL, Mis PL, and Rev PL can gradually improve the DR grading performance. There is an insight to explain this issue. Although the mixup is harmful to DR grading performance, it can make the model more robust, which facilitates the identification of generated mixup samples and gives a higher accuracy pseudo label. The extra information introduced by the pseudo label can positively feed back into the training of the model, thus compensating for the performance degradation caused by robustness.

4.7. Potential limitation

Assisting ophthalmologists in DR Diagnosis, the DR grading model trained by our SFADA may face some potential limitations, as follows: (1) Different from the unsupervised method, our SFADA requires professional ophthalmologists to provide labeling information of selected samples during the training. Although the amount of labeling is very small, it still needs the intervention of professional ophthalmologists to work well. (2) The DR grading model trained by our SFADA cannot provide ophthalmologists with more detailed information about fundus lesions, such as category and location. Further analysis of UWF fundus images by professional ophthalmologists is still needed to make final conclusions.

5. Conclusion

This paper proposes a novel approach to achieve superior DR grading performance on unannotated UWF fundus images while considering data privacy and computational efficiency. Specifically, it mainly includes three parts: source feature generation, active local representation matching, and lesion-based prototype domain adaptation. Extensive experimental results show that our proposed method achieves the state-of-the-art DR grading performance and the necessity of each component. Furthermore, given the potential of the proposed framework to establish a powerful DR computer-aided diagnostic system with sustainable performance improvement, we are committed to conducting further research and exploration based on this foundation in the future, addressing more practical clinical ophthalmic issues.

CRediT authorship contribution statement

Jinye Ran: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Guanghua Zhang:** Data curation, Funding acquisition, Methodology, Resources, Supervision. **Fan Xia:** Formal analysis, Visualization. **Ximei Zhang:** Data curation, Formal analysis, Writing – review & editing. **Juan Xie:** Data curation, Formal analysis, Writing – review & editing. **Hao Zhang:** Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 21806131), Research Funds of Shanxi Transformation and Comprehensive Reform Demonstration Zone (No. 2018KJCX04), Shanxi 1331 Project, and Shanxi Province Natural Science Foundation (No. 202203021211006). This work was also partially sponsored by Shanxi Provincial Key Research and Development Project (No. 201903D311009).

References

[1] W. Wang, A.C. Lo, Diabetic retinopathy: pathophysiology and treatments, Int. J. Mol. Sci. 19 (6) (2018) 1816.
[2] F.I. Himasa, M. Singhal, A. Ojha, B. Kumar, Prospective for diagnosis and treatment of diabetic retinopathy, Curr. Pharm. Des. 28 (7) (2022) 560–569.
[3] J.L. Leasher, R.R. Bourne, S.R. Flaxman, J.B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, R.A. White, T.Y. Wong, et al., Global estimates on the number of people blind or visually impaired by diabetic retinopathy: a meta-analysis from 1990 to 2010, Diabetes Care 39 (9) (2016) 1643–1649.
[4] N.H. Cho, J. Shaw, S. Karuranga, Y. Huang, J. da Rocha Fernandes, A. Ohlrogge, B. Malanda, IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes Res. Clin. Pract. 138 (2018) 271–281.
[5] L. Wang, P. Gao, M. Zhang, Z. Huang, D. Zhang, Q. Deng, Y. Li, Z. Zhao, X. Qin, D. Jin, et al., Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013, JAMA 317 (24) (2017) 2515–2523.
[6] Y.P. Liu, Z. Li, C. Xu, J. Li, R. Liang, Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network, Artif. Intell. Med. 99 (2019) 101694.
[7] J. Echouffo-Tcheugui, M. Ali, G. Roglic, R. Hayward, K. Narayan, Screening intervals for diabetic retinopathy and incidence of visual loss: a systematic review, Diabetic Med. 30 (11) (2013) 1272–1292.
[8] T.H. Tung, S.J. Chen, H.C. Shih, P. Chou, A.F. Li, M.P. Shyong, F.L. Lee, J.H. Liu, Assessing the natural course of diabetic retinopathy: a population-based study in Kinmen, Taiwan, Ophthalmic Epidemiol. 13 (5) (2006) 327–333.
[9] S. Stolte, R. Fang, A survey on medical image analysis in diabetic retinopathy, Med. Image Anal. 64 (2020) 101742.
[10] Z. Yang, T.E. Tan, Y. Shao, T.Y. Wong, X. Li, Classification of diabetic retinopathy: Past, present and future, Front. Endocrinol. 13 (2022) 1079217.
[11] M.Z. Atwany, A.H. Sahyoun, M. Yaqub, Deep learning techniques for diabetic retinopathy classification: A survey, IEEE Access 10 (2022) 28642–28655.
[12] P. Ansari, N. Tabasumma, N.N. Snigdha, N.H. Siam, R.V. Panduru, S. Azam, J. Hannan, Y.H. Abdel-Wahab, Diabetic retinopathy: an overview on mechanisms, pathophysiology and pharmacotherapy, Diabetology 3 (1) (2022) 159–175.
[13] W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: A review, Inform. Med. Unlocked 20 (2020) 100377.
[14] M.D. Abràmoff, J.M. Reinhardt, S.R. Russell, J.C. Folk, V.B. Mahajan, M. Niemeijer, G. Quellec, Automated early detection of diabetic retinopathy, Ophthalmology 117 (6) (2010) 1147–1154.
[15] K. Oh, H.M. Kang, D. Leem, H. Lee, K.Y. Seo, S. Yoon, Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, Sci. Rep. 11 (1) (2021) 1897.
[16] Diabetic Retinopathy Study Research Group, et al., Diabetic retinopathy study report number 6. Design, methods, and baseline results. Report number 7. a modification of the airlie house classification of diabetic retinopathy. Prepared by the diabetic retinopathy, Invest. Ophthalmol. Vis. Sci. 21 (1981) 1–226.
[17] Y. Matsui, A. Ichio, A. Sugawara, E. Uchiyama, H. Suimon, H. Matsubara, M. Sugimoto, K. Ikesugi, M. Kondo, et al., Comparisons of effective fields of two ultra-widefield ophthalmoscopes, optos 200Tx and clarus 500, BioMed Res. Int. 2019 (2019).
[18] K.G. Falavarjani, K. Wang, J. Khadamy, S.R. Sadda, Ultra-wide-field imaging in diabetic retinopathy; an overview, J. Curr. Ophthalmol. 28 (2) (2016) 57–60.
[19] E. Gupta, V. Gupta, M. Chopra, P.C. Chhipa, M. Liwicki, Learning self-supervised representations for label efficient cross-domain knowledge transfer on diabetic retinopathy fundus images, in: 2023 International Joint Conference on Neural Networks, IJCNN, 2023, pp. 1–7.
[20] Q. Wei, J. Yang, B. Wang, J. Wang, J. Zhao, X. Zhao, S. Yang, N. Manivannan, Y. Chen, D. Ding, X. Li, Cross-domain collaborative learning for recognizing multiple retinal diseases from wide-field fundus images, 2023, arXiv preprint arXiv:2305.08078.
[21] L. Ju, X. Wang, Q. Zhou, H. Zhu, M. Harandi, P. Bonnington, T. Drummond, Z. Ge, Bridge the domain gap between ultra-wide-field and traditional fundus images via adversarial domain adaptation, 2020, arXiv preprint arXiv:2003.10042.
[22] Y. Bai, J. Hao, H. Fu, Y. Hu, X. Ge, J. Liu, Y. Zhao, J. Zhang, Unsupervised lesion-aware transfer learning for diabetic retinopathy grading in ultra-wide-field fundus photography, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022, pp. 560–570.
[23] C. Zhang, T. Lei, P. Chen, Diabetic retinopathy grading by a source-free transfer learning approach, Biomed. Signal Process. Control 73 (2022) 103423.
[24] C. Zhou, W. Zhang, H. Chen, L. Chen, Domain adaptation for medical image classification without source data, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2022, pp. 2224–2230.
[25] M. Pourreza, Open-set source-free domain adaptation in fundus images analysis, 2023, Electronic Thesis and Dissertation Repository, URL: https://ir.lib.uwo.ca/etd/9241.
[26] I. Qureshi, J. Ma, Q. Abbas, Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning, Multimedia Tools Appl. 80 (2021) 11691–11721.
[27] M.A. Ahsan, A. Qayyum, A. Razi, J. Qadir, An active learning method for diabetic retinopathy classification with uncertainty quantification, Med. Biol. Eng. Comput. 60 (10) (2022) 2797–2811.
[28] Z. Qiu, Y. Zhang, H. Lin, S. Niu, Y. Liu, Q. Du, M. Tan, Source-free domain adaptation via avatar prototype generation and adaptation, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, 2021, pp. 2921–2927.
[29] M.I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, D. Hjelm, Mutual information neural estimation, in: Proceedings of the 35th International Conference on Machine Learning, Vol. 80, 2018, pp. 531–540.
[30] A. Sordoni, N. Dziri, H. Schulz, G. Gordon, P. Bachman, R.T. Des Combes, Decomposed mutual information estimation for contrastive representation learning, in: International Conference on Machine Learning, PMLR, 2021, pp. 9859–9869.
[31] Y. Lu, G. Zhang, S. Sun, H. Guo, Y. Yu, f-MICL: Understanding and generalizing InfoNCE-based contrastive learning, Trans. Mach. Learn. Res. (2023).
[32] T. Sun, C. Lu, H. Ling, Local context-aware active domain adaptation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2023, pp. 18634–18643.
[33] X. Li, Z. Du, J. Li, L. Zhu, K. Lu, Source-free active domain adaptation via energy-based locality preserving transfer, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 5802–5810.
[34] S. Hwang, S. Lee, S. Kim, J. Ok, S. Kwak, Combating label distribution shift for active domain adaptation, in: European Conference on Computer Vision, 2022, pp. 549–566.
[35] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (1) (2016) 2030–2096.
[36] K. Saito, D. Kim, S. Sclaroff, T. Darrell, K. Saenko, Semi-supervised domain adaptation via minimax entropy, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8050–8058.
[37] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C.A. Raffel, E.D. Cubuk, A. Kurakin, C.L. Li, Fixmatch: Simplifying semi-supervised learning with consistency and confidence, Adv. Neural Inf. Process. Syst. 33 (2020) 596–608.
[38] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017).
[39] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, K. Saenko, VisDA: The visual domain adaptation challenge, 2017, arXiv preprint arXiv:1710.06924.
[40] M.D. Zeiler, Adadelta: an adaptive learning rate method, 2012, arXiv preprint arXiv:1212.5701.
[41] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[42] S. Xu, H. Li, B. Zhuang, J. Liu, J. Cao, C. Liang, M. Tan, Generative low-bitwidth data free quantization, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16, 2020, pp. 1–17.

[43] M. Xie, S. Li, R. Zhang, C.H. Liu, Dirichlet-based uncertainty calibration for active domain adaptation, 2023, arXiv preprint arXiv:2302.13824.
[44] B. Xie, L. Yuan, S. Li, C.H. Liu, X. Cheng, G. Wang, Active learning for domain adaptation: An energy-based approach, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 8, 2022, pp. 8708–8716.
[45] M. Xie, Y. Li, Y. Wang, Z. Luo, Z. Gan, Z. Sun, M. Chi, C. Wang, P. Wang, Learning distinctive margin toward active domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7993–8002.
[46] J.C. Su, Y.H. Tsai, K. Sohn, B. Liu, S. Maji, M. Chandraker, Active adversarial domain adaptation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 739–748.
[47] J.T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, A. Agarwal, Deep batch active learning by diverse, uncertain gradient lower bounds, 2019, arXiv preprint arXiv:1906.03671.
[48] S. Desai, H.G. Ramaswamy, Ablation-CAM: Visual explanations for deep convolutional network via gradient-free localization, in: 2020 IEEE Winter Conference on Applications of Computer Vision, WACV, 2020, pp. 972–980.
[49] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).
[50] X. Ye, calflops: a FLOPs and params calculate tool for neural networks in pytorch framework, 2023, URL: https://github.com/MrYxJ/calculate-flops.pytorch.