# DOMAIN ADAPTATIVE RETINAL IMAGE QUALITY ASSESSMENT WITH KNOWLEDGE DISTILLATION USING COMPETITIVE TEACHER-STUDENT NETWORK

*Yuanming Lin*[1,2], *Heng Li*[1,*], *Haofeng Liu*[1,2], *Hai Shu*[3], *Zinan Li*[2], *Yan Hu*[2,*], *Jiang Liu*[1,2]

[1] Research Institute of Trustworthy Autonomous Systems,
Southern University of Science and Technology, Shenzhen 518055, China;
[2] Department of Computer Science and Engineering,
Southern University of Science and Technology, Shenzhen 518055, China;
[3] Department of Biostatistics, School of Global Public Health, New York University, New York, USA

## ABSTRACT

Retinal image quality assessment (RIQA) is essential for retinal examinations, as it impacts the certainty of both manual and intelligent diagnosis. Unfortunately, domain shifts, such as the variance of colors and illumination, are prone to confuse RIQA. Though efficient domain adaptation solutions have been proposed, properly transferring RIQA models to new domains remains a troublesome task. This paper presents a domain adaptative RIQA algorithm with knowledge distillation using competitive teacher-student network (CTSN) to address the above issue. The main structure consists of a teacher network, a student network, and a competition module. The teacher network provides pseudo-labels by adapting source and target domain features, and the student network learns features from target-specific pseudo-labels. The competition module boosts the fine-grained adaptation of RIQA. Comparison experiments and ablation studies demonstrate that our method performs outstandingly in RIQA with domain shifts.

***Index Terms—*** Retinal image quality assessment, domain adaptation, knowledge distillation

## 1. INTRODUCTION

Fundus images are frequently utilized for the early detection and diagnosis of many eye illnesses, such as age-related macular degeneration (AMD), glaucoma, and diabetic retinopathy (DR) [1, 2]. In a screening study, Philip et al. [3] showed that, due to a lack of adequate quality, 12% of fundus images from 5,575 sequential patients could not be suited for fundus illness diagnosis by ophthalmologists. Moreover, around 30% of retinal images were not of sufficiently high quality for accurate diagnosis, which has been found in a study based on UK BioBank [4]. Therefore, we need to focus on RIQA to ensure the automatic diagnosis system or ophthalmologists get credible fundus images before fundus diagnosis.

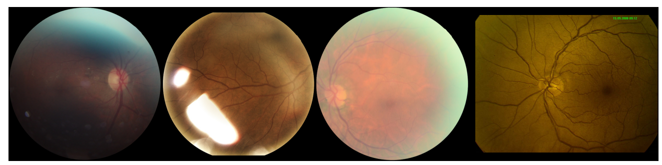

**Fig. 1**. Due to the presence of illumination and artifact, the quality of the images has an influence on the diagnosis. The first three images are from EyeQ [5] and the last one comes from DRIMDB [6].

Recently, several methods used convolutional neural networks (CNNs) for RIQA [5, 7]. Fu et al. [5] took a combination of retinal image features among HSV, RGB, and LAB color spaces to predict the quality of fundus images. Xu et al. [7] utilized blood vessel segmentation and detection of the optic disc region to forecast the quality of fundus images. However, these researchers did not consider domain adaptation in the RIQA task. In a recent study, Liu et al. [8] found that image blurring causes a significant decline in diagnosis performance. As shown in Fig. 1, different fundus image datasets were built by a variety of levels of experience people and various fundus cameras under different imaging conditions, which may result in different distributions of data causing the domain shift problem. Existing domain adaptation methods used the discriminator to align the common features between the source domain and the target domain, such as [9, 10, 11]. However, existing conventional domain adaptation approaches are not suitable for RIQA due to the failure of fine-grained classification. Hence, there are two challenges in the assessment of retinal quality: one is domain adaptation, and the other is failure to fine-grained classification.

A novel RIQA algorithm called CTSN is proposed to address the above challenges. We design a teacher-student competition mechanism for domain adaptation, which comprises the teacher and student networks, and the competition module. The teacher network first extracts the common features between the source and target domain, and then the student

∗ Correspondence: lih3, huy3@sustech.edu.cn

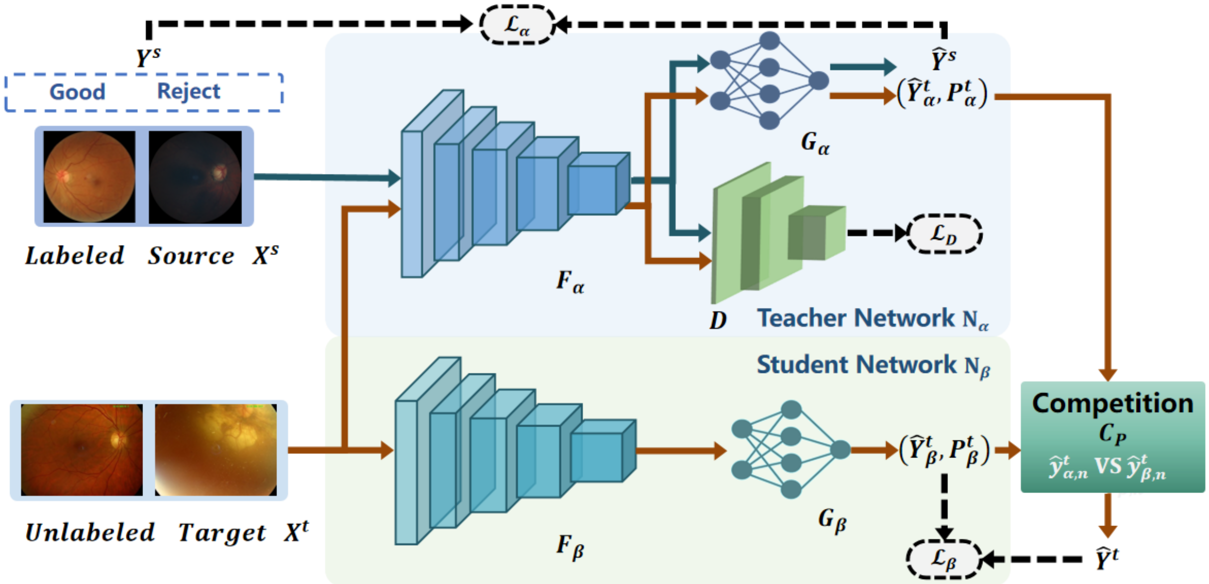


**Fig. 2**. The structure of CTSN, where the teacher network $N_\alpha$ (blue) and the student network $N_\beta$ (light green) feed their pseudo-labels and the corresponding probabilities ($Y^t_\alpha$ , $P^t_\alpha$) and ($Y^t_\beta$ , $P^t_\beta$) into the competition module (deep green) to choose the more reliable pseudo-labels $\hat{Y}^t$ for the student network training. $P^t_\alpha$ , and $P^t_\beta$ are the probability of $Y^t_\alpha$ , $Y^t_\beta$ respectively. Firstly, $N_\alpha$ extracts their common features to provide reliable pseudo-labels. Secondly, $N_\beta$ learns the features of the target domain. Finally, $C_P$ chooses the more credible pseudo-labels to alleviate the source-bias problem.

network learns the characteristics of the target domain. Finally, the competition module selects more credible pseudo-labels, which could break through the source bias issue to achieve fine-grained classification. The main contributions of the paper are as follows:

- A domain adaptative method is proposed to alleviate the domain shift problem in RIQA.
- we design a knowledge distillation network for domain adaptation to compete for pseudo-labels by introducing the teacher network, the student network, and a competition mechanism, leading to select more reliable target-specific labels.
- Comparison experiments and ablation study demonstrate that our method has a relatively good performance on domain adaptation in RIQA.

## 2. METHODOLOGY

The overall structure of our method is shown in Fig. 2, which includes three parts, the teacher network $N_\alpha$, the student network $N_\beta$, and the competition model $C_P$. $N_\alpha$ sends ($\hat{Y}^t_\alpha$,$P^t_\alpha$) and $N_\beta$ sends ($\hat{Y}^t_\beta$,$P^t_\beta$) to $C_P$ respectively, and then $C_P$ based on competition mechanism chooses more credible pseudo-labels $\hat{Y}_t$ to $N_\alpha$.

### 2.1. Teacher Network

The teacher network is a global unsupervised domain adaptation model, which globally aligns the features extracted from the network across the source and target domains, without any labeled target data. Besides, the teacher network also can be treated as an excellent leader to guide the learning of the student network by providing RIQA results to the student network as references.

The framework of the teacher network is displayed in Fig. 2, the objective function of which can be formulated as:

$$\min_{F_\alpha,G_\alpha} \max_{D} \mathcal{L}_\alpha(F_\alpha, G_\alpha) - \mathcal{L}_D(F_\alpha, D), \tag{1}$$

where $F_\alpha$ is the feature extraction network, $G_\alpha$ is the classification network and $D$ is the discriminator. With the guidance of the discriminate network $D$, the feature extraction network $F_\alpha$ can extract the common feature between source data and target data. The teacher network can exploit an unsupervised domain adaptation method named DANN [10], whose loss functions are listed as follows:

$$\mathcal{L}_\alpha(F_\alpha, G_\alpha) = \mathbb{E}\left[\gamma(G_\alpha(F_\alpha(x^s), y^s)\right], \tag{2}$$

$$\begin{aligned}\mathcal{L}_D(F_\alpha, D) = &-\mathbb{E}\left[\log[D(F_\alpha(x^s)]\right] \\ &- \mathbb{E}\left[\log[1 - D(F_\alpha(x^t)]\right],\end{aligned} \tag{3}$$

where $L_{\alpha}(F_{\alpha}, G_{\alpha})$ applies cross-entropy loss $\gamma$ to reduce the source classification error, and $L_D(F_{\alpha}, D)$ decreases the distance between source samples and target samples by handling their distribution.

Both the teacher network and the student network are not pretrained, so their classification accuracy is low at the beginning. As the training progresses, the teacher network provides more and more correct pseudo-labels to the student network with the help of the true labels of the source domain that is close to the target domain.

### 2.2. Student Network

However, the teacher network can not eliminate source domain dependencies and achieve class-level feature alignment. Therefore, we introduce knowledge distillation by adding a student network, a simple classification network with the same structure as the teacher network, so as to pay more attention to the features of the target domain and class-level feature alignment between the source and target domain. At the training process, the student network balances the preferences of pseudo-labels $\hat{y^t}$ by learning the features of the target domain. As the student network learns more and more characteristics of the target domain, the pseudo-label it provides becomes target-specific.

In our proposed CTSN, the student network puts its predictions and the corresponding probabilities ($\hat{Y_{\beta}^{t}}$ , $P_{\beta}^{t}$) into the competition module, which is crucial to avoid the influence of some incorrect pseudo-labels from the teacher network. The student network can achieve class-level alignment of the source and target domains, When the final pseudo-label is determined by $\hat{Y_{\beta}^{t}}$ of the student network instead of $\hat{Y_{\alpha}^{t}}$ from the teacher network. The loss function of the student network is written as:

$$\mathcal{L}_{\beta}(F_{\beta}, G_{\beta}) = \mathbb{E}\left[\gamma(G_{\beta}(F_{\beta}(x^t), \hat{y^t})\right], \tag{4}$$

where $\hat{y^t}$ is the prediction of the target domain.

### 2.3. Competition Module

In the proposed CTSN framework, the teacher network and the student network predict the results of the target domain samples to compete for the pseudo-label. The output pseudo-labels $\hat{y_t}$ of the competition module are listed as follows:

$$\hat{y_t} = \begin{cases} \hat{y_{\alpha}^{t}}, & p_{\alpha}^{t} > p_{\beta}^{t} \\ \hat{y_{\beta}^{t}}, & \text{otherwise} \end{cases} . \tag{5}$$

However, since both the teacher and student networks are trained from scratch, the pseudo-labels provided by the two networks during the initial training phase are unreliable. As the training progresses, the teacher network performs better than the student network, and has a certain generalization ability to provide slightly more reliable pseudo-labels, with the help of the source domain labels. In the subsequent training process, the student network gradually learns the characteristics of the target domain and needs to master the initiative to compete for pseudo-labels to obtain more reliable pseudo-labels. Hence, we introduce $T_p$ , a probability threshold, to provide priority for selecting pseudo-labels with the student network in a probability interval [$T_p$,1], which is formulated as:

$$T_p = \frac{1}{1 + \exp(-\alpha p)}, \tag{6}$$

where $\alpha$ is a hyper-parameter, and $\alpha$ is set to 10. $p$ represents the training process, and it increases as the training proceeds. $\alpha$ increases to 1 from 0 and $T_p$ increases to 1 from 0.5, as the training continues. At the same time, the pseudo-labels provided by the competition module will become more and more reliable. We combine $T_p$ with competition module to get the following new competition equation :

$$\hat{y_t} = \begin{cases} \hat{y_{\alpha}^{t}}, & \text{if } p_{\alpha}^{t} > p_{\beta}^{t} \text{ or } p_{\alpha}^{t} > T_p \\ \hat{y_{\beta}^{t}}, & \text{otherwise} \end{cases} . \tag{7}$$

The teacher network imparts the knowledge to the student network, when the pseudo-label selects $\hat{Y_{\alpha}^{t}}$ as the prediction result. Since the teacher network is trained under the supervision of source domain labels related to the target domain, the guidance of the teacher network is meaningful. However, the inherent source-domain dependency error negatively affects the student network at the same time. So it is significant to choose $\hat{Y_{\beta}^{t}}$ as supervision when it is more reliable than $\hat{Y_{\alpha}^{t}}$ in later training phase, which will introduce the knowledge of target domain to address the source bias problem.

## 3. EXPERIMENTS AND EVALUATION

### 3.1. Datasets and Settings

$1)Datasets$ : Our method was evaluated on two RIQA datasets: Eye-Quality (EyeQ) dataset [5] and diabetic retinopathy image database (DRIMDB) [6]. EyeQ comprises 28,792 retinal images, which uses a three-level quality grading system (i.e., 'Good', 'Usable' and 'Reject'). DRIMDB has 216 fundus images which are classified to three classes: good, bad and outlier. To guarantee experiment results immune to the numbers of both datasets, we selected the same number of images from EyeQ and DRIMDB respectively. Moreover, to construct a binary classification task, we selected 69 'bad' images and 125 'good' images from DRIMDB, 69 'reject' images, and 125 'good' images from EyeQ. The experiment followed the Declaration of Helsinki and obtained approval from the local ethics committee.

$2)Implementation\ Details$ : We follow the standard protocols in unsupervised domain adaptation, using labeled source domain data and unlabeled target data domain at the

**Table 1**. Ablation study of the proposed method.

| $D$ | $N_\beta$ | $C_P$ | EyeQ → DRIMDB | | DRIMDB → EyeQ | |
|---|---|---|---|---|---|---|
| | | | Acc | F1 | Acc | F1 |
| | | | 0.701 | 0.568 | 0.758 | 0.431 |
| √ | | | 0.711 | 0.595 | 0.768 | 0.474 |
| √ | √ | | 0.742 | 0.659 | 0.820 | 0.684 |
| √ | √ | √ | **0.825** | **0.791** | **0.902** | **0.868** |

training phase. At first, we took EyeQ as the source domain and DRIMDB as the target domain and then we reversed them. The epochs of all the experiment were set to 100. Our method was train with Stochastic Gradient Descent optimizer implementing on Pytorch, a minibatch size of 4, and learning rate is 0.01. In our CTSN, the teacher network and the student network compete for pseudo-labels $\hat{y}^t$ every 10 epochs. As epoch increases, $p$ increases linearly from 0 to 1. And we selected accuracy (Acc), and F1 score as metrics to measure performance.

### 3.2. Ablation Study

The ablation study was conducted to investigate how the modules contribute to the network generalization ability in RIQA. As shown in Table 1, our proposed model was trained using EyeQ as the source domain and DRIMDB as the target domain, and then we selected EyeQ as the target domain and DRIMDB as the source domain. The baseline of the RIQA model was trained with a base network without the discriminator, the student network, and the competition module. Then, the discriminator $D$ for aligning the features of the source and target domains, the student network $N_\beta$ and competition module $C_P$ were added to the RIQA model gradually.

With the facilitation of the discriminator, the model extracted to the domain-invariant feature, which improved the domain generalization ability. After introducing $N_\beta$, the model payed more attention to the features of the target domain and obtains better generalization ability. Finally, when $C_P$ was added, the model can achieve fine-grained feature alignment, and thus, class-level feature alignment, and the generalization ability of the model was further enhanced.

### 3.3. Comparison with Previous Methods

We conducted comparative experiments to demonstrate the usefulness of our proposed method. We compared our method with existing fundus image quality assessment methods and current domain adaptation methods, respectively in table 2: multiple color-space fusion network (MCF) [5], domain-invariant interpretable fundus image quality assessment (WLDFIQA) [12], deep subdomain adaptation network (DSAN) [13], adversarial discriminative domain adaptation (ADDA) [9], cycle-consistent adversarial domain adaptation (CyCADA) [11], batch nuclear-norm maximization (BNM) [14] and domain adversarial neural network (DANN) [10].

**Table 2**. Comparison on cross-domain classification between EyeQ and DRIMDB.

| Algorithms | EyeQ → DRIMDB | | DRIMDB → EyeQ | |
|---|---|---|---|---|
| | Acc | F1 | Acc | F1 |
| MCF [5] | 0.701 | 0.568 | 0.758 | 0.431 |
| WLDFIQA [12] | 0.769 | 0.772 | 0.732 | 0.788 |
| DSAN [13] | 0.706 | 0.711 | 0.794 | 0.847 |
| ADDA [9] | 0.799 | 0.779 | 0.711 | 0.693 |
| CyCADA [11] | 0.753 | 0.672 | 0.778 | 0.593 |
| BNM [14] | 0.742 | 0.781 | 0.732 | 0.785 |
| DAAN [10] | 0.753 | 0.789 | 0.768 | 0.828 |
| CTSN (ours) | **0.825** | **0.791** | **0.902** | **0.868** |

Table 2 shows the cross-domain classification results which demonstrate that our proposed method achieves the best performance. It can be observed that the evaluation metric F1 of MCF was not high due to the variance of colors and illumination across domains. Compared with MCF considering multiple color spaces (i.e., HSV, RGB and LAB), our method mitigates the effects of domain shifts in the task of fundus image quality assessment. Though WLDFIQA takes domain adaptation into account, they are still underperforming on generalization in RIQA. When CTSN is compared with domain adaptation methods which can only use an adversarial loss to align global features, our proposed method can achieve intra-domain category alignment through knowledge distillation using competitive teacher-student network (CTSN). Since DSAN can align subdomain features between the source domain and target domain, it can be observed that the performance of DSAN is not bad when we choose DRIMDB as the source domain and EyeQ as the target domain. However, intra-domain categories are more fine-grained features than sub-domain features, so our proposed method performs better in domain adaptation for classification tasks.

## 4. CONCLUSIONS

To mitigate the domain shifts in RIQA tasks, this paper proposes a domain adaptative network, named CTSN, based on the knowledge distillation with a teacher-student competition. Pseudo-labels of target data are firstly inferred by the teacher network $N_\alpha$ of CTSN using unsupervised domain adaptation, and thereby a target-specific model is then learned by the student network $N_\beta$. Moreover, a competition mechanism is designed between the teacher and student networks to achieve a fine-grained feature alignment across domains. Experiments are implemented with EyeQ and DRIMDB, which demonstrate the outstanding performance of CTSN in RIQA with domain shifts.

## Acknowledgment

This work was supported in part by Basic and Applied Fundamental Research Foundation of Guangdong Province (2020A1515110286), The National Natural Science Foundation of China (8210072776), Guangdong Provincial Department of Education (2020ZDZX3043), Guangdong Provincial Key Laboratory (2020B121201001), and Shenzhen Natural Science Fund (JCYJ20200109140820699, 20200925174052004).

## 5. REFERENCES

[1] Heng Li, Haofeng Liu, Yan Hu, Huazhu Fu, Yitian Zhao, Hanpei Miao, and Jiang Liu, "An annotation-free restoration network for cataractous fundus images," *IEEE Transactions on Medical Imaging*, 2022.

[2] Heng Li, Haofeng Liu, Huazhu Fu, Hai Shu, Yitian Zhao, Xiaoling Luo, Yan Hu, and Jiang Liu, "Structure-consistent restoration network for cataract fundus image enhancement," in *MICCAI*. Springer, 2022, pp. 487–496.

[3] Sam Philip, LM Cowie, and JA Olson, "The impact of the health technology board for scotland's grading model on referrals to ophthalmology services," *British Journal of Ophthalmology*, vol. 89, no. 7, pp. 891–896, 2005.

[4] Thomas J MacGillivray, James R Cameron, Qiuli Zhang, Ahmed El-Medany, Carl Mulholland, Ziyan Sheng, Bal Dhillon, Fergus N Doubal, Paul J Foster, Emmanuel Trucco, et al., "Suitability of uk biobank retinal images for automatic analysis of morphometric properties of the vasculature," *PLoS One*, vol. 10, no. 5, pp. e0127914, 2015.

[5] Huazhu Fu, Boyang Wang, Jianbing Shen, Shanshan Cui, Yanwu Xu, Jiang Liu, and Ling Shao, "Evaluation of retinal image quality assessment networks in different color-spaces," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 48–56.

[6] Ugur Sevik, Cemal Kose, Tolga Berber, and Hidayet Erdol, "Identification of suitable fundus images using automated quality assessment methods," *Journal of biomedical optics*, vol. 19, no. 4, pp. 046006, 2014.

[7] Ziwen Xu, Qing Liu, et al., "A deep retinal image quality assessment network with salient structure priors," *arXiv preprint arXiv:2012.15575*, 2020.

[8] Haofeng Liu, Haojin Li, Xiaoxuan Wang, Heng Li, Mingyang Ou, Luoying Hao, Yan Hu, and Jiang Liu, "Understanding how fundus image quality degradation affects cnn-based diagnosis," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 438–442.

[9] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.

[10] Chaohui Yu, Jindong Wang, Yiqiang Chen, and Meiyu Huang, "Transfer learning with dynamic adversarial adaptation network," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 778–786.

[11] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*. Pmlr, 2018, pp. 1989–1998.

[12] Yaxin Shen, Bin Sheng, Ruogu Fang, Huating Li, Ling Dai, Skylar Stolte, Jing Qin, Weiping Jia, and Dinggang Shen, "Domain-invariant interpretable fundus image quality assessment," *Medical image analysis*, vol. 61, pp. 101654, 2020.

[13] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He, "Deep subdomain adaptation network for image classification," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 4, pp. 1713–1722, 2020.

[14] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Qi Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3941–3950.