

# Efficient Fundus Image Gradeability Approach Based on Deep Reconstruction-Classification Network

Saif KHALID <sup>a,b,1</sup>, Saddam ABDULWAHAB <sup>a</sup> Hatem A. RASHWAN <sup>a</sup>  
 Julián CRISTIANO <sup>a</sup> Mohamed ABDEL-NASSER <sup>a,c</sup> and Domenech PUIG <sup>a</sup>

<sup>a</sup>*DEIM, Universitat Rovira i Virgili, Tarragona, Spain*

<sup>b</sup>*University of Al-Qadisiyah, Iraq*

<sup>c</sup>*Department of Electrical Engineering, Aswan University, 81528 Aswan, Egypt*

## Abstract.

Quality of retinal image is vital for screening of ailments pertaining to eye such as glaucoma, diabetic retinopathy (DR) and age related macular degeneration. Therefore, assessing quality of retinal image prior to any kind of diagnosis has assumed significance in Computer Aided Design (CAD) applications. The rationale behind this is that reliability of retinal image is to be guaranteed to have dependable diagnosis. In this paper, we propose a novel retinal fundus image quality assessment (RIQA) method based on autoencoder network to assess retinal images if the image is acceptable for screening or not. The autoencoder network architecture is well suited to precisely to properly represent the key features of the image quality, especially when the network can correctly reconstruct the input image. The proposed model consists of encoder and decoder successive networks. The encoder will be used for representing the features of the input image. In turn, the decoder will be used for reconstruct the input image. The features get from encoder network will then be fed to a classifier in order to classify the quality of retinal image to two classes: gradable or ungradable. The experimental results revealed more useful assessment and the proposed deep model provides a superior performance for RIQA. Thus, our model can serve real-world Clinical Decision Support Systems in the healthcare domain.

**Keywords.** Retinal Image, Quality Assessment, Autoencoder Network, Ocular Diseases, Deep Learning

## 1. Introduction

Retinal diseases are on the rise due to the increase in the diabetic population and increased life expectancy. A good example of the effect of age on retinal health is age-related macular degeneration (AMD). The number of people having AMD disease is estimated to be 196 million in 2020, and is expected in next twenty years will increase to 288 million. In addition, Glaucoma is a progressive optic neuropathy that causes blindness in industrialized countries, and ocular hypertension is the main risk factor for glaucoma [1].

<sup>1</sup>Corresponding Author: Saif khalid. E-mail: saif.khalid@qu.edu.iq

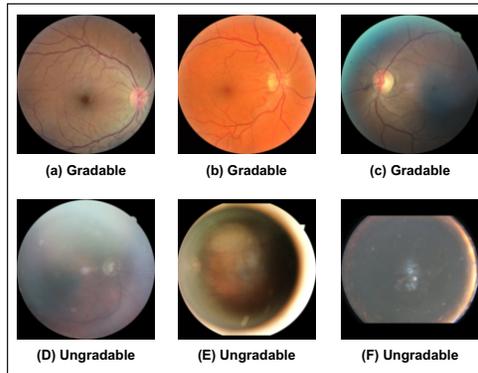
Poor quality retinal fundus images can lead to an incorrect automatic diagnosis of eye diseases and classification of its severity. Therefore, an expert must first visually classify the images as gradable or non-gradable. Visual analysis of large databases of retinal images is a time-consuming task that can be distributed among a large number of experts. However, this may lead to intervariability between the results in the quality assessment or diagnosis of different experts. Therefore, automatic analysis of retinal images is a possible solution to the lack of human experts. Moreover, to better understand the cause and progression of diseases, it may be necessary to analyze many images over a long period of time. However, the accuracy of these tasks required high quality retinal images. Automated quality analysis of retinal images can reduce the need for human intervention and create better conditions for further studies, increasing the functionality of retinal disease diagnosis based on computer aided design (CAD) systems.

Therefore, it is essential to have a measure to assess the quality of the retinal image prior to processing it for diagnosis [2]. Different techniques are found in the literature to assess the quality of retinal images. There are different metrics also used for measuring the quality of retinal images. Most of previous works depend on hand-crafted computer vision techniques, such as [3,2,4]. Nowadays with the progress of deep learning techniques, recent works use different deep learning networks architectures to develop a accurate retinal images quality assessment, such as [5,6,7,8].

All of the aforementioned eye diseases diagnosis systems scientifically depend on the quality of retinal images. Thus, there is a need for an improved approach for reliable and trustworthy retinal quality assessment for improving early detection of eye diseases. Towards this end, in this work, we propose a deep learning framework, which consists of two successive networks: an auto-encoder network depend on reconstruction image and a CNN-based classifier. The encoder will used for extract the key features related the quality of retinal images. These features get from encoder network are then fed to the classifier in order to classify the quality of the input retinal images. Our contributions in this paper are as follows:

- We propose an autoencoder network to correctly learn representative deep features of the retinal fundus images via the encoder network. The decoder part is used to reconstruct the input fundus image.
- We propose a CNN classifier fed by the features learned by the encoder network to classify the input fundus images as gradable or ungradable.
- We propose the use of the mean-square-error metric (MSE) as a loss function to train the autoencoder network. The MSE loss function calculates the sum of squared distance between input image and reconstructed image by the decoder. Also, we use a binary cross-entropy loss function to train the CNN classifier.
- We propose to integrate the losses of the two autoencoder network and the CNN classifier into a single learning framework to solve the fundus image gradeability problem.

The remainder of the paper is structured as follows. Section 2 reviews recent related literature to retinal image gradeability assessment. Section 3 presents the proposed method. Section 4 presents experimental results. Section 5 concludes the paper and gives directions for future scope of the research.



**Figure 1.** Example of gradable (top) and ungradable (bottom) retinal fundus images.

## 2. Related Work

In this section, we present a quick review of previous works related to retinal image assessment by using classical computer vision techniques and deep learning techniques. Many methods are based on binary quality labels (i.e., Accept and Reject), and others are based on a three-level quality grading system (i.e., Good, Usable and Reject). To our knowledge, no method has been proposed so far to classify the quality of retinal images through an autoencoder network like our model.

Just to name a few, [2] proposed a system for automatically detecting quality of retinal images based on the RGB color-space. Their system uses vessel density, textural features, global histogram features besides a metric known as non-reference perceptual sharpness. They also concentrated on three Regions of Interest (ROI) such as lower retinal hemispheres, upper retinal hemispheres and optic disc region. [3] also proposed a novel method for retinal image registration based on a concept named Salient Feature Region (SFR). The authors defined a measure called region saliency measure that is used to obtain SFR using gradient field entropy and local adaptive variance. They also used another method that combines gradient field distribution and computation of SFRs. However, it needs further enhancement for dealing with multimodal images. Besides, [9] introduced to Distortion Identification-based Image Verity and INtegrity Evaluation (DI-IVINE) to understand quality of distorted images. That identifies distortions and then assess quality based on the estimated distortions. [10] also proposed a Blind Image Quality Assessment (BIQA) method based on gradient magnitude and Laplacian of Gaussian response. [11] proposed a method named no-reference quality assessment metric for assessing quality of video encoding. The metric has two models, coefficient distribution model resulting in local error estimation and perceptual model resulting a quality score for a test image. One of the drawbacks of this approach is that needs further enhancement to avoid transmission errors.

Based on machine and deep learning, many methods have been proposed. For instance, [12] proposed a quality assessment method based on S3D INtegrated Quality (SINQ) Predictor that based on both univariate and bivariate statistical features obtained from image. In turn, [13] proposed a method for BIQA based on learning quality lookups and receptive fields. It used both local and global receptive fields. Their method needs improvement to consider different distortions in the training images. [14] developed a

framework for assessing 3D quality assessment of images containing filters to improve the image quality and algorithms to assess this quality. It is based on a technique called no-reference quality assessment. They used Artificial Neural Network (ANN) to achieve the assessment task. [5] also proposed a framework for synthesis of retinal colour images using an adversarial setting with generator and discriminator. Their framework needs further enhancement to overcome exhibiting abnormal interruptions. Recently, [15] analyzed the influences of different color-spaces on the retinal images assessment, and proposed a deep network, called Multiple Color-space Fusion Network (MCF-Net). MCF-Net integrated the different color-space representations (i.e., RGB, HSV and LAB) at a feature-level and prediction-level to predict image quality grades.

### 3. Proposed Model

Fig. 1 represents the training and testing phase of the proposed model. In the training model, we have uses an autoencoder network that consists of two serial networks: encoder and decoder. The encoder will be used for extracting the high-level features of the input images. The extracted features will be then fed to the decoder network to reconstruct again the same input image. Afterwards, another network, a classifier, will be fed by the features gotten from the encoder network to classify the quality of a retinal image into two classes: gradable and ungradable. The size of the input image is reshaped to 224x224. In the testing model, we used only the trained encoder and the classifier network in order to classify the quality of a retinal image.

#### 3.1. Network Architecture

Our model is based on the UNet [16] network. It is an encoder-decoder deep network architecture. UNet network is a full convolution network that does not include a fully connected layer and is not demanding on the amount of dataset. This network is simple, efficient, and easily used and adapted. It consists of two sub-networks, the first sub-network is an encoder that obtains different image feature levels continuously sampled through five convolution layers. The second one is a decoder that performed five deconvolution layers on the top-level feature map and combined different feature levels in the down-sampling process to restore the feature map to the original input image size and complete the end-to-end segmentation task of the image. Besides, it uses the skip connection operation to connect each pair of down-sampling layers and the up-sampling layer, which makes the spatial information directly applied to much deeper layers and a more accurate segmentation result.

The main task of using the UNet network is for semantic segmentation. However, in this paper, we used UNet for reconstructing the input image. We believe that if the autoencoder network succeeded in reconstruct the same input image, that means the network succeeded to learn the key features of the input image including visual quality features. Thus, the input to our UNet network in a RGB image and the target is the same RGB image.

For the quality classification network, we used very simple network. This classifier consists of one fully connected layer followed with Sigmoid as an activation function. This layer classify the features extracted by the encoder network to two quality labels that are gradable or ungradable.

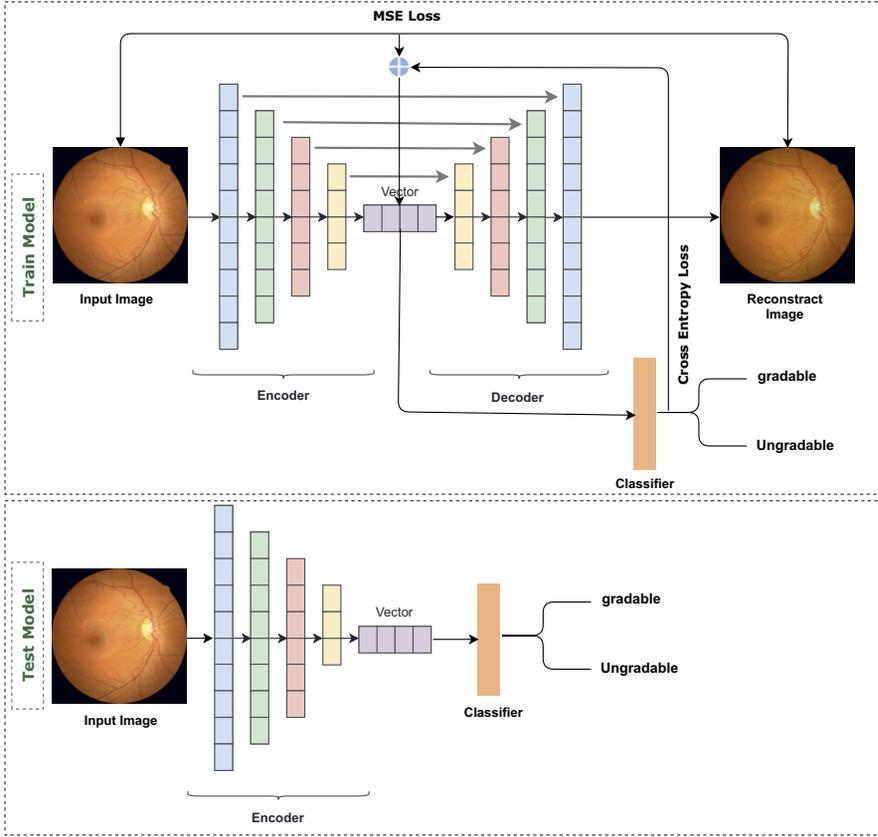


Figure 2. General overview of the proposed model in train and test stage.

### 3.2. Training

To assess the performance for optimizing the training of the network, we tried to use different loss function to compare the input image to the reconstructed image. This loss function is named  $L_{rec}$ . The first tested loss function is a Mean squared error (MSE) depends on the features extracted from both the input real image  $A$  and the reconstructed image  $\hat{A}$  from the autoencoder network. MSE is the most commonly used a loss function for regression tasks. The loss is the mean overseen data of the squared differences between true and predicted values,  $MSE$  is defined as follows:

$$L_{rec} = \frac{1}{n} \sum_{i=1}^n (A_i - \hat{A}_i)^2, \quad (1)$$

where  $A_{(i)}$  is the input image of pixel  $i$ ,  $\hat{A}_{(i)}$  is the reconstructed image and the  $n$  is the numbers of pixels in an image.

The second loss function is Mean absolute error (MAE). MAE is the mean overseen data of the absolute differences between true and predicted values, which depends on the features extracted from both the input real image  $A$  and the reconstructed image  $\hat{A}$  from auto-encoder network, and it can be defined as:

$$L_{rec} = \frac{1}{n} \sum_{i=1}^n |A_i - \hat{A}_i|. \quad (2)$$

The third loss function structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital images. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric for measuring the quality of reconstructed images compared to input images. SSIM can be defined as:

$$SSIM(\hat{A}, A) = \frac{(2\mu_{\hat{A}}\mu_A + c_1)(2\sigma_{\hat{A}A} + c_2)}{(\mu_{\hat{A}}^2 + \mu_A^2 + c_1)(\sigma_{\hat{A}}^2 + \sigma_A^2 + c_2)}, \quad (3)$$

where  $\mu_{\hat{A}}$  is the mean of  $\hat{A}$ ,  $\sigma_{\mu_{\hat{A}}}$  is the standard deviations of  $\hat{A}$ ,  $\mu_A$  is the mean of  $A$ ,  $\sigma_{\mu_A}$  is the standard deviations of  $A$ ,  $\sigma_{\hat{A}A}$  is the covariance of  $\hat{A}$  and  $b$ ,  $c1 = 0.01^2$ ,  $c2 = 0.03^2$ , respectively.

The fourth loss function multi-scale structural similarity index measure (MS-SSIM). The MS-SSIM approach is based on modeling of image luminance, contrast and structure at multiple scales and Multi-scale method is a convenient way to incorporate image details at different resolutions. MS-SSIM can be defined as:

$$MS-SSIM(\hat{A}, A) = [L_M(\hat{A}, A)]^{\alpha m} \cdot \prod_{j=1}^m [C_j(\hat{A}, A)]^{\beta j} \cdot [S_j(\hat{A}, A)]^{y_j}, \quad (4)$$

where  $m$  represents the image quality assessment scale, which is obtained after  $M_i$  iterations. At the  $j$ -th scale, the contrast comparison  $C_j(\hat{A}, A)$  and the structure comparison  $S_j(\hat{A}, A)$  and The luminance comparison  $L_M(\hat{A}, A)$ , and the exponents  $\alpha m, \beta j$ , and  $y_j$  are used to adjust the relative importance of different components.

For the quality labelling task, we used the standard loss function  $L_c$ , cross-entropy (CE), which depends on the predicted class from the classifier  $\hat{y}$  and corresponding target value  $y$ .  $CE$  is defined as follows:

$$L_c = - \sum_{i=1}^n y_i \cdot \log(\hat{y}_i), \quad (5)$$

where  $\hat{y}_i$  is the  $i$ -th scalar value in the model output,  $y_i$  is the corresponding target value, and output size is the number of scalar values in the model output. This loss is a very good measure of how distinguishable two discrete probability distributions are from each other.

In this context,  $y_i$  is the probability that event  $i$  occurs and the sum of all  $y_i$  is 1, meaning that exactly one event may occur. The minus sign ensures that the loss gets smaller when the distributions get closer to each other.

The final objective function, during the training for the model, is defined as follows:

$$Loss = (L_{rec} + L_c)/2. \quad (6)$$

#### 4. Experiments and Results

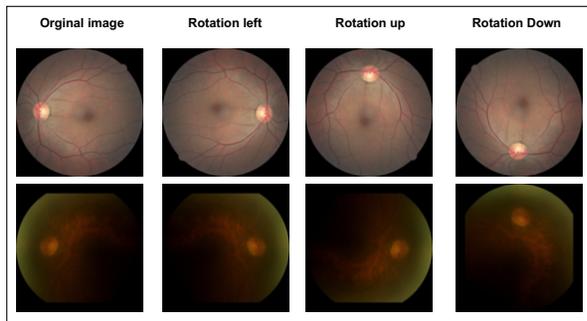
The experiments performed to evaluate the proposed model are described in this section, datasets and the evaluation measures used in the experiments.

#### 4.1. Datasets

There are several publicly available RIQA datasets with manual quality annotations, such as Eye-Quality (EyeQ) [15], HRF [17], DRIMDB [18] and DR2 [19]. Among them, the EyePACS- Quality dataset is a publicly available database with about 31,032 images labelled with gradable and ungradable without Data augmentation. we selected 29,033 for the training set and 1999 for the testing set. The second (EyeQ) dataset with 28,792 retinal images with manual quality labels (good, reject). Among them, the dataset is divided into 12,543 for training and 16,249 for testing.

#### 4.2. Data augmentation

CNN networks require large data sets in order to avoid overfitting. A class balanced dataset is also desirable as well. One of the proven approaches that can yield good results is to get more data from small datasets. In this study, we applied data augmentation techniques to the training images as proposed in [20,15] to increase the number of training samples under different conditions. Figure 3 shows the transformations applied to every input image. The applied data augmentation is done in two steps. The first steps a copy of the training examples of the small classes is done until they have the same number of images as the biggest class. This generates an equilibrated training set. In the second step, to every image, different transformations are applied in order to diversify the training, such as random uniform rotation and random flipping. Training data has two class first class (1) gradable contain 21,812 image and(0) ungradable class contain 7,218,After augmentations class 0 will be 28,872 and class 1 will be 29,083 , total for training for each class (i.e., gradable and ungradable), we have 57,957 images



**Figure 3.** Transformations (flipping and rotation) applied to every real image in all transformations.

#### 4.3. Parameter settings

We trained our networks. We used the Adam optimizer [21] with  $\gamma = 0.1$ , and an initial learning rate of 0.001. A batch size of 2 and 50 epochs yielded the best combination. All experiments were run on a 64-bit Core I7-6700, 3.40GHz CPU with 8GB of memory, as well as one NVIDIA GTX 1080 GPU on Ubuntu 16.04 and the PyTorch [22] deep learning framework. The computational time of the proposed method for the training process takes around 1 hour and 13 minutes and 37 second for each epoch with a batch size of 2 with train stage. In turn, the our model has a performance of around and 0.076 second for each image in train stage and 0.026 second for each image in test stage.

#### 4.4. Evaluation Metrics

In this work, to assess the proposed model, we use five different metrics to measure the resulting performance that are Accuracy, Sensitivity (Recall), Specificity, Precision and F1 Score. In medical diagnosis, sensitivity measures the model ability to correctly identify high quality images, whereas specificity measures the model ability to correctly identify low quality images.

#### 4.5. Results and Discussion

To evaluate the performance of the proposed model, we compared its results against a recent fundus image gradeability method called Multiple Color-space Fusion Network(MCF-Net) [15]. Both models were trained with the two datasets (i.e., subsets from Eyepacs and Eye-Quality). We tested different variations of loss functions with the proposed model: *MSE*, *MAE*, *SSIM* and *MS – SSIM* (Section 3) to find the best loss function that can help converge the network. Based on the two datasets, Table 1 and 2 show the results (i.e, Accuracy, Sensitivity, Specificity, Precision and *F1* score) of MCF-Net and four variations of the proposed systems with the four loss functions. As shown in Table 1, the proposed model with its four variations outperformed the performance of MCF-Net in terms of the five evaluation matrices. Among them, our model with MSE as a loss function yielded the best results with *F1* score, sensitivity and specificity of 0.88, 0.83 and 0.91, respectively. For instance, our model with MSE yielded an improvement of 8% with *F1* score compared to the MCF-Net. In turn as shown in Table 2 and with the second dataset EyeQ, the proposed model and its variations also outperformed the results with MCF-Net. Our model with MSE achieved significant improvements of 16%, 10% and 38% with *F1* score, precision and specificity, respectively. Besides, a small improvement of around 1% with sensitivity.

**Table 1.** Comparison between the proposed model and MCF-Net [15] on the Eypces dataset [20]

	Accuracy	Sensitivity	Specificity	Precision	F1-Score
<b>MCF-Net Model</b>	0.81	0.64	<b>0.95</b>	0.84	0.80
Our Model - SSIM Loss	0.815	<b>0.95</b>	0.65	0.84	0.82
Our Model - MS-SSIM Loss	0.86	0.94	0.76	0.87	0.86
Our Model - MAE Loss	0.85	0.84	0.86	0.85	0.85
<b>Our Model - MSE Loss</b>	<b>0.875</b>	0.83	0.91	<b>0.88</b>	<b>0.88</b>

**Table 2.** Comparisons of the proposed model and state-of-the-arts on (EyeQ) dataset [15]

	Accuracy	Sensitivity	Specificity	Precision	F1-Score
<b>MCF-Net Model</b>	0.865	0.946	0.51	0.80	0.75
Our Model - SSIM Loss	0.93	0.94	0.90	0.88	0.90
Our Model - MS-SSIM Loss	0.935	0.93	<b>0.91</b>	<b>0.94</b>	<b>0.93</b>
Our Model - MAE Loss	0.94	0.95	0.88	0.90	0.91
<b>Our Model - MSE Loss</b>	<b>0.942</b>	<b>0.954</b>	0.89	0.90	0.91

For validating the gradability of retinal fundus image classification performance, we also computed the confusion matrix and the overall classification accuracy on the test set of the two datasets.

This allows more detailed analysis than the mere proportion of high classification rate. As shown in Table 3, the numbers of true positives (TPs) and true negative (TNs) of the proposed model with the EyePACS dataset are 960 and 765 out of 1999 images, respectively. In the case of the EyeQ dataset, the numbers of TPs and TNs obtained by the proposed model are 12432 and 2867 out of 16249 images, respectively.

**Table 3.** Two resulting confusion matrices with the test sets of the EyePACS and EyeQ datasets

EyePACS datasets			EyeQ datasets		
n=1999	Predicted NO	Predicted YES	n=16249	Predicted NO	Predicted YES
Actual: NO	TN=765	FP=177	Actual: NO	TN=2867	FP=353
Actual: YES	FN=97	TP=960	Actual: YES	FN=597	TP=12432

Based on the results shown in Tables 1 and 2, we recommended to use the MSE function as the main loss function for calculating the the error between the input image and the reconstructed image. In general, we can say that the proposed model based an autoencoder network yields a promising results on improving RIQA for more accurate ocular diseases classification systems.

## 5. Conclusions

In this work, we proposed a supervised deep learning model based on an autoencoder network. The autoencoder is able to construct the same input image to correctly learn the visual features of fundus image quality. The model also include a classifier that is fed by the extracted features to classify the quality of a retinal image into gradable and ungradables. The proposed model can generate more interest in the biomedical community to improve the performance of the RIQA tasks, which plays a critical role in applications such as retinal image segmentation and automatic disease diagnosis. Future work aims to use the developed RIQA model for improving the accuracy of multi-task ocular diseases classification.

## Acknowledgement

The Spanish Government partly supported this research through project PID2019-105789RB-I00.

## References

- [1] G. M. Leggio, C. Bucolo, C. B. M. Platania, S. Salomone, and F. Drago, "Current drug treatments targeting dopamine d3 receptor," *Pharmacology & Therapeutics*, vol. 165, pp. 164–177, 2016.
- [2] H. Yu, C. Agurto, S. Barriga, S. C. Nemeth, P. Soliz, and G. Zamora, "Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening," in *2012 IEEE Southwest symposium on image analysis and interpretation*, pp. 125–128, IEEE, 2012.

- [3] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang, and M. Xu, "Salient feature region: a new method for retinal image registration," *IEEE transactions on information technology in biomedicine*, vol. 15, no. 2, pp. 221–232, 2010.
- [4] S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, and D. Qian, "Human visual system-based fundus image quality assessment of portable fundus camera photographs," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1046–1055, 2015.
- [5] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 781–791, 2018.
- [6] Y. Wang and S. Shan, "Accurate disease detection quantification of iris based retinal images using random implication image classifier technique," *Microprocessors and Microsystems*, vol. 80, p. 103350, 2021.
- [7] M. Ortega, N. Barreira, J. Novo, M. G. Penedo, A. Pose-Reino, and F. Gómez-Ulla, "Sirius: a web-based system for retinal image analysis," *International journal of medical informatics*, vol. 79, no. 10, pp. 722–732, 2010.
- [8] V. K. Singh, H. A. Rashwan, A. Saleh, F. Akram, M. M. K. Sarker, N. Pandey, and S. Abdulwahab, "Refuge challenge 2018-task 2: Deep optic disc and cup segmentation in fundus images using u-net and multi-scale feature matching networks," *arXiv preprint arXiv:1807.11433*, 2018.
- [9] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
- [10] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.
- [11] T. Brandao and M. P. Queluz, "No-reference quality assessment of h. 264/avc encoded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1437–1447, 2010.
- [12] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Processing: Image Communication*, vol. 58, pp. 287–299, 2017.
- [13] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, and Q. Dai, "Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images," *IEEE transactions on cybernetics*, vol. 46, no. 3, pp. 730–743, 2015.
- [14] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.
- [15] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, and L. Shao, "Evaluation of retinal image quality assessment networks in different color-spaces," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 48–56, Springer, 2019.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [17] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proceedings of the 26th IEEE international symposium on computer-based medical systems*, pp. 95–100, IEEE, 2013.
- [18] U. Sevik, C. Kose, T. Berber, and H. Erdol, "Identification of suitable fundus images using automated quality assessment methods," *Journal of biomedical optics*, vol. 19, no. 4, p. 046006, 2014.
- [19] R. Pires, H. F. Jelinek, J. Wainer, and A. Rocha, "Retinal image quality analysis for automatic diabetic retinopathy detection," in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 229–236, IEEE, 2012.
- [20] B. Graham, "Kaggle diabetic retinopathy detection competition report," *University of Warwick*, 2015.
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [22] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration," 2017.