A deep retinal image quality assessment network with salient structure priors

Ziwen Xu1 · Beiji Zou1,2 · Qing Liu1

Received: 25 September 2021 / Revised: 11 February 2022 / Accepted: 5 February 2023 / Published online: 10 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Retinal image quality assessment is an essential prerequisite for diagnosis of retinal diseases. Its goal is to identify high quality retinal images in which anatomic structures and lesions attracting ophthalmologists’ attention most are exhibited clearly and definitely while reject poor quality images. Motivated by this, we mimic the way that ophthalmologists assess the quality of retinal images and propose a method termed SalStructuIQA. First, two salient structures are detected for retinal quality assessment. One is large-size salient structures including optic disc and exudates in large-size. The other is tiny-size salient structures mainly including vessels. Then the proposed two salient structure priors are incorporated with deep convolutional neural network (CNN) to enforce CNN pay attention to these salient structures. Accordingly, two CNN architectures named Dual-branch SalStructIQA and Single-branch SalStructIQA are designed for the incorporation, respectively. Experimental results show that the $F_{score}$ of our proposed Dual-branch SalStructIQA and Single-branch SalStructIQA are 0.8723 and 0.8662 respectively on the public Eye-Quality dataset, which demonstrates the effectiveness of our methods.

Keywords Convolution neural network · Optic disc · Retinal image quality assessment · Vessels
1 Introduction

Retinal images are the most common modality used in the diagnosis of retinal diseases such as diabetic retinopathy, age-related macular degeneration and glaucoma etc [20, 29, 32]. High quality retinal images ensure the salient structures, such as optic disc, vessels, and lesions etc clearly visible, which help ophthalmologists make correct clinical decisions efficiently [38]. On the contrary, low quality images may disturb the judgments. For example, in over-exposed retinal images, the contrasts between target regions (such as vessels and lesions etc.) and the background are low, which may confuse the ophthalmologists when accessing the fundus. Furthermore, for computer-aided retinal image analysis systems, low quality retinal images would be a catastrophe. According to [24], approximately 25% of the retinal images can not used for diagnosis due to their poor quality. Thus retinal image quality assessment (Retinal-IQA) is desired. In clinical, it is performed by a trained optometrist manually, which heavily depends on the operator’s experience and is time-consuming. To improve the efficiency of retinal image acquisition, automated Retinal-IQA becomes necessary.

Early Retinal-IQA methods grade the image quality via solving the problems of how to measure the similarity to good quality retinal images or how to characterise quality distortions. In [17, 18], template matching is adopted for Retinal-IQA, in which intensity histogram and intensity-gradient histogram are calculated from a set of good quality retinal images as templates. In [6, 44], quality distortions are characterised by colour and illumination distortion, focus fluctuation and contrast sensitivity, with which a classifier is trained for retinal image quality grading. Although those methods are simple and intuitive, both templates based on histograms and features manually designed for quality distortions are insufficient to encode quality measures with high discriminative ability, which limits the performances of Retinal-IQA.

Driven by the successes achieved by deep convolutional neural networks (CNNs) in the field of computer vision, Retinal-IQA methods have updated to modern CNNs-based. In [49], Inception v3 [40] pre-trained on natural scene image dataset ImageNet [5] is fine-tuned in an end-to-end way for Retinal-IQA, which takes colour retinal images as input. To make use of both deep features and hand-crafted features, a two-stage learning method is proposed in [48]. In detail, deep features are learned in the first stage via fine-tuning an AlexNet [16]. With the deep features and hand-crafted features based on a saliency map, a SVM is trained for quality classification in the second stage. To further exploit powerful representation for Retinal-IQA, complex frameworks are proposed. Fu et al. propose a framework, called Multiple Colour-space Fusion Network (MCF-Net) [7], in which three parallel CNN branches are unified to learn complementary informative contexts from three different colour-spaces, i.e. RGB, HSV and Lab colour space, for Retinal-IQA. In [37], Shen et al. claim that auxiliary tasks contribute to Retina-IQA and propose a multi-task framework, named MFIQA. MFIQA [37] consists of a ResNet50 [9] for optic disc and macular detection, a VGG16 [39] for refinement of optic disc and macular location and two encoders to encode the global retinal images in Cartesian coordinate system and the local optic disc windows in polar coordinate system for images quality classification and three relevant classification tasks, i.e., artifact, clarity and field definition. It is further improved by introducing domain invariance and interpretability in [38]. Although learning multiple CNN branches jointly like [7] and multiple CNNs independently like [37, 38] exploit rich features and facilitate
classification, the parameters to be optimised rapidly multiply. Furthermore, in [37, 38], learning auxiliary tasks requires extra annotated data, which are always expensive.

Going back to the basics, the goal of capturing retinal images by fundus camera is to assist ophthalmologists to assess anatomic structures on fundus in a non-inVASion manner. From the view of saliency detection, those anatomic structures such as optic disc in large size and vessels in tiny size attract most of attention from ophthalmologists. Thus, a good quality retinal image is required to exhibit those salient structures clearly and definitely. To intuitively illustrate this, three examples from Eye-Quality dataset [7] are shown in Fig. 1. They are graded as ‘Good’, ‘Usable’ and ‘Reject’ respectively. Obviously, in good quality retinal image as shown in the left column in Fig. 1, the optic disc region and vessels are clearly visible. In usable quality retinal image as shown in middle column in Fig. 1, the optic disc region and most of the vessels are clearly visible except for vessel extending to the boundary of field of view which are distorted by slight uneven illumination. In retinal image labelled as ‘Reject’, due to the serious uneven illumination, the optic disc region and vessels are difficult to be distinguished.

Based on above observations, we argue that the clarity of salient structures in both large-size and tiny-size significantly contributes to the image quality assessment. Usually, more clarity that salient structures in retinal images are implies higher quality that retinal images have. To this end, we propose to incorporate the salient structure priors with deep CNN, named SalStructIQA. Figure 2 illustrates the flowchart of our proposed SalStructureIQA. It involves three components: (1) large-size salient structure detector which aims to detect salient structures in large-size such as optic disc region, large-size hard exudates and soft exudates etc., (2) tiny-size salient structure detector which aims to detect vessels, (3) a deep classification network which takes the retinal image together with the salient structures as input for deep feature learning and quality classification.

Accordingly, two deep CNN architectures are designed to incorporate salient structure priors with CNNs. One contains two parallel CNN branches for deep feature learning, which learns deep features from retinal image with large-size salient structures and retinal image with tiny-size salient structures separately, then fuses the features for quality assessment. We term it as Dual-branch SalStructIQA. The other only contains a single CNN branch.

![Fig. 1](image1)

**Fig. 1** Examples for retinal images with different quality grades. From left to right, the quality grade is ‘Good’, ‘Usable’ and ‘Reject’, respectively.
the contributions of this paper can be summarized as follows:

- We propose two salient structure priors in both large-size and tiny-size for retinal image quality assessment. They enable the CNN to pay more attention to informative regions with high contrast to background and learn discriminative features, which further facilitate to retinal image quality assessment.
- We propose two ways to incorporate our proposed salient structure priors with deep CNN and develop two deep CNN architectures termed Dual-branch SalStructIQA and Single-branch SalStructIQA for retinal image quality assessment.
- Experimental results on a public dataset Eye-Quality [7] show that both the Single-branch and Dual-branch SalStructIQA are able to achieve competitive performances compared with the state-of-the-arts and the scale of parameters to be optimised is much less. Moreover, experiments on our local dataset also demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 reviews related works on Natural-IQA and Retinal-IQA. Section 3 details our proposed quality assessment method. Section 4 gives the experimental results and analysis, Section 5 discusses several limitations about our methods and Section 6 concludes this paper.

2 Related work

Following the evolution of natural scene image quality assessment (Natural-IQA), retinal image quality assessment methods are developed from the two-step framework with handcrafted features to the end-to-end framework based on CNN in general. As retinal image quality assessment is closely relative to natural scene image quality assessment, in this
section, we first review previous works on Natural-IQA briefly, then elaborate preceding Retinal-IQA methods.

2.1 Natural-IQA

Based on the type of extracted features, Natural-IQA methods can be briefly classified into two groups: hand-crafted feature based methods and CNN-based methods. The hand-crafted based approaches mainly focus on designing effective hand-crafted features while the CNN-based methods automatically learn image features by designing deep networks.

Hand-crafted feature based methods These methods usually assume that pristine natural images possess potential statistic regularities and these regulations will be changed with the emergence of distortion [26]. For example, [26] proposes a two-stage framework called DIIVINE for quality assessment, which consists of distortion classification and distortion-specific quality estimation. BLINDS-II [34] utilizes DCT coefficients to model statistics of images, and then combines the estimated parameters of the model with Bayesian approach for quality score prediction. BRISQUE [25] adopts local normalization brightness coefficients for image quality. To better improve the performance, a new NBIQA [30] simultaneously considers the features from both the spatial and transform domains. Although NBIQA [30] performs better, it is also relatively time-consuming due to the complexity of extracted features. In addition, [19] designs a joint natural scene statistics model based on joint log-contrast statistics for Natural-IQA, which utilizes two natural image properties, i.e., the non-Gaussian statistics and image correlations across spatial and direction. As a whole, the hand-crafted feature based methods heavily rely on the artificial features design, which is too complex to achieve efficient and appropriate features.

CNN-based methods Due to the remarkable achievements by the CNNs in image classification tasks, researchers have introduced the deep learning method into Natural-IQA. In [14], the CNN only containing one convolutional layer is proposed for Natural-IQA. To better characterize features, [23] proposes a multi-task end-to-end optimized network for quality prediction called MEON. Similar to DIIVINE [26], MEON focuses on two subtask, distortion identification and quality prediction. In [21, 22], rank learning is adopted for Natural-IQA. In rankIQA [21], the traditional image processing method is firstly utilized to generate ranked images which are then used to train a siamese network for ranking. Subsequently, the branch of trained network is fine-tuned on IQA dataset for predicting the absolute quality score. In dipIQ [22], a large number of quality-discriminable image pairs are used to train a RankNet [2], generating an inferred image quality. However, the above CNN-based methods ignore the salient structures attracting human attention most, which have been demonstrated to be important for image quality assessment [45–47]. Yan et al. [45] proposes a two-stream network for Natural-IQA, which includes an image stream focusing on gray information and a structure stream focusing on image gradient details. Considering that image saliency can present image structure to some extent, [46] proposes a saliency-guided Natural-IQA network called SGDNet, which is trained with saliency ground truth and quality label. Similar to approach [13], SGDNet regards the saliency map as local weighted function. Specifically, the predicted saliency map is fused with original
features in the form of spatial attention mask, perceptually consistent to quality prediction. Yang et al. [47] proposes a deep feature importance awareness method to evaluate image quality named DFIA. They integrate the Squeeze-and-Excitation (SE) block [10] into deep IQA network, which emphasizes the deep features of image distortion and salient objects related to IQA.

2.2 Retinal-IQA

Retinal-IQA is an important task for the diagnosis of retinal diseases. However, Natural-IQA methods may not perform well in Retinal-IQA. More specifically, the hand-crafted based methods for Natural-IQA mainly rely on a hypothesis that pristine natural images present certain statistic regularities which are disturbed by distortions. Those statistics of natural images are not consistent with those of retinal images. In addition, several CNN-based methods for Natural-IQA usually assign the global quality score for all image patches cropped from the same image, resulting the label noise problem. This strategy is also not suitable for Retinal-IQA. Therefore, it is essential to design effective methods for Retinal-IQA. Similar to Natural-IQA, the current researches in Retinal-IQA can be roughly grouped to two categories: hand-crafted based methods and CNN-based methods.

Hand-crafted based methods Early methods such as [17, 18] for Retinal-IQA pursue a template for good quality images and grade the image quality via template matching while [4, 6] extract statistical characteristics such as focus, lumination and contrast features etc. and train a classifier for image quality assessment. Based on the assumption that the more vessels the better retinal image quality, [42] first calculates the area of vessels to define image clarity, in which matched filtering and region segmentation algorithm are applied to segment the vessel structures. In [15], Kohler et al. assume that patches located on vessel edges are more sensitive to image blur and noise, and propose to use the variance of vessels in patches to discriminate good retinal images from poor quality retinal images. In [12], retinal image quality is graded by comparing the vessel clarity and contrast between the macular area and background. From the perspective of human visual system (HVS) characteristics, [44] proposes to apply HVS features, including multi-channel sensation, just noticeable blur, and the contrast sensitivity measure, to estimate quality. However, the limitation of these hand-crafted features may still lead to performance bottlenecks.

CNN-based methods Motivated by the success of CNN in computer vision field, CNN has been introduced to Retinal-IQA. Several methods for Retinal-IQA are either training a shallow classification CNN network from scratch [41] or fine-tuning existing classification networks, such as AlexNet [16], ResNet [9] and Xception [3], from pre-trained models [33, 35, 49]. To make use of the complementary of different colour spaces, Fu et al. [7] propose a multiple colour-space fusion network, in which multiple parallel CNN branches are used to learn features from different colour-spaces. Although [7] is able to capture rich information such that the classification performances are improved significantly, the scale of parameters rapidly multiply. Instead, Yu et al. [48] propose to combine salient maps produced by [1] and deep features extracted by Alexnet [16] and train an SVM for retinal image quality assessment. In [37], Shen et al. claim that auxiliary tasks contribute to Retinal-IQA and propose a multi-task framework named MFIQA, in which a ResNet50 [9] is trained for the detection of optic disc and macular and a VGG16 [39] is trained for refinement of optic
disc and macular location and two encoders are used to encode the entire retinal images in Cartesian coordinate system and the local optic disc windows in polar coordinates systems for images quality classification. Additionally, three relevant classification tasks, i.e., artifact, clarity and field definition are proposed to assist the learning of quality assessment. To further improve the performances, domain invariance and interpretability are introduced in [38]. Learning auxiliary tasks do facilitate Retinal-IQA, but requires extra annotated data, which are expensive.

2.3 Limitation

A brief summary of Retinal-IQA methods is provided in Table 1. In summary, although there are many works focusing on Retinal-IQA, there still suffer from many limitations. (1) Hand-crafted based methods still lead to performance bottlenecks; (2) CNN-based methods are suffer from data hungry and lack of large public dataset. One possible reason is that

| Year | Authors | Pros | Cons |
|------|---------|------|------|
| 1999 | Lee and Wang [18] | Simple via template matching of intensity histogram | Ignore structure and local features |
| 2001 | Lalonde et al. [17] | Introduce edge and local properties | Sensitive to distortion types |
| 2003 | Usher et al. [42] | Easy via matched filter | Sensitive to vessels |
| 2011 | Hunter et al. [12] | Consider vessels, perform well on Gaussian noise and blur | Sensitive to distortions e.g. uneven illumination |
| 2013 | Köhler et al. [15] | Effective under suffering from Gaussian noise and blur | Ignore the diversity of distortion |
| 2014 | Dias et al. [6] | Focus on color and contrast features | Limited on pre-defined science |
| 2015 | Wang et al. [44] | HVS features is effective for quality grade | The performance relies on parameters setting |
| 2016 | Tennakoon et al. [41] | Simple neural network | Tested only on small-scale datasets |
| 2017 | Yu et al. [48] | Incorporate saliency information to CNN | Rely heavily on saliency information |
| 2018 | Zago et al. [49] | Simple pre-trained models | Ignore quality prior of retinal image |
| 2018 | Shen et al. [37] | Introduce auxiliary tasks | Requires extra annotated data, expensive |
| 2019 | Fu et al. [7] | Multiple colour-space fusion network | The scale of parameters rapidly multiply |
| 2020 | Raj et al. [33] | Ensemble model | Multiple parameters |
| 2020 | Shen et al. [38] | Domain invariance and interpretability | Requires extra annotated data, expensive |
| 2021 | Muddamsetty et al. [27] | Combine generic features with CNN | Lack of intuitive and interpretable presentation |
the grade of image quality does not have a unified standard, leading the unavailability of comparative performance. (3) Most CNN-based methods adopt multiple CNN branches or need extra auxiliary tasks, increasing memory usage and causing labor-expensive.

3 Methodology

The purpose of photographing for retina fundus is to provide an invasive way for ophthalmologist to assess the fundus. Commonly ophthalmologists pay more attention to informative structures. Among them, two kinds of informative structures attract most attentions from ophthalmologists: structures in large-size such as OD and large-size hard exudates and soft exudates etc. and structures in tiny-size, i.e., vessels. Whether these salient structures are clearly and definitely exhibited plays an important role in determining whether retinal images have good quality for ophthalmologists. To mimic the quality assessment procedure of ophthalmologists, we propose a deep retinal image quality assessment method termed SalStructIQA, which first detects two kinds of salient structures, then incorporates them with deep CNN for quality grading. Figure 2 illustrates the flowchart of our proposed SalStructIQA. It consists of three components: large-size salient structure detector, tiny-size salient structure detector and deep CNN classifier. In the following, we will detail each of them.

3.1 Large-size salient structure detector

Within the context of retinal images, large-size structures mainly consists of OD and large-size exudates etc. They are in bright yellow. In good quality retinal images, those structures are easy to be recognised. On the contrast, in poor quality retinal images, those structures are difficult to be recognised. The major reason is that the contrasts of those large-size salient structures in different quality images are different. Particularly, for good quality retinal images, large-size structures naturally have high contrast to their surrounding regions as images are usually captured with even illumination. For poor retinal images, they are usually captured with uneven illumination. Defect regions with strong illumination exhibit high contrast while the contrasts of those bright yellow structures with weak illumination are suppressed. To this end, we propose to detect large-size salient structures via contrast for retinal image quality assessment. It is estimated by Difference of Gaussian (DoG) operator, which can be expressed as:

$$\text{DoG}(x, y, \sigma, \rho) = \frac{1}{2\pi} \left[ \frac{1}{\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}} - \frac{1}{(\sigma/\rho)^2} e^{-\frac{x^2+y^2}{(\sigma/\rho)^2}} \right]$$

$$= G(x, y, \sigma) - G(x, y, \sigma/\rho), \quad (1)$$

where $\sigma$ is the standard deviation of Gaussian function and $\rho$ is a constant which is larger than 1. From the view of frequency domain, the DoG operator is a band-pass filter and $\rho$ is the bandpass width. Considering that the salient structures are diverse in scale, we propose to combine multiple DoGs with $K$ different bandpass widths for contrast estimation:

$$\text{MultiDoG}(x, y) = \sum_{k=1}^{K} \left\{ \text{DoG}(x, y, \sigma_k, \rho_k) - \text{DoG}(x, y, \sigma_k, \rho_k^{-1}) \right\}$$

$$= G(x, y, \sigma) - G(x, y, \sigma/\rho^K). \quad (2)$$

Here we directly set $\sigma$ be a large constant to maintain most of the frequency and filter out the high frequency noises. Thus the first item in (2) is approximated to a low-pass filter
with a large cut-off frequency. For computational simplicity, we follow [1] and use a $5 \times 5$ Gaussian filter to approximate it in spatial domain. For the second item in (2), we set $K$ be infinity, thus it is approximated to a low-pass filter at zero frequency. More specifically, with the colour retinal image $I \in \mathcal{R}^{H \times W \times 3}$ in RGB colour space, we first transform it into Lab colour space as Lab is designed to approximate human vision. We denote the Lab image as $I^{Lab}$. Then in spatial domain, we estimate the contrast via:

$$R_{\text{contrast}}(x, y) = \left\| I_{\mu}^{Lab} - I_{lp}^{Lab}(x, y) \right\|_2^2,$$

where $I_{\mu}$ is the global mean vector of $I^{Lab}$, $I_{lp}^{Lab}(x, y)$ is the corresponding filter response at location $(x, y)$ with the $5 \times 5$ Gaussian filter, and $\| \cdot \|_2$ is the Euclidean distance. Finally,

![Fig. 3](image-url) Examples for detecting large-size salient structures from retinal images with different quality grade. From top to bottom in (a), there are examples of retinal images of different quality. (b) are their corresponding probability map, and (c) are corresponding masks of detected large-size salient structures.
we rescale the contrast map into [0, 1] via max-normalisation, and obtain the probability map of large-size salient structures denoted by $P_{LS}$:

$$P_{LS}(x, y) = \frac{R_{\text{contrast}}(x, y)}{\max_{(x,y)\in \mathbb{R}^{H\times W}}(R_{\text{contrast}}(x, y))}. \quad (4)$$

By thresholding $P_{LS}$, we obtain the mask of detected large-size salient structures, and we denote it as $M_{LS}$.

Figure 3 illustrates three retinal images from Eye-Quality dataset [7], their corresponding probability map $P_{LS}$ of large-size salient structures and binarized results. From top to bottom, retinal images are graded as ‘Good’, ‘Usable’ and ‘Reject’ by experts respectively. Obviously, their detection results of large-size salient structures are significantly different: (1) in retinal image graded as ‘Good’, large-size salient structures are well highlighted in $P_{LS}$; (2) in retinal image graded as ‘Reject’, defects are well highlighted in $P_{LS}$ while the salient structures are suppressed; (3) in retinal image graded as ‘Usable’, although slight detect is highlighted in $P_{LS}$, its adverse effect is weak and the salient structure is still well highlighted. These imply that the detection results of large-size salient structures do contribute to image quality grading.

3.2 Tiny-size salient structure detector

Vessels are important structures in retinal image. Their abnormal morphology is a significant manifestation for eye diseases such as diabetic retinopathy and age related macular degeneration. Although they are in tiny size, they are salient for ophthalmologists. Thus we propose to detect those tiny-size salient structures for retina image quality assessment.

Here we directly adopt multi-scale line detector [28] as our tiny-size salient structure detector due to its simplicity. Multi-scale line detector [28] treats the vessels as line-shaped and assumes that there are high intensity contrasts between pixels along vessel line and rest pixels among multi-scale windows. As the vessels is darker than the background in the gray channel $I_{\text{Gray}}$ of a retinal image, the multi-scale line detector is performed on its inverted gray channel to highlight the vessels. For a pixel at location $(x, y)$, the multi-scale line detector measures the line response by:

$$R_{\text{line}}(x, y) = \frac{1}{|S|} \left( \sum_{s \in S} \max_{\theta \in \Theta} \left( \text{Avg}(L_{x,y}^{s,\theta}) - \text{Avg}(\Omega_{x,y}^{s}) \right) \right), \quad (5)$$

where $S$ is the set of window size, and $|S|$ is the cardinality of set $S$, $\Theta$ is the set of line direction, and $\Omega_{x,y}^{s}$ is the local window of size $s \times s$ centred on $(x, y)$, and $L_{x,y}^{s,\theta}$ is the pixel set of line along direction $\theta$ passing location $(x, y)$ in $\Omega_{x,y}^{s}$, and Avg($\cdot$) is an average function. We follow the setting of multi-scale line director [28] and set $S = \{1, 3, 5, 7\}$ and $\Theta = \{0^\circ, 15^\circ, 30^\circ, \ldots, 165^\circ\}$. Since the contrast of vessels to other structures such as lesions is apparent in the original inverted gray image, multi-scale line detector [28] also utilizes the information of the inverted gray channel to enhance the line response:

$$R'_{\text{line}}(x, y) = \frac{1}{|S| + 1} \left( \sum_{s \in S} \max_{\theta \in \Theta} \left( \text{Avg}(L_{x,y}^{s,\theta}) - \text{Avg}(\Omega_{x,y}^{s}) \right) + \left(1 - I_{\text{Gray}}(x, y)\right) \right). \quad (6)$$

Finally, we follow [28] and standardize the enhanced line response map $R'_{\text{line}}$ via $Z$-score standardization. We denote the standardized line response map as $R''_{\text{line}}$. By thresholding the $R''_{\text{line}}$ with threshold value 0.56, we obtain the mask of detected tiny-size salient structures and denote it as $M_{TS}$. 

 Springer
Figure 4 shows three examples of retinal images and the results by tiny-size salient structure detector. Obviously, the tiny-size salient structures in good quality retinal image are well detected by our proposed tiny-size salient structure detector. Most of the tiny-size salient structures in retinal image graded as ‘Usable’ quality are detected by our proposed detector. On the contrary, few tiny-size salient structures in retinal image graded as ‘Reject’ quality are detected by our proposed detector. These indicate that the density of detected tiny-size salient structures is an important cue for the quality classification of retinal images. To better illustrate this, we provide a statistic analysis on the training set of Eye-Quality [7]. For each retinal image, we account the number of pixels which are detected as tiny-size salient structures and denote it as $|M_{TS}|$. Then for each quality level, we separately plot the cumulative distribution of $M_{TS}$. As shown in Fig. 5, there are 70% images in
class of ‘Reject’ whose $|M_{TS}|$ is less than 1513 while the percentage in class of ‘Usable’ is 31.82%. This implies that it is easy to discriminate the class of ‘Reject’ and ‘Usable’ according to $|M_{TS}|$. There are 70% images in class ‘Usable’ whose $|M_{TS}|$ is less than 2160 while the percentage in class of ‘Good’ is 23.44%. This implies that it is easy to discriminate the class of ‘Usable’ and ‘Good’ according to $|M_{TS}|$.

### 3.3 Network architectures

To incorporate two salient structure priors with the convolutional neural network for retinal image quality grading, we implement a deep network, named SalStructIQA network. For these two priors, they can be separately concatenated with the RGB retinal image forming two tensors with four channels as shown in Fig. 6(a) or jointly concatenated with the RGB retinal image forming one tensor with five channels as shown in Fig. 6(b). This derives two versions of SalStructIQA: Dual-branch SalStructIQA and Single-branch SalStructIQA. In what follows, the RGB retinal image is denoted as $I \in \mathbb{R}^{H \times W \times 3}$, and the masks of large-size salient structures and tiny-size salient structures are denoted as $M_{LS} \in \mathbb{R}^{H \times W}$ and $M_{TS} \in \mathbb{R}^{H \times W}$.

**Dual-branch SalStructIQA network** In this architecture, there are two parallel branches of CNN equipped with a global average pooling layer (GAP) for feature learning, as shown in Fig. 6(a). One branch takes the combination of $I$ and $M_{LS}$ as input and produces a feature vector denoted as $f_{LS}$. The other branch takes the combination of $I$ and $M_{TS}$ as input and produces a feature vector denoted as $f_{TS}$. Then these two feature vectors are concatenated and fed into a full connection (FC) layer to predict the quality grade:

$$\hat{y} = \text{softmax} \left( w^T [f_{LS}, f_{TS}] + b \right),$$

(7)
The proposed deep CNN architectures incorporating with our proposed large-size salient structures and tiny-size salient structures, where $w$ and $b$ are weights and bias to be learned in FC layer. Parameters in both parallel CNN backbones and FC layers are learned in an end-to-end way and cross-entropy loss is used in training phase.

One thing worth noting is that arbitrary existing CNN backbones such as VGGNet, ResNet and DenseNet etc. can be used for feature learning in each branch in Fig. 6(a) after revising the first dimension of convolution kernel in the first convolutional layer to four. In our experiments, performances of two famous backbones, i.e., ResNet-50 and DenseNet-121 are reported.

**Single-branch SalStructIQA network** The architecture is shown in Fig. 6(b). It consists of one CNN branch together with a GAP for feature learning and a FC layer for classification. This architecture takes $[I, f_{LS}, f_{TS}]$ as an input and predicts the quality grade. Same to dual-branch one, the network can be trained in an end-to-end way and cross-entropy loss is used in training phase.

In short, given a retinal image $I$, the large-size and tiny-size salient structures are firstly detected by the proposed two structure detectors, respectively. Then, the detected two salient structures are incorporated with deep CNN to predict quality grade $\hat{y}$. Furthermore, the pseudocode of our proposed method is provided in Algorithm 1.

### 4 Experimental results and analysis

In this section, we first describe the details of experimental configurations and evaluation metrics in detail. Then, the proposed method is evaluated and compared with the state-of-the-art image quality assessment methods on the public dataset Eye-Quality [7].
extensive ablation experiments are conducted on Eye-Quality [7]. Furthermore, we perform experiments on a local dataset for illustrate the effectiveness of the proposed SalStructIQA method.

**Datasets** There are two datasets used in this paper: a public dataset Eye-Quality [7] and a local dataset re-annotated by an ophthalmologist. The Eye-Quality [7] collects 28,792 retinal images from Kaggle and each retinal image is re-annotated with three grades, i.e., ‘Good’, ‘Usable’ and ‘Reject’. The image size ranges from $211 \times 320$ to $3456 \times 5184$. Those images are divided into two subsets: 12,543 for training and 16,249 for testing. To test the model generalization, we re-annotated a local dataset including 2560 retinal images. The local dataset is taken from Retinal Fundus Multi-disease Image Dataset [31], and then re-annotated by an expert from the First People’s Hospital of Changde for retinal image quality assessment. Among them, 2239 images are annotated as ‘Good’, 194 images are annotated as ‘Usable’ and 127 images are annotated as ‘Reject’.

**Data preprocessing and augmentation** With the retinal images, we follow [7] and first crop the field of view (FoV) via Hough Circle Transform, then we pad the short side with zeros such that the width and height of the cropped FoV regions are equal length. Finally, we rescale padded regions to $224 \times 224$. The masks of large-sized salient structures $M_{LS}$ and tiny-sized salient structures $M_{TS}$ are detected from rescaled images. To increase the diversity of training data, we use vertical flipping, horizontal flipping and random rotation to augment the training data.

---

**Algorithm 1** The proposed method.

---

**Input:** Retinal image $I$

**Output:** Quality grade $\hat{y}$

1: **procedure** LARGE-SIZE SALIENT STRUCTURE DETECTOR($I$)
2: Transform RGB image $I$ into $Lab$ image $I_{Lab}$
3: Estimate the contrast map $R_{contrast}$ by (3)
4: Calculate the probability map $P_{LS}$ by (4)
5: Acquire the mask $M_{LS}$ of detected large-size salient structures by thresholding $P_{LS}$
6: **return** large-size salient structures $M_{LS}$
7: **end procedure**

8: **procedure** TINY-SIZE SALIENT STRUCTURE DETECTOR($I$)
9: Transform RGB image $I$ into gray image $I_{Gray}$
10: Calculate the multi-scale line response $R_{line}$ by (6)
11: Estimate the standardized line response $R_{line}'$ of response map $R_{line}$
12: Acquire the mask $M_{TS}$ of detected tiny-size salient structures by thresholding $R_{line}'$
13: **return** tiny-size salient structures $M_{LS}$
14: **end procedure**

15: **procedure** NETWORK ARCHITECTURES($I$, $M_{LS}$, $M_{TS}$)
16: Fuse image $I$ and $M_{LS}$ and fuse image $I$ and $M_{TS}$ as the inputs of dual branches of Dual-Branch SalStructIQA Network, respectively
17: Extract and fuse features maps from dual branches
18: Predict image quality grade $\hat{y}$
19: **return** Quality grade $\hat{y}$
20: **end procedure**
Experimental setup For Eye-Quality [7] dataset, parameters in backbone network are initialized with the pre-trained models on ImageNet [5]. Parameters associated with full connection layers are initialized with kaiming uniform initialization [8]. We adopt stochastic gradient descent to optimize the network with initial learning rate of 0.01. We train the model 20 epochs and reduce the learning rate to 0.001 after 10 epochs. For the local dataset, parameters are initialized with the corresponding pretrained models based on Eye-Quality [7], and then are optimized with the constant learning rate 0.001. The experiments of our proposed methods are performed on one GTX1080 Ti GPU.

Evaluation metrics For fair comparison, we follow the evaluation method in the work [7], and adopt average accuracy (Acc), precision (Pre), recall (Rec) and $F_{\text{score}}$ as performance metrics. Since $F_{\text{score}}$ considers precision and recall at the same time, it is a comprehensive value to assess the performance of our approaches. To describe these performance metrics, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are calculated. TP and TN refer to the number of images whose predicted quality and actual quality are correctly matched. FP and FN mean the number of images whose predicted quality and actual quality are oppositely matched. Hence, these evaluation metrics are formulated as follows:

$$\text{Acc} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$  

$$\text{Pre} = \frac{TP}{TP + FP},$$  

$$\text{Rec} = \frac{TP}{TP + FN},$$  

$$F_{\text{score}} = \frac{2 \times \text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}}.$$  

4.1 Performances on Eye-Quality dataset

We compare our SalStructIQA with seven image quality assessment methods: BRISQUE [25], NBIQA [30], TS-CNN [45], HVS-based method [44], DenseNet121-MCF [7], Multivariate-Regression CNN (MR-CNN) [33] and Combine model [27].

The first three methods are designed for Natural-IQA, among which BRISQUE [25] and NBIQA [30] are hand-crafted feature based and TS-CNN [45] is deep feature based. Their results are obtained by adapting the codes from authors to Retinal-IQA task with Eye-Quality dataset [7]. In detail, those three Natural-IQA methods aim to regress the quality scores of natural images while Retinal-IQA in this paper aims to classify the quality of retinal images into three classes. Thus to make those Natural-IQA methods adapt to Retinal-IQA, regression is replaced with classification. Particularly, in the implementations of BRISQUE [25] and NBIQA [30], we directly predict retinal image quality label instead of regressing the quality score. They are implemented by Matlab 2014A on the platform of a PC with an i5 CPU and 16 GB RAM. In the implementation of TS-CNN [45], the regression layer with one-dimensional output is replaced with a classification layer with three-dimensional outputs and the loss is replaced by multi-class cross-entropy loss.

The last four methods are specifically designed for Retinal-IQA, among which HVS-based method [44] is hand-crafted feature based and DenseNet121-MCF [7], MR-CNN [33] and Combine model [27] are deep feature based. The results of HVS-based method [44] and DenseNet121-MCF [7] are from [7]. For fair comparison, the Acc results of HVS-based method [44] and DenseNet121-MCF [7] are not given here because they do not report their corrected result. The results of MR-CNN [33] and Combine model [27] are directly
taken from the original paper. We also reproduce the results of DenseNet121-MCF [7] and DenseNet121-RGB [7].

4.1.1 Comparisons with other methods

Table 2 reports the performances comparing with state-of-the-arts on Eye-Quality dataset [7]. Results from the original papers or reproduced by others are marked with * while those without any mark are reproduced by ours. Our reproduction of previous methods and implementation of our methods are conducted with three independent trails and the average performances are reported.

From Table 2, we have the following observations.

- Comparing with hand-crafted feature based methods, i.e. BRISQUE [25], NBIQA [30] and HVS-based method [44], deep feature based methods achieve better performances on Retinal-IQA. The possible reason is that deep features are more powerful than hand-crafted features.
- For deep feature based methods, Retinal-IQA methods significantly outperforms Natural-IQA methods, i.e. TS-CNN [45].
- Comparing with deep features based Retina-IQA method MR-CNN* [33], our Single-branch SalStructIQA achieves 88.47% in Acc which slightly outperforms MR-CNN* [33] by 0.04%. In terms of F_{score}, our Single-branch SalStructIQA achieves 86.62%, which is slightly inferior to MR-CNN* [33] by 0.31%. One thing worth to mention is that the parameters of our Single-branch SalStructIQA are approximately 1/15 of MR-CNN* [33].
- Comparing with Combine model* [27], our Single-branch SalStructIQA and Dual-branch SalStructIQA are slightly inferior to Combine model* [27] in terms of F_{score}, however, the parameters of our Single-branch SalStructIQA and Dual-branch SalStructIQA are approximately 1/4 and 1/2 of Combine model* [27], respectively.

Table 2 Comparisons of the proposed methods and state-of-the-arts on Eye-Quality [7]

| Methods            | Acc    | Pre    | Rec    | F_{score} | Param   |
|--------------------|--------|--------|--------|-----------|---------|
| Natural-IQA        |        |        |        |           |         |
| BRISQUE [25]       | 0.7692 | 0.7608 | 0.7095 | 0.7112    | –       |
| NBIQA [30]         | 0.7917 | 0.7641 | 0.7509 | 0.7441    | –       |
| TS-CNN [45]        | 0.7926 | 0.7976 | 0.7446 | 0.7481    | 1.44M   |
| Retinal-IQA        |        |        |        |           |         |
| HVS-based* [44]    | –      | 0.7404 | 0.6945 | 0.6991    | –       |
| MR-CNN* [33]       | 0.8843 | 0.8697 | 0.8700 | 0.8694    | 101.80M |
| Combine model* [27]| –      | 0.878  | 0.880  | 0.878     | ~29.3M  |
| DenseNet121-MCF* [7]| –     | 0.8645 | 0.8497 | 0.8551    | 28.26M  |
| DenseNet121-MCF [7]| 0.8722 | 0.8563 | 0.8482 | 0.8506    | 28.26M  |
| DenseNet121-RGB* [7]| –     | 0.8194 | 0.8114 | 0.8152    | 6.96M   |
| DenseNet121-RGB [7]| 0.8568 | 0.8481 | 0.8239 | 0.8315    | 6.96M   |
| Ours               |        |        |        |           |         |
| Single-branch SalStructIQA | 0.8847 | 0.8715 | 0.8645 | 0.8662    | 6.96M   |
| Dual-branch SalStructIQA | 0.8897 | 0.8748 | 0.8721 | 0.8723    | 13.92M  |

The best result is in bold whereas the second best one is underlined under each case.
Comparing with DenseNet121-MCF* [7], our Single-branch SalStructIQA outperforms it by 1.11% in $F_{score}$. Additionally, our Single-branch SalStructIQA is much lighter than DenseNet121-MCF* [7], whose parameter scale is approximately 1/4 of DenseNet121-MCF* [7]. Comparing with DenseNet121-RGB* [7], ours Single-branch SalStructIQA significantly outperforms it with the same parameter scale.

Comparing with the Single-branch one, our Dual-branch SalStructSal further improves the performances consistently while the parameter scale is double. Comparing with both Natural-IQA methods and Retinal-IQA methods listed in Table 2, our Dual-branch SalStructIQA achieves best performances.

### 4.1.2 Comparisonsofconfusionmatrices

We also report the confusion matrices of methods BRISQUE [25], NBIQA [30], TS-CNN [45], DenseNet121-MCF [7], our Single-branch SalStructIQA and Dual-branch SalStructIQA in Fig. 7. For brevity, we report the confusion matrices of those methods according to the trial achieving best $F_{score}$. From confusion matrices we have following observations:

- Natural-IQA methods BRISQUE [25], NBIQA [30] and TS-CNN [45] work well in classifying retinal images with ‘Good’ quality. But they tend to wrongly classify retinal images graded as ‘Usable’ into ‘Good’ and ‘Reject’ and classify retinal images graded as ‘Reject’ into ‘Good’ and ‘Usable’.

- Comparing with Natural-IQA methods, DenseNet121-MCF [7] improves the classification accuracy of retinal images graded as ‘Usable’.

- Comparing with Natural-IQA methods, both our Single-branch method and Dual-branch method achieve higher classification accuracies of retinal images graded as ‘Usable’ and ‘Reject’ significantly. Comparing with DenseNet121-MCF [7], ours method performs better in classification on retinal images graded as ‘Reject’, which is more important than correctly identifying the retinal images graded as ‘Usable’ and ‘Good’ in clinical.
Fig. 8  ROC comparison of different approaches. Our Dual-branch SalStructIQA has achieved the highest auc value

4.1.3 Comparison of ROC

Receiver operating characteristic (ROC) curves of several methods are shown in Fig. 8. The Area Under Curve (AUC) of the proposed Dual-branch method as shown in Fig. 8 is 0.91, which is the highest among the methods BRISQUE [25], NBIQA [30] and DenseNet121-MCF [7]. The results demonstrate that incorporating the proposed two salient structure priors into network could effectively improve the prediction performance for retinal image quality assessment.

4.1.4 Comparison of class activation maps

In addition, we show three examples of gradient-weighted class activation maps of three models, i.e., DenseNet121-RGB [7], our Single-branch SalStructIQA and Dual-branch SalStructIQA in Fig. 9. Those gradient-weighted activation maps are generated according to [36]. As our Dual-branch SalStructIQA contains two branches, i.e., the large-size salient structure branch and the tiny-size salient structure branch, thus for each input there are two class activation maps as listed in the last two columns in Fig. 9.

With regard to these three examples graded as ‘Good’, ‘Usable’ and ‘Reject’ from top to bottom, DenseNet121-RGB [7] wrongly classifies them into ‘Usable’, ‘Good’ and ‘Usable’ while our two SalStructIQA models are able to correctly classify them. Gradient-weighted class activation maps of DenseNet121-RGB [7] (the second column in Fig. 9) indicate that DenseNet121-RGB may fail to capture the most informative information for classification. On the contrary, our two models are able to capture the most informative information. As shown in the third column in Fig. 9, for the classification of images graded as ‘Good’ and ‘Usable’, our Single-branch SalStructIQA are able to capture informative
Fig. 9 Examples of gradient-weighted class activation maps of different models generated according to [36]. These three retinal images are mis-classified by DenseNet121-RGB [7] but correctly classified by our Single-branch SalStructIQA and Dual-branch SalStructIQA. Obviously, with salient structures, the models tend to learn discriminate information from salient regions for image quality assessment.

### Table 3

The influences of backbone networks and our proposed salient structure priors

| Model               | Backbone     | Acc   | Pre   | Rec   | Fscore |
|---------------------|--------------|-------|-------|-------|--------|
| Baseline            | ResNet50     | 0.8809| 0.8670| 0.8616| 0.8609 |
| +LS                 | ResNet50     | 0.8834| 0.8626| 0.8703| 0.8642 |
| +TS                 | ResNet50     | 0.8833| 0.8663| 0.8661| 0.8646 |
| +LS+TS (Single-branch) | ResNet50   | 0.8840| 0.8656| 0.8680| 0.8657 |
| +LS+TS (Dual-branch)| ResNet50     | 0.8857| 0.8694| 0.8708| 0.8695 |
| Baseline            | DenseNet121  | 0.8834| 0.8711| 0.8624| 0.8641 |
| +LS                 | DenseNet121  | 0.8851| 0.8706| 0.8661| 0.8672 |
| +TS                 | DenseNet121  | 0.9229| 0.8843| 0.8631| 0.8653 |
| +LS+TS (Single-branch) | DenseNet121 | 0.8847| 0.8715| 0.8645| 0.8662 |
| +LS+TS (Dual-branch)| DenseNet121  | 0.8897| 0.8748| 0.8721| 0.8723 |

The best result is in bold
focus on the vessels for Retinal-IQA. Comparing with DenseNet121-RGB [7] without considering the salient structures in retinal images, the ways that our proposed two models with salient structures for Retinal-IQA are much more consistent to that ophthalmologists perform quality assessment.

4.2 Analysis of backbone networks and salient structures

We further look into the influences of different backbone networks and our two proposed salient structures, i.e., large-size salient structures and tiny-size salient structures. To this end, ablation study is first conducted. Two prevalent backbone networks ResNet50 [9] and DenseNet121 [11] are chosen. For each backbone network, we separately train four models: the baseline model which is trained with $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$, the model denoted as ‘+LS’ which is trained with $[\mathbf{I}, \mathbf{M}_{LS}] \in \mathbb{R}^{H \times W \times 4}$, the model denoted as ‘+TS’ which is trained with $[\mathbf{I}, \mathbf{M}_{TS}] \in \mathbb{R}^{H \times W \times 4}$ and the model denoted as ‘+LS+TS (Single-branch)’ which is trained with $[\mathbf{I}, \mathbf{M}_{LS}, \mathbf{M}_{TS}] \in \mathbb{R}^{H \times W \times 5}$. As mentioned before, $\mathbf{I}$, $\mathbf{M}_{LS}$ and $\mathbf{M}_{TS}$ are the RGB retinal image, the mask of large-size salient structures and the mask of tiny-size salient structures respectively. The performances are reported in Table 3. The performances of Dual-branch SalStructureIQA are also listed. Obviously, DenseNet121 [11] outperforms ResNet50 [9] in Retinal-IQA. Moreover, from Table 3 we can see that (1) combining RGB retinal image with either large-size salient structures or tiny-size salient structures is able to improve the performances of Retinal-IQA, (2) our Dual-branch SalStructIQA achieves superior performances than the Single-branch one which indicates that separately learning features from $[\mathbf{I}, \mathbf{M}_{LS}]$ and $[\mathbf{I}, \mathbf{M}_{TS}]$ works better than learning features from $[\mathbf{I}, \mathbf{M}_{LS}, \mathbf{M}_{TS}]$.

4.3 Analysis of parameter settings

Generally speaking, network parameters i.e. learning rate may cause the performance of final result to some extent. Therefore, the influences of different learning rate is analyzed on dataset Eye-Quality [7]. We give some result on Table 4 and select the learning rate which has the best $F_{score}$. In Table 4, lr0.01(0.001 after 10 epoches) means that models are first trained 10 epoches with learning rate 0.01 and then trained another 10 epoches with decreased learning rate 0.001, and lr0.01(0.001 after 15 epoches) means that the initial learning rate is set to 0.01, and the learning rate is decreased to 0.001 after 15 epoches.

From Table 4, we can observe that (1) both our Single-branch SalStructIQA and Dual-branch SalStructIQA have achieved the better performance undering lr0.01(0.001 after 10 epoches); (2) with the same setting, the proposed Dua-branch SalStructIQA is superior to Single-branch SalStructIQA.

| Learning rate | Network                  | $Pre$ | $Rec$ | $F_{score}$ |
|---------------|--------------------------|-------|-------|-------------|
| lr0.01        | Single-branch SalStructIQA | 0.8715| 0.8645| 0.8662      |
| (0.001 after 10 epoches) | Dual-branch SalStructIQA | **0.8748**| **0.8721**| **0.8723**|
| lr0.01        | Single-branch SalStructIQA | 0.8718| 0.8609| 0.8648      |
| (0.001 after 15 epoches) | Dua-branch SalStructIQA | 0.8701| 0.8682| 0.8668      |

The best result is in bold
4.4 Assessment time

To demonstrate the efficiency and applicability of our proposed approaches, we compare them with other methods in terms of assessment time. Compared with other methods in Table 5, the proposed methods are much more efficient except for DenseNet121-RGB [7] and TS-CNN [45]. However, our methods have a significant performance improvement compared with these two methods. Moreover, the speed of our methods is fast, since they adopt simple filtering operations during detecting large-size and tiny-size salient structures. And on a GTX1080 Ti GPU, the average assessment time of both our Single-branch SalStructIQA and Dual-branch SalStructIQA is less than 0.5s, which meets the requirement for real-time assessment. Totally, considering the balance of efficiency and performance, our methods are more suitable for application in the real world.

4.5 Performances on the local dataset

To further verify the model generalization, we compare our approaches with DenseNet121 [11] and DenseNet121-MCF [7] on the local dataset. The dataset include 2239 good images, 194 usable images and 127 reject images. Considering the dataset is far smaller than Eye-Quality [7], we initial parameter weights from pretrained models based on Eye-Quality [7]. We also run three times for each model, and report their average results in Table 6. Compared with DenseNet121 [11], our method is obviously superior in the terms of the comprehensive evaluation metric $F_{score}$. The method DenseNet121-MCF [7] is slightly superior to our methods over $F_{score}$, however, our methods require less memory use than it. That is to say, our models are relatively light-weighted and also can obtain considerable performance.
5 Discussion

The proposed method achieves considerable performance in predicting retinal image quality grade, however, there are several aspects deserve to be further considered: (1) In this paper, we incorporate the proposed two salient structures in simple ways, it can be explored with an effective fusion way for better feature discrimination for image quality assessment; (2) In this paper, we discussed the influence of backbone and parameters setting simply in Sections 4.2 and 4.3. It is necessary to conduct more experiments to further improve the performance of our method; (3) The test result on the local dataset is not performed very well. Therefore, it is essential to built a large retinal image quality database to improve the robust and applicability of the proposed method.

6 Conclusion

Retinal image quality assessment is of great importance to both ophthalmologist and computer-aided analysis systems for disease diagnosis. In this paper, we propose a deep network architecture with salient structures priors for retinal image quality assessment. First, large-size salient structure and tiny-size salient structure priors are detected, separately. Next they are incorporated with deep network to make our model learn informative information. To achieve the embedding of salient structure priors, two network architectures termed Dual-branch SalStructIQA and Single-branch SalStructIQA are developed. The former uses the proposed two priors to separately guide the learning of two branches, and achieve superior performances than other methods on Eye-Quality dataset [7]. The latter concatenates two priors into one branch, and it is much light-weight compared with other methods but still has competitive performances on dataset Eye-Quality [7]. Moreover, the two proposed networks still have considerable performance on our local dataset. In the future, we will further explore better fusion way and verify on more data to improve the feasibility and applicability of our approaches.

Abbreviations CNN, convolution neural network; IQA, image quality assessment; Natural-IQA, natural image quality assessment; Retinal-IQA, retinal image quality assessment; SalStructIQA, our salient structure IQA method; FR, full-reference; RR, reduced-reference; NR, no-reference; LS, large-scale; TS, tiny-scale.

Symbols \(I\), retinal images; \(R_{\text{contrast}}\), contrast map; \(P_{\text{LS}}\), probability map; \(M_{\text{LS}}\), the mask of detected large-size salient structures; \(R_{\text{line}}\), the multi-scale line response; \(R_{\text{line}}^{\prime}\), the standardized line response; \(M_{\text{TS}}\), the mask of detected tiny-size salient structures; \(\text{Acc}\), accuracy; \(P\), precision; \(R\), recall; \(F_{\text{score}}\), a comprehensive evaluation index.

Acknowledgements This work of Prof. Beiji Zou is partially supported by the National Key R&D Program of China (NO. 2018AAA0102100); Dr. Qing Liu is partially supported by the National Natural Science Foundation of China (No. 62006249) and Changsha Municipal Natural Science Foundation (kq2014135). This work was supported in part by the High Performance Computing Center of Central South University.

Declarations None declared.

References

1. Achanta R, Hemami S, Estrada F et al (2009) Frequency-tuned salient region detection. In: 2009 IEEE conference on computer vision and pattern recognition, pp 1597–1604
2. Burges CJC, Shaked T, Renshaw E et al (2005) Learning to rank using gradient descent. In: Machine learning, proceedings of the 22nd international conference (ICML 2005), pp 89–96
3. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
4. Davis H, Russell S, Barriga E et al (2009) Vision-based, real-time retinal image quality assessment. In: 2009 22nd IEEE international symposium on computer-based medical systems, pp 1–6
5. Deng J, Dong W, Socher R et al (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248–255
6. Dias JMP, Oliveira CM, da Silva Cruz LA (2014) Retinal image quality assessment using generic image quality indicators. Inf Fusion 19:73–90
7. Fu H, Wang B, Shen J et al (2019) 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
8. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
9. Zhou J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
10. Huang G, Liu Z, Laurens VDM, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
11. Hunter A, Lowell JA, Habib M et al (2011) An automated retinal image quality grading algorithm. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society, pp 5955–5958
12. Jia S, Zhang Y (2018) Saliency-based deep convolutional neural network for no-reference image quality assessment. Multimedia Tools Appl 77(12):14,859–14,872
13. Kang L, Ye P, Li Y et al (2014) Convolutional neural networks for no-reference image quality assessment. In: 2014 IEEE conference on computer vision and pattern recognition, pp 1733–1740
14. Köhler T, Budai A, Kraus MF et al (2013) 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
15. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
16. Lalande M, Gagnon L, Boucher MC et al (2001) Automatic visual quality assessment in optical fundus images. In: Proceedings of vision interface, pp 259–264
17. Lee SC, Wang Y (1999) Automatic retinal image quality assessment and enhancement. In: Medical imaging 1999: image processing, international society for optics and photonics, pp 1581–1590
18. Li Q, Lin W, Gu K et al (2019) Blind image quality assessment based on joint log-contrast statistics. Neurocomputing 331:189–198
19. Li XM, Hu XW, Qi XJ et al (2021) Rotation-oriented collaborative self-supervised learning for retinal disease diagnosis. IEEE Trans Med Imaging 40(9):2284–2294
20. Liu X, Weijer JVD, Bagdanov AD (2017) Rankiqa: learning from rankings for no-reference image quality assessment. In: Proceedings of the IEEE international conference on computer vision, pp 1040–1049
21. Ma K, Liu W, Liu T et al (2017a) Dipiq: blind image quality assessment by learning-to-rank discriminable image pairs. IEEE Trans Image Process 26(8):3951–3964
22. Ma K, Liu W, Zhang K et al (2017b) End-to-end blind image quality assessment using deep convolutional neural networks. IEEE Trans Image Process 27(3):1202–1213
23. MacGillivray TJ, Cameron JR, Zhang Q et al (2015) Suitability of uk biobank retinal images for automatic analysis of morphometric properties of the vasculature. PLoS One 10(5):e0127914
24. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 21(12):4695–4708
25. Moorthy AK, Bovik AC (2011) Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans Image Process 20(12):3350–3364
26. Muddamsetty SM, Moeslud SB, Tuck SE, Nordsletten DA, Barriga E et al (2012) Vision-based retinal image quality assessment using generic image quality indicators. Inf Fusion 19:73–90
27. Niu YH, Gu L, Zhao YT, Lu F (2022) Explainable diabetic retinopathy detection and retinal image generation. IEEE J Biomed Health Inform 26(1):44–55
30. Ou FZ, Wang YG, Zhu G (2019) A novel blind image quality assessment method based on refined natural scene statistics. In: 2019 IEEE international conference on image processing (ICIP), pp 1004–1008
31. Pachade S et al (2021) Retinal fundus multi-disease image dataset (rfmid): a dataset for multi-disease detection research. In Data 6(2):14
32. Parashar D, Agrawal D (2021) 2-D Compact Variational Mode Decomposition-Based Automatic Classification of Glaucoma Stages From Fundus Images. IEEE Trans Instrum Meas 70:1–10
33. Raj A, Shah NA, Tiwari AK et al (2020) Multivariate regression-based convolutional neural network model for fundus image quality assessment. IEEE Access 8:57,810–57,821
34. Saad MA, Bovik AC, Charrier C (2012) Blind image quality assessment: a natural scene statistics approach in the dct domain. IEEE Trans Image Process 21(8):3339–3352
35. Saha SK, Fernando B, Cuadros J et al (2018) Automated quality assessment of colour fundus images for diabetic retinopathy screening in telemedicine. J Digit Imaging 31(6):869–878
36. Selvaraju RR, Cogswell M, Das A et al (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626
37. Shen Y, Fang R, Sheng B et al (2018) Multi-task fundus image quality assessment via transfer learning and landmarks detection. In: International workshop on machine learning in medical imaging. Springer, pp 28–36
38. Shen Y, Sheng B, Fang R et al (2020) Domain-invariant interpretable fundus image quality assessment. Medical Image Anal 61(12):101654
39. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv: 14091556
40. Szegedy C, Vanhoucke V, Ioffe S et al (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
41. Tennakoon R, Mahapatra D, Roy P et al (2016) Image quality classification for dr screening using convolutional neural networks. In: Ophthalmic medical image anal. 3rd int. workshop, pp 113–120
42. Usher D, Himaga M, Dumskyj M et al (2003) Automated assessment of digital fundus image quality using detected vessel area. In: Proceedings of medical image understanding and analysis. Citeseer, pp 81–84
43. Wang Z, Bovik AC (2006) Modern image quality assessment. Synth Lect Image Video Multimed Process 2(1):1–156
44. Wang S, Jin K, Lu H et al (2015) Human visual system-based fundus image quality assessment of portable fundus camera photographs. IEEE Trans Medical Imaging 35(4):1046–1055
45. Yan Q, Gong D, Zhang Y (2018) Two-stream convolutional networks for blind image quality assessment. IEEE Trans Image Process 28(5):2200–2211
46. Yang S, Jiang Q, Lin W et al (2019) Sgdnet: an end-to-end saliency-guided deep neural network for no-reference image quality assessment. In: Proceedings of the 27th ACM international conference on multimedia, pp 1383–1391
47. Yang XH, Li F, Liu HT (2020) Deep feature importance awareness based no-reference image quality prediction. Neurocomputing 401:209–223
48. Yu F, Sun J, Li A et al (2017) Image quality classification for dr screening using deep learning. In: 2017 39th Annual international conference of the ieee engineering in medicine and biology society (EMBC), pp 664–667
49. Zago GT, Andreão RV, Dorizzi B et al (2018) Retinal image quality assessment using deep learning. Comput Biol Med 103:64–70

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.