# Image Quality Classification for DR Screening Using Deep Learning

FengLi Yu, Jing Sun, Annan Li, Jun Cheng, Cheng Wan, Jiang Liu

*Abstract*— **The quality of input images significantly affects the outcome of automated diabetic retinopathy (DR) screening systems. Unlike the previous methods that only consider simple low-level features such as hand-crafted geometric and structural features, in this paper we propose a novel method for retinal image quality classification (IQC) that performs computational algorithms imitating the working of the human visual system. The proposed algorithm combines unsupervised features from saliency map and supervised features coming from convolutional neural networks (CNN), which are fed to an SVM to automatically detect high quality vs poor quality retinal fundus images. We demonstrate the superior performance of our proposed algorithm on a large retinal fundus image dataset and the method could achieve higher accuracy than other methods. Although retinal images are used in this study, the methodology is applicable to the image quality assessment and enhancement of other types of medical images.**

***Keywords— image quality classification, saliency map, convolutional neural networks.***

## I. INTRODUCTION

Digital fundus photography is a common procedure in ophthalmology and provides non-invasive diagnostic information for retinal pathologies, such as diabetic retinopathy (DR), glaucoma, age-related macular degeneration, and vascular abnormalities. The symptoms of the above diseases are well defined and visible in fundus images [1]. Research communities have put great effort towards the automation of a computer screening system which is able to promptly detect DR in fundus images [2, 3]. The evaluation of fundus image quality involves a computer-aided retinal image analysis system that is designed to assist ophthalmologists to detect eye diseases. Consequently, automated evaluations of ophthalmopathy can be performed to support the diagnosis of doctors [4]. However, the success of these automatic diagnostic systems heavily relys on the input image quality. In reality, due to some unavoidable disturbances in the stage of image acquisition, e.g. the operator's expertise, the type of image acquisition equipment, the situation of different individuals, the images we received will become blurred and affect the follow-up diagnosis. Therefore, the process of image quality assessments (IQA) plays an extremely important role in the computer-aided DR screening system. Figure 1 shows four instances of poor quality images which restricted the subsequent analysis and DR diagnosis.

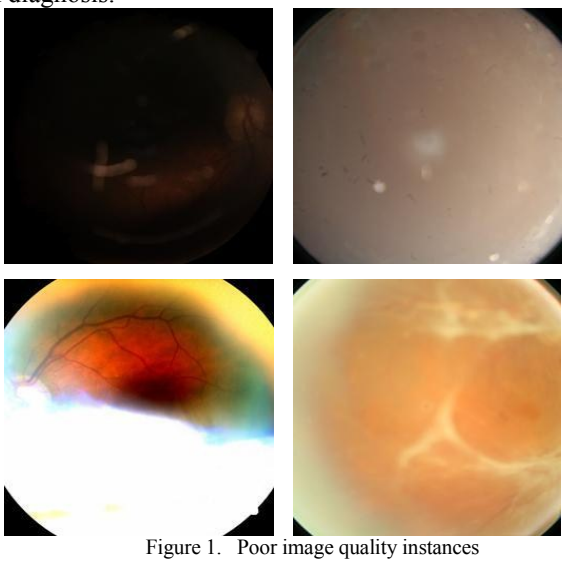


Figure 1. Poor image quality instances

As we all know that many classification methods used so far for image quality classification of digital fundus images rely on some kind of hand-crafted features that are based on either geometric or structural quality parameters which do not generalize well to new datasets. In the context of retinal image analysis, image quality classification is used to determine whether an image is useful or the quality of a retinal image is sufficient for the subsequent automated diagnosis. Different approaches are presented in literature for image quality classification, Lee et al. [5] use a quality index Q which is calculated by the convolution of a template intensity histogram to measure the retinal image quality. Lalonde et al. [6] adopt the features which are based on the edge amplitude distribution and the pixel gray value to automatically assess the quality of retinal images. Traditional feature extraction methods with low computational complexity only can obtain some characteristic that represents image quality rather than always acquiring diversity factors that affect image quality. For the past decade, a deep architecture [7] has gained a great attention in various fields and deep learning is a new breakthrough due to its representational power. Different from the traditional handcraft-feature extracted methods, a deep learning model can find the hidden or latent high-level information inherent in the original features, which can be helpful to build a more robust model.

Image saliency is an important visual feature to an image and it reflects the degree of human eyes' attention to some special regions of an image. People assess image quality

* Research supported by Chinese National Natural Science Foundation (GBA1604401); Natural Science Foundation of Jiangsu Province (PAF16022); Priority Academic Program Development of Jiangsu Higher Education Institutions.

F. L. Yu is with Nanjing University of Aeronautics and Astronautics and Ningbo Institute of Materials Technology and Engineering, China (email: email_yufl@163.com ).

J. Sun, C. Wan are with Nanjing University of Aeronautics and Astronautics, China (email: sunjing3747@163.com, wanch@nuaa.edu.cn).

J. Cheng, A. is with the Institute for Infocomm Research, A*STAR, Singapore (email: jcheng@i2r.a-star.edu.sg, Li.annan.cn@gmail.com).

J. Liu is with Ningbo Institute of Material Technology and Engineering, Chinese Academy of Sciences, China (email: jimmyliu@nimte.ac.cn).

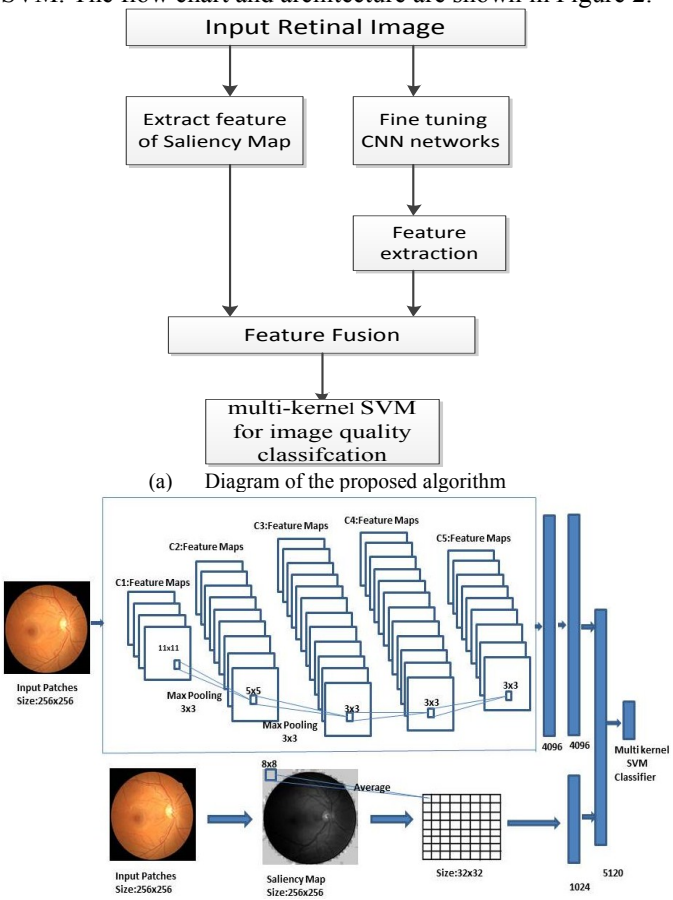
A. N. Li is with School of Computer Science and Engineering, Beihang University, China (email: liannan@buaa.edu.cn)

relying on the human visual system to identify whether a retinal image is in good quality. Achanta et al. [8] calculate the saliency map of an image and extract features from corresponding saliency map with full-resolution, which is different from traditional methods that make the problem of saliency transformed into the detection of the special nature of the target, such as color, brightness and texture features etc. Achanta use global saliency map that can obtain global information related to image quality.

In the paper, we propose a novel method for retinal image quality assessment that uses computational algorithms imitating the work of the human visual system. In our work, we extract unsupervised features from visual saliency maps as in [8] and fuse them with the supervised features learned from CNN network. In the fusing features step, the feature vectors from saliency and CNNs are firstly normalized respectively. Our analysis shows that the proposed method can learn the necessary information relevant for IQC and the result on a relatively large dataset shows that the overall method can achieve high accuracy.

## II. PROPOSED TECHNIQUE

A new image quality classification method proposed in the paper makes the features extracting from saliency map and the features learned from the trained CNN networks fused. The fused features are used for classification and tested in SVM. The flow chart and architecture are shown in Figure 2.




(a) Diagram of the proposed algorithm

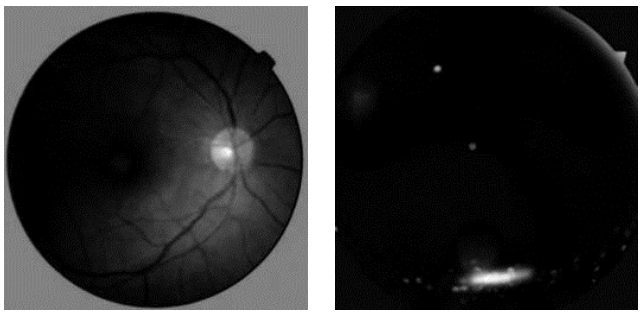
(b) The architecture of the proposed algorithm

Figure 2. Flow chart and architecture of the proposed algorithm

Saliency map reflects the degree to which a particular region is different from its neighbors with respect to image features and CNN could directly learn the high-level information hidden in the picture's low-level features. Finally, the SVM classifier is trained based on the fused features from unsupervised information of saliency map and supervised information of features learned from CNN networks.

### A. Features of saliency map

Saliency is defined to describe the best part of the image which can attract users and could be a visible performance. The blurred saliency map is produced because there are some certain losses to the image frequency in the scale space. The saliency map selected in this paper was with relatively complete image information of full-resolution with well-defined boundaries of salient objects [9], and the discriminatory information of the original image such as edge, texture was clearly extracted. The saliency maps of good quality and poor quality fundus images in this paper are shown in Figure 3.



(a) Good quality (b) Poor quality

Figure 3. Saliency maps for good quality image and poor quality image

In this paper, the size of all the original images is resized to $256\times256$ as input images, and we extract the saliency maps employing the method in [8] for the resized retinal images. The extracted saliency maps are of full-resolution with size of $256\times256$. Next, we use a window with size of $8\times8$ sliding on the saliency map with non-overlapping type to extract salient features from saliency map for input images. As a result, for an input retinal image, the saliency map is divided into $32\times32$ blocks. The mean pixel value of each block is calculated in the region within the window and these mean values are composited to the feature vector from the saliency map. Thus a saliency map feature vector of $1024(32\times32)$ dimensions is obtained from a retinal image.

### B. Fine tune CNNs

In practice, deep convolutional neural networks (DCNN) would not be randomly initializing trained from the beginning completely for the reason that the dataset with sufficient size to meet the needs of deep networks are quite rare. As a substitute, it is common to pre-train a deep CNN based on a large dataset, and the weight of the trained DCNN are used as initial settings or as fixed feature extractors for related tasks. In this work, the Alexnet network is trained by the method of transfer learning, then remove the output layer and use the trained CNN as feature extractor for SVM classifier. Alexnet [9] architecture consists of five convolutional layers, three pooling layers, two local response normalization layers and two fully connected layers. The network was originally trained on natural images from the ImageNet competition with more than one million images and the trained network weight are available through caffe model. Before training Alexnet by

using the retinal images, we firstly use the weight that trained on the Imagenet as the initial weighting of Alexnet. This process is so-called transfer learning.

In this paper, the training data is directly put into the network with pre-training weight parameters and the training is carried on a workstation with a NVIDIA-GTX1080 GPU. The final connection layer of Alexnet is taken as the feature vector. In Alexnet, the output of the final fully connection layer (fc7) is a vector of 4096 dimensions which is the feature vector extracted from CNN networks. The data for fine-tuning the networks is defined in section III and those extracted features are fused with the features extracted from saliency map and then those fused features are used to train the SVM classifier for the image quality classification and assessment.

### C. Feature Fusion

Before fusing the features, it is necessary to respectively normalize the features learned from CNN networks and the features coming from saliency map since the two types of features are two different features. The two feature vectors are normalized between 0 and 1 respectively and then the two feature vectors are merged into a vector with 5120 dimensions which is fed to the SVM classifier in the next step.

### D. Image quality classification

In this paper, we adopt the multi-kernel SVM [7] [10] for the image quality classification. For the linear SVM, a support vector machine constructs a hyperplane or sets of hyperplanes in a high-level or infinite-dimensional space, which can be used for classification, regression and other tasks, that is given a set of data $\{x_i, y_i\}_{i=1}^{N}$, where $y_i = \{0,1\}$ is the label respect to the sample of $x_i$ and N is the number of samples to train the SVM classifier. We are able to find the maximum-margin hyperplane $WX + b = 0$ that can linearly separate the data $x_i$. However in practice, it often happens that the sets to some occasion are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space by a nonlinear kernel function, presumably making the separation easier in that space, that is given the features selected from training samples $\{x_i, y_i\}_{i=1}^{N}$, our objective is to learn a function of the form $f(x) = w^t \Phi_d(x) + b$ with the kernel $k_d(x_i, x_j) = \Phi_d(x_i)\Phi_d(x_j)$ representing the dot product in feature space $\phi$ parameterized by $d$. The decision function of the multi-kernel SVM is defined as follows:

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i \sum_{k=1}^{K} \beta_K K_K(x_i, x) + b) \quad (1)$$

where $K_K(x_i, x)$ is the k-th kernel function, $\alpha_i$ is the Lagrangian multiplier, $\beta_K$ is the weight coefficient of the respect kernel function and b is the bias.

The classification effect of the algorithm in this paper is evaluated by the proposed technique in the next step. The quality classification of the retinal fundus images is classified with multi-kernel SVM (we use RBF kernel in this work). The result of the classification is compared with the results of traditional methods which proved that the method proposed in the paper achieved the best performance of highest accuracy with highest sensitivity and specificity.

## III. EXPERIMENTAL RESULTS

The dataset used to verify the effectiveness of the proposed method is provided by the Kaggle coding website [11] (http://www.kaggle.com) and contains over 80000 images of diabetic retinopathy. The data we used is based on the training set with 3000 original samples and the test set with 2200 original samples randomly selected (the proportion of the poor quality images in all images of kaggle is about 0.04). For the training set there are 1715 samples with label 1 and 1285 samples with label 0. The randomly selected test set contains 1101 samples with label 1 and 1099 samples with label 0. All images are tagged by the professionals to identify the labels of the dataset, in which 1 represents the image of good quality with the attributes for carrying on the following DR screening and analysis, and 0 is the label that stands for the poor quality images with the opposite attributes. The resolution of the original sample is $2592 \times 1944$ with plenty of redundant information if it is used to train CNNs directly. In the paper, the original images are resized to $256 \times 256$. These resized retinal images are used for salient features extraction and the input of CNN networks.

Results of our method denoted as SM+Alexnet+SVM (SM is Saliency Map) are compared with the following methods: HOG+SVM where the hog features extracted from the retinal image to train the SVM classifier with multi-kernel, SM+SVM where using the features from Saliency map to train the SVM classifier, Alexnet+SVM denoting that support vector machine uses features from the last fully connection layer (fc7) of Alexnet for classification and HOG+Alexnet+SVM representing fusing the hog features and features learned from CNNs of Alexnet in the same way as our proposed method for predicting the image quality classification. In this work, we extract HOG features with 2304 dimensions of each retinal image. The dimension of fused feature vector with HOG features and features learned from CNNs is 6400 for one retinal image. The experiment results of this work are shown in TABLE 1 and Figure 4.

TABLE 1. Sensitivity (SEN), Specificity (SPE), Accuracy (ACC) and Area Under Curve (AUC) for different methods

| Algorithm | Overall Acc. | SEN | SPE | AUC |
|---|---|---|---|---|
| **HOG+SVM** | 89.64% | 0.9445 | 0.8556 | 0.9599 |
| **SM+SVM** | 80.41% | 0.8854 | 0.7748 | 0.8898 |
| **Alexnet+SVM** | 94.80% | 0.9572 | 0.9237 | 0.9794 |
| **HOG+Alexnet+SVM** | 94.93% | 0.9636 | 0.9246 | **0.9828** |
| **SM+Alexnet+SVM** | **95.42%** | **0.9663** | **0.9310** | 0.9819 |

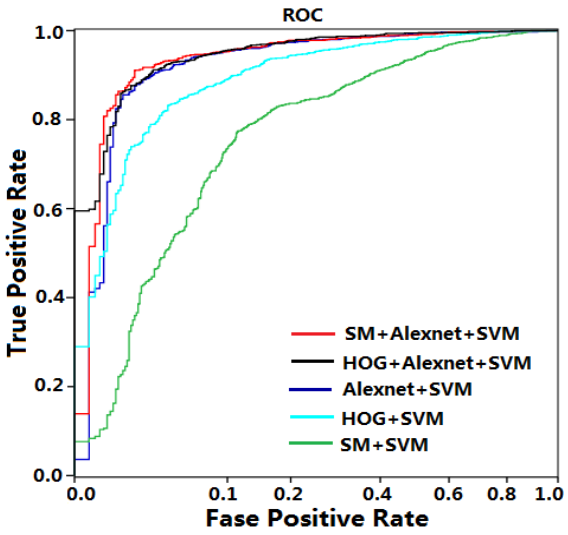
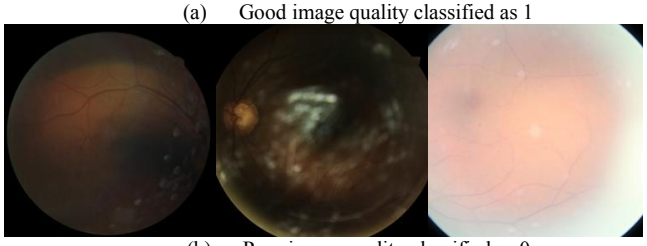
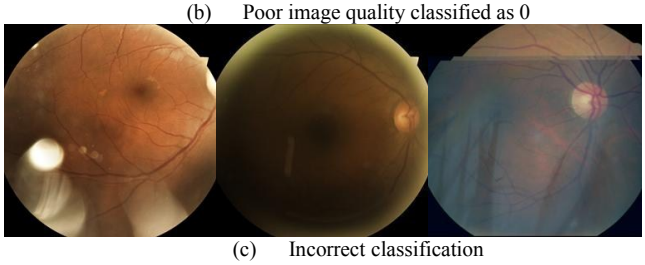



Figure 4. The ROC curves of different methods

(a) Good image quality classified as 1



(b) Poor image quality classified as 0



(c) Incorrect classification

Figure 5. Nine examples from the classification results

The classification results evaluated using the test set are presented in Table 1. The results show that fusing two kinds of features and concatenating them in a single feature vector pare better than only using a single kind of features to train the SVM classifier. The proposed algorithm has achieved accuracy of 95.42%. We obtain sensitivity (0.9663) and specificity (0.9310) which outperforms current methods we utilized on our dataset. This indicates that the proposed method fusing the saliency map features and the features learned from convolutional neural networks has been able to learn the necessary information for image quality classification in retinal images.

Figure 5 shows some classification results of fundus images. (a) and (b) show the correct classification results that the good quality images are classified as 1 and the poor quality images are classified as 0. (c) shows the incorrect classification results that good quality images are classified as 0 and the poor quality images are classified as 1. It is worth noting that despite a few erroneous labels, our approach could learn a reliable feature representations and separates different image classes.

## IV. CONCLUSIONS

A new algorithm of retinal image quality classification was proposed by fusing unsupervised information from visual saliency map and supervised information coming from trained CNNs in this paper. Our key contributions are the use of deep learning and leveraging the functioning of the human visual system for image quality classification. The accuracy of the final classification was improved based on the fusion of saliency map and the features learned from CNN that would be a very important step in the diagnostic process of large-scale diabetic retinopathy. We demonstrated the proposed algorithm on a large scale retinal image dataset and the results showed that fusing two kinds of features from saliency map and CNNs outperformed others methods using the learned knowledge of the baseline HOG method and Saliency map method.

## REFERENCES

[1] H. Jelinek, M. J. Cree, "Automated Image Detection of Retinal Pathology", Crc Press , 2009.

[2] D. Mahapatra, "Retinal Image Quality Classification Using Neurobiological Models Of The Human Visual System", 2016 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2016.

[3] A. D. Fleming, D. Alan, "Automated grading for diabetic retinopathy: A large-scale audit using arbitration by clinical experts", British Journal of Ophthalmology, vol. 94, no. 12, pp. 1606-1610, 2010.

[4] R. Tennakoon, D. Mahapatra, P. Roy, S. Sedai, R. Garnavi, "Image quality classification for DR screening using convolutional neural networks", 2016 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2016.

[5] S. C. Lee, C. Samuel, Y. Wang, "Automatic retinal image quality assessment and enhancement, " Proc Spie, vol. 3661, pp. 1581-1590, 1999.

[6] M. Lalonde, L. Gagnon, M. C. Boucher, "Automatic visual quality assessment in optical fundus images", Proceedings of Vision Interface, vol.18, no.4, pp. 437-450, 2001.

[7] H. I. Suk, D. Shen, "Deep Learning-Based Feature Representation for AD/MCI Classification", Medical Image Computing & Computer-assisted Intervention: Miccai International Conference on Medical Image Computing & Computer-assisted Intervention Med Image Comput Comput Assist Interv, vol. 16, pp. 583-90, 2013.

[8] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, "Frequency-tuned salient region detection", IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2009. vol. 22, pp. 1597-1604, 2009.

[9] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, "Frequency-tuned salient region detection", IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2009. vol. 22, pp. 1597-1604, 2009.

[10] M. Varma, B. R. Babu, " More generality in efficient multiple kernel learning", 2009 International Conference on Machine Learning (ICML), vol. 134, 2009.

[11] H. Pratt, F. Coenen, D. M. Broadbent, "Convolutional Neural Networks for Diabetic Retinopathy", Procedia Computer Science, vol. 90, pp. 200-205, 2016.