Retinal Image Quality Classification Using Fine-Tuned CNN

Jing Sun[1], Cheng Wan[1](✉), Jun Cheng[2], Fengli Yu[1], and Jiang Liu[3]

[1] Nanjing University of Aeronautics and Astronautics, Nanjing, China
wanch@nuaa.edu.cn
[2] Institute for Infocomm Research, A*STAR, Singapore, Singapore
[3] Ningbo Institute of Material Technology and Engineering,
Chinese Academy of Sciences, Ningbo, China

Abstract. Retinal image quality classification makes a great difference in automated diabetic retinopathy screening systems. With the increase of application of portable fundus cameras, we can get a large number of retinal images, but there are quite a number of images in poor quality because of uneven illumination, occlusion and patients movements. Using the dataset with poor quality training networks for DR screening system will lead to the decrease of accuracy. In this paper, we first explore four CNN architectures (AlexNet, GoogLeNet, VGG-16, and ResNet-50) from ImageNet image classification task to our Retinal fundus images quality classification, then we pick top two networks out and jointly fine-tune the two networks. The total loss of the network we proposed is equal to the sum of the losses of all channels. We demonstrate the super performance of our proposed algorithm on a large retinal fundus image dataset and achieve an optimal accuracy of 97.12%, outperforming the current methods in this area.

Keywords: No-reference image quality assessment (NR-IQA) · Convolutional neural networks (CNN) · Retinal image · Fine-tuning

1 Introduction

Retinal fundus images play an important role in ophthalmology diagnosis. In screening systems for diseases such as diabetic retinopathy (DR), glaucoma, age-related macular degeneration (AMD), and vascular abnormalities, a clear fundus image is a prerequisite for the right diagnosis of the disease. Research communities have put great efforts towards the automation of computer screening systems which are able to promptly detect DR in fundus images. The success of these automatic diagnostic systems heavily rely on the quality of input images. However, in reality, due to some unavoidable disturbances, for instance, differing lighting condition, the type of image acquisition equipment, the situation of different individuals, the images we acquired will be blurred and affect the final accuracy of diagnosis. Consequently, it is indispensable to conduct image quality assessment (IQA) in the computer-aided screening system for ophthalmology diagnosis. Figure 1 shows four instances of poor quality images which will restrict the

M.J. Cardoso et al. (Eds.): FIFI/OMIA 2017, LNCS 10554, pp. 126–133, 2017.
DOI: 10.1007/978-3-319-67561-9_14

subsequent analysis and DR diagnosis. These images are caused by occlusion, patients movements, underexposure or overexposure.

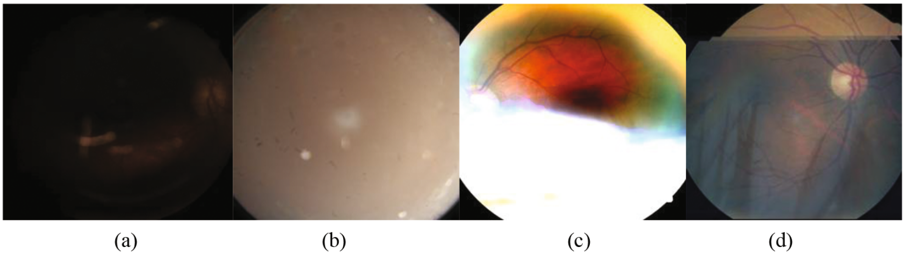


Fig. 1. Four instances of poor quality images in the retinal fundus image dataset.

Subjective evaluation and objective evaluation are two existing image quality evaluation methods [1]. In subjective method, quality is evaluated by organized groups of human observers to mark the distorted images, which is time-consuming and expensive. In general, objective image quality measures can be classified into three categories: full reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. But, in practical applications, ideal image selected as the reference image is often not available or it costs too much, so NR-IQA is desirable. Many algorithms have been proposed in the literature for no-reference retinal fundus images quality assessment [2]. Earlier methods adopt hand-crafted features. Lee et al. [3] 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. [4] adopt the features which are based on the edge amplitude distribution and the pixel gray value to automatically assess the quality of retinal images. Yu et al. [5] propose a no-reference image quality assessment method to extract features and introduce the support vector machine (SVM) into image quality assessment. All these methods do not generalize well to a new dataset since they rely on some kind of hand-crafted features that are based on either geometric or structural quality parameters.

For the past decade, a deep architecture [6] has gained a great attention in various fields and convolutional neural networks 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. [7–9] propose a new method for no-reference image quality assessment, and the structure of Le Kang's CNN has one convolutional layer with max and min pooling, two fully connected layers and an output node. However, these methods do not apply to retinal fundus images. [8, 9] leverage on learned supervised information using convolutional neural networks achieving high accuracy. Ruwan Tennakoon et al. adopt a shallow CNN architecture learning features for image quality classification, and use transfer learning achieving the same classification accuracy but they only fine-tune the AlexNet [10] architecture. Mahapatra D et al. propose a CNN architecture with five layers of convolution and max pooling operations. It needs a huge number of data to train the network otherwise it is going to lead to overfitting.

In this paper, we aim to conduct accurate classification for retinal fundus images quality. In our work, we explore four CNN architectures (AlexNet, GoogLeNet, VGG-16, and ResNet-50) from ImageNet image classification task to our Retinal fundus images quality classification, then we pick top two networks out and jointly fine-tune the two networks. The total loss of the network we proposed is equal to the sum of the losses of all channels. Our analysis shows that the proposed method can learn the necessary information relevant for IQA, and we demonstrate the superior performance of our proposed algorithm on a large retinal fundus image dataset.

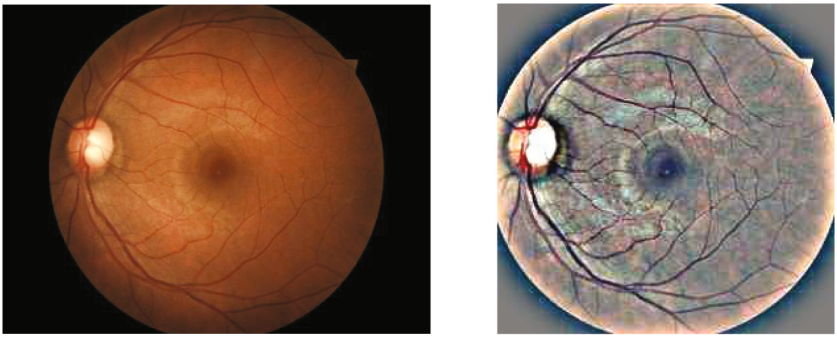
2 Method

2.1 Image Preprocessing

The resolution of the original sample is 2592 × 1994 to 4752 × 3168. First, all the images are resized to 256 × 256 pixels. And then, in order to avoid the negative effects of different conditions such as lighting on the fundus images, the images are normalized as follows:

$$I(x,y) = \alpha I^{o}(x,y) + \beta Gaussian(x,y,\omega) * I^{o}(x,y) + \gamma \tag{1}$$

Where * denotes the convolution operator, $Gaussian(x,y,\omega)$ represents the Gaussian filter with a standard deviation of ω, and the size of the Gaussian lowpass filter is $1 + floor(1 \times \omega)$. Where $floor(X)$ called the greatest integer function gives the largest integer less than or equal to X. The value of $\alpha, \beta, \omega, \gamma$ are designed empirically as $\alpha = 4, \beta = -4, \omega = 10, \gamma = 128$ respectively. In addition we clip the images to 90% size to reduce the black space on both sides of the retinal fundus images. We will evaluate the effect of image preprocessing in Sect. 3 showing that the preprocessed dataset achieves higher classification accuracy than the original dataset. The preprocessed images are shown in Fig. 2.



(a) Original image

(b) Preprocessed image

Fig. 2. One example from the training set.

2.2 Data Augmentation

Data augmentation is widely used in training a robust CNN network. We process the images by image rotation and horizontal reflections to increase the number of images. The training set are augmented with: random rotation 0–360°, random horizontal and vertical flips and random horizontal and vertical shifts, while the test set is only preprocessed without any augmentation. Through these operations, the training set is increased about 8 times.

2.3 Network Architecture

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. Therefore, it is common to pre-train a deep CNN based on a large dataset, and the weights of the trained DCNN are used as initial setting. In this work, the networks we used are all trained by the method of transfer learning. The four state-of-the-art CNN architectures and the Jointly Fine-tuned CNN model we proposed used for retinal fundus image quality classification are outlined below.

AlexNet: The AlexNet, proposed in [10] achieves significantly great performance in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 and wins by a large margin with the next non-CNN method. The network consists of 5 convolutional layers, maxpooling layers, dropout layers, and 3 fully connected layers.

GoogLeNet: This is an architecture used by Szegedy et al. [11], which uses several "Inception" modules to create a deeper network with 22 layers while having much fewer parameters than other networks such as VGG and AlexNet.

VGG-16: The VGG-16 only uses 3 × 3 filters in convolutional layers and combine them as sequence of convolution to emulate the effect of lager receptive fields and decrease the number of parameters. Overall, VGG-16 is made up of 13 convolutional layers, five maxpooling layers and three fully-connected layers.

ResNet-50: ResNet, the winner of ILSVRC2015 with an incredible error rate of 3.6%, presents residual learning framework that each layer consists of a residual block and a skip connection bypassing to ease the training of networks. This architecture substantially deeper than those used previously with 50, 101 or 152 layers. In this paper, we evaluate the performance of ResNet-50. ResNet-50 has six modules called conv1, conv2 x, conv3 x, conv4 x, conv5 x and fc. Conv1 is a convolutional layer. Conv2 x, conv3 x, conv4 x and conv5 x consist of residual blocks with the number of 3, 4, 6, 3 respectively. And fc is a fully-connected layer.

Jointly Fine-tuned CNN: We pick top two networks (GoogLeNet and VGG-16) out and jointly fine-tune the two networks. This method was proposed in [12]. The total loss (Loss_All) of the network we proposed is equal to the sum of the losses of all channels. It is given by:

$$\text{Loss\_All} = \text{Loss\_V} + \text{Loss\_G} \tag{2}$$

Where Loss_V denotes the loss of VGG-16 and Loss_G denotes the loss of GoogLeNet. Since all networks update weights in parallel through backpropagation according to the total loss, they all influence each other's weight and bias values. The Jointly Fine-tuned CNN architecture we proposed is illustrated in Fig. 3. Through the two-channel CNN architecture, we get two accuracy rates: GoogLeNet accuracy rate and VGG-16 accuracy rate of Jointly Fine-tuned CNN architecture. We use JCNN_GoogLeNet Acc and JCNN_VGG-16 Acc denote them respectively.

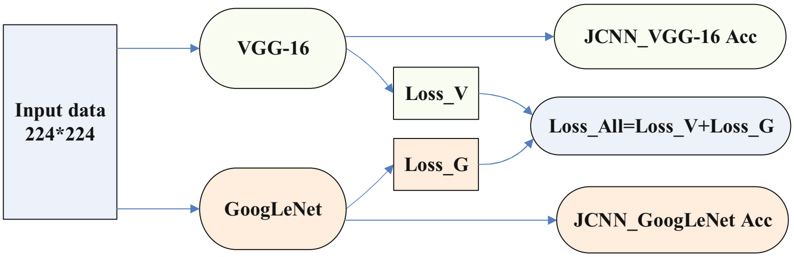


Fig. 3. The jointly fine-tuned CNN architectures.

2.4 Training

In this work, we train networks by the method of full fine-tune. It is done by removing the last fully connected layer and being replaced by a new one with 2 outputs and the learning rate of the last fully connected layer is increased ten times. In full fine-tune, the learning rate of every layer is left untouched except the last one. The purpose is to make the network has a good initial setting and iterate new data for the new fully connected layer for better learning. For AlexNet, GoogLeNet, VGG-16, and ResNet-50 architectures, the training data is directly put into the networks with pre-training weight parameters and the training process is carried on a workstation with a NVIDIA-GTX1080 GPU. The size of input data is 227 × 227 for AlexNet, and other networks is 224 × 224.

For the Jointly Fine-tuned CNN architectures, we fine-tune the GoogLeNet and VGG-16 networks at the same time, and each channel of CNN has the same data of input. The final loss is equal to the sum of the losses of all channels. It allows the backward propagation training method to broadcast the classifier gradients to all networks.

3 Experimental Result

The dataset used to verify the effectiveness of the proposed method is provided by the Kaggle coding website [13] (http://www.kaggle.com). The images in the dataset come from different models and types of cameras. Some images are shown as one would see the retina anatomically. Others are shown as one would see through a microscope condensing lens. It contains over 80000 images of diabetic retinopathy and a resolution of 2592 × 1994 to 4752 × 3168, but the proportion of the poor quality images in all images of Kaggle is small. We randomly select 2894 original samples and 2170 original samples from the dataset as training set and test set respectively. For the training set there are 1607 samples with label 1 and 1287 samples with label 0. After data augmentation, total 26046 images are used to train the CNN. The test set contains 1085 samples with label 1 and 1085 samples with label 0. All images are tagged by the professionals including doctors and experts in fields concerned based on if we can make diagnosis with the image, and the labels are determined under the majority's rule, in which label 1 represents the image with good quality and the attributes for carrying on the following DR screening and analysis, and label 0 stands for the poor quality images with the opposite attributes. The experiment results of this work are shown in Tables 1 and 2. We evaluate the performance of four state-of-the-art networks and the methods we proposed (denoted by JCNN) in this work.

Table 1. Accuracy (Acc) and area under curve (AUC) for different methods.

Algorithm	Acc	AUC
AlexNet	96.53%	0.993
GoogLeNet	97.04%	0.994
VGG-16	96.87%	0.995
ResNet-50	96.20%	0.992
JCNN_GoogLeNet	97.00%	0.995
JCNN_VGG-16	**97.12%**	**0.995**

Table 2. Comparison of image quality classification accuracy rate of GoogLeNet, GoogLeNet-NP, GoogLeNet-NA.

Algorithm	GoogLeNet	GoogLeNet-NP	GoogLeNet-NA
Accuracy	**97.04%**	96.12%	96.49%

We select the optimal accuracy rate (JCNN_VGG-16 Acc) as the final result of the JCNN architecture we proposed. The results show that jointly fine-tuning two-channel CNN architecture can achieve better accuracy than only fine-tuning a single convolutional neural networks, and GoogLeNet is superior to the other single channel networks. JCNN_GoogLeNet Acc achieves 97.00% which is close to GoogLeNet Acc, and JCNN_VGG-16 has achieved optimal accuracy of 97.12%, increasing 0.25% relative to VGG-16. Furthermore, all the fine-tuned networks have good performance, outperforming current methods we utilized on our dataset. This indicates that the knowledge learned from natural image can still transfer to make medical image quality classification effectively. The methods we proposed has been able to learn the necessary information for image quality classification in retinal images from convolutional neural networks.

For the Jointly Fine-tuned CNN (VGG16 + GoogLeNet), it makes the best network. This is a sensible choice because (1) the two network are among the best available networks, and (2) they are constructed based on two different architectural assumptions, making them relatively uncorrelated from the misclassification behavior standpoint.

We use GoogLeNet as the default architecture and evaluate the impact of image preprocessing and data augmentation. GoogLeNet has fewer parameters and higher accuracy than other single channel networks. Table 2 illustrates the accuracy rates of GoogLeNet, GoogLeNet-NP (model without preprocessing), GoogLeNet-NA (the model without data augmentation). We find that with the help of good preprocessing, the accuracy of the model increases about 1%. Data Augmentation is beneficial in our experiments, as evidenced by GoogLeNet(97.04%) versus GoogLeNet-NA (96.49%).

Figure 4 shows some classification results of fundus images. (a) and (b) show the correct classification results that the good-quality-images are classified as 1. (c) and (d) show the poor-quality-images are classified as 0. (e, f) and (g, h) shows the incorrect classification results that good-quality-images are classified as 0 and the poor-quality-images are classified as 1, respectively. It is worth noting that despite a few erroneous labels, our approach could learn a reliable feature representations and separates different image classes.

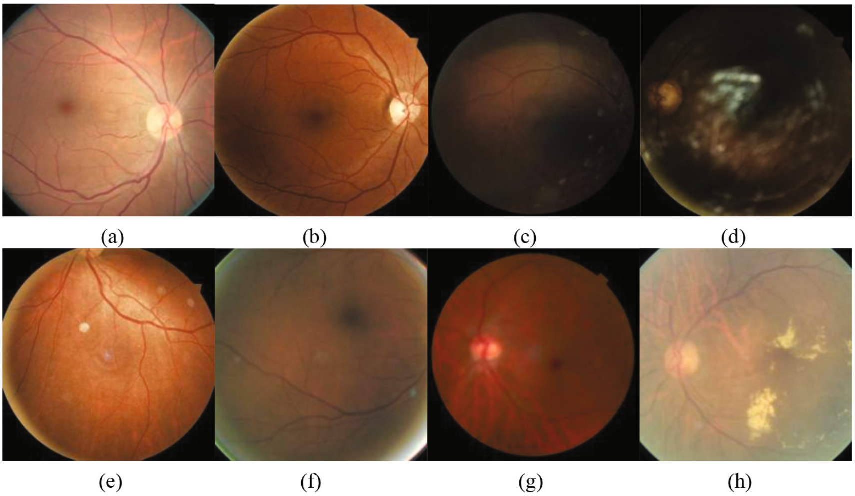


Fig. 4. Eight examples from the classification results.

4 Conclusions

In this paper, we evaluate the performance of different CNN architectures in retinal image quality classification and extensively evaluate two important factors on CNN architectures, preprocessing, data augmentation. It is evident from the results that the GoogLeNet architecture fine-tuned from ImageNet, with good image preprocessing and data augmentation performs better accuracy in the four state-of-the-art CNN architectures. Data augmentation and preprocessing is essential for medical image applications. Our experiments show the method we proposed that we pick top two networks out and jointly fine-tune the two networks is more useful for medical image analysis, with better performance than the other four CNNs. More importantly, the results of classification demonstrate that the knowledge learned from natural image can still transfer to make medical image quality classification effectively even though the disparity between them.

References

1. Ye, P.: Unsupervised feature learning framework for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 157(10), pp. 1098–1105 (2012)
2. Wang, J.: A novel contourlet-based no-reference image quality assessment metric. In: International Research Association of Information and Computer Science 2014, pp. 3339–3352 (2014)
3. Lee, S.C., Wang, Y.: Automatic retinal image quality assessment and enhancement. In: Proceedings Spie, pp. 1581–1590 (1999)
4. Lalonde, M., Gagnon, L., Boucher, M.C.: Automatic visual quality assessment in optical fundus images. In: Proceedings of Vision Interface, vol. 18, pp. 437–450 (2001)
5. Yu, L., Tian, X., Li, T., Tian, J.: No-reference image quality assessment based on svm for video conferencing system. In: Lei, J., Wang, F.L., Li, M., Luo, Y. (eds.) Communications in Computer & Information Science 2016, vol. 345, pp. 555–560. Springer, Heidelberg (2012). doi:10.1007/978-3-642-35211-9_70
6. Suk, H.-I., Shen, D.: Deep learning-based feature representation for AD/MCI classification. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 583–590. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40763-5_72
7. Kang, L., Ye, P.: Convolutional neutral networks for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition 2014, pp. 1733–1740 (2014)
8. Tennakoon, R., Mahapatra, D., Roy, P.: Image quality classification for DR screening using convolutional neural networks. In: Chen, X., Garvin, K. (eds.) OMIA 2016, pp. 113–120 (2016)
9. Mahapatra, D.: Retinal image quality classification using neurobiological models of the human visual system. In: Chen, X., Garvin, K. (eds.) OMIA 2016, pp. 97–104 (2016)
10. Krizhevsky, A., Sutskever, I.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems 2012, vol. 25, pp. 1097–1105 (2012)
11. Szegedy, W., Liu, Y.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition 2015, pp. 1–9 (2015)
12. Mohammadi, M., Das, S.: SNN: Stacked Neural Networks. arXiv:1605.08612 (2016)
13. Pratt, H., Coenen, F.: Convolutional neural networks for diabetic retinopathy. Proc. Comput. Sci. **90**, 200–205 (2016)