Image Quality Classification for DR Screening Using Convolutional Neural Networks

Ruwan Tennakoon, Dwarikanath Mahapatra, Pallab Roy, Suman Sedai, and Rahil Garnavi

IBM Research Australia
{truwan,dwarim,pallroy,ssedai,rahilgar}@au1.ibm.com

Abstract. The quality of input images significantly affects the outcome of automated diabetic retinopathy screening systems. Current methods to identify image quality rely on hand-crafted geometric and structural features, that does not generalize well. We propose a new method for retinal image quality classification (IQC) that uses computational algorithms imitating the working of the human visual systems. The proposed method leverages on learned supervised information using convolutional neural networks (CNN), thus avoiding hand-engineered features. Our analysis shows that the learned features capture both geometric and structural information relevant for image quality classification. Experimental results conducted on a relatively large dataset demonstrates that the overall method can achieve high accuracy. We also show that effective features for IQC can be learned by full training of shallow CNN as well as by using transfer learning.

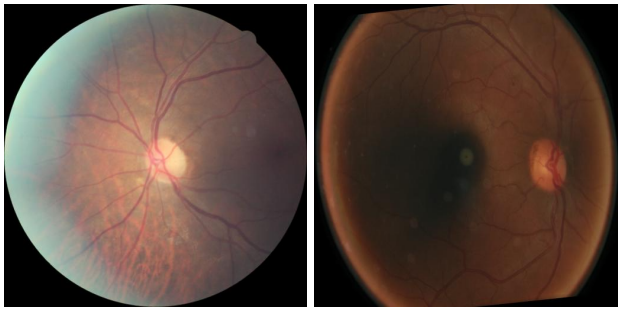
Keywords: Digital fundus images, Diabetic retinopathy, Image quality classification, CNN, retinal imaging.

1 Introduction

Digital fundus photography enables non-invasive diagnosis of retinal related conditions like diabetic retinopathy (DR), age-related macular degeneration (AMD) and glaucoma. The symptoms of the above conditions are well defined and visible in fundus images [6]. Consequently, automated evaluations can be performed to support the diagnosis and documentation of these conditions [3], [15]. The success of these automatic diagnostic systems heavily rely on the quality of images presented to them. Due to factors like level of operator expertise, type of equipment used and patient conditions, the acquired retinal image might not have the minimum quality that would facilitate feature extraction, leading to incorrect analytics. Therefore, image quality classification (IQC) is an important component in automated screening systems for diseases like diabetic retinopathy. Figures 1 shows two examples of un-gradable images that hamper reliable feature extraction.

In the context of retinal image analysis, IQC is used to grade an image according to its usefulness to the diagnosis. The Atherosclerotic Risk in Communities

X. Chen, M. K. Garvin, J. Liu, E. Trucco, Y. Xu (Eds.): OMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, Iowa Research Online, pp. 113–120, 2016. Available from: http://ir.uiowa.edu/omia/2016_Proceedings/2016/



(a) (b)

Fig. 1. Example of un-gradable images that hamper reliable analytics.

(ARIC) [1] study identified reliable factors for IQC that were grouped into two major categories: generic image quality parameters (e.g. contrast, clarity, etc) and structural quality parameters (such as visibility of the optic disc and macula). Methods using generic image information include histogram matching [10] and distribution of edge magnitudes [9]. Despite low computational complexity, these methods do not always capture diversity of conditions affecting image quality. Structural quality parameters based methods use retinal landmarks like the vasculature [17] and multi scale filter banks [11]. They require anatomical landmark segmentation which is complex and error prone, especially for poor quality images. The method proposed by Paulus et al. [12] combined generic and structural image features but rely heavily on accurate landmark segmentation.

To the best of our knowledge all the classification methods used so far for IQC of digital fundus images rely on some kind of handcrafted features that are based on either generic or structural quality parameters which do not generalise well to new datasets. On the other hand, human experts rely on the capabilities of the human visual system (HVS) to identify poor quality images and have the capability to adapt to new scenarios based on presented data. However, the evaluation may be subjective as it depends on a user's perception of good quality. Current approaches based on handcrafted features do not leverage the functioning of the HVS to improve IQC. This necessitates solving the problem using computational principles behind the working of the HVS, thus minimising subjectivity and bias of existing algorithms.

We propose a novel method for retinal IQA that uses computational algorithms imitating the working of the HVS. The proposed method leverages learned supervised information using convolutional neural networks (CNNs) thus avoiding hand crafted features. Our analysis shows that the learned features capture both geometric and structural information that is relevant for IQC and the result on a relatively large dataset shows that the overall method can achieve high accuracy.

2 Method

We propose a data driven approach for image quality classification of retinal fundus images. Given a fundus image I_i the intention here is to find a mapping $f : I_i \rightarrow c_i; \quad c_i \in [0, 1]$ where, 0 - is of sufficient quality for automated analysis (Gradable), 1 - does not have sufficient quality for automated analysis (un-gradable). Existing methods first extract hand-crafted geometric or structural features, $l_i = \mathcal{F}(I_i)$, from the image and use them for the mapping, instead of using the image information directly. In contrast, our approach employs deep convolutional neural networks (CNN) to learn features from data that can be used for dichotomising poor quality images. Basic CNNs model the feature extractor with consecutive layers of convolution, nonlinear activations and pooling. These features are then connected to a multilayer neural network classifier and the model parameters of the feature extractor and the classifier is learned end to end by maximising the data likelihood:

$$\underset{w}{arg\ max} \bigcup_{i \in \mathcal{D}} p\left(C = c_i \mid I_i, w\right) \tag{1}$$

where $\mathcal{D} = [I_i, c_i]_{i=1}^{n}$ is the training dataset and w are the parameters or weights of the CNN model with a specified architecture.

Since winning the ImageNet competition in 2012 [8], CNN's have gained wide popularity in computer vision. There success is mainly attributed to faster processing (GPUs), rectified linear units, dropout regularisation, and effective data augmentation [13]. Currently there are three main techniques used in training CNN's for medical imaging applications [13]: training the CNN from scratch, use trained off-the-shelf CNN's (i.e. AlexNet [8], VGG-Net [14]) to extract features, unsupervised pre training on large image datasets.

Training deep CNN (i.e. AlexNet, VGG-Net) from scratch requires large amount of labeled data, which is difficult to obtain for medical applications due to limited availability of resources (experts) for image annotations and patient privacy issues. Therefore, in this work we investigate two CNN architectures with two different model parameter estimation techniques. The first is a shallow CNN (with less parameters than conventional CNNs like AlexNet) that we train from scratch using the colour fundus image dataset described in Section 3.1. The second network uses transfer learning, where instead of training a complete deep network, we used pre-trained filters (trained on natural images from ImangeNet competition) to extract features and used them in different classifiers.

2.1 Network Architecture

ShallowNet: The architecture of the shallow network trained from scratch is illustrated in Figure 2. The network consists of three convolutional layers with $96, 256, 256$ convolutional filter of kernel size $11 \times 11, 5 \times 5, 3 \times 3$ respectively. Each Convolutional layer is immediately followed by rectified linear activations ($\sigma(x) = max(0, x)$) and max pooling layers (each max pooling layer has kernel

size 3×3). The feature maps produced by the convolution layers is then fed in to two consecutive fully connected layers of which the output is classified using a soft max classification layer. Training of deep network is known to be hampered by the phenomenon called internal covariate shift (the distribution of each layer's inputs changes during training). To overcome the above effect, following [5] we applied "batch normalisation" to the outputs of the first two convolutional layers. Drop-out regularisation [16] was also used at the last two fully connected layers to prevent over-fitting.

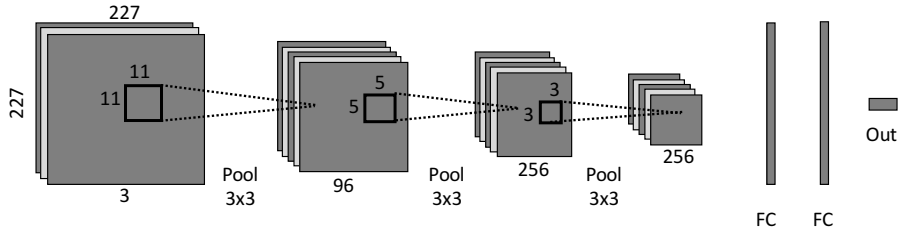


Fig. 2. The architecture of the shallow network used in fundus image quality assessment.

The network was trained by optimising the cross-entropy loss function given by:

$$E = \sum_{i=1}^{n} log\left(\hat{p}_{i,c_i}\right) \tag{2}$$

where $\hat{p}_{i,k} = e^{x_{ik}} / \sum_k e^{x_{ik}}$ and x_{ik} is the k^{th} output of the last layer (output layer) of the network. The above loss function was optimised using stochastic gradient decent with momentum. The proposed network architecture was implemented in caffe framework [7] and the training was carried out on a workstation with a NVIDIA-Tesla K40 GPU. The CNN weights were initialised with random Gaussian distributions and training was run for 30 epochs. The hyper-parameters of the network was set as follows: momentum: 0.9; weight decay: 0.0005; learning rate: 0.01 decreased by a factor of 10 at every 10 epochs.

AlexNet: Published in [8] AlexNet architecture consists of five convolutional layers, three pooling layer, two local response normalisation layers and two fully connected layers. The network was originally trained on natural images form the ImageNet competition and the network weight are available through caffe model zoo. Several medical imaging application has either used these trained weights as a feature extractor or to initialise there network [2]. This practice is called transfer learning and it helps overcome the problem of limited training data. In this work, the training images were directly fead to the network with pertained weights and the output of the last fully connected layer ("fc7" - with dimensions 1×4096) was extracted as features. Those extracted featured (of training data

defined in Section 3.1) were then used to train four seperate classifiers: 1) single layer Neural Network (AlexNet-FT - this is equivalent to fine tuning the last layer ("fc8") of the AlexNet) 2) linear support vector machine (AlexNet-SVM), 3) Boosted trees (AlexNet-BT), 4) k-Nearest Neighbour (AlexNet-KNN). The hyper-parameters of the three classifiers were tuned using five fold cross validation on the features extracted from training dataset.

2.2 Data Augmentation

Data augmentation is commonly used in training deep CNNs to make the network robust to slight input variances (e.g. translation, rotation). In this work, we rotated each training image by a set of fixed angles (6° to 210° with a resolution of 6°) to make the network rotational invariant. It is well known that the pooling layers used in the CNN incorporate some level of translational equi-variance.

3 Results & Discussion

3.1 Data

The dataset (D1) contain 908 un-gradable retinal images and 944 gradable images. All of the images are non-mydriatic and have a 45° FOV and a resolution of 2812×2442 pixels. All the images were graded by human graders, thus confirming their gradability labels. The dataset was randomly split into 75-25% training and test segments (training set: 681 un-gradable and 708 gradable images). The training images was then augmented to generate the dataset (see Section 2.2) for training the networks while the test set without any augmentation was used in our evaluations.

3.2 Relevance of CNN descriptors

CNN' are known to act as hierarchical feature extractors, where the first level extract simple features like edges and local contrast whereas the consecutive layers extract complex features that combines the several low level features. To gain an understanding of the information that the network has learned, we have shown several feature maps at two levels of the network (extracted after the first two convolutional layers) in Figure 3. The features extracted after the first convolutional layer shows that the network has learned to extract geometric information like edges of vessel structure (two left maps in the top row) as well as relevant structural features like the macular and optic disk (Two right maps in the top row). The features extracted after second convolutional layer (bottom row), show clear localizations of the optic disk and macula and also show segmentation of low and high bright regions of the retina.

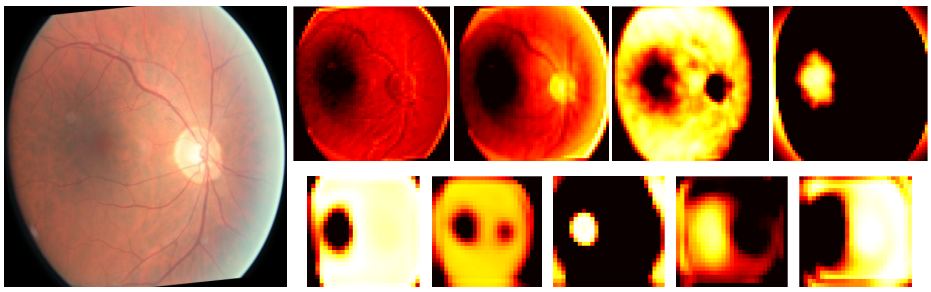


Fig. 3. Features extracted at two levels of the shallow network together with the original input image (left). Top row shows selected feature maps extracted after the first convolutional layer and bottom row shows selected feature maps extracted after the second convolutional layer.

3.3 Classification Results

The classification results evaluated using the test set is presented in Table 1. The results show that the shallow network trained from scratch has achieved very high accuracy (8 errors in 463 tested images) with over 99% sensitivity (the probability that the network would correctly identify a poor quality image). This indicates that the shallow network with only three convolutional layers has been able to learn the necessary information for image quality classification in retinal images, from data.

The classification accuracy for fine tuning the last layer of the AlexNet (Alexnet-FT) was similar to that obtained for the shallowNet. However, the other three classifiers (Alexnet-SVM, Alexnet-KNN and Alexnet-BT) working on features extracted from AlexNet (with pre trained weights) showed slightly lower classification accuracies of around 97%. This shows that while the features extracted from the pre-trained AlexNet is well descriptive for the task of IQC, the final results do depend on the classifier used. It is noted here that, the computation time for the shallow network (forward pass: 168ms, backward pass: 297ms) was significantly lower than the computation time for deeper network i.e. Alexnet (forward pass: 518ms, backward pass: 1344ms).

The classification result also shows that extracting features using pre-trained neural network filters can be as affective as fully training a network from scratch for the assessment of image quality in retinal fundus images. Comparing our results with [12, 11, 4] we come very close to the best performing methods of [4, 11]. However, we have used a much larger dataset that has images from various machines and imaging centres, thus demonstrating our method's robustness.

To obtain a qualitative idea of the performance, we have depicted several images in Figure 4 where the trained classifier (shallowNet) has failed. These and similar cases represent instances of wrong labels. Clearly the images in the top row are actually of poor quality but have been labelled as gradable. Similarly, the images in the bottom row are of good quality but have been incorrectly labeled as un-gradable. These results also indicate that despite a few erroneous

Table 1. Classification accuracy results on the test set.

Network	Accuracy (%)	Specificity (%)	Sensitivity (%)
Proposed shallowNet	98.27	97.46	99.12
Alexnet-FT	98.27	97.03	99.55
Alexnet-SVM	97.19	95.38	99.12
Alexnet-BT	96.98	99.15	94.71
Alexnet-KNN	96.98	96.19	97.80

labels our CNN based approach is able to learn a reliable feature representations that can separate different image classes.

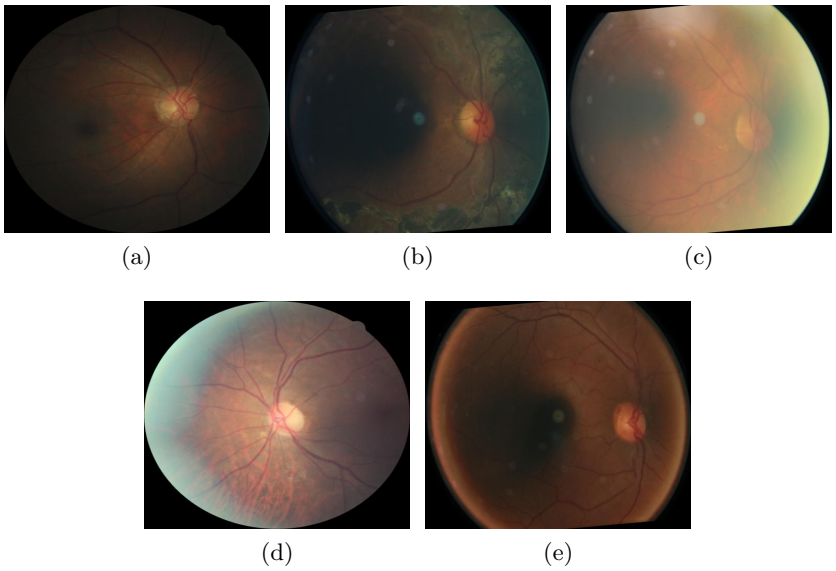


Fig. 4. Examples of incorrectly classified images. (a-c) Incorrectly classified as ungradable. (d-e) Incorrectly classified as gradable.

4 Conclusion

We have proposed a novel method based on a shallow CNN for the purpose of retinal image quality classification. This is an important step in large screening programs for diabetic retinopathy. The proposed shallow CNN architecture matches the performance of a deeper architecture, but has significantly reduced computational complexity. Experimental results on a very large dataset demonstrate that our shallow architecture outperforms other methods that use the learned knowledge of the baseline AlexNet method.

References

1. The atherosclerosis risk in communities (ARIC) study: design and objectives. The ARIC investigators. am j epidemiol. 1989 apr; 129(4), 687-702.
2. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H.: Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). pp. 294–297. IEEE (2015)
3. Bock, R., Meier, J., Nyúl, L.G., Hornegger, J., Michelson, G.: Glaucoma risk index: automated glaucoma detection from color fundus images. Medical image analysis 14(3), 471–481 (2010)
4. Dias, J.M.P., Oliveira, C.M., da Silva Cruz, L.A.: Retinal image quality assessment using generic image quality indicators. Information Fusion 19, 73–90 (2014)
5. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
6. Jelinek, H., Cree, M.J.: Automated image detection of retinal pathology. CRC Press (2009)
7. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. pp. 675–678. ACM (2014)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
9. Lalonde, M., Gagnon, L., Boucher, M.: Automatic visual quality assessment in optical fundus images. In: Proc. Vision Interface. pp. 259 – 264 (2001)
10. Lee, S., Wang, Y.: Automatic retinal image quality assessment and enhancement. In: Proc. SPIE Medical Imaging. pp. 1581–1590 (1999)
11. Niemeijer, M., Abramoff, M., van Ginneken, B.: Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Med. Imag. Anal. 10(6), 888–898 (2006)
12. Paulus, J., Meier, J., Bock, R., Hornegger, J., Michelson, G.: Automated quality assessment of retinal fundus photos. Intl. J. Comp. Assisted Radiology and Surgery . 10(6), 888–898 (2006)
13. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging 35(5), 1285–1298 (2016)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
15. Sinthanayothin, C., Boyce, J., Williamson, T., Cook, H., Mensah, E., Lal, S., Usher, D.: Automated detection of diabetic retinopathy on digital fundus images. Diabetic medicine 19(2), 105–112 (2002)
16. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15(1), 1929–1958 (2014)
17. Usher, D., Himaga, M., Dumskyj, M.: Automated assessment of digital fundus image quality using detected vessel area. In: Proc. Medical Image Understanding and Analysis. pp. 81–84 (2003)