DRNet-Q: A lightweight ConvNet for image quality classification for diabetic retinopathy screening

German Pinedo-Diaz[1], Susana Ortega-Cisneros[1],
E. Ulises Moya-Sanchez[2,3], Abraham Sanchez[2], Baowei Fei[3]
[1]Cinvestav Unidad Guadalajara, Mexico [2]Gobierno de Jalisco, Mexico [3]University of Texas, Dallas
{german.pinedo, susana.ortega}@cinvestav.mx
{eduardo.moya, abraham.sanchez}@jalisco.gob.mx
bfei@utdallas.edu

Abstract

Diabetic retinopathy (DR) is the primary cause of blindness in developing and developed countries and the early-stage DR detection reduces the risk of blindness in Diabetes Melitus (DM) patients. The image classification systems based on Deep Learning (DL) could improve the timely recognition of Diabetic Retinopathy (DR) features by a medical specialist. However, these benefits have yet to be demonstrated real-world clinical applications. One possible reason is that machine learning systems are trained and tested only on high-quality datasets, while in practical applications, the input images cannot be assumed to be of high quality. In this context, we propose a new simple light weight Convolutional Neural Network (ConvNet) architecture capable of classifying Retinal Fundus Photograph (RFP) quality for DR screening training from scratch. In particular, we take care not only in the classical image-quality degradations such as noise, blur, contrast but also in the visibility of the specific anatomical regions (optical nerve, fovea area, and arcade) which are very important to DR detection screening. In addition, we collected images from different datasets to improve the robustness of the model. The accuracy of the proposed model is 98.7 % with fewer parameters, less training, and inference time (78 % faster) in comparison with three state-of-the-art Convolutional Neural Network architectures.

1. Introduction

Image quality is an essential challenge to practical implementation [8], and this is commonly overlooked in the design of DL systems, as these systems are trained and tested on high-quality image datasets [3, 11]. These shortcomings could be more dramatic for medical applications [3]. The global population in 2019 with DM is 463 million and the estimation to 2045 is up to 700 million [20]. DR is a complication of DM [1] and is the primary cause of blindness in developing and developed countries [6]. The DR global prevalence in Africa is 35.90%, North American and the Caribbean with 33.30% and in South and Central America 13.37% [20]. According to International Council of Ophthalmology (ICO) [1], the early-stage DR detection reduces the risk of vision loss and blindness in DM patients. Due to some RFP of DR showed outstanding lab classification performance [21], the model's accuracy in real applications decreases significantly [3]. To face the problem of DR-RFP quality classification, a new lightweight Convolutional Neural Network (CNN) is proposed. Note that, we take into account the image quality metrics blur, noise and low contrast, in addition we evaluate visibility of the most important anatomical regions for DR for example, macula, optic nerve and arcades. Moreover, we selected the most representative severe DR lesions (drusen, microaneurysm, exudades, and hemorrhages) to avoid the DR-lesion/artifact confusion.

2. Data and labeling

The data was collected from different public data sources to improve the robustness of the model: Kaggle [9], MESSIDOR II [7], DRIMDB [17], IDRID [5], DRIVE [22]. The Label 0 corresponds to low quality images, i.e. images where the optic nerve, macula or arcades are not visible due to blur, noise, illumination problems, artifacts, and Label 1 corresponds to acceptable quality images even with severe DR lesions (drusen, microaneurysm, exudades, and hemorrhages), see examples for each class Figure 1). The training dataset was split using the hold-out strategy.

2.1. Data selection for labeling

The bad quality RFPs are difficult to find due to the images are heterogeneous in terms of size, Blur (B), Con-

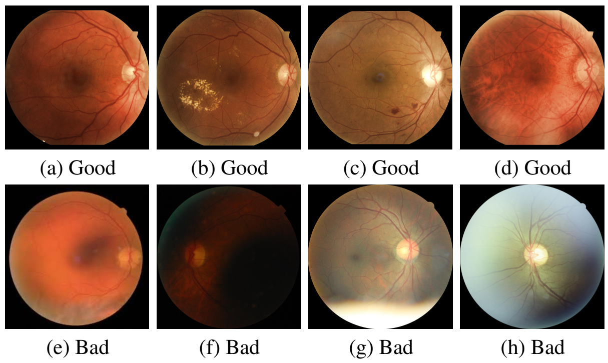


Figure 1. Examples for human labeling based on visual content considerations, a) Visible optical nerve, macula and vessels b) Visible exudates , c) Visible hemorrhages , d) Visible Neo-vessels and capillaries, e) Blurred image, f) Dark image, g) Flash artifact, f) Macula not visible and flash artifact

trast (C), and Signal-to-Noise Ratio (SNR). To avoid label all the images (around of 35,000 images), we pre-selected some images (using a k-means) and then labeled the RFP (by a human) as the diagram presented in Figure Fig. 2. In order to pre-select the images, first, we measured three non-reference image quality measurements (B, C and SNR) using OpenCV-Python library [13]

$$B = \sigma^2(\Delta(I)), \quad C = \sigma(I_{gray}), \quad SNR = \frac{\mu(I)}{\sigma(I)},$$

where $I = I(r, g, b)$, $r, g, b \in \{0, 255\}$, r, g, b are the red, green and blue components, I_{gray} is the gray representation of the image, Δ is the Laplacian operator and μ, σ and σ^2 are the mean, standard deviation, and variance of the image, respectively.

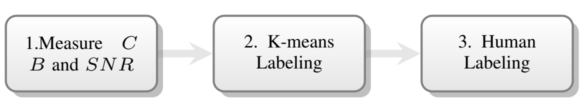


Figure 2. Flow diagram of the data pre-selection and labeling. The final label is assigned by a human based on the visual criteria proposed in [14].

Once the measurement is done, we group with K-means [15] (k=3), and we obtain the criteria ranges for label 0 and label 1, detailed in expression 1 and 2

$$Label\ 0 = \begin{cases} SNR < 1, \\ 10 > B > 64, \\ 0.12 > C > 0.27 \end{cases} \tag{1}$$

$$Label\ 1 = \begin{cases} SNR > 1, \\ 10 < B < 64, \\ 0.12 < C < 0.27 \end{cases} \tag{2}$$

Finally, the pre-selected images (based on the K means) were exterminated by a human and define the final label. We consider the visibility of the anatomical content (optic nerve, macula and arcade) of the RFP based on [14]. In addition, we considerate for high-quality image label examples from all DR- levels, specially examples, with microaneurysms, neovessels, hemorrhages, and exudates.

2.2. DRNet-Q and other ConvNet Benchmarks

Tab. 1 shows a summary of the proposed architecture. We selected a 5×5 window convolution as the top layer to detect the relatively big size artifacts features. Then we use a batch layer normalization to speed up the training process [12]. The rest of the architecture was proposed based in our experience. The performance of the proposed ConvNet were compared to three pre-trained state-of-the-art CNNs i)Inception v3 [19], w ii) Inception [18] ; and iii) ResNet-50 [10]. For all CNNs, we train under the same conditions, which are SGD optimizer, 64 batch size, 0.001 Learning rate and 100 epochs. Hardware specifications for training and testing are Intel Core i7 @4GHz, 64GB RAM and 2 Nvidia 1080 GPUs. For testing purposes only, a system based on Intel core i3 CPU @2.4GHz, 8GB RAM has been used. As well as, an integrated system Jetson TX2 Denver CPU, 8GB RAM DDR4 and GPU-Nvidia Pascal.

Table 1. DRNet-Q summary architecture.

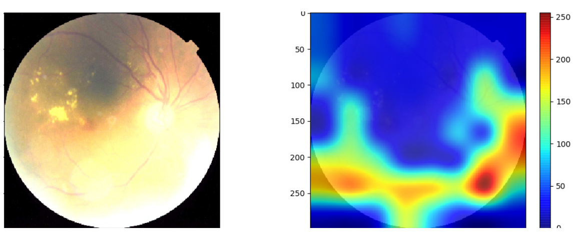
Layer	**Size**	**Dimensions**	**Activation**	**Parameters**
conv2d_1	5x5	256x256x96	ReLU	7296
conv2d_2	5x5	256x256x64	ReLU	153664
Max pooling	2x2	128x128x64	NA	0
batch norm.	NA	128x128x64	NA	256
conv2d_3	3x3	128x128x128	ReLU	73856
conv2d_4	3x3	128x128x128	ReLU	147584
Max pooling	2x2	64x64x128	NA	0
batch norm.	NA	64x64x128	NA	512
conv2d_5	3x3	64x64x128	ReLU	147584
conv2d_6	3x3	64x64x128	ReLU	147584
Max pooling	2x2	32x32x128	NA	0
batch norm.	NA	32x32x128	NA	512
conv2d_7	3x3	32x32x64	ReLU	73792
conv2d_8	3x3	32x32x64	ReLU	36928
Max pooling	2x2	16x16x64	NA	0
batch norm.	NA	16x16x64	NA	256
conv2d_9	3x3	16x16x32	ReLU	18464
Max pooling	2x2	8x8x32	NA	0
batch norm.	NA	8x8x32	NA	128
Flatten	NA	1x2048	NA	0
Dense	NA	1x64	ReLU	131136
Dropout(0.5)	NA	1x64	NA	0
Dense	NA	1x32	ReLU	2080
Dropout(0.4)	NA	1x32	NA	0
Dense	NA	1x2	Softmax	66

3. Results and analysis

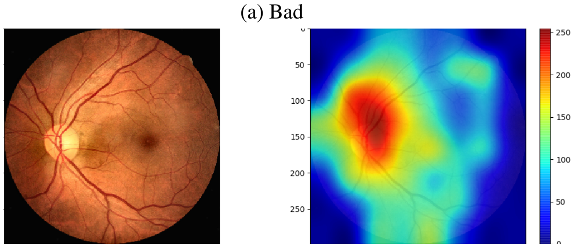
Tab. 3 presents the Area Under Curve (AUC) of the ROC curve, sensitivity (Sen), and specificity (Spe) from different models. The first three rows show similar works mainly based on transfer learning techniques [2, 4, 23] and with a mixture of public and provate datasets. According to the results The DRNet-Q has similar or better performance when compared to Inception V3, Inception V4 and ResNet-50 using the same data and hardware conditions.

On the other hand, the computational performance is compared in the Tab. 2. DRNet-Q achieves the shortest training time, with a factor between 1.25 to 2.76 times faster respect to the other CNNs. As well as the inference time of DRnet-Q on low-performance hardware with respect to traditional models, is 1.78 to 12 times faster.

Regarding the explicability of the model, we compute the activation maps generated by the Grad CAM [16] over two examples (see Fig. 3) of each class using DRNet-Q in layer *conv2d_9* . The activation map of class 0, highlights the area with more brightness (artifact). In contrast, the activation map of the class 1 images focuses on two main anatomical areas, such as the optic nerve and the macula, important anatomical areas to detect DR.



(a) Bad



(b) Good

Figure 3. DRNet-Q Gradient-weighted activation map for each class.

4. Conclusions

In this work, a new simple ConvNet architecture trained from scratch to classify the image quality of RFP is presented. The activation maps of the proposed ConvNets confirm that CNN finds the most important anatomical features of the RFP. According to the results, performance (AUC, sensitivity, specificity) of the proposed CNN does not significantly differ from three state-of-the-art ConvNets. However, the DRNet-Q used fewer computational resources and improved the training and the inference time, by a factor of at least 1.78 in comparison to the other ConvNets using the same data, software and hardware. We hope that this model could help to reduce inconvenient for a patient to return to a medical center to repeat the fundus photography exam.

Table 2. Computational resources and time processing comparison.

Model	Train (Hrs)	Params (M)	Mem (GB)	Speed (sec)		
				GPU	CPU	TX2
InceptionV3	1.21	115	1.29	2.24	5.47	5.87
InceptionV4	2.29	154	1.46	5.05	13.9	14.4
ResNet50	1.04	126	1.09	1.40	3.00	2.66
DRNet-Q	0.83	0.94	0.08	0.51	1.16	1.49

Table 3. State of the art review and model comparison.

Work	Dataset	Results AUC	Sen	Spec
Zago et al. [23]	DRIMDB, ELSA-Brasil.	98.5%	92.0%	96.0%
Chalakkal, et al. [4]	DRIMDB, HRF, MESSIDOR, UoA, IDRID, DR1-DR2 and Kaggle.	97.47%	98.38%	95.19%
Alais et al. [2]	OPHDIAT	97.1%	99.0%	95.3%
Inception V3 (same conditions)	MESSIDOR, IDRID and Kaggle.	98.5%	99.0%	98.3%
Inception V4 (same conditions)	MESSIDOR, IDRID and Kaggle.	98.6%	99.0%	99.8%
ResNet50 (same conditions)	MESSIDOR, IDRID and Kaggle.	98.7%	98.1%	98.0%
DRNet-Q (same conditions)	MESSIDOR, IDRID and Kaggle.	98.7%	98.2%	98.1%

References

[1] International council ophthalmology, guidelines for diabetic eye care, 2017. 1

[2] Robin Alais, Petr Dokládal, Ali Erginay, Bruno Figliuzzi, and Etienne Decencière. Fast macula detection and application to retinal image quality assessment. *Biomedical Signal Processing and Control*, 55:101567, 2020. 3

[3] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. A human-centered evaluation of a deep learning system deployed in clinics for the detection of

diabetic retinopathy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020. 1

[4] Renoh Johnson Chalakkal, Waleed Habib Abdulla, and Sinumol Sukumaran Thulaseedharan. Quality and content analysis of fundus images using deep learning. *Computers in Biology and Medicine*, 108:317 – 331, 2019. 3

[5] Eye Clinic. Diabetic Retinopathy Segmentation and Grading Challenge. https://idrid.grand-challenge.org/Data/, 2012. [accessed 04-September-2019]. 1

[6] Emily Dawn Cole, Eduardo Amorim Novais, Ricardo Noguera Louzada, and Nadia K Waheed. Contemporary retinal imaging techniques in diabetic retinopathy: a review. *Clinical & experimental ophthalmology*, 44(4):289–299, 2016. 1

[7] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, Aug. 2014. 1

[8] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks, 2016. 1

[9] EyePACS. Diabetic Retinopathy Detection competition. https://www.kaggle.com/c/diabetic-retinopathy-detection/, 2015. [accessed 04-September-2019]. 1

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2

[11] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations, 2019. https://arxiv.org/pdf/1903.12261.pdf. 1

[12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015. 2

[13] Itseez. Open source computer vision library. https://github.com/itseez/opencv, 2015. 2

[14] Cédric Lamirel, Beau B Bruce, David W Wright, Kevin P Delaney, Nancy J Newman, and Valérie Biousse. Quality of nonmydriatic digital fundus photography obtained by nurse practitioners in the emergency department: the foto-ed study. *Ophthalmology*, 119(3):617–624, 2012. 2

[15] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press. 2

[16] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2):336–359, Oct 2019. 3

[17] Ugur Sevik. Drimdb (diabetic retinopathy images database) database for quality testing of retinal images, 04 2014. 1

[18] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016. 2

[19] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision, 2015. 2

[20] Zhen Ling Teo, Yih-Chung Tham, Marco Yu, Miao Li Chee, Tyler Hyungtaek Rim, Ning Cheung, Mukharram M. Bikbov, Ya Xing Wang, Yating Tang, Yi Lu, Ian Y. Wong, Daniel Shu Wei Ting, Gavin Siew Wei Tan, Jost B. Jonas, Charumathi Sabanayagam, Tien Yin Wong, and Ching-Yu Cheng. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology*, 128(11):1580–1591, 2021. 1

[21] Daniel Shu Wei Ting, Carol Yim-Lui Cheung, Gilbert Lim, Gavin Siew Wei Tan, Nguyen D. Quang, Alfred Gan, Haslina Hamzah, Renata Garcia-Franco, Ian Yew San Yeo, Shu Yen Lee, Edmund Yick Mun Wong, Charumathi Sabanayagam, Mani Baskaran, Farah Ibrahim, Ngiap Chuan Tan, Eric A. Finkelstein, Ecosse L. Lamoureux, Ian Y. Wong, Neil M. Bressler, Sobha Sivaprasad, Rohit Varma, Jost B. Jonas, Ming Guang He, Ching-Yu Cheng, Gemmy Chui Ming Cheung, Tin Aung, Wynne Hsu, Mong Li Lee, and Tien Yin Wong. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*, 318(22):2211–2223, 12 2017. 1

[22] XX. XX. https://www.isi.uu.nl/Research/Databases/DRIVE/, X. [X]. 1

[23] Gabriel Tozatto Zago, Rodrigo Varejão Andreão, Bernadette Dorizzi, and Evandro Ottoni [Teatini Salles]. Retinal image quality assessment using deep learning. *Computers in Biology and Medicine*, 103:64 – 70, 2018. 3