The AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI (R^{2}HCAI)

Adaptive Histogram Equalization in Diabetic Retinopathy Detection

Daniela Angela Parletta[1] **Giovanna Purgato**[2]

[1] Akkodis, Data Science and Artificial Intelligence Division, Rome, Italy [2] Akkodis, Rome, Italy
danielaangela.parletta@akkodisgroup.com, giovanna.purgato@akkodisgroup.com

Abstract

Diabetic retinopathy is a dangerous pathology that can ultimately lead to permanent blindness. Recent studies have proved the feasibility of automatic diagnosis systems, supporting specialists, from eye fundus images, based on deep networks. A careful image equalization turned out to be important in obtaining good performances. However, state-of-the-art algorithms for image equalization require expensive parameter tuning which may limit the adoption of such support systems in practice. In this paper, we propose a learning-based approach to adaptively select the right equalization parameters for each image. This approach allows significant reduction in the inference time without necessarily sacrificing the accuracy. A preliminary empirical evaluation confirms the advantages of the proposed method.

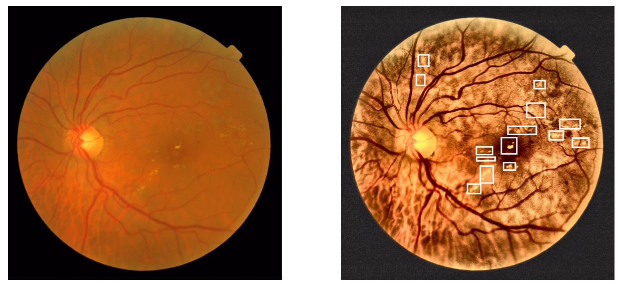

Figure 1: **Image equalization**. On the left, a raw eye fundus image of a subject suffering from a severe diabetic retinopathy. On the right, the equalized image where there are visible and highlighted in the white edged boxes many of the typical lesions caused by the pathology.

Introduction

Diabetic retinopathy is a dangerous pathology that can ultimately lead to permanent blindness (Cheung and Wong 2008; Stitt et al. 2016). Hence, a timely diagnosis is fundamental to enable an appropriate treatment and limit the consequences of the pathology. The problem of designing a learning based system for automatic retinopathy detection supporting the specialist dates back at least to (Sinthanayothin et al. 1999) (see also (Gupta and Chhikara 2018) for an historical account). However, traditional machine learning system suffers from the drawback of requiring a careful feature engineering phase that usually results in difficult to compute features. More recently, deep learning methodologies have been used to face this problem (Abràmoff et al. 2016; Badar, Haris, and Fatima 2020; Pour et al. 2020; Pratt et al. 2016). One aspect that turned out to be important in obtaining good performances with deep-learning methods is of image equalization (Pour et al. 2020). Roughly speaking, image equalization consists of making the distribution of the pixel intensity somehow uniform, resulting in images with improved contrast (see (Zuiderveld 1994) and the references therein). Figure 1 shows a comparison between a raw eye fundus image and its equalized counterpart. As it is possible to see, in the equalized image, many fine-grained details are made evident revealing the presence of the retinopathy (e.g., the highlighted lesions). State-of-the-art methods for image equalization, including the popular Contrast Limited Adaptive Histogram Equalization (CLAHE) (Akram et al. 2014; Zuiderveld 1994), require a careful parameter tuning in order to obtain an acceptable contrast. Unfortunately, automating this tuning is difficult due to the fact that there are no commonly accepted objective functions for evaluating the choice of the parameters and the existing proposals are non-convex; which makes their optimization computationally intractable. Furthermore, the existing proposals are related to some notion of image quality, rather than to the accuracy of the inference that can be made *a posteriori* on the equalized image. Hence, carefully selecting the parameters for an appropriate image equalization is a difficult and time consuming process (Campos et al. 2019; Reddy Eswar 2019; Fawzi, Achuthan, and Belaton 2021; Joseph et al. 2017; Min et al. 2013).

Contributions. Inspired by the recent work on dynamic resolution networks (Zhu et al. 2021), in this work we propose a novel approach to image equalization in the context of diabetic retinopathy detection. The proposed learning-based method consists of a pair of networks, that given an image determines the appropriate parameters for the equalization step and, after the equalization, predicts its label respectively. The networks are trained in an end-to-end fashion as described in Section 2. The main advantage of this method is that of substantially reducing the overall inference time of the system while retaining near optimal accuracy. This in-

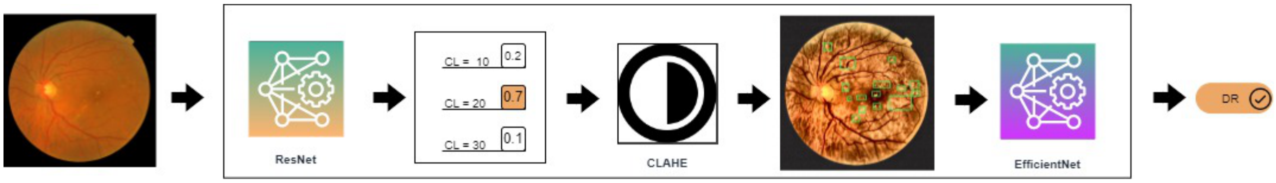

Figure 2: **Proposed method**. The overall system is composed of a series of two networks, a ResNet and an EfficientNet. After the system is trained in an end-to-end fashion, a raw image is fed to ResNet which selects the CL for the CLAHE. Second, the image is equalized with the tuned CLAHE method; finally, the processed image is fed into EfficientNet for retionapthy classification. The green bounding boxes reported on the equalized images have an illustrative purpose and are not applied in practice: they illustrates the many finer details highlighted by the equalization process.

creases the practical usability of such systems in real-world scenarios. Additionally, while many classical cost functions for pre-processing algorithms are not directly related to the accuracy, our proposal leads to an integrated solution that is optimized for the predictions. As a result, the proposed method, can even obtains better performances than expensive grid-search procedures; especially in the small data-regime. We illustrate our proposal in the context of CLAHE. In Section 3 we report an experimental evaluation on real-world data and finally draw the conclusions in Section 4.

Methodology

Our method takes inspiration from the dynamic resolution network developed in (Zhu et al. 2021) and the fact that an accurate pre-processing tunes the image equalization algorithm optimally for each image. We illustrate the proposal in the context of a specific image equalization method, the popular CLAHE algorithm (Zuiderveld 1994). CLAHE requires the user to specify two parameters: the *tile size* and the *clip limit*, denoted by TS and CL respectively. Motivated by the fact that the *tile size* plays a minor role on the equalization performances in comparison to the *clip limit* (Reddy Eswar 2019; Fawzi, Achuthan, and Belaton 2021; Joseph et al. 2017; Min et al. 2013) and to favor simplicity, we limit the tuning to CL and keep fixed TS.

As shown in Figure 2, the proposed method consists mainly of two components. The first one is the CL value predictor, implemented as a ResNet network; its goal given an image as input is to find the better *clip limit* value for processing that image with CLAHE. More specifically, given a grid of possible values for CL, this network outputs a probability distribution over such grid. Next, the image is processed with the CL value corresponding to the mode of this distribution. The second is a classification network implemented as via EfficientNet which takes the equalized image, with the selected CL, and predicts its class.

As for the training, both networks are trained in an end-to-end fashion. Due to the hard predictions of the first network (i.e. a given CL value chosen among a pre-defined grid), its output layer turn out to be non-differentiable; which makes the pair of networks not trainable via the standard back-propagation algorithm. To overcome this issue, only for training the networks, we replace the output layer of the first model with a Gumbel soft-max layer. Specifically, if the predicted probabilities for the possible N values of CL are denoted with $p = (p_1, \ldots, p_N)$, the Gumbel-soft-max output is computed as

$$y_{CL} = \mathbb{I}\left(\arg\max_{j\in[N]} \log_2(p_j) + g_j\right), \quad (1)$$

where the function $\mathbb{I} : [N] \to \{0,1\}^N$ provides the one-hot encoding of a given integer in $[N] \coloneqq \{1, \ldots, N\}$ as a binary vector in $\{0,1\}^N$, and each $g_j \coloneqq -\log_2(log_2(u_j))$ with u_j uniformly distributed in $(0,1)$. Notice that in this case y_{CL} is a binary vector, with a single 1 corresponding to the chosen CL value. Since the $argmax_j$ function introduces a non-differentiability, at training time we replace its Gumbel soft-max activation

$$y_{CL,j} = \frac{\exp(\log_2(p_j) + g_j)/\tau}{\sum_{j=1}^{N} \exp(\log_2(p_j) + g_j)/\tau}. \quad (2)$$

We make the following observations. First, $y_{CL,j}$ is now a real-valued vector whose mode some-how denote the preferred CL value. The Gumbel soft-max activation of 2 is parametrized by the scalar $\tau > 0$: small value of τ makes 2 a very good approximation of 1 with equality at the limit $\tau \to 0$; on the other hand $\tau \to \infty$ leads to a poor approximation with the function begin almost uniform. The choice of τ determines a trade-off between the approaximation quality and the numerical stability of the resulting gradients (see (Zhu et al. 2021) and the reference therein for further details on this trick). This layer is differentiable and then allows for a joint training of both network via back-propagation. The overall system, on the other hand, is trained to minimize the cross-entropy loss. We notice that one drawback of this proposal is an increasing training time w.r.t. a single network; this is however balanced by a substantial improvement in the prediction time, which is the goal of this paper.

Experiments

We compare the performance of our method (Predicted-CLAHE) against that of a network where images are not equalized (No-CLAHE) and a network where each image is equalized through a costly tuning of CLAHE (Tuned-CLAHE). The former is included only to provide a baseline for the inference time and further highlighting the benefits of an suitable pre-processing in the detection of diabetic retinopathy. The latter, for each image, performs a costly

Method	DS Tr	DS Ts	AUC	Se	Sp	Ac	Avg. Time (s)
No-CLAHE	M2	I	0.752	0.523	0.981	0.810	**0.799**
Tuned-CLAHE	M2	I	**0.816**	**0.679**	0.954	**0.859**	20.656
Predicted-CLAHE	M2	I	0.788	0.580	**0.997**	0.842	1.753
No-CLAHE	I	M2	0.656	0.751	0.562	0.702	**1.359**
Tuned-CLAHE	I	M2	0.755	**0.892**	0.619	0.721	38.844
Predicted-CLAHE	I	M2	**0.878**	0.726	**0.849**	**0.869**	1.879

Table 1: **Experimental evaluation**. **DS Tr** and **DS Ts** denote the training and the test dataset respectively, with M2 for MESSIDOR–2 and I for IDRiD. **AUC** denotes for Area Under Curve. **Se** stands for sensitivity. **Sp** stands for specifity. **Ac** stands for accuracy. **Avg. Time** is the average inference time across the predictions made on the test set.

grid-search over $\{2, 4, 6, \dots, 32\} \times \{1, 2, 3, \dots, 30\}$ for the pair (TS, CL) to optimize the entropy objective proposed in (Min et al. 2013). For each model, we measure several accuracy metrics and the average prediction time on the test data.

Datasets description. We used two real-world datasets, MESSIDOR-2 (Decencière et al. 2014) and IDRiD (Porwal et al. 2018). MESSIDOR-2 contains 1744 images while the IDRiD 516 testing images. Both datasets are made of images at different resolutions ranging from 1440×960 up to 4288×2848. Images in both datasets are classified according to the International Diabetic Retinopathy Scale into five classes that range from 0 up to 4 with 0 denoting healthy subjects and the higher levels denoting advanced stages of retinopathy.

Training details. First, all the images have been cropped to remove the black bands on the edges. Then, for computational purposes, we resized all the images to the resolution of 456×456 as also done in previous work (Pour et al. 2020). Finally, we reduced the problem to binary classification by merging the classes 0 and 1 and the classes $2, 3$ and 4. This is justified by the fact that level 1 denotes a fundamentally benign stage of the pathology and do not require any treatment, similarly to level 0. Moreover, a stage 1 diabetic retinopathy does not usually present symptoms at level of the eye fundus; which makes this class essentially equivalent to class 0. On the other hand, levels 2, 3 and 4 all require medical attention and a form of treatment. As classification networks, in all the experiments we resorted to a pre-trained EfficientNet $b5$. During training, we performed data augmentation with random horizontal and vertical flipping. Optimization is performed using SGD with batch-size 1 and learning rate is 0.01. Each model is trained for a maximum of 200 epochs and the drop-out rate is set to 0.1. On the other hand, the CL predictor has been designed to decide among thirty possible CL values in $\{1, 2, \dots, 30\}$ (this range is the same suggested in the seminal work (Zuiderveld 1994)) and has been implemented as a four layers ResNet. The TS is kept fixed to 8 in order to limit the complexity of the prediction network. Moreover, it has been observed that the TS plays a minor role w.r.t. CL (e.g.(Min et al. 2013; Joseph et al. 2017)); this is also supported in our experiments. Models are implemented in PyTorch and training is performed on a 16 GB NVIDIA Tesla P100 GPU accessed via Google Colab.

Results. Table 1 shows the experimental results. The first three rows report the performances of the considered methods when they are trained on the MESSIDOR-2 and tested on IDRiD; the last three rows instead of referring to the converse case. For the first setting, our proposal Predicted-CLAHE strikes a very good trade-off among predictive performances and inference time. As for the accuracy for example, it gets closer to the more expensive Tuned-CLAHE while predicting at a speed which is one order of magnitude faster. In the second case, instead of Predicted-CLAHE obtains superior predictive performances while retaining a small inference time. The large performance gap may be explained by the fact that, differently from Tuned-CLAHE, the equalization made by Predicted-CLAHE is optimized for the predictions. Indeed, the objective proposed in (Min et al. 2013) is related to the entropy of the equalized image. Specifically, the image pixel intensities are quantized into an histogram with a specified number of bins. The histogram is normalized into a (discrete) probability distribution and its entropy is calculated. The observation is that an image with a poor contrast will have most of its pixel intensities concentrated around a certain value (e.g. very bright image), this will result in a very low entropy. On the other hand, a well contrasted image will result in an almost flat histogram with maximal entropy. Notice however, that a flat histogram image, may have some of its details not suitably highlighted. Authors in (Min et al. 2013), observe that a *subjectively* good contrast can be obtained at the point where the entropy curve reached by varying CL and TS has maximal curvature. We notice, that while this approach is beneficial to improve the perceived quality of the image, it may lead to sub-optimal predictive performance, as the right level of contrast may be different from the one that is acceptable for a human eye. To further illustrate this phenomenon lets consider Figure 3. The image on the left, is an eye fundus image without equalization that is miss-classified by Tuned-CLAHE and correctly classified by our method. The central and the right images have been equalized with the grid search procedure of Tuned-CLAHE and the predicted CL from Predicted-CLAHE respectively. As it is possible to note, the central image, has a higher perceived quality than the one on the right which appears to be over-contrasted. However, in the right image, due to the high level of contrast many fine details are more evident, including the typical aneurysms of the diabetic rethinopathy. This is beneficial to the prediction task and indeed allows the correct classification.

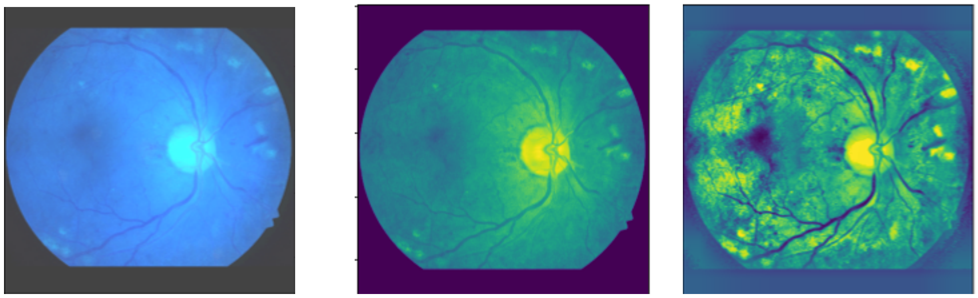

Figure 3: **Limitations of entropy based criteria**. On the left, a raw eye fundus image of a subject suffering from a severe diabetic retinopathy. In the center, the equalized image with $CL = 2$ as tuned by Tuned-CLAHE. On the right, the equalized image with $CL = 19$ as predicted by our method Predicted-CLAHE. Notice that Tuned-CLAHE, differently from Predicted-CLAHE, miss-classifies this image as that of an healthy subject. This comparison shows that while the first approach is beneficial to improve the perceived quality of the image, it may lead to sub-optimal predictive performance, since it doesn't sufficiently highlight the finer details (e.g. the many micro-aneurysms present in the eye fundus of this patient) that are important to detect the pathology.

Conclusions and future work

We have proposed a novel learning based approach to the problem of appropriately tuning the equalization parameters. The approach relies on a pair of networks that are trained in an end-to-end fashion. The proposed method strikes a trade-off between accurate predictions and reduced inference time. Promising preliminary experimental results support the validity of our proposal. We believe that these ideas can be further elaborated and ultimately lead to a better user experience of automatic diagnosis support systems. Besides, we believe this methodology is general enough to be applied also to other pre-processing mechanisms. As future work it would be interesting to extend the approach to the prediction of other equalization parameter, TS for example. It would be also interesting to evaluate the trade-off between the pre-processing network complexity, the accuracy and the resulting inference time; this will sheds light on the limiting performances of the proposed approach. Finally, a more extensive and comprehensive experimental evaluation, also including other equalization algorithms can serve a more solid benchmark for the proposed approach.

References

Abràmoff, M. D.; Lou, Y.; Erginay, A.; Clarida, W.; Amelon, R.; Folk, J. C.; and Niemeijer, M. 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative ophthalmology & visual science*, 57(13): 5200–5206.

Akram, M. U.; Khalid, S.; Tariq, A.; Khan, S. A.; and Azam, F. 2014. Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in biology and medicine*, 45: 161–71.

Badar, M.; Haris, M.; and Fatima, A. 2020. Application of deep learning for retinal image analysis: A review. *Computer Science Review*, 35: 100203.

Campos, G. F. C.; Mastelini, S. M.; Aguiar, G. J.; Mantovani, R. G.; Melo, L. F. d.; and Barbon, S. 2019. Machine learning hyperparameter selection for contrast limited adaptive histogram equalization. *EURASIP Journal on Image and Video Processing*, 2019(1): 1–18.

Cheung, N.; and Wong, T. Y. 2008. Diabetic retinopathy and systemic vascular complications. *Progress in retinal and eye research*, 27(2): 161–176.

Decencière, E.; Zhang, X.; Cazuguel, G.; Lay, B.; Cochener, B.; Trone, C.; Gain, P.; Ordonez, R.; Massin, P.; Erginay, A.; Charton, B.; and Klein, J.-C. 2014. Feedback on a publicly distributed database: the Messidor database. *Image Analysis & Stereology*, 33(3).

Fawzi, A.; Achuthan, A.; and Belaton, B. 2021. Adaptive Clip Limit Tile Size Histogram Equalization for Non-Homogenized Intensity Images. *IEEE Access*, 9: 164466–164492.

Gupta, A.; and Chhikara, R. 2018. Diabetic Retinopathy: Present and Past. *Procedia Computer Science*, 132: 1432–1440. International Conference on Computational Intelligence and Data Science.

Joseph, J.; Sivaraman, J.; Periyasamy, R.; and Simi, V. 2017. An objective method to identify optimum clip-limit and histogram specification of contrast limited adaptive histogram equalization for MR images. *Biocybernetics and Biomedical Engineering*, 37(3): 489–497.

Min, B. S.; Lim, D. K.; Kim, S. J.; and Lee, J. H. 2013. A novel method of determining parameters of CLAHE based on image entropy. *International Journal of Software Engineering and Its Applications*, 7(5): 113–120.

Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; and Meriaudeau, F. 2018. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data*, 3(3): 25.

Pour, A. M.; Seyedarabi, H.; Jahromi, S. H. A.; and Javadzadeh, A. 2020. Automatic detection and monitoring of diabetic retinopathy using efficient convolutional neural networks and contrast limited adaptive histogram equalization. *IEEE Access*, 8: 136668–136673.

Pratt, H.; Coenen, F.; Broadbent, D. M.; Harding, S. P.; and Zheng, Y. 2016. Convolutional neural networks for diabetic retinopathy. *Procedia computer science*, 90: 200–205.

Reddy Eswar, R. R. 2019. Dynamic Clipped Histogram Equalization Technique for Enhancing Low Contrast Images. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 89(2250-1762).

Sinthanayothin, C.; Boyce, J. F.; Cook, H. L.; and Williamson, T. H. 1999. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British journal of ophthalmology*, 83(8): 902–910.

Stitt, A. W.; Curtis, T. M.; Chen, M.; Medina, R. J.; McKay, G. J.; Jenkins, A.; Gardiner, T. A.; Lyons, T. J.; Hammes, H.-P.; Simo, R.; et al. 2016. The progress in understanding and treatment of diabetic retinopathy. *Progress in retinal and eye research*, 51: 156–186.

Zhu, M.; Han, K.; Wu, E.; Zhang, Q.; Nie, Y.; Lan, Z.; and Wang, Y. 2021. Dynamic Resolution Network. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 27319–27330. Curran Associates, Inc.

Zuiderveld, K. 1994. Contrast limited adaptive histogram equalization. *Graphics gems*, 474–485.

ACKNOWLEDGEMENT

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 877056. The m receives support from the European Union's Horizon 2020 research and innovation programme and by Spain, Italy, Austria, Germany, France, Finland, and Switzerland member states.