ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Numerical Technique For Testing the Coherence of Reported Performance Scores Of Image Segmentation Methods

Attila Fazekas[a], **György Kovács**[b]

[a]Faculty of Informatics, University of Debrecen
attila.fazekas@inf.unideb.hu

[b]Analytical Minds Ltd.
gyuriofkovacs@gmail.com

Abstract

Nowadays, there is a tremendous amount of research work done in explicitly (using online research platforms) or implicitly (based on publications about a specific dataset) organized competitions, where the goal is to find the "best" method to solve a specific, well-defined scientific problem. The researchers entering the competition are provided with the necessary data and benchmarks providing the basis for the evaluation and ranking of solutions. The mass appearance of this type of challenge has started a kind of "war of numbers", where the numerical values determined by the benchmarks quantify the efficiency and qualify the publication value of the method as a kind of objective measure.

However, these explicit or implicit competitions raise numerous questions, most importantly, are the publicly shared performance figures really comparable; is there a way to infer the actual way of evaluation; is there a way to check the consistency of publicly shared scores and the benchmark datasets?

In a previous paper of ours [4], we started this pioneering work by investigating the coherence of reported performance scores in the segmentation of vessels in retinal images. Based on the results of the consistency tests, we attempted to remove the identified biases to make a new baseline ranking.

The numerical techniques we have developed in [4] can be easily generalized and adapted to other problems where similar methodological flaws may appear. The primary conditions of applicability can be summarized as follows: a publicly avail-

able ground-truth manual annotation is available; the automated segmentations are evaluated and compared to the manual annotations in terms of accuracy, sensitivity and specificity (or other variations of the true/false positive/negative pixel counts). Preferably, the scores are shared at the image level, however, aggregations (like average scores) also enable some consistency testing with less power. The tests are based on the fact that the aforementioned commonly reported scores are based on pixel counts that need to be coherent with the foreground and background pixel counts of the benchmark images. In certain cases, for example, accidentally evaluating in smaller or larger areas than the original images – the bias can be estimated and removed.

In this paper, this numerical technique is presented in detail, defining the conditions of applicability more precisely. The proposed concistency check is illustrated on the problem of exudate detection in retinal images, the importance of which comes from the reliance of various diagnostic procedures on the presence of these anatomical structures in the retina. We investigate the coherence of the performance scores of 10 well-known exudate detection algorithms [1–3, 5–11] from the literature. The consistency checks enable the recognition of improper evaluations and once recognized, they can be adjusted or removed to derive a more realistic ranking of segmentation techniques in the field.

References

[1] K. Adem: *Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks*, Expert Systems With Applications 114 (2018), pp. 289–295.

[2] M. U. Akram, S. Khalid, S. A. Khan: *Identification and classification of microaneurysms for early detection of diabetic retinopathy*, Pattern Recognition 46 (2013), pp. 107–116.

[3] S. Banerjee, D. Kayal: *Detection of hard exudates using mean shift and normalized cut method*, Biocybernetics and biomedical engineering 36 (2016), pp. 679–685.

[4] A. Fazekas, G. Kovács: *A new baseline for retinal vessel segmentation: Numerical identification and correction of methodological inconsistencies affecting 100+ papers*, Medical Image Analysis 75 (2022), p. 1023.

[5] S. W. Franklin, S. E. Rajan: *Diagnosis of diabetic retinopathy by employing image processing technique to detect exudates in retinal images*, IET Image Processing 8 (2014), pp. 601–609.

[6] M. Fraza, W. Jahangira, S. Zahida, M. M. Hamayuna, S. A. Barman: *Multiscale segmentation of exudates in retinal images usingcontextual cues and ensemble classification*, Biomedical Signal Processing and Control 35 (2017), pp. 50–62.

[7] B. Harangi, A. Hajdu: *Automatic exudate detection by fusing multiple active contours and regionwise classification*, Computers in Biology and Medicine 54 (2014), pp. 156–17.

[8] P. Khojasteha, L. A. P. Júnior, T. Carvalhoc, E. Rezended, B. Aliahmada, J. P. Papae, D. K. Kumara: *Exudate detection in fundus images using deeply-learnable features*, Computers in Biology and Medicine 104 (2019), pp. 62–69.

[9] M. Maity, D. K. Das, D. M. Dhane, C. Chakraborty, A. Maiti: *Fusion of Entropy-Based Thresholding and Active Contour Model for Detection of Exudate and Optic Disc in Color Fundus Images*, Journal of Medical and Biological Engineering 36 (2016), pp. 795–809.

[10] M. Mateen, J. Wen, N. Nasrullah, S. Sun, S. Hayat: *Exudate Detection for Diabetic Retinopathy Using Pretrained Convolutional Neural Networks*, Hindawi Complexity (2020), pp. 1–11.

[11] D. Welfera, J. Scharcanskia, D. R. Marinhob: *A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images*, Computerized Medical Imaging and Graphics 34 (2010), pp. 228–235.