Low-Resolution Retinal Images: Detection and Mosaicing using Deep Learning Methods

Tales Veríssimo Souza Correia
School of Technology and Management, Polytechnic of Leiria

António Cunha
https://www.inesctec.pt/pt/pessoas/antonio-cunha
University of Trás-os-Montes and Alto Douro &
Institute for Systems and Computer Engineering, Technology and Science

Paulo Jorge Coelho
https://www.cienciavitae.pt/portal/pt/3818-FA4F-CC36
School of Technology and Management, Polytechnic of Leiria &
Institute for Systems Engineering and Computers at Coimbra

Abstract

Glaucoma is a severe eye disease that is asymptomatic in the initial stages and can lead to blindness due to its degenerative characteristics. This paper presents a framework for detecting the retinal fundus that is applied to lower-resolution images taken with a smartphone equipped with a D-EYE lens. A private dataset was assembled, annotated, and applied to several versions of the well-known YOLO object detector to evaluate their performance. Furthermore, some mosaicing techniques were evaluated and applied to the lower-resolution frames to verify their usefulness as a video summarization tool. Both YOLO v5 and v8 had similar performances, over 98% mAP(0.5) and 92.2%(0.5:0.95).

1 Introduction

Retinal imaging is a process that enables digital recording of the back of the eye. Pricey devices like fundus cameras often capture these, which deliver retinal images with superior detail and resolution for examination. With accurate retinal fundus detection, it is possible to extract meaningful information from such videos to process further to identify and locate important anatomical structures, such as veins, optic disc, fovea, macula, etc., to obtain an accurate diagnosis. Investigations of automatic or semi-automatic methods for retinal detection have been evolving to assist specialists. On the other hand, adopting simple lenses, such as D-EYE [1], can offer a variety of advantages, such as greater mobility, ease of use, enhanced patient comfort, and reduced costs. Therefore, it can be used to evaluate eye-related diseases in disadvantaged or remote populations. The drawback is that these images present lower quality when compared to those produced by professional fundus cameras, which makes it harder to identify anatomical features. Additionally, methods that can provide an in-place summary of the video captured by the examiner are relevant to alert, if necessary, of the need for the individuals to seek for specialized medical assistance.

The latest trends in research show the extensive use of convolution neural networks (CNN) for detecting the eye's fundus and detecting the disease(s). Nevertheless, these methods are focused on high-resolution retinal images, and there is still a lack of studies to evaluate the effectiveness of automatic methods to detect retinal regions in these low-resolution and low-quality retinal images. A similar situation occurs with the techniques applied to video summarization methods, mainly applied to high-quality images, and further investigation into lower quality and resolution is required.

This paper presents a framework focused on evaluating the performance of several versions of the You Only Look Once - YOLO network [2]–[6] to the fundus detection task on lower-resolution retinal images taken with a smartphone equipped with D-EYE lens. From these, it also evaluates the possibility of achieving a summary image (or small set of images) in a mosaicing-based approach [7]–[10] that translates the main information extracted from each individual video, aiming to provide a pre-diagnosis that can refer people, if necessary, to seek examination with the specialist. The dataset was created from 48 retina videos around the optic disc, with lower-resolution images. The ACRIMA pubic available dataset [11] was also used to improve the model's performance.

2 Methodology

This work is divided into two experiences (see the pipeline in Figure 1), applied to the dataset as follows.

Dataset

A dataset of 48 low-resolution videos of the optic papilla under myosis (undilated pupil) was captured from the left and right eyes. The videos were split into single images and organized in a dataset containing 380 frames. 150 frames are from the custom and private D-Eye dataset mentioned, with 1920x1080 pixels to be used to detect the visible retinal area (100 frames to train and 25 frames to the validation process). Additionally, from the ACRIMA dataset [11], 259 frames were used, 201 for training and 57 for validation. These present resolutions of 577x577 pixels. This small dataset, dully annotated, was applied solely to train the YOLO models to detect and extract the retinal region.

Setup

In the first part of the experience, the detection of the retinal visible area (dashed in blue in Figure 1) consists of computing the location of a rectangle that encloses the visible area in the image (the area of interest). The input images have 1920x1080 pixels. The dataset annotation must be divided into image and label folders, and inside each, separated into both train and validation data. For this purpose, the YOLO models from version 5 to version 8 were used [2]–[6], following previous contributions [12]. The mean average precision (mAP) is a popular metric between object detection methods to evaluate the model. Therefore, it will be used in this work as the main metric for evaluating the YOLO's performance. It is important to mention that several architectures present different labeling data formats, so it was required to re-format the information to meet all the versions.

In the second part of the experience, also depicted in Figure 1, involved by the dashed orange line, the mosaicing methods tested will process the cropped regions of interest from the previous part of the experience to extract as much information as possible, enhancing the details of the various retina frames available and delivering a single image as a result. The mosaicing technique can be divided into two main steps: image registration, where the image's key points are found and images are warped, and image blending, where image borders are smoothed. In this part of the experience, some studies were evaluated, namely Unsupervised Deep Image Stitching (UDIS) [7], Deep Image Stitching [8], Super Retina [9], and Multi-image Stitching [10]. In these tests, only the visual performance (subjective evaluation) of the methods was performed.

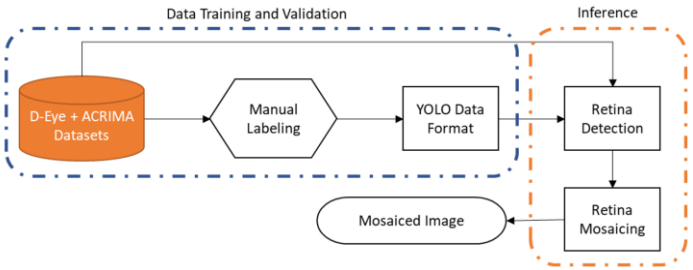

Figure 1: Pipeline diagram for the proposed low-resolution retinal detection and mosaicing framework.

For each version of YOLO models, six different tests were performed with the learning rate and epochs as variable parameters and keeping the same dataset for all. Other parameters were kept due to the time-limit operation of the machine used, which was a virtual machine from Google Colab. The momentum parameter was set to 0.937 for all tests, the image size normalized to 640 pixels, and the batch size selected was 32 due to speed and processing limitations. All tests were performed on a Tesla T4 with 12 GB RAM, using Python version 3.8.10, PyTorch 1.13.1, and Cuda version 116.

The mosaicing tests were run in a 10-core 10th generation Intel I9 10900 KF CPU, with 32GB of RAM, 1TB SSD, and a Nvidia RTX 3070 GPU.

3 Results and discussion

The framework was evaluated for the retinal visible area detection provided by various versions of YOLO and the mosaicing performance of state-of-the-art mosaicing methods.

For each image in the analysis, each YOLO method predicts the coordinates of the bounding box, the object class, and the confidence score of the prediction. The confidence score represents the probability that the prediction is correct and is used to filter out false positive detections. The mAP is a metric over different Intersection over Union - IoU thresholds (a pre-defined threshold for IoU between the predicted and ground truth bounding boxes to determine true positives and false positives). In these tests, the evaluation was performed using mAP metric (for both IoU higher than 0.5 and IoU values ranging from 0.5 until 0.95, in steps of 0.05).

Table 1: YOLO's mean average precision for each model.

Model	mAP(0.5)	mAP(0.5:0.95)
YOLO v5	**99.13**	92.20
YOLO v6	-	-
YOLO v7	95.58	56.74
YOLO v8	98.89	**92.26**

Table 1 displays the best result from each model for the mean average precision. Although YOLO v5 had a similar result to YOLO v8, the latter has a significantly better user interface, more parameters for tuning, better inference performance, and a faster training process. On the other hand, YOLO v6 did not perform as expected since the model couldn't converge after applying the aforementioned dataset, so no valid distinction between retinal and non-retinal areas was obtained. Additionally, these YOLO v6 metrics aren't available either. A larger dataset may be necessary to fulfill its potential. Furthermore, YOLO v7 model was unable to converge to have sufficient accuracy when comparing to YOLOv8, evident in the mAP(0.5:0.95).

Considering YOLO v8 as the best model globally, inference retina detection tests from the private dataset were executed in four videos (that were excluded from the training process). The videos were randomly selected and labeled as Samples 1 to 4, as depicted in Figure 2.

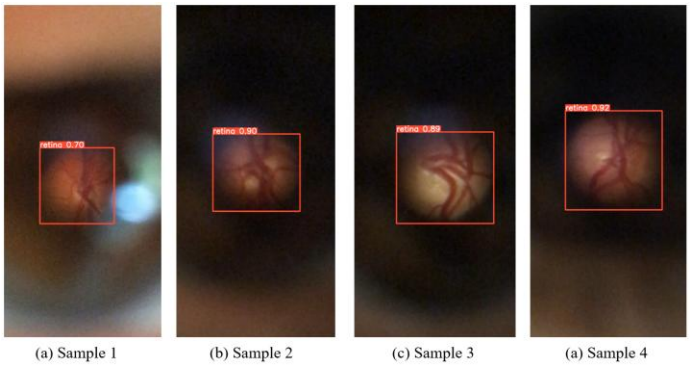

Figure 2: Example of the YOLO v8 retinal detection inference.

After the inference in the samples, a total of 2666 cropped images (the red rectangles) were acquired and are ready to feed the summarization methods.

To provide some summarization to the examiner, several mosaicing methods [7]–[10] were compared visually and the results were generally unsatisfactory. Figure 3 presents some results of the methods reported to perform well in high-quality images when applied to the lower resolution/quality ones.

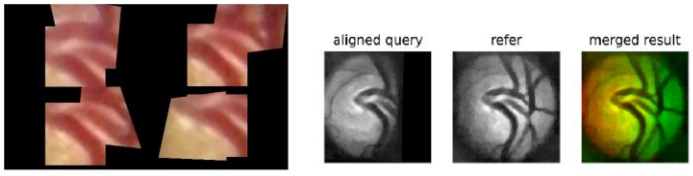

Figure 3: Examples of resulting images for the stitching process in low-quality retinal images. At left using Deep Stitching method [8], and on the right using Super Retina method [9].

Applying the 8 crops together with the Deep Image Stitching method [8], the results obtained from this inference are shown in the leftmost picture of Figure 3, and the method did not perform well enough to fit the objective of this work.

The best result in all the tested methods is depicted in Figure 3 on the right for the Super Retina method [9]. In this method, the pairs of images must be fed consecutively, one pair each time. In this example, 27 good key point matches were obtained between them, which is a reasonable amount for the quality of the images. The final stitched image, Figure 3 on the right, could show details from the two images but still not enough to fully expand the retina.

4 Conclusions

This paper evaluated a framework for retinal fundus detection based on YOLO methods on lower-resolution retinal images. A dataset of cropped images was assembled to evaluate the quality of the mosaicing technique in such lower-quality images. For the framework, both YOLO v5 and v8 had similar performances, over 98% mAP(0.5) and 92.2(0.5:0.95). Overall, the YOLO v8 has a faster training and user interface. The merging of the resulting images is still challenging due to the difficulty of establishing reliable key points through the selected images outputted from the mosaicing methods. Further developments in this subject are necessary.

References

[1] «Digital retinal camera | The Direct Ophthalmoscope for Your iPhone | Portable digital retinal camera | D-EYE». https://d-eyecare.com/ (acedido 16 de março de 2023).

[2] C.-Y. Wang, A. Bochkovskiy, e H.-Y. M. Liao, «Scaled-YOLOv4: Scaling Cross Stage Partial Network», em *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun. 2021, pp. 13024–13033. doi: 10.1109/CVPR46437.2021.01283.

[3] M. Horvat, L. Jelečević, e G. Gledec, *A comparative study of YOLOv5 models performance for image localization and classification*. 2022.

[4] C. Li *et al.*, «YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications». arXiv, 7 de setembro de 2022. Acedido: 15 de fevereiro de 2023. [Em linha]. Disponível em: http://arxiv.org/abs/2209.02976

[5] C.-Y. Wang, A. Bochkovskiy, e H.-Y. M. Liao, «YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors». arXiv, 6 de julho de 2022. Acedido: 15 de fevereiro de 2023. [Em linha]. Disponível em: http://arxiv.org/abs/2207.02696

[6] G. Jocher, A. Chaurasia, e J. Qiu, «YOLO by Ultralytics». janeiro de 2023. Acedido: 16 de fevereiro de 2023. [Em linha]. Disponível em: https://github.com/ultralytics/ultralytics

[7] L. Nie, C. Lin, K. Liao, S. Liu, e Y. Zhao, «Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images», *IEEE Trans. on Image Process.*, vol. 30, pp. 6184–6197, 2021, doi: 10.1109/TIP.2021.3092828.

[8] L. Nie, C. Lin, K. Liao, M. Liu, e Y. Zhao, «A view-free image stitching network based on global homography», *Journal of Visual Communication and Image Representation*, vol. 73, p. 102950, nov. 2020, doi: 10.1016/j.jvcir.2020.102950.

[9] J. Liu, X. Li, Q. Wei, J. Xu, e D. Ding, «Semi-Supervised Keypoint Detector and Descriptor for Retinal Image Matching». arXiv, 16 de julho de 2022. Acedido: 17 de março de 2023. [Em linha]. Disponível em: http://arxiv.org/abs/2207.07932

[10] R. Hu, R. J. Chalakkal, G. Linde, e J. S. Dhupia, «Multi-image Stitching for Smartphone-based Retinal Fundus Stitching», em *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, jul. 2022, pp. 179–184. doi: 10.1109/AIM52237.2022.9863260.

[11] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. Mossi, e A. Navea, «CNNs for automatic glaucoma assessment using fundus images: An extensive validation», *BioMedical Engineering OnLine*, vol. 18, mar. 2019, doi: 10.1186/s12938-019-0649-y.

[12] J. Camara, B. Silva, A. Gouveia, I. M. Pires, P. Coelho, e A. Cunha, «Detection and Mosaicing Techniques for Low-Quality Retinal Videos», *Sensors*, vol. 22, n.º 5, p. 2059, mar. 2022, doi: 10.3390/s22052059.