Improving Speed-Accuracy Trade-off in Face Detectors for Forensic Tools by Image Resizing

Deisy Chaves
Departamento IESA
Universidad de León
Researcher at INCIBE
dchas@unileon.es

Eduardo Fidalgo
Departamento IESA
Universidad de León
Researcher at INCIBE
efidf@unileon.es

Enrique Alegre
Departamento IESA
Universidad de León
Researcher at INCIBE
enrique.alegre@unileon.es

Pablo Blanco
Departamento IESA
Universidad de León
Researcher at INCIBE
pblanm@unileon.es

***Abstract*—During forensic material analysis, accurate and fast face detection is required prior to facial recognition of fugitives or children in sexual abuse scenes. However, this is not easy due to common limitations in the image quality or face pose. Moreover, real-time performance is expected in some applications as the forensic ones. There are several methods to address the face detection problem, but most of them are not suitable for real-time applications due to its computational complexity. In this work, we propose a strategy based on image resizing especially valid for Child Sexual Abuse crimes, oriented to improve the trade-off between the speed and the performance of three deep-learning-based face detectors. The results showed that the proposed approach is able to speed up face detection with a small reduction in accuracy. The best speed-accuracy trade-off is achieved using images resized to 50% of the original image size.**

***Index Terms*—Face detection, Forensic images, CSA, Deep learning**

Contribution type: *Ongoing Research*

I. Introduction

Face detection has applications in different fields such as security, bio-metrics and health-care. This process is crucial in some forensics and law enforcement [1] activities, since an accurate and fast face detection is required as an previous step in other tasks such as surveillance, fugitives recognition, sexual abuse detection, among others. However, several factors difficult a correct face detection, such as are variations in pose, expression, image resolution, and illumination [2], which are common issues found in some forensic images.

There are several approaches to address the face detection problem. In [3], Viola and Jones presented the first framework for real-time face detection based on a sliding window search using the AdaBoost algorithm with hand-crafted features (Haar descriptors). Recently, most face detectors focus on using features learnt from a Convolutional Neural Network (CNN) [4], [5], [6], [7], [8], which increase performance significantly in complex detection conditions, e.g. low/high illumination, blur and face occlusion. Nevertheless, most of these face detectors are focused on improving detection accuracy under challenging conditions without taking into account the processing time. Thus, their application for analysing large amounts of data where real-time performance is desired, e.g. forensics images, requires the development of strategies to speed up detection.

In this work, we evaluated the trade-off between processing time and accuracy detection through an image resizing strategy. We assessed three of the best, in terms of speed or accuracy, face detection methods presented in the last three years —Multi-Task Cascade CNN (MTCNN) [4], PyramidBox [8] and Dual Face Shot Detector (DSFD) [7]— using a set of images selected from the dataset UFDD [9] with similar characteristics in regards to the number of people found in Child Sexual Abuse (CSA) images. We are interested in this problem because this work is framed on the European project Forensic Against Sexual Exploitation of Children (4NSEEK) and in the research lines defined by the Framework agreement between INCIBE and the University of León. Preliminary results showed that our strategy improves the processing time of the evaluated face detector methods with a moderate reduction in accuracy.

II. Methodology

Fig. 1 shows the data flow of the strategy used to improve face detection processing time. First, images are resized to a percentage of their original size to keep the proportions of faces and objects contained on the image. Second, face detection is performed on the resized image using a state-of-art method. Finally, detected bounding boxes containing face locations are scaled back to the original image dimensions and returned as output.

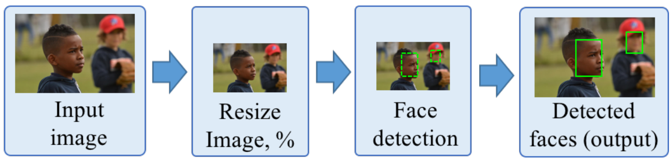
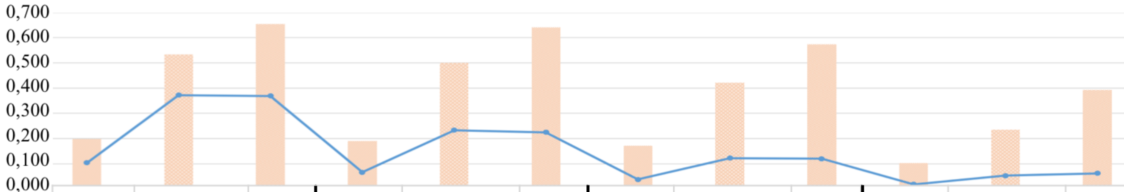


Figure 1. Proposed face detection strategy

We selected the three deep-learning-based methods indicated previously for face detection. The first method, MTCNN [4] performs both face detection and alignment using a multi-task training. Custom networks are used for proposing and refining regions that contain faces. MTCNN detects effectively faces not initially aligned, and it is widely used due to their fast detection performance. In particular, this method is the one currently integrated on the Evidence Detector software, provided from INCIBE to the *Policia y Guardia Civil Española* (Police and Law Enforcement of Spain). The second method, PyramidBox [8] combines context semantic with hierarchical features from an extended VGG16 [10] architecture and an anchor scale design strategy. This method performs better at detecting small faces. The last method used, DSFD [7] integrates features obtained from a VGG16 architecture with enhanced features to detect faces accurately.



	MTCNN	PyramidBox	DSFD	MTCNN	PyramidBox	DSFD	MTCNN	PyramidBox	DSFD	MTCNN	PyramidBox	DSFD
mAP	0,199	0,534	**0,654**	0,190	0,500	**0,641**	0,171	0,422	**0,573**	0,103	0,236	**0,392**
Speed (sec)	0,104	0,372	0,369	0,066	0,234	0,225	0,038	0,123	0,120	0,019	0,053	0,062
Resized image	**100%**			**75%**			**50%**			**25%**		

Figure 2. Speed and accuracy (mAP) trade-off results for MTCNN, PyramidBox and DSFD face detection methods using four different image resolutions

PyramidBox and DSFD methods were selected due to their accurate performance on complex detection conditions.

III. Experimental Results

The evaluation of the face detectors —MTCNN, DSFD and PyramidBox— was carried out using a subset of 5672 images chosen from the UFDD dataset [9], containing less than four people per scene which is the average maximum number of individuals observed on images of CSA. These images were taken considering several acquisition conditions to evaluate the robustness of the detectors: rain, snow, haze, blur, high/low illumination, lens distortion and distractors, i.e. images without any human faces. Moreover, four image sizes —original resolution, and images resized to 75%, 50% and 25% of the original size— were considered to determine the best trade-off between speed and accuracy in face detection. The accuracy was evaluated using the mean Average Precision (mAP) metric [11], which combines the precision and the recall measures by summarising the shape of the precision-recall curve considering different overlapping thresholds. All the experiments were performed on a GNU/Linux machine box running Ubuntu 16.04, with 32GB RAM, using a 6Gb GTX-1060 NVIDIA Card.

Fig. 2 presents the mAP values and processing times computed for the input images with the four evaluated sizes. As it can be observed, MTCNN is the faster detector, while DSFD is the most accurate for the evaluated image sizes. Furthermore, the use of resized images improves detection speed with a reduction of the mAP values related to the percentage of image resizing. A large resized image percentage leads to a substantial mAP decrease in comparison to the mAP values obtained using original images.

The use of images resized to 75% improved the detection speed more than a 36,37% when compared to the original images, with a slight reduction of mAP values —a maximum decrease of 6,38%. The analysis of images resized to 25% increased the detection speed more than 82,05% in comparison to using original images, but there is a significant reduction of mAP values —a maximum decrease of 55,80%. Finally, the best trade-off between speed and mAP is achieved using images resized to 50%. In this case, the detection speed improved more than 63,44% in comparison to the original images with a maximum decrease of 20,97% for mAP values.

IV. Conclusions

In this work, we proposed an image resizing strategy to speed up three face detector methods —MTCNN, DSFD and PyramidBox— while minimizing the performance drop. The experimental results showed that the use of resized images to 75% and 50% of the original image size allows for a significant improvement in processing time with a small reduction of the mAP values. However, further image resizing decreases face detection performance. All in all, the proposed strategy makes it possible to use complex face detectors such as DSFD and PyramidBox on real-time forensic applications such as CSA detection.

As future work, a dataset with problem domain images will be created to fine-tune face detectors and improve accuracy (mAP) performance.

Acknowledgement

This work was supported by the framework agreement between the Universidad de León and INCIBE (Spanish National Cybersecurity Institute) under Addendum 01. Also, this research has been funded with support from the European Commission under the 4NSEEK project with Grant Agreement 821966. This publication reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

References

[1] A. Gangwar, E. Fidalgo, E. Alegre, and V. González-Castro, in *ICDP*, 2017, pp. 37–42.

[2] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput Vis Image Underst.*, vol. 138, pp. 1–24, 2015.

[3] P. Viola and M. J. Jones, "Robust real-time face detection," *Int J Comput Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[4] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Proc Let*, vol. 23, no. 10, pp. 1499–1503, 2016.

[5] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "S^{3}FD: Single shot scale-invariant face detector," in *ICCV*, 2017, pp. 192–201.

[6] J. Zhang, X. Wu, J. Zhu, and S. C. H. Hoi, "Feature agglomeration networks for single stage face detection," *CoRR*, vol. abs/1712.00721, pp. 1–12, 2017.

[7] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, "DSFD: dual shot face detector," *CoRR*, vol. abs/1810.10220, pp. 1–10, 2018.

[8] X. Tang, D. K. Du, Z. He, and J. Liu, "Pyramidbox: A context-assisted single shot face detector," in *ECCV*, 2018, pp. 1–17.

[9] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, "Pushing the limits of unconstrained face detection: a challenge dataset and baseline results," *CoRR*, vol. abs/1804.10275, pp. 1–10, 2018.

[10] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *ACPR*, 2015, pp. 730–734.

[11] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserma, "The pascal visual object classes (VOC) challenge," *Int J Comput Vis.*, vol. 88, no. 2, pp. 303–338, 2010.