Detecting faces in specific scenarios: Systematic Literature Review*

Bruno Gonçalves Dias[1], **Victor Soares Ivamoto**[1], **Clodoaldo Aparecido de Moraes Lima**[1]

[1]Escola de Artes, Ciências e Humanidades (USP)
São Paulo, SP

brunogdias@usp.br, vivamoto@usp.br, c.lima@usp.br

Abstract. *Facial detection is a base component for multiple applications in the fields of biometrics, surveillance, human-robot interaction and others. Although significant progress has been made in the field over the past decade, there are still gaps to be addressed, particularly in specific scenarios as the presence of partial occlusion, variations of lighting, pose, and scale among others. This work aims to provide a comprehensive evaluation of recent studies on facial detection in the wild through a systematic literature review. The review includes a focus on the use of scenario-specific information within the field. A total of forty-five papers were analyzed to provide an overview of the field, incorporating information on scenarios.*

1. Introduction

Facial detection plays a crucial role in various applications such as biometrics, surveillance, and human-robot interaction systems among others. Despite its early development in the 1960s and 1970s, significant progress was only made in the 2000s with the introduction of the Viola and Jones face detector [Viola and Jones 2001]. However, this detector had limitations in detecting non-frontal faces, partial occlusions, and variations in expression, scale, and lighting. To overcome these challenges, the field of facial detection focused on developing robust detectors for uncontrolled environments.

Advancements in facial detection research have been driven by the availability of databases and benchmark protocols, open-source code repositories, and improved classification and feature extraction techniques. Databases like FDDB and WIDER FACE have contributed to the field by providing a wide range of images for training and evaluation. Open-source code repositories such as OpenCV, CAFFE, and PyTorch have facilitated the development and adoption of more efficient algorithms. Feature extraction techniques like Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) have been refined, and classification algorithms based on boosting and deep neural networks have shown significant improvements.

Convolutional Neural Networks (CNNs) have played an important role in advancing facial detection in uncontrolled environments. These networks have benefited from the "third wave" [Venkatesan and Li 2018] of research around deep neural networks and have been instrumental in handling the challenges posed by complex scenarios. Object

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

detection competitions, such as ImageNet, have also contributed to innovations in face detection research.

While progress has been made, facial detection in uncontrolled environments remains an unresolved problem. Scenarios involving variations in skin color, expression, makeup, and especially partial occlusions, including the increased use of masks during the COVID-19 pandemic, pose ongoing challenges. This paper aims to explore the literature on face detectors and their performance in different scenarios, focusing on the impact of partial occlusion, lighting variation, pose, and facial expression on detector performance. By addressing these challenges, researchers hope to enhance the effectiveness of facial detection systems in uncontrolled environments.

This review aims to benefit any researcher in the field of face detection in the wild, providing a comprehensive evaluation of recent studies. With the focus on scenarios, this review also wants to promote a possible path for research efforts where a gap was found. The present literature review does not intend to be an exhaustive overview of the field, it aims instead to be a quality resource for relevant aspects of face detection in the wild, with a special focus on the impact of scenarios.

The paper is divided into three phases of Systematic Literature Review: (1) the planning, (2) the conducting, and (3) the report phase. In the end, there is a conclusion summarizing the main points.

2. Plan

Based on the instructions provided by Kitchenham and Charters [Kitchenham and Charters 2007], the following aspects will be presented in this phase: (1) justification; (2) research questions; (3) primary study search strategy (4) quality assessment criteria. These aspects compose the systematic and structured approach used to conduct the review.

2.1. Justification

The present Systematic Literature Review (SLR) is justified for two reasons: (1) a gap in the literature regarding secondary studies considering the impact of scenarios; (2) providing a foundation for further studies in the field that aspire to consider scenarios in their investigation efforts. Although there are other secondary works in the area of face detection in uncontrolled environments, such as the papers by Feng et al. [Feng et al. 2022] and Zafeiriou et al. [Zafeiriou et al. 2015], these studies do not have an understanding on the use of scenario information in the field.

2.2. Research Questions

Follow the research questions: (Q1) What are the databases used?; (Q2) Which scenarios are explicitly considered?; (Q3) What metrics are used?; (Q4) What feature extraction techniques are employed?; (Q5) What classification techniques are utilized?

2.3. Primary Study Search Strategy

The search engines used were Scopus, IEEE, and ScienceDirect. The canonical search string used in the search engines is provided below:

(face OR facial) AND (detect OR segment*) AND (unconstrained OR uncontrolled OR wild)*

2.4. Primary Study Selection Criteria and Procedures

The studies were selected using a set of inclusion criteria (hereafter referred to as ICs) and exclusion criteria (hereafter referred to as ECs). For a study to be considered relevant for the review, it must meet all the ICs and not meet any of the ECs.

The specific ICs used in the selection process are described as follows: (IC-1) the study focuses solely on facial detection as a specific topic; (IC-2) the study exclusively addresses facial detection in uncontrolled environments; (IC-3) the study proposes a new technique.

The ECs: (EC-1) the reference is not available electronically; (EC-2) the study was published before or in the year 2001, which is the year when the Viola-Jones (VJ) detector was published. The VJ detector represents a paradigm shift in the field of facial detection; (EC-3) the study is not a primary study; (EC-4) the study uses specific cameras or techniques based on 3D imaging.

The papers were partially read for the application of ICs and ECs. The reading included the title, abstract, and conclusion.

2.5. Quality Assessment Criteria and Procedures.

For the evaluation of the quality of primary studies, criteria, and procedures were used. The criteria aim to examine the contribution and replicability of the works. The quality criteria are: (QC-1) are the research objectives clearly defined?; (QC-2) Were the face detection techniques clearly presented?; (QC-3) Were the used hyperparameters detailed to allow experiment reproduction, if necessary?; (QC-4) Was the utilized code made available?; (QC-5) Were the training techniques clearly presented? (QC-6) Are the used databases publicly available?; (QC-7) Is the article's proposal compared to other proposals?

For each criterion, a particular study can score in three ways: zero "0" if the criterion was not addressed; half a point "0.5" if the criterion was partially addressed; and one point "1" if the criterion was fully addressed. For the quality assessment, the articles were read in their entirety.

3. Conduction

3.1. Execution of the search string

The canonical search string was executed on three search engines: (1) Scopus, (2) IEEE Xplore, and (3) ScienceDirect. The search took place in February 2022. A total of 187 records were retrieved from Scopus, 153 records from ScienceDirect, and 92 records from IEEE Xplore, resulting in a total of 432 unique records. Subsequently, a manual analysis of the 432 identified records was conducted to ensure their relevance and alignment with the scope of the SLR.

3.2. Summary of the selection and quality assessment

The resulting records from the execution of the search string underwent a selection process involving three steps: (1) application of exclusion criteria, (2) application of inclusion criteria, and (3) application of quality criteria.

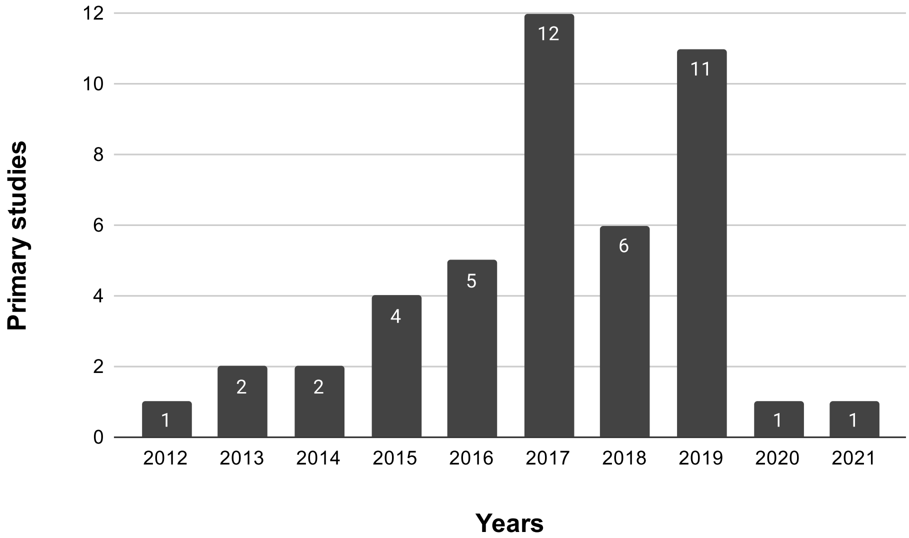

Figure 1. Primary studies per year

In the first step, the exclusion criteria were applied. This resulted in the elimination of 37 studies. In the second step, the inclusion criteria were applied. This resulted in 350 studies being eliminated, with IC-1 and IC-2 being the most relevant criteria for study exclusion. The most common case of exclusion was studies that focused on face detection along with other tasks such as alignment and/or identification. After the first two steps, 45 studies remained. The third step was the quality assessment and the results are presented in Appendix.

4. Report

This section presents the results observed in the conducted SLR. The presented results are structured according to the research questions outlined in subsection 2.2. Before presenting the research questions outcome, the chart for the primary studies per year will be presented in Figure 1.

The chart illustrates a rise from the year 2012 until the years 2017-2019. This rise coincides with the "third wave" of neural network research [Venkatesan and Li 2018], specifically with the publication of significant works in the field of Object Detection using CNNs such as AlexNet [Krizhevsky et al. 2012], VGGNet [Simonyan and Zisserman 2014], and ResNet [He et al. 2016]. The impact on the object detection field had a spin-off effect on face detection research.

The following is the structured analysis by research question.

4.1. The databases (Q1)

The first research question focuses on the databases used in the primary studies. Table 1 presents the records of the databases used. Records here stand for citation of each database in primary studies. A considerable number of studies mention more than one database, which is why the sum of the records in the chart exceeds the number of analyzed primary studies. The table also contains the number of images in each dataset, the number of

Table 1. Analysis of databases

Reference	Records	Images	Faces	E. I.
FDDB [Jain and Learned-Miller 2010]	33	2,845	5,171	No
WIDER FACE [Yang et al. 2016]	20	32,203	393,703	Yes
Others	14	-	-	-
AFW [Zhu and Ramanan 2012]	12	205	468	Yes
PASCAL FACES [Yan et al. 2014]	9	4,087	8,566	Yes
Own Database	4	-	-	-
AFLW [Martin Koestinger and Bischof 2011]	4	21,997	25,993	Yes
LFW [Huang et al. 2007]	4	13,233	13,233	No
AR [Martinez and Benavente 1998]	4	3,000	3,000	Yes
GENKI [`http://mplab.ucsd.edu`]	4	7,500	-	Yes
CMU-MIT [Rowley et al. 1998]	4	130	511	Yes

labeled faces, and if a dataset presents "Extra Information" (E.I.). "Extra Information" stands for any information beyond the face label that can be used to characterize some sort of scenario as partial occlusion, variation of pose, lighting, or scale.

The table presents the databases in order, from the most cited ones to the least. The categories "Others" and "Own Database" deserve an explanation. The "Others" category includes any database that was only mentioned once within the scope of the review. These databases were considered not relevant for describing the state of the art. The category "Own Database" encompasses databases mentioned in the studies but not publicly available.

Among the datasets considered, there is a size variation: from the smallest with 468 labeled faces (AFW) to the largest with 393,703 labeled faces (WIDER FACE). This brings non-uniformity to the detectors' validation. However, WIDER FACE is the most recent and the second-largest in terms of records - that indicates a trend for the largest database to become the main reference in the field. This is a positive trend in terms of the comparison between detectors and for robustness of assessment.

There is also a variation on the "Extra Information" provided beyond the labels of the faces. AFW, AFLW, and PASCAL FACES present face landmarks indicating face features position such as eyes, nose, or center of eyes - each dataset has its pattern. AFW, AFLW, and GENKI present head pose descriptions by providing yaw, pitch, and row directions. WIDER FACE, GENKI, and AR present binary information about specific characteristics of the labeled faces. For example, WIDER FACE presents information about the degree of blur and occlusion of a labeled face; and binary information about other characteristics such as variation of face expression, illumination, and pose. The GENKI dataset presents information if there is a smile or not in the face. And AR lists thirteen characteristics - for example, smile, anger, left light on, both sides light on, sunglasses, and more.

Differences can be found in how the datasets provide face labels. For example, FDDB provides face labels in an ellipse format. The WIDER FACE and PASCAL FACES provide face labels in a box format, but describe the box with different values. WIDER FACE describes the box using these four values: the x and y values of the top-left point

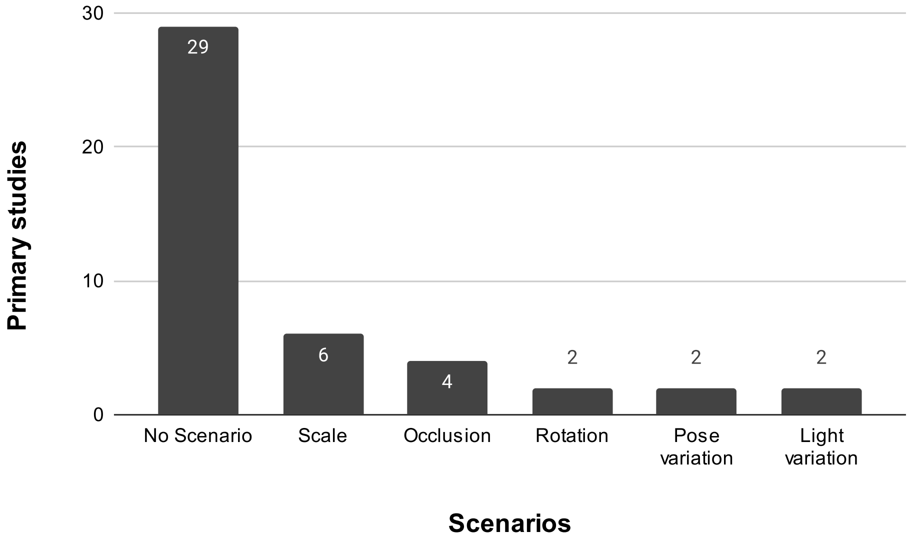

Figure 2. Primary studies per scenario

of the box, width, and height. While PASCAL FACES describes the box by these other four values: the x and y values of the bottom-left point of the box; and the x and y of the right-top point of the box. There is the AFLW, that provides face labels as ellipses and as boxes.

Regarding color and grayscale images, only AR and CMU-MIT provide only grayscale images. The rest of the datasets provide color images.

4.2. Usage of scenarios (Q2)

To create the chart presented in Figure 2, each study analyzed in this review was classified into one of the categories. The category "No scenario" encompasses all studies that do not differentiate between scenarios, in the sense that the developed classifier aims to address face detection in general and any scenario info was used to assess performance. This category has the highest number of studies.

The category with the second-highest number of studies is "Scale" with all studies that have detectors dedicated to detecting faces in scenarios where scale is the main challenge. The detectors aim to overcome scale challenges, meaning environments where very small and/or very large faces are encountered. The "Occlusion" category, the third largest, houses studies focused on overcoming the challenge of detecting partially occluded faces. The detectors here focus on detecting partially occluded faces. For example, in the case of the study by Lin et al. [Lin et al. 2016].

The last three categories have an equal number of primary studies in each. In the "Rotation" category, detectors focused on the issue of faces in different rotational axes than regular. For example, the study by Shi et al. [Shi et al. 2018]. The "Pose Variation" category contains studies that propose detectors dedicated to scenarios with significant pose variation. For example, the work by Ravidas S. [Ravidas 2019]. And the "Illumination" category houses works that propose techniques focused on overcoming the challenge of illumination variation. For example, as in the work by Li et al. [Li et al. 2017b].

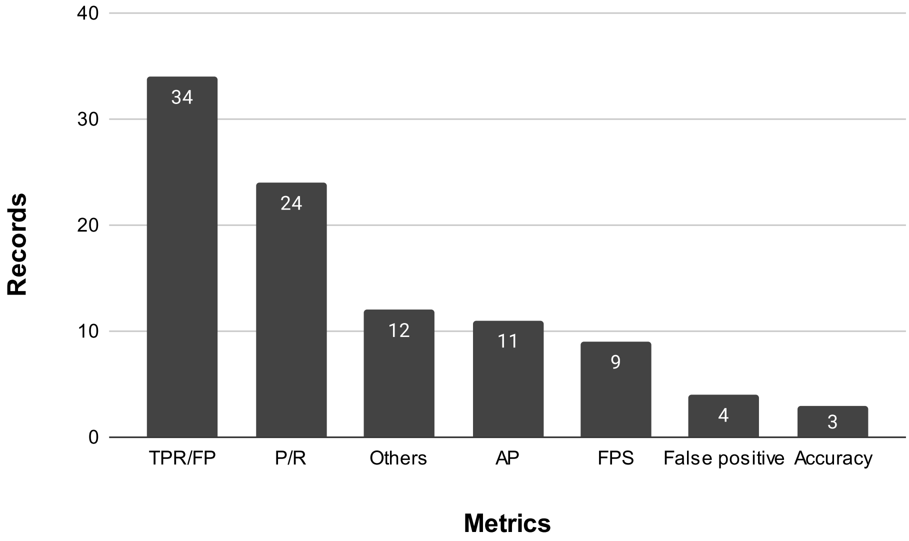

Figure 3. Records per metric

It is observed that most studies developed detectors using no scenario information. That means, these detectors do not use scenarios in the development or the analysis of results. For example, it is not analyzed whether the classifier performs better or worse in specific scenarios. This highlights a gap in the use of scenarios in detectors development and analysis. Also, the information beyond the labels provided by the databases is also not utilized by the majority of detectors analyzed in this review.

4.3. Metrics (Q3)

In the chart in Figure 3, there is the usage of metrics among the studies. Must be considered that most articles use more than one type of metric. Therefore, the sum of the records will be greater than the number of primary studies analyzed in this review.

The first and most widely used metric is described in the chart as “TPR/FP”, representing the True Positive Rate (TPR) to False Positive (FP) ratio. This metric is used by the FDDB benchmark. The second metric is the Precision and Recall “P/R” ratio. This metric is used in the WIDER FACE benchmark. In the third position, the “Other” category is observed, which includes all metrics with only one record. These metrics were considered not relevant to describe the literature. “AP” or Average Precision is the fourth most frequently cited metric.

As the fifth category with the most records, “FPS” or Frames per Second is observed. Among the studies that used FPS as the speed measurement unit, the most performant detectors include the study by Li et al. [Li et al. 2017b] with 100 FPS, the study by Zeng et al. [Zeng et al. 2019] with 60 FPS, and the study by Liao et al. [Liao et al. 2016] with 29.28 FPS. The least performant in terms of speed, measured in FPS, was the detector from the study by Zhang et al. [Zhang et al. 2017b] with 10.2 FPS. All values were measured in CPU tests, but they cannot be directly compared because not all CPU configurations and image sizes used were provided. The values shown serve as a reference to understand the speed of detectors in the state of the art.

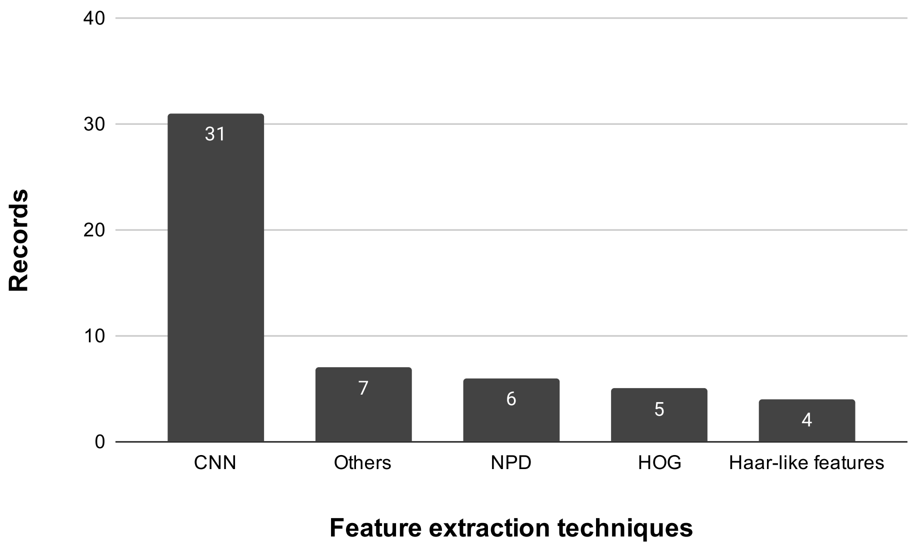

Figure 4. Records per feature extraction technique

4.4. Feature Extraction (CQ4)

The chart shown in Figure 4 analyzed the number of records per feature extraction technique. The sum is greater than the number of studies in the review because some works use more than one feature extraction technique. The work by Nanni et al. [Nanni et al. 2019] is an example of a study that uses multiple feature extraction techniques.

The technique with the highest number of records is the CNN-based technique. CNNs extract visual features from images through stages of filter and image reduction. It can be observed that feature extraction using CNNs is dominant in the state-of-the-art: the number of records for CNN exceeds the sum of all other categories.

In the second category with the most records, there is “Other,” referring to all extraction techniques with only one record. In the third category, there is “NPD,” which stands for Normalized Pixel Difference [Liao et al. 2016]. In the fourth category, with four records, there is “HOG” or Histogram of Oriented Gradients. These are widely used in combination with Support Vector Machine-based detectors. In the fifth position, there are Haar-like features, such as those used in the Viola-Jones detector.

4.5. Face or Non-face Classifiers (CQ5)

The chart shown in Figure 5 analyzed the number of records per face or non-face classification technique. The sum is greater than the number of works in the review because some works use more than one classification technique.

In the category with the most records, there is CNN. The CNNs can be used both as feature extractors and as classifiers. In the second category, there is the Deep Quadratic Tree classifier. This classifier is commonly used in combination with the NPD technique for feature extraction. In the third category, there is “Boosting Classifiers”, which represents classifiers generated using algorithms such as Adaboost or Gentle AdaBoost. In the

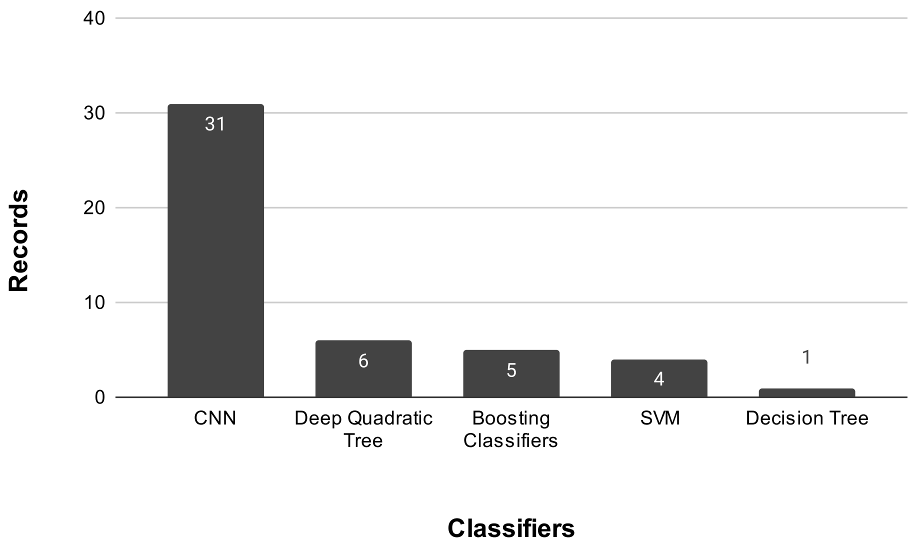

Figure 5. Records per classifier

fourth category, there is “SVM” or Support Vector Machines. And in the fifth position, with one record, there is the Decision Tree classifier.

5. Conclusion

This work had as its main goal to provide a comprehensive evaluation of recent studies on facial detection in the wild, with a special focus on the use of specific scenario information. After the analysis, it was noticeable that under 36% of the studies (16 from 45) made use of specific scenario information in order to develop or evaluate their propositions. A considerable gap in the exploration of this information, taking into account that among the nine datasets most cited, seven provide information that can be used to describe scenarios.

Beyond the scenario information analysis, it is noticeable that WIDER FACE has a perspective to become the main database in the field of face detection in the wild. The dataset is the most recent and has the second-largest number of records. Also, our work confirms a tendency observed by many over the last decade: CNN is a widely used technique in face detectors present in the state of the art.

The present RSL described an overview of the literature on face detection in uncontrolled environments, while calling attention to the gap in the use of scenario-specific information on the proposition of new techniques. This work stands as a reference for future works that intend to fill this gap or other research efforts in need of a quality source of resources.

References

Alafif, T., Hailat, Z., Aslan, M., and Chen, X. (2017). On detecting partially occluded faces with pose variations. In *2017 14th International Symposium on Pervasive Systems, Algorithms and Networks 2017 11th International Conference on Frontier of Computer Science and Technology 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC)*, pages 28–37.

Bai, Y. and Ghanem, B. (2017). Multi-scale fully convolutional network for face detection in the wild. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2078–2087.

Chai, Z., Zhang, Y., Du, Z., Wang, D., and Méndez-Vázquez, H. (2014). Learning flexible block based local binary patterns for unconstrained face detection. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

Chen, Q., Shen, F., Ding, Y., Gong, P., Tao, Y., and Wang, J. (2018). Face detection using r-fcn based deformable convolutional networks. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 4165–4170.

Deng, J. and Xie, X. (2017a). Detect face in the wild using cnn cascade with feature aggregation at multi-resolution. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4167–4171.

Deng, J. and Xie, X. (2017b). Nested shallow cnn-cascade for face detection in the wild. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 165–172.

El-Barkouky, A., Shalaby, A., Mahmoud, A., and Farag, A. (2014). Selective part models for detecting partially occluded faces in the wild. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 268–272.

Feng, Y., Yu, S., Peng, H., Li, Y.-R., and Zhang, J. (2022). Detect faces efficiently: A survey and evaluations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1):1–18.

Ge, S., Li, J., Ye, Q., and Luo, Z. (2017). Detecting masked faces in the wild with lle-cnns. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 426–434.

Gul, S. and Farooq, H. (2015). A machine learning approach to detect occluded faces in unconstrained crowd scene. In *2015 IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC)*, pages 149–155.

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

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

Jain, V. and Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst.

Jiang, F., Zhang, J., Yan, L., Xia, Y., and Shan, S. (2018). A three-category face detector with contextual information on finding tiny faces. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2680–2684.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. 2.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Li, J., Karmoshi, S., and Zhu, M. (2017a). Unconstrained face detection based on cascaded convolutional neural networks in surveillance video. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pages 46–52.

Li, J., Liu, L., Li, J., Feng, J., Yan, S., and Sim, T. (2019). Toward a comprehensive face detector in the wild. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1):104–114.

Li, J., Zhang, D., Zhang, K., Hu, K., and Yang, L. (2017b). Real-time face detection during the night. pages 582–586.

Li, X.-X., Liang, R., Gao, J., and Wang, H. (2015). Facial occlusion detection via structural error metrics and clustering. In He, X., Gao, X., Zhang, Y., Zhou, Z.-H., Liu, Z.-Y., Fu, B., Hu, F., and Zhang, Z., editors, *Intelligence Science and Big Data Engineering. Image and Video Data Engineering*, pages 118–127, Cham. Springer International Publishing.

Li, Z., Tang, X., Wu, X., Liu, J., and He, R. (2020). Progressively refined face detection through semantics-enriched representation learning. *IEEE Transactions on Information Forensics and Security*, 15:1394–1406.

Liao, S., Jain, A. K., and Li, S. Z. (2016). A fast and accurate unconstrained face detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):211–223.

Lin, S., Cai, L., Lin, X., and Ji, R. (2016). Masked face detection via a modified lenet. *Neurocomputing*, 218:197–202.

Liu, Y. and Levine, M. D. (2017). Multi-path region-based convolutional neural network for accurate detection of unconstrained "hard faces".

Lv, J.-J., Feng, Y.-J., Zhou, X.-D., and Zhou, X. (2016). Face detection using hierarchical fully convolutional networks. In Tan, T., Li, X., Chen, X., Zhou, J., Yang, J., and Cheng, H., editors, *Pattern Recognition*, pages 268–277, Singapore. Springer Singapore.

Magalhaes, J. P., Ren, T. I., and Cavalcanti, G. D. C. (2012). Face detection under illumination variance using combined adaboost and gradientfaces. In Yin, H., Costa, J. A. F., and Barreto, G., editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2012*, pages 435–442, Berlin, Heidelberg. Springer Berlin Heidelberg.

Martin Koestinger, Paul Wohlhart, P. M. R. and Bischof, H. (2011). Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization. In *Proc. First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies*.

Martinez, A. and Benavente, R. (1998). *The AR Face Database: CVC Technical Report, 24*.

Marčetić, D., Hrkać, T., and Ribarić, S. (2016). Two-stage cascade model for unconstrained face detection. In *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, pages 1–4.

Micheal, A. A. and Geetha, P. (2017). Multi-view face detection using normalized pixel difference feature. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pages 0988–0992.

Nanni, L., Brahnam, S., and Lumini, A. (2019). Face detection ensemble with methods using depth information to filter false positives. *Sensors*, 19(23).

Nguyen, D.-L., Nguyen, V.-T., Tran, M.-T., and Yoshitaka, A. (2015a). Adaptive WildNet Face network for detecting face in the wild. In Verikas, A., Radeva, P., and Nikolaev, D., editors, *Eighth International Conference on Machine Vision (ICMV 2015)*, volume 9875, pages 152 – 156. International Society for Optics and Photonics, SPIE.

Nguyen, D.-L., Nguyen, V.-T., Tran, M.-T., and Yoshitaka, A. (2015b). Boosting speed and accuracy in deformable part models for face image in the wild. In *2015 International Conference on Advanced Computing and Applications (ACOMP)*, pages 134–141.

Ravidas, S. (2019). Deep learning for pose-invariant face detection in unconstrained environment. *International Journal of Electrical and Computer Engineering (IJECE)*, 9:577.

Rowley, H., Baluja, S., and Kanade, T. (1998). Rotation invariant neural network-based face detection. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*, pages 38–44.

Sawat, D. D., Hegadi, R. S., Garg, L., and Hegadi, R. S. (2020). Pixel encoding for unconstrained face detection. *Multimedia Tools Appl.*, 79(47–48):35033–35054.

Shi, X., Shan, S., Kan, M., Wu, S., and Chen, X. (2018). Real-time rotation-invariant face detection with progressive calibration networks.

Shu, H., Chen, D., Li, Y., and Wang, S. (2017). A highly accurate facial region network for unconstrained face detection. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 665–669.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

Triantafyllidou, D., Nousi, P., and Tefas, A. (2018). Fast deep convolutional face detection in the wild exploiting hard sample mining. *Big Data Research*, 11:65–76. Selected papers from the 2nd INNS Conference on Big Data: Big Data Neural Networks.

`http://mplab.ucsd.edu`. The MPLab GENKI Database.

Venkatesan, R. and Li, B. (2018). *Convolutional Neural Networks in Visual Computing: A Concise Guide*. Convolutional Neural Networks in Visual Computing: A Concise Guide. CRC Press.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I.

Wang, S., Xu, T., Li, W., and Sun, H. (2019). Cssd: Cascade single shot face detector. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Yan, J., Zhang, X., Lei, Z., and Li, S. Z. (2013). Real-time high performance deformable model for face detection in the wild. In *2013 International Conference on Biometrics (ICB)*, pages 1–6.

Yan, J., Zhang, X., Lei, Z., and Li, S. Z. (2014). Face detection by structural models. *Image and Vision Computing*, 32(10):790–799. Best of Automatic Face and Gesture Recognition 2013.

Yang, S., Luo, P., Loy, C. C., and Tang, X. (2016). Wider face: A face detection benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yang, S., Wiliem, A., and Lovell, B. C. (2018). It takes two to tango: Cascading off-the-shelf face detectors. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 648–6488.

Zafeiriou, S., Zhang, C., and Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, 138:1–24.

Zakaria, Z., Suandi, S. A., and Mohamad-Saleh, J. (2018). Hierarchical skin-adaboost-neural network (h-skann) for multi-face detection. *Applied Soft Computing*, 68:172–190.

Zeng, D., Liu, H., Zhao, F., Ge, S., Shen, W., and Zhang, Z. (2019). Proposal pyramid networks for fast face detection. *Information Sciences*, 495:136–149.

Zhang, J., Wu, X., Hoi, S. C., and Zhu, J. (2020). Feature agglomeration networks for single stage face detection. *Neurocomputing*, 380:180–189.

Zhang, K., Guo, X., He, Y., Wang, X., Guo, Y., and Ding, Q. (2019a). IMS-SSH: multi-scale face detection method in unconstrained settings. *Journal of Electronic Imaging*, 28(1):1 – 10.

Zhang, K., Zhang, Z., Wang, H., Li, Z., Qiao, Y., and Liu, W. (2017a). Detecting faces using inside cascaded contextual cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3190–3198.

Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., and Li, S. Z. (2017b). Faceboxes: A cpu real-time face detector with high accuracy.

Zhang, Y., Ding, M., Bai, Y., and Ghanem, B. (2019b). Detecting small faces in the wild based on generative adversarial network and contextual information. *Pattern Recognition*, 94:74–86.

Zheng, Y., Zhu, C., Luu, K., Bhagavatula, C., Le, T. H. N., and Savvides, M. (2016). Towards a deep learning framework for unconstrained face detection. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8.

Zhou, L., Zhao, H., and Leng, J. (2022). Mtcnet: Multi-task collaboration network for rotation-invariance face detection. *Pattern Recognition*, 124:108425.

Zhou, Z., He, Z., Jia, Y., Du, J., Wang, L., and Chen, Z. (2020). Context prior-based with residual learning for face detection: A deep convolutional encoder–decoder network. *Signal Processing: Image Communication*, 88:115948.

Zhu, C., Zheng, Y., Luu, K., and Savvides, M. (2018). Enhancing interior and exterior deep facial features for face detection in the wild. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 226–233.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2879–2886.

6. Appendix

The quality analysis is presented in Table 2:

Table 2. Analysis of primary study quality.

Reference	C1	C2	C3	C4	C5	C6	C7	Total
[Zeng et al. 2019] [Zhou et al. 2020]	1	1	1	1	1	1	1	7
[Liao et al. 2016]	1	0.5	1	1	1	1	1	6.5
[Chen et al. 2018] [Shi et al. 2018] [Zhang et al. 2017a] [Zhou et al. 2022] [Zhang et al. 2020] [Zhang et al. 2017b] [Zhu et al. 2018] [Zhang et al. 2019b] [Liu and Levine 2017] [Li et al. 2019] [Zhang et al. 2019a] [Triantafyllidou et al. 2018] [Micheal and Geetha 2017]	1	1	1	0	1	1	1	6
[Li et al. 2020] [Jiang et al. 2018] [Sawat et al. 2020] [Wang et al. 2019] [Zheng et al. 2016] [Alafif et al. 2017] [Deng and Xie 2017b] [Zakaria et al. 2018]	1	1	0.5	0	1	1	1	5.5
[Nanni et al. 2019]	1	1	0.5	1	0	1	1	5.5
[Deng and Xie 2017a]	1	1	1	0	0.5	1	1	5.5
[Bai and Ghanem 2017]	0.5	1	1	0	1	1	1	5.5
[Yan et al. 2013]	1	0.5	1	0	1	1	1	5.5
[Lin et al. 2016]	1	1	1	0	1	0	1	5
[Ge et al. 2017] [El-Barkouky et al. 2014]	1	1	0.5	0	0.5	1	1	5
[Yan et al. 2014]	1	0.5	0.5	0	1	1	1	5
[Nguyen et al. 2015b]	0.5	1	1	0	0.5	1	1	5
[Li et al. 2017a]	1	1	0.5	0	1	0	1	4.5
[Yang et al. 2018]	1	0.5	1	0	0	1	1	4.5
[Shu et al. 2017]	0.5	1	0.5	0	0.5	1	1	4.5
[Lv et al. 2016]	0.5	1	0.5	0	0.5	1	1	4.5
[Chai et al. 2014]	1	0.5	0.5	0	0.5	1	1	4.5
[Marčetić et al. 2016]	1	0.5	1	0	0.5	1	0	4
[Li et al. 2015]	1	1	0.5	0	0.5	1	0	4
[Nguyen et al. 2015a]	0.5	1	0.5	0	0	1	1	4
[Ravidas 2019] [Gul and Farooq 2015]	1	0.5	0.5	0	0.5	1	0	3.5
[Magalhaes et al. 2012]	0.5	0.5	0.5	0	1	1	0	3.5
[Li et al. 2017b]	1	0.5	0.5	0	0.5	0	1	3.5