Engineering Science and Technology, an International Journal 61 (2025) 101893

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Human face localization and detection in highly occluded unconstrained environments

Abdulaziz Alashbi [a,*], Abdul Hakim H.M. Mohamed [a,*], Ayman A. El-Saleh [b], Ibraheem Shayea [c], Mohd Shahrizal Sunar [d], Zieb Rabie Alqahtani [d], Faisal Saeed [e], Bilal Saoud [f,g,c]

[a] Information Systems and Business Analytics Department, A'Sharqiyah University, (ASU), Ibra 400, Oman
[b] Department of Electronics and Communication Engineering, College of Engineering, A'Sharqiyah University (ASU), Ibra 400, Oman
[c] Electronics & Communications Engineering Department, Faculty of Electrical and Electronics Engineering, Istanbul Technical University (ITU), 34469, Istanbul, Turkey
[d] Media and Game Innovation Centre of Excellence (MaGICX), Institute of Human Centered Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia
[e] College of Computing and Digital Technology, Birmingham City University, B4 7XG, Birmingham, UK
[f] Electrical Engineering Department, Faculty of Sciences and Applied Sciences, University of Bouira, 10000, Bouira, Algeria
[g] LISEA Laboratory, Faculty of Sciences and Applied Sciences, University of Bouira, 10000, Bouira, Algeria

ARTICLE INFO

Keywords:
Occluded face detection
Facial landmark detection
Deep learning
Artificial intelligence
Computer vision

ABSTRACT

Significant advancements have been achieved in the field of computer vision pertaining to the detection of human faces. This technological development holds great potential for a wide range of applications including but not limited to identification, surveillance and expression recognition. Unconstrained face identification has been significantly improved by the advancements in Deep Learning algorithms (DL). However, the presence of severe occlusion is an ongoing obstacle particularly when it obstructs a substantial section of the facial area, resulting in the absence of crucial facial characteristics. Furthermore, the limited availability of comprehensive datasets containing substantially obscured faces exacerbates the problem, impeding the efficacy of face detection programs. This study presents a new methodology, which incorporates an advanced occluded face detection (OFD) model, in order to enhance feature extraction and detection network. A dataset was developed specifically for training and testing the model. The new dataset includes faces with significant occlusion. The utilization of contextual-based annotation approaches improves the depiction of crucial facial characteristics. The OFD model exhibits exceptional performance and attaining a notable accuracy rate of 57.84%, a precision rate of 73.70% and a recall rate of 42.63%. These results surpass those achieved by alternative methods such as YOLO-v3 and Mobilenet-SSD. This study shows the capacity to make substantial progress in detecting occluded faces, hence offering the ability to make a positive influence on the domains of identification, surveillance and expression recognition.

1. Introduction

The primary and fundamental phase of any automated system for face processing and facial analysis involves the detection of faces [1,2]. As a result, scholars are motivated to investigate approaches that can improve the accuracy of these systems. The overall effectiveness and precision of face-related applications, such as security monitoring, facial expression recognition, face identification, and human–computer interaction, are significantly influenced by the accuracy of the face detection algorithm [3]. Nevertheless, the phenomenon of occlusion, defined as the partial or complete blocking of facial features by objects or clothing, is a considerable obstacle in the field of face detection. Occlusion frequently arises in real-world situations as a result of diverse causes, including the presence of face masks, facial hair, sunglasses, or religious clothes such as the niqab [4]. The existence of occlusion poses are a significant obstacle to the precise detection of faces, resulting in a decline in the performance of face detection algorithms. The presence of occlusion hinders the visibility of important facial characteristics, hence increasing the difficulty of extracting distinguishing information

* Corresponding authors.
E-mail addresses: abdulaziez.hm@gmail.com (A. Alashbi), abdulhakim.mohamed@asu.edu.om (A.H.H.M. Mohamed), ayman.elsaleh@asu.edu.om (A.A. El-Saleh), shayea@itu.edu.tr (I. Shayea), shahrizal@utm.my (M.S. Sunar), zralqahtani@graduate.utm.my (Z.R. Alqahtani), Faisal.Saeed@bcu.ac.uk (F. Saeed), bilal340@gmail.com (B. Saoud).

https://doi.org/10.1016/j.jestch.2024.101893
Received 1 December 2023; Received in revised form 30 September 2024; Accepted 1 November 2024
Available online 29 November 2024
2215-0986/© 2024 The Authors. Published by Elsevier B.V. on behalf of Karabuk University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

and accurately distinguishing faces from the surrounding environment. Hence, the development of robust face identification models capable of effectively handling obstructed settings is of utmost significance.

Recent advances in face detection have achieved notable success in controlled environments with minimal occlusion. However, the unresolved matter of developing a robust face identification algorithm lies in its ability to accurately detect faces despite arbitrary changes in position and occlusion. One of the main challenges in designing face detection systems is the limited performance observed when detecting faces that are obstructed. Existing facial detection methods exhibit a substantial discrepancy in accuracy compared to the expected effectiveness in scenarios involving significant occlusion [5]. This gap underscores the need for more robust solutions, particularly in environments where occlusion is prevalent.

Several approaches have been proposed to address the challenge of partially covered facial photographs, as de-tailed in prior research [6–8]. However, the persistent issue remains in effectively addressing facial images that are heavily occluded, when a substantial section of the face is obscured. There are unsolved inquiries that necessitate deeper research within the realm of occlusion. Although contemporary face detectors are at the forefront of technological breakthroughs, they have difficulties when faced with faces that are severely obscured [9]. The resolution of this challenge has not been fully achieved. The detection rate of face detection is found to be inversely proportional to the degree of occlusion, as higher levels of occlusion result in a decrease in performance.

In recent years, the utilization of Deep Learning (DL) has played a crucial role in advancing the field of object detection and face detection. The primary factor contributing to this phenomenon can be attributed to the implementation of DL and Convolutional Neural Network (CNN) methodologies. These approaches have demonstrated exceptional efficacy in extracting relevant features and effectively employing learning techniques. The aforementioned technological improvements have resulted in the development of exceptionally accurate models for the purpose of face detection [10]. The categorization of face detection methods utilizing DL-CNN may be generally divided into two primary categories. The first category includes multi-stage approaches, exemplified as Faster-RCNN [11]. The second category consists of single-stage approaches, such as YOLO [12,13]. The multi-stage approach encompasses a sequence of procedures, including region proposal, feature extraction and classification. In contrast, the single-stage technique executes all of these stages within a singular iteration. Although both systems possess their own set of pros and limitations. The single-stage strategy has garnered more attention in recent years owing to its straightforwardness and effectiveness. In this methodology, the neural network makes direct predictions of both the bounding boxes and class labels, eliminating the requirement for supplementary procedures. This enhancement results in improved speed and increased suitability for real-time applications. The YOLO (You Only Look Once) model is widely recognized in the field of computer vision for its effectiveness in face identification. It belongs to the single-stage approach, which is characterized by its ability to efficiently execute regression-problem tasks. The system is notable not just for its ability to provide real-time performance indicators, but also for its sophisticated architecture that incorporates a variety of novel and enhanced components. The aforementioned features encompass residual skip connections, up sampling, and the capability to execute detection at three distinct scales. Anchor boxes are established for detecting objects of varying sizes on each scale. Every anchor is comprised of a bounding box, which is defined by its coordinates, objectness score, and class scores.

Although numerous models for face identification have made notable progress in detecting faces in unconstrained environments, they generally struggle when faced with instances where the face is partially covered. Previous studies have shown that occlusion poses are a challenge for existing models, such Tiny Face [14], Yolo-Face [15], and Ultra-Light [16], resulting in decreased performance. The models predominantly depend on visual signals and contextual information, which can become less trustworthy when important facial regions are obscured. As a result, the accuracy of detection diminishes, leading to a rise in both false positives and false negatives, so impeding the overall performance. Hence, it is imperative to develop innovative methodologies that explicitly tackle the difficulties presented by occlusion.

The drop in performance and poor outcomes observed in current face detectors when encountering substantial obstruction can be attributed to various probable sources. For instance, the absence of an annotated dataset of strongly obscured faces contributes to the widening of the gap. Furthermore, the available datasets lack a significant quantity of images that depict faces with varying degrees of obstruction, which is essential for training occluded face detection models. Moreover, the existence of a scarcity of unique attributes in substantially obscured facial images introduces intricacy to the feature extraction network and constrains its ability to acquire an adequate amount of distinguishing characteristics from the training data during the learning phase. This study introduces the Occlusion-Aware Face Detector (OFD) as a solution to the constraints faced by current face identification algorithms in contexts with occlusions. The OFD model has been specifically developed to address the difficulties presented by occlusion, allowing for precise and resilient face detection, even in situations with significant occlusion. The performance of the OFD model was enhanced through the integration of innovative methodologies, such as the utilization of the Generalized Intersection Over Union (GIoU) metric as a regression loss function. This improvement resulted in more precise predictions of bounding boxes. OFD has demonstrated its ability to enhance accuracy even in scenarios where there is either no overlap or only partial overlap. The primary objective of the proposed model is to revolutionize face detection in situations when occlusion is present. This is achieved by providing a solution that transcends the constraints of current methodologies and substantially improves the accuracy of detection.

The main objective of this research is to develop an effective and robust face detection model that can accurately detect faces in highly occluded scenarios, with a specific focus on niqab-occluded faces. These cases involve faces that are partially or fully covered by objects or clothing; which is a critical challenge in unconstrained environments, making detection more challenging. The proposed model aims to overcome the limitations of existing face detection algorithms by addressing the challenges posed by niqab occlusion, improving detection accuracy and reducing false positives and false negatives. While this study concentrates on niqab occlusion, we acknowledge the existence of various other types of facial occlusion, and future research could explore these areas to evaluate the model's robustness across different occlusion types.

The key contributions of this work are as follows:

- Development of a novel face detection model, termed the Occlusion-Aware Face Detector (OFD), specifically designed to handle scenarios with occluded faces.
- Proposal of a Generalized Intersection Over Union (GIoU) metric for regression loss function, which significantly improves the accuracy of bounding box predictions, even in cases of no or partial overlap.
- Construction and utilization of the Niqab-Face dataset, a comprehensive dataset containing images of highly occluded faces, for training and assessing the effectiveness of the proposed OFD model.
- Comparative evaluation of the OFD model against state-of-the-art face detection models, including MTCNN, Mobilenet-SSD, Tiny-Face, Ultra-Light and Yolo-Face, demonstrating its superior performance in terms of accuracy, precision, recall and F-measure.

Fig. 1. Sample of extensively covered faces in various niqab fashion variants.

The remaining sections of this paper are organized as follows: Section 2 presents background and shows the related work. Section 3 illustrates methodology. Section 4 presents Occluded Face Detection Construction. Section 5 illustrates results and comparison. Analysis and discussion are presented in Section 6. Finally, Section 7 gives a conclusion.

2. Related work

2.1. Occluded face detection

The presence of occluded faces, which refers to faces that are partially or fully obscured, can arise due to a multi-tude of factors. For example, the use of medical masks is mandated in some contexts, such as healthcare facilities or during periods of increased vigilance over the COVID-19 outbreak. Furthermore, the phenomenon of occluded faces can be observed in some Muslim countries, where women choose to wear the niqab. The niqab is a form of facial covering utilized as a religious custom to obscure individuals' faces while in public or in the company of others who are not related to them [17–19]. The illustration presented in 1 showcases instances of prominently concealed facial features of Muslim women who observe the tradition of wearing the niqab. The textile garment commonly referred to as the niqab is also recognized by alternative names such as the burqa or khimar. The niqab effectively conceals the wearer's full face, rendering it nearly imperceptible and entirely obstructed from visual observation. Consequently, their facial features are significantly obscured, leading to pronounced occlusion.

A person's face holds significant importance and serves as a rich source of information regarding an individual's race, sex, identity, age, emotions, and more. It acts as a crucial entry point for various face processing applications, such as face recognition, face verification, face tracking, and facial expression detection [20,21]. The aforementioned technology serves as the fundamental component for the creation of advanced systems designed for consumer-oriented items such as digital cameras, mobile phones, and other related entities [22]. The development of powerful feature extraction techniques, such as HOG [23], Local Binary Patterns (LBP) [24], and Integral Channels [25], has made a substantial contribution to enhancing the precision of face detection and has be-come popular in real-life applications [26]. The basic and simple handcrafted features of Haar-cascade enabled this framework to perform well on frontal faces under stable conditions, where illumination and lighting were mini-mal variants. However, for non-frontal faces and in unconstrained conditions, such as highly occluded faces in which images were taken under arbitrary conditions, accuracy decreases, and the false-positive rate increases dramatically [27].

2.2. Deep learning approaches

The emerging paradigm of computer vision employs DL and CNN-based approaches for image recognition, classification, and object localization due to their capability to mimic and automatically extract features without the need for handcrafted engineering or manual selection of appropriate features, as was done in traditional ML approaches [28]. The golden age of DL began in 2012 when a CNN-based architecture called AlexNet achieved unprecedented success in the ImageNet competition [29]. The current paradigm of computer vision is empowered by the utilization of DL and CNN for image classification and object detection due to their ability to learn and extract patterns from training examples and generalize to similar, unseen [30].

DL and CNN approaches for object detection can be grouped into two main categories: multi-stage and single-stage approaches. Multi-stage detection, like Region-Proposal-based Networks (RPN), generates proposals and redirects the output to a second stage for classification. The RPN architecture comprises two different networks: the first generates Region of Interest (ROI) proposals around 2000 for an image, and the second classifies the proposed ROI. Most region-based approaches are variations of those introduced in [31,32]. The second category is the single-stage approach, where the coordinates of bounding boxes and face object scores are regressed in one single pass, making it faster than RPN. The models You Only Look Once (YOLO) [18] and Single Shot MultiBox Detector (SSD) [33] are the most common face detection models belonging to this category. RPN and CNN were introduced by Xia and Zhang [34] as a successful implementation of DL CNN for detecting occluded faces. An effective face detector proposed in [35] utilized anchor-level attention focusing on features in face regions for detecting occluded faces with masks and sunglasses.

2.3. Applications of face detection

Face detection, as a part of facial recognition, is integrated into artificial intelligence (AI) technologies applied in various fields, including security, law enforcement, biometrics, health, safety, banking, and retail [36]. The market value of facial recognition was USD 3.72 billion in 2020 and is estimated to reach USD 11.62 billion by 2026 [37]. The COVID-19 pandemic has accelerated the development of these emerging technologies. Research by the National Institute of Standards and Technology (NIST) indicates that since the pandemic began, several facial recognition algorithms have improved rapidly in detecting occluded and masked faces, with an error rate reduced to ten times less than before the pandemic [38]. This derived technology leaders' companies to allocate a competitive position in this field. They have a great impact to the improvement of face detection and facial recognition as a part of their AI system. Giant companies such as Google for example was behind the development of single shot face detection model SSD [39].

Facial recognition project (Facenet) introduced by google and achieved state-of-the-art results on LFW bench-mark dataset [40]. This technology is embedded in Google Photos apps and it is used to automatically categorize and classify photos based on people's faces, which is highly important in the field of biometrics. Facebook also announced the DeepFace program, which was able to determine with 97.25% of accuracy whether the two photographed faces belong to the same person [41].

2.4. Current directions

Recent advancements in face detection technology focus on addressing ongoing challenges and vulnerabilities. For instance, in [42], researchers developed an optimal combinatorial detector designed to effectively tackle issues related to large-scale variations, occluded faces, and imbalanced samples caused by small faces. Additionally, [43] highlighted the vulnerability of modern face recognition systems to

backdoor attacks, underscoring the necessity for further investigation to address these threats and enhance the security of face recognition technology. Furthermore, [44] proposed a solution aimed at improving face recognition in the presence of occlusions by integrating occlusion detection and reconstruction techniques, which enhances the accuracy and reliability of recognition systems. Lastly, [45] introduced a single-stage face swapping model that achieves competitive performance by incorporating an adaptive Feature Fusion Attention and Interpreted Feature Similarity Regularization, allowing for the adaptive fusion of attribute features and features conditioned on identity information.

3. Methodology

3.1. Defining degree of occlusion

This research aims to develop an improved face detection model that can efficiently recognize faces with significant occlusion, such as those obscured by niqabs. The extent of facial blockage in individuals wearing niqab varies between 50% and 90% and more where most of the features are hidden. Researchers attempted to establish a definition for occluded faces based on the extent of occlusion. For instance, in the work of Yang and Luo [9], they divided faces into three classifications: faces with no occlusion, partially obstructed faces and heavily obstructed faces. Partial occlusion referred to cases where 1% to 30% of the face area was covered, while heavy occlusion was defined as situations where more than 30% of the face area was obscured or blocked. In another study [46], the face was segmented into four primary regions, as depicted in Fig. 2, namely the chin, mouth, nose and eyes. They measured the level of occlusion by quantifying the number of occluded regions. Weak occlusion was defined as one to two occluded regions, medium occlusion was characterized by three occluded regions and heavy occlusion was identified when all four regions were obstructed. However, there is a concern that the classification may be too loose when categorizing extensively obstructed faces as those with over 30% coverage. Describing highly occluded faces in this manner may not adequately capture the full spectrum of heavily and fully occluded faces. To address this issue, following the method outlined in [46], we propose extending the categorization of facial parts into five equal segments, including the forehead, both eyes, the nose, the mouth and the chin. This allows for a more precise differentiation between heavily occluded and fully occluded faces based on the number of occluded areas.

Fig. 2 illustrates the level of occlusion. The four facial regions defined by [46] are illustrated in Fig. 2(a). In Fig. 2(b), an extended five-region model is presented. In Fig. 2(c), we overlay the extended five regions with the four regions from [46] to emphasize and highlight the differences between the two definitions. In addition, an example of a heavily occluded face is shown in Fig. 2(c), with four occluded regions based on [46]. However, when overlaid with the extended five regions, the degree of occlusion is approximately 50%. This suggests that utilizing of the extended five face regions enables a clearer measurement of occlusion.

3.2. Dataset construction and preprocessing

In order to train DL CNN models for face detection, a significant dataset containing a wide range of samples is necessary. Nevertheless, current face identification datasets are deficient in terms of having an adequate number of substantially occluded face samples. In order to fill this void, we generated a Niqab-Face collection by extracting photos from diverse web sources. The gathered dataset underwent meticulous cleansing to exclude extraneous and substandard photos. This dataset serves as the basis for training and assessing our OFD model. The collection had at least 12 000 photos showing faces with a significant amount of occlusion. Social media platforms including Pinterest, Facebook, Instagram, and YouTube were the specific sources for photos. Nevertheless, acquiring a large number of photographs from many online sources simultaneously proved to be a challenging endeavor.

The search and download process was managed using the image scraping technique. Image scraping is a method employed to search for a substantial amount of data/images on indexed webpages, get and save the pertinent images according to a systematic indexing system. Scrapy, a widely utilized open-source framework for Python, is extensively employed for the purpose of locating and retrieving images [47]. Employing keyword and picture similarity searches for web crawling and scraping yielded a substantial dataset of images, a significant portion of which were duplicates or unrelated. This set comprised photos devoid of facial features and photographs solely focused on frontal faces, amounting to around 140 000. A substantial quantity of images required a clearing process, which was carried out.

The initial stage of dataset cleaning entailed a meticulous inspection and scanning of images to detect any extraneous, substandard or distorted images. The subsequent action is eliminating superfluous and undesirable images. For example, in [48], two individuals were employed to manually inspect images and verify that each image contained a face. They also eliminated any images that fell outside the scope of their dataset. In the study [49], researchers manually eliminated undesirable images, including those lacking facial features and those with low quality and inadequate resolution. Furthermore, [49] employed manual cleaning to eliminate images without occluded faces. Hence, the process of dataset cleaning primarily entails the use of filters and the manual elimination of undesirable images that are incongruous with the dataset's intended purpose, such as low-resolution, distorted and duplicated images. Fig. 3 depicts typical sample images from the Niqab dataset. The images illustrate various occlusion levels, including niqab-occluded faces, which are central to our study.

The gathered images were refined to eliminate any image that solely featured an unobstructed, straightforward-facing face. Nonetheless, the task of manually cleaning such an extensive quantity of images was a substantial obstacle. In order to tackle this issue, we utilized the haar-cascade face identification algorithm [50] to automatically eliminate straightforward frontal faces from the dataset.

3.3. Annotation and labeling

Manual dataset annotation and labeling are critical tasks. Thye involve a substantial amount of manual and repetitive work. The performance of any DL face detection model heavily depends on the accuracy of the training dataset. In addition, precise labeling is a crucial factor for maintaining dataset integrity. Various annotation methods and techniques exist. For instance we can find bounding box, polygon annotation, cuboid annotation and semantic segmentation [51]. However, bounding box annotation is the most commonly used method in practical and industrial applications [52]. It has proven to be very useful for object detection labeling and it is easy to implement. State-of-the-art face detectors rely on accurate bounding box annotations provided in famous and standard face detection datasets such as wider face [46] and malaf [53]. A bounding box is a rectangular shape manually drawn on the target face in an image using certain annotation tools. The annotator draws a rectangle on the targeted object from the upper-left corner to the lower-right corner, determined by (x, y) coordinates. It can be represented by two points (x_1, y_1) and (x_2, y_2) or by one coordinate point (x_1, y_1) and width and height (w, h).

3.4. Contextual-based labeling

The contextual information has been given less attention in the existing face detection models [54]. However, contextual information such as head, head's pose and shoulders can play an important role in detecting difficult faces such as on heavily occluded faces, where faces are mostly covered. Highly occluded faces are difficult to be

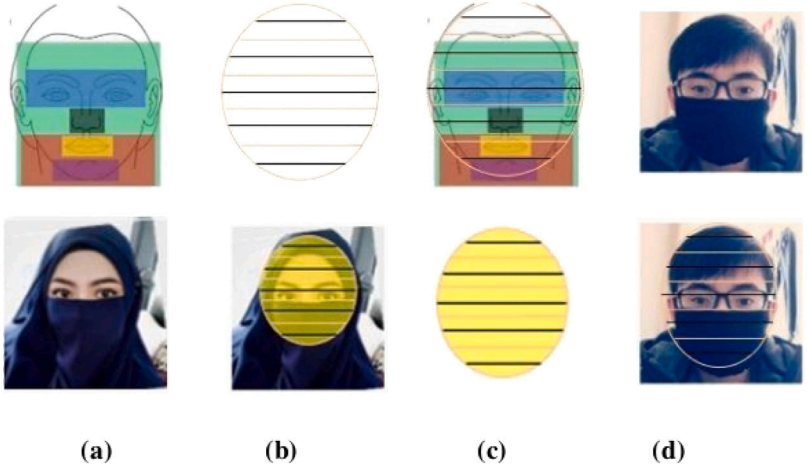

Fig. 2. Faces heavily obscured in varying degrees of occlusion.

Fig. 3. Sample images from the Niqab dataset showing high levels of occlusion.

detected due to the lack of visual information, while the larger regions within their context can provide more information about the face and its context which could be used as features representation for better detection. As [55] has highlighted, contextual information plays a critical role in addressing the occlusion problem in face detection.

Contextual information includes features that surround the occluded face, such as the head pose, shoulders and some background portions near the face region. Therefore, it extends beyond information within the face alone, encompassing body-related details. For instance, faces are often accompanied by the human body and even when faces are occluded. They can still be located by considering the entire body [56]. This contextual information becomes crucial in detecting challenging faces, particularly those heavily occluded.

The concept of using contextual information involves providing additional information around the face and its context. This allows feature extraction networks in DL CNN models to learn not only from facial features but also from contextual features [57–59]. Fig. 4 illustrates an example of an occluded face extracted in two different ways: Fig. 4(a) with no contextual information and Fig. 4(b) with contextual information about the head's pose and shoulders.

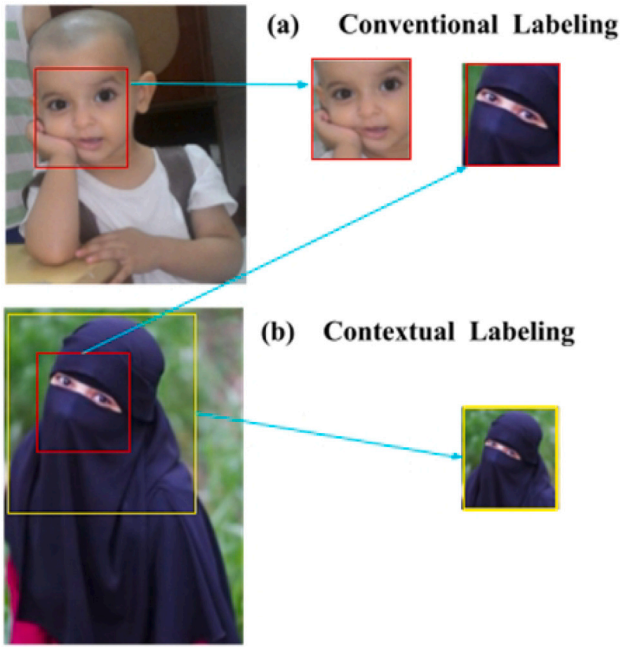

Fig. 4. An occluded face: (a) no contextual information; and (b) with contextual information.

4. Occluded face detection construction

Occluded Face Detection model (OFD) consists of two interconnected networks, which are feature extraction and detection network. Darknet-53 network is used as the backbone of feature extraction of the proposed OFD model. It was introduced in YOLO-v3 object detection

by [60]. It comprised of two major components: Feature extraction network and detection network. Feature extraction network is fed with images as inputs and obtains feature embeddings at three different scales. Then, the obtained features are fed into three branches of the detection network to get bounding boxes and class information.

4.1. Features extraction network

The main distinction between ML and CNN is that while features are manually handcrafted and designed in ML, in CNN features extraction is generated automatically and combined with classifier [61]. Darknet-53 consists of 53 layers that use the residual network as shortcut connections. The network uses (3×3) and (1×1) convolutional layers each layer is followed by batch normalization layers and Leaky ReLU activation layer. Regardless of the great performance of darknet-53 compared to dark-net-19, the overall performance of face detection is negatively affected in conditions with high levels of occlusion [62]. The challenge of high occluded faces is due to two main issues, the first issue is the limitation of learned representative features due to the obstruction which masks most of the salient face features [34]. The second issue is related to the network structure that can play a major role in feature extraction [63]. Representative features in occluded faces are inherently limited, as discussed previously and are further reduced on the feature map through multiple dimension reductions. Therefore, subsequent layers may struggle to capture sufficient information, leading to less representative features and consequently a degradation in the effectiveness of classification.

4.2. Improved feature extraction network

Feature extraction holds great significance in deep models [64,65]. Effective feature extraction is not merely a supplementary process but a foundational aspect of deep learning models that can make or break the model's performance [66]. Our methodology builds upon these insights. The feature map on darknet-53 is reduced due to several dimension, which are remarked as an advantage characteristic of the darknet-53 for object detection. Reduction leads to reduce the computation processing, which was a drawback of darknet-19. However, the feature map of occluded faces already suffers from limited representative features. As a result, several dimension reductions diminish the extraction network's ability to gather these features. Making it unsuitable for single-class object detection, such as in occluded faces [63]. To address this, enabling the feature extraction network on darknet-53 to extract available features before the reduction of the feature map can aid in obtaining more representative features. Thus, in this work, the architecture of the initial Darknet-53 network is enhanced by adding more layers to the first two residual networks, aiming to attain further facial representative features. Fig. 5 shows the two network structures. Darknet-53 is represented in Fig. 5(a) and the improved network structure is illustrated in Fig. 5(b).

4.3. Improved detection network

The prediction of YOLO is regressed as a vector which is defined as:

$$S \times S \times (B \times 5 + C) \tag{1}$$

where $S \times S$ is the grid cell i of the input image, B representing the number of bounding-boxes prediction for each grid cell i (i.e., x_{ij}, y_{ij}, h_{ij} and w_{ij} and the confidence score c), C is the abjectness class.

Therefore, the loss function is categorized into three sections: coordinate loss, confidence loss and object-class loss. Coordinate loss involves the coordinate points (x, y) and the width and height (w, h) loss of the predicted bounding box. Confidence loss indicates how certain the algorithm is that the box contains an object. The loss associated with the object-class score denotes the predicted object in multi-class object detection. However, in the case of face detection, there is only one object (a face). YOLO-v3 uses Intersection over Union (IoU) as a measurement distance for confidence loss to evaluate how closely the predicted bounding box aligns with the ground truth (GT). IoU is the most commonly used evaluation metric in both object detection and face detection [67]. It calculates a scale-invariant normalized measure of two bounding boxes:

$$IoU = |A \cap B| / |A \cup B| \tag{2}$$

There are two cases for IoU:

1. when there is overlapping between GT and predicted bounding box.
   - $IoU = 1$ this indicates the best fit as the predicted and GT are almost the same.
   - $IoU < 1$ and > 0, a threshold of $IoU \geq 0.5$ indicates positive prediction.
2. If no overlapping the value of $IoU = 0$.

The problem of YOLO-v3 lays in this case when there is no overlapping i.e., (IoU set to 0). The limitation is that in the no overlapping case, the value is set to 0 regardless of how close the two bounding boxes are to the GT. This is important as the CNN loss function uses backward propagation to adjust the wights based on loss decrease. When IoU is set to 0 regardless of how close the bounding boxes from the GT, it causes CNN to ignore the instance which degrades the detection performance and increases the complexity of the training in terms of long convergence and training time. Eq. (3) below shows the two cases with three IoU states (case 1: $IoU = 0$, case 2: IoU is less than 1 and greater than 0 and case 3: $IoU = 0$).

$$IoU = \begin{cases} 1 & \text{best fit} \\ \langle 1 and \rangle 0 & \text{overlapping } (IoU \geq 0.5 \\ & \text{positive prediction)} \\ 0 & \text{no overlapping} \end{cases} \tag{3}$$

Therefore, there is a way for loss improvement by adopting Generalized Intersection Over Union ($GIoU$) proposed by [68]. $GIoU$ has a gradient in all possible cases, including non-overlapping situations that considerably improved its performance. $GIoU$ is represented in Eq. (4).

$$GIoU = IoU - \frac{|C\ (A \cup B)|}{|C|} \tag{4}$$

where C is the smallest ellipsoids rectangle of A and B. The reason for using $GIoU$ over IoU as the regression loss function for the prediction box is driven by its superior ability to precisely capture the overlap between the two rectangular boxes [68].

The overall concept is illustrated in Fig. 6. The green rectangle represents the ground truth bounding boxes and the red one is for the predicted bounding boxes.

5. Results and comparison

To obtain a precise evaluation of the performance of the proposed OFD model, we conducted a comparison with numerous cutting-edge face detection models that are recognized for their high effectiveness in different situations. We conducted our selection process by evaluating models that have consistently achieved high performance on benchmark datasets, with a specific focus on their ability to handle occlusion. The selection of models, including MTCNN, Mobilenet-SSD, TinyFace, Ultra-Light and YOLOv3 were based on their popularity, extensive usage, and the accessibility of pre-trained weights.

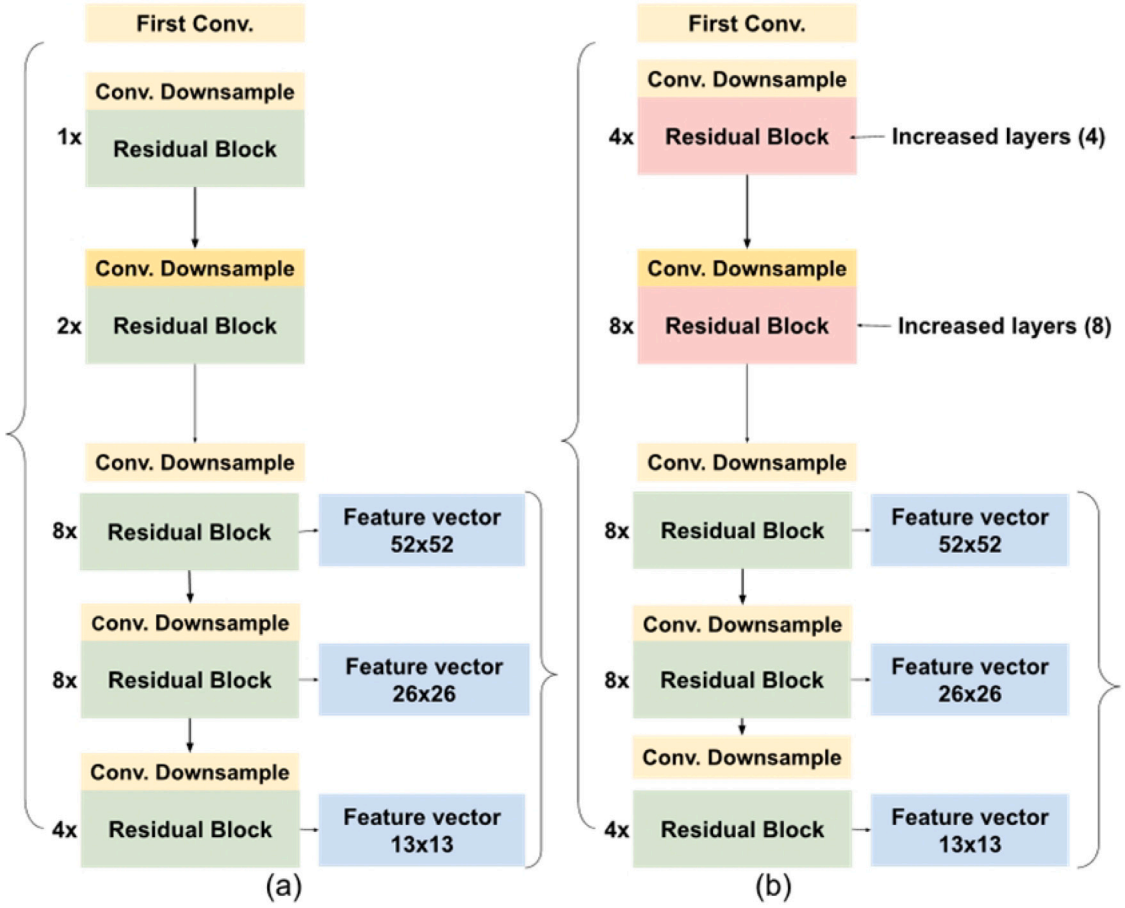

Fig. 5. Network architecture of the feature extraction network: (a) Original Darknet-53; (b) Improved structure achieved by increasing the network layers of the first two residual networks.

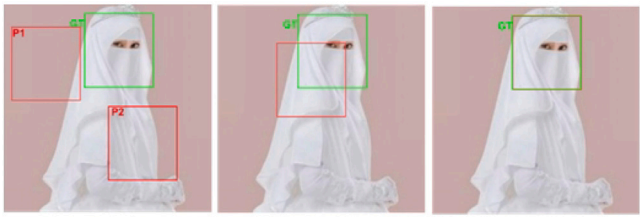

Fig. 6. The placement of IoU categorized into: (a) non-overlapping; (b) positive overlapping and (c) fit $IoU = 1$.

5.1. Training and validation

The experiments were carried out using the available hardware and software during the evaluation and testing phase. A single machine (desktop PC) with the following specifications was used for implementing and testing: Intel Core i7-6700 CPU @ 3.40 GHz with 16 GB RAM, NVIDIA GeForce 750 GPU with 512 CUDA cores and 1 GB memory. The PC was equipped with the Ubuntu 16.0 operating system. Additionally, Python with TensorFlow and other relevant libraries were utilized.

The Niqab-dataset, created to train and test the suggested model, was divided into three subcategories: training, validation, and testing. The images are divided into three sets, with each group accounting for 50%, 10% and 40% of the total. The training and validation sets were employed in the training procedure. The training setups included standard and effective techniques in DL, including transfer learning and data augmentation. These strategies enhance the efficiency of the training process and optimize the performance of the models. Data augmentation is a method that involves applying transformations to images while preserving their labeling. It is commonly used to enhance the accuracy of CNN-Based models and prevent overfitting of the model during training by applying it to the training images [69]. Four image augmentation techniques were employed: image rotation, scaling, horizontal flipping and cropping. A random selection of rotation angles ranging from 5 to 25 degrees was employed.

However, despite the benefits of transfer learning and data augmentation, these techniques may still face limitations when dealing with heavily occluded faces. Data augmentation, while effective in introducing variability, may not fully capture the complexity of severe occlusions. Similarly, transfer learning, which relies on pre-trained models, may be limited if the original dataset lacks sufficient occlusion examples.

The OFD model underwent training, validation and fine-tuning using the training set of the Niqab-Face dataset. The supplied image dimensions were set to the default size of (416×416). The given batch size is 64, indicating that 64 images must be inputted to the network for each iteration. It was then partitioned into a subdivision of four. This is because the limited memory of the GPU makes it challenging to accommodate all batches of images simultaneously. The stochastic gradient descent algorithm (SGD) was trained using the optimizer, with an initial learning rate of 0.001 and a momentum of 0.9. Following 10 000 cycles, the learning rate underwent a slow decrease. The total number of iterations needed for one epoch was 87. The model was trained over a span of 2000 epochs, divided into 10 separate experiments, with each experiment consisting of 200 epochs. An epoch refers to a single round of passing all training datasets through the network.

Table 1
Results of Precision, Recall, F-Measure and AP.

Face-detection models	Precision	Recall	F-Measure	AP
Mobilenet-SSD	57.59%	13.38%	21.71%	21.83%
TinyFace	41.65%	20.54%	27.51%	17.58%
Ultra-Light	23.10%	14.92%	18.13%	5.57%
MTCNN	52.93%	6.85%	12.13%	15.38%
YOLO-v3	68.98%	7.79%	14.%	33.71%
OFD	73.70%	42.63%	54.02%	50.34%

5.2. Experimental results

We conducted extensive experiments to evaluate the performance of the OFD model. The evaluation metrics included Precision, Recall, F-measure and Average Precision (AP). Precision quantifies the ratio of correctly detected faces to the total number of detected faces. Recall captures the proportion of true positive faces that were successfully detected and F-measure combines precision and recall into a single score and providing a balanced assessment of overall performance. AP is calculated by taking the area under the precision–recall curve. It summarizes the performance across various levels of precision and recall and provides a single scalar value that represents the overall quality of the model.

The benchmark detectors included in the evaluation are:

- Multitask Cascaded face detection CNN (MTCNN): A Python library based on [70].
- Mobilenet-SSD which is a real-time object detection model from Google optimized for mobile devices [71].
- Tiny Face, A face detection model designed specifically for detecting small faces [14].
- YOLO-v3: A real-time object detection model [60].
- Ultra-light detector: A small and fast face detection model trained on the VOC dataset [72].

The comparative evaluation, as depicted in Table 1 and Fig. 7, demonstrated that the OFD model surpassed state-of-the-art face detection methods in terms of many evaluation metrics. The precision of the OFD model is 73.70%. The first model outperformed the second-best model by an increase of 4.72%. In addition, the OFD model demonstrated outstanding recall, with a rate of 42.63%, which indicates its capacity to recognize a significant percentage of obscured faces. The F-Measure, which is a harmonic mean of precision and recall, was determined for the OFD model at a value of 54.02%, as shown in Table 1. The OFD model demonstrated exceptional AP, with a remarkable rate of 50.34%. This surpassed the nearest comparable model, YOLO-v3 by a significant margin of 16.63%. The results showcase the efficacy of our model in reliably detecting heavily occluded faces, as compared to MTCNN, TinyFace, Ultra-Light, Mobilenet-SSD and YOLO-v3. The suggested OFD model exhibits superior effectiveness compared to existing state-of-the-art algorithms.

The comparison results for correctly detected faces True Positives (TP) between the proposed OFD and the five related detector models are presented in Table 2 and Fig. 8. OFD demonstrated the highest performance among the other five detector models, accurately detecting 42% of the total ground-truth dataset used for testing. However, among the five compared detector models, TinyFace performed the best with a detection rate of only 20%. In contrast, Ultra-Light, MobileNet-SSD, YOLO-v3 and MTCNN achieved detection rates of 14%, 13%, 8% and 7% respectively.

In terms of false negatives (FN), OFD exhibited the lowest rate compared to the other models. Regarding false positives (FP), while OFD outperformed Ultra-Light and TinyFace, YOLO-v3, MTCNN and MobileNet achieved better results in this aspect, respectively. In accuracy, OFD surpasses Mobilenet-SSD, TinyFace, Ultra-Light, MTCNN and YOLO-v3 in precise face detection, particularly across varying sizes, orientations and occlusion levels. Finally, in term of accuracy, OFD model achieved an impressive 57.84% accuracy, surpassing the YOLO-v3, by 50%. Its standout capability lies in identifying both visible and partially obscured faces, making it a robust choice for addressing challenges in face detection scenarios.

While the proposed methodology has shown strong performance in controlled environments, it is crucial to as-sess its effectiveness in real-world scenarios, where challenges such as varying lighting, occlusions, and complex backgrounds and blurry or distorted races exist. To this end, we evaluated the OFD model on a diverse range of real-world scenes, including crowded settings and scenes with blare significant occlusion. The model achieved good precision and recall values. However, in certain challenging conditions, such as heavy occlusion, the false positive (FP) rate remained elevated. The processing speed was recorded at 30 FPS on an NVIDIA GTX-RTX 2080, underscoring the trade-off between accuracy and speed.

Fig. 9 illustrates the model's performance in real-world scenarios, highlighting both successful detections and instances where the model produced FPs. These examples provide a clear picture of how the model performs in practical applications and the specific challenges it faces.

To ensure a comprehensive and up-to-date evaluation, we included a recent detection method published in 2023 in our comparison. Specifically, a RetinaNet-based single-stage face detector, proposed by Mamieva et al. (2023) [73]. The comparison, shown in Table 3, highlights the performance of YOLO-3, which serves as the backbone for the OFD model achieved an AP of 33% and a detection speed of 19 FPS on our dataset. In contrast, Mamieva et al. (2023) reported a higher AP of 37, but with a slower speed of 11.1 FPS on the wider face dataset. Since the implementation code for Mamieva's model was unavailable, we relied on their published re-sults for comparison. Despite the difference in performance, the faster detection speed of the YOLOv3 model makes it more suitable for real-time applications.

Fig. 10 demonstrates how increasing levels of occlusion affect the model's detection performance, high-lighting the strengths and limitations of our approach in handling different occlusion scenario.

5.3. Ablation experiment

To evaluate the effectiveness of the proposed contextual-based labeling technique, we conducted a comparative analysis with the conventional labeling method, which is the standard technique used for face detection annota-tion. For this evaluation, we utilized a subset of 1200 randomly selected images from the Niqab-Face dataset. The images were divided into training (50%), validation (10%), and testing (40%) sets. Two distinct labeling methods were applied to the training images: (a) Traditional Labeling: (no contextual) A bounding box was drawn solely around the face. (b) Contextual-based Labeling: In addition to the face, surrounding contextual information such as parts of the head, neck, and shoulders was included within the bounding box to enrich the feature set.

These two labeling techniques were used to train two face detection models: Model-1, trained using the traditional labeling method (no contextual information). And Model-2, trained using the proposed contextual-based labeling method. Both models were tested on the same test set, and their performance was evaluated using common metrics, including TP, FP, FN, Accuracy, Precision, Recall, and F-measure. The results shown in Table 4. demonstrate a significant improvement in performance when contextual infor-mation is included. Model-2 achieved a 13.3% improvement in F-measure (74.8% vs. 63.6%) and a 13.6% increase in accuracy (59.1% vs. 45.5%). Additionally, the false positive rate was reduced by 4.4% in the context-based model, indicating fewer incorrect detections.

These results support the hypothesis that incorporating contextual information around the face can enhance the performance of occluded face detection models. This is particularly beneficial in cases of heavy occlusion, where parts of the face are obscured. By including surrounding features such as the neck and shoulders, the model is better able to extract meaningful representations, leading to improved detection accuracy and reduced errors.

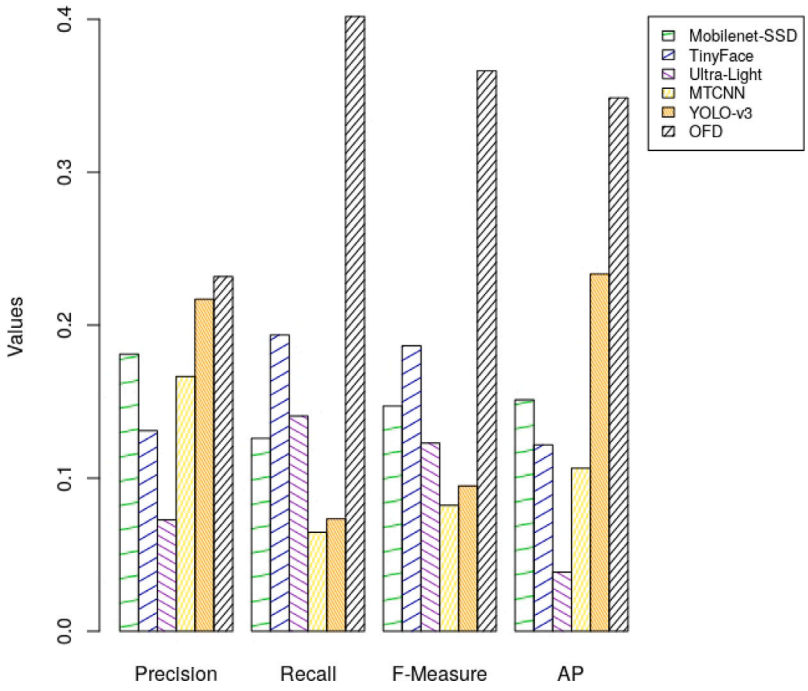

Fig. 7. Results in terms of Precision, Recall, F-Measure and AP.

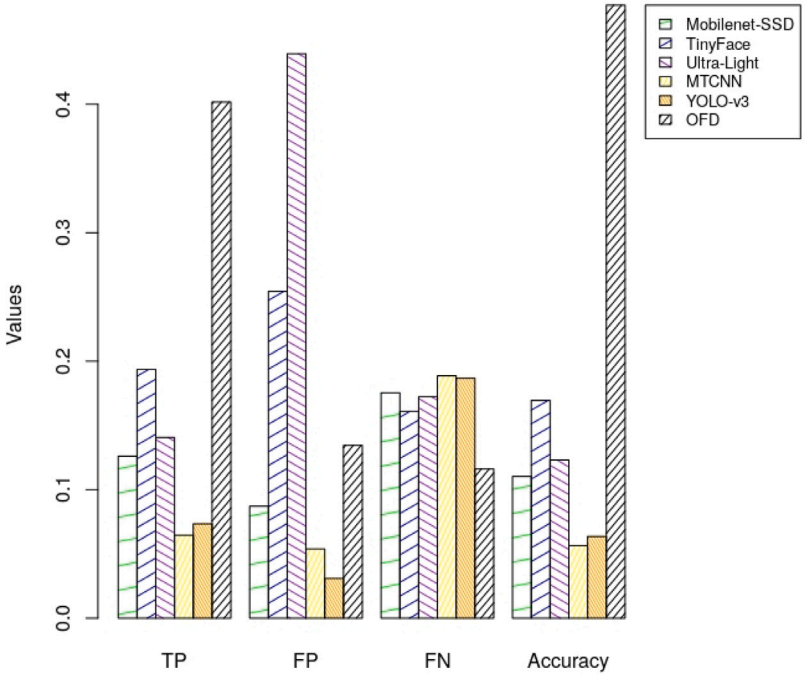

Fig. 8. Results in terms of TP, FP, FN and Accuracy.

Table 2
Comparison result of TP, FP, FN, Accuracy, FLOPs, FPS and detection time.

Face-detection models	TP	FP	FN	Accuracy	FLOPs	FPS	Detection time
Mobilenet-SSD	512	377	3314	13.38%	∽1.14 GFLOPs	30–60	30 ms
TinyFace	786	1101	3040	20.54%	∽1.2 FLOPs	40–50	40 ms
Ultra-Light	571	1901	3255	14.92%	∽109 MFLOPs	50–60	25 ms
MTCNN	262	233	3564	6.85%	∽1.4 GFLOPs	50–60	45 ms
YOLO-v3	298	134	3528	7.7%	∽140 GFLOPs	45–55	22 ms
OFD	1631	582	2195	57.84%	∽140 GFLOPs	55–60	24 ms

Fig. 9. Examples of detection result from test dataset.

Table 3
Comparison with RetinaNet model.

Model	AP	FPs	Notes
YOLOv3 (OFD Backbone)	33%	19	Tested on custom-dataset dataset.
RetinaNet [73]	37%	11.1	Results obtained from published work.

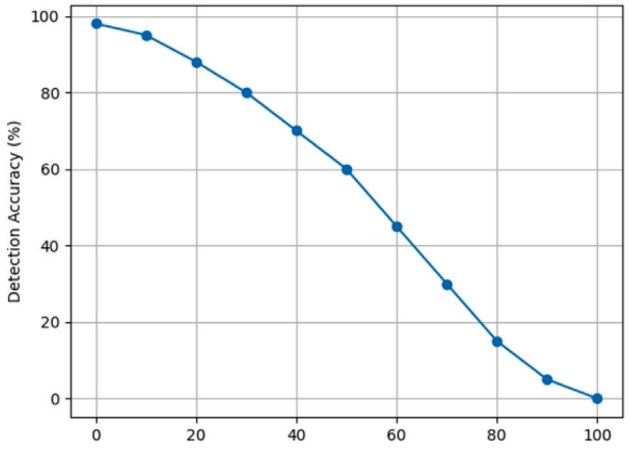

Fig. 10. Relationship curve of facial occlusion rate and detection rate.

Table 4
Comparison result of traditional labeling with contextual -base labeling.

Method	TP	FP	FN	Accuracy	Precision	Recall	F-Measure
Traditional labeling	201	10	230	45.5%	95.2%	51.5%	63.6%
Context-based labeling	261	1	179	59.1%	99.6%	59.9%	74.8%

6. Analysis and discussion

The OFD model demonstrates strong performance in accurately identifying TP when compared to other models. However, this comes at the expense of a higher rate of FP. Striking a balance between achieving a higher TP rate while maintaining a manageable FP rate is crucial. One remarkable strength of OFD lies in its capacity to successfully detect more than 40% of correctly identified faces, underscoring its effectiveness in face detection. Conversely, the model's relatively elevated FP rate indicates a tendency to identify non-faces or irrelevant patterns as faces. One contributing factor to this higher FP rate may be linked to the complexity of the dataset used for training. This dataset includes heavily occluded faces, which pose a significant challenge to detection. Occluded faces obscure a substantial portion of facial features, making their detection difficult due to reduced visibility caused by the occlusion. An in-depth analysis was conducted to explore how various adjustments to the model's parameters and training methods impacted the balance between TP and FP. We adjusted some key parameters such as detection thresh-olds and non-maximum suppression to refine the sensitivity of the model to distinguish between true facial features and non-face regions, therefore reducing FP but not sacrificing TP. Another impact was the Occlusion degree, We further analyzed the impact of different levels of occlusion on the model's performance. We identified that the FP rate was particularly high in cases where over 60% of the face was occluded. Metrics such as Precision, Recall, and F1-Score were used to assess the impact of the change, providing a comprehensive understanding of how the TP/FP trade-off could be optimized. This challenge highlights the model's ability to distinguish truly visible facial characteristics. While OFD excels in producing more TP, its increased FP rate may result in unnecessary processing of non-face regions, potentially affecting the overall system efficiency. It is important to recognize that achieving an optimal trade-off between TP and FP requires careful tuning and iterative experimentation. Addressing these challenges in the field of ML demands ongoing refinements to the model, its parameters and the input data.

The evaluation results highlight the insufficient performance of all the compared face detection models when tested on Niqab-benchmark dataset. Notably, while TinyFace achieved the highest TP result among the compared models. It accurately detected only 21% of the total number of ground-truth images. This performance is significantly lower than that of OFD, which demonstrates twice the accuracy. This examination brings to light several key assumptions that provide insights into the challenges faced by contemporary face detection models when addressing extensively obscured faces, ultimately leading to a decline in their performance.

The extent of occlusion itself, as seen in cases involving faces concealed with niqabs, poses a substantial obstacle to detection. These concealed faces obscure a significant portion of distinct facial features, making it challenging for face detectors to distinguish important features that differentiate faces from the background. Consequently, the performance of these detectors diminishes. In addition, the shortcomings are intertwined with the training datasets used for the face detection models, along with the examples employed during their training phase. The models under consideration, including TinyFace, were trained on publicly available datasets such as Widerface and FDDB. However, these datasets inadequately represent images featuring heavily occluded faces, such as those concealed with niqabs. The lack of relevant training examples for highly occluded faces inevitably results in reduced performance when detecting faces in heavily occluded scenarios. It is reasonable to anticipate less favorable detection outcomes from models lacking exposure to training examples that resemble the scenarios encountered during testing.

Furthermore, the performance of models on specific datasets may not directly generalize to other types of occluded facial images, which can limit the model's overall ability to generalize across various occlusion patterns, lighting conditions, and image quality. To address these limitations, it is essential to incorporate a more diverse set of training images that represent a wider range of occlusion scenarios. Techniques such as domain adaptation may also be explored to enhance the model's robustness and generalization capabilities.

7. Conclusion

Existing face detection models struggle to detect heavily occluded faces. This study proposes a new method that addresses this challenge by first creating a dataset of heavily occluded faces and then using a context-based annotation technique to improve feature representation. The proposed method is then used to train a DL CNN model tailored for detecting occluded faces. The results demonstrate that the proposed model outperforms state-of-the-art face detection models, such as MTCNN, Mobilenet-SSD, TinyFace, Ultra-Light and YOLO-v3, in terms of accuracy, precision, recall and F-measure. The evaluation outcomes highlight the limitations of current face detection models in excelling within scenarios involving significant occlusion, particularly those with niqab-covered faces. While these models manage to achieve reasonably good true positive rates, their performance is hindered by the intricate nature of occluded faces and the limitations inherent in their training data. Addressing these challenges necessitates the creation of refined training datasets and specialized techniques tailored to highly occluded faces, thereby enabling the development of more robust face detection models suited for such scenarios. This study highlights the potential of the proposed approach to improve the detection of occluded faces, which could benefit applications such as face identification, safety surveillance and facial expression recognition.

While this study contributes significantly to detecting heavily occluded faces, it is important to acknowledge its limitations. The dataset may not cover the full variety of occlusion types or real-world scenarios, potentially impacting generalizability. Environmental factors, such as lighting conditions and background clutter, were not extensively tested, and the model's computational complexity may hinder real-time application. Future work should focus on expanding the dataset to include a broader range of occlusion patterns, optimizing the model for improved speed and efficiency, and exploring the integration of contextual information to enhance detection accuracy. Additionally, incorporating face image super-resolution techniques as noted in [74] could improve performance by enhancing feature extraction, reducing false positives, and increasing resilience to variations in image quality. Super-resolution techniques elevate the resolution of low-quality images, making facial features clearer and enabling more accurate detection, especially in challenging conditions like low-resolution surveillance footage or poor lighting. While there are computational trade-offs, the potential gains in detection accuracy make this approach valuable. Another notable work incorporates Deep Fusion Network, which focuses on global and local facial features to provide a more comprehensive understanding of facial structure, further improving feature extraction and detection accuracy [75]. By addressing these limitations, subsequent research can build on these findings and further advance the field of face detection in challenging conditions.

CRediT authorship contribution statement

Abdulaziz Alashbi: Investigation, Software, Writing – original draft. **Abdul Hakim H.M. Mohamed:** Conceptualization, Supervision. **Ayman A. El-Saleh:** Conceptualization, Methodology, Supervision. **Ibraheem Shayea:** Methodology, Validation, Writing – review & editing. **Mohd Shahrizal Sunar:** Formal analysis, Methodology. **Zieb Rabie Alqahtani:** Formal analysis, Methodology. **Faisal Saeed:** Formal analysis, Visualization. **Bilal Saoud:** Formal analysis, Methodology, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research leading to these results has received funding from A'Sharqiyah University in the Sultanate of Oman through Research Project under grant number (BFP/RGP/ICT/22/490).

References

[1] D. Mamieva, A.B. Abdusalomov, M. Mukhiddinov, T.K. Whangbo, Improved face detection method via learning small faces on hard images based on a deep learning approach, Sensors 23 (1) (2023) 502.
[2] M. Tamilselvi, S. Karthikeyan, An ingenious face recognition system based on HRPSM_ CNN under unrestrained environmental condition, Alex. Eng. J. 61 (6) (2022) 4307–4321.
[3] Y. Kortli, M. Jridi, A. Al Falou, M. Atri, Face recognition systems: A survey, Sensors 20 (2) (2020) 342.
[4] Sang-In Bae, et al., Machine-learned light-field camera that reads facial expression from high-contrast and illumination invariant 3D facial images, Adv. Intell. Syst. 4 (4) (2022).
[5] Y. Chen, L. Song, Y. Hu, R. He, Adversarial occlusion-aware face detection, in: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS, IEEE, 2018, pp. 1–9.
[6] M. Mathias, R. Benenson, M. Pedersoli, L. Van Gool, Face detection without bells and whistles, in: European Conference on Computer Vision, Springer, 2014, pp. 720–735.
[7] T. Alafif, Z. Hailat, M. Aslan, X. Chen, On detecting partially occluded faces with pose variations, in: 2017 Third International Symposium of Creative Computing, ISPAN-FCST-ISCC, IEEE, 2017, pp. 28–37.
[8] M. Opitz, G. Waltner, G. Poier, H. Possegger, H. Bischof, Grid loss: Detecting occluded faces, in: European Conference on Computer Vision, ECCV, 2016.
[9] S. Yang, P. Luo, C. Loy, X. Tang, Faceness-net: Face detection through deep facial part responses, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 1845.
[10] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, Int. J. Comput. Vis. 128 (2020) 261–318.
[11] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. (2015) 91–99.
[12] J. Redmon, A. Farhadi, Yolo9000: Better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271.
[13] W. Dong, L. Pan, Q. Zhang, W. Zhang, Athlete target detection method in dynamic scenario based on nonlinear filtering and YOLOv5, Alex. Eng. J. 82 (2023) 208–217.
[14] P. Hu, D. Ramanan, Finding tiny faces, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2017, pp. 1522–1530.
[15] W. Chen, H. Huang, S. Peng, C. Zhou, C. Zhang, Yolo-face: A real-time face detector, Vis. Comput. (2020) 1–9.
[16] X. Liang, X. Zhao, C. Zhao, N. Jiang, M. Tang, J. Wang, Task decoupled knowledge distillation for lightweight face detectors, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2184–2192.
[17] M. Khan, Niqab: An Approach Based on the Proofs, The Institute for the Revival of the Traditional Islamic Sciences, 2016, https://bukhari2013.files.wordpress.com/2016/12/niqab-an-approach-based-on-the-proofs.pdf. Accessed on, 30.
[18] I. Zempi, 'It's a part of me, I feel naked without it': Choice, agency and identity for muslim women who wear the niqab, Ethn. Racial Stud. 39 (2016) 1738–1754.
[19] N.A. Chowdhury, H.S.A. Bakar, A.A. Elmetwally, Probing niqab wearing as an islamic identity, cultural piety and women's empowerment: A phenomenological approach, Int. J. Ethics Soc. Sci. 5 (2017) 57–76.
[20] S. Zhang, L. Wen, H. Shi, Z. Lei, S. Lyu, S.Z. Li, Single-shot scale-aware network for real-time face detection, Int. J. Comput. Vis. (2019) 1–23.
[21] M. Sajjad, F.U.M. Ullah, M. Ullah, G. Christodoulou, F.A. Cheikh, M. Hijji, . . ., J.J. Rodrigues, A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines, Alex. Eng. J. 68 (2023) 817–840.
[22] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: A brief review, Comput. Intell. Neurosci. 2018 (2018).
[23] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, 2005.
[24] A. Chehrehgosha, M. Emadi, Face detection using fusion of Lbp and Adaboost, J. Soft Comput. Appl. 2016 (2016) 1–10.
[25] X. Wang, T.X. Han, S. Yan, An Hog-Lbp human detector with partial occlusion handling, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 32–39.

[26] B. Yang, J. Yan, Z. Lei, S.Z. Li, Fine-grained evaluation on face detection in the wild, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG, IEEE, 2015, pp. 1–7.

[27] N. Wang, X. Gao, D. Tao, H. Yang, X. Li, Facial feature point detection: A comprehensive survey, Neurocomputing 275 (2018) 50–65.

[28] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V.M. Patel, C.D. Castillo, R. Chellappa, Deep learning for understanding faces: Machines may be just as good, or better, than humans, IEEE Signal Process. Mag. 35 (2018) 66–83.

[29] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.

[30] G. Guo, N. Zhang, A survey on deep learning based face recognition, Comput. Vis. Image Underst. 189 (2019) 102805.

[31] R. Ranjan, V.M. Patel, R. Chellappa, Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2017).

[32] H. Jiang, E. Learned-Miller, Face detection with the faster R-Cnn, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2017, IEEE, 2016, pp. 650–657.

[33] M. Najibi, P. Samangouei, R. Chellappa, L.S. Davis, Ssh: Single stage headless face detector, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4875–4884.

[34] Y. Xia, B. Zhang, F. Coenen, Face occlusion detection using deep convolutional neural networks, Int. J. Pattern Recognit. Artif. Intell. 30 (2016) 1660010.

[35] J. Wang, Y. Yuan, G. Yu, Face attention network: An effective face detector for the occluded faces, 2017, arXiv preprint arXiv:1711.07246.

[36] I. Azhar, M. Raza, M. Sharif, S. Kadry, S. Rho, Union is strength: Improving face sketch synthesis by fusing outcomes of fully-convolutional-networks and random sampling locality constraint, Alex. Eng. J. 61 (12) (2022) 10727–10741.

[37] I. Morder, Facial recognition market | growth, trends, and forecasts, 2020, pp. 2020—2025.

[38] M.L. Ngan, P.J. Grother, K.K. Hanaoka, Ongoing face recognition vendor test part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms, 2020.

[39] Y. Shuang, Face Detection with Mobilenet-Ssd[Online], GitHub, 2018, Available: https://github.com/bruceyang2012/Face-detection-with-mobilenet-ssd (Accessed 10 December 2019).

[40] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.

[41] Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.

[42] Z. Yu, H. Huang, W. Chen, Y. Su, Y. Liu, X. Wang, Yolo-facev2: A scale and occlusion aware face detector, Pattern Recognit. 155 (2024) 110714.

[43] Q. Le Roux, E. Bourbao, Y. Teglia, K. Kallas, A comprehensive survey on backdoor attacks and their defenses in face recognition systems, IEEE Access (2024).

[44] I. Lee, E. Lee, S.B. Yoo, Latent-OFER: detect, mask, and reconstruct with latent vectors for occluded facial expression recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1536–1546.

[45] F. Rosberg, E.E. Aksoy, F. Alonso-Fernandez, C. Englund, Face dancer: Pose-and occlusion-aware high-fidelity face swapping, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 3454–3463.

[46] S. Ge, J. Li, Q. Ye, Z. Luo, Detecting masked faces in the wild with lle-cnns, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2682–2690.

[47] Scrapy, _Scrapy_Open_Source, An open source and collaborative framework for extracting the data you need from websites, 2020.

[48] S. Yang, A. Wiliem, B.C. Lovell, To face or not to face: Towards reducing false positive of face detection, in: Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on, IEEE, 2016, pp. 1–6.

[49] S. Liao, A.K. Jain, S.Z. Li, A fast and accurate unconstrained face detector, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 211–223.

[50] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, IEEE, 2001, p. I.

[51] V. Le, J. Brandt, Z. Lin, L. Bourdev, T.S. Huang, Interactive facial feature localization, in: Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12, Springer Berlin Heidelberg, 2012, pp. 679–692.

[52] B. Adhikari, J. Peltomaki, J. Puura, H. Huttunen, Faster bounding box annotation for object detection in indoor scenes, in: 2018 7th European Workshop on Visual Information Processing, EUVIP, IEEE, 2018, pp. 1–6.

[53] S. Yang, P. Luo, C.C. Loy, X. Tang, Wider face: A face detection benchmark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5525–5533.

[54] X. Tang, D.K. Du, Z. He, J. Liu, Pyramidbox: A context-assisted single shot face detector, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 797–813.

[55] Z. Yu, H. Huang, W. Chen, Y. Su, Y. Liu, X. Wang, Yolo-facev2: A scale and occlusion aware face detector, Pattern Recognit. 155 (2024) 110714.

[56] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, 2005.

[57] C. Zhu, Y. Zheng, K. Luu, M. Savvides, Cms-rcnn: Contextual multi-scale region-based cnn for unconstrained face detection, in: Deep Learning for Biometrics, Springer, 2017.

[58] M. Kang, K. Ji, X. Leng, Z. Lin, Contextual region-based convolutional neural network with multilayer fusion for sar ship detection, Remote Sens. 9 (2017) 860.

[59] S. Segui, M. Drozdzal, P. Radeva, J. Vitria, An integrated approach to contextual face detection, 2012.

[60] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, 2018, arXiv preprint arXiv:1804.02767.

[61] M. Jogin, M. Madhulika, G. Divya, R. Meghana, S. Apoorva, Feature extraction using convolution neural networks (Cnn) and deep learning, in: 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, RTEICT, IEEE, 2018, pp. 2319–2323.

[62] Y. Zhang, X. Xu, X. Liu, Robust and high performance face detector, 2019, arXiv preprint arXiv:1901.02350.

[63] H. Ma, Y. Liu, Y. Ren, J. Yu, Detection of collapsed buildings in post-earthquake remote sensing images based on the improved Yolov3, Remote Sens. 12 (2020) 44.

[64] Y. Xiao, Q. Yuan, J. He, Q. Zhang, J. Sun, X. Su, . . ., L. Zhang, Space–time super-resolution for satellite video: A joint framework based on multi-scale spatial–temporal transformer, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102731.

[65] K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Wang, X. Wang, . . ., C.W. Lin, Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining, IEEE Trans. Image Process. 30 (2021) 7404–7418.

[66] Y. Xiao, Q. Yuan, K. Jiang, X. Jin, J. He, L. Zhang, C.W. Lin, Local–global temporal difference learning for satellite video super-resolution, IEEE Trans. Circuits Syst. Video Technol. (2023).

[67] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: European Conference on Computer Vision, Springer, 2014, pp. 740–755.

[68] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 658–666.

[69] I. Masi, A.T. Trân, T. Hassner, G. Sahin, G. Medioni, Face-specific data augmentation for unconstrained face recognition, Int. J. Comput. Vis. 127 (2019) 642–667.

[70] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (2016) 1499–1503.

[71] S. Chen, Y. Liu, X. Gao, Z. Han, Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices, in: Chinese Conference on Biometric Recognition, Springer, 2018, pp. 428–438.

[72] L. Llinzai, Lightweight facedetection model designed for edge computing devices, 2019.

[73] D. Mamieva, A.B. Abdusalomov, M. Mukhiddinov, T.K. Whangbo, Improved face detection method via learning small faces on hard images based on a deep learning approach, Sensors 23 (1) (2023) 502.

[74] Y. Xiao, Q. Yuan, J. He, Q. Zhang, J. Sun, X. Su, . . ., L. Zhang, Space–time super-resolution for satellite video: A joint framework based on multi-scale spatial–temporal transformer, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102731.

[75] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, J. Jiang, ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination, IEEE Trans. Multimed. 22 (10) (2019) 2734–2747.