Computer Modeling in
Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.064857

REVIEW

A Comprehensive Review of Face Detection Techniques for Occluded Faces: Methods, Datasets, and Open Challenges

Thaer Thaher[1,*], Majdi Mafarja[2], Muhammed Saffarini[3], Abdul Hakim H. M. Mohamed[4] and Ayman A. El-Saleh[5]

[1]Department of Computer Systems Engineering, Arab American University, Jenin, P.O. Box 240, Palestine
[2]Department of Computer Science, Birzeit University, Birzeit, P.O. Box 14, Palestine
[3]Department of Computer Systems Engineering, Faculty of Engineering and Technology, Palestine Technical University–Kadoorie, Tulkarm, P.O. Box 7, Palestine
[4]Information Systems and Business Analytics Department, A'Sharqiyah University (ASU), Ibra, 400, Oman
[5]Department of Electrical Engineering and Computer Science, College of Engineering, A'Sharqiyah University (ASU), Ibra, 400, Oman

*Corresponding Author: Thaer Thaher. Email: thaer.thaher@aaup.edu

Received: 25 February 2025; Accepted: 30 May 2025; Published: 30 June 2025

ABSTRACT: Detecting faces under occlusion remains a significant challenge in computer vision due to variations caused by masks, sunglasses, and other obstructions. Addressing this issue is crucial for applications such as surveillance, biometric authentication, and human-computer interaction. This paper provides a comprehensive review of face detection techniques developed to handle occluded faces. Studies are categorized into four main approaches: feature-based, machine learning-based, deep learning-based, and hybrid methods. We analyzed state-of-the-art studies within each category, examining their methodologies, strengths, and limitations based on widely used benchmark datasets, highlighting their adaptability to partial and severe occlusions. The review also identifies key challenges, including dataset diversity, model generalization, and computational efficiency. Our findings reveal that deep learning methods dominate recent studies, benefiting from their ability to extract hierarchical features and handle complex occlusion patterns. More recently, researchers have increasingly explored Transformer-based architectures, such as Vision Transformer (ViT) and Swin Transformer, to further improve detection robustness under challenging occlusion scenarios. In addition, hybrid approaches, which aim to combine traditional and modern techniques, are emerging as a promising direction for improving robustness. This review provides valuable insights for researchers aiming to develop more robust face detection systems and for practitioners seeking to deploy reliable solutions in real-world, occlusion-prone environments. Further improvements and the proposal of broader datasets are required to develop more scalable, robust, and efficient models that can handle complex occlusions in real-world scenarios.

KEYWORDS: Occluded face detection; feature-based; deep learning; machine learning; hybrid approaches; datasets

1 Introduction

1.1 Background and Motivation

Face detection is one of the most popular, fundamental, and practical tasks in computer vision. It involves detecting human faces in images and returning their spatial locations through bounding boxes [1], serving as a critical foundation for various advanced vision-based applications [2]. As a vital first step in facial analysis systems, face detection enables subsequent activities such as face alignment, recognition, verification, parsing, emotion detection, and biometric authentication. Its primary purpose is to determine the presence of faces in an image and, if detected, provide their location and extent for further analysis [3–6]. This preprocessing step reduces the amount of data to process and improves the accuracy of the next stages by concentrating on a smaller, relevant portion of the image. It is particularly helpful when dealing with images that have different backgrounds, lighting, and orientations [5]. By removing non-face data, face detection boosts both the speed and accuracy of recognition, making it ideal for large-scale, real-world applications.

Historically, the effectiveness of face recognition technologies has relied on improvements in face detection [7,8]. Early models, such as Viola-Jones [9], used basic features and simple classifiers, while today's deep learning methods utilize advanced convolutional neural networks (CNNs) to achieve greater accuracy. This evolution highlights how progress in detection enhances the overall field of facial analysis. Fig. 1 provides representative examples of typical challenges encountered in face detection tasks. These diverse challenges include variations in scale, atypical poses, occlusions, exaggerated expressions, and extreme illumination. These challenges highlight the need for robust face detection models that can perform accurately in real-world, unconstrained environments. Accurate face detection in such unpredictable environments ensures high-quality images, allowing for enhanced feature extraction and better matching accuracy. This capability is crucial in high-security applications–such as surveillance, biometric identification, law enforcement, airport security, and access control systems–where reducing false positives and improving reliability are paramount [10].

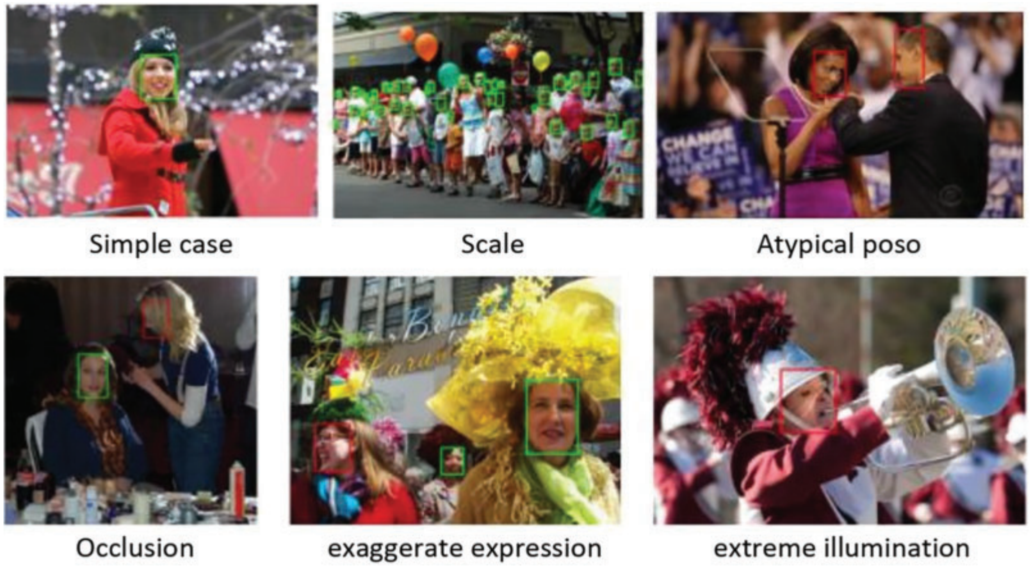


Figure 1: Illustrative examples of face detection challenges including simple cases, variations in scale, atypical poses, occlusions, exaggerated expressions, and extreme illumination. These images, sourced from [1] (WIDER FACE dataset), are provided for illustration purposes and are not linked to any specific models reviewed in this paper

Face detection technology has improved a lot, but it still faces challenges in complex and unpredictable environments [11]. This is why ongoing research is so important to make it more reliable and valuable in practical use. Finding obscured faces is more challenging since important facial traits are sometimes obscure and difficult to identify. The difficulty is to identify faces without depending on clear landmarks, deal with differences in appearance, and even estimate missing elements of the face [12]. One main problem is that occlusions can obscure vital face traits as the lips, nose, or eyes. External objects, body parts, or ambient elements including clothes, hands, sunglasses, or masks [13] can all cause these obstructions (as illustrated in Fig. 2).

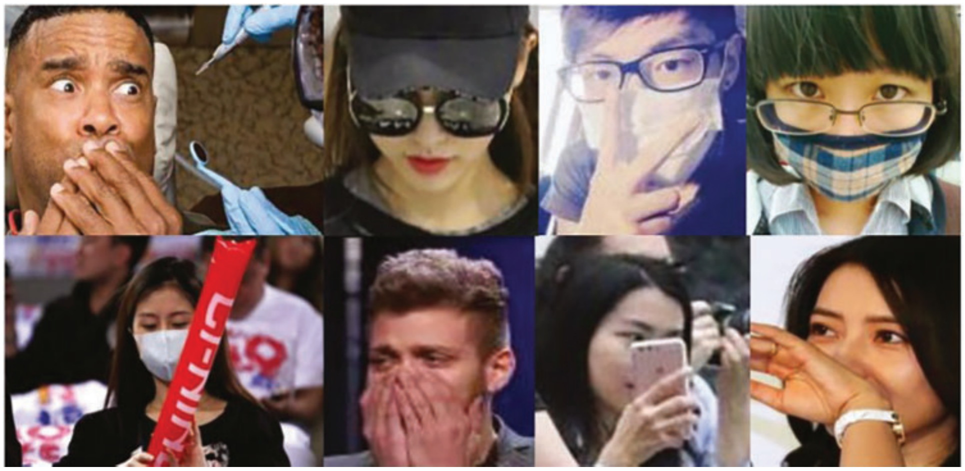


Figure 2: Example of occluded face images from the MAFA dataset [14]

Usually, depending on the evaluation of the complete face or particular landmarks, standard face detection techniques fail when these features are obscured, resulting in missed detections or large false positives [15]. For real-time applications like surveillance and security, where failing to identify obstructed faces can lead to major mistakes, this is particularly problematic. Accordingly, more advanced methods are required that can identify and infer facial features even in cases when significant portions of the face are covered in order to enhance occluded face detection.

Another big challenge in face detection is the way occlusions change a face's appearance. In specific, the size, shape, and texture of an occlusion can be different even when it covers the same part of the face [13,16]. This makes it harder for traditional face detection systems, which expect faces to look consistent. Occlusions also cause confusion by altering the usual relationships between facial features. Deep learning models, especially CNNs, have been very effective in handling these challenges by learning patterns in both clear and occluded faces [17,18]. These models require large datasets of faces with different levels of occlusion for training, but there is one major problem: there is a lack of well-annotated datasets. As for the models, they fail to detect the occlusions from different angles and other conditions due to the insufficient variety of examples [6].

Furthermore, a crucial part in the case of an occluded input is to preserve the balance between the detection of the visible features and the comprehension of the entire face. Traditional methods are based on the recognition of the complete facial structures, and, in case of partial obfuscation, the models learn to complete the sequence based on the available information. Recent techniques such as attention mechanisms and region-based detection are applied to focus on the regions of interest and context to infer missing regions [19,20]. However, such approaches are computationally expensive and therefore infeasible for real-time applications [21].

These challenges highlight the unique difficulties of detecting occluded faces compared to regular face detection. Continued research in this area is essential for developing more robust face detection systems capable of handling diverse and unpredictable real-world conditions. The emphasis on occluded face detection has gained importance in the AI and computer vision domain due to the rising need for reliable detection in practical applications. Facial analysis technology, integrated into fields like healthcare, security, retail, and social media, must function effectively in unregulated environments where occlusions are common. The COVID-19 pandemic, with widespread mask usage, underscored the need for algorithms capable of efficient detection under partial visibility [22–24]. In security and surveillance, detecting occluded faces is critical, as individuals often obscure their faces with items like hats, scarves, sunglasses, or masks, whether intentionally or not [25,26]. Improved occluded face detection can enhance AI's effectiveness in such critical applications [27]. In healthcare, occlusions from medical devices or environmental factors can obstruct critical facial regions, complicating tasks like emotion detection, gaze tracking, and diagnosing neurological disorders. Advances in AI-based facial analysis are contributing to more reliable telemedicine and assistive technologies, supporting consistent and accurate assessments even in challenging visual conditions [28]. Similarly, in social media, marketing, and retail, occluded face detection is vital for analyzing expressions, demographics, and engagement in dynamic situations. Retail environments often involve occlusions due to product displays or interactions, while social media platforms must detect faces obstructed by accessories or filters. Enhancing detection in these scenarios ensures AI systems perform accurately and fairly and improve their effectiveness across diverse use cases. These advancements are vital for addressing the challenges outlined across diverse applications, particularly in hidden face detection.

1.2 Objectives and Contributions of the Review

As highlighted earlier, the past few years have seen significant growth in research on face detection under occlusions. This increase reflects a growing demand to review and assess the impact of advancements in this field. The analogous challenges and notable progress in occluded face detection have motivated us to conduct this review study. The main objective of this study is to provide a comprehensive resource for researchers and practitioners interested in this topic. Face detection under occlusion remains critical for improving the reliability of real-world applications, such as surveillance, biometric authentication, and access control, where partial facial visibility is common. By providing an organized analysis of current methods, challenges, and datasets, this review aims to guide future research efforts and support practitioners in developing more robust face detection systems. To achieve this objective, we make the following key contributions:

1. We comprehensively review recent state-of-the-art approaches in the domain of occluded face detection, categorized into traditional feature-based methods, machine-learning-based approaches, advanced deep learning techniques, and hybrid methodologies.
2. We highlight key advancements, persistent challenges, and gaps in the field of occluded face detection, providing valuable insights into utilizing emerging technologies for diverse research directions.
3. We summarize and compare the reviewed approaches under varying conditions, offering a clear understanding of their strengths and limitations.
4. We analyze and compare benchmarking datasets commonly used to evaluate the performance of face detection systems under occlusions, emphasizing their characteristics and applicability.
5. We outline current challenges and promising research directions, inspiring further innovation and progress in this important area.

This paper focuses exclusively on face detection techniques, distinguishing them from face recognition methods. Specifically, it addresses the unique challenges and methodologies associated with detecting faces under occlusions. By narrowing the scope to this critical task, the review offers an in-depth analysis of improved detection approaches, contributing to the broader domain of face analysis. A specialized review paper on detecting occluded faces is essential, as most existing review studies on face detection and recognition approaches overlook the challenges posed by occlusions. Current review articles primarily concentrate on general face detection techniques or face recognition methodologies, with few explicitly addressing obstructed faces. For instance, some research examines face recognition algorithms in the context of occlusion [14,29–32], but their main focus is on identity verification rather than the foundational task of identifying the existence and location of occluded faces. Reviews on general face detection, such as [5,12,33–35], often assume full facial visibility and inadequately address the unique challenges of partial visibility or feature masking caused by occlusions. Our review paper, as a recent contribution to the field, addresses this gap by examining both past and recent studies on occluded face detection. It provides a focused resource delivering insights into detection strategies designed to solve diverse occlusion difficulties, offering a timely and essential reference for advancing this critical area of research. To sum up, the main contributions of this timely study are as follows.

1.3 Paper Structure

The rest of this paper is organized as follows: Section 2 explains the detection of faces under occlusion, focusing on the source, type, and level of occlusions. Section 3 presents the structured methodology used to prepare this review study. Section 4 analyzes, categorizes and compares state-of-the-art methods for the detection of occluded faces. Section 5 lists and compares the benchmark datasets used for occluded face detection. The challenges and future directions in the detection of occluded faces are highlighted in Sections 6 and 7, respectively. Finally, Section 8 concludes the study.

2 Background and Foundations of Occluded Face Detection

Before presenting a detailed review of detection techniques, it is important to first establish the necessary background and terminology that will be used throughout the rest of the paper. Section 2.1 clarifies the differences between face detection and face recognition. Section 2.2 discusses the various sources of occlusions, and Section 2.3 presents the types and levels of occlusions. Understanding these sources, types, and severity levels is crucial for accurately assessing and comparing face detection methods.

2.1 Face Detection vs. Face Recognition

Face detection and face recognition are two distinct but interrelated activities within the realm of computer vision. Face detection is the process of recognizing and localizing faces in an image or video frame, typically preceding additional facial analysis [7]. Detection systems focus on precisely identifying the facial region, regardless of identity, pose, or expression, and are essential for applications such as security monitoring, photo tagging, and human-computer interaction [15,36]. Face detection can be represented as a function. Given an input image x, the function $f(x, \theta)$ generates an output vector that indicates the location of a detected face. This vector (x, y, w, h) specifies the coordinates of the top left corner (x, y) and the width and height (w, h) of the bounding box around the face [37]. The parameter θ includes factors such as thresholds, settings, or contextual information that guide the detection process. The function may also return a confidence score that indicates the likelihood that the detected region contains a face.

In contrast, face recognition involves the identification or verification of the identity of a recognized face. Recognition jobs often depend on extracting robust features from the identified face to compare it with established identities in a database [38]. Although face recognition is based on detection, its objectives and challenges are significantly different. Recognition systems emphasize feature extraction and comparison, frequently utilizing methods such as feature embedding [39], while detection systems concentrate on rapid localization and generalization under diverse situations, including variations in illumination and occlusions. Fig. 3 illustrates the general face analysis pipeline, highlighting the steps of detection, feature extraction, and recognition. To clarify the difference between face detection and recognition, Table 1 presents a comprehensive comparison of the two tasks across various aspects.

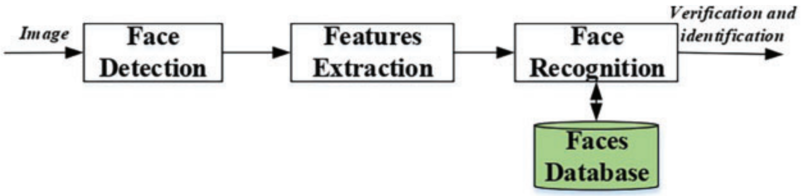


Figure 3: General pipeline illustrating the difference between face detection and face recognition [27]

Table 1: Comparison between face detection and face recognition

Aspect	Face detection	Face recognition
Definition	Identifying and locating faces in an image or video frame.	Identifying or verifying the identity of detected faces.
Purpose	Acts as a preprocessing step for further facial analysis.	Determines or confirms the identity of individuals.
Input requirements	Raw images or video frames.	Face regions detected from a face detection process.
Output	Bounding boxes or coordinates of detected faces.	Identity labels or verification scores.
Key challenges	Variations in illumination, occlusions, pose, and background clutter.	Feature similarity among individuals, occlusions, and spoofing attacks.
Techniques used	Haar cascades, HOG, CNNs, YOLO, Faster R-CNN.	Embedding-based methods (e.g., FaceNet, DeepFace), Siamese networks.
Applications	Security surveillance, human-computer interaction, photo tagging.	Biometric authentication, access control, identity verification.

2.2 Sources of Occlusions

Occlusions in face detection arise from a variety of sources, each introducing unique challenges for detection algorithms [14]. To organize these challenges, occlusions can be broadly categorized based on their source and context. These categories reflect whether the obstruction arises from personal accessories, physical objects, environmental conditions, artificially created elements, or severe real-world challenges such as disguises or crowding. These sources, summarized in Table 2, are detailed as follows:

Table 2: Sources of occlusions and their characteristics

Category	Examples	Challenges
Facial accessories	Sunglasses, eyeglasses, hats, scarves, helmets	Obscures key facial landmarks; static but variable across users

(Continued)

Table 2 (continued)

Category	Examples	Challenges
Objects (External)	Books, mobile phones, microphones	Partial/full blockage; unpredictable sizes and locations
Objects (Self-Imposed)	Hands, arms, gestures (e.g., covering mouth, shielding eyes)	Dynamic occlusions; frequent and hard to predict
Environmental (Shadows)	Uneven lighting casting shadows on the face	Mimics occlusions, causing misclassification
Environmental (Reflections)	Glare from glasses or shiny surfaces	Obscures details in certain regions
Environmental (Lighting)	Backlighting, poor visibility, sudden lighting changes	Distorts facial features, reducing accuracy
Artificial occlusions	Digital masks, stickers, graphical elements	Adds complexity for training and real-world applications
Severe occlusions (Crowding)	Faces blocked by other people in crowds	Partially hidden faces; common in crowded places
Severe occlusions (Disguises)	Masks, wigs, veils to conceal identity	Deliberate concealment challenges detection
Severe occlusions (Privacy)	Cultural or personal face coverings for privacy	Highly occluded faces

1. Facial accessories: Everyday accessories such as sunglasses, eyeglasses, hats, scarves, and helmets often block important facial features. For instance, sunglasses obscure the eyes, while masks cover the nose and mouth, disrupting algorithms that depend on these features for detection [40,41]. Although these occlusions are typically static, their variety across individuals makes them challenging to handle.
2. Objects: Occlusions caused by objects can be divided into two types:
   - External objects: Items like books, phones, or microphones can partially or fully block the face, especially during activities like reading or speaking [12,42].
   - Self-imposed obstructions: Hands, arms, or gestures, such as covering the mouth or shielding the eyes, create dynamic occlusions. These vary in size, shape, and location, making them particularly hard to predict and handle [16].
3. Environmental factors: The environment can create occlusions that interfere with detection systems in various ways:
   - Shadows: Uneven lighting can cast shadows on the face, making it appear partially occluded [34].
   - Reflections: Glare from glasses or shiny surfaces can hide important facial details [43].
   - Lighting variations: Sudden changes in lighting, such as backlighting or low visibility, can distort facial features and lower detection accuracy [35].
4. Artificial occlusions: These are intentionally created occlusions, often used to test algorithms or ensure privacy. Examples include digital masks, stickers, or other graphical overlays on faces [14]. While useful for training models, artificial occlusions can make real-world detection more challenging.
5. Severe occlusions in specialized scenarios: Some occlusions are more extreme and deliberate, particularly in real-world contexts like security or surveillance [44]. These include:

- Crowding: Faces may be partially blocked by other people in crowded places like public transport or events.
- Disguises: Deliberate obstructions such as masks, wigs, or veils are often used to conceal identity.
- Privacy measures: Cultural or personal practices, such as wearing face coverings for religious or privacy reasons, can make detection systems less effective [45].

To better illustrate the different sources and forms of occlusions, Fig. 4 presents a range of occluded face examples, including real-world occlusions (e.g., sunglasses, scarves), synthetic occlusions (e.g., digital masks), and unrelated obstructing objects.

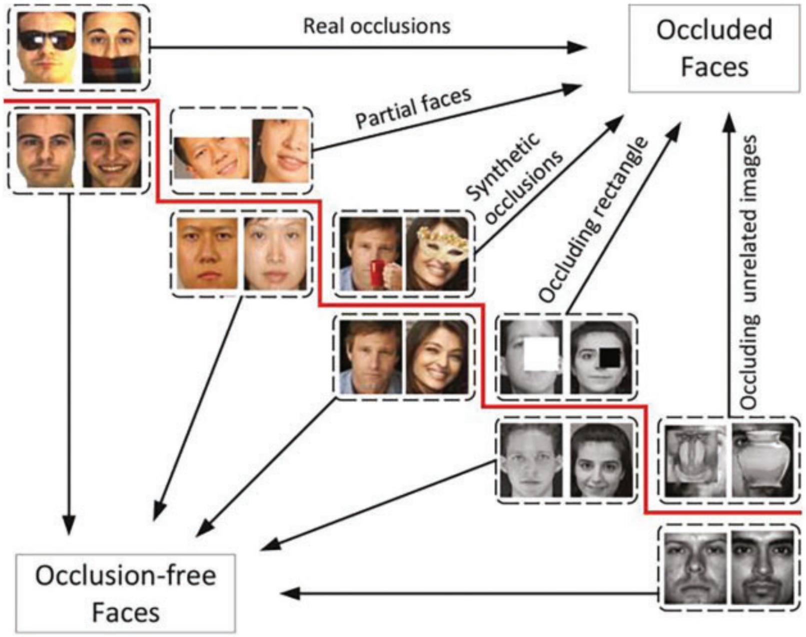


Figure 4: Examples of occlusion types commonly encountered in face detection tasks, including real-world occlusions, synthetic occlusions, partial faces, and unrelated occluding objects. Adapted from [14]

2.3 *Face Occlusion Types and Levels*

Handling occlusion in face detection requires a comprehensive understanding of the different types and degrees of occlusion. These elements directly affect the design and effectiveness of detection algorithms. This section examines the basic types of occlusion, their characteristics, and their consequences for detection methodologies, with reference to current literature. Occlusion can be categorized from two complementary perspectives: the overall level of coverage (partial and high occlusion) and the areas that are most impacted (spatial occlusion). Together, these classifications provide a comprehensive and practical foundation for addressing the various issues posed by occlusion.

2.3.1 Partial and High Occlusion

Occlusion can be categorized based on the degree of coverage into partial and high occlusion. For example, studies like [4,46] classify faces into three groups: non-occluded, partially occluded, and heavily occluded, depending on the percentage of the face area covered. Partial occlusion is defined as 1% to 30% coverage, while heavy occlusion exceeds 30%. Although this method offers a clear framework for classification, it may fail to account for extreme cases, such as fully obscured faces where 100% of the face is covered. Another study by [13] divides the face into four main regions: eyes, nose, mouth, and chin. They categorized occlusion levels based on how many regions are covered. Faces with one or two covered regions are classified as weakly occluded, three regions as medium occlusion, and all four as heavily occluded. While this approach adds more detail, it struggles to distinguish between different degrees of heavy occlusion, such as 70% vs. 100% coverage. To address this, reference [37] refines the classification by dividing the face into five regions: the forehead, two eyes, nose, mouth, and chin. This finer division helps differentiate between heavily occluded and fully occluded faces, which is particularly useful in scenarios like faces covered with niqabs (a cultural or religious head covering worn by some Muslim women [6]). By accounting for occlusion levels from 70% to 100%, this method provides a more detailed understanding of high occlusion.

According to the aforementioned studies, occlusions can be classified into partial occlusion and high occlusion, with additional distinctions based on the extent of covered and certain facial parts that are obscured.

- **Partial occlusion** arises when a segment of the face is obstructed, for instance, when the eyes are concealed by glasses or the mouth is obscured by a hand [47]. Partial occlusion disturbs the symmetry of face landmarks, upon which numerous algorithms depend for precise detection. Traditional feature-based approaches, such as Haar cascades and HOG, frequently struggle to generalize effectively under partial occlusions due to their reliance on the total visibility of essential face features.
- **High occlusion** denotes instances in which over 50% of the face is concealed. Typical instances encompass faces obscured by masks, scarves, or environmental obstacles such as foliage [13]. High occlusion is a considerable obstacle for conventional techniques and certain deep learning methodologies, as the visible areas may lack adequate information for dependable detection. Recent advancements in deep learning, including attention mechanisms and occlusion-aware models, have demonstrated the potential to tackle these scenarios.

2.3.2 Spatial Occlusion

In addition to the overall degree of coverage, occlusion can also be categorized based on the specific regions of the face that are obstructed. Spatial occlusion examines how particular areas of the face are obscured, which can have distinct impacts on detection algorithms. For example:

- Upper Occlusion: Obstructions of the forehead and eyebrows, such as those caused by hats or hair, can interfere with alignment algorithms that rely on these features.
- Lower Occlusion: Covering the mouth and chin, as with masks or scarves, poses challenges for recognition tasks that depend on these regions.

2.3.3 Levels of Occlusion

To better understand the impact of occlusion, it is essential to categorize it into levels:

- Low Occlusion: Less than 25% of the face is obscured. Examples include glasses or slight shadows.
- Medium Occlusion: Between 25% and 50% of the face is obscured. Examples include medical masks or objects partially blocking the face.
- High Occlusion: More than 50% of the face is obscured, further subdivided into:
  - Heavily Occluded: 50% to 70% of the face is covered, such as by scarves or environmental barriers.
  - Fully Occluded: 70% to 100% of the face is covered, as in cases like niqabs or full veils.

Fig. 5 illustrates examples of faces with varying degrees of occlusion, ranging from no occlusion to fully occluded faces. Occlusion levels in Fig. 5 are based on the approximate percentage of the facial area covered by occluding objects (e.g., hands, masks, glasses). High occlusion is defined as more than 50% of the face being obscured by objects such as masks, hands, or other barriers. The visualization highlights the progressive challenges introduced as occlusion levels increase, demonstrating the necessity for robust algorithms capable of handling each scenario effectively.

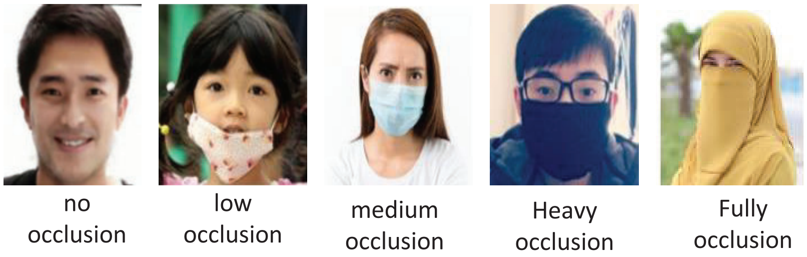


Figure 5: Examples illustrating different levels of occlusion, ranging from no occlusion to full occlusion. The images are sourced from publicly available benchmark datasets [6,13]. Low occlusion covers less than 25% of the face, medium occlusion covers 25%–50%, heavily occluded faces have 50%–70% coverage, and fully occluded faces exceed 70%. The images highlight the increasing challenge for face detection as occlusion levels rise

3 Review Scope and Methodology

This section describes the methodology used to choose and evaluate the existing literature on face detection under occlusion. It also includes a summary and gap analysis of previously published review studies in this domain. This helps justify the need for the present review and clarify its distinct contributions. This review study followed a structured methodology to ensure a comprehensive evaluation of face detection techniques for occluded faces. The methodology was divided into several stages, as shown in Fig. 6, which illustrates the process from the literature search to classification and summarization of results.

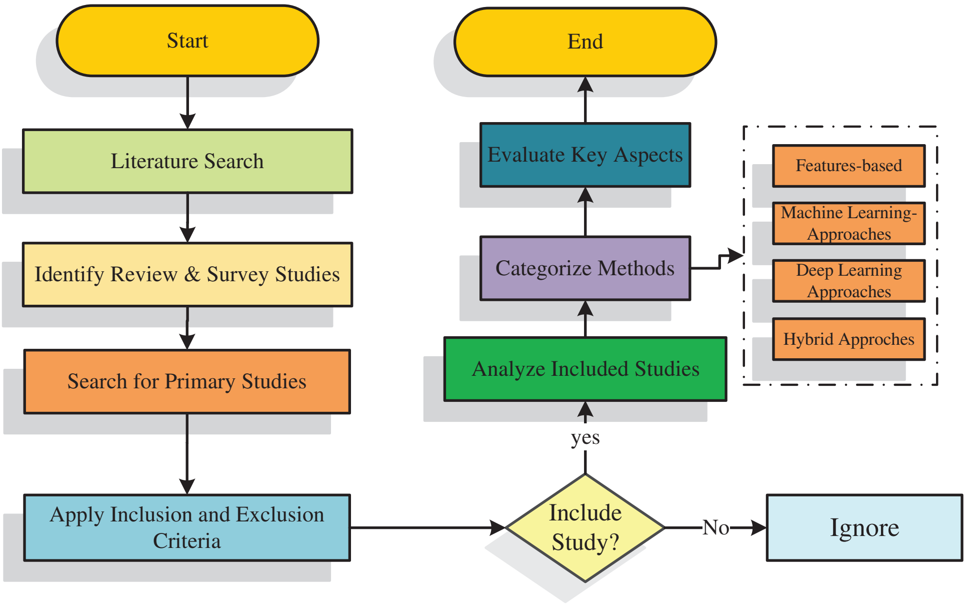


Figure 6: Structured methodology for evaluating face detection techniques under occlusion

3.1 Search Strategy and Inclusion Criteria

To ensure comprehensive coverage, our review was conducted in two phases. In the first phase, we focused on identifying and analyzing existing review and survey studies related to face detection and recognition. After identifying gaps in previous surveys, we proceeded to the second phase, where a systematic search was conducted across multiple academic databases, including IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Google Scholar. Keywords used during the search included "face detection," "face recognition," "occlusion," "occluded face detection," "partially occluded faces," "masked face detection," "face detection with masks," "face detection under unconstrained environments," "survey," and "review paper." For the second phase, the scope included both partial and heavy occlusions to cover a wide range of scenarios. We applied specific criteria to include or exclude studies in our review. Studies published between 2010 and 2024 were considered. In addition, the inclusion criteria focused on studies that address face detection techniques under occlusion, papers proposing novel algorithms, frameworks, or datasets, and studies providing experimental evaluations using standard benchmarks or self-customized datasets. In contrast, the exclusion criteria eliminated studies that focused entirely on face detection or face recognition rather than handling occlusion scenarios. Studies focusing purely on face recognition or identity verification under occlusion were excluded unless they included detection components.

3.2 Categorization and Evaluation of Methods

Based on the analysis of the retrieved studies, the selected papers were grouped into four primary categories: feature-based approaches, traditional machine learning approaches, advanced deep learning-based approaches, and hybrid approaches. Each category was further divided into subcategories to capture the specific methodologies employed. These subcategories allowed us to highlight key differences in techniques, such as handcrafted feature extraction, statistical modeling, ensemble learning methods, CNNs, recent advancements based on attention mechanisms, transformer architectures, and Generative Adversarial Networks (GANs), and combinations of traditional and modern approaches.

To evaluate and compare the reviewed methods, we analyzed several key aspects across all studies. Firstly, we captured the proposed approaches and methodologies employed in each study. This involved identifying whether the methods relied on handcrafted features, machine learning algorithms, deep learning frameworks, or hybrid techniques. After that, we examined the validation datasets, including standard benchmarks and custom datasets designed for occlusion scenarios. Then, the occlusion levels that were considered in each study were evaluated in order to determine the adaptability of the proposed methods. The evaluation metrics used in the studies were also summarized, including precision, recall, mean average precision (mAP), F1 scores, and intersection over union (IoU). Finally, the advantages and disadvantages of each approach were evaluated, and its strengths were noted in a specific context, such as computational cost, generalization ability, and sensitivity to variations in occlusion type or severity. This methodology allowed us to determine the gaps and trends, especially in the complex scenarios of detecting covered faces. It also allowed us to distinguish the open problems, propose directions for detecting faces under partial and severe occlusions. It also enabled us to assess the potential of the reviewed techniques for real-time applications and challenging unconstrained environments.

3.3 Summary of Existing Review and Survey Studies

This section examines and classifies existing review articles and surveys related to face detection, particularly those that have systematically examined progress in this field. Based on our analysis of the existing literature, review studies can be classified into three main groups: face detection, face recognition, and a combination of the two approaches. Table 3 presents a detailed summary of previously published review and survey studies related to face detection and recognition. It highlights each study’s focus area, specificity to occlusion challenges, dataset coverage, evaluation metrics, open challenges, and comparative results. This table helps to clearly identify the gaps in existing surveys, reinforcing the motivation for a focused and dedicated review of occluded face detection.

Table 3: Summary of previously published survey and review studies on face detection and recognition

Study	Focus area	Specific to occlusion	Key limitation regarding occluded face detection	Dataset coverage	Evaluation metrics	Open challenges & future directions	Comparative results
Sharifara et al. [48] 2014	Face detection	No	Primarily addresses fully visible faces without occlusion	Limited	Yes	No	No
Kumari and Kaur [34] 2023	Face detection	No	Limited focus on occlusion, centered on visible faces	Yes	Yes	Yes	Yes
Hire and Satone [49] 2018	Face detection	No	Discusses techniques only for visible faces, no occlusion handling	Yes	Yes	Yes	Yes
Thazheena and Aswathy Devi [50] 2017	Face detection	Yes	Limited to accessory-induced occlusions, does not consider advanced CNN techniques	Yes	Yes	Yes	No
Rao et al. [31] 2015	Face recognition	Yes	Primarily focused on identity verification, limited exploration of detection aspects	Yes	Yes	Yes	Yes
Jafri and Arabnia [8] 2009	Face recognition	No	No occlusion-specific focus	Yes	Yes	Yes	Yes
Zafeiriou et al. [12] 2015	Face detection	No	Focus on detection in unconstrained environments, minimal occlusion focus	Yes	Yes	Yes	Yes
Zeng et al. [14] 2021	Mixed	Yes	Emphasizes recognition, limited focus on recent advancements in detection	Yes	Yes	Yes	Yes
Zhang and Zhang [5] 2010	Face detection	No	Limited to pre-2010 techniques, does not cover modern deep learning approaches or occlusion handling	Yes	Yes	Yes	Yes
Wang et al. [43] 2021	Face recognition	Minimal	Primarily focuses on deep learning for recognition, minimal occlusion handling in detection	Yes	Yes	Yes	Yes
Deep Learning-based Occluded Person Re-ID	Face recognition	Yes	Focuses on Re-ID post-detection, limited discussion on occluded face detection	Yes	Yes	Yes	Yes
Ruvinga et al. [35] 2019	Face detection	Minimal	Limited coverage on occlusion and lacks recent deep learning advancements for occlusion	Yes	Yes	Yes	Yes
Budiarsa et al. [30] 2023	Face recognition	Yes	Focuses on recognition under occlusion with limited discussion on detection techniques	Yes	Yes	Yes	No
Zhang et al. (2018) [32] 2018	Facial expression analysis	Yes	Focused on FEA rather than face detection or broad recognition techniques	Yes	Yes	Yes	No
[51] 2021	Face detection	Minimal	Focuses on deep learning advancements for detection, limited occlusion-specific handling	Yes	Yes	Yes	Yes
Alzu’bi et al. [33] 2021	Face recognition	Yes	Primarily focuses on masked face recognition, limited exploration of other occlusion types	Yes	Yes	Yes	Yes
Mondal et al. [52] 2020	Face detection	Minimal	Focuses on traditional techniques, lacks extensive occlusion handling, does not list prior studies per approach	Yes	Yes	Yes	Yes

(Continued)

Table 3 (continued)

Study	Focus area	Specific to occlusion	Key limitation regarding occluded face detection	Dataset coverage	Evaluation metrics	Open challenges & future directions	Comparative results
Dagnes et al. [53] 2018	Face recognition	Yes	Limited to 3D face recognition, no focus on general face detection methods	Yes	Yes	Yes	Yes
Hasan et al. [54] 2021	Face detection	No	Primarily focuses on traditional and early deep learning methods, lacks occlusion-specific advancements	Yes	Yes	Yes	Yes
Kortli et al. [27] 2020	Face recognition	Minimal	Limited analysis on occlusion-resilient techniques, focuses on general recognition	Yes	Yes	Yes	Yes
Singh et al. [44] 2024	Face recognition	Yes	Emphasizes challenges in disguise and crowd scenarios, limited detection focus	Yes	Yes	Yes	Yes
Zhang et al. [29] 2021	Mixed	Yes	Primarily focused on recognition with limited detection strategies for occlusion	Yes	Yes	Yes	Yes

As summarized in Table 3 and highlighted in Fig. 7, most existing reviews do not specifically focus on occlusion-related techniques or provide only minimal insights. This highlights the need for a dedicated review of face detection under occlusion. Reviews such as those of [34,48,49], focus primarily on challenges such as pose, lighting, and background clutter, assuming full facial visibility. These studies offer limited insight into scenarios where occlusions significantly reduce detection accuracy. On the other hand, face recognition reviews, such as those of [30,31], focus on managing occlusions during identity verification. While these studies explore feature extraction and reconstruction techniques to mitigate occlusion effects, they do not address the critical challenge of detecting occluded faces. Similarly, mixed approaches, as shown by [14,29], touch on detection methods but lack an in-depth analysis of techniques specifically tailored to occluded face detection. In addition, many studies fail to adequately address data sets and evaluation metrics for occluded face detection. Furthermore, challenges such as reducing false positives caused by occlusion patterns, creating generalizable detection models, and handling different occlusion levels remain underexplored. Therefore, this review aims to address these gaps by providing a focused analysis of occluded face detection techniques, datasets, challenges, and future directions.

After defining the scope of existing review studies and confirming the necessity of dedicated occluded face detection analysis, the following section presents a thorough examination of current methodologies. The analysis groups detection methods according to their fundamental approaches, which range from traditional feature-based models to deep learning and hybrid strategies, including recent Transformer and GAN-based techniques.

4 Review of Detection Methods for Occluded Faces

In this part, the current studies in detecting occluded faces are systematically and comprehensively reviewed. It analyses and groups several strategies according to their basic approaches. Separate subsections in the section are arranged such that each one corresponds with a main category of techniques. Based on the reviewed literature, methods are categorized into three main groups: (1) classical techniques, including handcrafted feature-based and traditional machine learning-based models; (2) advanced deep learning-based methods, further divided into CNN-based, attention-driven, transformer-based, and GAN-based models; and (3) hybrid methods that integrate traditional and modern strategies. Based on their approaches, applications, degrees of occlusion handled, used datasets, advantages, constraints, and evaluation criteria, a thorough review of the relevant studies is offered for every category. Each section ends with a summary table methodically demonstrating outcomes to enable simple, easy comparisons of approaches. Visualizations are given to highlight important trends and developments in the area, including the publication distribution over time, the relative popularity of every method category, and comparative performance measures across common datasets.

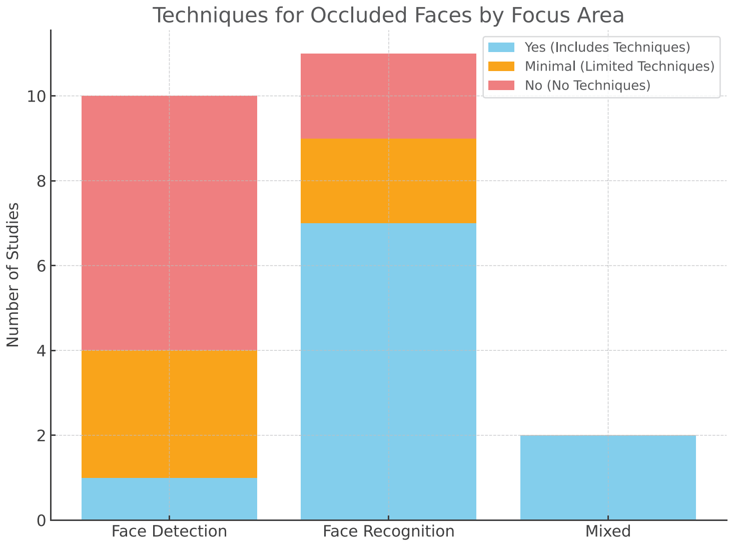


Figure 7: Distribution of review and survey studies on face detection and recognition techniques for occluded faces. The data is based on the studies summarized in Table 3 which are collected through manual database searches using relevant keywords

4.1 Classical Approaches for Occluded Face Detection

Before the advance of deep learning, face detection relied primarily on hand-crafted features and traditional machine learning [12,55,56]. These traditional methods were able to detect faces despite partial occlusion, but they struggled in complex real-world situations. This section briefly reviews the main types of these methods, including feature-based techniques and early learning models. Tables 4 and 5 present a comparative summary of the reviewed methods. It is worth mentioning that for traditional methods, the evaluation metrics are presented as reported in the original studies, whereas standardized benchmarks such as mAP are used where available for recent methods.

4.1.1 Traditional Feature-Based methods

Early approaches to face detection, such as those based on Haar-like features [9], Local Binary Patterns (LBP) [57], Histogram of Oriented Gradients (HOG) [58], edge-based techniques (e.g., Canny and Sobel filters) [59,60], and statistical models like Active Shape Models (ASM) [61] and Active Appearance Models (AAM) [62], relied heavily on manually designed features to extract facial patterns such as edges, textures, and geometric relationships [54]. These techniques are particularly helpful in cases when computational efficiency is more crucial, since they focus on identifying fundamental face components and apply predetermined descriptors to help in detection. Although computationally efficient and easy to implement [63], their performance significantly deteriorates in the presence of partial occlusions, lighting changes, and complex backgrounds. Some solutions attempted to enhance robustness through part-based or occlusion-aware models. For instance, Guo et al. [64] introduced an AdaBoost cascade classifier that utilized Haar-like features and facial proportion rules to detect occluded faces on the MAFA dataset. Similarly, Bade and Sivaraja [65] proposed enhancements to the Viola-Jones framework using heuristic boosting and decision tree tuning, showing improved results on partially occluded faces from WIDER FACE. On the other hand, Ganguly et al. [66] developed two geometric feature-based methods that employed 3D depth information to localize occlusions, achieving strong detection results on the Bosphorus database. Despite these efforts, traditional methods remain limited in flexibility and robustness, especially when handling transparent or irregular occlusion patterns, and have largely been surpassed by deep learning models. Table 4 presents a detailed summary of the reviewed feature-based methods for occluded face detection.

Table 4: Summary of reviewed feature-based methods for occluded face detection

Year	Paper	Category	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2015	Ganguly et al. [66]	Geometric feature-based	Threshold-Based and Block-Based depth analysis	Bosphorus Database	Partial, transparent occlusions	Detection Accuracy	Threshold: 91.79%; Block-Based (5 × 5): 99.71%	Effective in detecting occlusions and localizing regions	Struggles with transparent occlusions (e.g., glasses); fixed block sizes
2018	Guo et al. [64]	Haar-Like Feature-Based	Adaboost cascade classifier with Haar-like features and facial physiology relationships	MAFA	Partial, High	Detection Rate, false positive rate	Detection rate: 57.3%; False positive rate: 4.7%	Utilizes physiological heuristics for occlusion robustness	Limited precision; Higher false positive rates; ineffective when eyes and mouth are occluded; lacks generalization for complex occlusions
2020	Bade & Sivaraja [65]	Haar-Like Feature-Based	Enhanced Haar cascade with heuristic boosting and optimized CART depth	WIDER FACE	Partial	Accuracy, F1 Score	65.25% accuracy; 77.17% F1 score; 23.66% better than Haar-frontalface-default	Improved accuracy and F1 score for partially occluded faces; computational efficiency through grayscale preprocessing	Struggles with extreme occlusions and profile faces; limited robustness for diverse occlusion patterns

4.1.2 Traditional Machine Learning-Driven Methods

This section reviews traditional machine learning-based approaches used for occluded face detection. These approaches typically rely on predefined features extracted from facial images, followed by the application of classical machine learning algorithms to classify regions as faces or non-faces. Unlike feature-based methods, which depend solely on handcrafted features, machine learning-based methods combine feature extraction with learning algorithms to improve adaptability and accuracy. The studies in this category focus on techniques that learn patterns from labeled training data to detect occlusions and distinguish them from normal facial structures. These methods are particularly effective when paired with dimensionality reduction techniques, such as Principal Component Analysis (PCA). In this paper, the reviewed techniques are categorized into three main groups: (1) statistical and clustering models, (2) ensemble and boosting methods, and (3) Support Vector Machine (SVM)-based methods.

Statistical and clustering models, such as the method by Jabbar and Hadi [67], use fuzzy clustering, pixel similarities, or probabilistic analysis to segment occluded regions and reconstruct missing parts. Although this method was an early effort in addressing occluded face recovery, it is now considered outdated and is not applicable to modern real-time face detection systems. Ensemble and boosting methods, including the works of Gul and Farooq [68], Arunnehru et al. [69], and Liao et al. [70], improve detection by combining multiple weak classifiers using AdaBoost or decision tree ensembles to focus on hard-to-detect samples and manage occlusion and background complexity. Over time, researchers increasingly explored SVM-based techniques due to their strong discriminative power and ability to handle partial occlusions. These methods, such as those proposed by Hotta [71], Priya and Banu [72], SuvarnaKumar et al. [73], Zohra et al. [74], and Yang et al. [75], apply discriminative learning using handcrafted features, local descriptors, or depth-based cues to classify occluded regions. While these machine learning approaches showed improved detection over earlier basic feature-based methods, they still face challenges in handling diverse occlusion types and adapting to complex real-world settings due to their reliance on static feature representations. Table 5 summarizes the key characteristics of the reviewed machine learning-based techniques.

Table 5: Summary of reviewed machine learning-driven approaches

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2007	Hotta [71]	SVM with local summation kernel and Gabor filters	HOIP, MIT+CMU, PIE	Partial occlusion (sunglasses, scarves, shadows)	TPR, FPR	High accuracy in occluded settings; outperformed global-kernel-based SVMs	Robust to partial occlusion; effective use of local features	High computational cost due to multiple local kernels
2010	Jabbar & Hadi [67]	Skin segmentation, eye template matching, Fuzzy C-Means clustering	Custom dataset	Partial	Detection accuracy, recovery quality	93% detection accuracy (clean background), 73% recovery quality for 40% occlusion	Effective for symmetric occlusions; novel use of fuzzy clustering for segmentation	Struggles with non-frontal views; limited database size for asymmetric recovery
2012	Kumar et al. [73]	Circular Hough Transform, SVM for occlusion detection, HSPCA for recognition	Custom dataset	Partial	Accuracy	Achieved 94% accuracy in controlled environments	Effective for controlled environments; robust to skin-tone background challenges	Limited scalability; low adaptability to unconstrained environments; small dataset size
2012	Priya et al. [72]	MBWM feature extraction with SVM	MIT Face Database	Partial occlusion	Classification accuracy	98.75% accuracy with overlapping RBF SVM on two segments	Outperformed LBP and SLBM; effective in detecting partial occlusions	Relies on handcrafted features; segmentation limited to fixed regions

(Continued)

Table 5 (continued)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2015	Gul et al. [68]	AdaBoost with free rectangular features, skin color detection	FDDB	Partial occlusion	TPR, FPR	TPR: 33% at 19% FPR; improved accuracy over traditional Viola-Jones in detecting occluded faces	reduced false positives through skin color detection	Struggles with blurred or small-scale faces in distant or low-resolution regions
2015	Arunnehru et al. [69]	SFTA with tree-based classifiers (Random Forest, Decision Tree)	PNNL Parking Lot	Partial, full occlusion	Accuracy, precision, recall	Random Forest: 98.3% (SET-1), 98.2% (SET-2), 83.7% (SET-3); highest among tested classifiers	Effective in surveillance settings; robust to partial occlusion	Limited to handcrafted features; struggles with extreme occlusions
2016	Liao et al. [70]	NPD features with soft-cascade AdaBoost	FDDB, GENKI, CMU-MIT	Arbitrary occlusions, poses	TPR, false positives	Improved detection rate; 6x faster than Viola-Jones	Fast and efficient; robust to illumination and occlusions	Limited adaptability to complex datasets; struggles with severe occlusions and non-standard face appearances
2016	Zohra et al. [74]	LBP for feature extraction with SVM	EURECOM Kinect Face Dataset	Partial occlusion	Classification accuracy	Achieved 98.50% accuracy for occluded vs. non-occluded face detection	Effective in detecting and localizing occlusions in depth images	Struggles with low-quality depth data; limited in handling glasses-induced occlusion
2016	Yang et al. [75]	SVM-based FP classifier with GIST, HoG, and other features	AFLW, FDDB, IJB-A	Various	Precision, recall, ROC	Improved precision from 67.60% to 71.75% (NPD-FDDB) and 55.36% to 61.26% (NPD-IJBA)	Reduced false positives significantly, improved precision of existing detectors	Slight reduction in recall, dependency on pre-trained detectors, computational cost for feature extraction
2019	Qezavati et al., 2019 [25]	Haar Cascade, LBPH, SVM with skin-tone histogram	Custom surveillance dataset	Partial (headscarf, poses)	Precision, recall	Improved precision over standalone Haar Cascade or LBPH methods; effective in detecting headscarf occlusions	Improved precision via skin-tone histogram; works on low-resolution videos	Limited adaptability to dynamic occlusions; lower accuracy for side-view faces

4.2 Advanced Deep Learning-Based Methods

This section reviews advanced deep learning methods for face detection under occlusion. It focuses on approaches that use CNNs and other deep learning architectures. Unlike traditional methods that depend on handcrafted features, deep learning techniques automatically learn patterns and representations from data, which makes them more flexible in handling complex occlusions, lighting variations, and pose changes. These methods have demonstrated strong performance, particularly in detecting occluded faces under unconstrained conditions. Many of the reviewed methods leverage pretrained networks like VGG16, ResNet, and YOLO, while others introduce custom architectures optimized specifically for occlusion scenarios. Several approaches incorporate enhancements such as attention mechanisms, multi-task learning, multi-scale learning, and context-aware processing to further improve accuracy and generalization across datasets. In addition to these enhancements, recent works have introduced Transformer-based and GAN-based models as distinct subcategories. These emerging directions expand the capabilities of deep learning methods in handling complex occlusion scenarios. To highlight key advancements, the studies in this section are categorized based on their architectures and methodologies as presented in Table 6. Meanwhile, Tables 7 and 8 present a detailed summary and comparison of reviewed advanced deep learning based studies.

Table 6: Summary of subcategories and their focus in advanced deep learning approaches

Category	Primary focus	Key features
Attention mechanism-based approaches	Prioritizes visible facial regions and suppresses irrelevant areas using attention modules.	Enhances feature representation and localization using modules like CBAM, SENet, and SEAM for better robustness against occlusions and noise.
Multi-task learning approaches	Handles multiple tasks, such as face detection and occlusion classification, simultaneously to improve performance.	Uses shared learning across tasks, making them effective for detecting and classifying occlusions or mask compliance.
Multi-scale learning approaches	Focuses on detecting faces at varying scales and resolutions to handle small and large faces effectively.	Uses feature pyramids, scale-specific detectors, or multi-branch networks to process different scales and occlusion levels.
Single-stage detection approaches	Performs detection in a single forward pass for faster processing without requiring multiple processing stages.	Prioritizes real-time performance while integrating modules like context refinement and receptive field enhancements to boost accuracy.
Multi-stage approaches	Divides detection into multiple steps, such as region proposals and refinement, for higher accuracy.	Employs separate modules for proposals, classification, and refinement, often achieving better accuracy at the expense of higher computation.

(Continued)

Table 6 (continued)

Category	Primary focus	Key features
Context-aware approaches	Utilizes surrounding contextual information, such as head pose, body orientation, or spatial relationships, to assist detection.	Complements direct facial feature extraction by integrating spatial reasoning or contextual labeling to handle heavily occluded faces.
Other studies	Covers methods that propose unique loss functions, optimization techniques, or custom architectures that do not fit into the other categories.	Introduces novel algorithms, loss designs, and experimental approaches that contribute to enhancing occluded face detection.

4.2.1 Attention Mechanism-Based Approaches

Wang et al. [21] introduced the Face Attention Network (FAN), a single-stage face detector designed to handle occlusion challenges in face detection. FAN used an anchor-level attention mechanism to enhance features from facial regions while reducing focus on irrelevant areas, thus minimizing false positives. The model was built on the RetinaNet architecture and included a Feature Pyramid Network (FPN) to preserve both spatial resolution and semantic information which enables it to detect faces at different scales. Extensive data augmentation, including random cropping, was applied during training to simulate occluded faces. For evaluation, FAN was applied to WiderFace and MAFA datasets. It got 88.8% Average Precision (AP) on the hard subset of WiderFace, and 88.3% mean Average Precision (mAP) on MAFA, outperforming methods like Locally Linear Embedding Convolutional Neural Networks (LLE-CNNs) and Adversarial Occlusion-aware Face Detection (AOFD). The results showed that FAN effectively detects occluded faces while maintaining computational efficiency. However, its reliance on anchor-based methods could limit performance with complex occlusions in real-world scenarios. In 2022, Zhang et al. [20] proposed an improved RetinaNet model for detecting occluded faces by incorporating a Universal and Recognition-friendly Image Enhancement (URIE) module as a pre-network, along with an attention mechanism. The URIE network enhanced input images by highlighting visible facial regions and reducing distortions. The attention mechanism, meanwhile, boosted important features while keeping contextual information. The model was tested on the WiderFace and MAFA datasets. On MAFA, it attained a mAP of 89.7% surpassing techniques including FAN and LLE-CNNs. Strong performance was also shown across weak (84.5%), medium (75.8%), and moderate occlusion (26.1%). Although the technique maintained computing efficiency and handled occlusions well, problems with heavily occluded faces and large-scale variances were recognized as issues for further development. In another study, Qi et al. [76] suggested a modified form of YOLOv5 to identify mask-occluded faces in real-time applications. To highlight important facial features while reducing attention on irrelevant areas, the improvements included installing a Convolutional Block Attention Module (CBAM) to the backbone and neck of YOLOv5s. Moreover, focal Loss took the role of the binary cross-entropy loss function to solve sample imbalance and enhance identification in challenging scenarios. The model was investigated on the WIDER Face and AIZOO datasets. On Wider Face, it obtained a mAP50 of 95.9% and an F1-score of 92.8%; on AIZOO, it obtained a mAP50 of 96.5% and an F1-score of 94.3% surpassing the baseline YOLOv5s and other advanced approaches. The model also displayed enhanced sensitivity to finely occluded faces at small scales. Though it had advantages, the technique suffered with severe occlusions and needed further work on environmental conditions and changeable lighting.

Wang et al. [77] also presented EfficientFace, another attention-based model. EfficientFace is a lightweight deep-learning framework designed with a focus on occlusion handling, imbalanced aspect ratios, and feature representation difficulties. Three improvements were introduced: a Receptive Field Enhancement (RFE) module to handle facial aspect ratio variances, a Symmetrically Bi-directional Feature Pyramid Network (SBiFPN) to improve spatial accuracy and feature fusion, and an Attention Mechanism (AM) to concentrate on important areas for identifying occluded faces. Using only one-sixth of the computational resources compared to models like Dual Shot Face Detector (DSFD) and MogFace, EfficientFace achieved mAP scores of 95.1% (Easy), 94.0% (Medium), and 90.1% (Hard) evaluated on datasets including AFW, Pascal Face, FDDB, and WIDER Face. It also outperformed many state-of- the-art detectors. It highlighted areas for future development even with its great performance since it struggled with highly obscured faces and dense clusters in crowded settings.

Zhao et al. [78] proposed an attention-enhanced YOLOv4 framework to improve face mask detection under challenging conditions, such as occlusions and lighting variations. The approach incorporated three attention mechanisms: Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation Networks (SENet), and Coordinate Attention Networks (CANet) into the feature fusion and detection layers. The best version, YOLOv4-CBAM-A, integrated CBAM modules at key points in the network, achieved a 93.56% mAP on the MAFA and WIDER FACE datasets with a 4.66% improvement over the baseline YOLOv4. The CBAM modules enhanced feature extraction by focusing on relevant regions while suppressing irrelevant features, particularly for small or occluded faces. While the model showed higher accuracy, it faced limitations in real-time processing speed due to the computational overhead introduced by the attention mechanisms. To improve face detection and recognition under occlusions, Yuan [3] introduced a visual attention guided model. The model used a visual attention mechanism for concentrating on the visible facial components and excluding the cluttered background. It employed a feature extraction network of ResNet50 size with a spatial and channel attention module for enhanced feature representation. The model treated face detection as a high-level semantic feature detection task and used activation maps for localizing the face and its scale. It was evaluated on datasets including LFW, CMUFD, and UCFI and was found to be more accurate and efficient than the existing methods. In the case of severely occluded faces, it achieved a detection rate of 59.78%, which is better than several deep learning-based methods. Although the model has some good properties, it had a higher computational complexity of the attention modules that increased the training time.

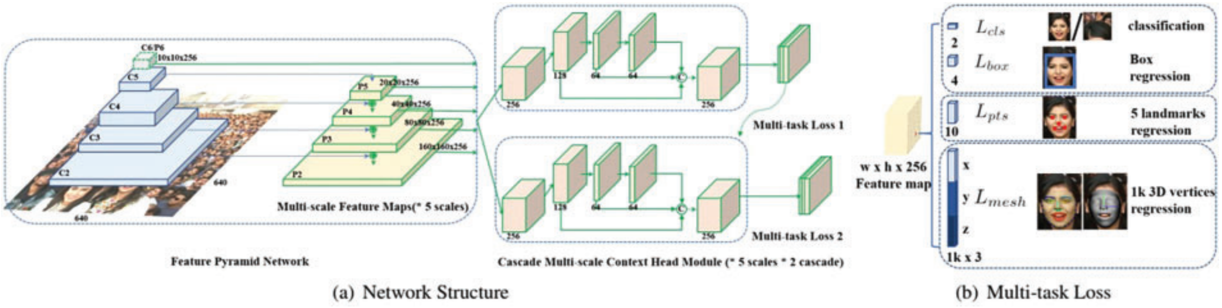
4.2.2 Multi-Task Learning Approaches

Xia et al. (2016) [17] presented an end-to-end framework for facial occlusion detection to enhance ATM security systems. The framework followed a coarse-to-fine strategy by employing two CNN models: one for head detection from upper-body images and another for categorizing occlusions in face parts (e.g., eyes, nose, and mouth). EdgeBoxes was used to create candidate areas; CNNs then were used for feature extraction and classification. Training on a bespoke face occlusion dataset, it was tested on extensively used datasets like LFW and AR. On the custom dataset, the framework attained high detection accuracy of 85.61%; on AR, 97.58%; and 100% for head detection with IoU > 0.5. It reported 94.55%, 98.58%, and 95.41%, respectively, for occlusion classification. However, the method struggled with illumination variations, complicated occlusion patterns, and head variability in real-world circumstances even if it was strong against many degrees of occlusion.

Ge et al. [13] proposed LLE-CNNs (Locally Linear Embedding CNNs), a deep learning framework for detecting masked faces. The authors addressed challenges like the lack of datasets and loss of facial cues caused by occlusions by introducing the MAFA dataset, which includes 30,811 images and 35,806 masked faces with varying orientations, occlusion levels, and mask types. The framework combined three main components. A proposal module first identified candidate facial regions using pre-trained CNNs. Next, an embedding module applied Locally Linear Embedding (LLE) to reconstruct missing facial cues and reduce occlusion noise. Finally, a verification module performed classification and regression to refine detections. Tested on the MAFA dataset, LLE-CNNs achieved an average precision (AP) of 76.4%, outperforming methods like Multi-task Cascaded Convolutional Neural Network (MTCNN) and Speeded Up Robust Features (SURF) Cascade by 15.6% or more. However, it struggled with extreme occlusions and side poses, achieving only 22.5% and 17.2% AP, respectively. The study demonstrated the value of combining feature refinement and contextual reasoning for occluded face detection and established the groundwork for future improvements in masked face detection.

Combining adversarial learning with segmentation, Chen et al. [79] presented a multi-task framework called AOFD to detect faces under significant occlusion. During training, the adversarial masking technique was applied to create occluded face features and push the detector to concentrate on visible facial areas. It also incorporated a segmentation branch to forecast blocked spots, which treated them as supplemental information rather than obstacles therefore enhancing feature extraction and detection accuracy. Outperformance of the model over FAN and LLE-CNNs was achieved with an 81.3% AP on MAFA. On FDDB at 1000 false positives, it attained a 97.89% recall rate, proving its robustness in identifying partially and highly occluded faces, even with low evident landmarks. Nevertheless, the approach depended on a limited manually labeled segmentation dataset (SFS), which restricts its scalability, and needed expensive CPU resources because of its segmentation technique. AOFD demonstrated in spite of these constraints the efficiency of adversarial learning and occlusion segmentation in enhancing occluded face detection.

In 2020, Deng et al. proposed RetinaFace [80], a single-stage, multi-task face detection framework that jointly predicts face bounding boxes, 2D landmarks, and 3D facial vertices. The overall architecture and multi-task design of RetinaFace are shown in Fig. 8. It introduces a unified regression target and enhances training with additional manual and semi-automatic annotations on datasets like WIDER FACE, AFLW, and FDDB. By combining these tasks into one inference process, RetinaFace improves detection robustness under occlusion, pose variations, and scale changes, while maintaining high efficiency.



(a) Network Structure (b) Multi-task Loss

Figure 8: Architecture of the RetinaFace model, including the feature pyramid network (left), cascade context head module (middle), and multi-task loss design (right), as proposed by Deng et al. [80]

Using a two-stage pipeline comprising face detection and classification to evaluate appropriate mask usage, Batagelj et al. [22] explored face-mask detection for COVID-19 compliance. To enable their investigations, the writers presented the Face-Mask Label Dataset (FMLD), constructed from the MAFA and Wider Face datasets. While the classification stage examined whether masks were worn correctly or incorrectly using CNN models, the detection stage concentrated on face identification and evaluation of mask effects on detection performance. Modern detectors struggled with masked faces, according to results; performance dropped by approximately 15% when compared to unmasked faces. RetinaFace emerged as the most robust detector, while ResNet-152 achieved the highest classification accuracy (over 98%) for identifying compliant and non-compliant mask placements. The combined pipeline achieved mAP values above 90%, outperforming baseline methods. Despite its effectiveness, the study faced limitations, including coarse dataset annotations that failed to capture varying occlusion levels, reliance on pre-trained models instead of custom architectures, and computational complexity, making real-time application challenging.

4.2.3 Multi-Scale Learning Approaches

To address the challenges of detecting faces with large variations in scale, pose, and occlusion, DSFD (Dual Shot Face Detector) was proposed by Li et al. [81] as an extension of the single-shot detectors (SSD) architecture. They introduced a Feature Enhance Module (FEM) and Progressive Anchor Loss (PAL) to improve multi-scale feature learning. An Improved Anchor Matching (IAM) method is also used for better training. Experiments on WIDER FACE and FDDB demonstrated that DSFD outperforms earlier methods like PyramidBox, especially under occlusion.

Jiang et al. [82] introduced 4AC-YOLOv5, an improved version of the YOLOv5 framework designed to detect small and occluded faces. The model featured three enhancements: a small target detection layer to capture low-level features for better detection of small faces, an Adaptive Feature Pyramid Network (AFPN) to dynamically adjust feature importance during multi-scale fusion, and a multi-scale residual module (C3_MultiRes) to improve multi-scale learning while maintaining efficiency. Tested on the WIDER Face and FDDB datasets, the model outperformed YOLOv5 and other methods, achieving mAP scores of 94.54% (Easy), 93.08% (Medium), and 84.98% (Hard) on WIDER Face, and a TPR of 0.99 at 1000 false positives on FDDB. While effective for occluded and small-scale faces, the model struggled with heavily occluded faces in dense scenes and required a balance between efficiency and accuracy. Jin et al. [83] proposed FSG-FD (Feature-Selective Generation for Face Detection), a deep learning model designed to detect occluded faces by combining multi-scale feature extraction and contextual information. The model introduced SG-net, a specialized feature-enhancement module that focuses on unoccluded regions while suppressing noise from occlusions. It then merges these enhanced features with high-level convolution outputs for classification and regression tasks. The model was evaluated on the WIDER Face dataset and a self-labeled surveillance dataset, achieving an average precision (AP) of 77.6% on WIDER Face, outperforming Faster R-CNN and other models. In real-world surveillance videos, it achieved a precision of 85.1% and a recall of 76.7%, demonstrating practical applicability. While the model showed effectiveness in multi-scale feature fusion and occlusion handling, its reliance on predefined feature generation limits adaptability to extreme occlusions and highly cluttered backgrounds.

Hu and Ramanan [84] proposed a face detection framework designed to handle small faces in complex environments. The method combined multi-scale representations, contextual reasoning, and a foveal descriptor that captured both local high-resolution features and global low-resolution context to improve detection accuracy. The framework used a multi-task model with scale-specific detectors trained on a coarse image pyramid, which enables it to detect faces across different scales. It leveraged large receptive fields to incorporate contextual information, which enhances performance for tiny faces as well as larger faces. Evaluated on the WIDER FACE and FDDB datasets, the method achieved state-of-the-art performance, with an AP of 81% on the WIDER FACE "hard" subset, outperforming earlier methods (29%–64% AP). It also showed robustness with respect to the scale, pose, and environmental conditions. Although the approach was effective, it had some drawbacks in terms of computational complexity and performance in crowded scenes, which point to further improvements for real-time applications. Tang et al. [85] proposed PyramidBox, a framework for detecting small, blurred, and partially occluded faces in complex scenarios. It enhanced the feature context by using pyramid anchors to include contextual information such as head and body contexts, which did not need any extra labels. The model incorporated Low-Level Feature Pyramid Networks (LFPN) to fuse the spatial detail information at different scales as coarse-level features and the semantic information as fine-level features to enhance the detection performance, especially for small faces. It also incorporated a Context Sensitive Prediction Module (CPM) to enhance the localization and classification performance. To enhance the training diversity and robustness, the authors proposed Data Anchor Sampling that involved face samples resizing and reshaping. The framework got the best results on the WIDER FACE and FDDB datasets, with mAP of 96.1% (easy), 95.0% (medium), and 88.9% (hard). However, there were some drawbacks to the method because it was costly and used semi-supervised anchor labeling.

To overcome the challenges of scale variation, occlusion, and imbalanced samples in training data, Yu et al. suggested an improved face detection algorithm based on YOLOv5 [86]. In this study, the model made several changes to boost the accuracy and robustness. The framework was also incorporated with a Receptive Field Enhancement (RFE) module that employed multi-branch dilated convolutions to deal with the multi-scale detection problem. To this end, it employed a Separated and Enhancement Attention Module (SEAM) to highlight the features in the occluded regions and a Repulsion Loss function to prevent the overlapping bounding boxes from affecting the detection performance in the case of occlusion. To address the sample imbalance, a Slide Loss was used to learn to dynamically rank hard samples; then, Normalized Wasserstein Distance (NWD) Loss was incorporated to improve the detection of small faces. The effectiveness of the model was validated by the experiments on the WIDER FACE dataset, and the mAP values of 98.7% (easy), 97.2% (medium), and 87.7% (hard) were achieved. However, the model proposed in this paper relied on anchor-based designs and had high computational costs, which may limit its applicability in environments with scarce resources.

Garg et al. [87] proposed a single-stage deep CNN for detecting partially occluded faces in video sequences. The approach used multi-scale anchor boxes to help capture the shapes and sizes of the facial regions. The approach reduced the computational costs by restricting the number of scales and the number of anchor boxes without sacrificing the accuracy. The network was designed to improve the detection performance in the occluded regions by using five max pooling layers for feature extraction and 22 convolutional layers for anchor-based identification of partially occluded faces. In this paper, the researchers have suggested a different Intersection-IoU threshold of 0.4 to eliminate the bounding boxes that are not relevant. The model was evaluated on the FDDB dataset and the results show that the model has an accuracy of 94.8%, precision of 98.7%, and F1-score of 98.25% at the frame rate of 21 fps. Its anchor-based approach, however, may limit the generality to non-standard face shapes or other datasets.

To enhance the accuracy and recall of face detection especially under occlusion, blurring, and at small scales, Mamieva et al. [88] presented a face detection method based on deep learning using the RetinaNet framework. The model design had a two-part architecture that consisted of a region-offering network (RON) to propose potential facial regions and a prediction network to further refine and classify these regions. To this end, the method adopted multi-scale features for robustness by employing a high and low feature generation pyramid that improves the ability to detect faces at different scales. The model was trained on the WIDER FACE dataset and fine-tuned on FDDB. It has an AP of 41.0 (single-scale) and 44.2 (multi-scale) and a detection accuracy of 95.6%. Though the method is powerful, it has a high computational cost, especially in multi-scale inference, which may be unfeasible in resource-constrained environments.

To improve the regression and classification performance in face detection from challenging poses and small sizes, Zhang et al. proposed a single-shot face detector [89]. Introducing five specialized modules; Selective Two-step Regression (STR) and Selective Two-step Classification (STC) to enhance bounding box localization and recall efficiency, Scale-aware Margin Loss (SML) to improve the scale, Feature Supervision Module (FSM) to improve feature alignment, and Receptive Field Enhancement (RFE) to increase the context for detection of faces at different scales. On WIDER FACE, MAFA, FDDB, AFW, and PASCAL Face datasets, the model achieved state-of-the-art results. At Video Graphics Array (VGA) resolution, using ResNet-18 as backbone, the detection rate and speed were very good with frame rate of 37.3 FPS. The model had some limitations, like reliance on anchor boxes and relatively high computational complexity due to extra modules, but it performed strongly.

In another interesting study, Tsai et al. [90] proposed a system based on SSH and feature extraction using VGG16 to detect and recognize partially-occluded faces. The SSH network had three detection branches, M1, M2, and M3, to detect small, medium, and large faces, respectively. It also incorporated a context module to improve feature maps by increasing the receptive field and feature pyramids for improved detection accuracy at low computational cost. The system was tested on the WIDER FACE and MS1M-ArcFace datasets and had very good accuracy for different levels of occlusion. However, it relied on VGG16, which might fail in extreme occlusion cases. Nevertheless, the method could process frames in real time with high precision for partially occluded faces.

4.2.4 Single-Stage Detection Approaches

Najibi et al. [91] proposed the SSH face detector, a single-stage, completely convolutional network for effective and accurate face detection. SSH was better than two-stage methods that rely on region proposals and classification because it performed classification and regression in a single pass, which reduces computational overhead while maintaining high precision. The model was able to work on all sizes and could find small, medium, and large faces by using many convolutional layers with different steps. It had a context module that, without an image pyramid, increased the receptive field to effectively capture contextual information. To increase robustness during training, SSH also employed online hard example mining (OHEM). Demonstration of modern performance on WIDER FACE, FDDB, and Pascal-Faces datasets. On Wider Face, it had mAP rates of 91.9% (easy), 90.7% (mid), and 81.4% (hard). It also increased mAP by 4% when coupled with an input pyramid. The model was also quite efficient, processing images GPU at 50 frames per second. However, there were some disadvantages: SSH was dependent on pre-trained backbones like VGG-16, which limited the ability to more recent architectures.

In 2020, Alashbi et al. [45] proposed the Niqab-Face-Detection model which is a deep learning framework for detecting mostly niqab-covered, highly occluded faces. The proposed framework, namely MobileNet-SSD, combines MobileNet for effective feature extraction and Single Shot Multiboxin Detector (SSD) for real-time detection. The approach of context-based labeling which pays attention to the context around the face rather than just the part of the face that is actually visible improves the detection accuracy. To improve performance, especially in challenging conditions, focus loss, and hard sample mining were employed. It was evident from the evaluation findings that current models including MTCNN, and MobileNet were outperformed by the proposed model with a precision of 99.6% and recall of 59.9%. The model however had a poor recall rate attributed to the small dataset and the high level of occlusion, which suggested the need for more data and optimization.

To improve the detection of partially-occluded faces, Zhao et al. [92] proposed an enhanced YOLOv5 framework. The method aimed at increasing the detection accuracy through changes in the loss function to replace the standard one with Distance Intersection over Union (DIoU) for faster convergence and better localization. In addition, it also applied data augmentation strategies such as flipping, scaling and brightness changes, and label smoothing to improve robustness. The model was trained on a large dataset which was collected from the MAFA dataset and web-sourced images, and it had six classes of occlusions:

masks, collars, hands, scarves, objects, and no occlusions. The enhanced YOLOv5 had an accuracy of 70.3%, which is better than the initial YOLOv5 accuracy of 64.1%. Nonetheless, the model had some difficulties with severe occlusion and irregular objects which are the areas for further optimization in terms of datasets and sophisticated loss functions. Nadhum et al. [93] proposed Ghost-YOLOv5, an improved version of the YOLOv5 deep learning algorithm for real-time detection of faces with and without masks. The model enhanced efficiency and effectiveness through the application of Ghost Convolution instead of the conventional convolution to enhance computation time and performance. A self-collected dataset of 219 images of masked and unmasked faces was used to train and test the model. The model achieved a mean Average Precision (mAP) of 96.6% which is higher than the baseline YOLOv5 model that achieved 89.11%. The model also has a fast inference time which makes it suitable for real time applications. But the study has some limitations like inability to detect faces with uncommon masks and in low light conditions, and the authors thus recommended dataset enhancement and architectural improvement for future work as well.

Kurniawan et al. assessed the performance of the YOLOv5 model for detecting masked and unmasked faces across different image resolutions [94]. For this purpose, the study employed three datasets: the M dataset which has real-world masked faces, the S dataset which has synthetic faces with masks, and the G dataset which is a combination of the M and S datasets. The performances were evaluated at 320 pixels and 640 pixels to determine the costs and benefits of increasing the size of the input image during training. The results indicated that the image resolution (pixels) affected the accuracy of the detection and that high resolution (640 pixels) provided better results at the expense of increased training time. In addition to achieving detection rates of 99.2%, 98.9%, and 98.5% on the G, M, and S datasets, respectively, the model also had some constraints in detecting small objects and in poor illumination. The study concluded that to enhance the detection accuracy, it is essential to employ an appropriate dataset and optimal image resolution. YOLOv5 is suggested for application in public health surveillance.

4.2.5 Multi-Stage Approaches

Li et al. [95] proposed a face detection model for handling the occlusions with the help of a double-channel network architecture. The framework entailed an occlusion perceptron network that learned the features from unoccluded regions and a residual network to learn the features from the entire face for a more complete representation. The output of both the networks were combined using a weighted scheme to improve the feature learning of the occluded faces. In order to address the problems of data scarcity and overfitting, the model employed transfer learning to pre-train convolutional layers. It was evaluated on the AR dataset (sunglasses and scarves' occlusions) and the MAFA dataset (diverse occlusions). On the AR dataset the model achieved 99.46% accuracy for sunglasses and 99.73% accuracy for scarves and on the MAFA dataset it achieved 80.2% accuracy with a frame rate of 39 FPS. But there were some issues e.g., high computational costs in training and the need to manually tune parameters for occlusion thresholds because they were not learned from the data.

4.2.6 Other Studies

In their paper, Alafif et al. [96] presented LSDL, a CNN based method for face detection in unconstrained environments with partial occlusions and pose variations, using a single CNN trained on a large scale dataset of occluded, posed and illuminated faces. First, it employs a sliding window approach for face localization and then uses a confidence score threshold and Non-Maximal Suppression (NMS) to further localize the detected faces. For training, the authors used four novel datasets (LSLF, LSLNF, CrowdFaces, and CrowdNonFaces) and used AFW and FDDB datasets for evaluation. The precision of AFW was 97.4% and it had a fairly good performance on FDDB. Nevertheless, LSDL was quite robust in detection but had some constraints such as slow inference time due to the sliding window and sensitivity to confidence thresholds that failed to detect in some instances.

Iqbal et al. [97] carried out a comparative analysis of the effectiveness of CNN-based face detection models for the detection of covered faces during the COVID-19 pandemic. The study established that most of the previous models that were trained on unmasked datasets were inefficient in detecting masked faces. The authors classified face detection models into two categories: Anchor-based vs. Anchor free and Single Stage vs. Two Stage architectures. Several models, RetinaFace, CenterFace, FaceBoxes, Extremely Tiny Face Detector (EXTD), TinyFaces, Light and Fast Face Detector (LFFD), and Multitask Cascaded Convolutional Networks (MTCCN) were evaluated for a modified WIDER Face dataset, in which the lower halves of the faces were erased to mimic masks. Of the evaluated models, it was observed that RetinaFace had the best performance at the Easy (84%) and Medium (80%) levels, while EXTD was the best at the Hard level (59%). The study also revealed that while the present models are appropriate for easy sets of data, their efficiency decreases when they are applied to difficult conditions, e.g., small, partially occluded, or complex images. Consequently, there is a requirement for new specific masked face datasets and models for occlusion scenarios.

To improve face detection under partial occlusion, Opitz et al. [98] proposed a grid loss function for CNNs. The method cuts the last convolutional layer's feature map into spatial blocks and uses a hinge loss for each block, to ensure that even partially visible regions remain discriminative. The approach was tested on FDDB, AFW, and PASCAL Faces datasets. It achieved a TPR of 86.7% at 0.1 FPPI on FDDB, outperforming traditional CNNs. Also reduced overfitting, improved mid-level feature learning, and was efficient with smaller datasets. The model supported real-time detection, processing at 20 frames per second. However, the method had some difficulties with extreme occlusions and cluttered backgrounds.

Table 7: Summary of reviewed deep learning approaches for occluded face detection (Part 1)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2016	Opitz et al. [98]	Grid loss: Dividing CNN feature maps into blocks with independent hinge loss	FDDB, AFW, PASCAL Faces	Partial	TPR, FPPI	TPR: 86.7% at 0.1 FPPI on FDDB	Robust under occlusions; reduced overfitting; real-time performance	Challenges with extreme occlusions and cluttered backgrounds
2016	Xia et al. [17]	Two-stage CNN: Head detection and occlusion classification using EdgeBoxes and multi-task learning	Custom face occlusion dataset, AR Face, LFW	Partial, High	Accuracy (Head Detection, IoU > 0.5), Accuracy (Occlusion Classification)	85.61%–100% (Head Detection IoU > 0.5); 94.55%-98.58% (Occlusion Classification)	Robust against various occlusions; end-to-end design	Sensitive to illumination and complex textures in real-world scenarios
2017	Najibi et al. [91]	Single-stage, headless CNN with context module, OHEM	WIDER FACE, FDDB, Pascal-Faces	Partial, High	mAP (Easy, Medium, Hard)	Achieved 91.9%, 90.7%, 81.4% mAP on WIDER FACE subsets. FDDB: Improved precision-recall curve	High efficiency (50 FPS), state-of-the-art performance, scale-invariant	Limited to older pre-trained backbones like VGG-16; dependency on specific architectures

(Continued)

Table 7 (continued)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2017	Alafif et al. [96]	Single CNN model (LSDL) with sliding window and NMS; trained on a large-scale dataset	AFW, FDDB, LSLF, CrowdFaces	Partial	Precision, ROC Curve	97.4% precision on AFW; competitive results on FDDB	Robust to occlusions, poses; no hand-crafted features required	Slower inference due to sliding window; sensitive to confidence thresholds
2017	Hu and Ramanan [84]	Multi-task scale-specific detectors with contextual foveal descriptors	WIDER FACE, FDDB	Partial, High	AP, Recall-Precision	AP: 81% (WIDER FACE, "Hard"); Significant improvement over prior methods	Robust for small faces; effective use of context and multi-scale features	Computationally intensive; struggles in crowded scenes
2017	Ge et al. [13]	LLE-CNNs: Feature embedding via dictionaries of masked and normal faces	MAFA	Partial, High	AP	AP: 76.4% (MAFA); Significant improvement over 6 baselines	Effective for partial occlusion; innovative embedding module	Limited performance on extreme occlusions and side poses
2017	Wang et al. [21]	FAN: Single-stage detector with anchor-level attention and data augmentation	WiderFace, MAFA	Partial, High	AP, mAP	88.8% AP (WiderFace Hard), 88.3% mAP (MAFA)	Robust feature enhancement for occlusions	Relies on anchor-based design
2018	Tang et al. [85]	Pyramid Anchors, LFPN, CPM, Data Anchor Sampling	WIDER FACE, FDDB	Partial, High	mAP (Easy, Medium, Hard)	Achieved 96.1%, 95.0%, 88.9% mAP on WIDER FACE	Superior performance under occlusions and small faces; innovative use of contextual information	High computational cost; semi-supervised anchor labeling process
2018	Chen et al. [79]	AOFD: Multi-task model with adversarial masking and segmentation	MAFA, FDDB, SFS	Partial, High	AP, Recall	81.3% AP (MAFA), 97.88% recall (FDDB)	Robust occlusion handling; integrates adversarial learning and segmentation	Small segmentation dataset; computationally intensive
2020	Yuan [3]	Visual attention-guided model with ResNet50 and attention mechanisms	LFW, CMUFD, UCFI	Partial, severe occlusions	Accuracy, MR, FPS	Achieved 59.78% accuracy on severely occluded faces, better than other deep learning models.	Robust to occlusions; integrates semantic features; suppresses background interference	Sensitive to parameter tuning; higher computational complexity
2020	Alashbi et al. [37]	Context-aware labeling and training with Niqab-Face dataset	Niqab-Face	High	Accuracy, Precision, Recall	TinyFace: 46.5% accuracy, YOLOv3: 33.6% accuracy	Highlights importance of contextual labeling for occlusion	Poor generalization of current detectors on heavily occluded faces
2020	Jin et al. [83]	FSG-FD: Region generation with SG-net for feature enhancement	WIDER Face, Self-labeled monitoring	Partial, High	AP, Precision, Recall	AP: 77.6% (WIDER Face); Precision: 85.1%, Recall: 76.7% (real-world dataset)	Robust feature enhancement for occluded regions; practical applicability	Limited adaptability to extreme occlusions and cluttered backgrounds

(Continued)

Table 7 (continued)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2021	Zhang et al. [89]	Single-shot detector enhanced with STR, STC, SML, FSM, and RFE	WIDER FACE, MAFA, FDDB, AFW, PASCAL Face	Extreme poses, tiny faces	Precision, Recall	State-of-the-art performance on WIDER FACE (Easy: 97.2%, Medium: 96.2%, Hard: 92.0%) and other datasets with significant improvements in AP scores	Effective regression and classification modules; robust against extreme poses and occlusions; achieves state-of-the-art results	Dependency on anchor boxes; relatively high computational cost due to added modules
2021	Zhao et al. [92]	Enhanced YOLOv5 with DIoU and data augmentation	MAFA, custom web-based dataset	Partial	Accuracy, Recall, Precision	Improved accuracy from 64.1% to 70.3%; Faster convergence with DIoU	Robust against diverse occlusions; faster training convergence	Limited performance on extreme occlusions and unconventional occluders

Table 8: Summary of reviewed deep learning approaches for occluded face detection (Part 2)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2021	Wang et al. [77]	EfficientFace: Lightweight framework with SBiFPN, RFE, and Attention Mechanism	AFW, Pascal Face, FDDB, WIDER Face	Partial, High	mAP, AP	mAP: 95.1% (Easy), 94.0% (Medium), 90.1% (Hard)	Real-time performance; robust against occlusion and unbalanced aspect ratios	Struggles in extreme occlusion and dense cluttered scenes
2021	Tsai et al. [90]	Integrated SSH network with VGG16 feature extraction; feature pyramids for scale invariance	WIDER FACE, MS1M-ArcFace	Partial	mAP, Precision-Recall	High precision; robust for small and medium faces	Real-time detection; effective scale invariance	Dependent on VGG16 for extreme occlusions, limited adaptability to extreme occlusions
2022	Alashbi et al. [45]	MobileNet-SSD with context-based labeling	Niqab-Face dataset	Heavy occlusion (niqabs, masks)	Precision, Recall	Precision: 99.6%, Recall: 59.9%; outperformed MTCNN and Mobilenet on heavily occluded datasets	Efficient real-time detection; superior precision for occluded faces	Limited training dataset; lower recall due to extreme occlusion
2022	Garg et al. [86]	Multi-scale anchor boxes with IoU adjustment	FDDB	Partial	Accuracy, Precision, F1 Score	94.8% accuracy, 98.7% precision, 98.25 F1 score; 21 FPS	High precision and speed; robust occlusion detection; tailored anchor box strategy	Dependency on anchor-based design; limited evaluation on other datasets
2022	Zhang et al. [20]	Enhanced RetinaNet with attention mechanism and URIE pre-network	MAFA, WiderFace	Weak, Medium, Heavy	mAP, AP	mAP 89.7% (MAFA); AP 84.5% (Weak), 75.8% (Medium), 26.1% (Heavy)	Enhanced visible regions; robust occlusion handling	Limited performance for high occlusion; large-scale variations
2023	Mamieva et al. [88]	RetinaNet with RON and multi-scale feature pyramids	WIDER FACE, FDDB	partial	AP, Precision, Recall, FPS	AP: 41.0 (single-scale), 44.2 (multi-scale); Accuracy: 95.6%	High accuracy; robust to occlusion and small scales; competitive speed	High computational cost for multi-scale inference strategies
2023	Zhao et al. [78]	YOLO-v4 with CBAM, SENet, and CANet attention mechanisms	MAFA, WIDER FACE	Partial, High	mAP, FPS	mAP: 93.56% (YOLOv4-CBAM-A); 4.66% improvement over baseline	Enhanced feature extraction; robust against occlusions	Increased computational cost; reduced real-time performance
2023	Kurniawan et al. [94]	YOLO-v5 with dataset variation and resolution assessment	M dataset, S dataset, G dataset	Partial	Detection Rate, mAP	G dataset (640px): 99.2%; M dataset (640px): 98.9%; S dataset (640px): 98.5%	High accuracy with diverse datasets; resolution-aware detection	Struggles with small objects and extreme lighting conditions

(Continued)

Table 8 (continued)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2023	Nadhum [93]	Ghost-YOLOv5: Lightweight convolution for efficient detection	Custom dataset (219 images)	Partial	mAP	Improved mAP: 96.6% vs. baseline YOLOv5 (89.11%)	Lightweight; real-time performance; robust detection	Struggles with unconventional masks and low-light scenarios
2023	Iqbal et al. [97]	Comparative evaluation of state-of-the-art face detection models on masked faces (RetinaFace, EXTD, etc.)	WIDER Face (blacked-out masked dataset)	Partial, High	Accuracy	84% (Easy), 80% (Medium), 59% (Hard)	Highlighted performance variations across models	Limited by lack of diverse masked datasets
2024	Yu et al. [86]	YOLOv5-based detector with RFE, SEAM, Slide Loss, and NWD Loss	WIDER FACE	Partial, High	mAP (Easy, Medium, Hard)	Achieved 98.7%, 97.2%, 87.7% mAP on WIDER FACE subsets	Robust occlusion handling; effective multi-scale detection; state-of-the-art performance	High computational cost; dependency on anchor-based designs
2024	Jiang et al. [82]	4AC-YOLOv5: Improved YOLOv5 with small target detection layer, AFPN, and C3_MultiRes	WIDER Face, FDDB	Partial, High	mAP, TPR	mAP 94.54% (Easy), 93.08% (Medium), 84.98% (Hard); TPR 0.99 (FDDB)	Effective for small faces; robust feature fusion; reduced computational overhead	Struggles with heavy occlusion in dense scenes
2024	Qi et al. [76]	Enhanced YOLOv5 with CBAM and Focal Loss	WIDER Face, AIZOO	Partial, High	mAP50, F1	mAP50: 95.9% (WIDER Face), 96.5% (AIZOO); F1: 92.8%, 94.3%	Real-time performance; improved detection of occluded faces	Limited in extreme occlusion and varying lighting
2025	Alashbi et al. [18]	Darknet-53 with contextual features	Niqab-Face	High	Precision, Recall, F1, AP	Precision: 73.70%, Recall: 42.63%, AP: 50.34%	Effective in highly occluded scenarios	High false positives; limited generalizability to other occlusion types; environmental factors untested

4.2.7 Context-Aware Approaches

By using contextual information around occluded areas, Alashbi et al. [37] proposed a CNN-based method to identify highly occluded faces. The Niqab-Face dataset, which comprises 10,000 images with high levels of facial occlusion (i.e., faces covered by niqabs) was first introduced by the authors. This dataset was specifically annotated to enable CNN models to train on the visible facial parts and their surroundings to improve detection. The work was evaluated on the Niqab-Face dataset against MTCNN, MobileNet, TinyFace, and YOLOv3. Among the models, TinyFace gave the best accuracy of 46.5%, YOLOv3 followed with 33.6% accuracy while MTCNN and MobileNet had a low accuracy of 18% and 20%, respectively. These results showed that existing detectors have a challenge with extreme forms of occlusion. The authors argued that context-aware labeling is necessary to improve the detection but also that there is a need for better models that are specifically meant for highly occluded faces.

Recently, Alashbi et al. [18] proposed an Occlusion-Aware Face Detector (OFD) to localize covered faces with high levels of occlusion such as niqab covers. To enhance the feature learning process, the model utilized contextual information such as head pose and shoulders, and body aspects. The authors enhanced the Darknet-53 backbone architecture to include more layers to enhance the feature learning of occluded faces. The OFD model outperformed other models including YOLO-v3, Mobilenet-SSD, and TinyFace with a precision of 73.70%, recall of 42.63%, and F-measure of 54.02%. It also offered a 50.34% average precision (AP). However, the model had low generalization capacity to non-niqab occlusions and had high rates of false positives, particularly in complex settings. It also did not really solve the problems of cluttered environments and poor illumination. Furthermore, it did not solve challenges like small dataset size, imbalanced dataset, overfitting, computational expense, and inaccurate detections. These problems are likely to hinder the adoption of this model in real-world applications.

4.3 Transformer-Based Models for Occluded Face Detection

In recent years, transformer architectures [99] and generative adversarial networks (GANs) [79] have gained substantial attention in computer vision because of their impressive ability to capture long-range dependencies and generate realistic imagery. Notably, when addressing occluded face detection, these models effectively manage complex spatial relationships and reconstruct plausible facial structures in areas that are partially obscured, thereby enhancing overall detection performance. Transformer-based models, such as Vision Transformer (ViT) [100–102], Swin Transformer [103–105], and Detection Transformer (DETR) [106], have been widely explored for face-related tasks. For example, SwinFace [107] employs a Swin Transformer backbone to address various face analysis tasks such as face recognition, facial expression recognition, age estimation, and attribute prediction. Similarly, DETR introduces a fully end-to-end transformer-based framework for object detection, which has been adapted in recent studies to improve face detection performance, particularly in challenging scenarios involving occlusions. These approaches highlight the capability of transformer-based models to extract comprehensive facial features and spatial relationships, even under difficult real-world conditions. A summary of the Transformer-based models discussed in this subsection is presented in Table 9.

Several recent works have leveraged the Swin Transformer to improve face detection and recognition robustness. For instance, Mao et al. [108] utilized a Swin Transformer backbone to enhance masked face detection performance. Their model, optimized through hyperparameter tuning, demonstrated superior results over classical CNN models but required higher computational resources. The workflow of their proposed model is illustrated in Fig. 9. Building upon YOLOv5, Yuan et al. [109] integrated Swin Transformer layers within a customized detection head, resulting in the DSH-YOLOv5 model, which achieved strong performance on WIDER FACE, FDDB, and PASCAL FACE datasets while maintaining practical efficiency. In another line of work, Zhou [110] designed YOLO-M, embedding Swin Transformer prediction heads within the detection framework to better address local occlusion challenges, achieving noticeable improvements on the WIDER FACE dataset. Furthermore, Zhao et al. [111] addressed masked face recognition by proposing the Masked Face Transformer (MFT). Their approach introduced Masked Face-compatible Attention (MFA) and a ClassFormer module to enlarge attention range and enhance intra-class consistency, outperforming prior methods on masked datasets.

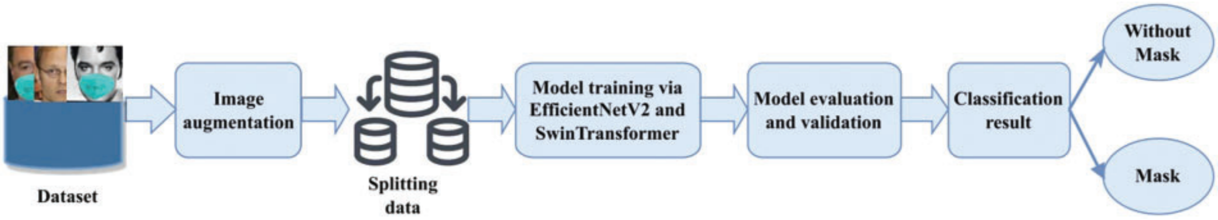

Dataset
Image augmentation
Splitting data
Model training via EfficientNetV2 and SwinTransformer
Model evaluation and validation
Classification result
Without Mask
Mask

Figure 9: Schematic diagram of the Swin Transformer-based mask detection model [108]

Parallel to Swin Transformer research, other studies have investigated Vision Transformer (ViT) backbones. Pandya et al. [112] explored the use of ViT for face mask classification, achieving 86% accuracy on a small custom dataset. Their analysis highlighted that smaller patch sizes preserved finer facial details, enhancing classification robustness. Despite the promising results, the study acknowledged that the limited dataset size posed challenges to generalization and scalability. In the context of facial expression recognition, Li et al. [113] developed the Mask Vision Transformer (MVT), introducing a mask generation network and a dynamic relabeling strategy to explicitly filter occluded or irrelevant regions, leading to improved robustness on RAF-DB, FERPlus, and AffectNet datasets. However, MVT's reliance on masking and relabeling strategies may limit its direct applicability to domains beyond expression recognition. In another application, Lee et al. [114] proposed Latent-OFER, a ViT-driven method for occluded Facial Expression Recognition (FER). By detecting and reconstructing occluded regions and extracting latent features via ViT and CNN hybrids, Latent-OFER achieved state-of-the-art results on occluded FER benchmarks. However, the authors noted that while the method showed strong performance, scalability across highly diverse datasets might require multi-dataset training strategies.

Detection Transformer (DETR) and its variants have also been adapted to handle occlusion challenges. Al-Sarrar and Al-Baity [115] combined a DETR face detector with an AlexNet-based mask classifier. Extensive experimental evaluations demonstrated that the proposed hybrid model surpassed previous CNN-based approaches. However, the model's execution speed, while acceptable for real-time applications, remains slower than lightweight CNN-only models. Beyond application, Zhao et al. [116] systematically analyzed DETR's behavior under occlusions and adversarial attacks. Their findings revealed DETR's strong resilience to moderate occlusion but exposed performance degradation under severe occlusion and heavy corruption due to a "main query" imbalance in attention. For real-time detection, Li et al. [117] developed DDR-DETR by optimizing RT-DETR with modules such as StarNet and CGRLFPN. Their model achieved improved mAP50-95 in classroom settings, offering an efficient solution for detecting faces under blur and occlusion.

Some studies specifically targeted occlusion-handling architectures beyond standard vision backbones. Chiang et al. [118] introduced ORFormer, an occlusion-robust transformer for facial landmark detection. By employing messenger tokens and dissimilarity evaluation, ORFormer selectively recovers non-occluded features, achieving strong performance on WFLW and COFW datasets. However, the authors noted that while the method showed strong performance, scalability across highly diverse datasets might require multi-dataset training strategies.

Hybrid architectures combining CNNs with Transformers have also shown promise. Zhang et al. [119] introduced E-CT Face, a lightweight face detector that fused CNNs for local detail preservation with ViT blocks for capturing global context. Their model maintained competitive performance on WIDER FACE and FDDB while using significantly fewer parameters than traditional heavyweight detectors. However, the method still fell slightly behind heavyweight detectors on extremely challenging conditions and complex datasets. Similarly, the Latent-OFER model by Lee et al. [114], although focused on facial expression recognition, employed a hybrid CNN-ViT design to handle occlusion recovery and feature extraction effectively.

Overall, the Transformer-based models show great potential in solving face detection and recognition problems under occlusion through their Swin Transformers, Vision Transformers, and Detection Transformer variants. These models surpass traditional CNN-based approaches because they use long-range dependencies to improve feature representation, which leads to better robustness and accuracy. The recent development of ORFormer and Latent-OFER demonstrates the increasing interest in designing systems that recover valuable information from partially hidden faces. Despite their promising performance, challenges such as computational complexity, generalization to various types of occlusion, and real-time applicability remain active areas for further research.

Table 9: Summary of reviewed transformer-based methods for occluded face detection

Year	Study	Methodology	Dataset	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2024	[108]	Swin Transformer for face mask detection with model tuning	RMFRD, SMFRD, Moxa3K	Partial occlusion (mask)	Accuracy, Precision, Recall, Specificity, F1-score, Kappa, MCC	Swin Transformer outperforms baselines	Superior feature extraction, handles occlusion better	Higher computational complexity compared to lightweight models
2024	[118]	ORFormer: Transformer with messenger tokens for occlusion detection and feature recovery in FLD	WFLW, COFW, 300W	Partial occlusion (occluded facial landmarks)	NME, FR, AUC	Outperformed baselines under occlusions	Explicit occlusion detection and feature recovery; improved landmark detection in challenging conditions	Heavy reliance on a well-trained quantized heatmap generator
2021	[113]	MVT: Pure transformer-based FER with mask generation and dynamic relabeling	RAF-DB, FERPlus, AffectNet-7/8, Occlusion-RAF-DB, Pose-RAF-DB	General occlusion (back-grounds, masks, pose)	Accuracy	Achieved 88.62% (RAF-DB), 89.22% (FERPlus), 64.57% (AffectNet-7);	Explicitly filters occlusions and background; robust to real-world FER challenges	May be limited to expression recognition tasks due to task-specific masking
2024	[119]	E-CT Face: Bi-Stream CNN and Transformer hybrid backbone with feature enhancement and multiscale aggregation	WIDER FACE, FDDB	General occlusion (blur, pose variation, small faces, crowded scenes)	AP	95.30% (easy), 94.20% (medium), 87.56% (hard) on WIDER FACE; strong performance with only 3.8M parameters	Combines local and global features; lightweight and fast	Still slightly behind heavy-weight models on very difficult conditions
2023	[115]	Hybrid DETR + AlexNet model for face mask detection	AIZOO FMD + MMD	Partial occlusion (face masks)	AP, Execution Time	Achieved 89.4% AP and 2.8 s execution time; better than YOLOv2+ResNet50 and LLE-CNNs	Combines Transformer detection strength with CNN classification speed; robust against masked faces	Slower execution compared to pure CNN lightweight models

(Continued)

Table 9 (continued)

Year	Study	Methodology	Dataset	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2023	[116]	Robustness study of DETR on occluded, adversarial, and corrupted images	COCO128 (custom occlusion & corruption scenarios)	Random occlusion, salient occlusion	mAP, mAP50	DETR outperformed Faster-RCNN and YOLOv5 on moderate occlusions; showed resilience against sticker attacks	Superior performance on partial occlusion; good adversarial robustness	Weaker performance under heavy corruption; slow convergence due to dominant query phenomenon
2024	[117]	DDR-DETR: Lightweight real-time DETR variant for classroom face detection	Custom classroom dataset	General occlusion (blur, low-res faces, partial occlusions)	mAP50-95	Improved mAP50-95; efficient real-time detection	Optimized real-time Transformer detection under occlusion	not validated on broader public datasets
2023	[114]	Latent-OFER: ViT-SVDD occlusion detection + hybrid reconstruction + latent feature extraction	RAF-DB, AffectNet, FED-RO, Occlusion-RAF-DB, Occlusion-AffectNet	Random real-world occlusions	Accuracy	Outperformed SOTA methods on occluded FER benchmarks	Full occlusion handling pipeline; strong FER boost	May require multi-dataset training for broader scalability
2024	[109]	DSH-YOLOv5: YOLOv5 with Swin Transformer and attention modules	WIDER FACE, FDDB, PASCAL FACE	Pose, occlusion, extreme light conditions, masks	AP	Achieved SOTA on results; competitive speed	Integrates Transformer attention and strong feature enhancement; practical extensions (mask, gender)	May introduce complexity compared to standard YOLOv5
2023	[112]	ViT for face mask recognition with patch size analysis	Custom small dataset (<1,000 images)	Partial occlusion (face masks)	Accuracy	86% accuracy; showed finer patches improve detection	Introduced ViT to face mask recognition	Limited by small dataset; generalization challenges
2024	[110]	YOLO-M: YOLOv5-based occluded face detector with Swin Transformer Prediction Head (STPH) and I-PANet	WIDER FACE	Complex local occlusion (lighting, obstruction, pose)	AP	Improved face detection accuracy under occlusion	Enhanced multi-scale fusion and global context modeling	Only evaluated on WIDER FACE; generalization to other datasets not discussed

(Continued)

Table 9 (continued)

Year	Study	Methodology	Dataset	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2023	[111]	MFT: Masked Face Transformer with MFA and ClassFormer	Simulated and real masked face datasets	Mask occlusion	–	Outperformed SOTA masked FR methods	Enlarged attention range; intra-class enhancement	Designed mainly for masked face recognition

4.4 *GAN-Based Methods*

Although this review focuses mainly on direct face detection under occlusion, several GAN-based methods have also been proposed to indirectly support this task. GANs contribute by reconstructing missing facial regions, restoring occluded faces, and augmenting datasets to improve the robustness of detection models. This subsection briefly highlights notable GAN-based approaches relevant to occlusion handling and face detection.

Restoring missing or occluded parts of the face has been a major focus of early GAN-based methods. Authors in [120] proposed a Deep Convolutional GAN (DCGAN) that restores blocked regions in facial images by learning from both occluded and unoccluded samples. Their model showed promising restoration capabilities for images with up to 50% occlusion. In a related approach, Lee and Han [121] introduced a three-stage GAN framework where occluded parts are first recognized and separated, then removed, and finally reconstructed, using dual discriminators to refine the restored regions. Their model demonstrated competitive results measured by FID, SSIM, and PSNR on the CelebA and FFHQ datasets. Similarly, Nelson et al. [122] combined feature extraction with an SR-SSA optimized GAN for occluded face recognition. By integrating Search and Rescue Optimization into a GAN training pipeline, they achieved a notable 95.6% accuracy rate on occlusion-affected datasets.

Moving beyond simple restoration, some studies integrated occlusion handling with other face-related tasks. For example, Duan et al. [123] proposed TSGAN, a two-stage GAN architecture that simultaneously performs face de-occlusion and frontalization. Their model utilized an occlusion mask-guided attention mechanism and dual triplet losses to preserve identity features throughout the recovery process. Evaluations on both constrained and unconstrained datasets confirmed TSGAN's effectiveness in synthesizing frontal, occlusion-free face images. Given the computational demands of traditional GANs, lightweight alternatives have emerged. For example, a Lightweight DCGAN (LW-DCGAN) was proposed by Lv et al. [124]. The suggested model aimed to reconstruct partially occluded faces with fewer parameters and faster inference times to baance speed with visual quality. Another work by Zhou and Lu [125] introduced a Masked Face Restoration Model based on a lightweight GAN, specifically targeting the challenge of restoring masked or obstructed facial images with minimal computational overhead.

Face inpainting has also been explored through GANs to deal with occlusion. The FD-StackGAN model introduced by Jabbar et al. [126] focused on generating complete face images by stacking multiple GAN stages to progressively refine the missing areas. Meanwhile, in the "Look Through Masks" study [127], a GAN model was trained to de-occlude masked faces, removing occlusion artifacts and recovering underlying facial textures while preserving identity information. Furthermore, Man and Cho [128] proposed T-GANs, a novel face restoration framework combining a Transformer module with GANs to improve occluded face inpainting, as illustrated in Fig. 10. Rather than focusing solely on restoration, some methods leveraged GANs to improve training data diversity. Qiu et al. [129] introduced a novel approach called FROM, where dynamically learned occlusion masks are applied to deep features during training to clean corrupted representations. Additionally, other studies [130] explored using GANs for face dataset augmentation, generating a variety of occluded and clean faces to enhance detector robustness.

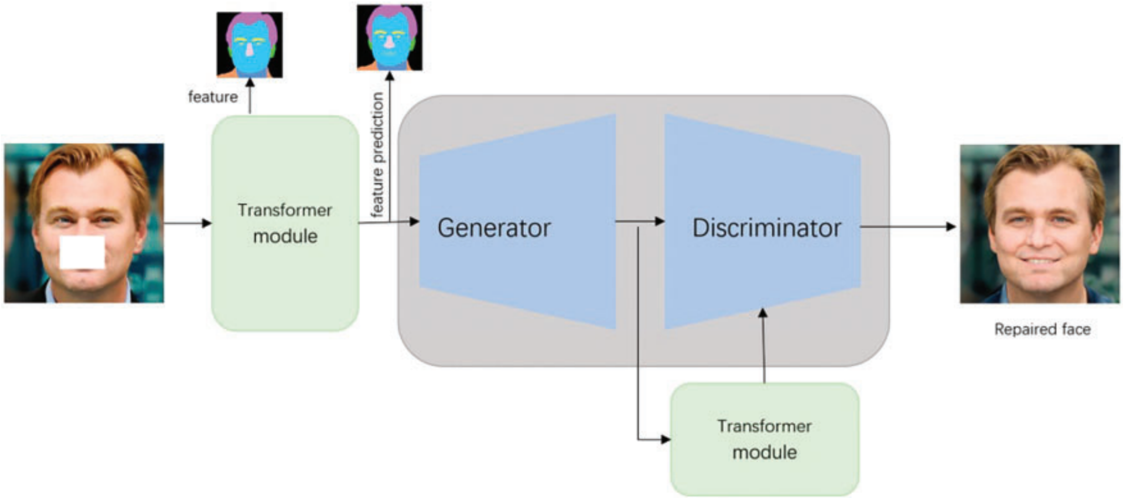


Figure 10: The architecture of the T-GANs framework proposed in [128], combining a Transformer module with GAN components for occluded face restoration

In general, GAN-based methods have substantially improved face detection and recognition under occlusion by providing solutions for restoration, frontalization, and data enhancement. Traditional GANs produced high-quality restorations, but recent works have focused on lightweight architectures and task-specific adaptations to better support real-world occluded face detection scenarios.

4.5 Hybrid Methods

This subsection discusses techniques that combine traditional feature-based approaches with modern deep learning techniques, leveraging the strengths of both methodologies to enhance robustness against occlusion. Some methods are combining traditional machine learning with deep learning, or using multiple techniques (for example, feature extraction + CNN). A detailed comparison of hybrid methods is presented in Table 10.

To improve the detection of faces in challenging conditions such as heavy occlusions, extreme poses, poor lighting, and low resolutions, Zhu et al. [131] proposed CMS-RCNN, a deep learning-based model. The model also learned multi-scale features and performed contextual reasoning by using both facial and body contexts for better detection results. It also incorporates a Multi-Scale Region Proposal Network (MS-RPN) which is responsible for producing likely face regions and a Contextual Multi-Scale CNN (CMS-CNN) for further processing of these regions, including facial and body context. The use of body features was also inspired by the concept of using body features to confirm the presence of faces or the absence of them in occluded or low-resolution images. The CMS-RCNN model was evaluated on WIDER Face and FDDB datasets. It achieved high accuracy and outperformed baseline models. On the WIDER Face dataset, it achieved AP of 90.2% (Easy), 87.4% (Medium), and 64.3% (Hard). It also had good recall rates on the FDDB dataset. However, it had some problems including the detection of faces in densely crowded scenes and the need for a large number of computational resources owing to the use of multiple feature streams and region proposals.

In a different hybrid approach, Zhang et al. proposed a face occlusion detection algorithm for ATM surveillance scenarios in [26]. The method addressed challenges such as restricted views, harsh lighting conditions, and severe occlusions due to masks, hats, or sunglasses. It combines techniques for head localization, tracking, and occlusion verification. The head localization was performed using the Omega shape which is the shape formed by the head and shoulders and the potential energy function was used to model it to detect heads even when the facial features are fully occluded. The detection framework was integrated with gradient and shape cues into a Bayesian tracking algorithm to improve computational efficiency. The system also uses a cascaded classifier for occlusion verification, which combines skin color analysis and face template matching, trained with the AdaBoost algorithm. The performance of the system was evaluated on a custom dataset of 120 video sequences and the head detection accuracy was 98.64% and the occlusion detection accuracy was 98.56% with the frame rate of 12 fps. However, it has some restrictions in dynamic lighting and complex occlusions, which are outside the omega-shaped region.

Table 10: Summary of reviewed hybrid approaches for occluded face detection

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2014	Shin and Kim [132]	Hybrid method combining discriminative and generative techniques	AR, LFPW, Talking Face	Partial, High	Normalized error rates	Reduced errors by 10% on AR; lower errors on LFPW and Talking Face	Robust against occlusions; handles pose and tracking effectively	Increased computational complexity due to hybrid optimization
2015	Liu and Graeser [133]	PMT, CST, OST with Haar-like features	Custom driver dataset	Eye occlusion	Detection rate, FPR	OST (4 sub-blocks): 91.3% accuracy, 0.3% FPR; OST (6 sub-blocks): 97.9% accuracy, 0.2% FPR	Robust to eye occlusion; suitable for real-time applications	Increased computational cost with sub-block count; PMT limited for non-uniform shadows
2015	El-Barkouky et al. [134]	Selective Parts Model (SPM) based on Deformable Parts Model (DPM), matching visible parts only.	FDDB, WIDER FACE	Partial and Heavy Occlusion	True Positive Rate (TPR)	Achieved 87% TPR for partially occluded faces, outperforming standard DPMs	Robust to heavy occlusions, effective for complex backgrounds	Struggled with extreme occlusions (<30% visibility), high computational cost, limited real-time use
2018	Zhang et al. [26]	Omega-shaped head localization with Bayesian tracking and AdaBoost	Custom ATM surveillance dataset	Partial, High	Accuracy, FPS	Face detection: 98.64%; Occlusion detection: 98.56%; 12 FPS	Robust for ATM environments; effective with severe occlusions	Limited generalization under dynamic lighting and complex occlusions
2019	Mahbub et al. [47]	Proposal-based (FSFD, SegFace, DeepSegFace), DRUID (end-to-end regression)	UMDAA-02-FD, AA-01-FD	Partial and heavy occlusion	Precision-Recall, ROC Curve	DRUID outperformed proposal-based methods in precision and recall, achieving TAR of 91.65% on AA-01-FD	Robust against occlusions, fast, scalable to mobile environments, effective augmentation techniques	Limited to single face detection; requires diverse data for generalization
2019	Qezavati et al. [25]	Hybrid approach combining Haar Cascade, LBPH, and SVM with color histogram analysis	Custom dataset with 10,000 images captured in a crowded office environment	Partial occlusion (headscarves, side-view faces)	Precision	Improved precision over standalone methods like Haar Cascade and LBPH	Effective under partial occlusion and low resolution; combines multiple techniques for improved performance	Limited adaptability to dynamic occlusion patterns; struggles with side views and low-resolution images

(Continued)

Table 10 (continued)

Year	Paper	Methodology	Datasets	Occlusion level	Evaluation metrics	Key results	Advantages	Limitations
2020	Balasundaram et al. Qezavati et al. [25,135]	Pivotal point analysis using Viola-Jones, PCA for dimensionality reduction, and SVM for classification.	AR Face Database, Real-time	Partial Occlusion	Accuracy, TPR, FPR, CIR	97% accuracy for occluded face detection	High accuracy, computational efficiency, robust for surveillance applications	Struggled with low-light conditions and tilted orientations, occasional misclassification
2021	Zhu et al. [131]	CMS-RCNN: Multi-scale feature extraction with contextual body reasoning	WIDER Face, FDDB	Partial, High	AP, Recall	AP: 90.2% (Easy), 87.4% (Medium), 64.3% (Hard); Competitive recall (FDDB)	Robust against occlusion; effective contextual reasoning	High computational demand; struggles with crowded scenes
2021	Batagelj et al. [22]	Two-stage pipeline with CNN models	MAFA, Wider Face	Masked faces	mAP, accuracy	mAP > 90% for detecting compliant and non-compliant masks; 98% accuracy	Robust to proper/improper mask placement detection	Dataset lacks granular occlusion representation; computational complexity
2022	Li [95]	Double-channel CNN with occlusion perceptron and transfer learning	AR, MAFA	Sunglasses, scarves, mixed occlusions	Accuracy, FPS	99.46% (sunglasses), 99.73% (scarf), 80.2% overall accuracy on MAFA	Robust to occlusions; high detection accuracy; fast processing speed	High computational cost; manual parameter tuning for occlusion thresholds

The study by Shin and Kim [132] proposed a hybrid approach that combined discriminative and generative methods to improve facial feature detection and tracking, especially under occlusions and pose variations. Discriminative techniques ensured accurate feature localization using local constraints, while generative methods minimized global appearance errors. The method worked in two stages. First, it estimated facial pose and initialized parameters using a multi-view face detector and the RANSAC method. Then, it refined the parameters by combining local shape errors and global appearance errors through iterative optimization. To handle occlusions, a shape-weighting matrix excluded occluded features from optimization. It also extended the framework for facial feature tracking by ensuring temporal continuity across video frames. Evaluations on datasets like AR Face Database (AR), Labeled Face Parts-in-the-Wild (LFPW), and Talking Face Video showed that the approach achieved lower error rates than existing methods, particularly under heavy occlusions, reducing errors by about 10% on the AR dataset. However, the process required more computation time due to the combined optimization. Another hybrid approach for detecting partially visible and occluded faces captured by mobile cameras was proposed by Mahbub et al. [47]. The study introduced two approaches: proposal-based detection and end-to-end regression-based detection. The proposal-based methods, including FSFD, SegFace, and DeepSegFace, generated facial segment proposals and classified them using SVMs or CNNs, but they were computationally intensive. To address this, the authors developed DRUID (Deep Regression-based User Image Detector), which bypassed proposal generation and directly predicted face and segment bounding boxes using a regression loss function. DRUID also utilized data augmentation and regularization to handle variations in visibility, scale, and lighting. Tests on UMDAA-02-FD and AA-01-FD datasets showed that DRUID outperformed other methods, achieving a True Acceptance Rate (TAR) of 91.65%. However, the model was limited to single-face detection and required diverse datasets for better generalization across different occlusion patterns. To protect drivers from dazzling light, the study by Liu and Graeser [133] introduced methods for detecting faces with eye occlusions caused by shadows, such as those created by systems like ShadeVision. To enhance the detection accuracy and to reduce the false positives the authors have suggested several strategies. One of the approaches, Partially Masked Training (PMT)–trained the model on shadowed images which helped the model to generalize better in the presence of occlusions. Another method Consecutive Sub-block Training (CST)–splits training images into sequential blocks to enhance the robustness to occlusions. Based on CST, Overlapped Sub-block Training (OST) was also introduced to enhance the performance by using overlapping image regions for better facial landmark coverage. OST detected the faces with 91.3% accuracy using four sub-blocks and 97.9% accuracy using six sub-blocks with false positive rates as low as 0.2%. Nevertheless, the methods had some problems with computational complexity, especially when using more sub-blocks, and PMT had a problem with the invariance to the shadow intensity, which limited its generality.

El-Barkouky et al. [134] presented a framework for detecting partially occluded faces in difficult conditions. It proposed the Selective Parts Model (SPM) as an enhancement of the Deformable Parts Model (DPM) that aimed at comparing only those parts of the face that are not covered by an occlusion, instead of using the entire face. The SPM-based face detection used small parts of the face and assigned confidence values to each part according to its visibility and combined these values into a single global detection score that gave higher weights to the visible parts. The results of the experiments on FDDB and WIDER FACE datasets were 87% of TPR for partially occluded faces, which is better than that of the standard DPMs for extreme occlusion and complicated backgrounds. Nevertheless, the method had some limitations, namely, it could not work effectively with the most severe occlusions (less than 30% of the visibility) and had a high computational complexity due to the description of the multiple parts, which prevents real-time operation. In the study conducted by [135], a technique was introduced to detect partially obscured faces using pivotal point analysis. Some random important facial features were picked from the list, including the eye pupils, nose tip, and mouth center (referred to as pivotal points), and then used to find out the extent of the occlusion. It applied the Viola-Jones algorithm for the feature extraction and used the PCA for the dimensionality reduction to improve the computational complexity. An SVM-based binary classifier was implemented to distinguish between an occluded and a non-occluded face based on the availability of the pivotal points. The method was tested on the AR Face Database and real-time images with 97% accuracy. Its robustness was shown by such metrics as TPR and FPR. However, the system had some drawbacks, including the sensitivity to low light, and the presence of unwanted faces with a tilted posture, which sometimes led to incorrect decisions. In 2019, reference [25] introduced a method for detecting partially uncovered faces in low-resolution surveillance videos, with a specific emphasis on Central Asian clothing, such as head coverings. The approach used Haar Cascade and Locally Binary Patterns Histogram (LBPH) to find features and SVM for classification. To improve the sensitivity, color histogram analysis was performed to enhance the probability of detecting skin color regions. A new dataset was also employed; this consisted of surveillance videos with more than 10,000 face images of people in a crowd in office environments with all poses and partial coverage. It was found that the use of the proposed method enhanced the precision of the detection as compared to the use of Haar Cascade or LBPH alone. However, the method had some limitations, such as sensitivity to dynamic occlusions and low accuracy in detecting side-view faces in low-resolution videos.

4.6 Summary of Reviewed Studies

This section presents a visual summary of the research studies to give a general view of the trends, preferred methodologies, and dataset choices in the reviewed studies.

Fig. 11 shows a noticeable increase in research interest, particularly after 2015. This reflects the growing relevance of this topic due to advances in deep learning and the increasing demand for robust face detection systems in challenging scenarios.

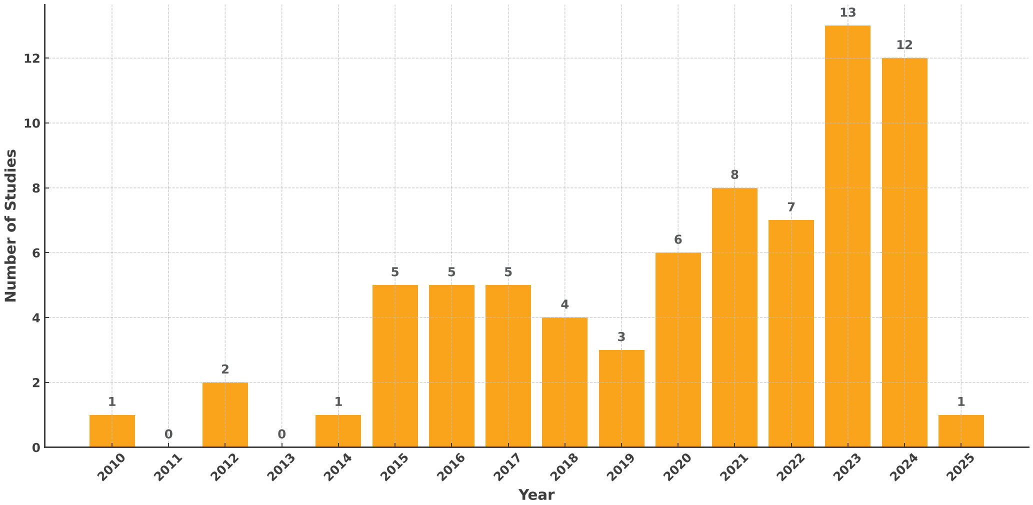


Figure 11: Number of publications per year that focused on face detection under occlusion

The pie chart in Fig. 12 shows that deep learning-based methods are the most popular (69.9%), while hybrid approaches that combine traditional and modern techniques are also widely used, indicating the attempt to combine different strategies for better occlusion handling.

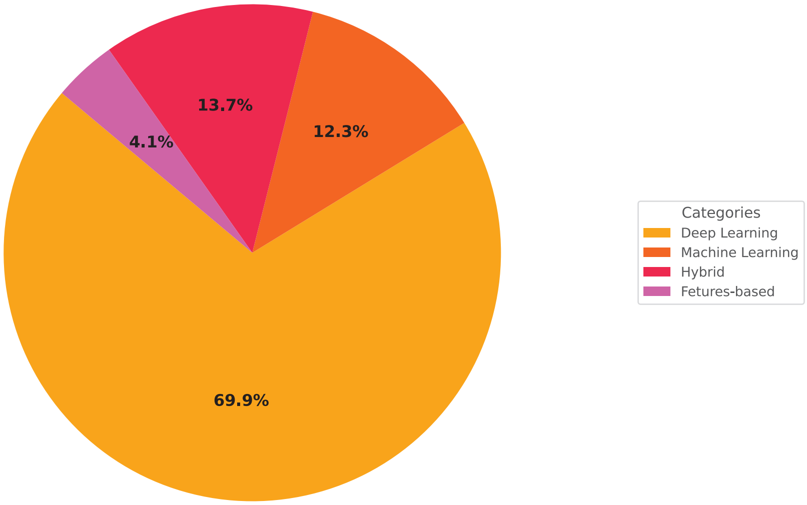


Figure 12: Percentage of research publications across categories

Fig. 13 highlights the predominance of WIDER Face, FDDB, and MAFA datasets, confirming their role as benchmarks. Other datasets, such as AFW, and PASCAL Face, also show notable usage, which indicates their relevance in specific research scenarios. It is important to note that the remaining 38% of the datasets were used only once. These datasets are often custom-made, designed for specific domains, or are not publicly available, which may limit their widespread adoption in face detection research. However, the appearance of specialized datasets like Niqab-Face suggests increasing attention to diverse and severe occlusion scenarios.

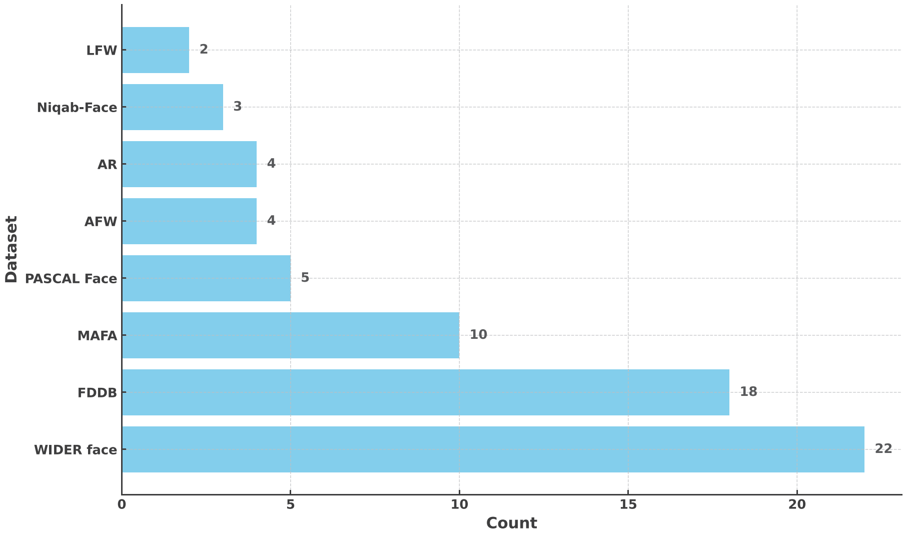


Figure 13: Common datasets used more than twice in the reviewed publications, representing 62% of all analyzed datasets

5 Benchmark Datasets for Occluded Face Detection

Having discussed the primary detection methodologies, we now review the benchmark datasets commonly used to train and evaluate occluded face detection models. Benchmark datasets play a crucial role in the development and evaluation of face detection algorithms, providing standardized benchmarks to compare the performance of various methods under diverse conditions. These datasets often encompass a wide range of variations, including changes in pose, illumination, expressions, and occlusions. The inclusion of such variation factors ensures that face detection models are robust and adaptable to real-world scenarios. Among the many datasets, not all of them are used for the same purpose; some are used for general face detection, such as the Face Detection Dataset and Benchmark (FDDB) [136], PASCAL Face [137], Annotated Faces in the Wild (AFW) [10], IARPA Janus Benchmark A (IJB-A) [138], Wider Face [4], and Masked Faces (MAFA) [13], which present images with various constraints and challenges. However, since occlusion is still a major problem in face detection, not all of these datasets have been designed to include only images that meet specific occlusion-level constraints, but some of them have, such as the Masked Face Detection Dataset (MFDD) [139], Niqab Dataset [6], Headscarf Partially Covered Face Dataset [25] and Face-Mask Label Dataset (FMLD) [22]. There are a few datasets which are particularly developed to evaluate the performance of occlusion detection algorithms, and these include the FaceOcc Dataset [140], and FSG-FD Dataset [83] with images having annotated occlusion types and levels.

This section presents an overview of the commonly used datasets for face detection benchmarking, with a particular emphasis on those that support the task of detecting occluded faces. A detailed comparison of these datasets is provided, which includes attributes such as size, image quality, levels of occlusion, types of occlusion, and other factors that are relevant when evaluating face detection systems in occluded conditions. Tables 11 and 12 offer a structured comparison across multiple dimensions, covering the dataset source, the number of images and labeled faces, the types and severity of occlusion, variation factors (e.g., pose, lighting, expression), annotation details, intended primary use (e.g., detection or recognition), and suitability for occlusion-aware face detection models. The table also contains information on whether the datasets are easily retrievable and accessed by researchers for their projects. Where applicable, “N/A” in the table means that the particular feature is not specifically discussed in the referred literature.

Table 11: Comprehensive comparison of existing datasets utilized for developing and testing occluded face detection models (Part 1)

Dataset Name	Year	Source	Size (images and faces)	Levels of occlusion	Types of occlusion	Variation factors	Annotation details	Applications	Primary Use	Suitability for occluded face detection models	Data accessibility
AR face database [141]	1998	Controlled environment (CVC, Universitat Autónoma)	Over 4000 images of 126 individuals (70 males, 56 females)	Partial	Sunglasses, scarves	Facial expressions, lighting conditions, with/without accessories	Manual organization; no detailed bounding box or occlusion labeling	Facial recognition testing, algorithm evaluation for expression and lighting variation	Face recognition	Partially suitable	Publicly available for non-commercial research
FDDB [136]	2010	Yahoo! News Website	2845 images, 5171 faces	Partial, heavy	Occlusions caused by objects, low resolution	Pose variations, lighting, low resolution, occlusion levels	Elliptical annotations for face regions, covering visible areas	Benchmarking face detection in unconstrained environments	Face detection	Partially suitable	Public
AFLW [142]	2011	Flickr Images	25,993 faces in 21,997 images	Partial, None	N/A	Pose, Lighting, Expression, Ethnicity, Age, Gender, Hairstyles	21 Landmarks, Bounding Boxes, Ellipses	Facial Feature Localization, Multi-view Face Detection, Coarse Head Pose Estimation	Face detection	Partially suitable.	Public
AFW [10]	2012	Flickr Images	205 images, 468 faces	Partial, None	Sunglasses	Pose, Lighting, Appearance, Skin Color, Expression, Makeup, Aging	Bounding Boxes, 6 Landmarks, Viewpoint Annotations	General Detection, Pose Estimation, Landmark Localization	Face detection	Limited suitability.	Public
MALF [46]	2015	Internet and Baidu image search	5250 high-resolution images with 11,931 labeled faces	Partial (occluded faces labeled as an attribute)	Not explicitly categorized	Pose, gender, resolution, wearing glasses, exaggerated expressions	Bounding boxes and multi-attribute labels (e.g., pose, occlusion, gender)	Fine-grained performance evaluation of face detection algorithms	Face detection	Partially suitable	Public
IARPA Janus Benchmark A (IJB-A) [138]	2015	Internet (Creative Commons-licensed images/videos)	5712 images and 2085 videos with over 67,183 manually localized faces	Partial (eyes, mouth/nose, forehead)	Diverse (facial hair, head coverings, glasses)	Pose, illumination, occlusion, geographic diversity	Bounding boxes, fiducial landmarks (eyes and nose), occlusion metadata	Evaluating face detection, recognition, and landmark localization algorithms	Both detection and recognition	Highly suitable	Public
UMDAA-02 Face Detection Subset (UMDAA-02-FD) [143]	2015	Front camera images from smartphone usage	33,209 images with face annotations from 43 users	Partial faces due to occlusion and pose variations	Hands, phones, lighting artifacts, and other common obstructions	Pose, illumination, occlusion, expression	Bounding boxes, face orientation, five fiducial landmarks (eyes, nose, mouth corners)	Face detection, partial face detection under real-world conditions	Face detection for active authentication on mobile devices	Highly suitable	Restricted; available for research upon request
Dazzling Avoidance Occluded Face Dataset [133]	2015	On-road driving scenarios using ShadeVision system	10,000 images of drivers	Partial occlusions primarily affecting the eye region	Shadows caused by selective-darkening panels to prevent dazzling effects	Pose, lighting variations, occlusion from shadows, and driver demographics	Bounding boxes for facial regions and shadowed eye areas	Robust face detection under occluded conditions, driver monitoring, and fatigue detection	Face detection	Highly suitable	Not publicly available; likely restricted to research purposes
WIDER FACE [4]	2016	WIDER dataset via search engines	32,203 images, 393,703 faces	No occlusion, Partial (1%–30%), Heavy (over 30%)	Masks, sunglasses, obstructions	Scale, pose (typical and atypical), occlusion levels, event categories (60 classes)	Bounding boxes, occlusion levels, poses, event categories	Benchmarking face detection in diverse, challenging scenarios	Face detection	Highly suitable	Public

(Continued)

Table 11 (continued)

Dataset Name	Year	Source	Size (images and faces)	Levels of occlusion	Types of occlusion	Variation factors	Annotation details	Applications	Primary Use	Suitability for occluded face detection models	Data accessibility
MAFA [13]	2017	Search engines and social networks	30,811 images, 35,806 faces	Weak (1-2 regions), Medium (3 regions), Heavy (4 regions)	Masks (simple, complex, hybrid, human body)	Pose (5 orientations), mask types, occlusion levels	Bounding boxes, mask locations, occlusion levels, face orientations, eyes locations	Designed for benchmarking masked face detection and training	Face detection	Highly suitable	Public
LSLF [96]	2017	YouTube videos	1,195,976 labeled face images of 11,459 individuals	Light and severe partial occlusion	Accessories (hats, glasses, sunglasses, scarves), hands, hair, microphones, overlapping objects	Pose, illumination, multi-view, backgrounds, facial expressions, gender, age, race	Bounding boxes, identity labels	Multi-view and partially occluded face detection, face recognition	Face detection and recognition	Highly suitable	Publicly available for research use

Table 12: comparison of existing datasets utilized for developing and testing occluded face detection models (Part 2)

Dataset name	Year	Source	Size (Images and faces)	Levels of occlusion	Types of occlusion	Variation factors	Annotation details	Applications	Primary use	Suitability for occluded face detection models	Data accessibility
UFDD [16]	2018	Internet platforms like Google, Bing, Flickr	6425 images with 10,897 annotated faces	Partial (due to environmental conditions)	Weather-based (rain, snow, haze), lens impediments, motion blur	Pose, illumination, environmental degradations, and distractors	Bounding boxes; annotations for environmental degradations	Evaluating robustness of face detection algorithms in challenging environments	Face detection	Highly suitable	Public
Wildest faces [144]	2018	YouTube videos	67,889 frames from 2186 shots of 64 celebrities, with 109,771 annotated faces	Partial (mixed, significant, no occlusion; 20% significant occlusion)	Varied; includes heavy occlusions and challenging poses	Scale, Pose, blur, lighting, expressions, age variance (up to 40 years)	Bounding boxes, occlusion levels, shot-based splits	Testing robustness of face detection and recognition in extreme scenarios	Both detection and recognition	Highly suitable	Public
Headscarf partially covered face [25]	2019	Surveillance video footage	5000 facial images out of 10,000 images cropped from surveillance videos	Partial occlusion due to headscarves and traditional clothing	Headscarves, partial face coverings	Pose, lighting conditions, head orientation, and occlusions due to traditional clothing	Cropped face regions annotated for training and testing	Surveillance monitoring, face detection in culturally specific environments, evaluation of occluded face detection methods	Face detection	Highly suitable	Not publicly stated; presumed restricted for academic purposes
FSG-FD self-monitored [83]	2019	N/A	300 labeled image, which includes 3236 faces	Partial and significant occlusions	Hands, masks, hair, glasses, and various accessories	Pose, occlusion types, lighting conditions, and background complexity	Bounding box annotations for face detection	Evaluation of face detection models, particularly under occlusion scenarios	Occluded face detection	Highly suitable	Not explicitly mentioned; likely restricted for research
Niqab dataset [6]	2020	Search engines and social networks	10,000 images, 12,000 faces	High (70%–100%)	Niqabs, veils	Pose variations, lighting conditions	Bounding boxes, occlusion metadata	Designed for occluded face detection, particularly for niqab-wearing individuals	Face detection	Highly suitable	Public upon request
MFDD [139]	2020	Internet	24,771 masked face images	Partial (mask)	Face masks	N/A	Mask presence and bounding boxes	Masked face detection	Face detection	Highly suitable	Public
Real-world Masked Face Recognition (RMFRD)	2020	Internet	5000 images of 525 people wearing masks, and 90,000 images of the same 525 subjects without masks	Partial (mask)	Face masks	N/A	Bounding boxes, identity labels	Masked face recognition	Face recognition	Highly suitable	Public
Simulated Masked Face Recognition (SMFRD)	2020	Simulated masks applied to existing datasets	500,000 images of 10,000 subjects	Partial (simulated masks)	Simulated Face masks	Depends on source datasets	Identity labels	Face recognition systems for robustness in masked face scenarios	Face recognition	Highly suitable	Public
ROF [145]	2021	Google image search	6421 neutral images, 4627 sunglasses images, 678 masked images; total subjects: 161	Partial (upper face, lower face)	Sunglasses, face masks	Pose, illumination, image quality	Bounding boxes; identity annotations	Testing robustness of face recognition systems against real-world occlusions	Face recognition	Highly suitable	Publicly available on GitHub
FMLD [22]	2021	MAFA and WIDER FACE datasets	41,934 images and 63,072 faces; includes training and testing splits	Paritial (masks)	Masks, incorrectly worn masks	Pose, illumination, ethnicity, gender, and background diversity	Bounding boxes, gender, ethnicity, pose, mask presence, and mask compliance	Face detection, mask compliance detection, and classification	Face detection and classification (mask compliance detection)	Highly suitable	Publicly available on GitHub
Custom occlusion dataset	2021	Wuhan University, MAFA, and web-sourced images	from 20,000 images, 3120 labeled for training and testing	Partial occlusions due to masks and other coverings	Masks (worn correctly/incorrectly), collars, hands, scarves, and other obstructions	Pose, lighting conditions, occlusion types, and variations in object coverings	Bounding boxes and six occlusion categories (correct mask, no mask, collar, hand, scarf, other)	Face occlusion detection, training models for mask compliance and occlusion recognition	Occluded face detection for mask and occlusion recognition	Highly suitable	Presumed restricted for research purposes; no explicit statement in the document

(Continued)

Table 12 (continued)

Dataset name	Year	Source	Size (Images and faces)	Levels of occlusion	Types of occlusion	Variation factors	Annotation details	Applications	Primary use	Suitability for occluded face detection models	Data accessibility
FaceOcc Dataset [140]	2022	CelebA-HQ and additional sources	Over 30,000 images	Partial and full occlusions	Sunglasses, spectacles, hands, masks, scarves, microphones, accessories	Pose, illumination, occlusion types, augmented textures	Manually labeled occlusion masks for facial regions; additional attributes from CelebA-HQ	Face occlusion detection and segmentation, training face extraction models	Face occlusion detection, face extraction, and segmentation	Not suitable for occluded face detection tasks that require bounding box annotations	Publicly available via GitHub
Facemask Detection Dataset (M, S, G datasets) [94]	2023	Real-world and synthetic masked face images	M: 853 masked face images, S: 262 synthetic masked face images, G: 1115 images (masked and unmasked faces)	Partial (facial area covered by masks)	Masks (real and synthetic)	Pose, illumination, real vs. synthetic masks	Bounding boxes and class labels (masked or unmasked)	Facemask detection, compliance monitoring	Face detection and facemask compliance classification	Highly suitable	Publicly available via Kaggle

This comprehensive comparison aims to assist researchers in understanding the strengths, limitations, and suitability of each dataset for various occlusion scenarios. In addition to the comparison, Fig. 14 below shows examples of images together with their descriptions from some of the common datasets. These visualizations show the variety of occlusions in terms of their type, severity and surrounding environment present in the datasets.

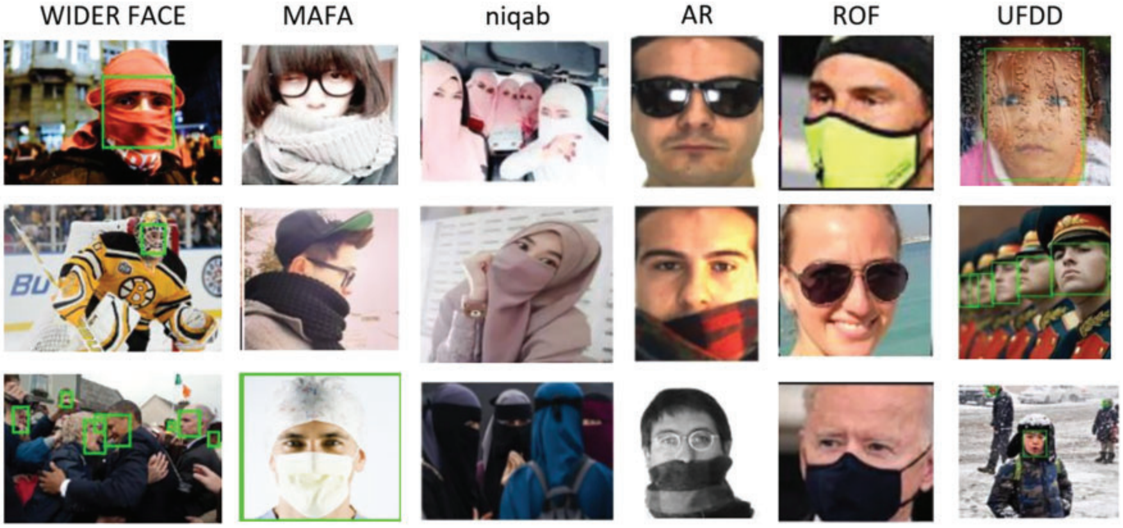


Figure 14: Sample images from benchmark datasets, including WIDER FACE, MAFA, Niqab Dataset, AR Face Database, ROF, and UFDD

5.1 Key Observations from Dataset Comparison

From the detailed comparison in Tables 11 and 12, several key observations can be made:

1. Dataset Primary Use: While some datasets, such as the AR Face Database and Real-world Masked Face Recognition (RMFRD), are primarily designed for face recognition, they can also be utilized for face detection tasks. This is possible because the annotation details (e.g., bounding boxes and identity labels) support face localization, which is a prerequisite for detection models.
2. Levels and Types of Occlusion: Most datasets provide partial occlusions, such as masks, sunglasses, hands, and scarves. However, datasets like Niqab Dataset and MAFA explicitly focus on heavy occlusions, making them particularly suitable for benchmarking models designed for challenging scenarios.
3. Variation Factors: The majority of datasets include multiple variation factors, such as pose, lighting conditions, and occlusion types, enhancing their utility for training robust models.
4. Suitability for Occlusion Detection: Datasets like WIDER FACE, MAFA, and Dazzling Avoidance Occluded Face Dataset are highly suitable for evaluating face detection under occluded conditions due to their diverse annotations and focus on occlusion-handling capabilities.
5. Accessibility: While many datasets are publicly available (e.g., WIDER FACE, UFDD, and FaceOcc Dataset), others are restricted or require requests for research purposes, which may limit their accessibility.
6. Recent Surge in Dataset Availability: There is a noticeable trend of increased interest in creating datasets for occluded face detection in recent years, particularly between 2019 and 2023. This surge reflects the growing importance of occlusion-aware face detection, especially with real-world demands such as

masked face detection due to the COVID-19 pandemic. Examples include the Masked Face Detection Dataset (MFDD), Face-Mask Label Dataset (FMLD), and the Headscarf Partially Covered Face Dataset.

5.2 Additional Observations: Diversity, Environment, and Realism

Alongside the structured comparison in Tables 11 and 12, there are several additional factors that affect how suitable and fair occluded face detection datasets are. These include demographic diversity, the range of environments, and image resolution; factors that standard evaluation metrics often overlook. For example, the WIDER FACE dataset contains images from 60 different event categories, showing a wide range of poses, scales, facial expressions, appearances, and levels of occlusion. It also features both indoor and outdoor scenes, making it a strong reflection of real-world conditions. However, it does not provide detailed demographic information such as gender or ethnicity, which makes it harder to assess fairness in detection performance.

Demographic diversity-such as gender, ethnicity, and age-is often either limited or not consistently recorded in many datasets. Some datasets, like LSLF, RMFRD, FaceOcc, and FMLD, include a range of participants or provide demographic labels, which allows for more comprehensive evaluations. On the other hand, datasets like AR Face, FDDB, and Niqab do not include clear demographic information, making it harder to evaluate how well models perform for different groups. This lack of demographic data can result in detection systems that are biased and less accurate for underrepresented communities.

The environment where images are collected, whether indoors, outdoors, or a mix of both, also influences how well models generalize. Datasets like WIDER FACE, FDDB, and UFDD include a variety of conditions, covering both indoor and outdoor scenes, which makes them more suitable for real-world applications. In contrast, datasets such as AR Face and the Dazzling Avoidance Occluded Face Dataset are gathered in controlled settings, which can limit their effectiveness in more dynamic or unpredictable environments.

Image resolution plays an important role in face detection. Some datasets, like LSLF and the Custom Covered Face Dataset, provide high-resolution images that keep facial details clear, which helps in identifying subtle occlusions. On the other hand, older datasets such as AR Face and AFW contain lower-resolution images, which can reduce detection accuracy, especially when using advanced deep learning models.

These challenges reveal potential biases that may impact the robustness, fairness, and generalizability of face detection models. To create more effective and inclusive systems, future datasets should include a wider range of demographics, diverse environmental conditions, and more consistent annotations for different types of occlusions.

6 Challenges and Limitations in Occluded Face Detection

Despite recent advancements, the development of reliable occluded face detection models faces multiple ongoing challenges. This section identifies the main limitations and open problems that limit current methods while highlighting specific areas that need additional innovation. These challenges arise due to partial or full occlusion of facial regions, loss of critical features, and generalization issues across different occlusion types. This section categorizes the major challenges into algorithmic challenges, dataset challenges, and performance and generalization issues.

6.1 Algorithmic Challenges

Algorithmic challenges involve the technical limitations faced by detection models when handling occluded faces:

6.1.1 Loss of Critical Feature Information

Occluded faces often lack visible key features (e.g., eyes, nose, mouth), making it difficult for both traditional methods (e.g., Haar cascades, HOG) and deep learning models to extract discriminative information. This significantly affects detection accuracy, especially under heavy occlusion.

6.1.2 High Rate of False Positives

Some models are wrongly triggered by occlusions resembling facial features (e.g., shadows, hands, patterned clothing), leading to high false positive rates. Traditional methods struggle due to limited feature discrimination, while deep learning models also fail if occlusion diversity is lacking in training data. Developing robust algorithms to filter out such non-face occlusions remains a key challenge.

6.1.3 Diversity and Complexity of Occlusions

Occlusions vary greatly in type, size, and location, arising from masks, sunglasses, hands, scarves, or environmental obstructions. Models trained on one type often generalize poorly to others, causing performance drops in unseen scenarios. Handling this complexity requires datasets and models that cover a wide range of occlusion variations.

6.1.4 Method-Specific Challenges

Different face detection methods face unique challenges with occlusion. Traditional feature-based approaches rely on handcrafted features and symmetry, but become ineffective when key features are hidden. Machine learning methods like SVMs struggle with complex occlusions and large datasets. Deep learning methods are more powerful but often overfit on small or occlusion-specific data and are computationally expensive. Their performance also drops with increased occlusion. Hybrid methods improve accuracy but add complexity, making real-time use more difficult.

6.2 Datasets Challenges

Robust datasets are crucial for training and evaluating occluded face detection models. However, several limitations persist:

1. **Annotation challenges:** Annotating occluded datasets requires precise labeling of occlusion regions and degrees. Manual annotations are labor-intensive, and inconsistencies across datasets hinder comparative evaluation.
2. **Lack of diversity:** Existing datasets often lack sufficient diversity in occlusion types, levels, and environments. For example, some datasets focus on masked faces (MAFA, RMFRD) while neglecting other occlusions like headscarves or hands.
3. **Imbalanced coverage of occlusion levels:** Heavily occluded faces (over 70%) remain underrepresented, leading to poor model generalization for extreme scenarios.
4. **Temporal and environmental constraints:** Most datasets are static images; real-world video streams with occlusions caused by motion blur or lighting variations remain underexplored.

6.3 Performance and Generalization Issues

1. **Performance drop with increasing occlusion:** As occlusion levels increase (e.g., from partial to heavy), detection accuracy drops significantly, especially in deep-learning models that rely on global visual features.

2. **Overfitting to specific occlusion types:** Models trained on specific occlusion types (e.g., masks) struggle to generalize to unseen occlusions like scarves or environmental barriers. Generalization across datasets remains a major challenge.
3. **Domain and real-world adaptation:** Many methods perform well in controlled environments but fail in unconstrained settings where occlusions occur unpredictably. Techniques such as transfer learning or domain adaptation are essential to address this gap.

7 Future Research Directions

Building upon the identified challenges, this section explores promising future research directions and emerging trends aimed at improving the robustness, efficiency, and scalability of occluded face detection systems. Although there has been great improvement in the detection of occluded faces, there are open issues that still need to be handled to find effective solutions. These issues will be discussed in this paper as future work. This final section focuses on the remaining issues, namely, algorithmic limitations, dataset shortcomings, and deployment challenges.

7.1 Robust and Efficient Deep Learning Models

Enhancing the robustness and efficiency of deep learning-based methods remains a key challenge in occluded face detection. Occlusions introduce variations that can severely degrade model performance, leading to false positives or missed detections. Addressing these challenges requires a multi-faceted approach as follows:

1. Designing occlusion-aware Architectures: Develop networks that explicitly handle occluded regions, such as dual-path or attention-based models that separate visible and occluded features.
2. Developing occlusion-invariant features: Design robust feature extraction techniques that work well independent of the level or type of the occlusion to improve the generalization across the various occlusion scenarios.
3. Involving optimization techniques: Employ hyperparameter tuning methods (e.g., Bayesian optimization) and metaheuristic algorithms (e.g., genetic algorithms) to improve model performance and generalization.
4. Ensuring lightweight and real-time models: To enable real-time deployment on edge devices, design computationally efficient architectures using model compression techniques such as pruning, quantization, and knowledge distillation.

Furthermore, emerging techniques from related fields, such as graph-based learning for human pose estimation and multi-view feature fusion for occlusion handling [146–148], may offer promising strategies for improving the robustness and adaptability of occluded face detection models. Exploring such cross-domain approaches could inspire the development of more effective architectures for challenging real-world scenarios.

7.2 Ensuring Development of Diverse and Annotated Datasets

The creation of robust datasets is essential to train and evaluate models under diverse occlusion scenarios:

1. Diverse occlusion types and levels: Build datasets with a variety of occlusion types (e.g., masks, hands, scarves) and granular annotations for partial and heavy occlusions.
2. Synthetic data generation: Use generative adversarial networks (GANs) and synthetic augmentation to simulate realistic occlusions, improving dataset size and balance.

3. Temporal and environmental variability: Incorporate video-based datasets and real-world scenarios with dynamic occlusions caused by lighting, motion blur, or environmental obstructions.

7.3 Hybrid and Multi-Scale Approaches

Combining the strengths of various methods can further improve occluded face detection:

1. Feature fusion: Merge handcrafted and deep-learning-based features to improve detection accuracy under varying occlusions.
2. Multi-scale analysis: Integrate multi-scale methods to enhance detection performance for small or partially visible faces.
3. Contextual information: Leverage auxiliary cues, such as body posture and environmental context, to infer occluded regions.

7.4 Application-Specific Solutions

Future work should address the unique requirements of real-world applications:

1. Surveillance systems: Develop models for detecting occluded faces in crowded, dynamic environments with low-resolution inputs.
2. Masked face detection: Enhance performance for masked face detection, a significant post-pandemic challenge in healthcare and security.
3. Cultural occlusions: Improve detection for culturally specific occlusions, such as niqabs and headscarves, which remain underrepresented in datasets.

8 Conclusion

This review provided a comprehensive analysis of face detection methods under occlusion and systematically categorized them into traditional feature-based, machine learning, deep learning, and hybrid approaches. The reviewed studies demosntrated that early occluded face detection methods, such as feature-based and traditional machine learning methods, provided important foundations by focusing on handcrafted features and statistical models. However, their limited ability to handle complex occlusions and variations led to the adoption of CNNs as a more powerful alternative. While CNN-based models improved detection robustness by learning hierarchical features, they still struggled with severe occlusions and generalization to unseen scenarios. Modern advances in deep learning models achieve better management of complex occlusion patterns through the implementation of multi-scale feature extraction and attention mechanisms. Traditional and learned features can be combined through hybrid feature fusion strategies to show potential in closing performance gaps when faces are partially occluded. The field has moved toward Transformer-based architectures, including Vision Transformer (ViT) and Swin Transformer, as well as GAN-based models, which provide new capabilities for context modeling and face restoration and occlusion-aware feature learning. Despite the progress made in detecting occluded faces, challenges such as high false positive rates, limited dataset diversity, and poor generalization to unseen occlusions remain. Many state-of-the-art models still struggle to achieve real-time performance and computational efficiency, limiting their deployment in resource-constrained environments. Future research should focus on the following directions: (1) Developing large-scale, occlusion-specific datasets with real-world diversity; (2) Developing lightweight, adaptive models that can balance detection accuracy with real-time efficiency; (3) Developing occlusion-aware learning frameworks, including partial feature modeling and face reconstruction networks; (4) Developing standardized benchmarks for evaluating occluded face detection performance; and (5)

Using synthetic data generation methods such as GANs and domain adaptation to enrich training datasets. Addressing these limitations is crucial to developing more reliable and scalable face detection solutions.

This study has some limitations that should be acknowledged. First, while we attempted to provide a comprehensive review, some recent developments and proprietary models may not have been covered due to access restrictions. Second, the model comparison was based primarily on reported metrics, which may not fully reflect differences in real-world performance. Additionally, the lack of a standardized criterion for hidden face detection limits the ability to make absolute comparisons between different approaches. Finally, the study does not include an experimental evaluation, meaning that the results are based on secondary sources rather than direct experimental validation.

Acknowledgement: The authors would like to acknowledge A'Sharqiyah University, Sultanate of Oman, for supporting this research under Research Project grant number (BFP/RGP/ICT/22/490).

Funding Statement: This research was funded by A'Sharqiyah University, Sultanate of Oman, under Research Project grant number (BFP/RGP/ICT/22/490).

Author Contributions: Conceptualization: Thaer Thaher, Majdi Mafarja, Muhammed Saffarini, Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. Literature review and critical analysis: Thaer Thaher, Muhammed Saffarini. Review structure and thematic organization: Thaer Thaher, Majdi Mafarja. Data collection and curation: Thaer Thaher, Ayman A. El-Saleh. Formal analysis: Thaer Thaher. Writing—original draft preparation: Thaer Thaher, Muhammed Saffarini, Majdi Mafarja, Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. Writing—review and editing: Thaer Thaher, Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. Visualization: Thaer Thaher. Supervision: Thaer Thaher, Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. Project administration: Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. Funding acquisition: Abdul Hakim H. M. Mohamed, Ayman A. El-Saleh. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Feng Y, Yu S, Peng H, Li YR, Zhang J. Detect faces efficiently: a survey and evaluations. IEEE Transact Biomet, Behav Iden Sci. 2022 Jan;4(1):1–18. doi:10.1109/TBIOM.2021.3120412.
2. Yang MH, Kriegman DJ, Ahuja N. Detecting faces in images: a survey. IEEE Transact Pattern Anal Mach Intell. 2002;24(1):34–58. doi:10.1109/34.982883.
3. Yuan Z. Face detection and recognition based on visual attention mechanism guidance model in unrestricted posture. Sci Program. 2020;2020(1):8861987. doi:10.1155/2020/8861987.
4. Yang S, Luo P, Loy CC, Tang X. WIDER FACE: a face detection benchmark. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. p. 5525–33.
5. Zhang C, Zhang Z. A survey of recent advances in face detection; 2010. MSR-TR-2010-66. [Internet]. [cited 2025 May 29]. Available from: https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/.
6. Alashbi AAS, Sunar MS. Occluded face detection, face in niqab dataset. In: Saeed F, Mohammed F, Gazem N, editors. Emerging trends in intelligent computing and informatics. Cham, Switzerland: Springer International Publishing; 2020. p. 209–15. doi:10.1007/978-3-030-33582-3_20.
7. Li C, Wang R, Li J, Fei L. Face detection based on YOLOv3. In: Jain V, Patnaik S, Popentiu Vlădicescu F, Sethi IK, editors. Recent trends in intelligent computing, communication and devices. Singapore: Springer Singapore; 2020. p. 277–84. doi:10.1007/978-981-13-9406-5_34.

8. Jafri R, Arabnia H. A survey of face recognition techniques. J Inform Process Syst. 2009;5(2):41–68. doi:10.3745/jips.2009.5.2.041.
9. Viola P, Jones M. 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; 2001 Dec 8–14; Kauai, HI, USA.
10. Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16–21; Providence, RI, USA. p. 2879–86.
11. Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. Marseille, France: Erik Learned-Miller and Andras Ferencz and Frédéric Jurie; 2008 [Internet]. [cited 2025 May 29]. Available from:https://inria.hal.science/inria-00321923.
12. Zafeiriou S, Zhang C, Zhang Z. A survey on face detection in the wild: past, present and future. Comput Vis Image Understand. 2015;138(4):1–24. doi:10.1016/j.cviu.2015.03.015.
13. Ge S, Li J, Ye Q, Luo Z. Detecting masked faces in the wild with LLE-CNNs. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 426–34.
14. Zeng D, Veldhuis R, Spreeuwers L. A survey of face recognition techniques under occlusion. IET Biometrics. 2021;10(6):581–606. doi:10.1049/bme2.12029.
15. Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett. 2016;23(10):1499–1503. doi:10.1109/LSP.2016.2603342.
16. Nada H, Sindagi V, Zhang H, Patel V. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS); 2018 Oct 22–25; Redondo Beach, CA, USA. p. 1–10.
17. Xia Y, Zhang B, Coenen F. Face occlusion detection using deep convolutional neural networks. Int J Pattern Recognit Artif Intell. 2016;30(9):1660010. doi:10.1142/S0218001416600107.
18. Alashbi A, Mohamed AHHM, El-Saleh AA, Shayea I, Sunar MS, Alqahtani ZR, et al. Human face localization and detection in highly occluded unconstrained environments. Eng Sci Technol Internat J. 2025;61(1):101893. doi:10.1016/j.jestch.2024.101893.
19. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 2980–8.
20. Zhang K, Zhu Q, Li W. Experimental research on occlusion face detection method based on attention mechanism. J Phy: Conf Ser. 2022;2258(1):012078. doi:10.1088/1742-6596/2258/1/012078.
21. Wang J, Yuan Y, Yu G. Face attention network: an effective face detector for the occluded faces. arXiv:1711.07246. 2017.
22. Batagelj B, Peer P, Štruc V, Dobrišek S. How to correctly detect face-masks for COVID-19 from visual information? Appl Sci. 2021;11(5):2070. doi:10.3390/app11052070.
23. Yahya SN, Ramli AF, Nordin MN, Basarudin H, Abu MA. Comparison of convolutional neural network architectures for face mask detection. Int J Adv Comput Sci Appl. 2021;12(12):83. doi:10.14569/IJACSA.2021.0121283.
24. Lad AM, Mishra A, Rajagopalan A. Comparative analysis of convolutional neural network architectures for real time COVID-19 facial mask detection. J Phys: Conf Ser. 2021;1969(1):012037. doi:10.1088/1742-6596/1969/1/012037.
25. Qezavati H, Majidi B, Manzuri MT. Partially covered face detection in presence of headscarf for surveillance applications. In: 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA); 2019 Mar 6-7; Tehran, Iran. p. 195–9.
26. Zhang T, Li J, Jia W, Sun J, Yang H. Fast and robust occluded face detection in ATM surveillance. Pattern Recognit Lett. 2018;107(15):33–40. doi:10.1016/j.patrec.2017.09.011.
27. Kortli Y, Jridi M, Al Falou A, Atri M. Face recognition systems: a survey. Sensors. 2020;20(2):342. doi:10.3390/s20020342.
28. Lei C, Dang K, Song S, Wang Z, Chew SP, Bian R, et al. AI-assisted facial analysis in healthcare: from disease detection to comprehensive management. Patterns. 2025;6(2):101175. doi:10.1016/j.patter.2025.101175.

29. Zhang Z, Ji X, Cui X, Ma J. A survey on occluded face recognition. In: Proceedings of the 2020 9th International Conference on Networks, Communication and Computing, ICNCC '20. New York, NY, USA: Association for Computing Machinery; 2021. p. 40–9. doi:10.1145/3447654.3447661.
30. Budiarsa R, Wardoyo R, Musdholifah A. Face recognition for occluded face with mask region convolutional neural network and fully convolutional network: a literature review. Int J Elect Comput Eng (IJECE). 2023;13(5):5662–73. doi:10.11591/ijece.v13i5.pp5662-5673.
31. Rao KS, Fernandes SL, Haniben P, Prajna, Pratheek, Devadiga RV, et al. A review on various state of art technique to recognize occluded face images. In: 2015 2nd International Conference on Electronics and Communication Systems (ICECS); 2015 Feb 26–27; Coimbatore, India. p. 595–601.
32. Zhang L, Verma B, Tjondronegoro D, Chandran V. Facial expression analysis under partial occlusion: a survey. ACM Comput Surv. 2018;51(2):25–49. doi:10.1145/3158369.
33. Alzu'bi A, Albalas F, AL-Hadhrami T, Younis LB, Bashayreh A. Masked face recognition using deep learning: a review. Electronics. 2021;10(21):2666. doi:10.3390/electronics10212666.
34. Kumari V, Kaur B. A review on comparative analysis of face detection algorithms. Int J Comput Applicat. 2023;185(20):17–21.
35. Ruvinga C, Malathi D, Jayaseeli JDD. Exploration of face detection methods in digital images. Int J Recent Technol Eng. 2019;8(4):12130–6. doi:10.35940/ijrte.d8014.118419.
36. Hasan Alhafidh BM, Hagem RM, Daood AI. Face detection and recognition techniques analysis. In: 2022 International Conference on Computer Science and Software Engineering (CSASE); 2022 Mar 15–17; Duhok, Iraq. p. 265–70.
37. Alashbi AAS, Sunar MS, Alqahtani Z. Context-aware face detection for occluded faces. In: 2020 6th International Conference on Interactive Digital Media (ICIDM); 2020 Dec 14–15; Bandung, Indonesia. p. 1–4.
38. Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. p. 1701–8.
39. Chen YN, Han CC, Wang CT, Fan KC. Face recognition using nearest feature space embedding. IEEE Transact Pattern Anal Mach Intellig. 2011;33(6):1073–86. doi:10.1109/tpami.2010.197.
40. Zhang Z, Luo P, Loy CC, Tang X. Learning deep representation for face alignment with auxiliary attributes. IEEE Transact Pattern Anal Mach Intell. 2016;38(5):918–30. doi:10.1109/tpami.2015.2469286.
41. Wang X, Zhang W. Anti-occlusion face recognition algorithm based on a deep convolutional neural network. Comput Elect Engi. 2021;96:107461. doi:10.1016/j.compeleceng.2021.107461.
42. Ghiasi G, Fowlkes CC. Occlusion coherence: localizing occluded faces with a hierarchical deformable part model. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14; 2014 Jun 23–28; Columbus, OH, USA. p. 1899–906. doi:10.1109/CVPR.2014.306.
43. Wang M, Deng W. Deep face recognition: a survey. Neurocomputing. 2021;429:215–44. doi:10.1016/j.neucom.2020.10.081.
44. Singh AK, Singh A, Sirohi H, Bhardwaj N. Review of challenges and innovations in occluded facial recognition in disguise and crowd. Int J Res Appl Sci Eng Technol (IJRASET). 2024 May;12(V):1481–7.
45. A.Alashbi A, Sunar MS, Alqahtani Z. Deep learning CNN for detecting covered faces with niqab. J Inf Technol Manag. 2022;14:114–23.
46. Yang B, Yan J, Lei Z, Li SZ. Fine-grained evaluation on face detection in the wild. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG); 2015 May 4–8; Ljubljana, Slovenia. p. 1–7.
47. Mahbub U, Sarkar S, Chellappa R. Partial face detection in the mobile domain. Image Vision Comput. 2019 Feb;82(C):1–17. doi:10.1016/j.imavis.2018.12.003.
48. Sharifara A, Mohd Rahim MS, Anisi Y. A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection. In: 2014 International Symposium on Biometrics and Security Technologies (ISBAST); 2014 Aug 26–27; Kuala Lumpur, Malaysia. p. 73–8.
49. Hire AN, Satone MP. A review on face detection techniques. Int J Trend Scient Res Develop. 2018;2(4):1470–6.

50. Thazheena T, Aswathy Devi T. A review on face detection under occlusion by facial accessories. Int Res J Eng Technol (IRJET). 2017;4(12):672–4.
51. Minaee S, Luo P, Lin Z, Bowyer K. Going deeper into face detection: a survey. arXiv:2103.14983. 2021.
52. Mondal SK, Mukhopadhyay I, Dutta S. Review and comparison of face detection techniques. In: Chakraborty M, Chakrabarti S, Balas VE, editors. Proceedings of International Ethical Hacking Conference 2019. Singapore: Springer Singapore; 2020. p. 3–14. doi:10.1007/978-981-15-0361-0_1.
53. Dagnes N, Vezzetti E, Marcolin F, Tornincasa S. Occlusion detection and restoration techniques for 3D face recognition: a literature review. Mach Vis Applicat. 2018;29(5):1–25. doi:10.1007/s00138-018-0933-z.
54. Hasan MK, Ahsan MS, Abdullah-Al-Mamun, Newaz SHS, Lee GM. Human face detection techniques: a comprehensive review and future research directions. Electronics. 2021;10(19):2354. doi:10.3390/electronics10192354.
55. Qais Abdulalla F, abduljabar Sadiq A, Hameed Shaker S. A Surveyof human face detection methods. J Al-Qadisiyah Comput Sci Mathem. 2018;10(2):108– 117. doi:10.29304/jqcm.2018.10.2.392.
56. Hjelmås E, Low BK. Face detection: a survey. Comput Vis Image Understand. 2001;83(3):236–74. doi:10.1006/cviu.2001.0921.
57. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 1996;29(1):51–9. doi:10.1016/0031-3203(95)00067-4.
58. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005 Jun 20–25; San Diego, CA, USA. p. 886–93.
59. Canny J. A computational approach to edge detection. IEEE Transact Pattern Anal Mach Intellig. 1986;8(6):679–98. doi:10.1109/tpami.1986.4767851.
60. Sobel I, Feldman G. A 3× 3 isotropic gradient operator for image processing. In: Pattern classification and scene analysis. 1st ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1973. p. 271–2
61. Cootes TF, Taylor CJ. Active shape models—'smart snakes'. In: Hogg D, Boyle R, editors. BMVC92. London, UK: Springer London; 1992. p. 266–75. doi:10.1007/978-1-4471-3201-1_28.
62. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE Transact Pattern Anal Mach Intellig. 2001;23(6):681–5. doi:10.1109/34.927467.
63. Srisuk S, Boonkong A, Arunyagool D, Ongkittikul S. Handcraft and learned feature extraction techniques for robust face recognition: a review. In: 2018 International Electrical Engineering Congress (iEECON); 2018 Mar 7–9; Krabi, Thailand. p. 1–4. doi:10.1109/ieecon.2018.8712272.
64. Guo Z, Zhou W, Xiao L, Hu X, Zhang Z, Hong Z. Occlusion face detection technology based on facial physiology. In: 2018 14th International Conference on Computational Intelligence and Security (CIS); 2018 Nov 16–19; Hangzhou, China. p. 106–9.
65. Bade A, Sivaraja T. Enhanced AdaBoost Haar cascade classifier model to detect partially occluded faces in digital images. ASM Sci J. 2020;13:1–6. doi:10.1007/978-3-319-08234-9_371-1.
66. Ganguly S, Bhattacharjee D, Nasipuri M. Depth based occlusion detection and localization from 3D face image. Int J Image, Graph Signal Proces. 2015;7(5):20–31. doi:10.5815/ijigsp.2015.05.03.
67. Jabbar EK, Hadi WJ. Face occlusion detection and recovery using fuzzy C-means. Eng Technol J. 2010;28(18):5744–56. doi:10.30684/etj.28.18.11.
68. Gul S, Farooq H. 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); 2015 Jul 6–8; Beijing, China. p. 149–55.
69. Arunnehru J, Kalaiselvi Geetha M, Nanthini T. Occlusion detection based on fractal texture analysis in surveillance videos using tree-based classifiers. In: Abawajy JH, Mukherjea S, Thampi SM, Ruiz-Martínez A, editors. Security in computing and communications. Cham, Switzerland: Springer International Publishing; 2015. p. 307–16. doi:10.1007/978-3-319-22915-7_29.
70. Liao S, Jain AK, Li SZ. A Fast and accurate unconstrained face detector. IEEE Transacti Pattern Anal Mach Intellig. 2016;38(2):211–23. doi:10.1109/tpami.2015.2448075.

71. Hotta K. Robust face detection under partial occlusion. Syst Comput Japan. 2007;38(13):39–48. doi:10.1002/scj.20614.
72. Priya GN, Banu RSDW. Detection of occluded face image using mean based weight matrix and support vector machine. J Comput Sci. 2012;8(7):1184–90. doi:10.3844/jcssp.2012.1184.1190.
73. SuvarnaKumar G, Reddy PVGD, Swamy S, Gupta S. Skin based occlusion detection and face recognition using machine learning techniques. Int J Comput Applicat. 2012;41(18):11–5. doi:10.5120/5640-7998.
74. Zohra FT, Rahman MW, Gavrilova M. Occlusion detection and localization from kinect depth images. In: 2016 International Conference on Cyberworlds (CW); 2016 Sep 28–30; Chongqing, China. p. 189–96.
75. Yang S, Wiliem A, Lovell BC. To face or not to face: towards reducing false positive of face detection. In: 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ); 2016 Nov 21–22; Palmerston North, New Zealand. p. 1–6.
76. Qi Y, Wang Y, Dong Y. An improved face mask detection simulation algorithm based on YOLOv5 model. Int J Gam Comput-Mediat Simulat. 2024;16(1):1–16. doi:10.4018/ijgcms.343517.
77. Wang G, Li J, Wu Z, Xu J, Shen J, Yang W. EfficientFace: an efficient deep network with feature enhancement for accurate face detection. Multimedia Syst. 2023;29(5):2825–39. doi:10.1007/s00530-023-01134-6.
78. Zhao G, Zou S, Wu H. Improved algorithm for face mask detection based on YOLO-v4. Int J Comput Intell Syst. 2023;16(1):104. doi:10.1007/s44196-023-00286-7.
79. Chen Y, Song L, Hu Y, He R. Adversarial occlusion-aware face detection. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS); 2018 Oct 22–25; Redondo Beach, CA, USA. p. 1–9.
80. Deng J, Guo J, Ververas E, Kotsia I, Zafeiriou S. RetinaFace: single-shot multi-level face localisation in the wild. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 5202–11.
81. Li J, Wang Y, Wang C, Tai Y, Qian J, Yang J, et al. DSFD: dual shot face detector. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 5055–64.
82. Jiang B, Jiang H, Zhang H, Zhang Q, Li Z, Huang L. 4AC-YOLOv5: an improved algorithm for small target face detection. J Image Video Process. 2024;2024(1):10. doi:10.1186/s13640-024-00625-4.
83. Jin Q, Mu C, Tian L, Ran F. A region generation based model for occluded face detection. Procedia Comput Sci. 2020;174(4):454–62. doi:10.1016/j.procs.2020.06.114.
84. Hu P, Ramanan D. Finding tiny faces. arXiv:1612.04402. 2017.
85. Tang X, Du DK, He Z, Liu J. PyramidBox: a context-assisted single shot face detector. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer vision–ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 812–28. doi:10.1007/978-3-030-01240-3_49.
86. Yu Z, Huang H, Chen W, Su Y, Liu Y, Wang X. YOLO-FaceV2: a scale and occlusion aware face detector. Pattern Recognit. 2024;155(10):110714. doi:10.1016/j.patcog.2024.110714.
87. Garg D, Jain P, Kotecha K, Goel P, Varadarajan V. An efficient multi-scale anchor box approach to detect partial faces from a video sequence. Big Data Cogn Comput. 2022;6(1):9. doi:10.3390/bdcc6010009.
88. Mamieva D, Abdusalomov AB, Mukhiddinov M, Whangbo TK. Improved face detection method via learning small faces on hard images based on a deep learning approach. Sensors. 2023;23(1):502. doi:10.3390/s23010502.
89. Zhang S, Chi C, Lei Z, Li SZ. RefineFace: refinement neural network for high performance face detection. IEEE Transact Pattern Anal Mach Intell. 2021;43(11):4008–20. doi:10.1109/tpami.2020.2997456.
90. Tsai AC, Ou YY, Wu WC, Wang JF. Integrated single shot multi-box detector and efficient pre-trained deep convolutional neural network for partially occluded face recognition system. IEEE Access. 2021;9:164148–58. doi:10.1109/access.2021.3133446.
91. Najibi M, Samangouei P, Chellappa R, Davis LS. SSH: single stage headless face detector. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 4885–94.
92. Zhao Y, Geng S. Face occlusion detection algorithm based on yolov5. J Phys: Conf Ser. 2021;2031(1):012053. doi:10.1088/1742-6596/2031/1/012053.

93. Nadhum A. Face detection method with mask by improved YOLOv5. J Image Process Intell Remote Sen. 2023;4:9–19. doi:10.55529/jipirs.41.9.19.

94. Kurniawan F, Astawa I, Sentana I, Atmaja I, Wibawa A. Facemask detection using the YOLO-v5 algorithm: assessing dataset variation and resolutions. Register: Jurnal Ilmiah Teknologi Sistem Informasi. 2023;9:95–102. doi:10.26594/register.v9i2.3249.

95. Li Y. Face detection algorithm based on double-channel CNN with occlusion perceptron. Comput Intell Neurosci. 2022;2022(1):3705581. doi:10.1155/2022/3705581.

96. Alafif T, Hailat Z, Aslan M, Chen X. 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); 2017 Jun 21–23; Exeter, UK. p. 28–37.

97. Iqbal SM, Shekar D, Mishra S. A comparative study of face detection algorithms for masked face detection. arXiv:2305.11077. 2023.

98. Opitz M, Waltner G, Poier G, Possegger H, Bischof H. Grid loss: detecting occluded faces. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision–ECCV 2016. Cham, Switzerland: Springer International Publishing; 2016. p. 386–402. doi:10.1007/978-3-319-46487-9_24.

99. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. ACM Comput Surv. 2022 Sep;54(10s):200. doi:10.1145/3505244.

100. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. In: 2021 International Conference on Learning Representations; 2021 May 4; Vienna, Austria.

101. Bi J, Zhu Z, Meng Q. Transformer in computer vision. In: 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI); 2021 Sep 24–26; Fuzhou, China. p. 178–88.

102. Elmi S, Morris B. Res-ViT: residual vision transformers for image recognition tasks. In: 2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI); 2023 Nov 6–8; Atlanta, GA, USA; 2023. p. 309–16.

103. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Los Alamitos, CA, USA: IEEE Computer Society; 2021. p. 9992–10002.

104. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, et al. Swin transformer V2: scaling up capacity and resolution. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24. New Orleans, LA, USA. p. 11999–2009.

105. Pradhan P, Das A, Kumar D, Baruah U, Sen B, Ghosal P. SwinSight: a hierarchical vision transformer using shifted windows to leverage aerial image classification. Multimed Tools Appl. 2024;83(39):86457–78. doi:10.1007/s11042-024-19615-9.

106. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: Computer Vision–ECCV 2020: 16th European Conference; 2020 Aug 23–28; Glasgow, UK. p. 213–29. doi:10.1007/978-3-030-58452-8_13.

107. Qin L, Wang M, Deng C, Wang K, Chen X, Hu J, et al. SwinFace: a multi-task transformer for face recognition, expression recognition, age estimation and attribute estimation. IEEE Transact Circ Syste Video Technol. 2024;34(4):2223–34. doi:10.1109/tcsvt.2023.3304724.

108. Mao Y, Lv Y, Zhang G, Gui X. Exploring transformer for face mask detection. IEEE Access. 2024;12(1):118377–88. doi:10.1109/access.2024.3449802.

109. Yuan S, Guo W, Yang F. A practical YOLOV5 face detector with decoupled swin head. In: 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2023 Oct 1–4; Oahu, HI, USA. p. 2171–7.

110. Zhou Z. YOLO-M: an improved YOLOv5 method for occlusion face. In: 2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST); 2023 Dec 8–10. Guangzhou, China. p. 918–21.

111. Zhao W, Zhu X, Guo K, Shi H, Zhang XY, Lei Z. Masked face transformer. IEEE Transact Inform Foren Secur. 2024;19:265–79. doi:10.1109/tifs.2023.3322600.
112. Pandya B, Patel D, Yow KC. Face mask detection using vision transformer. In: 2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE); 2023 Sep 24–27; Regina, SK, Canada. p. 268–72.
113. Li H, Sui M, Zhao F, Zha Z, Wu F. MVT: mask vision transformer for facial expression recognition in the wild. arXiv:2106.04520. 2021.
114. Lee I, Lee E, Yoo SB. Latent-OFER: detect, mask, and reconstruct with latent vectors for occluded facial expression recognition. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Los Alamitos, CA, USA: IEEE Computer Society; 2023. p. 1536–46.
115. Al-Sarrar HM, Al-Baity HH. A novel hybrid face mask detection approach using Transformer and convolutional neural network models. PeerJ Comput Sci. 2023;9(5):e1265. doi:10.7717/peerj-cs.1265.
116. Zou ZN, Zhang Y, Wijaya R. Investigating the robustness and properties of detection transformers (DETR) toward difficult images. arXiv.2310.08772. 2023.
117. Li T, Yu C, Li Y. DDR-DETR: real-time face detection algorithm for classroom scenarios. In: Proceedings of the 2024 International Conference on Artificial Intelligence of Things and Computing. AITC ’24. New York, NY, USA: Association for Computing Machinery; 2025. p. 192–7. doi:10.1145/3708282.3708317.
118. Chiang JC, Hu HN, Hou BS, Tseng CY, Liu YL, Chen MH, et al. ORFormer: occlusion-robust transformer for accurate facial landmark detection. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Los Alamitos, CA, USA: IEEE Computer Society; 2025. p. 784–93.
119. Zhang Z, Chao Q, Wang S, Yu T. A lightweight face detector via bi-stream convolutional neural network and vision transformer. Information. 2024;15(5):290. doi:10.3390/info15050290.
120. Kumar S, Bhardwaj S, Vishwakarma DK. Face restoration via generative adversarial networks. In: 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC); 2023 May 26–28; Jalandhar, India. p. 551–5.
121. Lee DG, Han DS. GAN-based two stage network for de-occlusion face image. In: 2024 IEEE International Conference on Consumer Electronics (ICCE); 2024 Jan 5–8. Las Vegas, NV, USA. p. 1–4.
122. Nelson A, Shaji R. A novel occluded face detection approach using Enhanced ORB and optimized GAN. Int J Wave, Multiresolut Inform Process. 2023;22(2):2350051. doi:10.1142/s0219691323500510.
123. Duan Q, Zhang L, Gao X. Simultaneous face completion and frontalization via mask guided two-stage GAN. IEEE Transact Circ Syst Video Technol. 2022;32(6):3761–73. doi:10.1109/tcsvt.2021.3111648.
124. Lv Y, Wang J, Gao G, Li Q. LW-DCGAN: a lightweight deep convolutional generative adversarial network for enhancing occluded face recognition. J Electron Imaging. 2024;33(5):053057. doi:10.1117/1.JEI.33.5.053057.
125. Yitong Zhou TL. Masked face restoration model based on lightweight GAN. Comput Mater Contin. 2025;82(2):3591–608. doi:10.32604/cmc.2024.057554.
126. Jabbar A, Li X, Iqbal MM, Malik AJ. FD-StackGAN: face de-occlusion using stacked generative adversarial networks. KSII Transact Inter Inform Syst. 2021;15(7):2547–67. doi:10.3837/tiis.2021.07.014.
127. Li C, Ge S, Zhang D, Li J. Look through masks: towards masked face recognition with de-occlusion distillation, MM ’20. New York, NY, USA: Association for Computing Machinery; 2020. p. 3016–24. doi:10.1145/3394171.3413960.
128. Man Q, Cho YI. Efficient face region occlusion repair based on T-GANs. Electronics. 2023;12(10):2162. doi:10.3390/electronics12102162.
129. Qiu H, Gong D, Li Z, Liu W, Tao D. End2End occluded face recognition by masking corrupted features. IEEE Transact Pattern Analy Mach Intellig. 2022;44(10):6939–52. doi:10.1109/tpami.2021.3098962.
130. Cong K, Zhou M. Face dataset augmentation with generative adversarial network. J Phys: Conf Ser. 2022;2218(1):012035. doi:10.1088/1742-6596/2218/1/012035.
131. Zhu C, Zheng Y, Luu K, Savvides M. In: Bhanu B, Kumar A editors. CMS-RCNN: contextual multi-scale region-based cnn for unconstrained face detection. Cham, Switzerland: Springer International Publishing; 2017. p. 57–79. doi:10.1007/978-3-319-61657-5_3.
132. Shin J, Kim D. Hybrid approach for facial feature detection and tracking under occlusion. IEEE Signal Process Lett. 2014;21(12):1486–90. doi:10.1109/lsp.2014.2338911.

133. Liu X, Graeser A. Robust face detection with eyes occluded by the shadow from dazzling avoidance system. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems; 2015 Sep 15–18; Gran Canaria, Spain. p. 2352–7.

134. El-Barkouky A, Shalaby A, Mahmoud A, Farag A. Selective part models for detecting partially occluded faces in the wild. In: 2014 IEEE International Conference on Image Processing (ICIP); 2014 Oct 27–30. Paris, France. p. 268–72.

135. Balasundaram A. Computer vision based detection of partially occluded faces. Int J Eng Adv Technol. 2020;9(3):2188–200. doi:10.35940/ijeat.c5637.029320.

136. Jain V, Learned-Miller E. FDDB: a benchmark for face detection in unconstrained settings. Amherst, MA, USA: University of Massachusetts Amherst; 2010. UM-CS-2010-009.

137. Yan J, Zhang X, Lei Z, Li SZ. Face detection by structural models. Image Vis Comput. 2014;32(10):790–9. doi:10.1016/j.imavis.2013.12.004.

138. Klare BF, Klein B, Taborsky E, Blanton A, Cheney J, Allen K, et al. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7–12. Boston, MA, USA. p. 1931–9.

139. Wang Z, Huang B, Wang G, Yi P, Jiang K. Masked face recognition dataset and application. IEEE Transact Biomet, Behav Iden Sci. 2023;5(2):298–304. doi:10.1109/tbiom.2023.3242085.

140. Yin X, Chen L. FaceOcc: a diverse, high-quality face occlusion dataset for human face extraction. arXiv:2201.08425. 2022.

141. Martinez A, Benavente R. The AR face database. In: CVC technical report no. 24. Barcelona, Spain: Universitat Autonoma de Barcelona; 1998.

142. Köstinger M, Wohlhart P, Roth PM, Bischof H. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops); 2011 Nov 6–13; Barcelona, Spain. p. 2144–51.

143. Mahbub U, Sarkar S, Patel VM, Chellappa R. Active user authentication for smartphones: a challenge data set and benchmark results. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS); 2016 Sep 6–9; Niagara Falls, NY, USA. p. 1–8.

144. Yucel MK, Bilge YC, Oguz O, Ikizler-Cinbis N, Duygulu P, Cinbis RG. Wildest faces: face detection and recognition in violent settings. arXiv:1805.07566. 2018.

145. ErakLn ME, Demir U, Ekenel HK. On recognizing occluded faces in the wild. In: 2021 International Conference of the Biometrics Special Interest Group (BIOSIG); 2021 Sep 15–17; Darmstadt, Germany. p. 1–5.

146. Lin S. A study on 3D human pose estimation with a hybrid algorithm of spatio-temporal semantic graph attention and deep learning. Inform Technol Cont. 2024;53(4):1042–59. doi:10.5755/j01.itc.53.4.37243.

147. He Y, Wan L. YOLOv7-PD: incorporating DE-ELAN and NWD-CIoU for advanced pedestrian detection method. Inform Technol Cont. 2024;53(2):390–407. doi:10.5755/j01.itc.53.2.35569.

148. Xu Y, Wei S, Yin J. Optimization of human posture recognition based on multi-view skeleton data fusion. Inform Technol Cont. 2024;53(2):542–53. doi:10.5755/j01.itc.53.2.36044.