ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor (RJIF): 8.4
IJAR 2025; 11(6): 270-275
www.allresearchjournal.com
Received: 01-04-2025
Accepted: 05-05-2025

Sapna Alha
Assistant Professor, Department of Computer Science Engineering, Shekhawati Institute of Engineering and Technology, Sikar, Rajasthan, India

Irfan Khan
Assistant Professor, Department of Computer Science Engineering, Shekhawati Institute of Engineering and Technology, Sikar, Rajasthan, India

Corresponding Author:
Sapna Alha
Assistant Professor, Department of Computer Science Engineering, Shekhawati Institute of Engineering and Technology, Sikar, Rajasthan, India

A comprehensive review of face detection using deep learning techniques

Sapna Alha and Irfan Khan

Abstract

Face detection is a crucial component of computer vision with applications ranging from biometric authentication and surveillance to social media and human-computer interaction. The primary objective of face detection is to accurately identify and localize human faces within digital images or video streams. Over the years, face detection has evolved from traditional hand-crafted feature-based methods to highly robust and efficient deep learning-based models. This shift has been driven by the need for higher accuracy, better generalization to diverse real-world conditions, and real-time processing capabilities. This review paper presents a comprehensive survey of face detection techniques, with a specific focus on advancements powered by deep learning. The paper begins with an overview of classical methods including Viola-Jones, HOG-SVM, and landmark-based detectors. It then delves into deep learning models such as Convolutional Neural Networks (CNNs), Region-based CNNs (R-CNNs), Single Shot Detectors (SSDs), YOLO (You Only Look Once), and Multi-task Cascaded Convolutional Networks (MTCNN), highlighting their architectures, performance, and deployment efficiency. Additionally, we explore the datasets commonly used for training and evaluation, along with the key performance metrics that benchmark model accuracy and robustness. The paper also outlines various real-world applications, current challenges like occlusion and lighting variation, and discusses future research directions including transformer-based detection and edge AI. By systematically reviewing the evolution and current state of face detection technologies, this paper aims to serve as a valuable resource for researchers, developers, and practitioners interested in the intersection of deep learning and face analytics.

Keywords: Face detection, deep learning, convolutional neural networks (CNN), R-CNN, YOLO, RetinaFace, benchmark datasets

Introduction

Face detection has become a pivotal task in modern computer vision, serving as the foundation for numerous real-world applications such as facial recognition, emotion analysis, identity verification, surveillance, and human-computer interaction. Unlike traditional object detection, face detection is a more complex and nuanced problem due to the variety of facial poses, expressions, lighting conditions, occlusions, and background clutter. With the advancement of artificial intelligence, particularly deep learning, face detection systems have reached unprecedented levels of accuracy and robustness [1]. Historically, face detection relied on hand-crafted features and machine learning techniques such as Haar cascades and support vector machines. These methods, while effective in controlled environments, failed to generalize well to real-world scenarios. The advent of deep learning, especially convolutional neural networks (CNNs), has transformed the field, enabling the development of models that can learn features automatically from large-scale datasets and perform robust detection under diverse conditions [1].

Deep learning models such as Multi-task Cascaded Convolutional Networks (MTCNN), Single Shot Detector (SSD), Faster R-CNN, and the YOLO family have shown exceptional performance in face detection tasks. These models leverage hierarchical feature extraction, transfer learning, and real-time inference to achieve high accuracy and speed. Furthermore, face detection has been extended into more complex tasks, such as face alignment, landmark detection, and 3D face modeling, where deep learning continues to play a significant role [2]. This review aims to provide a comprehensive overview of face detection techniques with an emphasis on deep learning methods.

It begins by outlining the classical approaches that predate deep learning, followed by an in-depth analysis of modern deep learning-based techniques. The paper also covers popular models, benchmark datasets, evaluation metrics, applications, current challenges, and future research directions in the field of face detection [2].

By the end of this paper, readers will have a thorough understanding of how deep learning has revolutionized face detection, what the current state-of-the-art methods are, and where the field is headed in the future [2].

Background of Face Detection

Face detection is a computer vision task that aims to locate human faces within digital images or videos. It serves as the fundamental step for various downstream applications, including facial recognition, emotion analysis, facial landmark detection, and face tracking. The significance of face detection has surged due to the proliferation of biometric systems, smart surveillance, and user-centric computing devices [3]. The challenge of face detection stems from a wide range of variations in facial appearance, pose, lighting, expression, occlusion, and image quality. Unlike other object detection tasks, face detection must cope with these complexities in real-time scenarios and across diverse populations.

Historically, face detection emerged in the 1990s with rule-based and statistical models that relied heavily on hand-crafted features. The most influential early work was the Viola-Jones detector introduced in 2001, which used Haar features and AdaBoost for rapid object detection. Although it revolutionized the field and became widely adopted, it had limitations in handling complex backgrounds and non-frontal face orientations [3]. The subsequent decade saw improvements with the integration of statistical methods such as Principal Component Analysis (PCA), Support Vector Machines (SVMs), and Local Binary Patterns (LBP). However, these methods still lacked robustness against occlusion and illumination variation [3].

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), marked a significant turning point. CNNs possess the ability to automatically learn hierarchical feature representations, making them well-suited for tasks like face detection that involve subtle patterns and high intra-class variation [4]. As computational resources, labeled datasets, and algorithmic innovations matured, deep learning-based face detection models surpassed classical methods in both accuracy and speed. Today, face detection systems must handle real-time video streams, accommodate edge computing environments, and maintain fairness across demographic groups—tasks that were unimaginable for earlier techniques [4] Understanding the evolution from classical to deep learning-based methods helps us appreciate the design choices, limitations, and future direction of modern face detection frameworks.

Classical Methods

Before the advent of deep learning, face detection was largely dominated by classical computer vision techniques that relied on handcrafted features and conventional machine learning algorithms. These approaches were grounded in human intuition about facial structures and employed statistical models to differentiate face and non-face regions in images. Although these methods laid the groundwork for modern advancements, they had notable limitations in terms of robustness and adaptability under challenging conditions such as occlusions, varying lighting, and extreme facial poses. Among the many classical techniques, a few stood out for their influence and effectiveness [5]. One of the most iconic and widely adopted methods in classical face detection is the Viola-Jones algorithm, introduced in 2001 by Paul Viola and Michael Jones. It was the first framework capable of real-time face detection, making it a landmark achievement in computer vision. The core innovation of Viola-Jones was its use of Haar-like features, which are simple rectangular features that capture contrast information in localized regions of the image—ideal for detecting patterns like eyes, nose bridges, and mouth contours. These features were calculated using an integral image representation, allowing extremely fast computation. To manage the vast number of potential features, the algorithm employed AdaBoost, a machine learning technique that selects the most relevant features and combines weak classifiers into a strong one. Moreover, Viola-Jones utilized a cascade of classifiers, where simple classifiers rapidly discard non-face regions, and more complex classifiers are applied only to promising regions. Despite its real-time speed and high performance on frontal faces, the algorithm struggled with non-frontal poses, occlusions, and significant variations in illumination [5].

Another important classical method is the combination of Histogram of Oriented Gradients (HOG) features with a Support Vector Machine (SVM) classifier. HOG was developed to capture the distribution of edge orientations within an image. It works by dividing the image into small cells and computing histograms of gradient directions within each cell. This technique effectively encodes the shape and structure of objects, including human faces. When paired with a linear SVM classifier, HOG features enabled the system to distinguish between face and non-face image patches. This method marked an improvement over Viola-Jones in terms of handling pose variations and facial expressions. However, HOG + SVM systems still faced difficulty in detecting partially occluded faces and adapting to scale changes or cluttered backgrounds [6]. Local Binary Patterns (LBP) emerged as another significant technique in the classical face detection era. LBP is a texture descriptor that compares each pixel with its surrounding neighborhood and encodes the result as a binary number. This approach captures fine-grained texture information, which is useful in characterizing facial skin patterns and expressions. LBPs are particularly robust to changes in lighting, making them suitable for detection in environments with varying illumination. Typically, LBP features were used in conjunction with sliding window techniques and classifiers like SVM or decision trees. While LBP-based detectors offered better performance in certain conditions, they were sensitive to scale variations and required precise face alignment for optimal accuracy [6].

In addition to feature-based approaches, template matching and Principal Component Analysis (PCA)-based methods were also explored in early face detection systems. Template matching involved comparing input images to a set of predefined facial templates using similarity measures. Although simple, this method was computationally expensive and performed poorly when faces deviated from the average template. PCA-based methods, such as the Eigenfaces approach, represented facial images in a reduced-dimensional subspace. These approaches assumed that face images lie on a low-dimensional manifold and aimed to reconstruct input images using linear combinations of principal components. While PCA helped reduce noise and dimensionality, it was limited in handling intra-class variability, such as differences in age, facial hair, and facial expressions. Moreover, the computational burden and sensitivity to alignment reduced its practicality in real-world applications [7]. In summary, classical face detection methods made substantial contributions to the field by introducing fundamental concepts such as feature extraction, boosting, and dimensionality reduction. However, they relied heavily on manual feature design and often lacked robustness against real-world variations. The limitations of these approaches ultimately paved the way for the adoption of deep learning techniques, which brought significant improvements in accuracy, generalization, and flexibility.

Deep Learning for Face Detection

The evolution of face detection has been revolutionized by the rise of deep learning, a branch of machine learning that enables models to automatically learn features from data rather than relying on handcrafted representations. The most significant development in this transformation has been the introduction and adoption of Convolutional Neural Networks (CNNs), which are exceptionally well-suited for image-related tasks due to their ability to learn spatial hierarchies and capture local patterns. In face detection, CNN-based models now dominate state-of-the-art performance benchmarks, providing highly accurate and robust solutions even under complex conditions such as extreme poses, occlusions, low lighting, and diverse backgrounds [8]. One of the primary reasons for the widespread success of deep learning in face detection is its clear advantage over classical methods. Traditional systems, which relied on manually engineered features such as Haar cascades or HOG descriptors, often struggled to generalize across varying environments and typically failed under non-ideal conditions. Deep learning, in contrast, enables automatic feature learning, allowing neural networks to discover the most relevant and discriminative features directly from raw pixel data. This leads to higher detection accuracy, especially on challenging datasets such as WIDER FACE and FDDB. Furthermore, deep learning-based models generalize better to unseen data and are inherently more adaptable to different tasks and domains. With the support of Graphics Processing Units (GPUs) and hardware accelerators, many of these models also achieve real-time performance, making them viable for live applications such as video surveillance and smartphone authentication [8].

A typical deep learning-based face detection pipeline includes several key components. At the heart of the system lies the backbone network, which is responsible for extracting multi-scale, hierarchical features from input images. Common backbones include pre-trained CNN architectures such as VGG16, ResNet, and MobileNet, which are chosen based on trade-offs between accuracy and computational efficiency. Once the features are extracted, the next stage involves a detection head or region proposal mechanism. There are two main families of detection architectures: two-stage detectors such as R-CNN, Fast R-CNN, and Faster R-CNN, which first generate region proposals and then classify them; and single-stage detectors like YOLO (You Only Look Once), SSD (Single Shot Detector), and RetinaNet, which perform detection in a single forward pass. Single-stage models are generally faster and suitable for real-time applications, while two-stage models often provide higher accuracy [9]. Another critical aspect of these systems is the design of loss functions, which guide the model during training. Typically, a combination of losses is used: classification loss (commonly cross-entropy) determines whether a region contains a face or not, while bounding box regression loss (often smooth L1 or IoU-based loss) refines the predicted location of the face. Some advanced models also include landmark localization loss, which helps the model predict key facial points such as eyes, nose, and mouth—useful for tasks like face alignment and pose estimation [9].

To maximize the performance and generalization ability of deep learning models, several training strategies are employed. Data augmentation is essential in this context, as it introduces variability into the training set through operations like rotation, scaling, flipping, blurring, and occlusion simulation. This helps the model learn to be robust against real-world challenges. Another widely used strategy is transfer learning, where models pre-trained on large datasets like ImageNet are fine-tuned for face detection tasks. This not only reduces training time but also improves performance, particularly when labeled face datasets are limited. Additionally, multi-task learning has gained popularity, where a single model is trained to perform multiple tasks—such as face detection, facial landmark localization, and head pose estimation—simultaneously. This joint training paradigm enhances the model's contextual understanding and often leads to improved results across all tasks. In terms of performance, deep learning methods have demonstrated remarkable success. On challenging benchmarks like FDDB (Face Detection Data Set and Benchmark) and WIDER FACE, deep models have consistently achieved detection accuracies exceeding 98%. These benchmarks present a wide variety of challenges, including small face sizes, dense crowd scenes, occlusions, and different lighting conditions, making them ideal for evaluating model robustness. Furthermore, lightweight CNNs such as MobileNet-SSD and EfficientDet allow for real-time inference even on mobile devices and embedded systems, without significant sacrifices in accuracy [10].

A number of noteworthy deep learning-based face detection models have been developed over recent years. MTCNN (Multi-task Cascaded Convolutional Neural Network) is one such model that performs face detection and facial landmark localization in a cascaded framework of three networks. Its multi-task nature improves both speed and precision, especially in face alignment. Another significant model is YOLOv5-face, a customized variant of the YOLOv5 object detection model, specifically fine-tuned for face detection in crowded or cluttered scenes. It achieves a strong balance between accuracy and speed. Similarly, RetinaFace is a state-of-the-art single-stage detector that enhances performance through joint face detection and facial landmark regression, achieving excellent results on both easy and hard subsets of WIDER FACE. It also incorporates context modules and attention mechanisms to better focus on face regions [10].

From a deployment perspective, deep learning models are increasingly being optimized for use in constrained environments. With tools like TensorFlow Lite, ONNX Runtime, and OpenCV DNN, face detection models can be quantized and pruned for efficient deployment on mobile apps, IoT devices, and edge computing platforms. These optimizations maintain a balance between performance and computational demand, enabling broader adoption in consumer electronics, surveillance systems, and real-time video analysis platforms [11]. Despite the tremendous progress brought by deep learning, some challenges still persist. Computational cost remains a concern, particularly when deploying on resource-limited hardware. Deep learning models also raise issues related to fairness and bias, as they may underperform on underrepresented demographic groups due to training data imbalances. Additionally, robustness to adversarial attacks—where small, imperceptible perturbations to the input image can cause misdetections—remains an open research problem. These issues underscore the need for continual refinement and ethical evaluation of face detection systems [11].

In conclusion, deep learning has fundamentally changed the landscape of face detection, introducing models that are more accurate, scalable, and capable of handling a diverse array of real-world scenarios. By leveraging CNNs, end-to-end training, and advanced optimization techniques, modern face detection systems now play a critical role in numerous applications—from biometric security and healthcare to social media and smart surveillance. While there are challenges yet to be addressed, the deep learning era has undoubtedly marked a new and promising chapter in the history of face detection.

Popular Architectures and Models

Over the years, various deep learning architectures have been developed and refined specifically for the task of face detection. These models are designed to balance accuracy, speed, and computational efficiency based on their intended application, ranging from high-precision surveillance to real-time mobile applications. The architectural choices and design optimizations have made a significant impact on the ability of these systems to detect faces under challenging conditions such as varying lighting, pose, occlusion, and scale. The R-CNN family represents one of the earliest and most influential advancements in object detection, including face detection. The original R-CNN (Regions with CNN features) architecture begins by using selective search to propose regions of interest (RoIs), which are then classified using a CNN. Although R-CNN achieved high detection accuracy, it suffered from slow inference speed due to its two-stage processing and lack of end-to-end training. Fast R-CNN addressed this by extracting convolutional features from the entire image once and then applying RoI pooling, allowing for faster and more accurate predictions. Faster R-CNN further improved upon this by introducing the Region Proposal Network (RPN), which allowed region proposals to be generated and refined within the same CNN architecture. Despite its high accuracy, the R-CNN family is typically resource-intensive and less suitable for real-time applications [12].

In contrast, single-stage detectors such as YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) offer significant speed advantages by skipping the region proposal stage. These models treat object detection as a regression problem, predicting bounding boxes and class probabilities directly from full images in a single forward pass. YOLO, particularly versions YOLOv3, YOLOv4, and YOLOv5, has been widely adapted for face detection due to its ability to balance speed and accuracy effectively. While these models are well-suited for real-time applications, they may struggle with detecting small or heavily occluded faces compared to two-stage detectors. SSD also provides competitive performance with the advantage of multi-scale detection via feature maps at different levels of abstraction, making it robust to scale variation. The MTCNN (Multi-task Cascaded Convolutional Networks) is another widely-used architecture specifically designed for face detection and alignment. It employs a cascade of three networks—P-Net, R-Net, and O-Net—to progressively refine face candidates while also predicting facial landmarks. This approach makes MTCNN particularly effective for preprocessing tasks such as face alignment, which is essential for applications like facial recognition and emotion detection [13].

RetinaFace is a more recent innovation in face detection. It is a single-stage detector that leverages the power of Feature Pyramid Networks (FPN) and strong backbone networks like ResNet and MobileNet. RetinaFace not only predicts the face bounding boxes but also includes five facial landmarks, enabling it to handle complex scenarios with occlusions and varied facial orientations. It has been shown to achieve state-of-the-art results on multiple benchmarks. DSFD (Dual Shot Face Detector) introduces a dual-shot structure that applies two detection stages at different feature levels. This approach enables it to handle faces at various scales with greater accuracy. DSFD leverages multi-level supervision to improve both the detection of small faces and the localization precision [13].

FaceBoxes is designed with speed as a primary goal. It simplifies the network design by using lightweight convolutional structures, allowing it to achieve real-time performance even on devices with limited computational resources. Although FaceBoxes may not outperform heavier architectures in terms of accuracy, its speed and simplicity make it highly suitable for embedded systems and edge devices. Each of these models reflects a different trade-off between speed, accuracy, and computational demand. Therefore, the selection of a face detection model depends largely on the application scenario. High-accuracy surveillance systems may prefer Faster R-CNN or RetinaFace, whereas mobile or real-time video applications might benefit more from YOLO variants or FaceBoxes due to their lower latency and computational overhead [13].

Benchmark Datasets and Evaluation Metrics

Benchmark datasets and evaluation metrics are crucial for the development and validation of face detection algorithms. They provide standardized environments that allow researchers to compare performance across different models and identify areas of improvement. These benchmarks incorporate a wide range of challenges including pose variation, occlusion, lighting conditions, and scale, helping ensure that face detection models are robust and generalizable [14]. One of the most well-known datasets in face detection research is the FDDB (Face Detection Data Set and Benchmark). It consists of 2,845 images containing a total of 5,171 faces. The faces are annotated using elliptical bounding boxes, which poses a unique challenge compared to standard rectangular annotations. FDDB is known for its realistic scenarios and variety of face scales, making it a strong benchmark for evaluating both traditional and deep learning-based face detectors [14].

Another highly influential dataset is WIDER FACE, which contains over 32,000 images and approximately 393,000 labeled faces. This dataset is particularly comprehensive as it includes faces with extreme variations in pose, expression, occlusion, and illumination. WIDER FACE is commonly divided into three difficulty levels: Easy, Medium, and Hard, based on face size and occlusion levels. The vast number of samples and diversity in conditions have made it a standard for evaluating the generalization capability of face detectors [14]. The AFW (Annotated Faces in the Wild) dataset is designed to represent real-world scenarios where faces appear in cluttered and uncontrolled environments. It contains annotated images collected from Flickr, providing a test bed for models aiming to operate in naturalistic settings. Although smaller than WIDER FACE, AFW is still useful for qualitative comparisons and performance visualization [14].

With the rise in face mask usage due to global health events, the MAFA (Masked Faces) dataset was created to specifically evaluate the ability of face detection models to recognize masked faces. This dataset includes over 30,000 images, focusing on occluded lower facial regions. It has become increasingly important in applications such as health surveillance and security monitoring during pandemics. In terms of evaluation metrics, multiple quantitative measures are used to assess the performance of face detection models. Precision, Recall, and F1-Score are foundational metrics used to assess the balance between false positives and false negatives. Precision measures the percentage of correctly predicted faces out of all predicted faces, while recall assesses the percentage of actual faces correctly detected. F1-Score is the harmonic mean of precision and recall, providing a single measure that balances both [15].

Average Precision (AP) is a more comprehensive metric that calculates the area under the precision-recall curve. It is particularly useful for evaluating performance across different confidence thresholds. Intersection over Union (IoU) is used to quantify the overlap between the predicted bounding box and the ground truth bounding box. A higher IoU indicates better localization performance. For more comprehensive evaluations, mean Average Precision (mAP) is used, which calculates the average AP across different object categories or difficulty levels. In face detection, mAP helps evaluate the consistency of detection across different scales and occlusion levels.

Together, these datasets and evaluation metrics form the backbone of modern face detection research. They ensure that proposed methods are rigorously tested and that improvements are clearly measurable. Moreover, consistent benchmarking facilitates the identification of model limitations, guiding future innovations in architecture design and training methodology.

Conclusion

Face detection has evolved significantly with the advent of deep learning, transitioning from hand-crafted features and traditional classifiers to sophisticated neural network architectures capable of robust and real-time performance. This review has highlighted the fundamental principles behind face detection, tracing its development from classical approaches to modern deep learning-based solutions. A detailed exploration of key components such as preprocessing techniques, CNN architectures, loss functions, and optimization strategies reveals how each contributes to improving detection performance.

The discussion on popular architectures—including R-CNN variants, YOLO, SSD, MTCNN, RetinaFace, and others—demonstrates that no one-size-fits-all solution exists. Each model presents trade-offs in speed, accuracy, and computational efficiency, making it essential to select the appropriate architecture based on specific application requirements. Furthermore, benchmark datasets such as WIDER FACE, FDDB, and MAFA, along with standardized evaluation metrics like IoU, mAP, and precision-recall curves, provide a consistent framework for model development and comparison.

Despite significant progress, challenges remain in detecting small, occluded, or masked faces in unconstrained environments. Future research is expected to focus on lightweight models for mobile deployment, multimodal fusion (e.g., thermal and RGB face detection), and increased robustness through self-supervised or few-shot learning techniques. Moreover, ethical considerations, including fairness, bias mitigation, and privacy protection, are becoming increasingly important as face detection technologies are widely adopted in surveillance, authentication, and human-computer interaction.

In summary, deep learning has revolutionized face detection, enabling higher accuracy, adaptability, and deployment versatility. With continued innovation in model design, training strategies, and real-world testing, face detection systems are poised to become even more efficient, accessible, and reliable across diverse domains and applications.

References

1. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2001 Dec; Kauai, HI, USA. p. 511-518.
2. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2005 Jun; San Diego, CA, USA. p. 886-893.
3. Ahonen T, Hadid A, Pietikäinen M. Face description with local binary patterns: Application to face recognition. IEEE Trans Pattern Anal Mach Intell. 2006 Dec;28(12):2037-2041.
4. Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell. 1997 Jul;19(7):711-720.
5. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2014 Jun; Columbus, OH, USA. p. 580-587.
6. Girshick R. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2015 Dec; Santiago, Chile. p. 1440-1448.
7. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017 Jun;39(6):1137-1149.

8. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, *et al.* SSD: Single shot multibox detector. In: Proceedings of the European Conference on Computer Vision (ECCV); 2016 Oct; Amsterdam, Netherlands. p. 21-37.
9. Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767 [cs.CV]. 2018.
10. Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multi-task cascaded convolutional networks. IEEE Signal Process Lett. 2016 Oct;23(10):1499-1503.
11. Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: Single-shot multi-level face localisation in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun; Seattle, WA, USA. p. 5203-5212.
12. Li S, Wang X. WIDER FACE: A face detection benchmark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun; Las Vegas, NV, USA. p. 5525-5533.
13. Jain V, Learned-Miller E. FDDB: A benchmark for face detection in unconstrained settings. University of Massachusetts Amherst, Technical Report UM-CS-2010-009; 2010.
14. Jiang H, Learned-Miller E. Face detection with the Faster R-CNN. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG); 2017 May; Washington, DC, USA. p. 650-657.
15. Wang Y, Ji X, Liang Z, Zhang G. FaceBoxes: A CPU real-time face detector with high accuracy. In: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN); 2017 May; Anchorage, AK, USA. p. 1-7.