Analyzing different Techniques for Face Detection and Recognition

Sahil Bisht
Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India

Sonal Fatangare
Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India

Apeksha Patil
Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India

Aakanksha Panadi
Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India

Dev Kulkarni
Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India

ABSTRACT
Face detection and recognition are two critical tasks in the realm of computer vision with a diverse array of real-world applications, encompassing surveillance, security, human-computer interaction, and biometric authentication systems. This survey paper offers an exhaustive review of the advancements in face detection and recognition techniques over the past years. The paper first delves into the fundamental concepts and challenges associated with face detection, like changing lighting scenarios, obstructed views, and pose fluctuations. Various traditional and deep learning face detection algorithms are then analyzed, highlighting their strengths and limitations. Subsequently, this study explores the intricacies of face recognition, emphasizing the significance of feature extraction, representation, and matching methods. It discusses the evolution from classical methods, such as eigenfaces and Fisherfaces, to cutting-edge deep learning techniques, such as advanced convolutional neural networks. Further this paper investigates the issues related to facial expression variations, aging, and demographic biases in face recognition systems. Finally, the survey concludes with a comprehensive comparative analysis of various benchmark datasets, evaluation metrics, and performance measures used in assessing the efficacy of algorithms. It also highlights the future research directions and emerging trends in the field, including multimodal fusion, cross-modal face recognition, and the integration of deep learning with generative models for robust and efficient face analysis.

Keywords
Face Recognition, Face Detection, Convolutional Neural Network, Deep Learning, Industrial Application.

1. INTRODUCTION
In recent years, the proliferation of digital imaging technologies and the exponential growth of data have accelerated the development and deployment of sophisticated computer vision techniques. Among these, face detection and recognition have emerged as crucial components, playing an integral role in numerous real-world applications. The automated identification and acknowledgment of human faces in visual data have catalyzed a multitude of inventive solutions, reshaping industries such as law enforcement, healthcare, marketing, and entertainment.

Face detection and recognition present unique challenges stemming from the intricate nature of human faces, including differences in body positioning, changes in ambient lighting, diverse facial gestures, obstructions, and the effects of aging. [18]. Over the years, researchers and practitioners have made significant strides in addressing these challenges, leveraging both traditional computer vision techniques and cutting-edge deep learning methodologies. These advancements have led to the development of robust algorithms which can accurately detect and recognize faces across diverse environments and conditions [22].

Moreover, [35] the widespread adoption of facial recognition technologies has sparked noteworthy ethical apprehensions, especially regarding issues of privacy and surveillance. The ethical implications and societal impact of deploying facial recognition systems in public spaces have prompted critical discussions surrounding the need for regulatory frameworks and responsible AI practices to ensure the protection of individual privacy and mitigate potential discriminatory biases

In light of these developments, this survey paper seeks to offer a thorough examination of the latest advancements, obstacles, and ethical aspects in the fields of face detection and recognition. By examining the evolution of methodologies,[30] benchmark datasets, evaluation metrics, and emerging trends, this paper aims to provide valuable perspectives on the current landscape and the future prospects of face detection and recognition technologies. The objective of this survey is to enhance comprehension of the complexities associated with developing robust and ethical face detection and recognition systems, thereby contributing to the advancement of responsible and inclusive AI applications in the realm of computer vision.

2. LITERATURE SURVEY
In this segment, the literature survey provides a concise exploration of face detection and recognition algorithms. Covering academic research and technological advancements, it offers insights into the evolution, trends, and challenges of

this dynamic field. From classical image processing methods to the latest in deep learning, the survey functions as a concise reference for researchers and practitioners navigating the diverse terrain of computer vision research centered around faces.

Muhammad Zeeshan Khan, et al [1] addressed the surge in data from internet-connected devices, proposing an algorithm for detecting and recognizing faces based on Convolutional Neural Networks (CNN). The algorithm outperformed traditional methods like SIFT and SURF. Validated in a smart classroom, it achieved 97.9% accuracy in student attendance. Leveraging IoT and edge computing, outperforming current architectures, this system demonstrated superior performance in terms of data latency and real-time responsiveness.

Billy Peralta, et al [2] explored the importance of gender recognition in cQA platforms, emphasizing its role in achieving gender parity and enhancing user engagement. Despite the nonmandatory gender field in enrollment forms, the research focused on extracting gender information from user activities, particularly profile pictures. Assessing three image processing techniques, the best configuration, Inception-ResNet-50, achieved an 81.68% accuracy, with a focus on determining silhouette edges. The study anticipates its findings will contribute to the design of efficient multi-modal strategies in cQA platforms.

Dalin Wang and Rongfeng Li [3] addressed limitations in face recognition models related to factors such as masks and glasses that cause occlusion were addressed by the newly introduced Occlusion-Aware Module Network (OAM-Net). outperformed other methods by incorporating an occlusion-aware sub-network with adaptive convolutional kernel weights and enhanced generalization performance and accuracy were achieved through the incorporation of a sub-network that is sensitive to key regions, employing a Spatial Attention Residual Block. The application of a meta-learning-based strategy further improved overall performance. The experimental outcomes validated OAM-Net's excellence in recognizing occluded faces, underscoring its practical utility.

Lina Li, et al [4] introduced ET-YOLOv5s, a solution aimed at enhancing student behavior recognition in classrooms. By combining ESRGAN for image enhancement and a tiny object detection module with YOLOv5s, the system demonstrated improved accuracy in detecting various student behaviors simultaneously. Experimental results showcased superior performance over YOLOv4, YOLOv5s, and other tiny target detection algorithms. The solution effectively addressed limitations in single-student-focused methods, offering enhanced detection capabilities for multiple students' behaviors in a classroom environment. Rajashree Tripathy and R N Daschoudhury [5] implemented real-time face detection and tracking head pose through a Haar Classifier on a Raspberry Pi BCM2835 CPU processor. Leveraging the GPU-based architecture and a 5-megapixel OV5647 CMOS image sensor, the system achieved high-definition video capture, supporting resolutions up to 1080p at 30fps. The utilization of the SimpleCV and OpenCV libraries for detecting faces and tracking head poses., demonstrated successful results with 30fps accuracy under 1080p resolutions.

Dostdar Hussain, et al [6] proposed an automated system for face mask detection to address the challenges posed by the rapid spread of COVID-19. Utilizing deep learning algorithms for COVID-19 detection, they employed two models: a Deep Convolutional Neural Network (DCNN) and a transfer learning approach based on MobileNetV2, evaluating them on real-world datasets. MobileNetV2 demonstrated accuracies of 98% and 99% on the respective datasets, while DCNN achieved a consistent 97% accuracy. The findings indicated that MobileNetV2 is a highly accurate alternative to DCNN for effective face mask detection in public places.

Table 1. Deep Literature Survey of Current Technologies

Ref no:	Methodology used	Dataset used	Accuracy	Research
[1]	Deep unified model for Face Recognition based on Faster region CNN.	Wild (LFW) dataset.	97.9%	Successful in detecting faces from clear images.
[2]	Neural network models to identify genders, Grad-Cam heat maps	Face images of male, female and children (19,000)	81.68%	Grad-Cam heat maps focus their attention on outlining body silhouettes.
[3]	OAM-Net Spatial Attention Residual Block meta-learning-based strategy	AR Face, CASIA-Web Face Dataset	92.5% (occlusion ratio: 30%)	OAM-Net shows superior performance in areas of privacy and data security.
[4]	YOLOv5s, ET-YOLOv5s, ESRGAN.	Surveillance videos in Changchun University.	0.964 precision	Successfully detected 11 kinds of students' behaviours in classrooms.
[5]	Real time Face detection, tracking Haar Classifier.	Viola and Jones data set.		Tracked the lost object under dynamic environment and depth analysis of face detection using image sensors.

[6]	Deep CNN (DCNN) , MobileNetV2 for face mask detection.	2500 images dataset from PyImage search Reader Prajna handary.	MobileNet V2: 98% (dataset-1) and 99% (dataset-2), DCNN: 97% on both datasets.	MobileNet V2 model is an alternative to the DCNN model.
[7]	ML tools TensorFlow, Keras, OpenCV and Scikit-Learn.	Face dataset (with/without mask) from aggle.com with 3,832 images.		Identifies faces within an image or video and subsequently assesses whether a mask is present or absent.
[8]	Wi-Fi channel state , 1-D CNN, LSTM, CLSTM	Wi-Fi CSI data (Sample rate: 1,000 Hz.)	92%	Processing data to categorize mixed types for facilitating identification of one or more subjects.
[9]	SVC, LDA, KNN, DT, logistic regression, VGG16, AlexNet, ResNet50	FERET dataset (1208 subjects)	96%	Impacts of genders and ages across algorithms.
[10]	IoT integration of (ICT), (PiDiNet), Stochastic Gradient Descent		Face feature extraction: 96% Face verification: 97%	Utilizing the registered ID of the driver to send messages and notifications pertaining to slot information.
[11]	CNN, RESNET 50, VGG16, SVM	ImageNet dataset	92.7%	Prediction-based facial recognition systems are possible.
[12]	CNN , SVM classifier	Photographs of children.	Classification: 99.41%	Creating an extensive dataset, extending identification for video.
[13]	CNN, SVM, Decision Tree, Logistic Regression, KNN.	100 images of different objects, nature, bike and car images.	93.2%	Enhance the precision and resilience of video recognition systems.
[14]	Product Quantization-based hash look-up, FaceNet, ArcFace, VGG-Face2.	LFW database, FERET, FEI.	95%	More stable hash generation may be achieved by combining PQ and binary representations and clustering in Hamming space

3. METHODOLOGY

3.1 Face Detection Methods

Detecting faces is like breaking down a picture into two parts: one with faces and the other without. It looks at images or videos and figures out where the faces are. The system identifies faces and supplies the coordinates of a bounding box for each detected face, if present. The face detection process typically begins by initially searching for eyes, which are relatively easier to discern in a human face, before progressing to the detection of eyebrows, mouth, nose, and other significant facial features. In a typical face detection system, you give it a picture, and it goes through some initial steps to make the picture better and remove any unwanted stuff. Then, it decides if there is a face or not based on what it learned before. Finally, it shows where the face is in the picture. The detection process exhibits significant variability, primarily stemming from the diverse methodologies employed by various face detection algorithms.

The face detection cascade introduced by Viola and Jones (Viola-Jones object detection framework [30]) employs Haar-Like features and AdaBoost for training cascaded classifiers, delivering effective performance.

However, there are quite a few instances [26] where this algorithm fails in extreme background conditions and face orientations limited to frontal faces, not effective for profile faces or different angles.

3.2 Face Recognition Methods

Face recognition methods are techniques and algorithms created to recognize and authenticate individuals through analysis of their facial characteristics. Various approaches are utilized for face recognition [36]:

1. Holistic Related Techniques:
   Employing the entire facial area as input, Comprehensive approaches such as Eigenfaces, PCA, LDA and Independent Component Analysis are prominent examples. These methods leverage facial features for natural system feedback.

2. Structural Techniques:
   This method includes the extraction of specific facial features like eyes, nose, and mouth before inputting their locations and associated measurements into a structural classifier. An essential consideration is "recovery," where the system strives to recover obscured features resulting from substantial alterations.
3. Hybrid Techniques:
   These systems combine components from both holistic and feature-based methods, frequently incorporating 3D images. This allows the system to discern facial contours and features, enhancing overall recognition accuracy. The 3D process typically involves Placement Recognition, Orientation, Measurement, Representation, and Matching stages. This encompasses capturing an authentic face, establishing head direction and size, assigning dimensions to facial features, converting the template to code, and conducting matching and research based on the acquired information in the existing database.

4. ALGORITHMS

The prevalence of the ANN and Viola-Jones algorithms is notably higher, constituting approximately 40% of the total usage. Authors commonly utilize the open-source library OpenCV as the primary database, given its advantageous features for conducting experiments. Moreover, OpenCV supports several algorithms like ANN, Viola-Jones, Eigenface, Fisherface, and LBPH, simplifying the development of various solutions. The precise determination of the most frequently employed method remains inconclusive, particularly in the context of different stages in real-time facial recognition.

4.1 Viola-Jones

In 2001, Paul Viola and Michael Jones introduced the Viola-Jones algorithm [42] which stands out for its low computational cost and high real-time performance in facial recognition. This algorithm relies on HAAR filters, utilizing HAAR features to represent images in a feature space, extracting unique features of faces or different objects. [43]. It is implemented in libraries such as OpenCV and OpenBR.

The facial recognition process using the Viola-Jones algorithm involves three main steps. Firstly, HAAR filters are employed to create a spatial image that covers the entire image. The formula for a Haar-like feature is represented as follows:

$$H(x, y, w, h) = \Sigma \text{ (white region) pixel value} - \Sigma \text{ (black region) pixel value}$$

In the second phase, training is conducted utilizing the Boosting classification technique, specifically AdaBoost. The objective in this stage is to identify and extract the most pertinent features from the integral image. The concluding step involves constructing the HAAR Cascade, manifested as a tree structure, alternatively termed cascade classifiers [43].

The Viola-Jones algorithm in OpenCV is often referred to as HAAR Cascade, leveraging HAAR features as masks that modify brightness levels to describe an object [44]. These masks capture variances in various directions and magnitudes, then undergo training using the AdaBoost algorithm to produce classifiers individually, each corresponding to a specific HAAR feature.

Viola and Jones [42] introduced three fundamental types of masks in their work:

1. Attributes of two rectangles: the numerical value represents the disparity between the totals of pixels in each rectangle, with equal area and adjacency.
2. Attributes involving three rectangles: the numerical value represents the product of the weight-adjusted disparity between the outer and inner rectangles, accounting for variations in their respective areas.
3. Attributes of four rectangles: the numerical value represents the variance between diagonal pairs of rectangles.

In summary, Viola-Jones, is an object detection algorithm utilizing combination of classifiers.

4.2 Eigenface

The Eigenface algorithm is a technique that involves a set of eigenvectors derived from the covariance matrix of a collection of faces. Described in [39], this method aims to capture characteristics independent of the geometric shapes of facial features like eyes, mouth, nose, and ears. It leverages facial representation information to achieve this goal.

Eigenface utilizes the Principal Component Analysis (PCA) algorithm for dimensionality reduction, a process that proves advantageous in decreasing data magnitude and adjusts the quantity of images in the dataset, is a key optimization strategy. In the context of facial recognition, the application of Karhunen-Loeve methods, commonly known as PCA, is specifically referred to as Eigenface. [40].

Principal Component Analysis (PCA), guided by statistical analysis of existing redundancy and variance in the data, efficiently diminishes dimensionality while preserving the fundamental information. As emphasized in, the Eigenface algorithm functions as an appearance-based approach. This means it doesn't necessitate prior knowledge about what is being recognized. A noteworthy aspect is that the algorithm actively seeks the main components during recognition, namely the eigenvalues that characterize an individual's facial features.

While the Eigenface algorithm has its advantages, it's essential to recognize its limitations. It is responsive to changes in lighting conditions and specific noise types, which can potentially undermine its effectiveness and precision, as elaborated in Ref. [41].

4.3 Fisherface

The Fisherface algorithm is an advancement of the Eigenface algorithm [46]. Introduced by Peter Belhumeur, João Hespanha, and David Kriegman, the algorithm aims to enhance the discriminatory power of Eigenfaces by considering class-specific information. Fisherface addresses the limitations of Eigenfaces, particularly its sensitivity to variations in lighting conditions and noise.

The algorithm utilizes Linear Discriminant Analysis (LDA). LDA aims to optimize the ratio between the scatter among different classes and the scatter within each class. The criterion here is designed to enhance the distinction between various classes, rendering Fisherface especially proficient in tasks related to recognizing faces.

In the Fisherface algorithm, the images in the training set are first preprocessed, often involving normalization and centering [46]. Then, the algorithm computes the mean face and the scatter matrices, capturing both within-class and between-class variations. The Fisherfaces are created by extracting the eigenvectors associated with the highest eigenvalues of the generalized eigenvalue problem.

In the process of recognition, a recently encountered face is mapped onto the Fisherface subspace, and its classification is determined using methods such as KNN classification or other appropriate techniques. Fisherface has demonstrated improved performance over Eigenfaces in scenarios with variations in lighting and other environmental factors.

In summary, the Fisherface algorithm enhances facial recognition by incorporating class-specific information through Linear Discriminant Analysis, making it robust to variations in lighting and noise, which are challenges faced by its predecessor, Eigenface. Ref. [45] also emphasized that the Fisherface algorithm serves as a valuable approach for diminishing dimensionality while retaining information to the greatest extent possible.

4.4 Local Binary Pattern Histogram

The Local Binary Pattern (LBP), introduced by Ref. [37] and refined by Ref. [38], functions as a descriptor by utilizing texture features extracted from facial regions through a binary pattern. The integration of this Local Binary Pattern (LBP) pattern into the Histogram of Oriented Gradients (HOG) classifier results in the nomenclature LBPH, indicating the amalgamation of the LBP algorithm with histogram computation.

In LBPH, each pixel value in an image receives a label. The operator compares pixel values, verifying whether the value of the neighboring pixel surpasses that of the central pixel. In such cases, it returns a '1'; otherwise, if the neighboring pixel is smaller, it returns '0'. After performing these comparisons, a binary number is produced, and when translated into decimal form, it forms the basis for constructing a histogram. The primary choice for "Feature extraction" is the histogram of linear binary patterns.

4.5 OpenCV

The open-source library OpenCV is crucial for interfacing and executing facial recognition algorithms like Eigenface, Fisherface, and LBPH [47]. These algorithms find widespread applications across various domains. Developed by Intel Corporation, OpenCV is specifically designed for computer vision applications, showcasing remarkable efficiency in real-time scenarios. It ensures compatibility with key operating systems such as Windows, Mac OS, and Linux, and further extends support to IOS and Android platforms. OpenCV provides interfaces for various programming languages, including Python, C++, and Java. [47].

OpenCV, boasting a collection of more than 500 functions, is organized into five primary categories: image processing, structural analysis, motion analysis and object tracking, pattern recognition, and camera calibration featuring 3D reconstruction capabilities. This extensive functionality makes OpenCV a versatile and comprehensive tool for implementing facial recognition systems and various other computer vision tasks. Its ability to seamlessly integrate with various platforms and programming languages has played a significant role in its widespread acceptance for crafting advanced applications and solutions in the realm of computer vision.

4.6 MTCNN

MTCNN, an acronym for Multi-task Cascaded Convolutional Networks, is a widely employed algorithm in the realm of face detection. Its operation involves three key stages: initial face detection, refining bounding box parameters, and pinpointing facial landmarks. The working of MTCNN is summarized as follows [35]:

Stage 1: Face Detection

1. The given input image is resized at multiple scales to create an image pyramid. This helps in detecting faces of different sizes.
2. P-Net, a compact convolutional neural network (CNN), receives an image patch as input and generates a collection of bounding box proposals or potential windows, each assigned a probability score. It uses a sliding window approach to generate candidate boxes at different positions and scales.
3. Overlapping bounding box proposals are filtered using non-maximum suppression to keep only the most confident and non-overlapping proposals.

Stage 2: Bounding Box Regression

1. The selected bounding box proposals from the P-Net stage are warped and resized to a fixed size.
2. R-Net is a more complex CNN that takes the warped and resized image patches as input. It improves upon the bounding box suggestions produced by the P-Net, enhancing the precision of bounding box coordinates and assigning a probability score to each refined box.
3. Just like in Stage 1, Non-Maximum Suppression (NMS) is utilized to eliminate redundant bounding boxes with overlaps, ensuring the preservation of the most confident and distinct ones.

Stage 3: Facial Landmark Localization

1. The surviving bounding boxes from the R-Net stage are warped and resized to a fixed size.
2. O-Net is another CNN that takes the warped and resized image patches as input. It anticipates and identifies key facial points, including the locations of eyes, nose, and mouth, within individual bounding boxes. It also refines the bounding box coordinates and provides a probability score for the final prediction.
3. Once more, Non-Maximum Suppression (NMS) is utilized to eliminate overlapping bounding boxes, retaining those with the highest confidence scores and ensuring they do not overlap with each other.

4.7 FaceNet

FaceNet, a deep learning model designed for face recognition, is renowned for its capacity to produce a concise and distinguishing representation (embedding) of facial features. The structure relies on a deep convolutional neural network (CNN) and employs a triplet loss function in its training process to acquire knowledge for face embeddings [20]. Detailed methodology for FaceNet as described in [20] is as follows:

1. Compile an extensive dataset of facial images that is both varied and encompasses a broad spectrum of poses, lighting scenarios, and facial expressions.
2. Data Preprocessing includes aligning face images after cropping, to ensure that the face is the primary focus in each image. Normalizing pixel values to bring consistency in the data is the second important step.
3. Using a deep CNN architecture for feature extraction. FaceNet often uses the inception module to capture hierarchical features. Design the structure to generate a consistent-dimensional vector (embedding) for every given

face image input.

4. The triplet loss function, vital in FaceNet training, aims to reduce the gap between the anchor image and a positive image (belonging to the same identity) while simultaneously maximizing the gap between the anchor image and a negative image (pertaining to a different identity). The loss function as referred in [28] is formulated as :
   L(A, P, N) = max(0, ||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + margin)
   where f(.) represents the embedding function.
5. The subsequent step involves training the FaceNet model by utilizing the triplet loss function and the gathered dataset. Employing a substantial volume of labeled data during the training process is essential to guarantee the model acquires a resilient facial representation. Conducting experiments with different hyperparameters, including learning rate, batch size, and margin in the triplet loss, is advisable.
6. It is essential to validate the model using a distinct dataset to ensure its ability to effectively generalize to faces not encountered during training.
7. Once trained, this face data can be used to train models to generate face embeddings for new face images.
8. In face recognition applications, the produced embeddings serve the purpose of assessing the likeness between faces. A prevalent method involves utilizing cosine similarity or Euclidean distance between embeddings to ascertain the degree of similarity.
9. Assessing the effectiveness of FaceNet involves evaluating its performance on established face recognition benchmarks. This evaluation utilizes metrics like accuracy, precision, recall, and F1 score to gauge the model's proficiency.

5. ALGORITHMIC SURVEY

Face detection and recognition technologies have witnessed remarkable advancements in recent years. This algorithmic survey delves into the diverse landscape of face detection and recognition algorithms, from classical methods like Viola-Jones to cutting-edge deep learning approaches showing their accuracies over a period of time.

Sameer Aqib Hashmi [16] addressed the challenge of face detection in challenging conditions, leveraging deep learning for enhanced performance. The proposed deep cascaded multi-task framework, consisting of three layers of carefully designed convolutional networks, demonstrated effective identification of faces and landmark regions in unconstrained environments. Additionally, a novel online hard sample mining method was introduced, autonomously improving performance without manual pattern selection during the learning process.
Md. Towfiqul Islam, et al [17] addressed the integration of machine learning for facial recognition in daily life scenarios, emphasizing the challenges posed by low-quality images. Concentrating on a dataset comprising 627 individuals from Bangladesh, images were taken from four different angles. Five machine learning approaches, namely CNN, Haar Cascade, Cascaded CNN, Deep CNN, and MTCNN, were utilized in the study. After model creation and execution, the Multi-Task Convolutional Neural Network (MTCNN) outperformed others, achieving a peak accuracy of 96.2% with the training data. The research showcased the effectiveness of MTCNN in enhancing facial recognition accuracy, particularly in scenarios with hardware cost constraints.

Laxmi Narayan Soni and Dr. Akhilesh A Waoo [18] addressed the underexplored area of face detection, emphasizing the lack of research in facial recognition for infants, the elderly, and individuals with dark skin. It highlighted the Viola-Jones algorithm's historical significance and popularity in various applications. The paper presented an in-depth exploration of methodologies for face detection and the cutting-edge advancements in face recognition systems based on deep learning. Convolutional Neural Networks (CNNs) have proven to be the most efficient artificial intelligence technique in the realm of face detection, capitalizing on their capacity to discern patterns within unlabeled data.

Thai-Viet Dang and Linh H. Tran [19] explored the importance of Artificial Intelligence (AI) and the Internet of Things (IoT) in the context of the Fourth Industrial Revolution, emphasizing their broad practicality. Highlighting AI's self-learning ability, the study proposed a secure two-step identification system utilizing MTCNN and FaceNet networks, enhanced with head pose estimation. The model achieved a competitive accuracy range of 92% to 95%, showcasing its usability in face recognition applications.

Sumaira Manzoor, et al [20] tackled the challenge of deploying deep learning models on edge devices for efficient inference, focusing on the GuardBot service robot powered by Jetson Xavier NX. The research showcased a practical case example that featured the implementation of an enhanced application for recognizing face masks in real-world scenarios. Utilizing a dual-stage architecture that incorporates MTCNN, a recently introduced CNN model, and custom transfer learning-based models, the framework displayed enhanced accuracy in face mask recognition compared to existing top-tier models. By optimizing the model with TensorRT, notable improvements in inference speed were achieved on the Jetson Xavier NX, demonstrating impressive throughput and minimal latency in real-world experiments conducted both indoors and outdoors. The CNN model proposed in this study outperformed alternative models, achieving accuracies of 94.5%, 95.9%, and 94.28% on the training, validation, and testing datasets, respectively.

Muhammad Shoaib Farooq, et al [21] reflected on the substantial progress in facial recognition technology over recent decades, widely employed for defence, security, and daily applications. While facial recognition systems demonstrate general efficiency, challenges have emerged, particularly with accuracy issues when processing images of individuals with darker skin tones. The article brought attention to the variability in accuracy among current algorithms when used on individuals with dark skin tones. It introduced a hybrid algorithm, combining Gaussian and Explicit rule models, with a focus on enhancing the accuracy of face detection specifically for dark-skinned individuals. Thorough experiments with a dataset of black faces demonstrated a significant accuracy enhancement, reaching 89% for dark skin.

Yibo Cao, et al [22], found that face recognition garnered attention for its non-contact benefits. Researchers preferred 3D faces for their spatial richness, but obtaining 3D data in normal conditions was challenging. To enhance 3D face recognition in weak-light environments, an algorithm which automatically detected and processed 3D faces, used RoPS in PointNet++ for feature description, and achieved successful results in simulated low-light scenarios with datasets like Bosphorus and CASIA-3D.

Table 2: Algorithmic Survey of Research Studies

Ref. No.	Algorithm Name	Accuracy
[15]	Voila Jones for Face detection	90%
[16]	MTCNN	97%
[16]	Haar Cascade	68.16%
[16]	Viola-Jones	61.81%
[16]	MTCNN+Facenet+ Head Pose	90-95%
[17]	MTCNN	96%
[18]	DeepID, DeepFace	92%
[18]	CNN, FaceNet	95%
[19]	LBP	84-88%
[20]	Caffe DNN	76.96%
[21]	Gaussian, Explicit Rule	60-80%
[22]	PointNet++ 3D-face recognition, RP-Net	90%
[23]	MTCNN, Haar Cascade	93%
[24]	ARCFACE, MTCNN	95%
[25]	FaceNet, MTCNN	90%
[26]	ArcFace, CosFace & FaceNet FR	70-80%
[27]	FaceNet, MTCNN	95%

Time Complexity: O(n) for all algorithms where n represents the number of features used for evaluation.

Figure 1 shows the comparison of the most commonly used algorithms for the purpose of face detection and recognition derived from the above algorithmic survey.

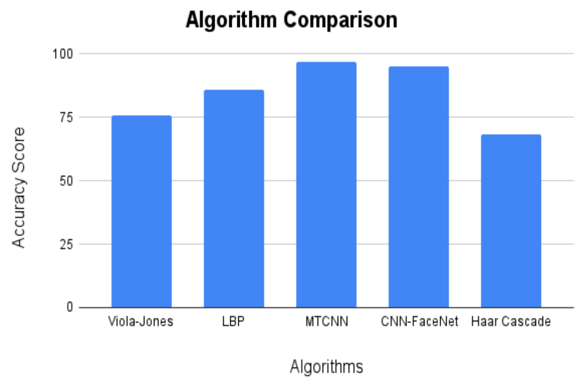

Fig 1: Comparison of Algorithms

Algorithms give variable accuracy scores depending on the dataset used with MTCNN and FaceNet giving the highest accuracy in fields of face detection and recognition respectively.

6. CONCLUSION

In conclusion, this survey paper provides an in-depth exploration of various face detection and recognition algorithms, with a particular focus on state-of-the-art approaches such as MTCNN and FaceNet. In recent years, facial analysis has experienced notable advancements propelled by the progress in deep learning and computer vision techniques. Referencing [18], the initial successful approach for face detection involved the Haar cascade classifier—a machine learning-based method trained on numerous positive and negative images, offering a swift and effective detection process. Another successful method is the Local Binary Pattern (LBP), a non-parametric approach utilizing predefined binary patterns for face detection based on local texture analysis. Renowned for its speed and reliability, LBP stands out. The third method, Histogram of Oriented Gradients (HOG), serves as a feature descriptor for object detection, relying on local gradient orientation and predefined orientation bins. HOG's advantage lies in its capability to operate with low-resolution images and its enhanced robustness compared to other methods [18]. Lastly, the fourth method involves Deep Learning based Face Detection, recognized as the most accurate and reliable approach. This method leverages convolutional neural networks and employs layers of convolutional filters for precise face detection. The MTCNN algorithm has demonstrated its effectiveness in accurately localizing faces in images with varying scales and orientations, making it a robust choice for real-world applications. Furthermore, the integration of FaceNet into face recognition has notably enhanced the precision and dependability of identity verification systems. The ability to generate compact and discriminative face embeddings, coupled with the triplet loss function, helped in addressing pivotal obstacles in face recognition, such as effectively managing variations in lighting conditions and pose. The examined algorithms have played a crucial role in advancing face detection and recognition systems. Ongoing improvements in these algorithms show considerable potential for the future of biometric applications, promising heightened security and enhanced user experiences across diverse domains.

7. FUTURE SCOPE

Promising opportunities lie ahead for the future development of face detection and recognition systems, with various areas offering exciting prospects for further research. Challenges related to occlusions continue to persist in the face recognition task [18]. Integration of multiple modalities, such as combining facial features with other biometric data like voice, gait, or iris recognition, can enhance the overall accuracy and robustness of identification systems, Real-time Processing and Edge Computing, developing systems that are resilient to adversarial attacks is a crucial area for future exploration, Dataset Diversity and Bias Mitigation are potential challenges for building more effective systems.

8. REFERENCES

[1] Khan, Muhammad Zeeshan, et al. "Deep unified model for face recognition based on convolution neural network and edge computing." IEEE Access 7 (2019): 72622-72633.

[2] Peralta, Billy, et al. "Gender identification from community question answering avatars." IEEE Access 9 (2021): 156701-156716.

[3] Wang, Dalin, and Rongfeng Li. "Enhancing Accuracy of Face Recognition in Occluded Scenarios with OAM-Net." IEEE Access (2023).

[4] Li, Lina, et al. "ET-YOLOv5s: toward deep identification of students' in-class behaviors." IEEE Access 10 (2022): 44200-44211.

[5] Tripathy, Rajashree, and R. Daschoudhury. "Real-time face detection and tracking using haar classifier on soc." International Journal of Electronics and Computer Science Engineering 3.2 (2014): 175-184.

[6] Hussain, Dostdar, et al. "Face mask detection using deep convolutional neural network and MobileNetV2-based transfer learning." Wireless Communications and Mobile Computing 2022 (2022): 1-10.

[7] Kaur, Gagandeep, et al. "Face mask recognition system using CNN model." Neuroscience Informatics 2.3 (2022): 100035.

[8] Mo, Hyunggeun, and Seungku Kim. "A deep learning-based human identification system with wi-fi csi data augmentation." IEEE Access 9 (2021): 91913-91920.

[9] James Coe and Mustafa Atay " Evaluating Impact of Race in Facial Recognition across Machine Learning and Deep Learning Algorithms" Department of Computer Science, Winston-Salem State University, Winston-Salem, NC 27110, USA

[10] Vankadhara Rajyalakshmi And Kuruva Lakshmanna "Intelligent Face Recognition Based Multi-Location Linked IoT Based Car Parking System" IEEE Access 11(2023): 84258-84269.

[11] M. Raghavendra, R. Neha et al. "Missing Child Identification using Convolutional Neural Network", IJRASET: 380-384.

[12] D. J. Samatha Naidu, R. Lokesh "Missing Child Identification System using Deep Learning with VGG-FACE Recognition Technique", International Journal of Computer Science and Engineering (2022), volume 9 issue 9, 1-11.

[13] Shaik Mohammed Zahid, Salman.K et al. "A Multi Stage Approach for Object and Face Detection using CNN" Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023), 798-803.

[14] Osorio-Roig, Dailé, et al. "Stable hash generation for efficient privacy-preserving face identification." IEEE Transactions on Biometrics, Behavior, and Identity Science 4.3 (2021): 333-348.

[15] Pilania, Urmila, and Prinima Gupta. "An Improvised Method for Detecting Face as ROI in Video."

[16] Hashmi, Sameer Aqib. "Face Detection in Extreme Conditions: A Machine-learning Approach." arXiv preprint arXiv:2201.06220 (2022).

[17] Md. Towfiqul Islam, Tanzim Ahmed et al. "Convolutional Neural Network based Partial Face Detection", Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh.

[18] Soni, Laxmi Narayan, and Akhilesh A. Waoo. "A Review of Recent Advances Methodologies for Face Detection." (2023).

[19] Thai-Viet Dang et al. "A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet", Journal of Robotics and Control (JRC), School of Mechanical Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam.

[20] Manzoor, Sumaira, et al. "Edge Deployment Framework of GuardBot for Optimized Face Mask Recognition with Real-Time Inference Using Deep Learning." Ieee Access 10 (2022): 77898-77921.

[21] Farooq, Muhammad Shoaib, et al. "A Hybrid Algorithm for Face Detection to Avoid Racial Inequity Due to Dark Skin." IEEE Access 9 (2021): 145109-145114.

[22] Cao, Yibo, et al. "RP-Net: A PointNet++ 3D face recognition algorithm integrating RoPS local descriptor." IEEE Access 10 (2022): 91245-91252.

[23] Wei, Jun. "Video face recognition of virtual currency trading systems based on deep learning algorithms." IEEE Access 9 (2021): 32760-32773.

[24] Nguyen, Duy Dieu, et al. "Smart Desk in Hybrid Classroom: Automatic Attendance System based on Face Recognition using MTCNN and ARCFACE." 2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR). IEEE, 2022.

[25] Jin, Rongrong, et al. "Face recognition based on MTCNN and Facenet." (2021).

[26] Shen, Meng, et al. "Effective and robust physical-world attacks on deep learning face recognition systems." IEEE

Transactions on Information Forensics and Security 16 (2021): 4063-4077.

[27] Dang, Thai-Viet, and Linh H. Tran. "A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet." Journal of Robotics and Control (JRC) ISSN 2715.5072 (2023): 2.

[28] Arun Pratap Singh; Vaishali Singh, Infringement of Prevention Technique against Keyloggers using Sift Attack,2018 International Conference on Advanced Computation and Telecommunication (ICACAT),19 December 2019, IEEE,10.1109/ICACAT.2018.8933805

[29] Danial Javaheri; Mehdi Hosseinzadeh; Amir Masoud Rahmani,Detection and Elimination of Spyware and Ransomware by Intercepting Kernel-Level System Routines,IEEE Access (Volume: 6),07 December 2018,IEEE,10.1109/ACCESS.2018.2884964

[30] Paul Viola and Michael J. Jones. 2014. Robust Real-Time Face Detection. International Journal of Computer Vision 57, 2 (May 2004), 137ś154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[31] 1Ag. Asri Ag. Ibrahim, 1,2Kashif Nisar, 1Yeoh Keng Hzou, 2Ian Welch, Review and Analyzing RFId Technology Tags and Applications, IEEE 2019, https://doi.org/10.1109/AICT47866.2019.8981779

[32] Xiulong Liu Xin Xie Xibin Zhao Keqiu Li Alex X. Liu Song Guo Jie Wu, Fast Identification of Blocked RFID Tags, (Volume:17, Issue: 9Sept. 1 2018), https://doi.org/10.1109/TMC.2018.2793219

[33] Dr. Harman Preet Singh Department of Management & Information Systems, College of Business Administration, University of Hail. EXPLOITING RFID FOR BUSINESS TRANSFORMATION: A STRATEGIC ANALYSIS VIS-À-VIS AGRICULTURAL BANK OF CHINA, Volume: 5 | Issue: 1 | Jan 2019

[34] Mugahid Omer1, Yachao Ran2, and Gui Yun Tian1, 2, Indoor Localization Systems for Passive UHF RFID Tag Based on RSSI Radio Map Database, Progress in Electromagnetics Research M, Vol. 77, 51–60, 2019

[35] Face Recognition Based on MTCNN and FaceNet Rongrong Jin, Hao Li, Jing Pan, Wenxi Ma, and Jingyu Lin, 2020, https://api.semanticscholar.org/CorpusID:231931832}

[36] R. Tyagi, G. S. Tomar, and N. Baik, "A survey of unconstrained face recognition algorithm and its applications," Int. J. Secur. its Appl., vol. 10, no. 12, pp. 369–376, 2016.

[37] Ojala, T., Pietikainen, M. I., Harwood, D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 1996; 29 (1): 51-59.

[38] Ahonen, T., Hadid, A., Pietikainen, M. Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 28 (12): 2037-2041. DOI: 10.1109/TPAMI.2006.244.

[39] Kshirsagar, V. P., Baviskar, M. R., Gaikwad, M. E. Face recognition using Eigenfaces. In: 2011 3rd International Conference on Computer Research and Development. IEEE, 2011, pp. 302-306.

[40] Ejaz, M. S., Islam, M. R., Sifatullah, M., Sarker, A. Implementation of principal component analysis on masked and non-masked face recognition. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, 2019, pp. 1-5.

[41] Mulyono, I. U. W., Susanto, A., Rachmawanto, E. H., Fahmi, A. Performance Analysis of Face Recognition using Eigenface Approach. In: 2019 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE, 2019, pp. 1-5.

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

[43] Lu, W. Y., Ming, Y. A. N. G. Face detection based on Viola-Jones algorithm applying composite features. In: 2019 International Conference on Robots & Intelligent System (ICRIS). IEEE, 2019, pp. 82-85.

[44] Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989; 11 (7): 674-693.

[45] Hegde, N., Preetha, S., Bhagwat, S. Facial Expression Classifier Using Better Technique: FisherFace Algorithm. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2018, pp. 604-610.

[46] Belhumeur, P. N., Hespanha, J. P., Kriegman, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In: European Conference on Computer Vision. Springer, Berlin, 1996, pp. 43-58.

[47] Taheri, S., Vedienbaum, A., Nicolau, A., Hu, N., Haghighat, M. R. Opencv. js: Computer vision processing for the open web platform. In: Proceedings of the 9th ACM Multimedia Systems Conference. 2018, pp. 478-483