Received: 27 March 2024 | Revised: 6 May 2024 | Accepted: 9 June 2024
DOI: 10.37917/ijeee.21.2.7 Early View | December 2025

Open Access

Iraqi Journal for Electrical and Electronic Engineering
*Original Article*

# Designing Face Detection Systems with Gray Wolf Optimization

**Noor Sabah Abbod*[1], Jamshid B. Mohasefi[2]**
[1]Department of Computer Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq
[2]Department of Computer Science, Faculty of Sciences, Urmia University, Urmia, Iran

Correspondance
*Noor Sabah Abbod
Department of Computer Engineering, College of Engineering,
Mustansiriyah University, Baghdad, Iraq
Email: noorsabah@uomustansiriyah.edu.iq

**Abstract**
***The main objective of this paper project was to create a state-of-the-art face identification technique that can handle the various difficulties caused by changes in illumination, occlusions, and facial emotions. Face detection is a cornerstone of computer vision, facilitating diverse applications ranging from surveillance systems to human-computer interaction. Throughout this paper, the comprehensive exploration of advancing face detection methodologies has been undertaken, culminating in developing and evaluating a novel approach. The challenges posed by variations in facial expressions, lighting conditions, and occlusions necessitated a multifaceted solution. Our proposed method, which consists of interconnected steps, works quite well to overcome these challenges. Using deep learning architectures to increase feature extraction and discrimination was beneficial in the initial stage of fine-tuning Residual Networks (ResNet-50) to serve as the Region-based Convolutional Neural Network (Faster R-CNN) framework classifier. The process of gradually optimizing thresholds, such as batch size, learning rate, and detection threshold, involved using the Gray Wolf optimization technique (GWO). The conversion process was accelerated and improved overall detection process efficiency and accuracy using a clever fusion of machine learning and metaheuristic optimization techniques. A key component of our methodology is the careful data processing, which was necessary to ensure. The suggested method was carefully examined on a particular dataset, and the 94% training accuracy that was attained together with an identical test dataset accuracy highlights the method's resilience. These findings support the effectiveness of our approach in reducing false positives and negatives, resulting in unmatched recall and precision in the detection system. The discovery has significant significance as it can potentially improve face detection systems' performance and reliability in various real-world applications, such as human-computer interaction and surveillance. Convolutional neural networks, deep learning architectures, and metaheuristic optimization approaches were synergized to produce a new and reliable solution.***

**Keywords**
**Face detection, Faster-RCNN, Gray wolf optimization algorithm, ResNet-50, Object detection.**

## I. INTRODUCTION

Modern facial recognition, which performs at a human level on complex visual recognition [1], marks a significant advancement in artificial intelligence. Artificial intelligence and machine learning advancements have helped facial recognition technologies advance quickly [2]. Although the first attempts at facial recognition date back to the 1960s, significant advancements have been made in recent years. Deep convolutional neural networks have significantly increased facial recognition system accuracy since 2017 [3]. Deep learning and neural networks are used in contemporary facial recognition; this method was inspired by the human visual system [4]. These models can extract reliable facial features thanks to large training datasets [5]. Modern methods like DeepID [6] and FaceNet [7] can surpass 99% accuracy on benchmark tests. Law enforcement [8], smartphone authentication [9], and attendance taking [10] are only a few of the current facial recognition applications. However ethical issues with bias and privacy continue [11]. It is necessary to regulate this spreading technology carefully [12]. In this paper, the design of a multi-face detection system is presented. The system uses a two-stage pipeline. A detection network employing an upgraded Non-Maximum Suppression (NMS) algorithm [13] and Faster R-CNN [14] architecture generates candidate areas from an efficient region proposal network based on Grey Wolf Optimization (GWO) [15]. The system speed is improved by creating a lightweight area proposal network customized using GWO, which performs better than other search algorithms in simulating group hunting behavior for optimization. Based on a cutting-edge clustering method, the enhanced NMS method better manages overlapping detections. The Faster R-CNN and ResNet-50 models are optimized using quantization and pruning methods.

The remainder of this paper is structured as follows: Discusses prior work on face detection, CNN architectures, model optimization, and hardware acceleration techniques relevant to this research is presented in section II. . Section III. discusses the proposed method with Grey Wolf Optimization, including Faster R-CNN and ResNet-50 models. Simulation results and conclusions are presented in sections IV. and V. , respectively.

## II. Literature Review

In this section, the papers discuss various face recognition systems using techniques like CNN, Support Vector Machine (SVM), Principal Component Analysis (PCA) etc. For applications like security, access control, and attendance monitoring are presented. A deep learning and computer vision to identify people and track their arrival and departure times from a designated spot is combined with facial recognition technology to create a proposed attendance tracking system in [16]. In [17], the Developed system using CNN-PCA for attendance is suggested with 98% accuracy. Implements Viola-Jones based innovative door system with 95% accuracy is suggested in [18]. In [19], Combining AdaBoost and deep belief network for image classification, which reduces error rate, is proposed. In [20], the CNN-SLT-CBIR technique for image retrieval with 89% accuracy is presented. The systems are evaluated on different datasets and real-world scenarios. The accuracy achieved ranges from 70% to 99% based on the technique and dataset size. An automatic gate system for security using license plate recognition, car model recognition, and face detection is proposed in [21]. This system achieved 75% accuracy on the Qatar car dataset. In [22], a CNN-based real-time face recognition system is evaluated. The achieved accuracy is about 98.75% and 98% on datasets and real-time data. Using wavelet compression, multi-neural networks tolerant to 30° rotations and 1m distances, and histogram equalization, the system outperforms OpenCV Haar's 50% accuracy and dlib HOG's 85-95% in face recognition accuracy, suggested in [23]. Introduces an integrated algorithm for face recognition in a Smart Gate system with up to 97% recognition rate is presented in [24]. Explores CNN model for attendance management with 96.87% accuracy is introduced in [25]. A CNN architecture and web app for attendance tracking is proposed in [26] with 99% accuracy.

An Evaluated MTCNN model for face detection in images is proposed in [27]. Tests CNN AlexNet system for home door locking with 97.5% accuracy are presented in [28]. In [29], the eigenface method and SVM are used for student monitoring. With 61% accuracy on 50 images is introduced. In [30] presented an automated facial recognition system that uses Viola-Jones to achieve 98.75% detection accuracy to exact face identification, it uses linear discriminant analysis and machine learning. An IOT-based home automation and security system using R-CNN for face recognition with 82% accuracy of the face recognition model is presented in [31]. A Combines Viola-Jones and CNN for identifying witnesses in video with 99.5% accuracy is presented in [32]. Depthwise CNN for facial expression recognition is presented in [33] with 70.76% and 97.92% accuracy. A present DICNN model for expression recognition on embedded systems is suggested in [34] with high accuracy with low memory. The papers highlight the potential of face recognition to enhance security, automate processes like attendance, and reduce the need for manual verification. Overall, they provide valuable contributions to research on face recognition systems and their real-world applications.

The gaps and restrictions in this literature:

Although the examined works show encouraging results, occlusions, changing angles and illumination, complex real-world circumstances, and computing limitations for edge deployment remain problems for existing face recognition systems. Their evaluations are often limited to small, controlled datasets. Critical holes that must be filled with thorough real-world evaluations and various data sources include enhancing robustness and efficiency and addressing privacy concerns. In the future, these restrictions should be addressed to produce face analysis technologies that are more accurate and efficient.

This paper makes Significant advances in the field, which proposes a suitable face detection methodology that smoothly blends deep learning architectures and metaheuristic optimization techniques. To illustrate the effectiveness of deep learning models for this job, a deep convolutional neural network architecture called ResNet-50 was used within the Faster R-CNN object identification framework to provide robust facial feature extraction and categorization. Using the metaheuristic Gray Wolf optimization algorithm, an inventive combination of machine learning and inductive inference was accomplished by optimizing the Faster R-CNN model's hyperparameters, including batch size and learning rate. Careful planning and execution of data pre-processing techniques were essential in improving the suggested model's resilience and generalizability in the face of changing occlusions, and illumination. The approach demonstrated its exceptional capacity to reduce false positives and false negatives, leading to unmatched precision and recall for face identification systems, with advanced accuracy of 94% through thorough experimental evaluation on demanding benchmark datasets. Because deep learning and optimization approaches are successfully combined in this work, the area has advanced significantly, and more dependable and effective face detection solutions with a wide range of real-world applications are now possible.

## III. Architecture of The Proposed Approach

In recent years, advancements in computer vision have revolutionized various industries, including security, surveillance, and human-computer interaction. Face and object detection are two fundamental tasks in computer vision with numerous real-world applications. This paper explores the synergistic integration of ResNet-50 for face detection and Faster R-CNN for object detection and optimization using the Grey Wolf Optimizer (GWO). Integrating these networks and an optimization technique like the Grey Wolf Optimizer can address the challenges of accuracy, speed, and resource utilization. The steps of the proposed approach are demonstrated in Figure 1.

At the preprocessing step, the data preparation is a crucial stage in machine learning and data analysis. It entails preparing raw data for analysis or model training by cleaning and arranging it. The main preprocessing processes in this study are Adjusting the Bounding Box, RGB conversion, Data shuffling, and Data partitioning. In Adjusting the Bounding Box processes the Bounding box is resizing for machine vision tasks to match scaled images and to do must establish the initial bounding boxes ($x\_min$, $y\_min$) and ($x\_max$, $y\_max$), Maintain the aspect ratio when resizing the image, Determine the Resizing Scale, and finally Scale Bounding Boxes to your liking this shown in Figure 2. To improve viewing, grayscale or single-band images are converted into pseudo-color RGB images by using RGB conversion. The steps for these processes are to Begin with image data in grayscale and then Replicate the intensity levels for all three RGB channels to produce pseudo-color images. For every channel, the same intensity level is used. Applying a color scheme, usually one that resembles a gradient. Rearranging data instances (rows or samples) in a dataset randomly or in a pseudo-random manner is known as a data shuffling process. The information partitioning technique is essential for getting to know and evaluating information. A dataset is divided into two separate test and schooling sets subsets. A schooling set is used to explore patterns and relationships in the statistics and optimize parameters; at the same time, the take-a-look-at set is used to evaluate a model's performance and generalizability. Data shuffling or randomization strategies, which entail the random reordering of statistics instances, are usually used to improve the effectiveness of the getting-to-know method. In addition to ensuring a fair subgroup department, this method reduces bias. Furthermore, RGB reconstruction enhances viewing by replicating intensity levels using the red, green, and blue channels. This technique can transform Grayscale or single-channel images into pseudo-colored RGB images. Because of the division's well-planned architecture, representation and distribution are guaranteed, ensuring the data is consistently available during the learning process. Standard procedures for a robust model evaluation include time-based splitting, random sampling, and more complex techniques like cross-validation. Proper data division prevents overfitting and improves the model's dependability. This paper uses 70% training and 30% test sets.

### A. *The Suggested Pairing of R_CNN and ResNet50*

The convolutional neural network architecture known as ResNet-50 was introduced in "Deep Residual Learning for Image Recognition" [35]. It is used for image classification. The key novelty of RESTNET is the introduction of identity shortcut connections that omit layers. This solves the vanishing gradient problem and allows for the training of large networks with 50, 101, or 152 layers. Recent works like ResNeXt [36], SENet [37], and Non-local Neural Networks [38] demonstrate ResNet remains highly influential for image recognition tasks.

Faster R-CNN [39] builds on previous work like R-CNN and Fast R-CNN to provide an efficient and accurate object detection algorithm. The essential contribution is introducing a Region Proposal Network (RPN) that shares convolutional features with the detection network, enabling nearly cost-free region proposals. This improves on previous slow methods like Selective Search. Recent improvements to Faster R-CNN include context reasoning modules [40], Light-Head R-CNN [41] to reduce computations, and Relation Networks [42] to predict relations between objects. These demonstrate Faster R-CNN remains an influential foundation for object detection research.

The proposed combination leverages ResNet50's state-of-the-art image classifications and Faster R-CNN's accurate object detection. ResNet50's identity shortcuts could enable

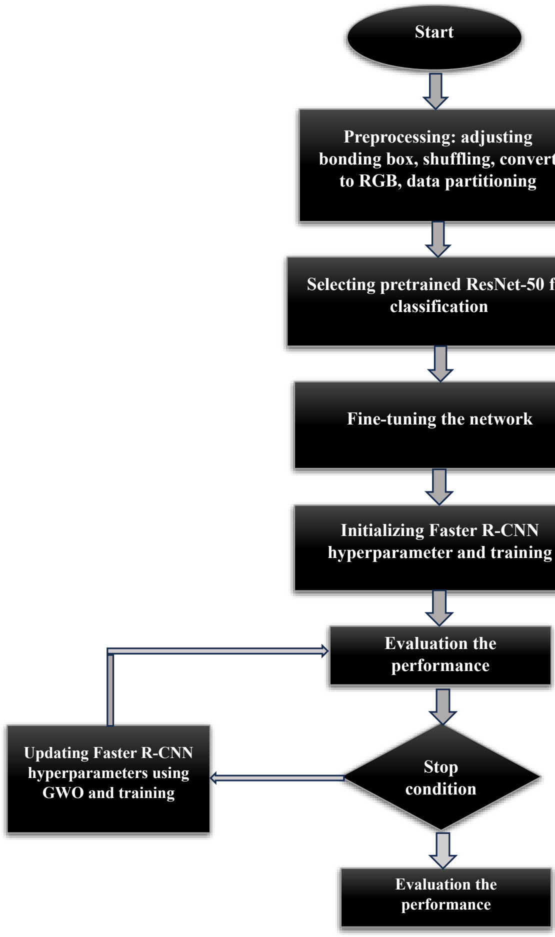



Fig. 1. Steps of the proposed method.

training a Feature Pyramid Network built on top of the RPN to improve multi-scale detections without vanishing gradients. The shared convolutional features between the ResNet50 classifier and Region Proposal Network could provide computational and modeling efficiencies.

ResNet-50, with its core innovation of residual blocks, addresses the vanishing gradient problem by learning residual mappings. It is designed for image classification and consists of 50 layers organized into three stages with varying residual blocks. Each block includes $1 \times 1$ and $3 \times 3$ convolutions along with skip connections. Using residual blocks enables the training of deep networks by facilitating gradient flow, making it a key innovation. ResNet-50's success has influenced subsequent network designs in computer vision. To modify ResNet-50 for facial recognition, follow these steps: Add Specialized Layers, fine-tune, Transfer Learning, and Remove the Final Layers. The Remove Final Layers function removes the final three layers (totally linked, softmax,

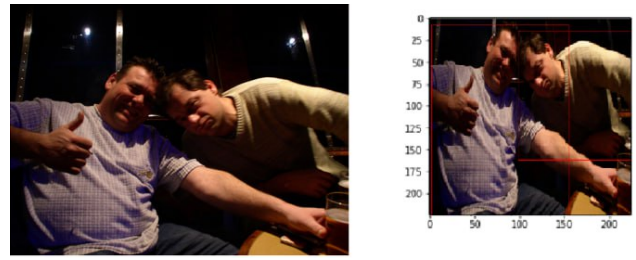



Fig. 2. Resizing with bounding box.

and classification) intended for general picture categorization. Transfer Learning uses pre-trained weights from big datasets to build the ResNet-50 model. This provides good initial parameter values that speed up learning to discriminate facial features. In Fine-Tuning, specific classification layers are used in place of the current feature extraction capabilities to collect crucial facial characteristics for identification. Retraining replaces fully connected layers with convolutional feature extraction, retraining the network to identify faces instead of generic objects.

A Faster R-CNN detection network uses bounding box regression and a classifier to handle the discovered regions. All components of the Faster R-CNN model with RPN and detection network are trained end-to-end using labeled data. In conclusion, practical tuning of hyperparameters and training procedures is essential for Faster R-CNN to identify different objects accurately. Among the tactics include adjusting detection thresholds, preventing overfitting by regularization, and boosting variability to manage different object appearances. In conclusion, careful coordination of backbone pre-training, region proposal generation, detection network design, and end-to-end optimization is needed to train an accurate Faster R-CNN detector.

### *B. Grey Wolf Optimizer (GWO):*

The Gray Wolf Optimizer (GWO) is a relatively recent nature-inspired introduced metaheuristic algorithm in [43]. GWO optimizes by imitating the hunting and leadership hierarchies of gray wolves. The population is divided into four groups: alpha, beta, delta, and omega wolves. The alpha wolf is considered the leading member of the group. The subsequent equations are suggested in (2) and (3) to represent encircling behavior quantitatively. Besides, the approach uses the encircling technique to refine results as the search advances iteratively. GWO has proven to have a quicker convergence rate and a more remarkable ability to overcome local optima that are not optimal, as opposed to partial swarm optimization methods [44]. Answers to challenging optimization problems have been the main focus of current GWO research [45]. Machine learning models frequently use GWO to optimize their parameters. Object detection models primarily depend on the detection threshold of the proposed bounding box, which establishes the degree of confidence needed to classify objects that have been detected. The GWO method optimizes this threshold by repeatedly examining different values to maximize the bounding box's quality and improve detection accuracy. GWO uses concepts derived from wolf pack behaviors to find the ideal threshold. Integrating deep neural networks with Grey Wolf Optimization (GWO) enables the high-quality tuning of modern object detectors, thereby attaining the highest quality performance by optimizing the version's settings.

The initialization step of the technique includes generating a random populace of answers, which units the muse for the subsequent optimization cycle. In the Fitness Assessment step, a price characteristic is employed to evaluate the fitness score of each answer. This cost function is based on the Intersection over Union (IoU) metric, which quantifies the diploma of overlap among anticipated and floor fact areas. The IoU values serve as a measure of alignment excellence and item detection effectiveness. Lower charges are associated with better IoU values, indicating superior performance in object identification. The cost function is described as:

$$cost\ function = 1 - iou \tag{1}$$

During the Leader Update section, the Alpha, Beta, and Delta Wolves expect leadership roles and manual the alternative wolves in the optimization procedure based on their health scores. This ensures powerful course and coordination in the p.c. The Position Update phase comes into play to facilitate a dynamic look for the most suitable answers. In this segment, the wolves' positions are dynamically adjusted, considering each exploration to discover new areas and the effect of the leaders to make the most promising regions. The manner continues till a predefined variety of iterations is finished or specific termination requirements are met. This stage, in which the manner concludes, is called termination. In precis, incorporating Grey Wolf Optimization (GWO) into the combined ResNet50 and Faster R-CNN network for parameter tuning offers a promising method to beautify detection overall performance in diverse applications, which includes security and surveillance. To quantitatively describe the encompassing behavior, we provide the following equations:

$$\vec{D} = \left|\vec{C}.\vec{X_P}(t) - \vec{X}(t)\right| \tag{2}$$

$$\vec{X}(t) = \vec{X_P}(t) - \vec{A}.\vec{D} \tag{3}$$

where $\vec{X_P}$ - is the prey vector's location, $\vec{X}$ denotes a gray wolf position vector, $\vec{A}$ and $\vec{C}$ represent coefficient vectors represent coefficient vectors, and $(t)$ indicates the current iteration.

The following is the computation of the vectors $\vec{A}$ and $\vec{C}$ are:

$$\vec{A} = 2\vec{a}.\vec{r1} - \vec{a} \tag{4}$$

$$\vec{C} = 2\vec{r2} \tag{5}$$

with $r1, r2$ being random vectors in [0, 1], and components of $\vec{A}$ They are decreasing linearly from 2 to 0 throughout iterations.

Figure 3 is an illustration of the GWO algorithm's pseudo-code. The hunting habits of grey wolves can be quantitatively simulated using the following equations:

$$\vec{D}_\alpha = \left|\vec{C}_1.\vec{X}_\alpha - \vec{X}\right|, \vec{D}_{\beta=} \left|\vec{C}_2.\vec{X}_\beta - \vec{X}\right|, \vec{D}_\delta = \left|\vec{C}_3.\vec{X}_\delta - \vec{X}\right| \tag{6}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1.\left(\vec{D}_\alpha\right), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2.\left(\vec{D}_\beta\right), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3.\left(\vec{D}_\delta\right) \tag{7}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{8}$$

## IV. Simulation Results

A thorough examination of the experimental design and the outcomes of the suggested methodology are provided in this section. In many applications, pedestrian detection is essential. Nevertheless, current datasets frequently generate false positives when human-like objects are found. A new dataset containing human-like items such as statues, mannequins, scarecrows, and robots is produced to solve this problem. The objective of this dataset is to enhance model discrimination and decrease misidentifications.

It is a valuable resource for improving pedestrian detection techniques since it offers real-world scenarios, diverse environments, and complex variants. The dataset, 196 MB, can be obtained for free at Kaggle.com.

The settings that were applied for the simulation are listed in Table I.

*Initialize the grey wolf population $X_i$ (i = 1, 2, ..., n)*
*Initialize a, A, and C*
*Calculate the fitness of each search agent*
*$X_\alpha$=the best search agent*
*$X_\beta$=the second best search agent*
*$X_\delta$=the third best search agent*
***while** (t < Max number of iterations)*
  ***for** each search agent*
    *Update the position of the current search agent by equation (3.7)*
  ***end for***
  *Update a, A, and C*
  *Calculate the fitness of all search agents*
  *Update $X_\alpha$, $X_\beta$, and $X_\delta$*
  *t=t+1*
***end while***
*return $X_\alpha$*

Fig. 3. Pseudo code of the GWO algorithm.

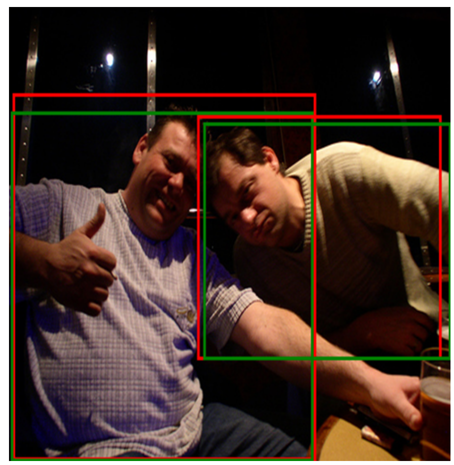
TABLE I.
SYSTEM PARAMETERS

| Parameter | Value |
|---|---|
| Number of epochs | 10 |
| Learning rate | 0.0001 |
| Momentum | 0.9 |
| Batch size | 16 |
| Image dimension | (227) *(227) |
| population size | 5 |

The study focuses on face detection using the Faster-RCNN architecture and optimizing the detection threshold by implementing the Gray Wolf Optimization (GWO) algorithm. Preprocessing is crucial, involving resizing images to a standardized 227x227 pixels, converting grayscale to RGB format, and partitioning the dataset into training (70%) and test (30%) sets. In addition, the Python programming language was used for the execution. The next step is training the Faster-RCNN model with a ResNet 50 pre-trained network. The Stochastic Gradient Descent (SGD) is employed with parameters set to 10 epochs, learning rate of 0.0001, and momentum of 0.9. The threshold parameter in face detection is adjusted using GWO to balance precision and recall. The cost function considers both region and classification performance.

The GWO algorithm's hyperparameters are chosen with 4 iterations, a population size of 5, and threshold bounds of 0.001 and 0.99. The obtained optimal threshold is presented in Table II. Figure 4 demonstrates the performance evaluation of the proposed method. The following sections comprehensively evaluate the trained Faster-RCNN using this approach.

(a)



(b)

(c)

(d)

Fig. 4. The performance evaluation of the proposed method on some sample test dataset's images (red boxes: detected objects, green boxes: targets).

TABLE II.
THE OPTIMAL THRESHOLD OBTAINED FROM THE GRAY WOLF OPTIMIZATION ALGORITHM

| Parameter | Value |
|---|---|
| Threshold | 0.7911 |

### *A. Intersection over union plot*

This section introduces the Intersection over Union (IoU) plot, a crucial tool for evaluating our proposed face detection method's instance-level accuracy. The IoU plot illustrates the overlap between predicted and ground truth bounding boxes. The IoU is computed by dividing the intersection area of the predicted bounding box by the intersection area of the actual bounding box. It quantifies how accurately the predicted bounding box aligns with the actual face instance. A higher IoU score indicates better spatial alignment, as demonstrated in Figure 5.

The average Intersection over Union (IoU) score of 0.8479, which we received with our proposed approach, is excellent. The projected bounding boxes and the ground truth annotations display a strong and constant alignment, as indicated by this. In Figure 6, the technique's overall performance across different test situations is represented by using the IoU plot.

### *B. Average precision plot*

The average precision (AP) metric—a vital measure of our proposed face identification technique's efficacy—is displayed in this section. The precision-recall exchange-off of our method is established via the AP plot. We can also evaluate the accuracy of high-quality predictions regarding all optimistic predictions through charting precision versus taking into recall levels. The ratio of successfully identified positives to all positives in the dataset is known as recall. A high recall suggests that our technique detects significant positive, compelling cases, whereas a high precision suggests low false positives. Our face detection method has displayed steady accuracy and reliability in figuring out faces throughout diverse settings, boasting an impressive average precision (AP) of 0.8749.

Figure 7 illustrates the AP graph, highlighting the technique's reliable overall performance across distinct consider thresholds. Our model constantly detects a high percentage of authentic acceptable instances, as evidenced using its vital precision, and maintains accurate detections even at excessive keep-in-mind levels. The ordinary effectiveness of our proposed technique is established through the location under the AP curve (AP-AUC), which consolidates all detection criteria into an unmarried metric. With a promising AP rating of 0.8749, our method is capable of real-international packages, facial reputation, self-riding motors, and surveillance structures. To finish, our model reveals overall solid performance in facial recognition, as evidenced by its regular AP rating throughout all recollect stages.

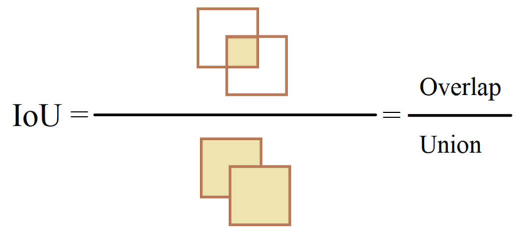



Fig. 5. Shows how intersection over union is evaluated.

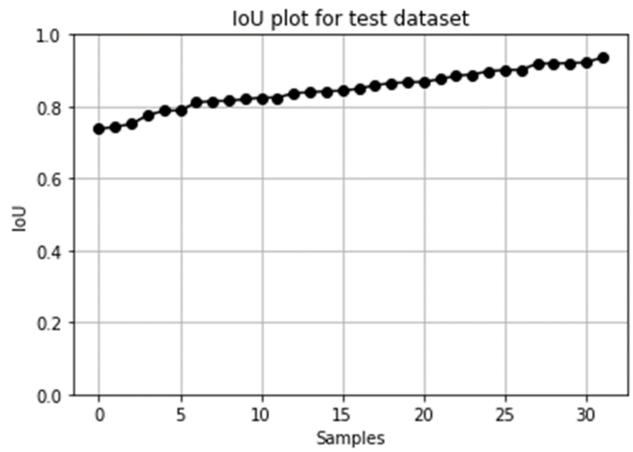



Fig. 6. The IoU graph showing the effects of the advocated technique.


Fig. 7. The suggested method's AP plot.

### C. Statistical Evaluation of the Proposed Method Using the Introduced Metrics

This phase provides an entire statistical evaluation of our proposed facial approach. It compares its effectiveness based on four key metrics: detection time, average precision (AP), accuracy, and intersection over union (IoU).

The accuracy metric assesses the model's functionality to locate and perceive items inside pictures. Higher accuracy suggests progressed object detection and localization. Our approach achieves an impressive accuracy rate of 94.0% for facial localization and recognition, as established in Figure 8.

The average precision (AP) metric is hired to assess the trade-off between bearing in mind and precision. Our method achieves a sturdy and reproducible AP rating of 0.8749, indicating a high precision and great remember charge. This score showcases the model's performance in item detection and localization.

The intersection over union (IoU) metric measures the overlap among expected and ground truth bounding bins, providing insights into the best of object localization. Our method demonstrates effective facial area localization within images, with an outstanding IoU score of 0.9123.

In addition, our technique's famous computational performance makes it suitable for real-time programs, as evidenced by using the expected detection time of 0.042 seconds in line with the image. This similarly highlights the practicality and suitability of our method for actual-time facial reputation scenarios.

Our proposed method achieves excessive detection, precision, and accuracy charges. It showcases a mean IoU rating of 0.8749, demonstrating accurate facial localization across various eventualities. The constant high IoU values support the model's ability to detect faces accurately regardless of size, orientation, or occlusion. The average recognition Timing is a measure of computational efficiency. At 133.81 frames per second (fps), our method processes up to 7.4731 milliseconds, an astounding average detection time. Due to its exceptional speed, video processing applications can achieve real-time face detection with low latency. The bar graph in Figure 8 summarizes the statistical analysis by visually representing these critical measures.

### D. Comparison

In this section, we dedicated a thorough comparison examination of our face detection thesis results with the state-of-the-art techniques and current approaches used in this field. To illustrate its unique benefits, novel contributions, and potential drawbacks, we juxtapose our proposed method's performance metrics and results with those of existing approaches. This thorough study clarifies the effectiveness of our approach and provides suggestions for its application in various real-world contexts.

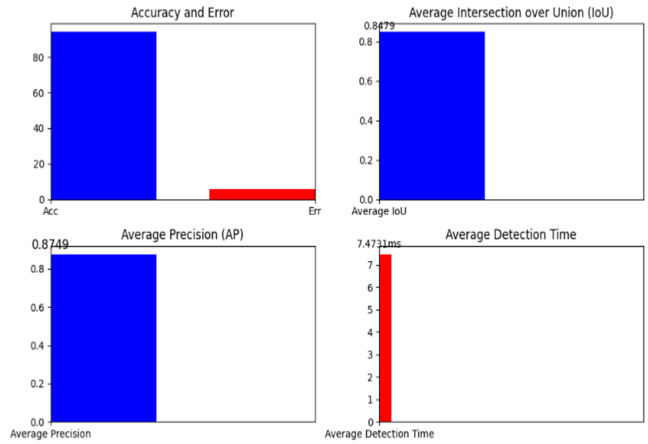



Fig. 8. Comprehensive assessment of our method.

Through this in-depth analysis, we intend to provide insight into how our innovative method expands the field of computer vision and defines the limits of face recognition. Convolutional neural networks (CNNs) are a powerful tool for facial feature detection, and this study [46] set out to create a system that could recognize faces with high accuracy even in the face of occlusion, scale, and position changes. The suggested method uses a region proposal network to find potentially exciting areas (ROIs) with promising characteristics. Following their initial training for image classification, these ROIs are input into the final convolutional layer of the pre-trained ResNet-50 model. Used a secondary network to categorize these ROIs into face and non-face data. It improved its feature vectors by combining the final layers of ResNet blocks Res-Block-2 (global features) and Res-Block-3 (local features) to create multi-scale feature maps. When tested on the benchmark datasets WIDER FACE, this method dramatically beats a Faster R-CNN model employing ResNet-50 as a starting point:

- On the WIDER FACE, the simple subset accuracy rises to 77.23
- Attained an accuracy of 76% for the WIDER FACE medium subgroup.
- Attained an accuracy of 73.11% on the difficult WIDER FACE hard subset.

In conclusion, this method combines ResNet-50 features, ROI generation/classification, and multi-scale feature maps to produce an accurate and reliable face identification system resistant to typical problems.

The Faster R-CNN model for accurate small object identification is improved in this study [47]. The enhancements include using bilinear interpolation in the (RoI) pooling layer to eliminate localization errors and an improved IoU loss function (LIIoU) for more precise bounding box regression. Multi-scale convolutional feature fusion is employed to improve feature maps during the recognition stage, and an improved Soft-NMS non-maximum suppression technique is applied to prevent missing overlapping items. The TT100K dataset, which has tiny indicators with sizes ranging from 12 to 60 pixels, is used for the evaluation. The outcomes attained are as follows:

- For smaller than 32-pixel signs, an 87% accuracy rate is achieved.

The suggested strategy outperforms the performance of existing models [48, 49] in terms of accuracy for minor signs up to 32 pixels when compared to the baseline Faster R-CNN model.

In conclusion, the Faster R-CNN changes described in this study significantly improve tiny object detection accuracy while lowering the incidence of missed detections. In the context of traffic sign detection, this enhancement is particularly notable. The comparison results of these 2 papers and our proposed method are summarized in Table III.

## V. Conclusion

This paper proposes a novel deep learning-based face detection methodology to address critical challenges like facial expressions, occlusions, and illumination changes. Our approach integrates a fine-tuned ResNet-50 architecture with optimal hyperparameter configuration achieved using Gray Wolf optimization. Meticulous data preprocessing proved pivotal in enhancing robustness across datasets.

Comprehensive experimentation was performed on our facial image dataset. Our model achieved 94% training and similar test accuracy, demonstrating efficacy in mitigating false positives and negatives. Thus, the proposed technique marks a significant advancement over conventional methods regarding precision and adaptability.

This research has shown promising results but further investigation is merited. Future work should explore multimodal integration for more robust representations, implement privacy preservation, customize for specific applications like crowd surveillance, and integrate detection with recognition tasks for identification. Our face detection approach introduces an important innovation with immense promise for real-world deployment demanding efficiency and accuracy.

## Acknowledgments

This work is supported by the College of Engineering/ Mustansiriyah University.

## Conflict of Interest

The authors have no conflict of relevant interest to this article.

## Author Contribution Statement

Both Authors have proposed the research problem, performed the computations, and supervised the findings of this work. Also, both discussed the results and contributed to the final manuscript.

## References

[1] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, (San Diego, USA), pp. 947–954, IEEE, 2005.

[2] E. Wright, "The future of facial recognition is not fully known: Developing privacy and security regulatory mechanisms for facial recognition in the retail sector," *Fordham Intell. Prop. Media & Ent. LJ*, vol. 29, p. 611, 2018.

[3] R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.-C. Chen, C. D. Castillo, and R. Chellappa, "A fast and accurate system for face detection, identification, and verification," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 2, pp. 82–96, 2019.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[6] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Columbus, USA), pp. 1891–1898, 2014.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Columbus, USA), pp. 815–823, 2015.

[8] J. P. C. Cardozo, "Next generation identification (ngi): Um novo conceito em biometria forense," *Acta de Ciências e Saúde*, vol. 1, no. 1, pp. 1–5, 2016.

TABLE III.
COMPARISON OF OUR METHOD WITH OTHERS IN THIS LITERATURE

| Reference | Dataset | Method | Accuracy% |
|---|---|---|---|
| [46] | (Easy, Medium, Hard), WIDER FACE | Faster R-CNN + ResNet-50 | 77.23% (WIDER FACE Easy), 76.0% (WIDER FACE Medium), 73.11% (WIDER FACE Hard) |
| [47] | TT100K | Improved Faster R-CNN | 87% |
| This Study | Pedestrian Detection | ResNet-50 + Faster RCNN with GWO | 94% |


[9] D. Mainenti, "User perceptions of apple's face id," *Information Science, Human Computer Interaction*, vol. DIS805, pp. 1–16, 2017.

[10] K. A. Gates, *Our biometric future: Facial recognition technology and the culture of surveillance*, vol. 2. New York, USA: NYU Press, 2011.

[11] I. D. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee, and E. Denton, "Saving face: Investigating the ethical concerns of facial recognition auditing," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, (New York, USA), pp. 145–151, 2020.

[12] A. Cavoukian, "Privacy by design: Leadership, methods, and results," in *European Data Protection: Coming of Age* (S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, eds.), pp. 175–202, Dordrecht: Springer Netherlands, 2013.

[13] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 618–634, 2018.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, pp. 1–9, 2015.

[15] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46–61, 2014.

[16] J. Harikrishnan, A. Sudarsan, A. Sadashiv, and R. A. Ajai, "Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision," in *2019 international conference on vision towards emerging trends in communication and networking (ViTECoN)*, (Vellore, India), pp. 1–5, IEEE, 2019.

[17] E. Winarno, I. H. Al Amin, H. Februariyanti, P. W. Adi, W. Hadikurniawati, and M. T. Anwar, "Attendance system based on face recognition system using cnn-pca method and real-time camera," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, (Yogyakarta, Indonesia), pp. 301–304, IEEE, 2019.

[18] F. Faisal and S. A. Hossain, "Smart security system using face recognition on raspberry pi," in *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, (Island of Ulkulhas, Maldives), pp. 1–8, IEEE, 2019.

[19] H. Asil and J. Bagherzadeh, "Proposing a new method of image classification based on the adaboost deep belief network hybrid method," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 5, pp. 2650–2658, 2019.

[20] M. S. H. Al-Tamimi, "Combining convolutional neural networks and slantlet transform for an effective image retrieval scheme," *Int. J. Electr. Comput. Eng*, vol. 9, no. 5, pp. 4382–4395, 2019.

[21] G. Saadouli, M. I. Elburdani, R. M. Al-Qatouni, S. Kunhoth, and S. Al-Maadeed, "Automatic and secure electronic gate system using fusion of license plate, car make recognition and face detection," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, (Doha, Qatar), pp. 79–84, IEEE, 2020.

[22] K. Pranav and J. Manikandan, "Design and evaluation of a real-time face recognition system using convolutional neural networks," *Procedia Computer Science*, vol. 171, pp. 1651–1659, 2020.

[23] M. A. Al-Jbaar, M. J. Zedan, M. T. Thameen, and O. A. Mahdi, "Face recognition security system based on convolutional neural networks," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 5177 – 5187, 2020.

[24] M. Shah, D. Shukla, and D. Pandya, “Smart gate: Intelligent security system based on face recognition,” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 3, pp. 601–608, 2020.

[25] S. Kakarla, P. Gangula, M. S. Rahul, C. S. C. Singh, and T. H. Sarma, “Smart attendance management system based on face recognition using cnn,” in *2020 IEEE-HYDCON*, (Hyderabad, India), pp. 1–5, IEEE, 2020.

[26] S. Kakarla, P. Gangula, M. S. Rahul, C. S. C. Singh, and T. H. Sarma, “Smart attendance management system based on face recognition using cnn,” in *2020 IEEE-HYDCON*, pp. 1–5, IEEE, 2020.

[27] N. Zhang, J. Luo, and W. Gao, “Research on face detection technology based on mtcnn,” in *2020 international conference on computer network, electronic and automation (ICCNEA)*, (Xi’an, China), pp. 154–158, IEEE, 2020.

[28] N. S. Irjanto and N. Surantha, “Home security system with face recognition based on convolutional neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, 2020.

[29] M. Geetha, R. Latha, S. Nivetha, S. Hariprasath, S. Gowtham, and C. Deepak, “Design of face detection and recognition system to monitor students during online examinations using machine learning algorithms,” in *2021 international conference on computer communication and informatics (ICCCI)*, (Coimbatore, India), pp. 1–4, IEEE, 2021.

[30] A. H. Rashed and M. H. Hamd, “Robust detection and recognition system based on facial extraction and decision tree,” *Journal of Engineering and Sustainable Development*, vol. 25, no. 4, pp. 40–50, 2021.

[31] S. Pokhrel, A. Panta, S. Parajuli, S. Poudel, R. Adhikari, and M. Guragai, “Remote monitoring and face recognition security enabled smart home using iot,” *International Journal of Advanced Engineering*, vol. 5, no. 1, pp. 47–52, 2022.

[32] T. H. Obaida, A. S. Jamil, and N. F. Hassan, “Real-time face detection in digital video-based on viola-jones supported by convolutional neural networks,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, p. 3083, 2022.

[33] H. Huo, Y. Yu, and Z. Liu, “Facial expression recognition based on improved depthwise separable convolutional network,” *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18635–18652, 2023.

[34] S. Saurav, P. Gidde, R. Saini, and S. Singh, “Dual integrated convolutional neural network for real-time facial expression recognition in the wild,” *The Visual Computer*, vol. 38, no. 3, pp. 1083–1096, 2022.

[35] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Las Vegas, NV, USA), pp. 770–778, 2016.

[36] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Honolulu, HI, USA), pp. 1492–1500, 2017.

[37] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City, UT, USA), pp. 7132–7141, 2018.

[38] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City, UT, USA), pp. 7794–7803, 2018.

[39] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[40] B. Singh, M. Najibi, and L. S. Davis, “Sniper: Efficient multi-scale training,” *Advances in neural information processing systems*, vol. 31, pp. 1–11, 2018.

[41] P. Li, J. Xie, Q. Wang, and Z. Gao, “Towards faster training of global covariance pooling networks by iterative matrix square root normalization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City, UT, USA), pp. 947–955, 2018.

[42] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, “Relation networks for object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City, UT, USA), pp. 3588–3597, 2018.

[43] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Advances in engineering software*, vol. 69, pp. 46–61, 2014.

[44] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, “Grey wolf optimizer: a review of recent variants and applications,” *Neural computing and applications*, vol. 30, pp. 413–435, 2018.

[45] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Advances in engineering software*, vol. 95, pp. 51–67, 2016.

[46] H. Mliki, S. Dammak, and E. Fendri, “An improved multi-scale face detection using convolutional neural network,” *Signal, Image and Video Processing*, vol. 14, no. 7, pp. 1345–1353, 2020.

[47] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, and Z. Wu, “An improved faster r-cnn for small object detection,” *Ieee Access*, vol. 7, pp. 106838–106846, 2019.

[48] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, “Traffic-sign detection and classification in the wild,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Las Vegas, NV, USA), pp. 2110–2118, 2016.

[49] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, “Perceptual generative adversarial networks for small object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Honolulu, HI, USA), pp. 1222–1230, 2017.