Pertanika J. Sci. & Technol. 33 (5): 2155 - 2176 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Face Detection and Gender Classification by YOLO Algorithm

Aseil Nadhim Kadhim*, Syahid Anuar and Saiful Adli Ismail

Fakulti Teknologi dan Informatik Razak, Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

ABSTRACT

Gender classification is a fundamental computer vision problem used in a wide range of applications in surveillance and marketing. The work in this paper is to determine the gender classification ability of the You Only Look Once (YOLO) algorithm using deep learning. YOLO is one of the most accurate object detection models that can detect multiple objects in a video or picture in real-time. In this work, various versions of YOLO (YOLOv3 to YOLOv9) were compared to determine the most accurate and efficient model for gender classification. The work utilized a collection of 361 test images of male and female subjects in different scenarios of their settings, and the models' performance was gauged in terms of key metrics of Precision, Recall, and F1-score. The analysis of performance also confirmed that YOLOv9 was even better compared to its counterparts, registering a mean average precision (mAP) of 97%, a Precision of 86.8%, a Recall of 86.1%, and an F1-score of 86.54%. The processing time of the model was 0.332 seconds per picture, or a frame per second (FPS) of 3.00. The confusion matrix also recorded 157 true positives (TP), 25 false negatives (FN), 23 false positives (FP), and 156 true negatives (TN), a reflection of a highly balanced classification performance. The results confirm that YOLOv9 is highly accurate and efficient to use in gender classification in practical applications.

Keywords: Face detection, gender classification, object detection, YOLO

ARTICLE INFO

Article history:
Received: 18 October 2024
Accepted: 17 April 2025
Published: 28 August 2025

DOI: https://doi.org/10.47836/pjst.33.5.06

E-mail addresses:
n-20@graduate.utm.my (Aseil Nadhim Kadhim)
syahid.anuar@utm.my (Syahid Anuar)
saifuladli@utm.my (Saiful Adli Ismail)
* Corresponding author

INTRODUCTION

Gender classification is one of the most important applications of computer vision in surveillance, security, and marketing (Xiao et al., 2020). The gender is automatically classified in terms of a human's face features (Varnima & Ramachandran, 2020) using deep learning to be more accurate and efficient (Jabraelzadeh et al., 2023; Tilki et

e-ISSN: 2231-8526

al., 2021). Unlike traditional methods that apply hand-coded features, more recent deep learning methods, such as YOLO, enable gender classification in real-time straight from pictures (Sonthi et al., 2023). YOLO is a widely used object detection approach that can identify multiple objects in a single picture. The more recent versions of YOLO, such as YOLOv9, execute face detection and gender classification in a single step of inference without separate preprocessing procedures. This enhances processing time, in addition to improving precision in classification (Tejaswi et al., 2023). There have been various studies that have explored gender recognition using deep learning, noting that models that employ convolutional neural networks (CNN) in face feature processing work (Azhar et al., 2022). This study compares different versions of YOLO in gender classification based on a set of tagged face pictures. The aim is to determine the most efficient YOLO model that can be applied in gender classification in a way that is applicable in real-time and in a diverse range of situations (Ayo et al., 2022).

RELATED WORKS

With the increasing use of artificial intelligence in various sectors, face detection and gender identification between men and women pose primary challenges in security, surveillance, and data analysis. Gender identification is of crucial significance in various situations, such as in identity recognition systems, behavioral analysis, and enhancing the user experience in smart services. From a security aspect, face detection and gender identification models of high efficiency need to be developed, mainly in areas of public use such as airports and stations (Awan et al., 2021; Ayo et al., 2022; Singh et al., 2021). Several models of gender identification in real-time via CNN have been put forward (Xiao et al., 2020). The traditional method of CNN is separated into two key stages (Sumi et al., 2021) and has evolved in different ways, i.e., R-CNN, Fast-RCNN, and Faster-RCNN. Although CNN is found to be extremely efficient in gender detection, it is also afflicted with a set of serious limitations, i.e., (1) multiple stages of training, (2) high memory and computational demands, and (3) slower detection compared to other methods, rendering detection speed enhancement a necessity (Raman et al., 2023; Ramya et al., 2022). On the contrary, YOLO models have gained immense popularity in face detection and gender recognition due to their high efficiency and speed. As compared to CNN-based approaches, YOLO processes one entire picture in a one-time step, in keeping with the "You Only Look Once" method, rendering it much more efficient (Raman et al., 2023; Ramya et al., 2022). Given these advantages, this work is focused on comparing and evaluating different YOLO models to determine the most efficient one to apply in real-time face detection and gender recognition. The latest one, YOLOv9, is selected here because it is more computationally efficient compared to previous versions of YOLO. However, challenges in gender classification accuracy in public places persist. Hence, it is important to evaluate YOLO models in such an environment

(Sonthi et al., 2023). Consequently, this work attempts to compare various implementations of YOLO to determine the most appropriate model for face detection in real-time in actual applications (Jabraelzadeh et al., 2023).

Face Detection

Face detection is a fundamental computer vision problem, and various methods of making it more accurate and efficient have been proposed (Sonthi et al., 2023). Even though face recognition is one of computer vision's fundamental challenges, there have been proposals in the literature to process face features. The face detection is a beginning point in this process, given that it facilitates gender identification using face-based properties (Tejaswi et al., 2023).

Gender Classification

Gender classification is a crucial aspect of social interaction in most intelligent applications, i.e., human-machine interaction (Ayo et al., 2022; Azhar et al., 2022). The most effective method of gender classification is face image analysis, and face detection is a fundamental starting process in gender classification (Awan et al., 2021; Singh et al., 2021). Research has extensively explored various gender classification methods, of which face-based methods predominate (Mittal et al., 2023). The face detection coupled with gender classification is of interest in scientific research (Jabraelzadeh et al., 2020; Shet et al., 2022). This study is distinct in using face feature analysis in gender identification (Shen et al., 2021), one of the most accurate yet practical approaches in real applications (Al-Bayati & Abood, 2023; Alsaif et al., 2023; Srinivasan, 2023).

Detection Algorithms

Object detection and classification models can be divided into two groups of interest: one is region-based classification using CNN, and the other is one-stage models like YOLO. The models using CNN follow a two-stage approach of classification (Dutta et al., 2022; Ullah, 2020). However, despite their accuracy, there also lie certain limitations of CNN models, such as multiple processes of training and high demands on computation, thus making it slower compared to other approaches (Nepal & Eslamiat, 2022; Srikar & Malathi, 2022). YOLO, on the other hand, detects and classifies in a single step without a separate phase of classification. The combined approach enhances speed in addition to precision, making YOLO more suitable in real-time applications (Eşer & Hardalaç, 2021; F. Wang et al., 2023; Gan et al., 2023; Jiang et al., 2022; Thakuria & Erkinbaev, 2023; W. Wang et al., 2023; Zhang et al., 2020). Due to its efficiency, YOLO is widely used in face detection and gender classification, where real-time performance is essential.

Research Goal

This study seeks to assess YOLO models in gender classification, i.e., whether a detected face is male or female. Considering that existing studies affirm YOLO's speed over CNN, this work aims to compare various versions of YOLO to determine the most accurate and efficient of these models in real-time applications for gender classification.

MATERIALS AND METHODS

This study attempts to determine the effectiveness of YOLO-based deep learning models in face detection and gender recognition. Compared to methods that use separate face detection and classification networks, YOLO brings the two processes in one system, making it more efficient and capable of real-time processing. The primary objective of this work is to find the best YOLO model for gender classification in terms of its accuracy, inference time, and resource requirements.

The methodology is a sequential process of ordered processes that encompasses data collection, image preprocessing, augmentation, model training, evaluation, and comparative analysis. The work is divided into two major phases: first, YOLO models are used to detect faces to accurately detect and cut faces out of pictures. Then, in phase two, the same YOLO model is used to classify gender directly, in which each face is automatically categorized as male or female without applying an external classifier. This integrated approach ensures a seamless and optimized classification process, making YOLO a robust solution for real-time gender classification.

Data Collection

The dataset for this study was collected from the students at the University of Babylon, Iraq, ensuring diverse real-world conditions to enhance model robustness. Images were captured under various lighting conditions (sunlight, cloudy, backlighting) and different facial angles (frontal, side, tilted) to improve detection accuracy. The dataset includes individuals with and without head coverings, varying age groups, and diverse facial expressions. Partial occlusions due to environmental factors were also considered to simulate real-world challenges. A total of 140 images were collected, ensuring the trained model's reliability across different scenarios. Figure 1 shows examples of photos of people's faces.

Image Processing and Data Segmentation with Augmentation

A systematic preprocessing and augmentation strategy was applied to optimize the dataset for training and evaluation. Initially, the dataset consisted of 140 high-quality images captured most within the university campus, encompassing diverse lighting conditions, facial orientations, and real-world variations. To standardize input dimensions, all images

Figure 1. Examples of student photos taken in various settings: (a) Normal frontal face image, (b) Natural light angle, (c) Heavy occlusion (e.g., glasses or mask), (d) Side light angle, (e) Overlapping group of students, and (f) Backlight angle

were resized to 640 × 640 pixels and converted to grayscale, with pixel values normalized within the 0 to 1 range for consistency in model interpretation. Facial features were emphasized by cropping each image to retain 25 to 75% of both horizontal and vertical regions, minimizing background interference. Additionally, images were segmented into a 2 × 2 grid (four equal parts) to enhance localized feature extraction, improving the precision of face detection and gender classification.

To further enhance model generalization and prevent overfitting, 25% of the dataset underwent augmentation, incorporating horizontal and vertical flipping, cropping adjustments, hue modifications (+25° to -25°), brightness variations (±25%), and Gaussian blurring (up to 2.5 pixels) to introduce controlled noise. The Static Crop function was also utilized to generate more variations by flipping images over specific areas and axes to enrich the dataset even further Figure 2.

As a result of expansion and segmentation, the dataset was increased to 361 images to give a more representative and diverse set of training. The final dataset was divided into three groups: 316 for training to learn to better optimize the YOLO model, 30 for validating to set hyperparameters and to prevent overfitting, and 15 for final testing to verify generalization performance. The new structured data preprocessing approach significantly improved gender recognition using YOLO to be more resilient in various practical situations Figure 3.

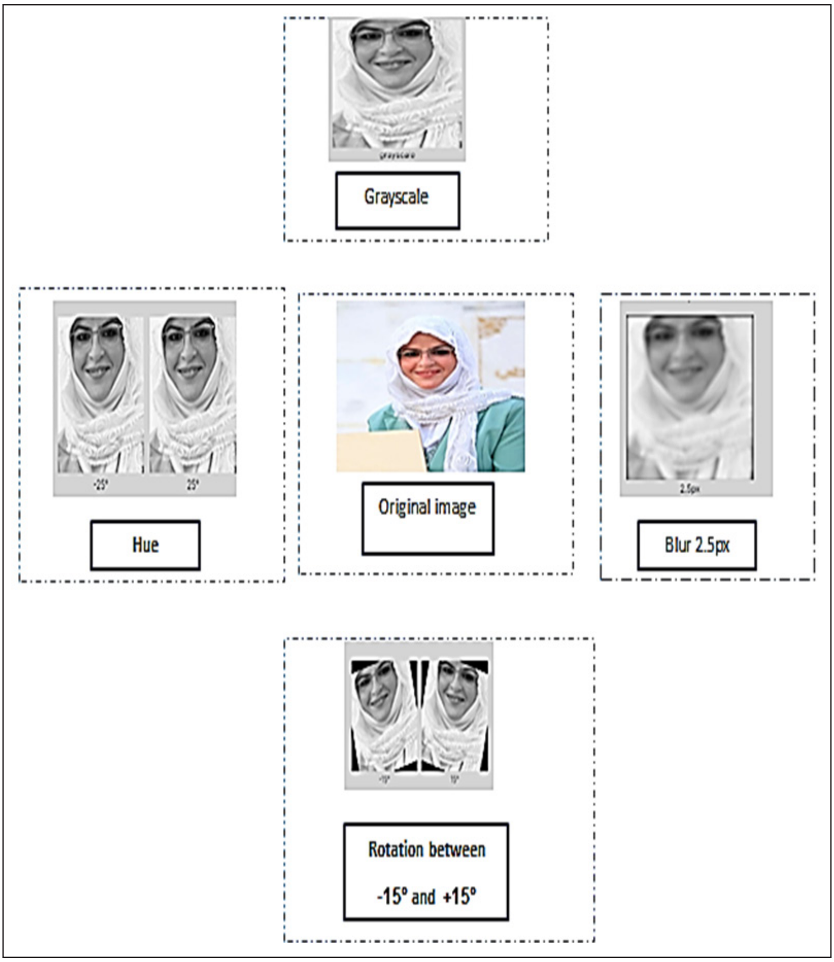


Figure 2. Image processing and data set segmentation

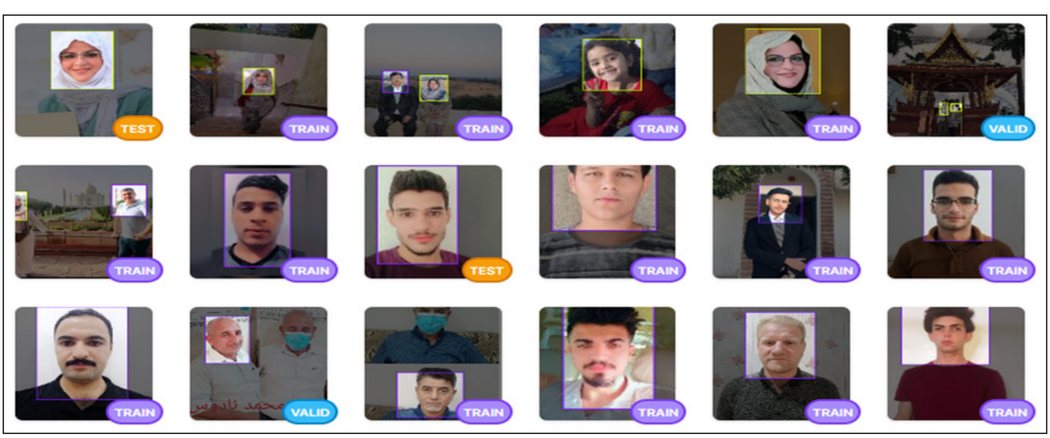


Figure 3. Augmented training set and distributed

YOLOv9 Network Architecture

YOLOv9 represents a milestone in one-stage object detection models that is intended to find a balance between speed and accuracy in real-time applications. As per benchmark tests, YOLOv9 can process between 5 and 160 frames per second (FPS), depending on hardware specifications and input resolution (Figure 4). Overview of the proposed research process, including training and testing phases using YOLO-based gender classification. (Acharya, 2024a; Boesch, 2024). The network architecture is structured into three primary components:

- Internal Network: Handles data preprocessing and feature extraction.
- Core Network: Obtains high-level representations to facilitate object recognition.
- Main Network: Carries out end detection and classification processes.

Key enhancements:

- Object detection in real-time: Maintains high detection speed without sacrificing precision.
- Programmable Gradient Information (PGI) integration: A stand-alone process in YOLOv9 that allows smooth flow of information and reduces loss of features to support better model efficacy and accuracy.
- Cross Stage Partial Network (CSPNet): Improves computational efficiency and reduces memory requirements by splitting features into two streams and merging them afterwards.
- Generalized Efficient Layer Aggregation Network (GELAN): An advanced version of ELAN that enhances computational efficiency, reduces resource use, and enhances accuracy and inference time.
- ELAN (Efficient Layer Aggregation Network): Enhances information flow within the network, reducing feature loss and improving deep learning.
- Improved performance: Outperforms state-of-the-art models on datasets like MS COCO, excelling in accuracy, speed, and overall efficiency for real-time gender classification (C.-Y. Wang et al., 2024).

Here is a step-by-step process of detection using the YOLOv9 model:

1. Input Processing: The model segments input images into a grid of cells, each responsible for detecting objects within its assigned region.
2. Bounding Box Prediction: YOLOv9 generates bounding boxes for detected objects, defining their position using x, y (center point), width, and height.

3. Object Classification: Each bounding box is assigned a classification probability, determining the likelihood that it belongs to a specific class.
4. Non-Maximum Suppression: To refine detections, redundant bounding boxes are filtered based on confidence scores, ensuring accurate localization.
5. Post-Processing: After suppression, the most reliable bounding boxes are retained, enabling precise real-time detection (Boesch, 2024; C.-Y. Wang et al., 2024).

YOLOv9 integrates PGI for reliable gradient generation and deep feature retention, while GELAN optimizes parameters, accuracy, and inference speed for enhanced efficiency (Boesch, 2024; C.-Y. Wang et al., 2024) as shown in Figure 5.

Training Platform and Parameter Settings

This study applied transfer learning with the use of the YOLOv9 algorithm in gender identification, leveraging its object detection in real-time. Training was done using Google Colab on a Tesla T4 GPU to provide high processing capability. To ensure that it was working to its optimal capacity, it was trained for 300 epochs at a batch size of 16 samples per iteration and an input image size of 640 × 640 pixels. The learning rate was initially set to 0.001 and was dynamically adjusted during training to facilitate better convergence and model stability.

Model Establishment and Assessment Indicators

Establishment of Model

After dataset preprocessing (as described in the methodology), the YOLOv9 model was trained using the allocated training set and optimized through validation. The final model performance was then assessed on the test set to ensure its effectiveness in gender classification. The evaluation process and performance results are illustrated in Figure 4.

Model Evaluation

For quantification, the Complete Intersection Over Union (CIOU) loss function was employed. The difference between the calibration and prediction squares was analyzed effectively (Zheng et al., 2020) with A representing the calibration bin, B representing the prediction bin, and L1 representing the distance. It shows the specifications based on the model's predicted and validation bins required for computing the CIOU. The diagonal length, or L2, is the separation between the centroids of minimum-Square Rectangle A and B.

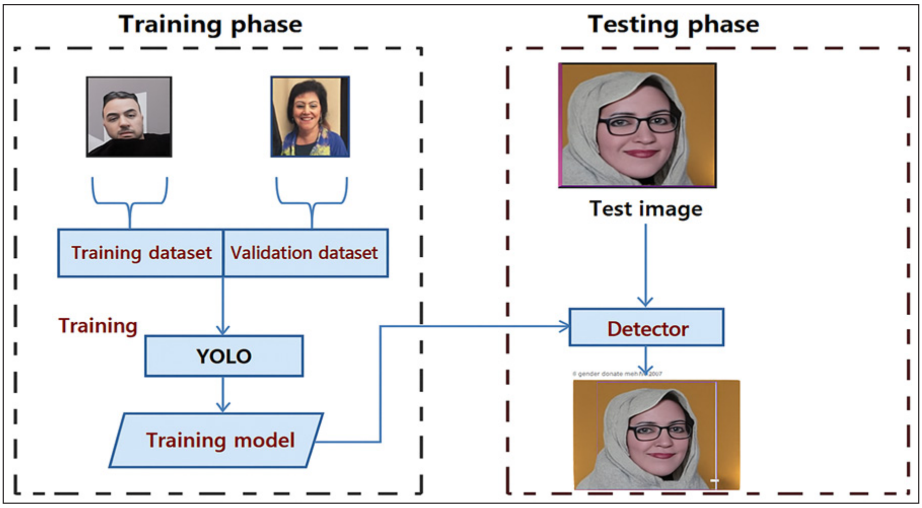


Figure 4. Process of the suggested research

A visual representation of the CIOU computation between the ground truth and the model prediction box is shown in Figure 6. The blue box represents the prediction box, while the calibration box is represented by the yellow box.

CIOU was calculated as follows (Equation 1):

$$Loss_{CIou} = 1 - IOU + \frac{L1^2}{\mathrm{L}2^2} + \alpha v \quad [1]$$

where α is the balancing component among the loss brought on by Intersection over Union (IOU) and V, and v is the aspect ratio similarity between boxes A and B. The study employed the following metrics F1 score, recall, precision, and mAP to evaluate model performance in an impartial and consistent manner. Accuracy, which is determined by dividing the total number of detected targets by the number of correct targets, is the most commonly used evaluation index. In general, accuracy increases detection efficacy. While high precision is necessary for grading, it is not usually indicative of genius. Therefore, to assess comprehensiveness, mAP, recall, and F1 are shown. This section displays the calculations for the mAP, F1, recall, precision, and F1 outcomes (Achary, 2024b):

$$\text{Precision}, p = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \times 100\% \quad [2]$$

$$\text{Recall}, \mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100\% \quad [3]$$

$$\text{Average Precision}, \mathrm{AP} = \mathrm{P} \int_0^1 .\,(\mathrm{r})\mathrm{dr} \quad [4]$$

$$\text{Mean Average Precision, mAP} = \frac{1}{n}\sum_{i=1}^{n} \text{AP1} \quad [5]$$

$$\text{F1 score, F1} = 2 \times \frac{P \times R}{P + R} \quad [6]$$

False negative (FN) refers to the total number of gender-related objects that were not identified or ignored. In contrast, false positive (FP) represents the percentage of objects that

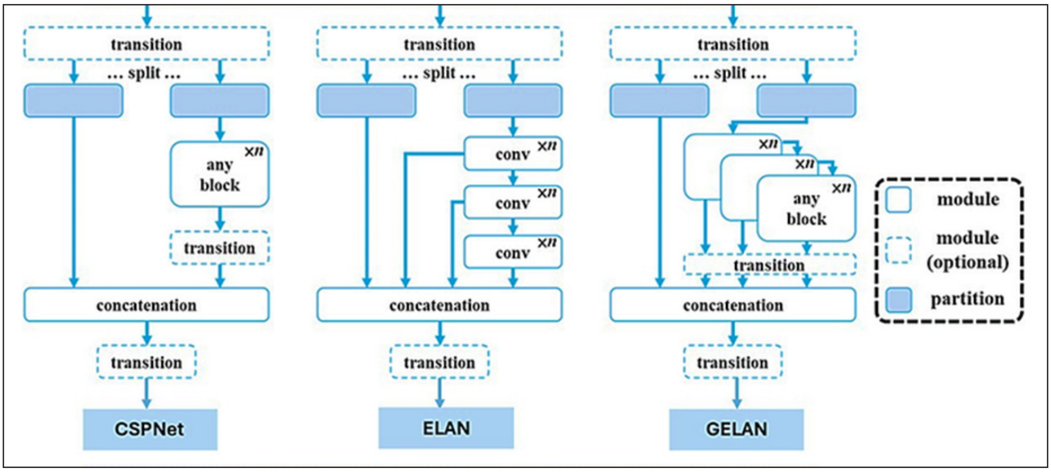


Figure 5. Architecture of YOLOv9 network (Yaseen, 2024)
Note. CSPNet = Cross Stage Partial Network; ELAN = Efficient Layer Aggregation Network; GELAN = Generalized Efficient Layer Aggregation Network

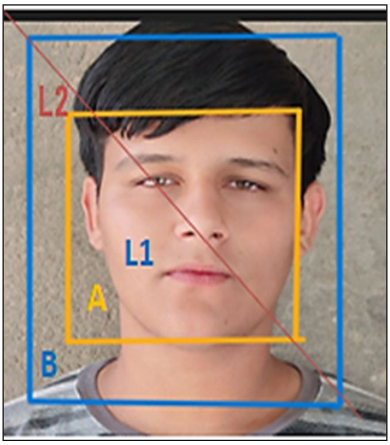


Figure 6. The prediction box
Note. A = Calibration (ground truth) box; B = Predicted box by the model; L1 = Euclidean distance between the centroids of boxes A and B; L2 = Diagonal length of the smallest enclosing box covering both A and B (used for Complete Intersection Over Union [CIOU] calculation)

were misidentified as gender. True positive (TP) refers to the number of objects correctly identified as a gender.

RESULTS AND DISCUSSION

Training Dataset of YOLO Models

Figure 7 shows an example of the complete pipeline used for gender classification and gender recognition results. The red color represents the category designated for men, and the blue color represents the category designated for women. During testing of images using YOLO models, it was discovered that the YOLOv9 algorithm achieved the highest detection accuracy.

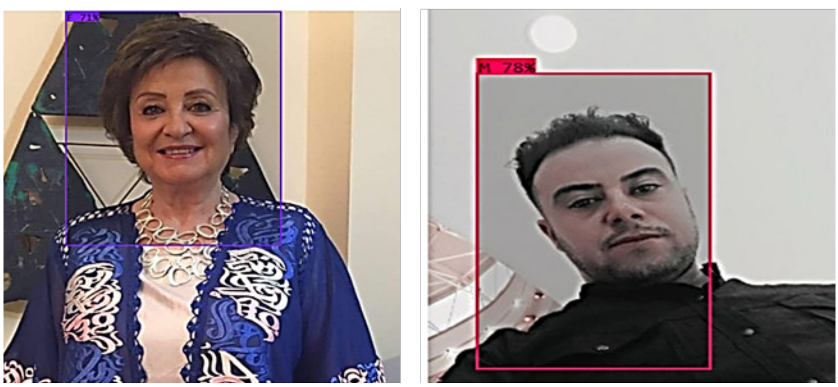


Figure 7. Results of the real-time gender classification

Using the YOLOv9 algorithm, gender detection has achieved accuracy rates of up to 97%. Figure 8 illustrates the mAP values for this model. Overall, YOLOv9 surpasses YOLOv8 and other YOLO versions in terms of speed, accuracy, and scalability. This performance enhancement is attributed to advancements in model architecture, improved training techniques, and the integration of new features such as PGI and GELAN.

The loss curves of YOLOv9 in Figure 9 reflect a constantly improving performance coupled with stabilization of models during learning. Both Box Loss and Object Loss exhibit a steep decline that is subsequently followed by stabilization, indicating improving precision in object localization. Similarly, Class Loss reduces steeply in early training stages and plateaus at a low value, suggesting the strength of the model in gender classification. The patterns also suggest YOLOv9's capability of learning in a more efficient manner and having a high performance, making it a reliable model for gender classification.

The loss curves of YOLOv8 in Figure 10 exhibit progressive learning and stabilization during training. Box Loss and Object Loss decline steeply in earlier epochs and stabilize to smaller values, suggesting improved object localization. Similarly, Class Loss also drops steeply and remains low, demonstrating effective gender classification. However, compared to YOLOv9, YOLOv8 displays slightly more oscillations, suggesting potential instability in a number of its phases. The results verify that even though YOLOv8 is efficient in learning and has a faster convergence, YOLOv9 is a better model due to its higher stability and overall performance in gender classification.

The loss curves of YOLOv7 in Figure 11 exhibit a smooth improvement in performance during training, though there is a visible oscillation. Both Box Loss and Object Loss decrease steeply initially, after that there is a slower stabilization, indicating improvement in object localization though there is a lack of stability. Class Loss has high fluctuations prior to stabilization, suggesting gender classification to be less stable in learning compared to newer models. Compared to YOLOv8 and YOLOv9, YOLOv7 is more volatile in loss values, suggesting less stable convergence. These trends highlight that while YOLOv7 achieves reasonable performance, YOLOv9 remains the most stable and accurate model for gender classification.

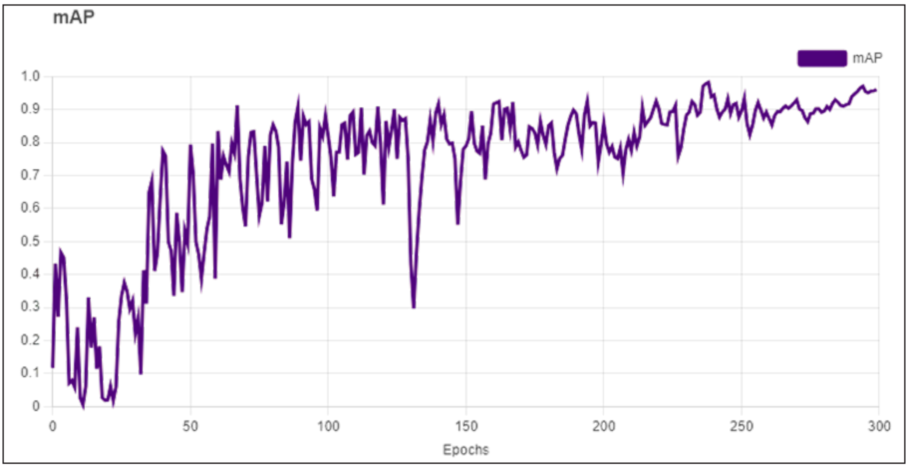
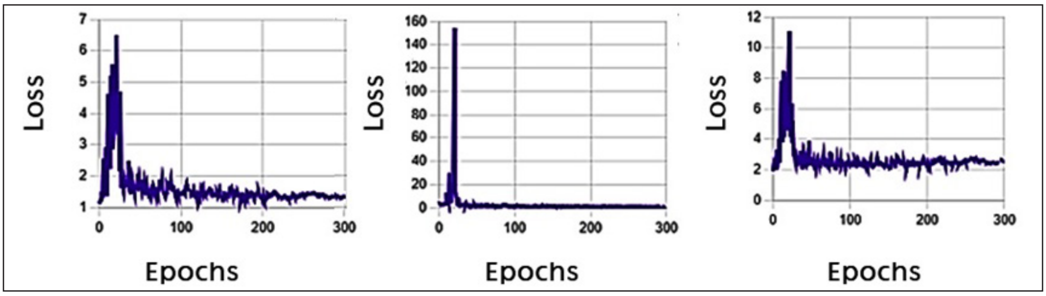
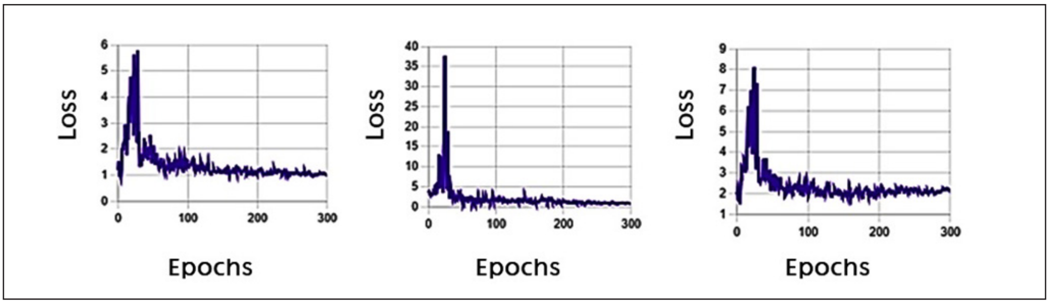


Figure 8. Mean average precision (mAP) accuracy of the YOLOv9 model for gender detection



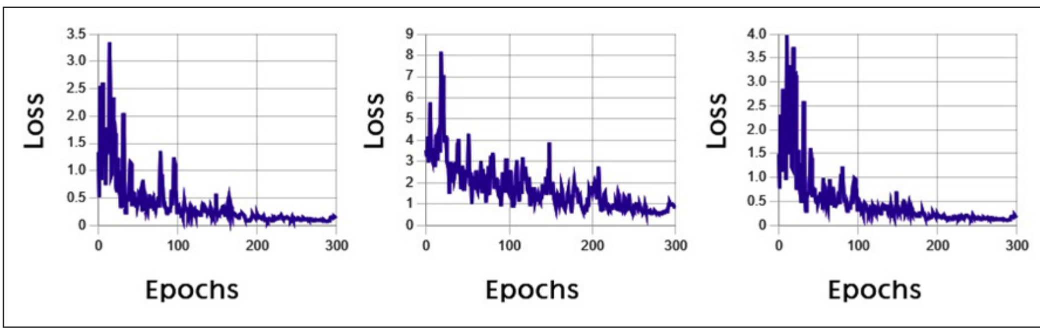
(a) (b) (c)

Figure 9. Loss curve of the Yolov9 model training: (a) Box Loss; (b) Class Loss; (c) Object Loss



(a) (b) (c)

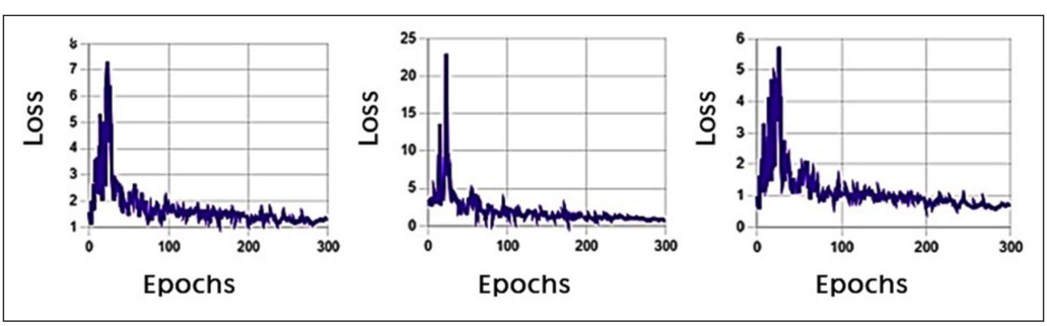
Figure 10. Loss curve of the Yolov8 model training: (a) Box Loss; (b) Class Loss; (c) Object Loss



(a) (b) (c)

Figure 11. Loss curve of the YOLOv7 model training: (a) Box Loss; (b) Class Loss; (c) Object Loss

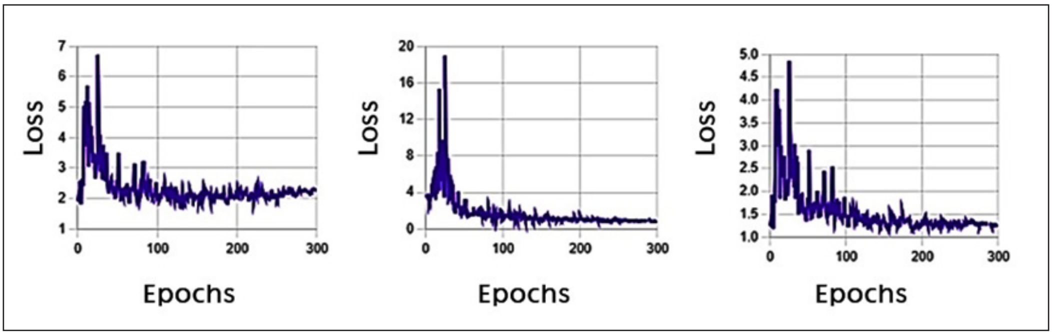
The loss curves of YOLOv5 in Figure 12 display a steep initial decline in Object Loss and Box Loss, subsequently levelling off, suggesting improvement in localization. The Class Loss oscillates to stabilize, suggesting learning progressively. However, compared to YOLOv9, YOLOv5 displays higher variance and a slower convergence rate, justifying its lower stability and accuracy in gender classification.



(a) (b) (c)

Figure 12. Loss curve of the Yolov5 model training: (a) Object Loss; (b) Class Loss; (c) Box Loss

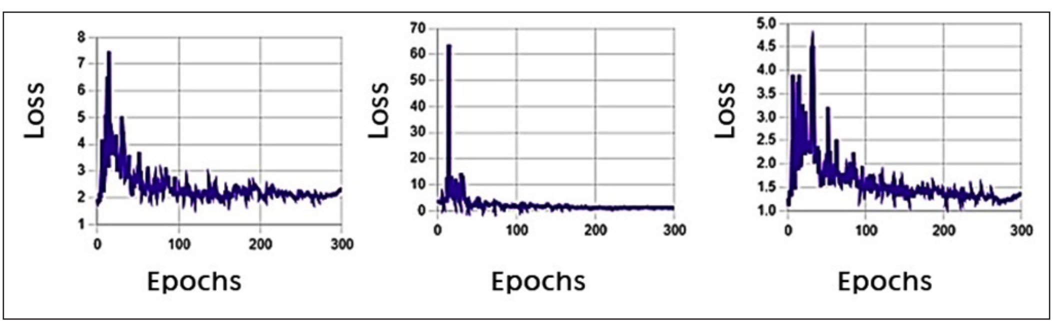
The loss curves of YOLOv4 in Figure 13 display a steep initial decline in Object Loss and Box Loss, after which they stabilize over time. Class Loss steeply drops but oscillates before it eventually converges, indicating a moderate learning efficiency. YOLOv4 is more variant and slower to converge than YOLOv9, validating that it is less accurate and more unstable in gender classification.



(a) (b) (c)

Figure 13. Loss curve of the Yolov4 model training: (a) Object Loss; (b) Class Loss; (c) Box Loss

The loss curves of YOLOv3 Figure 14 indicate a steep initial decline coupled with apparent oscillations, more evidently in Class Loss, indicating unstable learning. Object Loss and Box Loss take a slower time to converge compared to newer models of YOLO. The patterns confirm that YOLOv3 is of lower precision and stability, hence less accurate in gender classification.



(a) (b) (c)

Figure 14. Loss curve of the Yolov3 model training: (a) Object Loss; (b) Class Loss; (c) Box Loss

Comparison of Models

The results demonstrate a progressive improvement in the performance of YOLO models, from YOLOv3 to YOLOv9, with newer versions exhibiting enhanced accuracy, inference speed, and training efficiency. YOLOv9 achieved the highest mAP (97%) with the fastest inference time (0.332 seconds per image, FPS = 3.00), making it the most efficient for real-time applications. YOLOv8 (mAP = 96.2%) followed closely, offering the best balance between precision and recall (F1-score = 91.82%), ensuring reliable classification with minimal errors. YOLOv7 (mAP = 94.4%) performed well but was slightly less efficient than the more recent versions. In terms of inference speed, the latest versions demonstrated superior performance, with YOLOv9 being the fastest (0.332 seconds per image, FPS = 3.00), followed by YOLOv8 (0.379 seconds, FPS = 2.63). In contrast, older models such as YOLOv3 and YOLOv4 exhibited significantly slower inference times (0.759 and 0.648 seconds, respectively), making them less suitable for real-time applications requiring immediate processing. Regarding training time, YOLOv9 required the shortest duration (1:45:00 hours) compared to YOLOv3 (4:00:00 hours), reflecting advancements in algorithm optimization and architectural efficiency as shown in Table 1. These findings confirm that YOLOv9 is the most effective and reliable model for gender classification, followed by YOLOv8 as the optimal balance between accuracy and performance, while older versions demonstrate lower efficiency, making them less suitable for modern, high-performance applications, as shown in Figure 15.

The analysis of error rates across YOLO models confirms YOLOv9 as a highly reliable model, achieving a strong balance between precision and recall. While YOLOv8 recorded the lowest total error count (FP + FN = 29), YOLOv9 (FP = 23, FN = 25, total errors = 48) demonstrated superior stability and consistency, making it highly effective in maintaining classification accuracy. Additionally, YOLOv9's structure is highly optimized to limit false positives compared to older models, resulting in fewer incorrect detections. In contrast, YOLOv3 was associated with the highest error rate (FP + FN = 59), a reflection of inferior reliability. The results point to YOLOv9 to be a highly efficient model of high detection strength that is a competitive candidate for real-time classification, as shown in Figure 16 and Table 2.

Table 1
Performance comparison among different YOLO models

Target detection networks	**mAP (%)**	**Precision (%)**	**Recall (%)**	**F1 score (%)**	**Inference time per image**	**Frames per second**	**Training time**
YOLO v3	83.0	84.3	82.4	83.32	0.759	1.31	4:00:00 sec
YOLO v4	89.6	85.1	86.6	85.84	0.648	1.54	3:25:00 sec
YOLOv5	84.1	87.7	83.4	85.49	0.604	1.65	3:11:00 sec
YOLOv7	94.4	87.9	85.3	86.58	0.427	2.34	2:15:00 sec
YOLOv8	96.2	88.7	95.2	91.82	0.379	2.63	2:00:00 sec
YOLOv9	97	86.8	86.1	86.54	0.332	3.00	1:45:00 sec

Note. YOLO = You Only Look Once; mAP = Mean average precision

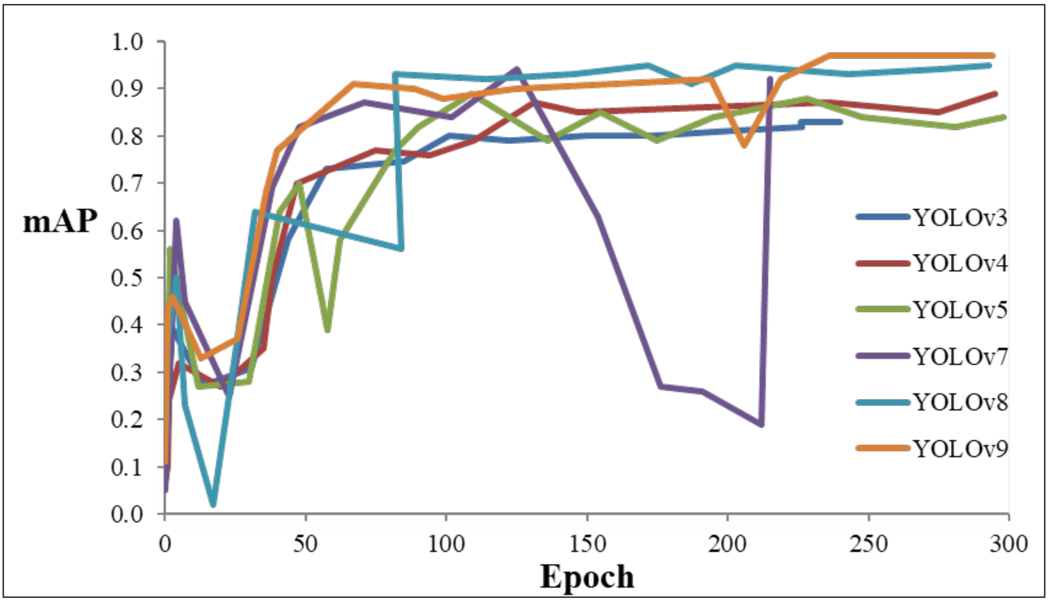


Figure 15. Mean average precision (mAP) accuracy of You Only Look Once (YOLO) models for gender detection

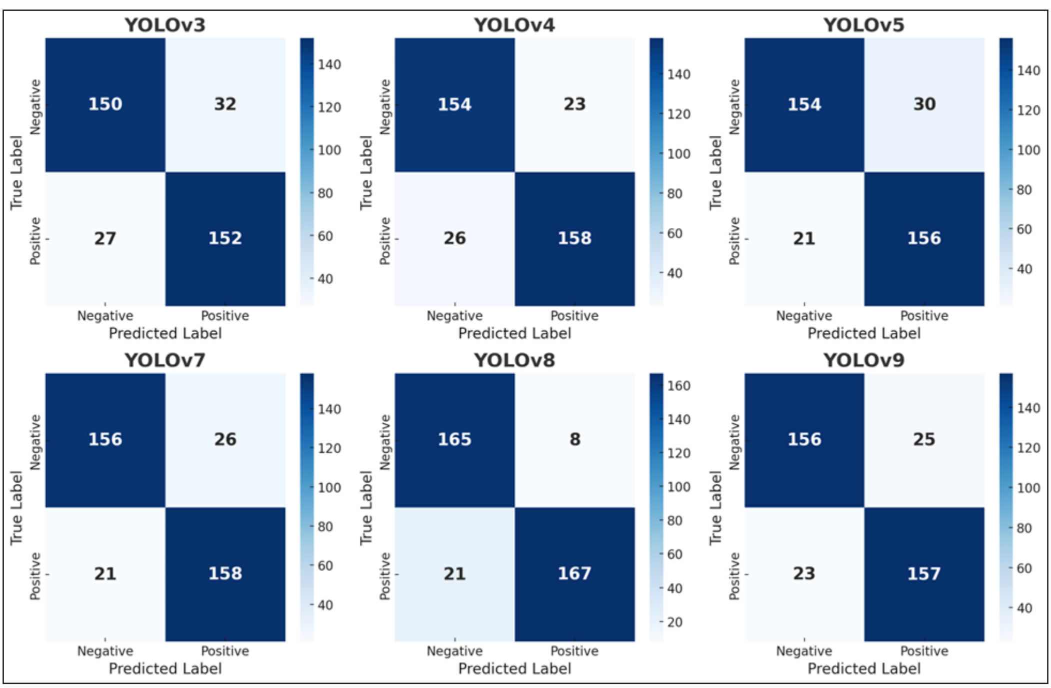


Figure 16. Confusion matrix analysis for the You Only Look Once (YOLO) models

Table 2
Confusion matrix analysis for the You Only Look Once (YOLO) models

Model	True positives (TP)	False negatives (FN)	False positives (FP)	True negatives (TN)
YOLO v3	152	32	27	150
YOLO v4	158	26	23	154
YOLOv5	156	30	21	154
YOLOv7	158	26	21	156
YOLOv8	167	8	21	165
YOLOv9	157	25	23	156

Used PyTorch Grad-CAM to visualize the most responsible areas in the YOLOv9 decision-making process in classifying gender. The results revealed that the model is primarily observing eyes and mouth, both of which are biological features used in classifying between genders. This proves that YOLOv9 relies on patterns learned that are in agreement with human vision, contributing to its classification accuracy.

In addition to that, Canny Edge Detection with OpenCV has been employed in analyzing structural features that are recognized by the model as critical. The outcome indicates that YOLOv9 is heavily dependent on the sharp edges of faces, particularly the jawline edge, eyebrow curve edge, and nose and mouth edges. This is a pointer that the model classifies based on patterns in face architecture and is quite sensitive to image quality changes, lighting conditions, and complexity of the environment.

The results revealed that Grad-CAM approximately 31.48% of the image consists of very active regions, which confirms that the model is focusing on significant face features such as eyes and mouth. Canny Edge Detection analysis, on the other hand, revealed that 5.70% of the image consists of sharp edges around the jawline, eyebrows, and mouth, which confirms that in its classification process, YOLOv9 is largely reliant on structural face features.

Although YOLOv9 is good at recognizing large-scale face features and effectively analyzing visual patterns for classifying gender. This reliance implies that the model's classification performance would be susceptible to low lighting conditions, noise, or distractions in the scene. The results, as observed in Figure 17, present a full picture of YOLOv9's feature selection mechanism and confirm that its decision process is both edge-dependent and biologically meaningful.

A one-way analysis of variance (ANOVA) was conducted using Python's SciPy library to determine if different YOLO models (YOLOv3 to YOLOv9) varied in their performance statistically. Five trial results for each model were utilized to provide consistency in results in terms of mean average precision (mAP) values. The analysis also confirmed statistically significant differences ($F = 8993.63$, $P < 0.05$), implying that observed gains in

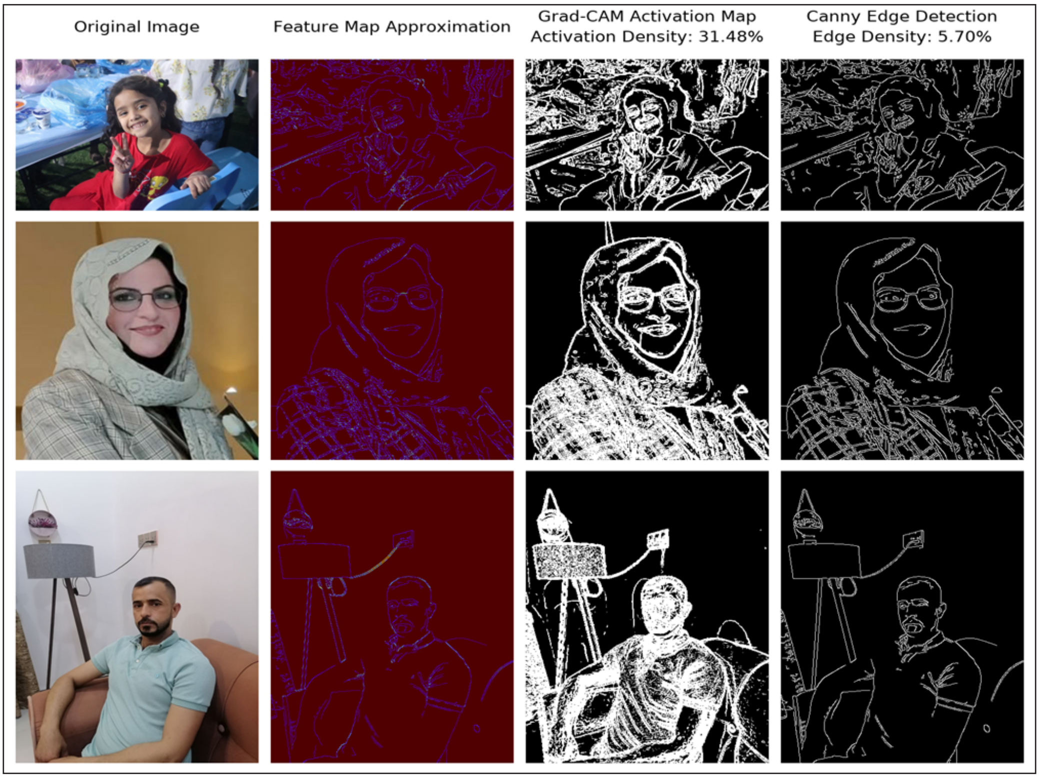


Figure 17. Feature visualization and model analysis for the YOLOv9 in gender classification

YOLOv9 and YOLOv8 are not a function of random variance but a genuine improvement in performance. The results affirm that architectural advances in newer versions of YOLO have been a primary force in detection precision, as shown in Table 3.

Table 3
Mean average precision (mAP) values across five trials for YOLO models

YOLOv9 (mAP %)	YOLOv8 (mAP %)	YOLOv7 (mAP %)	YOLOv5 (mAP %)	YOLO v4 (mAP %)	YOLO v3 (mAP %)	Epoch
97.3	96.4	93.9	84.4	89.2	83.0	1
96.9	96.1	94.2	84.1	89.8	82.8	2
97.1	95.8	94.1	83.8	89.1	83.1	3
97.2	96.3	94.5	84.3	88.9	82.9	4
97	96.2	94.4	84.1	89.6	83.0	5

Practical Applications and Future Directions

YOLOv9's real-time performance (FPS = 3.00) is ideal for applications that need quick decision-making, e.g., automated surveillance, biometric identification, and AI-based customer engagement. Deployment is still a challenge, though, mainly in terms of computational requirements and privacy-sensitive settings. Future research must explore optimization techniques, i.e., quantization and model pruning, to enable it to be used in low-power embedded systems.

While YOLOv9 is highly performing, pairing it with a more accurate approach would also increase classification precision and credibility. Further enhancement would be introduced by expanding examinations using real-world datasets to enable flexibility in different demographic and environmental settings.

CONCLUSION

This study showed a progressive improvement of the YOLO models (YOLOv3–YOLOv9) in gender classification, of which YOLOv9 was found to be most accurate (mAP = 97%) and most precise (86.8%) in nature, making it most efficient for use in real-time. Despite its advancement, YOLOv9 is also afflicted with high computational demands and misclassification in complex situations. The efforts in the coming times must be focused on optimization strategies and fusion of higher-accurate methods to enhance robustness and flexibility in a diverse set of real-world applications.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi Malaysia for their academic support throughout the completion of this research.

REFERENCES

Acharya, A. (2024a). *YOLOv9: SOTA object detection model explain*ed. Encord. https://encord.com/blog/yolov9-sota-machine-learning-object-dection-model/

Acharya, A. (2024b). *Comparative analysis of YOLOv9 and YOLOv8 using custom dataset on encord active*. Encord. https://encord.com/blog/performanceyolov9-vs-yolov8-custom-dataset/

Al-Bayati, R. E., & Abood, Z. M. (2023). Comparison between gait recognizing algorithms. *Mustansiriyah Journal of Pure and Applied Sciences*, *1*(1), 133–139.

Alsaif, O. I., Hasan, S. Q., & Maray, A. H. (2023). Using a structural model for human gait gender recognition. *IAES International Journal of Artificial Intelligence*, *12*(2), 974–983. https://doi.org/10.11591/ijai.v12.i2.pp974-983

Awan, M. J., Raza, A., Yasin, A., Shehzad, H. M. F., & Butt, I. (2021). The customized convolutional neural network of face emotion expression classification. *Annals of the Romanian Society for Cell Biology*, *25*(6), 5296–5304.

Ayo, F. E., Mustapha, A. M., Braimah, J. A., & Aina, D. A. (2022). Geometric analysis and YOLO algorithm for automatic face detection system in a security setting. In *Journal of Physics: Conference Series* (Vol. 2199, No. 1, p. 012010). IOP Publishing. https://doi.org/10.1088/1742-6596/2199/1/012010

Azhar, M., Ullah, S., Ullah, K., Rahman, K. U., Khan, A., Eldin, S. M., & Ghamry, N. A. (2022). Real-time dynamic and multi-view gait-based gender classification using lower-body joints. *Electronics*, *12*(1), 118. https://doi.org/10.3390/electronics12010118

Boesch, G. (2024). *YOLOv9: Advancements in real-time object detection*. Viso.ai. https://viso.ai/computer-vision/yolov9/

Dutta, A., Atik, A., Bhadra, M., Pal, A., Khan, M.A., & Chakraborty, R. (2022). Detection of objects using a fast R-CNN-based approach. In B. Das, R. Patgiri, S. Bandyopadhyay, & V. E. Balas (Eds.), *Modeling, simulation and optimization: Proccedings of CoMSO 2021*. Springer. https://doi.org/10.1007/978-981-19-0836-1_21

Eşer, E. E., & Hardalaç, F. (2021, December 9-11). *Performance of YOLO on a custom dataset of infrared laser images* [Paper presentation]. 1st International Conference on Informatics and Computer Science, Ankara, Turkey. https://webupload.gazi.edu.tr/upload/1104/2022/6/11/1f03cadc-5320-4546-86fb-c36b37c86b77-guncel_2021-1st-international-conference-on-informatics-and-computer-science_isbn.pdf

Gan, N., Wan, F., Lei, G., Xu, L., Xu, C., Xiong, Y., & Zhou, W. (2023). YOLO-CID: Improved YOLOV7 for X-ray contraband image detection. *Electronics*, *12*(17), 3636. https://doi.org/10.3390/electronics12173636

Jabraelzadeh, P., Charmin, A., & Ebadpore, M. (2020). Providing a hybrid method for face detection, gender recognition, facial landmarks localization and pose estimation using deep learning to improve accuracy. *Journal of Artificial Intelligence in Electrical Engineering*, *8*(32), 1–14.

Jabraelzadeh, P., Charmin, A., & Ebadpour, M. (2023). Providing a hybrid method for face detection and gender recognition by a transfer learning and fine-tuning approach in deep convolutional neural networks and the Yolo algorithm. *International Journal of Nonlinear Analysis and Applications*, *14*(1), 2373–2381. https://doi.org/10.22075/ijnaa.2022.26099.3661

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A review of YOLO algorithm developments. *Procedia Computer Science*, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Mittal, S., Thakral, K., Majumdar, P., Vatsa, M., & Singh, R. (2023). Are face detection models biased? In *IEEE 17th International Conference on Automatic Face and Gesture Recognition* (pp. 1–7). IEEE. https://doi.org/10.1109/FG57933.2023.10042564

Nada, A. M. A., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using CNN. *International Journal of Academic Engineering Research*, *4*(8), 21-24.

Nepal, U., & Eslamiat, H. (2022). Comparing YOLOV3, YOLOV4 and YOLOV5 for autonomous landing spot detection in faulty UAVs. *Sensors*, *22*(2), 464. https://doi.org/10.3390/s22020464

Raman, V., ELKarazle, K., & Then, P. H. (2023). Artificially generated facial images for gender classification using deep learning. *Computer System Science and Engineering*, *44*(2), 1341-1355. https://doi.org/10.32604/csse.2023.026674

Ramya, R., Anandh, A., Muthulakshmi, K., & Venkatesh, S. (2022). Gender recognition from facial images using multichannel deep learning framework. In P. P. Sarangi, M. Panda, S. Mishra & B. S. P Mishra (Eds.), *Machine learning for biometrics: Concepts, algorithms and learning* (pp. 105–128). Academic Press. https://doi.org/10.1016/B978-0-323-85209-8.00009-2

Shen, T. W., Wang, D., Cheung, K. W. K., Chan, M. C., Chiu, K. H., & Li, Y. K. (2021). A real-time single-shot multi-face detection, landmark localization, and gender classification. In *Proceedings of the 2021 3rd International Conference on Image Processing and Machine Vision* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3469951.3469952

Shet, A. V., Chinmay, B. S., Shetty, A. A., Shankar, T., Hemavathy, R., & Ramakanth, P. (2022). Face detection and recognition in near infra-red image. In *6th International Conference on Computation System and Information Technology for Sustainable Solutions* (pp. 1–6). IEEE. https://doi.org/10.1109/CSITSS57437.2022.10026378

Singh, A., Rai, N., Sharma, P., Nagrath, P., & Jain, R. (2021). Age, gender prediction and emotion recognition using convolutional neural network. In *Proceedings of the International Conference on Innovative Computing and Communication* (pp. 1–11). https://doi.org/10.2139/ssrn.3833759

Sonthi, V. K., Sikka, R., Anusha, R., Devi, P. B. S., Kamble, S. K., & Pant, B. (2023). A deep learning technique for smart gender classification system. In *3rd International Conference on Advance Computing and Innovative Technologies in Engineering* (pp. 983–987). IEEE. https://doi.org/10.1109/ICACITE57410.2023.10182683

Srikar, M., & Malathi, K. (2022). An improved moving object detection in a wide area environment using image classification and recognition by comparing You Only Look Once (YOLO) algorithm over Deformable Part Models (DPM) algorithm. *Journal of Pharmaceutical Negative Results*, *13*(Special Issue 4), 1701–1707. https://doi.org/10.47750/pnr.2022.13.S04.204

Srinivasan, L. (2023). Automatic gait gender classification using convolutional neural networks. In *Proceedings of the 2023 Fifth International Conference on Image Processing and Machine Vision* (pp. 34–40). Association for Computing Machinery. https://doi.org/10.1145/3582177.3582184

Sumi, T. A., Hossain, M. S., Islam, R. U., & Andersson, K. (2021). Human gender detection from facial images using convolution neural network. In M. Mahmud, M. S. Kaizer, N. Kasabov, K. Iftekharudin, & N. Zhing (Eds.), *Applied Intelligence and Informatics: First International Conference* (Vol. 1435, pp. 188–203). Springer. https://doi.org/10.1007/978-3-030-82269-9_15

Tejaswi, K., Mokshith, D., Kumar, C. M., & Kumar, M. (2023). Facial expression recognition using YOLO. In *International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering* (pp. 1–8). IEEE. https://doi.org/10.1109/RMKMATE59243.2023.10369028

Thakuria, A., & Erkinbaev, C. (2023). Improving the network architecture of YOLOV7 to achieve real-time grading of canola based on kernel health. *Smart Agricultural Technology*, *5*, 100300. https://doi.org/10.1016/j.atech.2023.100300

Tilki, S., Dogru, H. B., Hameed, A. A., Jamil, A., Rasheed, J., & Alimovski, E. (2021). Gender classification using deep learning techniques. *Manchester Journal of Artificial Intelligence and Applied Sciences*, *2*(1), 126–131.

Ullah, M. B. (2020). CPU based YOLO: A real time object detection algorithm. In *IEEE Region 10 Symposium* (pp. 552–555). IEEE. https://doi.org/10.1109/TENSYMP50017.2020.9230778

Varnima, E. K., & Ramachandran, C. (2020). Real-time gender identification from face images using You Only Look Once (YOLO). In *4th International Conference on Trends in Electronics and Informatics* (pp. 1074–1077). IEEE. https://doi.org/10.1109/ICOEI48184.2020.9142989

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. In A. Leonardis, E. Ricci, S. Roth, T. Sattler, & G. Varol (Eds.), Computer *vision – ECCV 2024: Lecture notes in computer science* (Vol. 15089, pp. 1-21). Springer. https://doi.org/10.1007/978-3-031-72751-1_1

Yaseen, M. (2024). *What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector*. arXiv. https://doi.org/10.48550/arXiv.2409.07813

Wang, F., Jiang, J., Chen, Y., Sun, Z., Tang, Y., Lai, Q., & Zhu, H. (2023). Rapid detection of Yunnan Xiaomila based on lightweight YOLOV7 algorithm. *Frontiers in Plant Science*, *14*, 1200144. https://doi.org/10.3389/fpls.2023.1200144

Wang, W., Chen, J., Huang, Z., Yuan, H., Li, P., Jiang, X., Wang, X., Zhong, C., & Lin, Q. (2023). Improved YOLOV7-based algorithm for detecting foreign objects on the roof of a subway vehicle. *Sensors*, *23*(23), 9440. https://doi.org/10.3390/s23239440

Xiao, M., Xiao, N., Zeng, M., & Yuan, Q. (2020). Optimized convolutional neural network-based object recognition for humanoid robot. *Journal of Robotics and Automation*, *4*(1), 122–130. https://doi.org/10.36959/673/363

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12993–13000). Association for the Advancement of Artificial Intelligence Publishing. https://doi.org/10.1609/aaai.v34i07.6999