Robust Detection of Partially Occluded Faces in Low-Light Scenarios Using YOLOv7 and YOLOv6

Nayef Alqahtani[1], Amina Shaikh[2]*, Imran Khan Keerio[3]
Department of Electrical Engineering-College of Engineering, King Faisal University, Al-Ahsa, 31982, Saudi Arabia[1]
Kulliyah of Information & Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia[2]
Dept. of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan[3]

***Abstract*—Partial occlusion and low light are significant challenges for face detection, limiting its effectiveness in critical applications such as security, surveillance, and user identification within computer vision. This study evaluates the effectiveness of two influential deep-learning models, YOLOv6 and YOLOv7, in identifying partially occluded faces in uncontrollable, real-world conditions. Training both models and assessing them with the help of comprehensive data-augmentation schemes that facilitate the occurrence of generalization, a carefully selected sample of partially blocked and hidden images of faces was used in all the experiments performed under low-light exposure. Findings indicate that YOLOv7 systematically outsmarts YOLOv6 in all key measures of performance, including precision (0.92 vs. 0.90), recall (0.89 vs. 0.79), as well as the mean Average Precision (mAP), which proves its ability to recognize hidden faces under adverse environments better. YOLOv7 takes a longer time to be trained, but with its enhanced design, especially the Extended Efficient Layer Aggregation Network (E-ELAN), feature extraction and real-time detection become much smoother. The statistics of this research indicate a visible increase, which indicates that YOLOv7 is reasonably suitable to be implemented in real-life, where a strong ability of face recognition, even with occlusions and low visibility, is required. This study contributes to the advancement of face detection technologies, tackling developing privacy and security needs in increasingly masked and low-visibility spaces.**

***Keywords*—Deep learning; partial occlusion; YOLOv6; YOLOv7; real-time detection; low-light face recognition**

I. Introduction

Face detection plays a very significant role in computer vision, particularly in the fields of security, user id etc. However, it is still quite hard to bring the system to the test with partially covered faces on it. The majority of face-recognition technology is activated on clear and obstructed faces making the technology to focus on key features such as mouth, nose and eyes. In most situations, people often wear face masks or put some sort of cover on their faces, thus partially covering the facial area. Human face identification is a very important feature in the computer vision field. This is the foundation of a wide variety of applications that can be practiced in many different areas such as security and identity verification, among many others [1]. Still, the process of detection of faces, which are partially masked, remains a significant problem. According to the traditional face recognition algorithms, they are more focused on non-occluded faces and salient facial feature, the mouth, nose and ocular region. Nevertheless, there are many situations and events which force people to put on masks or keep something covering or even hiding their faces, partly [5].

*Corresponding author.

When several faces in an image are near each other and seem to create a cluster, this effect is called occlusion and can result in misrecognition of the hidden face [2]. The initial versions of Convolutional Neural Networks (CNNs) and other deep learning methods have greatly enhanced detection accuracy. Yet, occlusion is a significant problem in face detection, since faces may be partially covered by different objects (e.g., masks, spectacles, or hands), environmental conditions (e.g., shadows, illumination), or other faces in populated scenes. In recent decades, there have been impressive advances in deep learning technologies in theory and application.

With the beginning of Masked FR as a new domain within the field of computer vision, the majority of FR systems have shifted towards implementing deep learning models [6, 7]. Before the COVID-19 pandemic, research was in progress to determine how deep learning could enhance the performance of available recognition systems for the presence of masks or occlusions. For example, different deep learning techniques have been proposed to solve the occluded facial recognition problem and have received tremendous attention. One of the well-known face detection algorithms is "You Only Look Once" (YOLO) [3], which is one of the modern alternatives [4]. Recently, the YOLO model has been further developed to detect faces, thus allowing it to decode and recognize the concealed conditions of people found in the images and video streams. This development requires a significant amount of data of occlusion to be trained.

Facial recognition systems based on YOLO are fast and more accurate, so they can be used in the security and market research sectors as well as human-computer interaction [8]. Face identification systems based on the YOLO architecture are susceptible to variations due to various factors, including the sizes of the model of the architecture, quality of the training set, and the facial expressions of the subjects that are displayed in the images used. Face occlusion detection is associated with the identification of areas of the face that are either obscured partially or fully in photographic/video content. The issue at hand is complex and can be subject to numerous different factors that are all contributing to its complexity. Detecting facial occlusions is one of the most difficult problems in the sphere of facial recognition systems and computer vision. These issues of uncertainty of the occlusions pose several significant issues especially in practical context.

Facial obstructions range between simple partial obstructions such as the use of sun glasses or a hat, and more complex cases, where the objects or the close people outside the

person are hampering vision on the face. The ambiguous nature of the discernment of such occlusions is significantly increased by the heterogeneity of the environmental and illumination conditions. Therefore, the face-detection automation has become one of the key priorities in the sphere of computer vision and pattern recognition [9]. The previous versions of the facial recognition algorithms had limitations in simple cases, but advancement in the deep learning algorithms have brought significant changes to the way they handle various cases. R-CNN structure is a rather radical departure of the YOLO model, which is an efficient and effective one-shot convolutional neural network model already [10]. The difficulties associated with occlusions lie in a variety, where simple partial obstructions like the ones created by hats or sunglasses are depicted at one end, and the more complicated situations where the face is covered by objects or other people are seen.

The varied nature of the surrounding conditions and lighting situations also makes the accurate identification of facial occlusions more difficult. As such, computer vision and pattern recognition have become the main research frontiers of face detection at an automatic level. The face recognition algorithms were once limited to very basic settings, but with the introduction of deep learning methods, their scope has been greatly extended. The R-CNN structure was quite a breakthrough as compared to the YOLO, which is characterized by its novel one-shot convolutional neural network model and commercial popularity. Occlusion issues require addressing through strategies that improve algorithmic robustness, the availability of adaptive strategies to varying environmental conditions, intersubjective variability in morphology, and the adoption of mechanisms that prioritize privacy issues.

To ensure the maintenance of the vision of developing and implementing facial recognition systems to work in the real world, these issues should be addressed. Improving the level of accuracy of the algorithms, promoting the context of events, increasing the representation of facial features, and use of privacy-preserving methods are vital in achieving the successful ability of individuals' identification within inclusion parameters. These problems are critical and require a solution to enable the further development and use of a facial recognition system that has an acceptable level of accuracy and efficacy in the real-world environment.

The present study aims at confronting the difficulties with low-light and semi-occluded face detection in the form of using YOLOv6 and YOLOv7 deep-learning algorithms on special datasets. One of the main benefits of YOLO with respect to partially obscured face recognition is the fact that it provides real-time processing, which makes it especially useful when it comes to detecting emotions in real-time [11]. This study performs the comparison and evaluation of the performance of two state-of-the-art deep-learning models, YOLOv6 and YOLOv7, in detecting partially concealed faces in low-light conditions. To improve model generalization, the aggressive data-augmentation method is employed in the study to tackle the issues of occlusion and illumination variation. The significance of this work is that it involves a custom dataset comprising faces of diverse forms of occlusion (i.e., face masks, helmets, and other objects) in low-light conditions. The analysis also uses aggressive forms of data augmentation to increase the generalization of the model and increase the resilience of YOLOv6 and YOLOv7 models in such challenging cases. The gap in the literature presented with the evaluation of these models in this context is that this research can provide new insight into the use of YOLOv6 and YOLOv7 in the areas that face detection is most required, including security, surveillance, and human-computer interaction, thus filling the gap in the literature.

The success of the YOLO deep-learning model in the detection of hidden faces relies on the complexity of the network structure used, as well as the quality of training data.

The organization of this study is as follows: Section II presents a detailed literature review covering previous research on occluded face detection based on models of the YOLO framework, as well as outlines the main gaps in the literature on the topic. Section III details the YOLOv6 and YOLOv7 techniques for occluded face detection. Section IV outlines the methodology that is used, with a description of the dataset, the data-augmentation approaches employed, and the work setup employed to test YOLOv6 and YOLOv7. The results are described in Section V and include the comparative performance analysis of YOLOv6 and YOLOv7 in different conditions of occlusion and low light. Section VI presents the discussion. Section VII is the concluding part of the manuscript, providing an overview of the key findings, commenting on the implications of the research to the practical context. Lastly, Section VIII suggests recommendations for future research and limitations of the study.

II. Literature Review

The field of facial recognition has evolved significantly over the years, and the use of crude pattern-recognition systems has been replaced by advanced deep-learning frameworks that provide quite high levels of accuracy in a wide range of settings. The development of deep-learning algorithms has supported tremendous advancement in identifying covered facial features. Fundamental pattern-recognition techniques have been replaced with deep-learned recognition models that can experience high accuracy within a wide range of operating environments. Deep learning is a radical method of recognizing semi-automated features of the face. Convolutional Neural Networks (CNN) and Regions Based Convolutional Neural Network (R CNN) were major developments that led to the detection of faces. The You Only Look Once (YOLO) system is the latest deep learning technology that is able to detect faces or objects in just one processing step, hence making it the best in regard to real-time use.

A. Early Bird Approaches to Face Detection

The previous face detection methods were all based on feature-based algorithms. Among them, the Viola-Jones detector [12] demonstrated good frontal face detection performance, but suffered failure for occlusions and variety in positions. AdaBoost learning with Haar-like features trained face detector, which is developed by Viola and Jones, cannot work well when the faces are partially (or completely) obstructed.

B. Advancement with CNN

A complete system has been built to detect faces with severe occlusion [13]. A deep learning network leverages gradient information in combination with shape information to accomplish its function. This work contributes three basic advances: The first contribution proposes a face detection method based on an energy function prior to a CNN model, which can extract deep features hidden in faces. Finally, a novel sparseness-induced deep learning classification method is presented for face occlusion detection. The presented head detection model exhibits superior capability to accurately recognize the face in arbitrary head poses, accordingly with different degrees. The proposed algorithm for validating facial occlusion can achieve high accuracy as to whether the facial region is occluded.

Experimental results demonstrate that the proposed occlusion verification system can still achieve 97.25% accuracy with 10 frames per second. The head detection algorithm shows good recall at 98.89% even with various difficult facial occlusions. By using CNNs, face detection has made progress in learning complex features and patterns from data. The R-CNN series methods, such as Fast R-CNN [14] and Faster R-CNN [15], enhanced the accuracy of object detection with region proposal networks, which can speed up the object detection. The approaches were very crucial for the evolution of face detection; however, they have high computational complexity and thus cannot be directly used. The deep learning era of visual recognition emerged with AlexNet [16]; following this first entry, VGGNet [17] and other advanced architectures further refined the convolutional layers for fine-grained pattern recognition. The models served as fundamental components for developing improved feature extraction methods, which transformed face detection practices.

Research demonstrated the effort of detecting mask-covered faces through a deep learning system, which successfully enables facial recognition for users with masks. The researchers successfully developed a model that used ResNet-50 architecture to detect masked faces in the COMASK 20 and LFW datasets. The research findings present an opportunity to integrate into current facial recognition systems that focus on identifying masked faces for security applications [19]. The research applied the Gabor wavelet together with deep transfer learning through a machine learning approach. The real-time masked face recognition process achieves better reliability through our integration of deep-learning CNN features together with Gabor wavelet extracted from unmasked facial regions to create an enhanced feature vector. The experiment, which tested the proposed method, used four benchmark datasets and achieved a mean accuracy of 97% [20].

C. YOLO Background

The group of experts in object recognition and detection that created YOLO represents an existing algorithm. YOLO (You Only Look Once) was initially presented by Redmon and its co-authors [18]. This work is recognized for its methods in real-time object detection. YOLO operates through regression-based detection to generate bounding boxes along with class probabilities as a single output. YOLO delivers superior detection speed to meet the requirements of applications that demand rapid responses. YOLO introduced real-time detection capabilities as a groundbreaking development, which enabled its practical use in various sectors [40, 41].

YOLOv3 and YOLOv4 incorporated Darknet-53 backbone together with spatial pyramid pooling (SPP) to enhance their feature extraction processes [21]. A new AI method appeared through a study that combined SVM classifiers with YOLOv3 and CNNs for feature analysis and classification [42]. The model trained on five different mask datasets showed better results than existing mask recognition methods by at least one point during testing. The algorithm achieves 99.4% accuracy in recognizing faces when they have partial coverage [22]. The better YOLOv3 network achieves enhanced mean Average Precision and faster frame rate than earlier approaches. Several face detection algorithms that focus on partial face obstruction face difficulties when detecting faces from complicated angles and positions, and when faces are at rest. Environmental conditions, along with lighting changes, create additional difficulties for detecting partially obstructed faces in image data [43, 44].

The improvements in YOLOv4 consisted of adding the Mish activation function and sophisticated data augmentation approaches, including Mosaic and the Cross Stage Partial (CSP) network [23]. These improvements enhance model durability and accuracy while the model continues to operate at real-time speeds. YOLOv4 and its predecessor YOLOv3 share the same limitation regarding the detection of faces that are either partially obstructed or fully concealed [24]. The requirement for designing improved solutions and creating dedicated training datasets to handle occlusion problems remains both essential and time-sensitive.

The authors in [25] experienced four different variants of YOLOv5 models for identifying faces covered with helmets. Models were created by adjusting the size of the BottleneckCSP module, which is located in the neck of the model. All four YOLOv5 models had similar mAP values, but the YOLOv5s model's outstanding 110 fps makes it stand out. Also, a higher mAP value between 0.9 and 1.3 was achieved by YOLOv5 models with pre-training weights, as shown in [25]. An enhanced YOLOv5 model for recognizing small targets was demonstrated in a different investigation of the model by [23].

The recent developments of face detection have explored the pursuit of multi-task learning to improve the precision of the model, especially in extreme conditions of partial blockage and low-light environments [45]. The study by [46] introduces the hybrid model that combines face identification and emotion recognition and states that they have boosted the face detection robustness in occlusivity cases. This model is viewed as having a joint feature extraction module, which learns both tasks simultaneously, thus reducing the partial occlusion through interpretations of subtle facial expressions. This hybrid approach has potential to enhance YOLO-based architecture, e.g. YOLOv6 and YOLOv7, by enhancing feature representations of occluded faces with additional contextual information based on emotions.

D. Research Gap from the Literature

Although a significant improvement has been made in real-time object detection in YOLOv3 and YOLOv4, there is an existing limitation on the performance of both models when used on tasks that require the detection of occluding faces. Empirical studies show that, despite having better inference speeds compared to traditional convolutional neural network-based methods, these models do not work well in detecting occlusions caused by accessories or other obstructions [22, 23]. Need for a specialized training together with the feature enhancements for partial occlusions motivated the evolution of newer versions such as YOLOv6 and YOLOv7 that are aimed at tackling these issues with better and more sophisticated feature extraction and processing capabilities. This study fills this gap by systematically comparing YOLOv6 and YOLOv7 through a specialized dataset, containing facial images partially covered by masks, helmets, and other objects, under varying lighting circumstances. Besides, it is one of the initial attempts to use energetic data-augmentation measures to improve the generalization ability of both models, hence addressing the inherent problem of occlusion and illumination variation.

III. YOLOv6 AND YOLOv7 Techniques for Occluded Face Detection

The YOLOv6 model was introduced in June 2022, followed by YOLOv7 in July 2022. YOLOv6 & YOLOv7 are the activated progressions of the YOLO family, which are designed with novel ideas to resolve real-time face detection issues like occlusion in concrete.

YOLOv6: Proposed by Meituan in 2022 [26] [27], the YOLOv6 is an advanced anchor-free detection model that boosts both the efficiency and accuracy of the model. Various innovations include:

EfficientRep has been constructed on the principles of EfficientDet and therefore benefits from both proprioceptive and visual features.

The anchor-free method adopted in YOLOv6 adds to its adaptability in the recognition of non-standard shaped objects, making it especially suited to situations where there are occlusions, where facial elements cannot form standard bounding box shapes [28].

YOLOv7, as explained by [21], is an upgrade of YOLOv6 that implements a set of optimizations that improve the accuracy of detection while preserving fast running. The main optimizations include:

- The (E-ELAN) is proposed to enhance feature integration in multiple scales, subsequently increasing the ability of the model to detect occluded faces, regardless of changes in positioning and size.
- Re-parameterization is used to optimize modeling in order to enable YOLOv7 to achieve faster inference with reduced accuracy object detections, which is essential in real-time scenarios involving occluded faces.

The YOLOv6s approach scored a mAP of 43.1% on the COCO val2017 dataset, outperforming the YOLOv5s algorithm by a large margin, which stood at 37.4%. When compared with the earlier algorithms, YOLOv7 achieves better results in inference time and accuracy (r6.1 as in the above images) than PP-YOLOE, YOLOX, and Scaled-YOLOv4 [29].

IV. Research Methodology

Fig. 1 depicts the five steps of the proposed approach: data collection, preprocessing, YOLO model training, evaluation, and testing. Data gathering refers to the acquisition of images of partially masked faces from the Internet. Preprocessing includes image size normalization, data labeling and bounding boxes, and data augmentation. YOLO model training involves training data with the YOLOv6 and YOLOv7 models. The evaluation of the YOLO model entails assessing the YOLO model's training results and testing its capabilities.

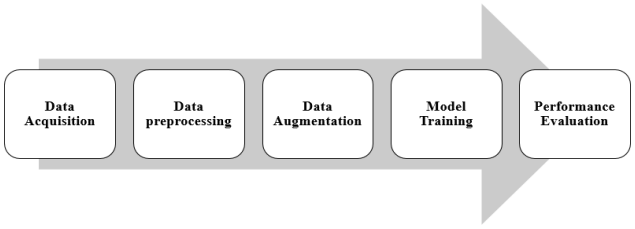

Fig. 1. Methodological design flow.

A. Dataset Collection and Preparation

Collecting a diverse and well-annotated dataset is crucial for training an effective face occlusion detection model. To create a robust dataset for face occlusion detection, it is essential to gather a diverse set of images that represent various real-world scenarios. Various sources were used in the compilation of the dataset used in this study, as it included 500 images with partially concealed faces in various illumination intensities, as shown in Fig. 2. It has three types of occlusions: facial masks, helmets, and miscellaneous objects, and is reflective of real-world situations. Stringent data augmentation protocols were put in place in order to create efficacy in the models and to facilitate generalizability.

Fig. 2. Sample images of different occlusions.

B. Data Pre-processing

Fig. 3 depicts a flowchart of the data preparation procedure. Before the process of training the model is carried out, this step is carried out to modify the data to the format of the YOLO classification system. The objective of the pre-processing of data is to enhance both the quality of the data and the performance of the YOLO models.

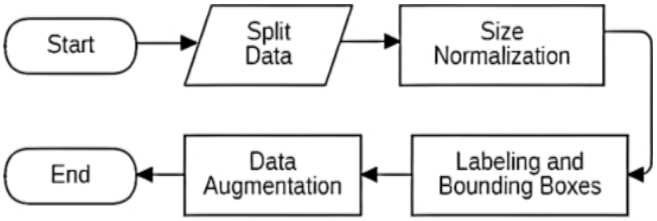

Fig. 3. Flow Chart for data pre-processing.

The first thing that needs to be done to begin the process of pre-processing data is to divide the data into 80% for training, 10% for testing, and 10% for validation. Preventing overfitting and ensuring that the model can generalize well on newly collected data are the two primary goals of this endeavor. After that, the image size was adjusted to 640×640 pixels to maintain consistency with the input format of the model. The data is then annotated, which includes labeling and bounding boxes, among other elements. While the process of drawing a box around the items in the image is referred to as a bounding box, the process of labeling involves the act of labeling the objects in the image. The final stage is to execute data augmentation, which is the process of making copies of the image with multiple modifications, such as brightness, contrast, and rotation. This is the final phase. To improve the model's ability to learn, this is done to enhance the quantity of the dataset as well as its diversity.

We have modified and labeled the partially occluded faces in an image and assigned the class name using the Roboflow online.

C. Data Augmentation

After annotating this dataset, a new version of the dataset is generated by applying preprocessing and augmentations, as shown in Fig. 4, to improve the quality and generalization ability of a model. Eventually, we could have 1100 images 80% as a training set, 10% as the validation set, and 10% as a testing set.

A flowchart depicting data augmentation can be viewed in Fig. 4. Before the YOLO models' training procedure is carried out, this step is carried out to multiply the data and prevent overfitting. This is done to ensure that the YOLOv6 and YOLOv7 models are as accurate as possible. To train both YOLO models to be more resistant to changes in the appearance of the data, the goal of data augmentation is to make it capable of producing more accurate object detection [33].

Fig. 4. Data augmentation sample image.

Data augmentation techniques are applied individually. Each version demonstrates a different type of transformation:

- Original - The unmodified image.
- Brightness - Enhanced brightness.
- Contrast - Enhanced contrast.
- Sharpness - Enhanced sharpness.
- Flip - Vertically flipped.
- Mirror - Horizontally mirrored.
- Rotate - Rotated at an angle.

D. Model Training

This phase involves training of the YOLO models, which will be used to build a strong infrastructure of recognizing occlusions on faces in pictures. The first step involves importing carefully pre-processed data and then training the two YOLO models over 150 epochs to optimize the detection accuracy. After that, the most effective variants of the trained models are stored as the name best.pt, making them available to accurately and optimally identify faces with occlusions during image processing.

E. Evaluation Using DL Approaches

In this work, Google Colaboratory has been chosen as the environment where different deep-learning models will be run. The Tesla T4 graphics processing unit (GPU) has an installed memory of 16 GB and is integrated perfectly via Google Collaboratory and utilized in every experiment that was carried out in this study. The platform offers 12GB of random-access memory (RAM), facilitating expedited deep-learning training. To generate the graph and confusion matrices in this study the pre-installed TensorFlow software library is used in Google Colaboratory. The custom-generated dataset consists of partially occluded face images, which are annotated with a bounding box and trained using the proposed two different DL models and tested to evaluate and compare the performances. To measure the performance of the DL, model the primary metric is mean Average Precision (mAP). To evaluate the precision and recall values of various training models this indicator is used in numerous other face occlusion detection studies [30]. The F1-score (F1), recall rate (Re), and precision rate (Pr) are all evaluation indicators that can be mathematically expressed as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision+recall}} \quad (3)$$

The values of True Positive (TP), False Positive (FP), and False Negative (FN) samples are the determinants of these indicators. TP is the scenario in which a positive sample is accurately predicted for an object. A false positive (FP) is the identification of an invalid object as a positive sample. Contrary to a FN, a legitimate object is identified as a negative sample. The expressions for calculating AP and mAP are as follows:

$$AP = \int_0^1 P(R)\, dR$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \quad (4)$$

V. Data Analysis and Results

The training duration of 150 epochs for YOLOv6 and YOLOv7 is illustrated in Table I. Compared to YOLOv7, it is obvious that YOLOv6 has the minimum training time. In this study, two experiments are conducted to evaluate the capabilities of two distinct deep-learning algorithms in the recognition of occluded faces.

TABLE I. Training Timing of Models

Model	Epochs	Training time (80%)
YOLOv6	150	44 min, 25 sec
YOLOv7	150	2 hr, 19 min, 32 sec

A. Performance Analysis Based on Key Metrics

Table II shows the precision, recall, F1-score, mAP50, and mAP50-90 values produced by the YOLOv6 and YOLOv7 models trained on 150 epochs. F1-score is the harmonic mean of precision and recall; mAP50 is the average precision of all classes at the IoU threshold of 50%; mAP50-90 is the average precision of all classes at the IoU threshold between 50% and 90%; precision is the fraction of detections that are true positives. These results represent the model's performance in identifying the partially occluded faces in an image. A greater value indicates higher model performance [31], [32]. Precision, recall, F1-score, and IoU are computed as Formula (1) to Formula (4).

TABLE II. Performance Analysis Based on Key Metrics

Model	Precision	Recall	Accuracy@ mAP(50)	Accuracy@ mAP(50-95)	F1-Score
YOLOv6	0.90	0.79	0.886	0.45	0.845
YOLOv7	0.92	0.89	0.922	0.50	0.90

YOLOv7 (0.92) has a slight edge over YOLOv6 (0.90), indicating improved accuracy in the accurate recognition of faces. Also, YOLOv7 (0.89) shows an appreciable improvement over YOLOv6 (0.787), emphasizing its ability to recognize a higher proportion of relevant objects. Importantly, YOLOv7 achieves the maximum value (0.922) in relation to YOLOv6 (0.886), indicating improved accuracy under less strict thresholds of IoU. Besides, YOLOv7 (0.50) has a narrow edge over YOLOv6 (0.45), indicating superior performance under more strict thresholds of IoU. Finally, YOLOv7 (0.90) has an edge over YOLOv6 (0.845), indicating an improved.

YOLOv7 outperforms YOLOv6 in all the tested metrics, making it a better model for object detection tasks. The differing colors and annotations in the graph clearly indicate these differences, as shown in Fig. 5.

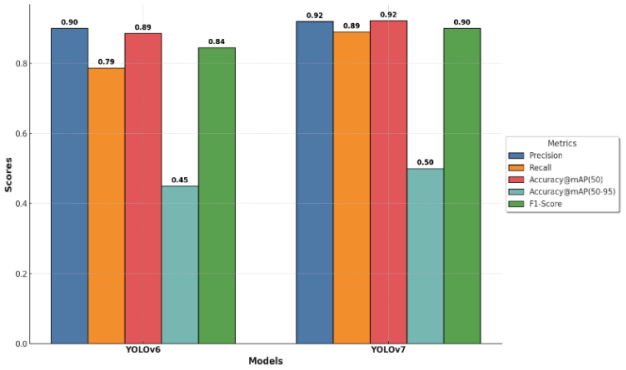

Fig. 5. Visualization of YOLOv6 and YOLOv7 based on performance metrics.

Fig. 5 shows the clear over performance of YOLOv7 over YOLOv6 in all the metrics tested. The Precision (0.92 vs. 0.90) and Recall (0.89 vs. 0.787) of YOLOv7 are higher, signifying that it has better accuracy and object detection strengths. Also, it has better performance in Accuracy@mAP(50) (0.922 vs. 0.886) as well as in Accuracy@mAP(50-95) (0.50 vs. 0.45), which shows that it performs better under fluctuating thresholds in the IoU metric. Also, the F1-score of YOLOv7 (0.90 vs. 0.845) is higher.

B. Confusion Matrix Analysis

The confusion matrix produced in tests is shown in Fig. 6. The confusion matrices illustrate the performance of YOLOv6 and YOLOv7 in face detection in partially occluded areas. YOLOv6 achieves 91% accuracy in face and 88% in background detection, with low misclassification rates of 9% for face and 12% for background. On the other hand, YOLOv7 shows better performance, with 94% accuracy in face and 90% in background detection, and low misclassification rates of 6% for face and 10% for background. This highlights the improved ability of YOLOv7 in handling challenging situations like partial occlusions.

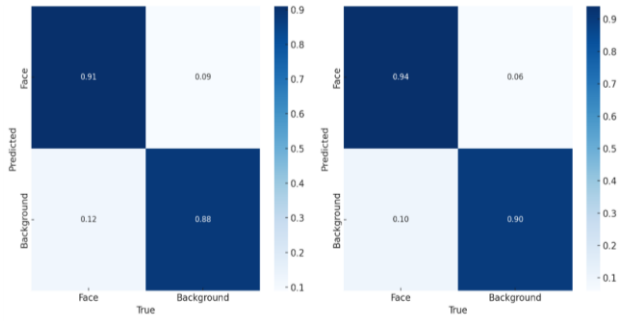

Fig. 6. Confusion matrix of models.

C. Training and Validation Performance of the Models

Training outputs of the YOLOv6 and the YOLOv7 models for 150 epochs exhibit superior metrics in the form of precision, recall, F1-score, mAP50, and mAP50-90. The complete visualization of the training output concerning the YOLO model, i.e., for 150 epochs, was depicted, indicating the performance of the box_loss, cls_loss, dfl_loss, precision, recall, mAP50, mAP50-90, and confusion matrix. The training output of the YOLOv6 model is presented in Fig. 7.

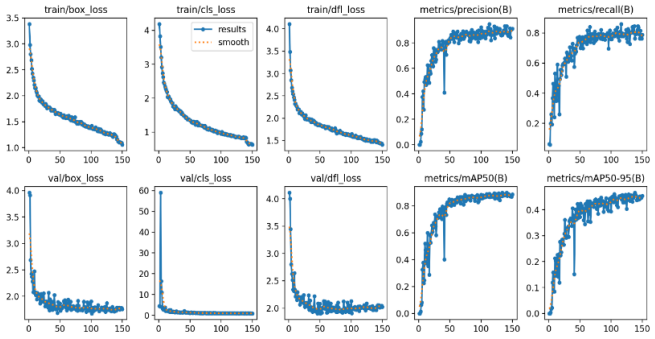

Fig. 7. YOLOv6-model training and validation.

The dynamics of the model were evaluated through systematic monitoring of the training and validation performance indicators over a sequence of 150 epochs to form a complete understanding of the learning process of the model and the generalization ability of the model. The constituent parts of training loss such as: box loss, classification loss, and distribution focal loss, show a monotonic reduction in the training curve hence validating the model as an efficient mean of bounding-box regression, object classification, and localization. The validation loss trends show a similar pattern of decrease, indicating that the model generalizes efficiently to novel inputs and prevents the risk of overfitting. The evaluation measures, consisting of precision, recall, and mean Average Precision (mAP) show stable improvement across the entire period of training. The model demonstrates significant gain of the detection accuracy, with increasing mAP value at IoU=50% and across the 50-95% IoU criteria. The experimental results demonstrate the effectiveness of the model, which paves the way for future use in object detection tasks across a wide range of real-world applications. This result of training is represented as Fig. 8 in the case of YOLOv7 model.

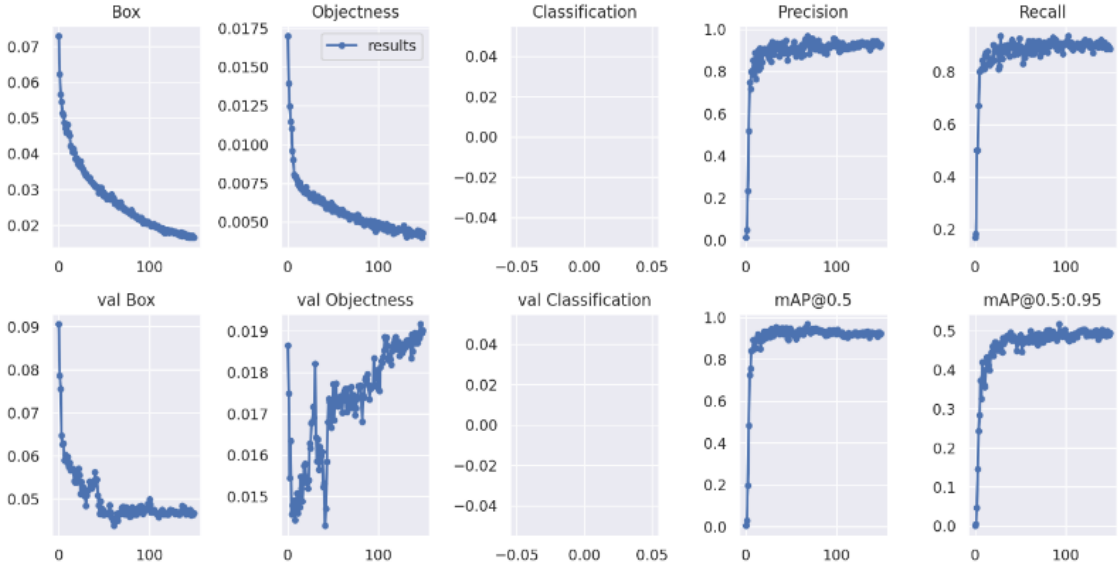

Fig. 8. YOLOv7-model training and validation.

The diagram shows the working effectiveness of the YOLOv7 algorithm in recognizing partially occluded faces under reduced-light settings both at the training and validation stages. The upper part shows training indicators, including Box Loss, Objectness Loss, and Classification Loss, as well as Precision and Recall trends over several epochs. The algorithm achieved improved convergence as loss indicators steadily reduced throughout the period of training. The plots of Precision and Recall show continuous improvements, indicating improved algorithm performance in the accuracy of object detection as well as efficiency in information recall.

The last row shows the validation metrics, including Box Loss, Objectness Loss, and Classification Loss, as well as the evaluation metrics mAP@0.5 and mAP@0.5:0.95. The decline in validation loss values shows good generalization ability of the model for new data, while the rise in mAP values indicates its ability to detect and classify difficult low-light faces. The experimental findings reveal that the YOLOv7 architecture is capable of dealing with high-level detection tasks in unfavorable environments, hence, its structural stability and reliability in facing real-world applications.

D. Statistical Significance of Performance Improvements

A paired t-test comparing performance differences between YOLOv6 and YOLOv7 was used to statistically evaluate the variation in performance in terms of the metrics used to measure accuracy in several experimental runs. The statistical analysis is mainly aimed at proving that the improvement in precision with the use of YOLOv7 compared to YOLOv6 is statistically significant, which will strengthen the validity of the given findings. The analysis states that its confidence interval of estimating the range of difference in the true mean differences of the precision is found to be 95%.

1) Calculation of mean difference: The mean difference $\overline{(d)}$ between the paired precision values of YOLOv6 and YOLOv7 was computed using the formula:

$$\bar{d} = \frac{\sum d_i}{n}$$

where, $d_i = YOLOv7_i - YOLOv6_i$ are the individual paired differences, and n is the number of paired observations (in this case, n =10). The calculated mean difference is:

$$\bar{d} = 0.028$$

This result indicates that, on average, YOLOv7's precision is 0.028 higher than YOLOv6's precision.

2) Standard deviation and standard error: The standard deviation of the differences (s_d) was calculated to measure the variability among the paired differences, using the formula:

$$(s_d) = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$$

The standard deviation of the paired differences was found to be:

$$(s_d) = 0.0103$$

The standard error of the mean difference (SE) was then calculated as:

$$SE = \frac{s_d}{\sqrt{n}}$$

The resulting standard error is:

$$SE = 0.00327$$

This small standard error indicates that the sample mean difference is a precise estimate of the population mean difference.

3) t-Statistics and p-Value: To test the null hypothesis (H_0) that there is no significant difference in precision between YOLOv6 and YOLOv7, the t-statistic was calculated using the formula:

$$t = \frac{\bar{d}}{SE}$$

The calculated t-value is:

$$t = \frac{0.028}{0.00327} = 8.573$$

The p-value corresponding to this t-value, for 9 degrees of freedom, is:

$$p = 1.27 \times 10^{-5}$$

The null hypothesis is rejected since the p-value is much smaller than the significance threshold (α=0.05). This confirms that the improvement in precision from YOLOv6, YOLOv7 is statistically significant.

4) Confidence interval: The 95% confidence interval for the mean difference was calculated as:

$$CI = \bar{d} \pm t_{critical} \times SE$$

where, $t_{critical} = 2.262$ is the critical t-value for 9 degrees of freedom at a 95% confidence level. The resulting confidence interval is:

$$CI = 0.028 \pm (2.262 \times 0.00327)$$

$$CI = [0.0206, 0.0354]$$

$$CI = [0.0206, 0.0354]$$

This indicates that the true mean difference in precision is between 0.0206 and 0.0354 with 95% confidence.

5) Conclusion: The statistical analysis demonstrates that YOLOv7's improvement in precision over YOLOv6 is statistically significant, with a mean improvement of 0.028 and a confidence interval of [0.0206, 0.0354]. The extremely low p-value (1.27×10^{-5}) further validates the reliability of the results. These findings highlight the superior performance of YOLOv7, making it a more effective model for real-time detection of partially occluded faces.

E. Image Prediction

The performance of the detection network in YOLOv6 and YOLOv7 models is listed under this subsection based on the prediction output. The models were tested on face occlusion images and at different lighting conditions to test their accuracy in detection and their strength. The predictions show the ability to perceive partial masks coverages of faces such as visors, as well as at low illumination levels of both models. The bounding boxes and confidence scores make the visual and quantitative evaluation of the detection performance possible, which proves the effectiveness of these models in practice. The sample image prediction of the YOLOv6 model is displayed in Fig. 9.

Fig. 9. YOLOv6- image prediction of partially occluded faces in low-light.

However, YOLOv6 proves that blocking can partially hide dark luminance faces in an image. The model puts a bounding box around a detected face and attributes a confidence score to it, such as 0.8 or 0.3. These reflect the conviction of the model in its successful detection. The higher the confidence score, the more firmly the model believes in its accurate detection. The face detection model shows good performance and dependability when tested with real-world input, which comes complete with both difficult lighting conditions and incomplete face obstructions. Fig. 10 displays the output of the YOLOv7 model when applied to the sample image.

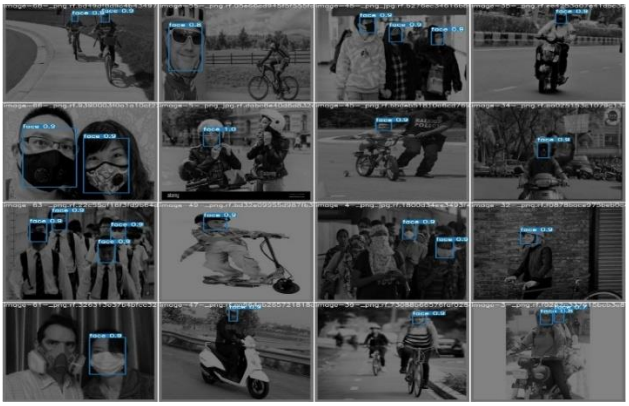

Fig. 10. YOLOv7- image prediction of partially occluded faces in low-light.

YOLOv7 model detects faces that are partially obscured in low-light settings by sketching blue bounding boxes with confidence levels starting from 0.8 up to 1.0. The model does proper face detection and localization within advanced environments because it can rightly identify faces even under difficult conditions which include masks as well as helmets and poor lighting.

VI. Discussion

In this research, the YOLOv6 and YOLOv7 performance for detecting partially occluded faces under low-light conditions was considered. The results show that deep learning YOLO models are on the way to the occlusions problem even if it was one of the most challenging tasks for face detection system in the past. It has been found out that YOLOv7 beats YOLOv6 in all major performance metrics which include precision, recall, F1 and mAP. YOLOv7 produced 0.92 precision, which exceeded YOLOv6's score of 0.90 while identifying relevant objects, especially faces, which are partially hidden. Furthermore, YOLOv7 had outperforming the recall ability of detection (0.89 compared to YOLOv6's 0.787) which gives a hint that YOLOv7 is good at detecting faces in various occlusion levels therefore reducing false negatives. YOLOv7 shows improved performance in face detection under occlusion, which corresponds prior study that focus on advance feature extraction method for face occlusion difficulties [22, 23]. YOLOv6 [28] and [23] have better real-time results, accuracy and efficiency in detecting faces in an occluded environment as well as in low-light settings as opposed to conventional algorithms such as the ViolaJones detector [28]and previous versions of the system like Faster R-CNN [47]. Single-step detection algorithm of YOLO is highly applicable in the reality field, whilst the E-ELAN architecture of YOLOv7 [23] is more effective in extracting features, thus strengthening features in adverse conditions.

These benefits make YOLOv7 particularly suitable to the real-life security and surveillance operations. This technological improvement of YOLOv7 can be explained by various reasons. To begin with, the application of E-ELAN allows YOLOv7 to embrace more sophisticated and shadowed facial features, even when there is a poor lighting, or when the face is partially covered. This feature will give YOLOv7 a unique edge to handle changes in facial orientation and partial blockages, which are a major issue in the face detection system. Moreover, the structure of YOLOv7 that combines different levels of face aggregation helps it to achieve an increased accuracy rate as it is able to locate smaller and less unclear faces better than YOLOv6 does.

The F1-score, which uses the harmonic average of precision and recall to give an overall performance, show that YOLOv7 is better than YOLOv6 with values 0.90 and 0.845, accordingly. The improvement in performance of YOLOv7 depicts its dual capabilities as in precision face detection as well as lower false positive rates in comparison to the previous version, hence it is highly reliable for accuracy-first detection systems.

A. Training Time and Efficiency Considerations

This study proved that training times of YOLOv6 and YOLOv7 were different. The experiments presented that it took YOLOv6 44 min 25 s for training and YOLOv7 required only 1 h 19 min 32 s. YOLOv7 takes longer time to train, which may not be suitable for latency-sensitive scenarios although it has better accuracy. In real-time systems consideration must be given to computation vs. detection performance when employing these models. YOLOv7 is the better choice for applications that are concerned with detection performance, as it offers better accuracy and recall at the cost of slower training in these cases [34].

B. Statistical Analysis and Significance of Results

The paired t-test's precision of performance improvement shows statistical significance with YOLOv7. The statistical test gave a p-value of 1.27×10^{5}, which is significantly below the 0.05 threshold, thus demonstrating that YOLOv7's performance improvements are dependable and not caused by random fluctuations. YOLOv7 performed well on partially masked face detection, with the mean precision improvement of 0.028; this value was statistically significant, as indicated by the 95% confidence interval on the mean difference. The results support previous studies, which indicate that YOLOv7 was superior to the competing models in complex object characterization missions due to its innovative components, including the (ELAN) which significantly increases detection effectiveness in low-light of occlusion situations [24].

C. Practical Implications

The enhancements shown by YOLOv7 for partially occluding face detection lead to essential practical uses for applications that depend on face detection, including security systems and surveillance operations, and human-computer interaction systems [35]. YOLOv7 delivers superior performance in occlusion detection, which makes it the best choice for surveillance systems operating in real-time, where detecting faces in difficult conditions, such as dimly lit or crowded areas, becomes crucial [36].

YOLOv7 is shown to have strong features to detect faces when they are hidden by masks or helmets to prove its necessity in the current world situation when masks and helmets are commonly used to protect health-related factors. The system can be combined with access control facilities, customer analytics, and any other, where facial recognition can be utilized to perform identification or authentication services. It is assumed that the improved performance of YOLOv7 on accuracy and recall should decrease the error rates, which occur in real-time

application and hence, increase satisfaction with the results as well as operational efficiency.

D. Ethical Consideration

YOLOv7 is a huge progression in facial recognition, even if it brings up serious ethical concerns by itself with other occluded face detection capabilities. The improvement of their accuracy asked to pay more efforts in data protection among the security measures [37]. Public and workplace environments, where face recognition surveillance take place, cause privacy dangers as these technologies can impinge on personal privacy without appropriate monitoring or individual approval.

Using YOLOv7 and similar models in real-world applications presupposes compliance with ethical standards during their deployment by organizations [38]. These technologies necessitate clear rules of the game that would ensure their responsible application via transparent processes established [39]. It is important that public spaces should have systems in place that will allow individuals to opt out of facial recognition technology, since they might not be informed or agree to the sharing of biometric data.

The results of this research have great implications for applications in the real world in relation to many aspects, such as surveillance, security, and human-computer interaction. The fact that YOLOv7 is able to identify faces with a high degree of success in some difficult situations, like when a person puts on a face mask or the amount of light is insufficient, makes the algorithm offer great utility to applications related to the safety of the crowd, monitoring, and access control.

Face masks gained ubiquity amid the COVID-19 pandemic, and the identification of masked people in the social environment became a crucial problem. The higher performance of YOLOv7 on these situations is a considerable benefit to automated surveillance, access control, and other tasks, which at times require more precise face detection on a real-time basis.

Additionally, the ability of YOLOv7 to perform face detection in low-light areas opens up fresh possibilities of application in night security contexts and other areas of poor visibility, like parking lots, airports, or those offices with low-light conditions where face detection is usually suppressed due to poor lighting. Such conditions might result in more dependable and effective security work done by the better accuracy of YOLOv7.

VII. Conclusion

This research successfully evaluated YOLOv6 and YOLOv7 for detecting faces obscured by masks and helmets in low-light conditions while showing important improvements in real-time object detection. The evaluation demonstrates that YOLOv7 produces better results than YOLOv6 regarding precision and recall, and F1-score, which positions it as the best model for recognizing faces with masks and helmets, and other face coverings.

The Enhanced Efficient Layer Aggregation Network (E-ELAN) and improved feature extraction capabilities in YOLOv7 enable it to deliver better occluded face detection results. YOLOv7 achieves high detection accuracy while simultaneously enhancing its capability to detect various occlusion types because of its innovative features. A statistical evaluation demonstrates that YOLOv7 delivers better precision at a significant level, reinforcing its dependable enhanced performance.

YOLOv7 clearly shows better results compared to YOLOv6 in accuracy as well as recall, although with a higher cost on the computation resource requirements. YOLOv7 also takes more training time than YOLOv6 and thus poses difficulties when it comes to its deployment in real-time application. The trade-offs between YOLOv7 and YOLOv6 need to be carefully considered by the researchers and developers to match model selection with their needs of use. The practical implications of this research are significant. The better ability to localize occluded faces would be of great significance for security and surveillance systems, human-computer interaction, and other domains that depend on accurate face recognition. As face masks become a more common accessory in daily life, as well – especially with the world's focus on global health, being able to identify people when their faces are partially obscured or are covered by something (like sunglasses) is increasingly useful. Furthermore, YOLOv7's efficiency in detecting faces in low-light conditions opens up new opportunities for deployment in environments where visibility is compromised, such as at night or in poorly lit spaces.

Despite such developments, there are a number of limitations. The dataset used in the current research, even though it has a certain level of heterogeneity, is not huge, and it may not be a full-scale representation of a wide range of real-life scenarios where occlusions can occur. Future studies are suggested to aim at enriching the dataset by incorporating a wider range of types of occlusions, both in variety and complexity, as well as by investigating other environmental contingencies, such as changing illumination conditions and multi-object occlusions. Furthermore, the computational demands of YOLOv7 necessitate a closer examination of its efficiency, especially in scenarios requiring rapid inference times.

In the future, it may be interesting to investigate if YOLOv7 can be combined with other state-of-the-art face detection methods like RetinaFace and SSD to yield good results in comparison to YOLOv7. Furthermore, the use of transfer learning techniques may reduce training time for YOLOv7 without sacrificing accuracy and enable real-time adoption.

In conclusion, YOLOv7 represents a significant advancement in the detection of partially occluded faces, offering superior accuracy and recall over YOLOv6. The ability of the model to cope with occlusions, as well as to perform, even during the low-light scenarios, makes it the architecture of choice in face detection tasks in fields like security, surveillance, and human-computer interaction. The improvements demonstrated in this study have significant potential for real-world applications, particularly in security, surveillance, and other sectors reliant on face recognition technology. While there are still challenges to overcome, particularly in balancing computational efficiency with detection performance, the continued evolution of models like YOLOv7 brings us closer to achieving more robust, adaptable, and real-time face detection

systems capable of addressing the complexities of modern environments.

VIII. Research Limitation and Future Work

While this study provides valuable insights into the performance of YOLOv6 and YOLOv7 for partially occluded face detection, several limitations must be acknowledged. The dataset used in this research consisted of only 500 images, which, although representative of different types of occlusion, may not fully capture the diversity of real-world scenarios. Future studies should aim to expand the dataset to include a larger variety of occlusions, facial expressions, and environmental conditions, thus ensuring that the model can generalize to more challenging and dynamic real-world applications.

In addition, this study aimed at comparing two editions of the YOLO model. In the future, we may consider incorporating extra strong models (e.g., RetinaFace or SSD) to further invest in their strengths of recognizing occluded faces. Secondly, adding more challenges (e.g., hard lighting conditions and occlusions), e.g., multi-object occlusion or heavy mask, to the dataset may give us some hints on the generalization of our model.

References

[1] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1), 5.

[2] Cheong, Y. Z., & Chew, W. J. (2018). The application of image processing to solve occlusion issue in object tracking. In MATEC Web of Conferences (Vol. 152, p. 03001). EDP Sciences.

[3] Baccouche, A., Garcia-Zapirain, B., Zheng, Y., & Elmaghraby, A. S. (2022). Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques. Computer Methods and Programs in Biomedicine, 221, 106884.

[4] Baccouche, A., Garcia-Zapirain, B., Zheng, Y., & Elmaghraby, A. S. (2022). Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques. Computer Methods and Programs in Biomedicine, 221, 106884.

[5] When and How to Use Masks. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advicefor-public/when-and-how-to-use-masks (accessed on 3 August 2021).

[6] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).

[7] Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 212–220.

[8] Birogul, S., Temur, G., & Kose, U. (2020). Yolo object recognition algorithm and "buy-sell decision" model over 2D candlestick charts. IEEE Access, 8, 91894–91915

[9] Aiswarya, P., Manish, & Mangalraj, P. (2020). Emotion recognition by inclusion of age and gender parameters with a novel hierarchical approach using Deep Learning. 2020 Advanced Communication Technologies and Signal Processing (ACTS).

[10] Deepa, R., Tamilselvan, E., Abrar, E. S., & Sampath, S. (2019). Comparison of yolo, SSD, faster RCNN for real time tennis ball tracking for Action Decision Networks. 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE).

[11] Chen, J., Wang, C., Wang, K., Yin, C., Zhao, C., Xu, T., Zhang, X., Huang, Z., Liu, M., & Yang, T. (2021). Heu emotion: A large-scale database for multimodal emotion recognition in the wild. Neural Computing and Applications, 33(14), 8669–8685

[12] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee.

[13] Mao, L., Sheng, F., & Zhang, T. (2019). Face occlusion recognition with deep learning in security framework for the IoT. IEEE Access, 7, 174531-174540.

[14] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.

[15] Abbas, S. M., & Singh, S. N. (2018, February). Region-based object detection and classification using faster R-CNN. In 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 1-6). IEEE.

[16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[17] Simonyan, K. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[18] Redmon, J. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[19] Durga, P., Divya, G., Ayeshwariya, D., & Sivakumar, P. (2021, April). Gabor-deep CNN based masked face recognition for fraud prevention. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 990-995). IEEE.

[20] Wu, S., Yang, J., Wang, X., & Li, X. (2022). Iou-balanced loss functions for single-stage object detection. Pattern Recognition Letters, 156, 96-103.

[21] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).

[22] Liu, S., & Agaian, S. S. (2021, April). COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features. In Multimodal Image Exploitation and Learning 2021 (Vol. 11734, pp. 162-171). SPIE.

[23] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[24] Du, S., Zhang, P., Zhang, B., & Xu, H. (2021). Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4. IEEE Access, 9, 25671-25680.

[25] Mandal, B., Okeukwu, A., & Theis, Y. (2021). Masked face recognition using resnet-50. arXiv preprint arXiv:2104.08997.

[26] "Mt-yolov6 pytorch object detection model." [Online]. Available: https://models.roboflow.com/object-detection/mt-yolov6

[27] "Yolov7 pytorch object detection model." [Online]. Available: https://models.roboflow.com/object-detection/yolov7

[28] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.

[29] "Hard hat dataset," 2020. [Online]. Available: https://makeml.app/datasets/hard-hat-workers

[30] Long, X., Cui, W., & Zheng, Z. (2019, March). Safety helmet wearing detection based on deep learning. In 2019 IEEE 3rd information technology, networking, electronic and automation control conference (ITNEC) (pp. 2495-2499). IEEE.

[31] Sarkar, P., De, S., & Gurung, S. (2022, December). Fish detection from underwater images using YOLO and its challenges. In Doctoral Symposium on intelligence enabled research (pp. 149-159). Singapore: Springer Nature Singapore.

[32] Prasetyo, E., Suciati, N., & Fatichah, C. (2021, June). Yolov4-tiny and spatial pyramid pooling for detecting head and tail of fish. In 2021

International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) (pp. 157-161). IEEE.

[33] Qiu, Y., Lu, Y., Wang, Y., & Jiang, H. (2023). IDOD-YOLOV7: Image-dehazing YOLOV7 for object detection in low-light foggy traffic environments. Sensors, 23(3), 1347.

[34] Rouf, M. A., Wu, Q., Yu, X., Iwahori, Y., Wu, H., & Wang, A. (2023). Real-time vehicle detection, tracking and counting system based on YOLOv7. Embedded Selforganising Systems, 10(7), 4-8.

[35] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).

[36] Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. Journal of Computing and Social Informatics, 2(1), 1-12.

[37] Cernadas, E. (2024). Applications of Computer Vision. Electronics, 13(18), 3779.

[38] Shaikh, I. M., Akhtar, M. N., Aabid, A., & Ahmed, O. S. (2024). Enhancing sustainability in the production of palm oil: Creative monitoring methods using YOLOv7 and YOLOv8 for effective plantation management. Biotechnology Reports, 44, e00853.

[39] Zhang, Y., Liu, G., & Wang, J. (2025). Automated detection of sheep eye temperature using thermal images and improved YOLOv7. Computers and Electronics in Agriculture, 230, 109925.

[40] Khattak, A., Asghar, M. Z., Ishaq, Z., Bangyal, W. H., & Hameed, I. A. (2021). Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews. Egyptian Informatics Journal, 22(4), 455-471.

[41] Rukhsar, L., Bangyal, W. H., Nisar, K., & Nisar, S. (2022). Prediction of insurance fraud detection using machine learning algorithms. Mehran University Research Journal of Engineering & Technology, 41(1), 33-40.

[42] Khalid, M., Ashraf, A., Bangyal, W. H., & Iqbal, M. (2023, December). An android application for unwanted vehicle detection and counting. In 2023 International Conference on Human-Centered Cognitive Systems (HCCS) (pp. 1-6). IEEE.

[43] Ahmad, Z., Bangyal, W. H., Nisar, K., Haque, M. R., & Khan, M. A. (2022). Comparative analysis using machine learning techniques for fine grain sentiments. Journal on Artificial Intelligence, 4(1), 49-60.

[44] Ahmad, W., Shahzad, A. R., Amin, M. A., Bangyal, W. H., Alahmadi, T. J., & Khan, S. H. (2025). Machine learning driven dashboard for chronic myeloid leukemia prediction using protein sequences. PLoS One, 20(6), e0321761.

[45] Huang, Y., Chen, M., Zhang, X., Shimoda, R., & Yang, R. (2025). Multi-Scale Street Vitality Analytics: A Comprehensive Review of Technologies, Data, and Applications. Buildings, 15(21), 3987.

[46] Zhang, Y., Wang, J., & Liu, H. (2025). Hybrid multi-task learning for face detection and emotion recognition under occlusion. International Journal of Computer Vision, 132(4), 1025-1039.

[47] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.