Al-Nahrain Journal for Engineering Sciences **NJES** 27(3)351-356, 2024
http://doi.org/10.29194/NJES.27030351

Enhancing Facial Identification Systems with YOLOv8: A Cutting-Edge Approach

Huda S. Mithkhal[1*], Ahmed H Y Al-Noori [2], Emad Tariq Al-Shiekhly[3]

Authors affiliations:
1*) Department of Computer Engineering, University of Al-Nahrain, Baghdad-Iraq.
st.compe.huda.s.mithkhal@ced.nahrainuniv.edu.iq

2) Department of Computer Engineering, University of Al-Nahrain, Baghdad-Iraq.
ahmed.hani.al-noori@nahrainuniv.edu.iq

3) College of Business Administration, Prince Mohammad Bin Fahd University, KSA.
etariq@pmu.edu.sa

Paper History:

Received: 5th May 2024

Revised: 29th Jul. 2024

Accepted: 6th Aug. 2024

Abstract

Face recognition and identification have recently become the most widely employed biometric authentication technologies, especially for access to persons and other security purposes. It represents one of the most significant pattern recognition technologies that uses characteristics included in facial images or videos to detect the identity of individuals. However, most of the traditional facial algorithms have faced limitations in identification and verification accuracy.

As a result, this paper presents a sophisticated system for face identification adopting a novel algorithm of deep learning, namely, You Only Look Once version 8 (YOLOv8). This system can detect the face identity of different individuals with different positions with high accuracy. The YOLOv8 model has been trained for several target face images classified as training and validation images of 1190 and 255, respectively. The experimental results show a significant improvement in face identification accuracy of 99% of mean average precision, which outperforms many state-of-the-art face identification techniques.

Keywords: Face-Identification, Artificial Intelligence, Deep Learning, Convolutional Neural Networks, You Only Look Once (YOLO).

تحسين أنظمة التعرف على الوجه باستخدام YOLOv8: نهج متطور

هدى صدام مثكال ، احمد هاني يوسف، عماد طارق الشيخلي

الخلاصة:

لقد أصبح التعرف على الوجوه وتحديد الهوية في الآونة الأخيرة من أكثر تقنيات المصادقة البيومترية استخدامًا على نطاق واسع، خاصة للوصول إلى الأشخاص ولأغراض أمنية أخرى. وهي تمثل إحدى أهم تقنيات التعرف على الأنماط التي تستخدم الخصائص المضمنة في صور الوجه أو مقاطع الفيديو للكشف عن هوية الأفراد. لقد واجهت معظم خوارزميات الوجه التقليدية قيودًا في دقة التحديد والتحقق.

نتيجة لذلك، تقدم هذه الورقة نظامًا متطورًا للتعرف على الوجه يعتمد خوارزمية جديدة للتعلم العميق، وهي "أنت تنظر مرة واحدة فقط" الإصدار ٨ (YOLOv8). يمكن لهذا النظام اكتشاف هوية الوجه لأفراد مختلفين بأوضاع مختلفة بدقة عالية. تم تدريب نموذج YOLOv8 على العديد من صور الوجه المستهدفة المصنفة على أنها صور تدريب وصور للتحقق من الصحة تبلغ ١١٩٠ و٢٥٥ على التوالي. تظهر النتائج التجريبية تحسنًا كبيرًا في دقة التعرف على الوجه بنسبة ٩٩٪ من متوسط الدقة، وهو ما يتفوق على العديد من تقنيات التعرف على الوجه الحديثة.

1. Introduction

Face recognition (or face biometrics) has become one of the most significant biometric techniques for identifying or verifying individuals since it is easy to capture and implement compared with other types of biometrics [1]. Face identification is a biometric technique that utilizes distinctive features of human faces to identify or verify an individual's identity. The face recognition system generally involves two phases [2].

Face Detection and Pre-processing: where the image or video is captured to find any individual face, then the image of the face is pre-processed, and features are extracted from it to use as a reference for this face. In the recognition phase, the identifying features in each face image are extracted in the same way as in phase one and compared with existing face features in the database to ensure the identity of each input face image [2,3].Several techniques and algorithms have been developed to improve face recognition performance, Such as Local Binary Pattern (LBP) [4], Principal Component Analysis (PCA) [5], Local Binary Pattern Histograms (LBPH) [6], K-Nearest Neighbor (KNN) [7], and Support Vector Machine (SVM) [7,16]. Recently, deep learning algorithms have been highly effective for computer vision applications because of their efficiency in detecting various types of objects. Deep learning is a subfield of Machine Learning (ML) and Artificial Intelligence(AI) that mimics the way that the human brain gains certain types of knowledge[9].

We used advanced model deep learning You Only Look Once (YOLO) to build the face identification system and test it on difficult datasets. The CNN deep learning model predicts class probabilities and bounding boxes for all the objects demonstrated in sample images by looking only once at the image. [9] For that reason, YOLO takes its name from You Only Look Once [10]. YOLO is one of the most effective algorithms in the field of object detection [11].

In this work, the faces of persons are identified using a You Only Look Once version 8 (YOLOv8) algorithm. This technique can identify the faces of individuals with high identification system accuracy [8].

2. Related Works

This section presents some face recognition and identification-related works, focusing on the state of the art in this field.

In 2018 Yan et al.[12], presenting a combination of two methods to create a new model for recognizing faces. Markov Stationary Features (MSF) and Vector Quantization (VQ) approaches. The VQ algorithm was performed to quantify the face image's sub-regions into 33 levels. The Multiple Image Sub-Regions (MSR) are combined with the MSF-VQ method to provide features with position information and relations. SVM was utilized for the classification stage. This approach used five types of datasets: FERET, ORL, CAS-PEAL-R1, Yale, and Yale-B face images. The recognition accuracy for each database is 99.16%, 100%, 98.40%, 100%, and 93.69, respectively.

The proposed model has high identification accuracy and processes more frames per second.

Allagwail et al., in 2019 [4], Introduced the study of face identification with symmetric training images of the face dependent on the Gabor filter, and Local Binary Patterns (LBP) were employed. In addition, local binary pattern-based 2D discrete wavelet transforms with a single-level Gaussian low-pass filter are. The datasets (ORL and Yale) were used. The performance results showed 100% identification for both datasets.

Ling et al., in 2019 [13], Propose using a Convolutional Neural Network (CNN) that relies on self-residual attention for selecting features to recognize faces. This method concentrates on the long-range dependencies of face images by decreasing information redundancies and focusing on the most significant components of the space function and channels. The datasets employed in this work were LFW, Age DB, and CFP. The results of the performance to recognize faces were LFW (99.83%), Age DB (98.47%), and CFP (95.6%), respectively.

Deeba, F., in 2019 [6], Presented a system based on the Local Binary Pattern Histograms (LBPH) technique to recognize faces in the proposed system. This technique is split into three main steps: face detection, facial feature extraction, and image categorization. The framework of the face recognition system uses the HAAR CASCADE algorithm for detection and uses image processing to prepare images. A dataset of face variants was used as a database to train pictures and save images of various persons (30 images per person). The result showed that the accuracy of the face recognition rate was 80 %.

Karanwal Sh. Al., in 2022 [14], Indicated a developed system for face identification by fusing features computed from the HE LBP, LBP, and RD-LBP approaches. After the fusion of these features, the PCA is done to reduce the number of features extracted. The Support Vector Machine (SVM) is used for classification, and images of the face are taken from the ORL and GT datasets. The dataset of the ORL and GT identification accuracy is 99.16% and 91.33%, respectively.

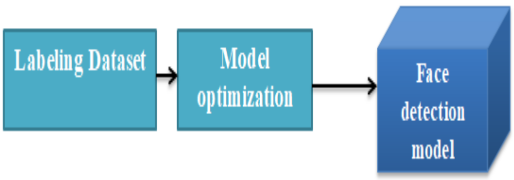
Gautam, G. et al., in 2023 [15], Proposed a model for autism spectrum disorder screening using facial images using YOLOv8 deep learning on a Kaggle. This method achieved a higher accuracy of 89.64% in classification. These results indicate that utilizing the YOLOv8 deep learning model for screening autism spectrum disorder In facial images of children is an effective and cost-efficient method (ASD) to facilitate early intervention.

Briefly, the above researchers attempted to recognize faces using different techniques. This study proposed a new model employed to detect and identify human faces based on YOLOv8 deep learning algorithm.

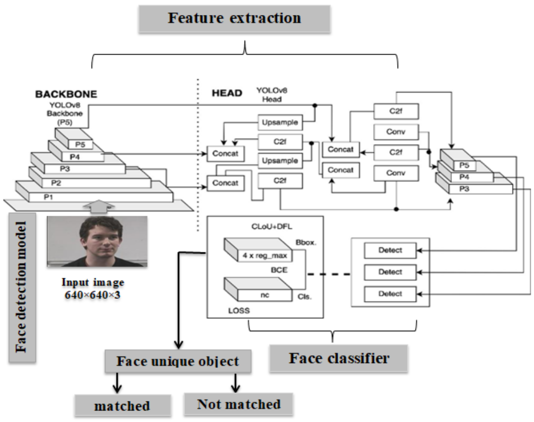
3. Facial Identification based YOLOv8 model

The model works with multiple images for the region of the face to be identified as the input images. labeling the dataset, optimizing the model, and training the model were performed using the YOLOv8 algorithm (Figure 1 a). The image containing the face is provided to the model file as input. Then, the region where the face is detected is set as output, and bounding box information is acquired.

Finally, the trained Yolo model extracts a 3D feature map. The features extracted are compared with the suspect individual features in the database, and then the classification of these features and the match final score is taken based on the similarity of facial features. The face identification is performed on the frames of the image files given as input to calculate the similarity score. This score is determined by whether the face image of the individual is matched; otherwise, it is not reached, considering there is little or no similarity. The identification is recorded in a database in (Excel) format so that it can search later for which persons were identified in the images. (Figure1b.) demonstrates the complete process steps.



(a)



(b)

Figure (1-a): The process of Face detection; (1-b): The proposed system of Face identification.

4. You Only Looks Once Model

You Only Look Once (shortly YOLO) is one of the most popular deep learning models for real-time object detection and image recognition. It represents a rising computer vision model since 2015. YOLO was created by Ali Farhadi and Joseph Redmon at the University of Washington.[10]

The YOLO algorithm has become one of the most common models in object detection. The work of the YOLO algorithm is mainly based on principles of regression, which predict and analyze the approaches in the entire image, such as speed, and accuracy in identifying objects. [16]

The YOLO algorithm uses a CNN to predict class probabilities and bounding boxes for all objects in sample images, doing so in a single pass.

Originally designed for object detection, YOLO has become popular in classification fields due to its fast inference.[10]

YOLOv8 represents the latest version of YOLO and uses cutting-edge developments in deep learning and computer vision to provide efficient, flexible, and high-performance real-time image classification and object detection.[17]

Its deep CNN architecture includes novel features and improvements such as the Cross Stage Partial Network (CSPNet) backbone architecture, the Feature Pyramid Network (FPN) neck structure, and the Pyramid Attention Network (PAN) head architecture, which enable it to handle scale variations and occlusion more robustly.[18]

YOLOv8 divides input images into a grid of cells and predicts a set of bounding boxes in each cell, then selects the most likely ones for each object.

To augment images during training, the model shows a little different of the images at each epoch. With a focus on size, accuracy, and speed, YOLOv8 is an excellent option for various visual AI applications. [8].

The full details of the YOLOv8 network are demonstrated in Figure (2)

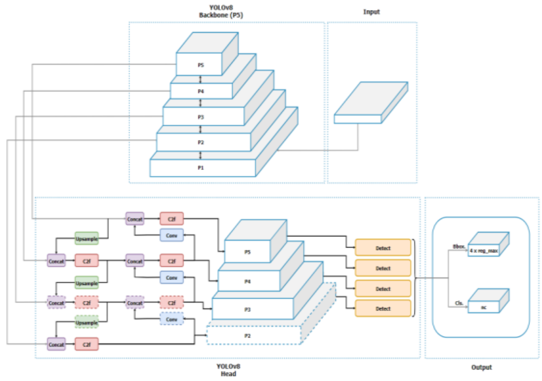


Figure (2): The architecture of YOLOv8 [17]

5. Experimental Setup

In this section, the proposed face identification system based on YOLOv8 has been presented with materials and methods that are used in this system:

a- Dataset and Pre-Processing

The head pose images (HPI) dataset was employed [19]. This dataset comprises 2790 monocular face images of 15 people with pan and tilt angle variations from +90 to -90 degrees. For every person, there are two groups of 93 images (93 different poses). The purpose of having two groups per person is to be able to train and test algorithms on known and unknown faces. People in the dataset have different skin colors and wear glasses or not. Furthermore, the background is neutral and uncluttered to focus on facial details. Face positions on each image are labeled in an individual text file. Figure (3) depicts a small sample of this data.

Figure (3): Samples of HPI dataset of different classes

b- Training Process for the Face Identification Model

The training of the facial identification model includes the following steps: data labeling, model optimization, and a training step. The HPI dataset is utilized considerably in face identification tasks. Images are organized based on 17 classes; much of this data is used to achieve better results. These classes names are (face_ID1, face_ID2, face ID3, face ID4, face ID5, face ID6, face ID7, face ID8, face ID8, face ID9, face ID10, face ID11, face ID12, face ID13, face ID14, face ID15, face ID16, face ID17) as shown in Table (1): The sample dataset of different classes.

Table (1): Sample dataset of different classes

Faces Class	Images Number
face ID1	100
face ID2	100
face ID3	100
face ID4	100
face ID5	100

face ID6	100
face ID7	100
face ID8	100
face ID9	100
face ID10	100
face ID11	100
face ID12	100
face ID13	100
face ID14	100
face ID15	100
face ID16	100
face ID17	100

The face identification process includes data labeling to classify the location of the object in a rectangular area.

Therefore, the location of the object in the image (in this case, the face of the person) is determined by drawing a rectangle with two points (right-top and left-bottom) in the coordinate system. The information about the object in the image is stored in data-containing structures such as TXT, and CSV, which include a file with the same name as the class and information of the image.

c- Performance Metrics

Many metrics have been adopted to evaluate the performance of the proposed model utilizing the YOLOv8 algorithm. The identification is evaluated based on precision, recall, F1-score, and mean average precision (mAP).

Equations 1 to 4 outline these metrics, where APk represents the average precision of class k, and n represents the total number of classes.

Equation (5) illustrates the computation of the F1-score. The F1-score is the harmonic mean of precision and recall, and it is a perfect performance measure for imbalanced data since it considers how distributed the data is, where (R) represents recall, and (P) represents precision.

$$\text{Preecision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \quad \ldots\ldots(1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \quad \ldots\ldots(2)$$

$$\text{Average Precision(AP)} = \int_0^1 \text{P (r) dr} \quad \ldots\ldots(3)$$

$$\text{mean Average Precision (mAP)} \frac{1}{\text{C}} \sum_{\text{k}=1}^{\text{k}=\text{n}} \Delta\text{R(k) R(k)}\ldots(4)$$

$$\text{F1 Score } \frac{1}{\frac{1}{2}\left(\frac{1}{\text{p}}+\frac{1}{\text{R}}\right)} = \frac{2*\text{P}*\text{R}}{(\text{P}+\text{R})} \quad \ldots\ldots(5)$$

where:

- **TP:** refers to the number of positive samples predicted as positive.
- **FP:** refers to the number of negative samples predicted as positive.
- **FN: represents** the number of negative samples predicted as negative.

6. Experimental Results

As mentioned before, seventeen target face images from the validation set were selected to represent the identification of faces. Specifically, 1700 images have been used. This dataset was divided into 70% training data (1190 photos), 15% validation data (255 photos), and 15% test data (255 photos) to achieve high performance, as shown in Table 2

Table (2): Dataset for faces

Training set	Testing set	Validating set
1190	25٥	25٥

The identification results can be increased to a certain extent to ensure the efficiency of the face recognition system in complicated scenes.

Fig. 4 represents the face identification for different persons with different positions, in addition to an identification accuracy rate for each one.

For instance, the system can identify the persons wearing classes with high performance, as seen in Face ID4 and Face ID6, with high identification accuracy of 97% and 94%, respectively.

Figure (4) On the other hand, the proposed system can identify faces with different positions, as seen in face ID 12, with an identification accuracy of 94%.

Figure (4): Accuracy of Face identification system based YOLOv8 model.

The precision and recall of the proposed face identification model for each individual are illustrated in Table (3). The results show how the precision values vary with the increase in recall values during training. It shows that the model attained an overall mean average precision (mAP) of 99% in all face classes. All results are close to each other and are concentrated in the upper right corner, indicating that the model can accurately predict the identity of the faces in input images.

In Figure (6-a, b): box loss (box_loss), target loss (obj_loss), and classification loss (cls_loss) are presented. Box loss indicates the extent to which the algorithm can locate the center of an object and the extent to which the predicted bounding box covers the object. Classification loss, on the other hand, allows the algorithm to predict the correct class of a given object. The model rapidly improves precision, recall, and average accuracy, leveling off after about 100 epochs. As shown in Figure (6-a,b).

Table (3): Results of the Training dataset

Face Class	P	R	mAP	F1-score
face ID1	0.97	1	0.995	0.787
face ID2	0.977	1	0.922	0.975
face ID3	0.975	1	0.995	0.980
face ID4	0.967	1	0.995	0.977
face ID5	0.972	1	0.995	0.977
face ID6	0.789	1	0.995	0.978
face ID7	0.99	1	0.995	0.987
face ID8	0.789	1	0.995	0.989
face ID9	0.972	1	0.995	0.783
face ID10	0.971	1	0.995	0.977
face ID11	0.967	1	0.995	0.971
face ID12	0.969	1	0.995	0.961
face ID13	0.963	1	0.995	0.964
face ID1٤	0.965	1	0.995	0.960
face ID1٥	0.973	1	0.995	0.970
face ID1٦	0.972	1	0.995	0.976
face ID1٧	0.971	0.75	0.945	0.970
All classes	0.963	0.985	0.992	0.97

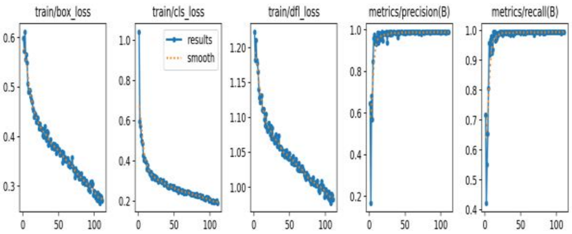


Figure (6-a): Plots of box loss, abjectness loss, classification loss, precision, recall and mean average precision (mAP) over the training epochs for the training set.

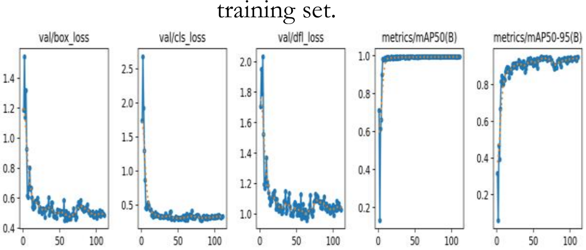


Figure (6-b): Plots of box loss, abjectness loss, classification loss, precision, recall, and mean average precision (mAP) over the training epochs for the validation set.

A normalized confusion matrix is also made to assess the effectiveness of the proposed model's classification, as seen in Figure (7).
The Performance Comparison curve exhibits the Precision-Recall plots for all face categories of the designated query image. This graphical representation serves as an effective tool to compare and contrast the precision and recall values of each face class. The curve offers a clear overview of the performance of the face classes in the selected image and demonstrates how to algorithm works in accuracy in different position in faces.

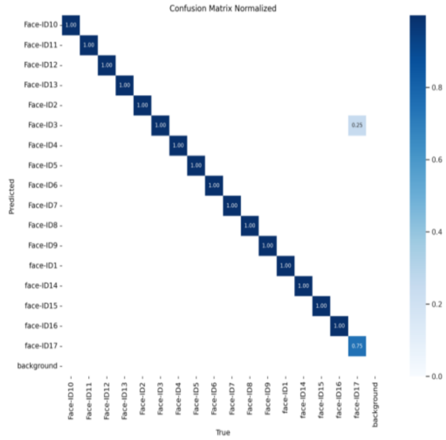


Figure (7): Confusion matrix normalized

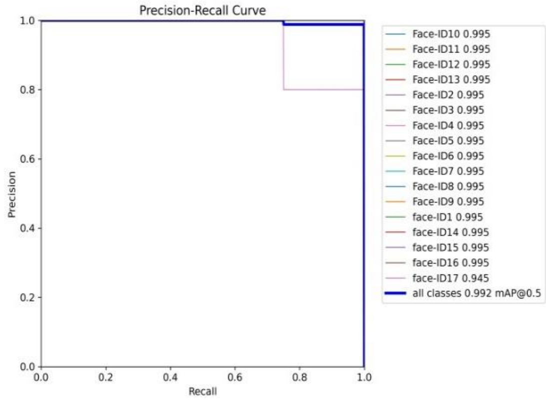


Figure (8): Precision-Recall curve

Finally, this work achieved the best results, as shown in Table (4) The total precision was 97 %, the Total Recall was 98%, and the Total mAP was 99%, which is very high compared to other works using the YOLOv8 model.

Table (4): The total precision, the Total Recall and the Total mAP.

Model Training	Time (min)	precision	F1-Score	recall	mAP
YOLOv8	56m and 2s	97%	99%	98%	99%

7. Conclusions

This study has presented a robust model for face identification, depending on the YOLOv8 algorithm. The YOLOv8 model was trained on the Head Pose Image Database. The network backbone developed an anchor-free detection head, and the function of a novel-reducing loss makes it much faster. As a result, the system shows that the performance of the YOLOv8 algorithm is exceptionally significant. The algorithm is working significantly in terms of face identification accuracy and false detection rate. Moreover, it has the advantage of being in real-time, which meets the requirements of high speed and high accuracy when identification faces complex scenes.

In future studies, the increase of intra-identity variation brings a much more significant performance gain for large-scale face recognition than simply increasing variation overall. There is a need to improve research and studies to develop and employ systems that can classify and identify biometric identification in general and face identification in particular.

8. References

[1] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, 'Face recognition systems: A survey', *Sensors (Switzerland)*, vol. 20, no. 2, 2020, doi: 10.3390/s20020342.

[2] M. Kasar and S. Patil, 'Study and Analysis of Facial Landmark Detection Techniques', no. August, 2022.

[3] R. Ullah *et al.*, 'A Real-Time Framework for Human Face Detection and Recognition in CCTV Images', *Mathematical Problems in Engineering*, vol. 2022, 2022, doi: 10.1155/2022/3276704.

[4] A. Elbkosh, 'Face Recognition with Symmetrical Face Training Samples Based on Histograms Face Recognition with Symmetrical Face Training Samples Based on Histograms of Oriented Gradients', no. December, 2020.

[5] A. A. Aburomman and M. B. I. Reaz, 'Ensemble SVM classifiers based on PCA and LDA for IDS', *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering, ICAEES 2016*, pp. 95–99, 2016, doi: 10.1109/ICAEES.2016.7888016.

[6] F. Deeba and A. Ahmed, 'LBPH-based Enhanced Real-Time Face Recognition', vol. 10, no. 5, pp. 274–280, 2019.

[7] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, 'A new method for face recognition using convolutional neural network', *Advances in Electrical and Electronic Engineering*, vol. 15, no. 4 Special Issue, pp. 663–672, 2017, doi: 10.15598/aeee.v15i4.2389.

[8] M. Talib, A. H. Y. Al-Noori, and J. Suad, 'YOLOv8-CAB: Improved YOLOv8 for Real-time object detection', *Karbala International Journal of Modern Science*, vol. 10, no. 1, 2024, doi: 10.33640/2405-609x.3339.

[9] N. P. Motwani and S. S, 'Human Activities Detection using DeepLearning Technique-YOLOv8', *ITM Web of Conferences*, vol. 56, p. 03003, 2023, doi: 10.1051/itmconf/20235603003.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You Only Look Once: Unified, Real-Time Object Detection', Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.02640

[11] R. Malisetty, 'Race Classification Using YOLO Based on Human Face'.

[12] Y. Yan, F. Lee, X. Wu, and Q. Chen, 'Face recognition algorithm using extended vector quantization histogram features', *PLoS ONE*, vol. 13, no. 1, pp. 1–24, 2018, doi: 10.1371/journal.pone.0190378.

[13] H. Ling, J. Wu, L. Wu, J. Huang, J. Chen, and P. Li, 'Self residual attention network for deep face recognition', *IEEE Access*, vol. 7, pp. 55159–55168, 2019, doi: 10.1109/ACCESS.2019.2913205.

[14] 'S. Karanwal and M. Diwakar, "Feature Fusion of LBP, HELBP & RD-LBP for Face Recognition," in Distributed Computing and Optimization Techniques, Springer, 2022, pp. 471 – 480.', p. 2022, 2022.

[15] S. Gautam *et al.*, 'Screening Autism Spectrum Disorder in childrens using Deep Learning Approach : Evaluating the classification model of YOLOv8 by comparing with other models'.

[16] M. T. Abdullah and J. H. Alameri, 'A Multi-Weapon Detection Using Synthetic Dataset and Yolov5', *Proceedings - CSCTIT 2022: 5th College of Science International Conference on Recent Trends in Information Technology*, pp. 100–104, 2022, doi: 10.1109/CSCTIT56299.2022.10145737.

[17] N. Karna, A. P. Putra, S. M. Rachmawati, R. Abiyev, and G. A. Sampedro, 'Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Towards Accurate Fused Deposition Modeling 3D Printer Fault Detection using Improved YOLOv8 with Hyperparameter Optimization', doi: 10.1109/ACCESS.2017.DOI.

[18] J. Terven, 'A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS', *Machine Learning and Knowledge Extraction*, vol. 5, no. 4. Multidisciplinary Digital Publishing Institute (MDPI), pp. 1680–1716, Dec. 01, 2023. doi: 10.3390/make5040083.

[19] N. Gourier, D. Hall, and J. L. Crowley, 'Estimating face orientation from robust detection of salient facial structures', *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)*, pp. 17–25, 2004. Available: http://www.homepages.inf.ed.ac.uk/rbf/CAVIAR/PAPERS/Pointing04-Gourier.pdf