J. Integr. Sci. Technol. 2025, 13(1), 1010 Article

Journal of Integrated
SCIENCE & TECHNOLOGY

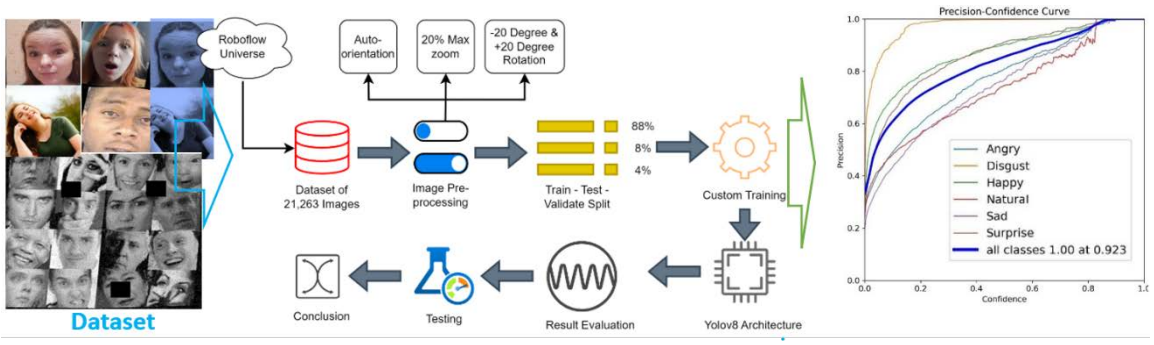
Enhancing emotion recognition in controlled environments with YoLoNetv8

Ekta Singh*, Parma Nand

Department of Computer Science and Engineering, Sharda University, Uttar Pradesh, India

Received on: 11-Jun-2024, Accepted and Published on: 25-Sep-2024

ABSTRACT



Traditional facial expression recognition (FER) approaches for understanding human emotional signals have limitations such as preprocessing, feature extraction, and multi-stage classification, which require significant processing power and computational complexity. Nevertheless, an advanced object detection model like YOLOv8 is favored for its calaboose and accuracy. The designed approach of facial emotion recognition in controlled environments employs the YOLOv8 model to enhance accuracy. The research employs a dataset of 21,263 images that are categorized into six emotions: There are six basic emotions Joy, Sorrow, Anger, Disgust, Neutral, and Surprise. To enhance the model's resilience, the images went through a preprocessing stage that included auto-orientation, magnification, rotation, and resizing. This dataset used to train the designed model on a Kaggle T4 GPU and the model yielded satisfactory accuracy within emotions' identification and categorization. The ability of the designed model in real-time emotion detection was given by the assessment of the model's performance by familiar parameters that include precision, recall, and mean Average Precision (mAP). This study utilizes various facial expressions and new participants to enhance human-computer interaction and mental health evaluation, contributing to the advancement of affective computing.

Keywords: YOLOv8, facial expression recognition, deep learning, human-computer interaction, emotions

INTRODUCTION

The way people interact, including verbal communication, is closely tied to the identification and understanding of gestures. Recognizing and interpreting Facial Emotional Responses (FERs) is a key issue in fields such as affective computing, computer vision, and human-computer interaction since faces are seen as the main way to express emotions.[1] This technology has applications in enhancing human-computer interactions, creating socially aware agents, improving customer relations, marketing strategies, and even mental health solutions. Previously, creating an automatic facial expression recognition system based on ACCD required a complex classification process with hand-crafted features, but deep learning, particularly Convolutional Neural Networks (CNN), has transformed the field.[2] These models require less feature engineering because they learn complex features from the data themselves. However, while deep learning models perform well in controlled environments, they face challenges in real-life applications. Factors such as subtle emotional expressions, occlusion, head positioning, and lighting conditions present significant barriers.[3] Most research in this field focuses on feature extraction and classification.[4] Feature extraction involves identifying attributes from images or videos that can aid classification, using methods like Gabor wavelet transform, Haar wavelet transform, LBP, and AAM, often followed by dimensionality reduction.[5] Facial expressions are then classified using algorithms such as HMM, SVM, AdaBoost, and ANN, with performance dependent on the created expression categories.[6] To simplify the steps for explicit feature extraction and data management, methods like Fast R-CNN and Feature Redundancy-Reduced CNN (FRR-CNN) have been introduced, using convolutional kernels to generate lower-dimensional hidden features for better performance.[7] The latest developments in facial expression recognition have led to improvements in CNN algorithms but also highlight issues like long training times and poor performance in complex environments. This paper introduces

*Corresponding Author: Ekta Singh, Department of Computer Science and Engineering, Sharda University, Uttar Pradesh, India. Email: 2021210063.ekta@dr.sharda.ac.in

Cite as: *J. Integr. Sci. Technol., 2025, 13(1), 1010.*
DOI: 10.62110/sciencein.jist.2025.v13.1010

a new method for estimating facial emotions using the YOLOv8 deep learning architecture,[8] which differs from previous approaches that relied on CNNs for image classification. YOLOv8 allows for face detection and emotion identification in a single process, rather than separating these tasks into two stages. This could enhance real-time results and provide robustness. YOLOv8's high throughput and accuracy make it ideal for real-time emotion recognition, which can be useful in applications such as robotics or video analysisThrough understanding and implementing this strategy, we hope that it will help to initiate research on more durable and practical effects for effective computing systems.[8] This work describes the different aspects of the model based on YOLOv8, including training processes, the data set, and the outcomes of the experiment, including the problems faced and possible future research that can be done in this forming field.[9] This paper presents facial expression recognition using YOLOv8 and image edge detection to tackle the stated challenges. The main contributions that signify this method are

- Taking the texture image's edge structure information and superimposing it on top of each feature map after extracting the edges of each input image layer.
- To decrease the training time of the convolutional neural network model, we use the maximum pooling method to decrease the dimension of the retrieved implicit features.
- The simulation tests were run with the Fer-2013 database of facial expressions and the LFW dataset (Labeled Faces in the Wild) to show how well the suggested strategy works in challenging environments.

RELATED WORK

The research proposed in the article is suggestive of a unique model named FER-YOLO-Mamba model that makes use of two distinct technologies called Mamba and YOLO for effective recognition of facial expressions and also helps in localization of the facial features that contribute to the expression recognition. The model proposes a VSS-based dual branch external module that strengthens the convolutional layers and aids the state space modularities in targeting and detecting the face image's local and distant dependencies. Several experimentations performed on RAF-DB and SFEW datasets are convincing the results given the proposed FER-YOLO-Mamba model outperforms the existing model reported in the domain.[10] Another study that has been reviewed to formulate this literature analysis talks about several deployments and developments in emotion recognition and identification using facial features and highlights the challenges posed by the traditional methodologies used for recognition. It is notable that traditional emotion recognition technologies mostly used to be dependent on visual signals. The research approach brings to light the caliber of conclusively using thermal imaging and other parameters-based studies. The study has analyzed specifically the deep learning models, from basic to advanced ones. The targeted upcoming emerging models of the domain, such as CNNs, and several versions of YOLONet, such as YOLOv3 and YOLOV5 have their specific relevances.[6,11] The precise elaboration provided on the impact of precision achieved in terms of performance for recognition of emotions and its influence on other social arenas like the automation sector, personal technologies and devices, the education sector, and other areas of marketing and healthcare. A brand-new method of recognizing cat face expressions by blending Canny edge detection with Convolutional Neural Networks (CNNs) is presented in this research paper. The CNN model employs dropout regularization to minimize overfitting. Additionally, the YOLO model integrates cat breed identification into the system. The suggested method demonstrates a remarkable average accuracy of 87% in detecting basic emotional states among cats. Besides, this conversation touches upon several practical applications of this technology, such as creating mobile and PC software for pet behavior analysis.[12] This article presents a state-of-the-art real-time face recognition network that merges the InsightFace 2D and 3D face analysis module with the YOLO-V7 deep learning model. The presented method aims to overcome the challenges of identifying hidden or disguised faces in current facial recognition systems. YOLO-V7 is known for its speed and accuracy for real-time applications, while InsightFace generates highly discriminative face embeddings, thereby enhancing recognition. In this regard, it is significant to mention that face recognition has become one of the techniques for non-invasive biometric identification because it is easy to use and keeps a high level of hygiene. By leveraging the advantages of YOLO-V7 and InsightFace, this innovative system can improve precision and reliability for real-time facial recognition in various applications.[13] This work presents YOLO-FaceV2, a face detector that runs in real time. YOLO-FaceV2 is built on the one-stage deep-learning architecture of YOLOv5. Some components have been added to improve the detection of faces, particularly hidden ones and those that are small-sized. Examples include RFE, NWD Loss, which aims to make IoU more sensitive in detecting small things, Repulsion Loss, SEAM attention module for dealing with occlusions, and Slide weight function to ensure balanced training Easy hard samples. The WiderFace dataset experiments show that YOLO-FaceV2 surpasses YOLO and variants for different difficulty levels, demonstrating its good performance in real-time face recognition scenarios.[14] This study looks at how music and emotions connect through brain science. It improves music suggestion systems by using Facial Emotion Recognition (FER) with the YOLOv8 algorithm. This mixes Content-Based and Collaborative Filtering to change in real-time. The system uses Spotify data and applies TF-IDF to study genres, cosine similarity to find similar songs, Singular Value Decomposition for User-Based Filtering, and k-nearest Neighbors for Item-Based Filtering. The YOLOv8 model got 86.20% accuracy with the FER+ dataset. Future work aims to improve emotion recognition, diversify datasets, and create better ways to measure success. Autism Spectrum Disorder (ASD) is a condition that affects growth. It causes problems with talking and social skills because of brain differences. Finding it is key but often hard with regular methods. This paper presents a deep convolutional neural network (DCNN)-based system that recognizes emotions in real time. It's made for autistic children, who often show unique facial expressions. The system spots six emotions: surprise, delight, sadness, fear, joy, and natural. This helps to find autism and start help sooner. The new

method, AutYOLO-ATT, makes the YOLOv8 model better by adding an attention mechanism. It works well, with 93.97% precision, 97.5% recall, 92.99% F1-score, and 97.2% accuracy. These results show it could work well in real-life situations.[8] Scientists have created a CNN model that predicts real-time mood states by looking at people's faces. The model focuses on six moods: Tension, Frustration, Anxiety, Fatigue, Neutrality, and Happiness. The team built and cleaned up a dataset, ensuring all images were the same size and lighting. They increased the number of training images from 2500 to 4000. The team used a step-by-step approach, making sure each version of the model worked well and was built on the last one. This helped them get the best accuracy and mean Average Precision, always keeping an mAP of 99.5% and an accuracy of 99%. They trained the model by hand using YOLOv8 object detection on Google Colab. After 155 training rounds, the YOLOv8 model did its best at round 129, with 94.3% accuracy. The outcomes achieved by this experimentation assures of model's capabilities to significantly analyze human emotions and respond to it. This study looks at how Smart AI Cameras can make security better in many industries. It pays special attention to spotting objects during the day and night. The research tests CNNs and RNNs with YOLOv8 to see how well they work with datasets with different objects, lighting, and camera angles. The team compares these methods to other studies, showing how important camera placement, lighting, and algorithm choice are for finding objects well. Studies show that smart AI cameras can detect and track moving objects and work with other security systems to enhance safety. The study results provide important insights into a project implementing security-focused solutions that demonstrate that student attendance is critical to effective learning.[15] Attendance plays an important role in student learning. Schools often use sign-ins, QR codes, and RFID tags to track attendance. However, these methods can lead to cheating and forgetfulness. This study suggests a new way to boost attendance accuracy with a custom recognition system. The plan is to put a small computer and camera in each classroom to take pictures. These images will then go through a process to identify faces and log attendance, complete with date and time. Teachers can access this info from a database, so they don't have to take roll calls by hand.[16]

Our faces tell a lot about us through features like eyes, nose, and mouth. Face recognition tech uses these to tell people apart. This tech is everywhere, from unlocking phones to keeping public spaces safe. Some even think facial features can hint at personality traits. This idea sparked us to make a new face data set for YOLO object detection models. Our dataset has 2,116 images with over 10,000 labels across six types: blue eyes, brown eyes, round noses, round eyebrows, pointy noses, and straight eyebrows. We tested different YOLOv8 versions and found the small one did best, scoring 89.9% on accuracy. We then tweaked a lighter YOLOv8 with 211 layers, which did even better at 91.3% accuracy.[17] This study describes a new YOLOv8 enhancement for the detection of distracted driving behavior and driver emotions, unlike the standard YOLOv8, which applied a multi-head self-attention mechanism (MHSA) and a convolutional neural network (CNN) module to improve accuracy and convergence. MHSA targets distracted driving, while the CNN module focuses on detecting driver emotions. The case study was developed employing the FER2013 dataset and custom dataset. The main conclusion is that the improved YOLOv8 method is more effective than standard YOLO-based approaches. The practicality and efficiency of this method are extended by implementing the FER2013 dataset and presenting the fine-tuned version of the pre-trained YOLOv8 models on the Jetson Nano platform. Additional speed and processing enhancements are done with TensorRT and DeepStream CUDA for the best results in a Jetson Nano platform.[18]

All the deep learning progress seen in computers and vision is related to using the YOLO series of pattern intelligence. This will explore the best way to perform automatic face detection using a YOLOv8 model implemented on OpenCV. The authors first found the settings that determine the confidence and the threshold for point detection that correspond to each other before configuring our network architecture. Another significant aspect is that the model can change the image size and padding of the input images, allowing the model to remain consistent in detection and precise with the results. Our idea uses a deep learning technique that first looks for the location of a face and then looks at those points on the face to do the landmark detection. Preliminary benchmarks demonstrate the model's speed and precision accuracy.[19] New previously trained YOLOv8 object recognition models-ranging from nano to very large-are used in this study. These images were trained on the Wider-Face dataset, and their goal was to identify suspects from digital images and videos in digital forensics. Obtaining mean accuracy (mAP) values ranging from 97.513% to 99.032%, the YOLOv8 architecture showed better performance than YOLOv5, exceeding 7.1% to 8.8% in mAP Real-time image analysis to assist digital forensic experts in identification and identify individuals with intellectual disabilities as part of this study developed capable Desktop applications.[20] In this study, pre-trained YOLOv8 object recognition models use the value from nano to extra-large. These images were trained on the Wider-Face dataset, and their goal was to identify suspects from digital images and videos in digital forensics.[21] This article proposes an enhanced spatio-temporal learning network (ESTLNet) for improving dynamic face recognition. This includes a spatial fusion learning module (SFLM) to obtain explicit spatial feature representations through dual-channel feature fusion and a time-varying enhancement module (TTEM) for auto-attentional gated feed-forward networks) to produce a spatial -a temporal model that extracts appropriate time-referenced context components by application Analyzes four dynamic context datasets (DFEW, AFEW, CK+, Oulu-CASIA).[22] Extensive testing shows that this method outperforms existing methods, and dramatic performance improvements occur in dynamic facial expression detection. Introduced a novel feature of hierarchical attentional networks with progressive feature fusion of facial expression recognition (FER) under unlimited conditions using different feature extraction modules to add high-quality, brightly complex features of the combination. To blend these features effectively, the hierarchical attention module gradually enhances discrimination from distinct facial regions. It suppresses irrelevant regions. This approach aims to overcome pose variations i.e. occlusion-and light changes in FER. Extensive experiments show that the proposed model

performs best among existing FER methods for unconstrained conditions.[23] The study presents a dual subspace multiple learning method based on a Graph Convolutional Network (GCN) for intensity-invariant facial expression recognition (FER). This addresses FER as a node classification problem. Learn multiple representations using locality-preserving projection (LPP) and additional vertex pilot LPP (PLPP) to enhance the stability of smooth descriptions. Two subspace fusion methods combine LPP and PLPP - one using weighted adjacency matrices function and the other using self-attention - to improve performance further. The proposed focused fusion methods achieve state-of-the-art accuracy on the CK+, Oulu-CASIA, and MMI datasets, especially for simple facial expression recognition.[24] A recent survey in the domain has shown up which targeted several parameters like success rate, latency and model size, key parameters for practical deployment in FER. The authors reviewed the best models with respect to latency and RMSE values achieved in their evaluation with that of MobileNetV3Large MINI, MobileNetV2, EfficientNetB0 and MnasNet. The authors also concluded that the best model depends on the specific needs of the application.[25] The study proposes a real-time emotion identification system specially targeting autistic children using their facial expressions and deep learning. A DCNN architecture with an autoencoder and pre-trained Xception model was deployed for achieving a high accuracy of 0.9523% in classifying emotions, demonstrating the system's potential for real-time emotion recognition and its potential benefits for medical experts and families of autistic children.[26]

METHODOLOGY

Facial expression recognition based on CV has emerged as another crucial application domain in numerous sectors including medical diagnosis and others. This study will use deep learning techniques to identify human emotions. Through using a massive dataset and constructing an efficient model structure, this study offers a comprehensive solution for emotion recognition. The designed methodology of the study is represented in Figure 1.

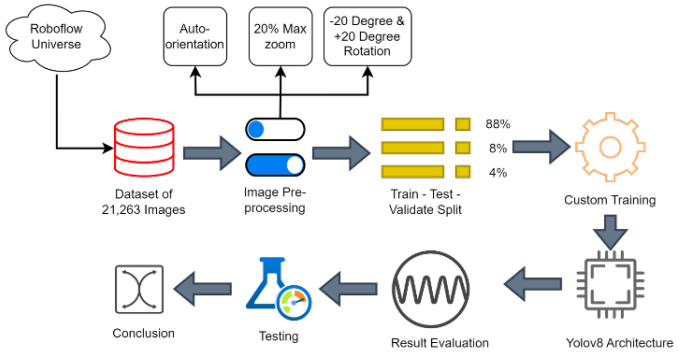


Figure 1: Designed Methodology

DATASET

The authors used 21,263 images collected from Roboflow.[27] These images are categorized into six distinct emotional classes: Happy, Sad, Angry, Disgust, Natural, and Surprise. The distribution of labels across these classes is as follows: 6,316 for Happy, 3,498 for Sad, 4,592 for Angry, 1,149 for Disgust, 4,124 for Natural, and 5,321 for Surprise. It should be noted that in some cases, several labels were assigned to one image, so the number of labels is greater than the number of images. This is because individual images may contain more than one facial expression, which results in multi-labeling. Namely, a single picture can be assigned more than one emotion if the picture depicts more than one emotion or a sequence of emotions. For instance, it may depict a person with a surprised and happy emotion.

Consequently, the label count for each emotion adds up to more than the total number of images, resulting in a higher aggregate label count despite the actual image count being 21,263. The dataset was split into three subsets to ensure robust model training and evaluation: training, validation, and testing. Specifically, 18,744 images (88%) were allocated for training, 1,696 images (8%) for validation, and 823 images (4%) for testing. These are split differently from the traditional 70-30 ratio to enhance the model's efficiency. This stratified split ensures the model has sufficient data to learn from while providing ample validation and testing samples to assess its performance and generalize to new, unseen data. Figure 2 represents the sample images used for the experiment.

Figure 2: Dataset Sample Images

PREPROCESSING

The preprocessing steps in preparing the image data for the emotion detection project using YOLOv8 were of utmost importance to increase the model's performance. These steps ensured that the images were in a consistent format and augmented to improve the model's robustness.[28] The key preprocessing steps included auto-orientation, zoom, rotation, and resizing. First, auto-orientation was applied to the images. Auto-orientation is a technique that corrects the orientation of an image to create diversity in data. Secondly, a 20% maximum zoom was applied to all images. By applying these zoom variations, the intention was to stimulate the image's effect from various distances, helping it perform better on real-time data. Rotation between -20 and +20 degrees was another step applied to the images.[29] This was done to ensure the model learns to recognize emotional expressions even when the faces are not well aligned. Images were also resized to 416x416 pixels during training. Together, these preprocessing steps—auto-orientation, zoom, rotation, and resizing—helped standardize the input data and augment it in ways that improved the robustness and accuracy of the YOLOv8 model for emotion detection.[30]

TOOLS AND HARDWARE USED

T4 GPU significantly accelerated the emotion detection project from Kaggle. It involved training the model on a large dataset of images to identify various emotional expressions, and It was computationally expensive due to its complexity.[31] The process was slow without a GPU, taking approximately 1.5 hours to complete one epoch. Switching to the T4 GPU reduced the training time significantly, allowing quicker iterations, experimentation, and model refinement. The purpose of using Kaggle's T4 GPU was influenced by the limitations and performance issues encountered with Colab's TPU/GPU offerings. Kaggle's platform provided a more efficient and stable environment, allowing 20 epochs to be completed within an hour. The authors used Python languages in this experiment.

YOLOV8 ARCHITECTURE

The "You Only Look Once" (YOLO)[32] series has come a long way from its first version, and the latest version, YOLOv8, The last modification of the original YOLO, called YOLOv8, gives some improvements, particularly for real-time object detection and identification. YOLOv8 is an advanced version of the previous YOLO models, which implements better feature extraction, better localization, and less computational power. Many techniques like anchor-free detection, dynamic head modules, and multi-scale feature aggregation enhance the precision and Recall of this model you've mentioned. By its nature, YOLOv8 can work on auto, surveillance, and medical imaging applications, among other fields, since it is optimized to be reliable and fast. Regarding speed and accuracy, it tops the list of other models that can be used in real-time object detection. Yolov8 architecture is shown in Figure 3.

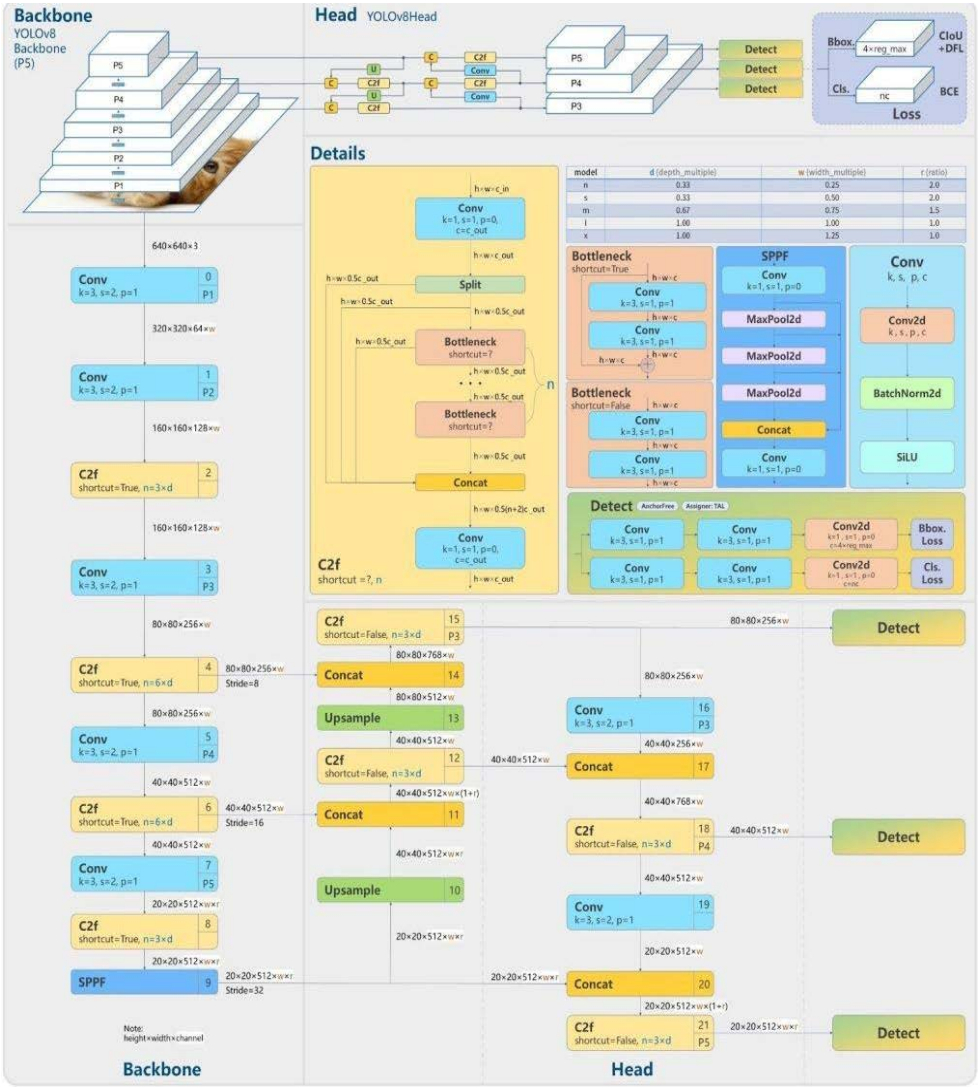


Figure 3: Yolov8 Architecture. Reproduced from Ref.[36] with permission.

PERFORMANCE EVALUATION

The training results of our emotion detection model using YOLOv8 indicate strong performance across several emotions, especially given that the model was trained for only five epochs. Here's a detailed summary, and the graphical results are represented in Figure 4:

EVALUATION METRICS USED

Evaluation is a critical step in any project, especially in computer vision-based tasks, as it directly impacts the effectiveness of computer vision-based projects. The results of the emotion detection project are evaluated using metrics like precision, Recall, and mAP (mean Average Precision).

Precision: The accuracy of the model's positive predictions is measured by precision. Precision measures how many anticipated emotions are correctly identified in emotion detection. With a high precision score, the model can make fewer predictions that are not accurate.[33]

$$Precision = \frac{True\ Positive}{True\ Positives + False\ Positives}$$

Recall: Recall is an evaluation metric popularly used for classification problem statements. It measures the ability of the model to identify all relevant instances of emotions. It is also known as sensitivity.[34]

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

High Recall means the model performs well for specific classes, while low recall means incorrectly classifying.

mAP (mean Average Precision): mAP is commonly used in object detection, specifically dealing with classification problems. It calculates the average precision across all classes (emotions) and provides a single measure of overall performance.[35]

It gives a detailed overview of the model's performance across all emotions.

$$mAP = \frac{1}{N}\sum_{i=1}^{N}(AP_i)$$

F1 Score: When working with imbalanced datasets, the F1 score is a useful indicator for evaluating the effectiveness of classification models. It offers a metric that harmonizes precision and recall, creating a harmonic mean.

$$F1\ Score = 2\ X\ \frac{Precision\ X\ Recall}{Precision + Recall}$$

Training Results Analysis

The current detection model has some interesting findings to share in emotions. Among all, the class that triggers the strongest reaction of disgust boasts the highest precision, Recall, and mean average precision (MAP) scores. It's exact at 0.99, with a recall of 0.96 and a remarkable MAP score of 0.99, represented in Table 1.

Table 1: Training Test Analysis

Class	Images	Instances	Box	R	mAP50
all	1696	4626	.817	.796	.87
Angry	1696	885	.764	.779	.849
Disgust	1696	233	.997	.966	.99
Happy	1696	1129	.871	.877	.941
Natural	1696	749	.686	.644	.728
Sad	1696	657	.709	.659	.773
Surprise	1696	973	.877	.855	.937

The happy class follows closely, with a precision and Recall at 0.87 and a MAP score of 0.94. Surprisingly, the surprise class also shows significant performance, with a precision of 0.87, a Recall of 0.85, and a MAP score of 0.893.

However, when it comes to anger, though recognizable, it's a bit more elusive. The precision is 0.76, the Recall is 0.78, and the MAP score is 0.84. The sad class follows with a precision of 0.70, Recall of 0.65, and MAP score of 0.77.

Lastly, there's the natural class, with a precision of 0.68, Recall at 0.64, and a MAP score of 0.72. Our model performs well with an overall precision of 0.81, a Recall of 0.8, and a MAP score of 0.87.

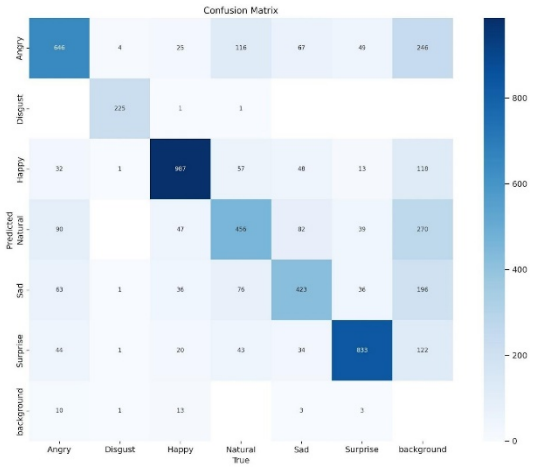


Figure 4: Confusion Matrix

The confusion matrix shown in Figure 4 illustrates the performance of our classification model. The matrix displays the total number of true positive predictions for each class. The intensity of the blue shading in each cell represents the accuracy of the model's predictions: darker blue cells indicate a higher number of correct, true positive predictions, whereas lighter blue cells suggest fewer correct predictions. This color gradient provides a clear visual indication of the model's strengths and weaknesses across different classes.

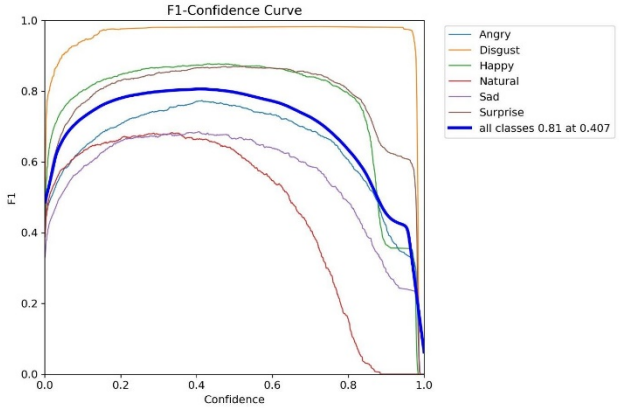


Figure 5: Confidence vs F1 Score

The curve in Figure 5 represents the F1 score with multiple classes plotted against the confidence interval on the x-axis, illustrating how the model's performance varies with prediction confidence across different categories. This helps assess the model's reliability, identify optimal confidence thresholds, and highlight class-specific performance issues. A steep drop in the F1 score at higher confidence intervals suggests the model is less reliable with lower confidence predictions, whereas a flatter curve indicates consistent performance. This visualization aids in fine-tuning and understanding the robustness of the model. It elucidates that all classes' overall F1 score is maximum at 0.40 confidence level.

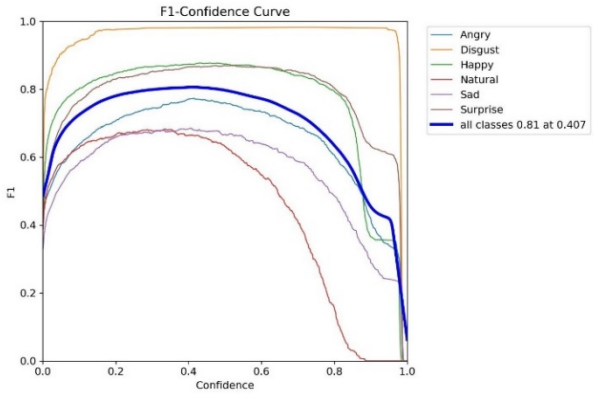


Figure 6: Recall vs Precision

The curve in Figure 6 shows the optimal balance between recall (the fraction of captured genuine positives) and precision (the proportion of true positives among positive predictions) at various classification thresholds. The PR curve peaks at 0.8 at 0.5 mAP when all classes are considered together.

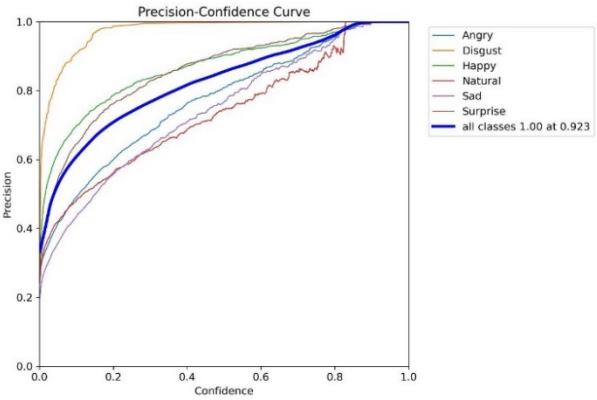


Figure 7: Confidence vs Precision

The curve in Figure 7 is of Precision with multiple classes plotted against the confidence interval on the x-axis, illustrating how the model's performance varies with prediction confidence across different categories. It seems from the plot that the overall highest Precision across all classes is a perfect one at the Confidence Threshold of 0.92.

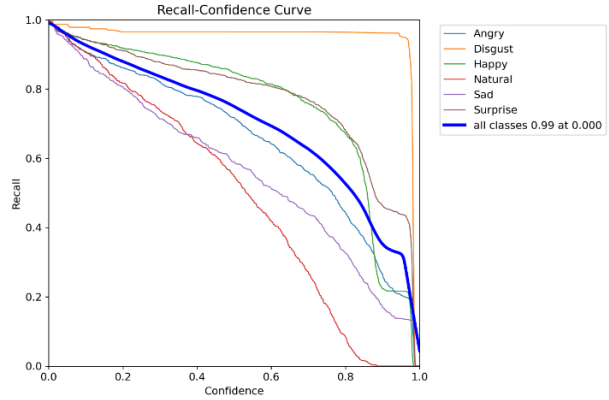


Figure 8: Confidence vs Recall

The curve in Figure 8 is of Recall with multiple classes plotted against the confidence interval on the x-axis, illustrating how the model's performance varies with prediction confidence across different categories. The plot shows the overall highest Recall score across all classes, which is 0.99, with a Confidence Threshold of 0.01.

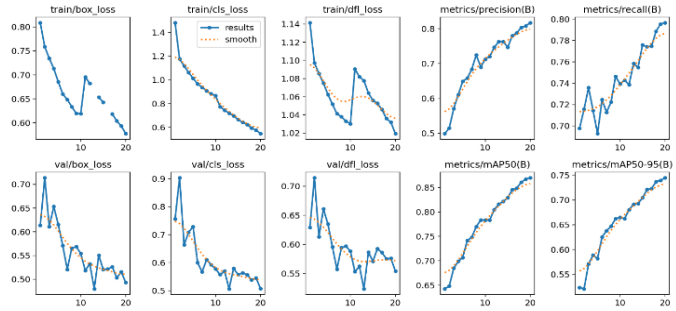


Figure 9: Relationship Between Loss and Mean Average Precision (mAP)

The pairplot in Figure 9 depicts the relationship between loss and Mean Average Precision (mAP) scores across 20 epochs. It reveals a clear trend: the loss decreases as the number of epochs increases while the mAP score successively improves. This visualization underscores the model's learning progress over time, indicating that with each epoch, the model becomes more adept at minimizing loss and improving its performance in terms of mAP score.

VALIDATION

After training, validation was conducted using the `best. The patient's weights on the validation data yielded good results, demonstrating the model's effectiveness in detecting emotions accurately. Post-training validation results are represented in Figure 10.

Figure 10: Post-Training Validation Results

The model exhibits consistent performance between the training and validation sets, with nearly identical results. This parity indicates that the model generalizes to unseen data, performing equally well on the validation and training sets. This alignment between training and validation performance suggests that the model has learned from the training data and can accurately predict outcomes on new, unseen data, demonstrating its robustness and reliability.

CUSTOM TESTING

Following the training phase, the model's performance was verified through custom testing, which demonstrated accurate emotion detection, as evidenced by the results shown in the figures below:

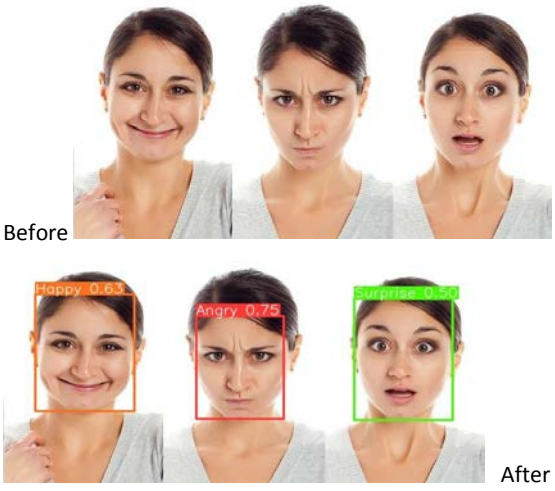


Figure 11: Before and After Custom Testing Results

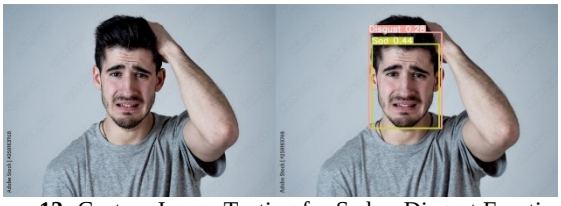


Figure 12: Custom Image Testing for Sad or Disgust Emotion

As seen in Figure 11, the model successfully predicts three classes even on custom images: "happy" with a confidence of 0.63, "angry" with 0.75, and "surprise" with 0.50 confidence. However, in Figure 12, there's an interesting observation where classes "disgust" and "sad" overlap, both predicted with confidence scores of 0.28 and 0.44, respectively. This overlapping suggests that features typical to both classes might have led to confusion in prediction, indicating a potential ambiguity in the dataset or model's decision boundaries.

CONCLUSION AND FUTURE SCOPE

In conclusion, this study successfully demonstrates the effectiveness of the YOLOv8 model in facial emotion recognition, achieving an overall precision of 0.81, recall of 0.8, and mAP score of 0.87. The model's ability to accurately detect and classify emotions in real time, even with limited training epochs, highlights its potential for various applications. The model excelled in recognizing disgust (precision: 0.99, recall: 0.96, mAP: 0.99) and happiness (precision: 0.87, recall: 0.87, mAP: 0.94). Using preprocessing techniques and a robust dataset further enhanced the model's performance, as evidenced by consistent results in training and validation sets. The findings of this study contribute to the growing field of affective computing and pave the way for future research in real-time emotion detection and its integration into diverse domains, such as human-computer interaction and mental health assessment. The limitation of this study is that dataset images are limited which and more emotions can be considered. In future authors will take large dataset with more emotion types and can apply another deep learning techniques to enhance the results. Future researchers can work on these limitation of this study which will advance the emotion detection of humans.

CONFLICT OF INTEREST STATEMENT

Authors declare that there is no conflict of interest for publication of this work.

REFERENCES

1. M.K. Chowdary, T.N. Nguyen, D.J. Hemanth. Deep learning-based facial emotion recognition for human--computer interaction applications. *Neural Comput. Appl.* **2023**, 35 (32), 23311–23328.
2. N. Singh, H. Sabrol. Convolutional neural networks-an extensive arena of deep learning. A comprehensive study. *Arch. Comput. Methods Eng.* **2021**, 28 (7), 4755–4780.
3. A.W. Salehi, S. Khan, G. Gupta, et al. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability* **2023**, 15 (7), 5930.
4. A. Oluwasammi, M.U. Aftab, Z. Qin, et al. Features to text: A comprehensive survey of deep learning on semantic segmentation and image captioning. *Complexity* **2021**, 2021 (1), 5538927.
5. R. Hammouche, A. Attia, S. Akhrouf, Z. Akhtar. Gabor filter bank with deep autoencoder based face recognition system. *Expert Systems with Applications*. Elsevier 2022, p 116743.
6. V. Pagire, A.C. Phadke, J. Hemant. A deep learning approach for underwater fish detection. *J. Integr. Sci. Technol.* **2024**, 12 (3), 765.
7. O.S. Ekundayo, S. Viriri. Facial Expression Recognition: A Review of Trends and Techniques. *IEEE Access*. IEEE 2021, pp 136944–136973.
8. R. Hosney, F.M. Talaat, E.M. El-Gendy, M.M. Saafan. AutYOLO-ATT: an attention-based YOLOv8 algorithm for early autism diagnosis through facial expression recognition. *Neural Comput. Appl.* **2024**, 36 (27), 17199–17219.
9. G. Wang, Y. Chen, P. An, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, 23 (16), 7190.
10. H. Ma, S. Lei, T. Celik, H.-C. Li. FER-YOLO-Mamba: Facial Expression Detection and Classification Based on Selective State Space. *arXiv Prepr. arXiv2405.01828* **2024**.
11. Y. Shuai, Z. Lin, W. Chen, W. Shenghuai, T. Yu. SF-YOLO: An Evolutionary Deep Neural Network for Gear End Surface Defect Detection. *IEEE Sens. J.* **2024**, 24 (13), 21762–21775.
12. P. Su, H. Han, M. Liu, T. Yang, S. Liu. MOD-YOLO: Rethinking the YOLO architecture at the level of feature information and applying it to crack detection. *Expert Syst. Appl.* **2024**, 237 (8), 121346.
13. N. Anjeana, K. Anusudha. Real time face recognition system based on YOLO and InsightFace. *Multimed. Tools Appl.* **2023**, 83 (11), 31893–31910.
14. Z. Yu, H. Huang, W. Chen, et al. Yolo-facev2: A scale and occlusion aware face detector. *Pattern Recognit.* **2024**, 110714.
15. F. Pashayev, L. Babayeva, Z. Isgandarova, B.K. Kalejahi. Face Recognition in Smart Cameras by Yolo8. *KHAZAR J. Sci. Technol.* **2023**, 67.

16. M.J.A. Daasan, M.H.I. Bin Ishak. Enhancing Face Recognition Accuracy through Integration of YOLO v8 and Deep Learning: A Custom Recognition Model Approach. In *Asia Simulation Conference*; **2023**; pp 242–253.
17. D. Al-Obidi, S. Kacmaz. Facial Features Recognition Based on Their Shape and Color Using YOLOv8. In *7th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2023 - Proceedings*; **2023**; pp 1–6.
18. B. Ma, Z. Fu, S. Rakheja, et al. Distracted Driving Behavior and Driver's Emotion Detection Based on Improved YOLOv8 With Attention Mechanism. *IEEE Access* **2024**, 12, 37983–37994.
19. N.S. Vemulapalli, P. Paladugula, G.S. Prabhat, S. Abhishek, T. Anjali. Face Detection with Landmark using YOLOv8. In *2023 3rd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*; **2023**; pp 1–5.
20. C. Dewi, D. Manongga, E. Mailoa, others. Deep Learning-Based Face Mask Recognition System with YOLOv8. In *2024 16th International Conference on Computer and Automation Engineering (ICCAE)*; **2024**; pp 418–422.
21. S. Karaku\cs, M. Kaya, S.A. Tuncer. Real-Time Detection and Identification of Suspects in Forensic Imagery Using Advanced YOLOv8 Object Recognition Models. *Trait. du Signal* **2023**, 40 (5).
22. W. Gong, Y. Qian, W. Zhou, H. Leng. Enhanced spatial-temporal learning network for dynamic facial expression recognition. *Biomed. Signal Process. Control* **2024**, 88, 105316.
23. H. Tao, Q. Duan. Hierarchical attention network with progressive feature fusion for facial expression recognition. *Neural Networks* **2024**, 170, 337–348.
24. J. Chen, J. Shi, R. Xu. Dual subspace manifold learning based on GCN for intensity-invariant facial expression recognition. *Pattern Recognit.* **2024**, 148, 110157.
25. M. Krumnikl, V. Maiwald. Facial Emotion Recognition for Mobile Devices: A Practical Review. *IEEE Access* **2024**, 12, 15735–15747.
26. F.M. Talaat, Z.H. Ali, R.R. Mostafa, N. El-Rashidy. Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. *Soft Comput.* **2024**, 1–14.
27. Roboflow. Face Emotion Dataset.
28. V. Nair, M. Kanojia. Integrating Facial Emotion Recognition into Music Recommendation Systems using YOLOv8. *J. Inf. Assur. \& Secur.* **2023**, 18 (6).
29. T. Ward. Development of detection and tracking systems for autonomous vehicles using machine learning; Morehead State University, **2023**.
30. F. Pérez-Garc\'\ia, R. Sparks, S. Ourselin. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput. Methods Programs Biomed.* **2021**, 208, 106236.
31. A. Alshahrani, M.M. Almatrafi, J.I. Mustafa, L.S. Albaqami, R.A. Aljabri. A Children's Psychological and Mental Health Detection Model by Drawing Analysis based on Computer Vision and Deep Learning. *Eng. Technol. Appl. Sci. Res.* **2024**, 14 (4), 15533–15540.
32. Y. Pratama, E. Rasywir, A. Sunoto, I. Irawan. Application of YOLO (You Only Look Once) V.4 with Preprocessing Image and Network Experiment. *IJICS (International J. Informatics Comput. Sci.* **2021**, 5 (3), 280.
33. C. Sharma, S. Shambhu, P. Das, S. Jain, Sakshi. Features contributing towards heart disease prediction using machine learning. In *CEUR Workshop Proceedings*; **2021**; Vol. 2823, pp 84–92.
34. P. Das, S. Jain, C. Sharma, S. Shambhu. Prediction of Heart Disease Mortality Rate Using Data Mining; **2021**.
35. J. Ding, N. Xue, G.-S. Xia, et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 44 (11), 7778–7796.
36. M. Lalinia, A. Sahafi. Colorectal polyp detection in colonoscopy images using YOLO-V8 network. *Signal, Image Video Process.* **2024**, 18 (3), 2047–2058.