

Real Time Human Facial Emotion Recognition Using YOLO With Confidence and IoU Threshold Optimization

Group1: Rahul Saini, Thuvarakan Mathipalan, Koti Reddy Kommireddy, Sri Sowharan Palanisamy Krishnamurthy

1. Introduction

Real time facial emotion recognition is a key part in modern intelligent systems, supporting applications in human computer interaction, security monitoring, mental health analytics, and smart environments[1]. Deep learning methods particularly YOLO (You Only Look Once) have shown exceptional speed and accuracy in face detection tasks, making them suitable for real time applications. This project aims to develop Human facial emotion recognition using YOLO [2] [3].

To assess the performance of the real time facial emotion recognition system, this project employs Intersection over Union (IoU) and confidence score as key evaluation metrics. These metrics help measure how accurately the model detects faces and how confidently it predicts emotions, giving a clear and reliable way to evaluate its performance in real time use [4].

2. Objectives

- To detect human faces and classify basic emotions in real time using YOLO
- To optimize YOLO performance by tuning confidence and IoU thresholds for improved detection stability and accuracy
- To demonstrate the complete system through a real time webcam based facial emotion recognition

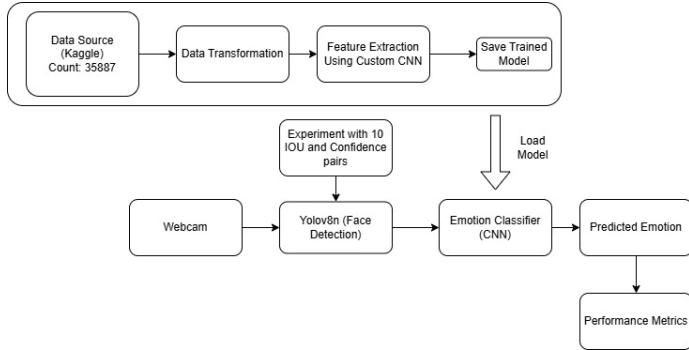
3. Literature Review

Recent research highlights the growing importance of facial emotion recognition in intelligent systems. The study in [1] emphasizes the need for reliable facial expression recognition frameworks to support real world applications.

With advancements in object detection architectures, YOLO based models have become widely adopted for real time tasks. Alshammari [2] demonstrate that YOLOv8 significantly improves detection speed and accuracy for emotion recognition

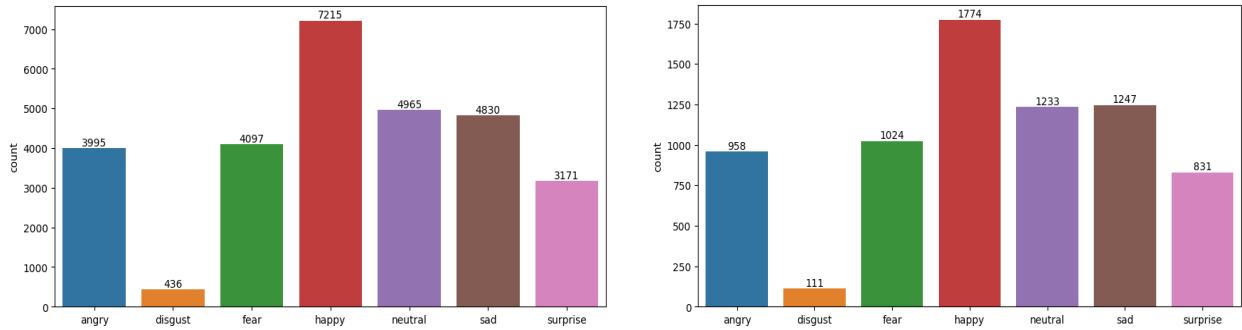
Furthermore, recent YOLO based study done by Alshammari [2] and Parambil et al. [3] indicate that fine tuning IoU thresholds, confidence values, and model hyperparameters significantly enhances detection quality, particularly in real time emotion recognition applications.

4. Methodology



4.1 Data Source

This project uses the FER-2013 (Facial Expression Recognition 2013) dataset. This dataset holds 35887 grayscale images (48×48 pixels), categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise [5]. Data split ratio is 80:20 for training (28709) and testing (7178). That distribution is shown in below diagram.



4.2 Data Transformation

The images were converted to grayscale, resized to 48×48 , transformed into tensors and normalized. These steps ensured uniform input quality and consistent preprocessing for the CNN model.

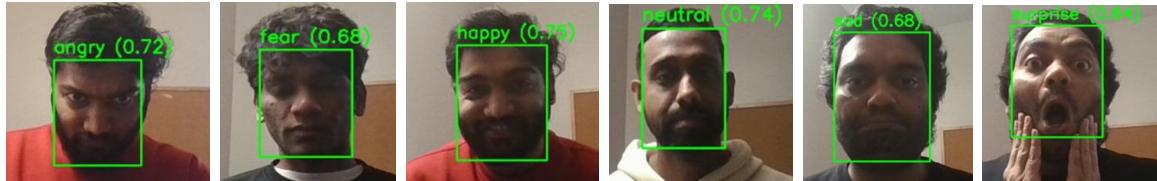
4.3 CNN Architecture

Layer (type:depth-idx)	Output Shape	Param #
Sequential	[1, 7]	--
—Conv2d: 1-1	[1, 32, 48, 48]	320
—ReLU: 1-2	[1, 32, 48, 48]	--
—MaxPool2d: 1-3	[1, 32, 24, 24]	--
—Conv2d: 1-4	[1, 64, 24, 24]	18,496
—ReLU: 1-5	[1, 64, 24, 24]	--
—MaxPool2d: 1-6	[1, 64, 12, 12]	--
—Conv2d: 1-7	[1, 128, 12, 12]	73,856
—ReLU: 1-8	[1, 128, 12, 12]	--
—MaxPool2d: 1-9	[1, 128, 6, 6]	--
—Flatten: 1-10	[1, 4608]	--
—Linear: 1-11	[1, 256]	1,179,904
—ReLU: 1-12	[1, 256]	--
—Dropout: 1-13	[1, 256]	--
—Linear: 1-14	[1, 7]	1,799
Total params:	1,274,375	
Trainable params:	1,274,375	
Non-trainable params:	0	
Total mult-adds (Units.MEGABYTES):	23.21	

4.4 Yolov8n

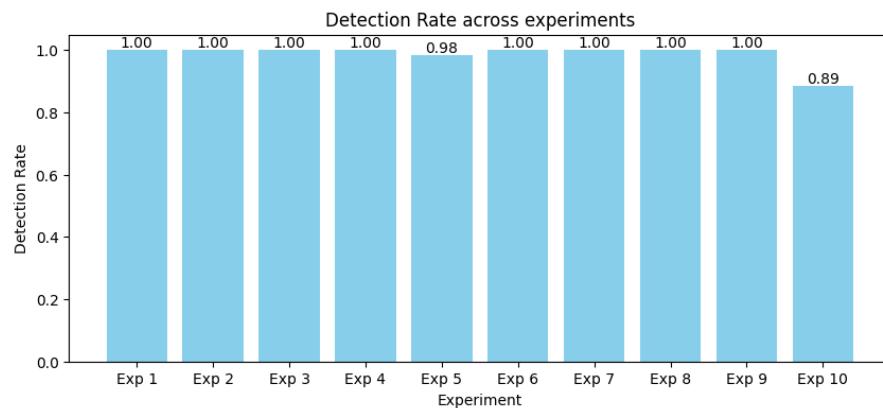
In this project, the YOLOv8n model is used for fast and accurate face detection in real time webcam input. YOLOv8n is the lightweight version of the YOLOv8 family, offering high speed and low computational cost while maintaining strong detection performance, making it suitable for real time facial emotion recognition [6].

5. Experiments

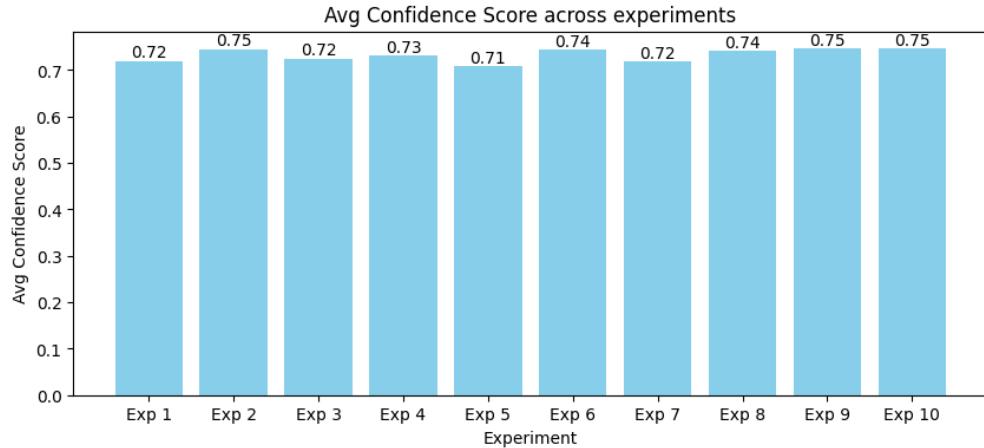


Ten different combinations of IoU and confidence thresholds were experimented.

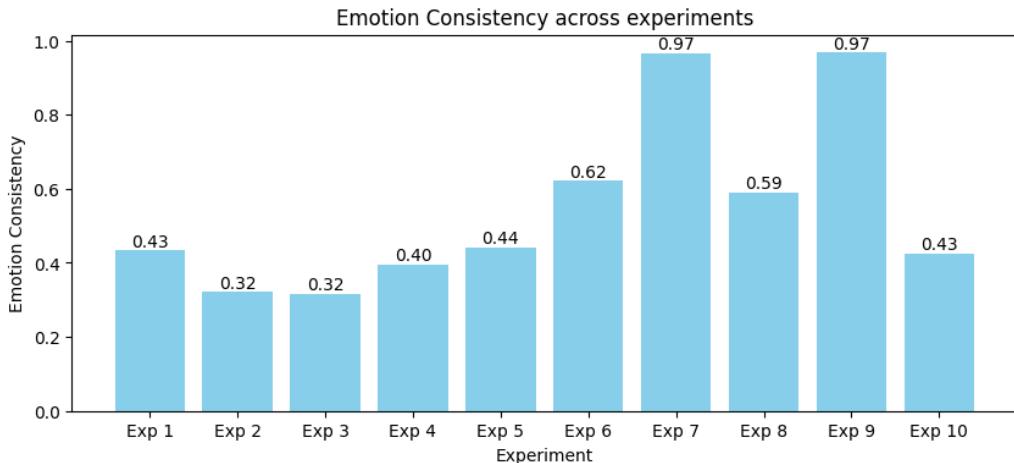
	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
Confidence	0.2	0.3	0.4	0.5	0.2	0.3	0.4	0.5	0.6	0.7
IoU	0.3	0.3	0.3	0.3	0.5	0.5	0.5	0.5	0.4	0.6



Across the ten IoU-confidence threshold combinations, the model demonstrated consistently strong detection performance, achieving a perfect detection rate in eight experiments and only a slight drop in two, confirming the reliability of YOLOv8n for face detection.



The average confidence score also remained stable between 0.71 and 0.75, indicating that the model maintained reliable prediction.



The emotion consistency metric showed greater variability, ranging from 0.32 to 0.97, with higher stability seen in experiments using more balanced threshold pairs.

6. Conclusion and Future Work

In summary, the experiments demonstrate that the face detection across varying confidence and IoU thresholds, with detection rates consistently near 1.0 and stable confidence scores around 0.72 - 0.75. Emotion consistency, however, showed greater sensitivity to parameter changes, with mid range thresholds (Confidence 0.4 - 0.6, IoU 0.5) showing the most reliable and stable emotion predictions.

In future work, improving generalization by expanding the dataset can be done and exploring more advanced architectures such as Vision Transformers or hybrid CNN Transformer models for enhanced emotion recognition [7].

References

- [1] M. Tanweer, M. Tanveer, A. Mayo, Q. Ain, and J. Ahmed, "Real-Time Intelligent Facial Expression Recognition System," *J. Independ. Stud. Res. Comput.*, vol. 20, Dec. 2022, doi: 10.31645/JISRC.22.20.2.4.
- [2] A. Alshammari and M. E. Alshammari, "Emotional Facial Expression Detection using YOLOv8," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 5, pp. 16619–16623, Oct. 2024, doi: 10.48084/etasr.8433.
- [3] M. M. A. Parambil *et al.*, "Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition," *IEEE Access*, vol. 12, pp. 109427–109442, 2024, doi: 10.1109/ACCESS.2024.3439346.
- [4] S. Arya, "Measuring Success: Comprehensive Evaluation Metrics for Object Detection Models," *Analytics Vidhya*. Accessed: Dec. 08, 2025. [Online]. Available: <https://www.analyticsvidhya.com/blog/2024/02/evaluation-matrix-for-object-detection-using-iou-and-map/>
- [5] "FER-2013." Accessed: Dec. 08, 2025. [Online]. Available: <https://www.kaggle.com/datasets/msambare/fer2013>
- [6] Ultralytics, "Explore Ultralytics YOLOv8." Accessed: Dec. 09, 2025. [Online]. Available: <https://docs.ultralytics.com/models/yolov8/>
- [7] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," June 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.