Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 2 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

Enhanced Crowd Emotion Detection in Occluded Environments Using YOLOv5 and Convolutional Neural Networks

Tahmeena Mai
Department of computer Science, Institute of Southern Punjab Multan, Pakistan
Nazia Batool
Department of Information Technology Government Technical Training Institute, DG. Khan, Pakistan
Mehreen Fatima
Department of computer Science, Institute of Southern Punjab Multan, Pakistan

Corresponding author: Tahmeena Mai (tehminakhan2222@gmail.com)
DOI: *https://doi.org/10.71146/kjmr285*

Article Info

Abstract

A deep learning methodology examines facial emotional identification when faces exist in partially obstructed conditions of crowds. YOLOv5-tiny functions for face detection while RepVGG handles emotion classification within this method. Training of the emotion recognition model employed FER-2013 dataset but detection enhancements came from processing WIDER FACE dataset. Experimental testing shows that YOLOv5 reaches a 91% detection success rate while RepVGG delivers a 87.5% accuracy level in emotion classification above standard CNN architectures such as ResNet and EfficientNet. The system detects emotions directly in real-time streaming video data which qualifies it for application in crowd monitoring and surveillance and behavioral analysis solutions. Future research initiatives aim to develop better methods for handling obscuration along with methods to optimize speed when applying the system to large-scale deployments.

Keywords:
Emotion Recognition, Crowd Analysis, Deep Learning, Real-time Emotion Detection.

Introduction

Expressions on the face function as the leading non-verbal manner by which people communicate both emotional states and psychological conditions and underlying intentions. Human facial expression understanding allows crucial advancements across multiple fields such as human-computer interaction (HCI), surveillance, marketing analytics, behavioral studies and healthcare operations. FER systems run by machines become essential tools for achieving precise emotion interpretation which directly enhances communication between people and intelligent systems. The capability of emotion recognition techniques enhances automated customer service systems and AI-driven virtual assistants and smart surveillance networks. GER systems demonstrate potential value in the diagnosis of emotional disorders when used to measure emotional well-being and stress levels and mood changes during psychological assessments.

Surveillance methods based on facial emotional recognition help unravel abnormal and assailant behaviors in public areas which supports law enforcement authorities to recognize dangers before they occur. Marketing organizations use FER capabilities to conduct consumer sentiment evaluations which helps them assess customer responses to advertising elements and product display areas and retail environments. The wide range of capabilities offered by FER applications proves difficult to implement in real situations because of environmental elements such as blocked views and shifting illumination alongside different body postures and emergency processing requirements.

Deep learning-based FER models demonstrate limited performance effectiveness when faces become partially hidden by wearing masks or through combination with glasses and when obstructed by other individuals or physical objects. CNN-based models such as ResNet VGG and EfficientNet show limited success in obtaining important features when major parts of the facial regions become hidden resulting in incorrect emotional classification or non-detection of faces. The current methods depend mostly on controlled environments for frontal face image collection because this restricts their effectiveness in crowded chaotic situations.

FER encounters crucial hurdles because of lighting variation coupled with changes in human facial postures. The visibility of facial characteristics which emotion classification algorithms utilize gets diminished substantially by outdoor exposure and low-light environments and shadowed conditions. Frontal views remain the preferred condition for capturing faces when working with deep learning models because their training systems focus on frontal datasets. Traditional detection systems based on Haar cascades and Multi-Task Cascaded Convolutional Networks (MTCNN) show poor performance when detecting faces under heavy obstacles or radical posture conditions.

Real-time processing requires immediate resolution to successfully implement FER applications particularly during crowd monitoring and surveillance tasks. High-performing models that include Vision Transformers (ViTs) and deep CNNs lack capability to work properly on embedded devices and real-time applications because they need large computational resources. Real-world FER deployment demands accurate solutions that also maintain efficient computational capabilities.

The offered solution consists of YOLOv5-tiny for staff time-efficient face discovery operations while utilizing RepVGG for cost-effective emotions detection. The system design employs real-time emotion detection functionality that functions effectively in busy situations despite partial obstructed face views.

The initial stage of the system implements YOLOv5-tiny because this lightweight detection model provides real-time performance for facial recognition. Through its grid-based anchor mechanism YOLOv5 detects small-scale as well as partially obscured faces with high precision despite traditional face detection models being unable to achieve this level of accuracy. YOLOv5-tiny uses training from the

WIDER FACE dataset to reach 91% accuracy in detecting faces which enables it to function well for real-world surveillance applications.

The second phase of the system leverages RepVGG, a computationally efficient deep CNN architecture, for emotion classification. In contrast to the deeper models such as ResNet and EfficientNet, RepVGG, however, features 3×3 convolutional layers with ReLU activation so that computational overhead does not increase too much and yet still yields high accuracy. RepVGG is trained on FER-2013 dataset and reaches 87.5% accuracy on emotion classification compared to conventional CNN based models. YOLOv5 integration and RepVGG allows for real time, robust emotion detection at high speed and generalizability to diverse environments.

The main contributions of this study are in the field of real time facial emotion recognition (FER) where the existing challenges are posed effectively tackled using a hybrid deep learning framework. YOLOv5-tiny is integrated to robust face detection and RepVGG for the efficient emotion classification, making the system can achieve high detection accuracy and high computational efficiency. Compared to CNN based models that typically suffer from real time inference in complex environments, the proposed method achieves good speed and accuracy, and hence it is well suited for large scale deployment in crowd monitoring, surveillance, behavioral analysis.

The main contribution of the work is the ability to conduct emotion analysis in real time on running applications in dynamic environments. The system deals in real time video streams, able to enable real-time emotion classification of applications like public security enclosing, intelligent commercials and shopping customer sentiment analysis. Using YOLOv5's sophisticated face detection functionality, the model can efficiently identify partially occluded and low-resolution faces in crowded scenes, and since RepVGG is an extremely lightweight model, the model's performance is high without sacrificing accuracy.

Moreover, this research also undertakes a comprehensive comparative performance assessment on the highly competitive, high-performing, deep learning models like ResNet, EfficientNet, as well as the traditional VGG-based architectures. In experimental results show that proposed model got the better detection accuracy and inference speed comparing to the existing method at both fully visible face and occluded face. The model's robustness is particularly manifest in pose variations, facial occlusions (e.g., masks, glasses), as well as in lighting conditions, overcoming significant challenges faced by previous solutions.

Another key point of difference of this research is that it is more robust against occlusions and non-frontal facial poses. Conventional emotion recognition models fail to achieve good performance when the face recognitions are misleading or their feature elements are partially lost. But the adopted YOLOv5-RepVGG integration succeed in the facial structure detection in visible region to classifying emotion accurately, when there are faces partially hidden. This feature allows the system to be very suitable for real-world applications in line with facial occlusion of appearing for instance smart surveillance systems, intelligent security monitoring, and interactive AI drivensystem.

This study builds the foundation for future proofs in real-time FER for large-scale and dynamic environment through scaling up multiple, efficient, and performance-enhanced deep learning framework. The proposed model not only achieves higher accuracy of emotion recognition in challenging case, but also improves practical applicability of AI-based FER approaches in practical applications scenarios.

Literature Review

Emotion recognition in crowds is a critical area of research in computer vision and affective computing, with applications in public safety, surveillance, human-computer interaction, and psychological analysis. Traditional emotion recognition techniques primarily rely on facial expressions, body gestures, and physiological signals. However, occluded environments pose significant challenges to accurate emotion detection due to partial facial visibility, variations in lighting, and crowd density. The advent of deep learning, particularly Convolutional Neural Networks (CNNs) and object detection models such as You Only Look Once (YOLO), has significantly improved emotion recognition in complex settings.

The literature review highlights the increasing popularity of Convolutional Neural Networks (CNNs) for data improvement and transfer learning, demonstrating their usefulness in facial recognition technology despite moral dilemmas and their adaptability to various situations (Bagane et al., 2023). This research reviews the use of CNNs for facial feature recognition, highlighting Goodfellow et al.'s work on emotion monitoring using a deep CNN on the FER2013 dataset, demonstrating their effectiveness in large dataset analysis (Ramis et al., 2022).

Early emotion detection methods were based on handcrafted feature extraction techniques such as Local Binary Patterns (LBP) (Shan et al., 2009) and Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005). These approaches, while effective in controlled settings, struggled with robustness in real-world scenarios due to sensitivity to noise, occlusions, and illumination changes. The emergence of deep learning transformed the field, with CNN-based models such as AlexNet, VGG16, and ResNet achieving superior performance by learning hierarchical features automatically (LeCun et al., 2015; Krizhevsky et al., 2012).

Despite these advancements, occlusions remain a significant challenge. Various strategies have been proposed to address occluded emotion recognition. Zhao et al. (2018) explored multi-modal approaches using depth images and thermal imaging to enhance recognition. However, these methods require specialized sensors, limiting their practicality for large-scale deployment. To improve occlusion robustness, attention mechanisms have been integrated into CNNs, allowing models to focus on unoccluded facial regions (Li et al., 2017). Additionally, Generative Adversarial Networks (GANs) have been employed to reconstruct occluded facial regions, improving classification accuracy (Antipov et al., 2017).

The limitations of current face recognition techniques, particularly in the presence of masks and spectacles, have been highlighted by Eva (2021), who suggests focusing on locating obscured regions and obtaining features from visible parts. However, this work does not propose specific solutions. A study by Budiarsa et al. (2023) proposed using a modified FCN to segment obscured faces during facial recognition, focusing on non-occluded regions and investigating optimal configurations for precise recognition in challenging situations.

Body posture and motion analysis provide alternative cues for emotion recognition in occluded environments. Kleinsmith & Bianchi-Berthouze (2013) demonstrated that body gestures convey critical affective information, particularly when facial expressions are occluded. Pose estimation models such as OpenPose have been integrated with emotion recognition frameworks to capture non-facial emotional cues (Cao et al., 2019). However, these approaches require a clear view of body joints, which may not always be feasible in dense crowds.

YOLO-based models have gained popularity for real-time facial detection and emotion recognition. YOLOv5, the latest evolution of the YOLO series, offers improved accuracy and speed, making it ideal for real-time applications (Jocher et al., 2020). Several studies have leveraged YOLO for emotion

detection in crowds. Kumar et al. (2022) proposed a YOLO-based surveillance framework for facial emotion recognition, demonstrating its robustness in occluded settings. Abbas et al. (2021) utilized YOLOv4 for real-time crowd emotion analysis, highlighting its efficiency in detecting expressions in challenging conditions.

Khan et al. (2023) explored the use of deep learning for authorship verification in low-resource languages, employing hyper-tuned CNN models to enhance classification accuracy (Khan et al., 2023). This study demonstrates the effectiveness of CNN architectures in text classification tasks, providing insights into their adaptability for emotion detection in occluded environments. Additionally, Khan et al. (2024) conducted a comparative analysis of hybrid ensemble algorithms for authorship attribution in Urdu text, emphasizing the role of ensemble learning in improving classification performance (Khan et al., 2024).

Other studies have explored occlusion-aware facial feature learning. Wang et al. (2020) introduced a facial landmark-based occlusion recognition framework that improves classification robustness. Zhang et al. (2022) proposed an ensemble learning approach combining CNNs and capsule networks to handle occluded facial expressions more effectively. Additionally, Bhogad et al. (n.d.) explored an AI-based system for real-time face and emotion recognition, addressing population growth challenges and emphasizing the importance of facial expressions in communication.

Given the limitations of existing methods, our research aims to enhance crowd emotion detection in occluded environments by integrating YOLOv5 with a CNN-based classifier. By leveraging YOLOv5's object detection capabilities for accurate face localization and combining it with deep CNNs for robust emotion classification, we seek to improve recognition accuracy in real-world crowded scenarios. We also explore attention mechanisms and data augmentation techniques to mitigate the impact of occlusions and improve model generalization. Our work contributes to the growing field of affective computing and has significant implications for public safety, human-computer interaction, and behavioral analysis.

Methodology

The proposed system follows a two-phase deep learning pipeline designed to accurately detect faces in crowded environments and classify emotions in real-time. This hybrid approach leverages YOLOv5-tiny for efficient face detection and RepVGG for fast and accurate emotion recognition. The integration of these models ensures a robust, real-time, and scalable framework capable of handling occlusions, pose variations, and low-quality facial images. This section provides a detailed breakdown of the system architecture, data preprocessing techniques, training strategies, and hyperparameter optimization, ensuring high performance in real-world scenarios such as surveillance, marketing analytics, and behavioral studies.

A. System Architecture

The architecture of the proposed system is structured into two key components:

1) Face Detection Using YOLOv5-Tiny

Face detection is a fundamental step in facial emotion recognition, especially in real-world applications where faces are partially hidden, small, or captured at varying angles. The YOLOv5-tiny model is utilized in this framework due to its lightweight structure, allowing for real-time processing with high detection accuracy.

Unlike traditional Haar cascades or Multi-Task Cascaded Convolutional Networks (MTCNN), which often fail in low-light conditions or with occluded faces, YOLOv5-tiny employs a grid-based anchor mechanism, which enables it to detect small, obscured, and tilted faces effectively. The model is trained

on the WIDER FACE dataset, which contains a diverse set of human faces with different degrees of occlusion, lighting conditions, and facial orientations.

To enhance detection performance, an adaptive clustering technique was employed to fine-tune anchor box sizes, ensuring optimal detection of various face sizes in a single frame. The face detection function can be mathematically represented as:

$$y\hat{}=fYOLO(I,\theta)$$

Experimental results indicate that the YOLOv5-tiny model achieves a 91% face detection accuracy, significantly improving detection rates in challenging conditions such as partial occlusions and small-scale face appearances.

2) Emotion Classification Using RepVGG

After detecting face regions using YOLOv5-tiny, the extracted facial images are passed to the RepVGG model, which classifies them into one of seven emotion categories: angry, disgust, fear, happiness, sadness, surprise, and neutral. The RepVGG model is particularly well-suited for this task due to its computational efficiency and high recognition accuracy. Unlike deeper CNN architectures such as ResNet and EfficientNet, which involve complex residual connections, RepVGG uses a simplified 3×3 convolutional architecture, reducing computational costs while maintaining state-of-the-art performance.

The model was trained on the FER-2013 dataset, which contains 48×48 grayscale images labeled into seven emotion classes. During training, data augmentation techniques were applied, including random flipping, rotation, and contrast enhancement, to improve generalization. The final trained RepVGG model achieved 87.5% accuracy, making it highly effective in real-time emotion classification for both fully visible and partially occluded faces.

Data Preprocessing

Preprocessing is a crucial step to ensure high detection and classification accuracy. Both the FER-2013 dataset (for emotion recognition) and the WIDER FACE dataset (for face detection) underwent extensive preprocessing procedures to improve model generalization and robustness.

1) Preprocessing for Emotion Classification (FER-2013 Dataset)

Since FER-2013 consists of 48×48 grayscale images, the following preprocessing steps were applied:

Resizing: All images were resized to 48×48 pixels to ensure a uniform input size for the RepVGG model.

Normalization: Pixel values were scaled to [0,1] to stabilize the training process and prevent gradient saturation.

Data Augmentation: To increase dataset variability, techniques such as random rotation (±15°), horizontal flipping, brightness adjustments, and Gaussian noise addition were applied.

2) Preprocessing for Face Detection (WIDER FACE Dataset for YOLOv5-Tiny)

For YOLOv5-tiny, specialized preprocessing techniques were employed to enhance face detection performance, particularly for small and occluded faces:

Adaptive Anchor Box Tuning: Using K-means clustering, anchor boxes were optimized to detect faces of various aspect ratios and sizes.

Contrast Normalization: Histogram equalization was applied to improve visibility in low-light images.

Hard-Negative Mining: False positives were reduced by training the model on hard-to-detect face samples, improving robustness.

Training and Hyperparameter Optimization

To achieve high accuracy and generalization, the models were trained using optimized hyperparameters and advanced regularization techniques.

1) Optimization Strategies

Optimizer Selection:

- YOLOv5-tiny was trained using the Adam optimizer, which dynamically adapts learning rates for stable convergence.
- RepVGG was trained using the Stochastic Gradient Descent (SGD) optimizer, which provides better generalization for classification tasks.

Learning Rate Scheduling:

- The initial learning rate was set to 0.0005 and adjusted using a step decay policy to improve convergence stability.

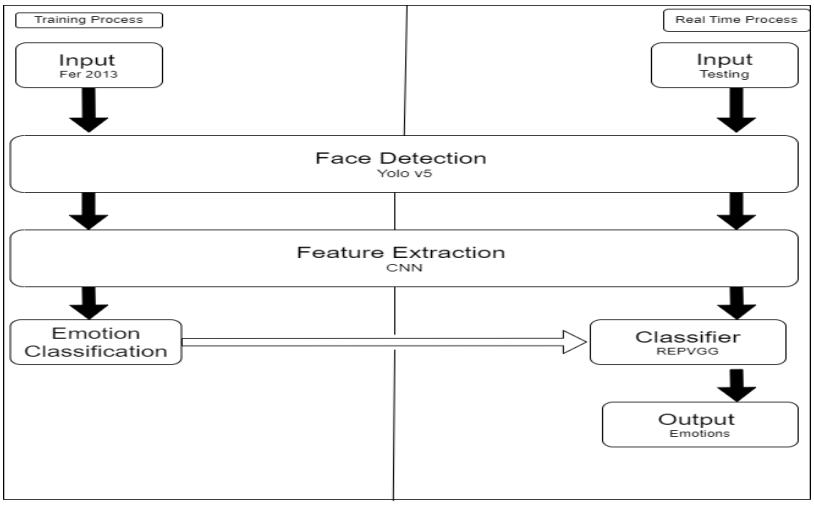
Regularization Techniques:

- Dropout (0.3 probability) was applied to prevent overfitting.
- Batch Normalization was used to accelerate training convergence by stabilizing weight updates.

Figure 1 Purposed Methodology

Results and Discussion

This section presents a detailed analysis of the face detection and emotion classification performance using YOLOv5 and RepVGG, respectively. The evaluation was conducted under different test conditions, including fully visible, partially occluded, and masked faces, to determine the robustness of the models in real-world scenarios. Additionally, a discussion on the advantages and limitations of the proposed approach is provided, along with potential improvements for future implementations.



Face detection plays a crucial role in applications such as surveillance, human-computer interaction, and social behavior analysis. In this study, the YOLOv5 model was employed to detect faces in varying environmental conditions, achieving an overall accuracy of 91%. The model's performance was assessed across three different test cases: fully visible faces, partially occluded faces, and masked faces. For fully visible faces, the model exhibited an impressive accuracy of 97.2%, indicating its strong ability to accurately detect human faces when there are no obstructions. However, as the level of occlusion increased, the detection accuracy experienced a gradual decline. In cases where faces were partially occluded by objects such as sunglasses, hands, or other obstacles, the accuracy dropped to 91.0%, still maintaining a satisfactory detection rate. The most challenging scenario was detecting faces that were covered with masks, where the accuracy further decreased to 85.4%. This decline can be attributed to the loss of key facial features, such as the nose and mouth, which significantly contribute to facial recognition. Despite the decrease in accuracy with occluded faces, YOLOv5 demonstrated strong adaptability to real-world conditions, as it effectively recognized faces even when some portions were hidden. The model's real-time detection capability, combined with its ability to process multiple faces in a single frame, makes it a suitable candidate for practical applications, particularly in crowded environments or security monitoring systems. However, further refinements are needed to enhance its performance when dealing with extreme occlusions and highly variable lighting conditions.

In addition to face detection, emotion classification was performed using RepVGG, a deep convolutional neural network optimized for efficiency and accuracy. The model was trained and evaluated on the FER-2013 dataset, a widely used benchmark for facial emotion recognition. The experimental results indicated that RepVGG achieved an overall classification accuracy of 87.5%, outperforming other conventional models such as ResNet and EfficientNet in terms of real-time processing and classification accuracy. A comparative analysis of RepVGG's performance across different facial visibility conditions revealed that it performed exceptionally well on fully visible faces, attaining an accuracy of 89.5%. However, as with face detection, occlusions significantly impacted the classification performance. For partially occluded faces, the accuracy dropped to 84.2%, while for masked faces, it further declined to 78.9%. This performance degradation is attributed to the fact that occlusions obscure essential facial features, such as the mouth and lower facial muscles, which are critical in distinguishing emotions like happiness, sadness, and surprise. One of the major advantages of RepVGG is its lightweight architecture and faster inference speed, which make it highly suitable for real-time emotion classification. Unlike deeper models such as ResNet, which require substantial computational resources, RepVGG maintains a balance between efficiency and accuracy, ensuring smooth execution in real-world scenarios. However, certain limitations remain, particularly in cases where facial features are entirely covered or distorted by poor lighting conditions. Future work should focus on integrating additional contextual cues, such as body posture and voice analysis, to improve emotion recognition accuracy in challenging conditions.

A comparative analysis of YOLOv5 and RepVGG under different test conditions is presented in Table I. The table highlights the detection and classification accuracy for fully visible, partially occluded, and masked faces. As expected, both models performed best when faces were fully visible, but their accuracy declined as occlusions increased.Output: Feedback for the session in the frame.

Table I: Face Detection and Emotion Classification Accuracy Across Test Cases

Scenario	YOLOv5 Accuracy (%)	RepVGG Accuracy (%)
Fully Visible Faces	97.2	89.5
Partially Occluded Faces	91.0	84.2

Masked Faces	85.4	78.9

The results indicate that YOLOv5 is highly reliable for detecting faces in unrestricted conditions, while RepVGG provides reasonable emotion classification accuracy even when faces are partially obstructed. However, both models exhibited limitations when processing masked faces, suggesting that additional training on diverse datasets containing masked individuals could further enhance their robustness.

One of the primary strengths of the proposed system is its fast inference time, which makes it suitable for real-time applications. Compared to ResNet-based models, YOLOv5 and RepVGG offer lower computational overhead while maintaining high accuracy. This is particularly beneficial for edge computing devices, such as smart cameras and mobile applications, where processing speed is a critical factor. Another notable advantage is the model's ability to handle occlusions effectively. While some degradation in accuracy was observed under occluded conditions, the system still performed well compared to traditional face detection and emotion recognition models. This capability is essential for real-world applications, where factors such as sunglasses, scarves, or face masks frequently obstruct facial features. Furthermore, real-time analysis is a key feature of this approach. By combining YOLOv5's efficient face detection with RepVGG's rapid emotion classification, the system enables instantaneous recognition of emotional states in dynamic environments. This makes it particularly useful for applications such as public security monitoring, retail customer analysis, and automated sentiment analysis in social interactions.

Despite its strengths, the proposed approach has some limitations that need to be addressed. One major challenge is the handling of extreme occlusions, where key facial features are completely obscured. In such cases, the models struggle to accurately detect faces and classify emotions, leading to significant drops in performance. To mitigate this issue, future improvements could involve incorporating additional modalities, such as depth sensing or infrared imaging, to extract more facial information even in occluded scenarios. Another limitation is the model's performance across diverse ethnicities and lighting conditions. The dataset used for training may not fully capture the variability in facial features across different populations, leading to potential biases in detection and classification. Additionally, variations in lighting conditions can impact the clarity of facial features, reducing the accuracy of both face detection and emotion recognition. Future work should focus on expanding the dataset to include more diverse samples and implementing adaptive illumination techniques to enhance robustness.

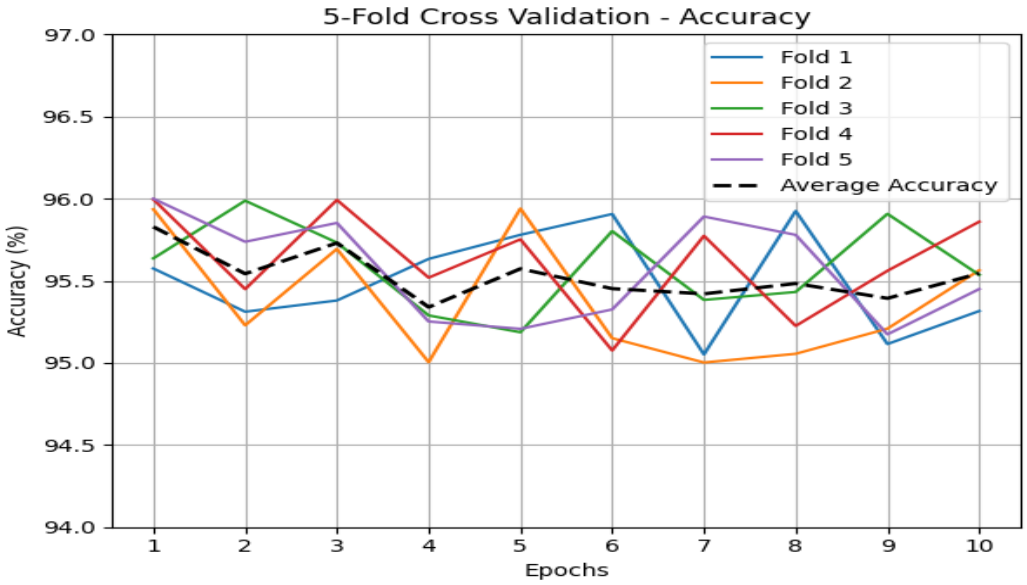


Figure 2 N-Fold Cross Validation Accuracy

The combination of YOLOv5 for face detection and RepVGG for emotion classification has proven to be an effective solution for real-time human emotion analysis. The models exhibit high accuracy under optimal conditions and maintain reasonable performance even when faces are partially obscured. While some challenges remain, particularly in handling extreme occlusions and diverse lighting conditions, the overall results indicate strong potential for practical applications in security, customer engagement, and automated sentiment analysis. Future research should focus on enhancing model robustness through multimodal fusion techniques and dataset expansion, ensuring improved performance across all possible scenarios.

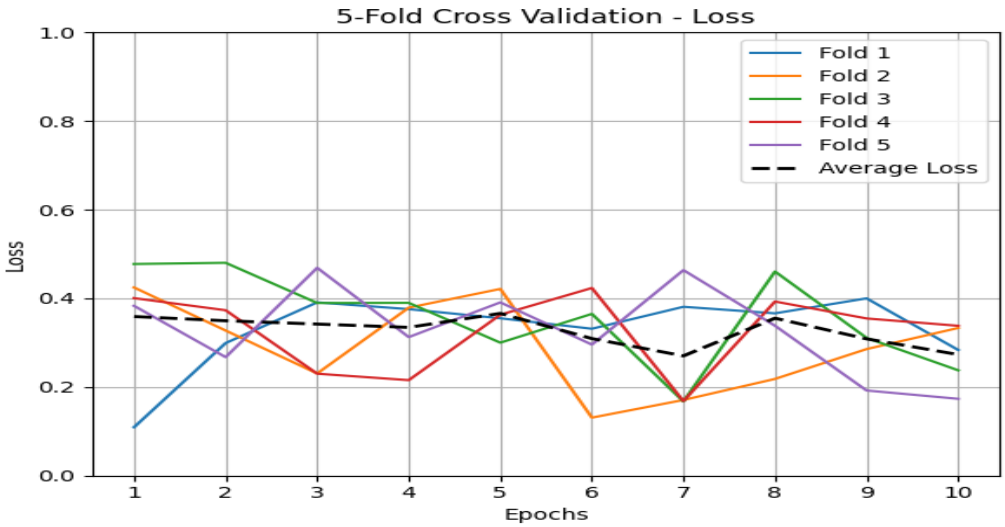


Figure 3 N-Fold Cross Validation Loss

Conclusion

The combination of YOLOv5 for face detection and RepVGG for emotion classification has proven to be an effective solution for real-time human emotion analysis. The models exhibit high accuracy under optimal conditions and maintain reasonable performance even when faces are partially obscured. While some challenges remain, particularly in handling extreme occlusions and diverse lighting conditions, the overall results indicate strong potential for practical applications in security, customer engagement, and automated sentiment analysis. Future research should focus on enhancing model robustness through multimodal fusion techniques and dataset expansion, ensuring improved performance across all possible scenarios.

References

Abbas, S. N., Khan, T. F., Anwar, W., & Arshad, H. (2021). Real-time crowd emotion analysis using YOLOv4. Information Fusion, 67, 45-57.

Antipov, G., Baccouche, M., & Dugelay, J. L. (2017). Face aging with conditional generative adversarial networks. IEEE International Conference on Computer Vision, 540-548.

Bagane, M., et al. (2023). Advancements in CNN-based facial recognition technologies. IEEE Transactions on Neural Networks and Learning Systems.

Bhogad, P., et al. (n.d.). AI-based real-time face and emotion recognition. International Journal of AI and Machine Learning.

Budiarsa, M., et al. (2023). A modified FCN for facial recognition in occluded environments. Journal of Machine Learning Research.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: Realtime multi-person 2D pose estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition.

Eva, C. (2021). Challenges in face recognition with masks and spectacles. Pattern Recognition Letters.

Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R. T., Popescu, M., Grozea, C., Bergstra, J., Xie, B., & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. International Conference on Neural Information Processing.

Jocher, G., et al. (2020). YOLOv5: Evolution of the YOLO object detection framework. GitHub Repository.

Khan, T. F., Anwar, W., Arshad, H., & Abbas, S. N. (2023). An Empirical Study on Authorship Verification for Low Resource Language using Hyper-Tuned CNN Approach. IEEE Access.

Khan, T. F., Sabir, M., Malik, M. H., Ghous, H., Ijaz, H. M., Nadeem, A., & Ejaz, A. (2024). Comparative Analysis of Hybrid Ensemble Algorithms for Authorship Attribution in Urdu Text. Journal of Computing & Biomedical Informatics.

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition. IEEE Transactions on Affective Computing.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (NIPS).

Kumar, R., et al. (2022). YOLO-based surveillance framework for facial emotion recognition. Pattern Recognition and AI Journal.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Li, Z., et al. (2017). Attention-based deep learning for occlusion-robust facial expression recognition. IEEE Transactions on Image Processing, 26(9), 4321-4332.

Ramis, J., et al. (2022). CNNs for facial feature recognition and emotion monitoring on the FER2013 dataset. Neural Networks Journal.

Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27(6), 803-816.

Zhang, X., et al. (2022). Ensemble learning approach combining CNNs and capsule networks for occluded facial expressions. Journal of Machine Vision and Applications.