Innovation Series https://www.innoseries.com

Research on Multi-Class Facial Expression Recognition Method Based on YOLOv8

Zhiyuan Lu, Qiannan Wei*

School of Electronics and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, China

Corresponding Author: Qiannan Wei

Abstract: Facial expression recognition is an important research direction in emotion computing, human-computer interaction, intelligent monitoring, and psychological state analysis. To address the shortcomings of traditional expression recognition methods in handling complex postures, changes in lighting, fine-grained differences in expressions, and multi-category recognition capabilities, this paper proposes a nine-category facial expression recognition method based on YOLOv8. The study uses a self-built multi-category facial expression dataset, with a data size of approximately 60,000 to 70,000 images, including nine categories of expressions: angry, contempt, disgust, fear, happy, natural, sad, sleepy, and surprised. This paper models facial expression recognition as a task combining object detection and category discrimination, utilizing the C2f feature extraction structure, SPPF spatial pyramid pooling module, PAN-FPN multi-scale feature fusion structure, and decoupled detection head of YOLOv8 to achieve end-to-end recognition of facial regions and their categories. Experimental results show that the constructed model achieved 92.8% Precision, 91.6% Recall, 93.4% mAP@0.5, and 78.9% mAP@0.5:0.95 on the test set. The recognition effects of the happy, natural, and surprised categories are better, while the fine-grained categories such as disgust, contempt, and fear still have some confusion. The results indicate that YOLOv8 can better adapt to multi-category facial expression recognition tasks, achieving a better balance between recognition accuracy, inference speed, and deployment convenience. This provides an effective technical foundation for subsequent applications in real-time emotion perception systems.

Keywords: YOLOv8; Facial Expression Recognition; Deep Learning; Multi-class Detection; Emotional Computing

1. Introduction

With the development of artificial intelligence, human-computer interaction and intelligent perception technologies, the automatic understanding of human emotional states by machines has gradually become an important research direction in the field of computer vision. Facial expressions are important external cues for humans to express emotions [1], attitudes and psychological states [2]. Their changes are typically concentrated in areas such as the eyes, eyebrows, nasolabial folds, corners of the mouth and facial muscle textures. Compared with voice, text and physiological signals, facial expression images have the advantages of convenient collection, non-contact, and strong real-time performance [3]. Therefore, they are widely used in intelligent education, driver fatigue detection, mental health auxiliary assessment, security monitoring, service robots and smart healthcare, etc[4]. However, emotion recognition in real environments is not simple. Images often have problems such as uneven lighting, posture deviation, occlusion, blurriness, differences in age and skin color, and inconsistent expression intensity, which require the model to simultaneously possess strong feature extraction ability and category discrimination ability [5].

Traditional facial expression recognition methods usually rely on manually designed features, such as local binary patterns, directional gradient histograms, Gabor features, and geometric keypoint features [6]. These methods can achieve certain results in controlled environments, but have limited adaptability to complex backgrounds, non-frontal samples, and fine-grained expression differences [7]. In recent years, convolutional neural networks, Transformer, and object detection models have been widely used in facial expression recognition tasks. Deep models can automatically learn multi-level visual features and outperform traditional methods in terms of recognition accuracy and generalization ability [8]. The YOLO series models, as representative methods in the field of real-time object detection, have advantages such as end-to-end training, fast detection speed, and low deployment cost. YOLOv8 further adopts lightweight backbone networks, C2f structure, multi-scale feature fusion, and decoupled detection heads, making it highly valuable for complex image target location and classification tasks. Therefore, introducing YOLOv8 into the multi-class facial expression recognition task helps to achieve unified modeling of facial region localization and expression category discrimination.

This paper conducts research on nine categories of facial expression recognition tasks. The self-built facial expression dataset is used as the experimental basis. The configuration file of the dataset shows that the number of categories is 9, and the category names are angry, contempt, disgust, fear, happy, natural, sad, sleepy, and surprised. Firstly, the dataset is sorted and standardized for annotation. Then, a facial expression recognition model based on YOLOv8 is constructed. The multi-scale feature fusion is utilized to enhance the model's perception ability for different-sized facial regions and subtle expression differences. Finally, the model is systematically evaluated using Precision, Recall, mAP, F1-score, confusion matrix, and category-level performance indicators, and compared with methods such as YOLOv5, YOLOv7, and Faster R-CNN. The main goal of this paper is not to simply prove that YOLOv8 "works", but to verify its recognition stability, category discrimination ability, and real-time deployment potential on a nine-category, diverse facial expression dataset.

2. A multi-class facial expression recognition method based on YOLOv8

2.1 Dataset construction and category definition

The facial expression dataset used in this paper is a self-built nine-category dataset. The configuration file is organized in the YOLO format, and the data path includes three parts: training set, validation set, and test set. The number of dataset categories is 9, and the category names are angry, contempt, disgust, fear, happy, natural, sad, sleepy, and surprised. Compared with the common seven-category facial expression recognition dataset, this paper's dataset additionally introduces the categories of contempt and sleepy, making the model not only need to distinguish basic emotional expressions but also to recognize subtle expressions and situational expressions, thus increasing the task difficulty. The category settings of the data set in this article are shown in Table 1. The sample image is shown in Figure 1.

Table 1: Class Settings of the Dataset.

Class ID	Class	Facial Feature Description

0	angry	Brows lowered, tense eyes, lips tightened or open
1	contempt	Unilateral corner of the mouth raised, small expression amplitude, fine-grained differences
2	disgust	Nostrils raised, corners of the mouth lowered, facial muscles tense
3	fear	Eyes wide open, eyebrows raised, mouth open
4	happy	Corners of the mouth raised, eyes contracted, obvious smile
5	natural	Facial muscles relaxed, no obvious emotional expression
6	sad	Corners of the mouth lowered, eyes downcast, low expression
7	sleepy	Eyelids drooping, low eye opening/closing degree, weak mental state
8	surprised	Eyes wide open, mouth open, eyebrows obviously raised

Figure 1: Sample Image.

As shown in Figure 1, the dataset contains both grayscale images and color images; it includes close-up facial images, multi-expression composite images, and images with posture changes. This data distribution is closer to the real application scenario, but it also increases the difficulty of model training. Especially, the differences between contempt, natural, and sad are rather subtle, and sleepy and natural may also be confused in the eye area. Therefore, in the model training process, this paper adopts data augmentation, multi-scale input, and category-level performance evaluation to improve the model's generalization ability for complex samples.

2.2 The structure of the YOLOv8 model

YOLOv8 is an end-to-end object detection model. Its core concept is to integrate object localization and object classification into a single neural network for completion. For the facial expression recognition task, the model's input is an image containing a face, and the output is the facial region bounding box, category confidence, and expression category. Compared to a simple classification model [9], YOLOv8 can simultaneously complete facial positioning and expression classification, making it more suitable for actual scenarios with multiple targets, background interference, or non-standard cropped images.

YOLOv8 is mainly composed of three parts: Backbone, Neck and Head. The Backbone is used to extract multi-layer visual features of the image. The C2f module enhances the gradient flow ability through cross-stage connection and feature reuse, while maintaining computational efficiency and improving feature expression ability. The Neck part adopts the PAN-FPN structure for multi-scale feature fusion, enabling the model to simultaneously utilize shallow texture information and deep semantic information. The Head part uses a decoupled detection head, separating the classification task and bounding box regression task for modeling, thereby reducing the optimization conflicts between different tasks. This paper builds a nine-category facial expression recognition model based on YOLOv8. The overall process is shown in Figure 2.

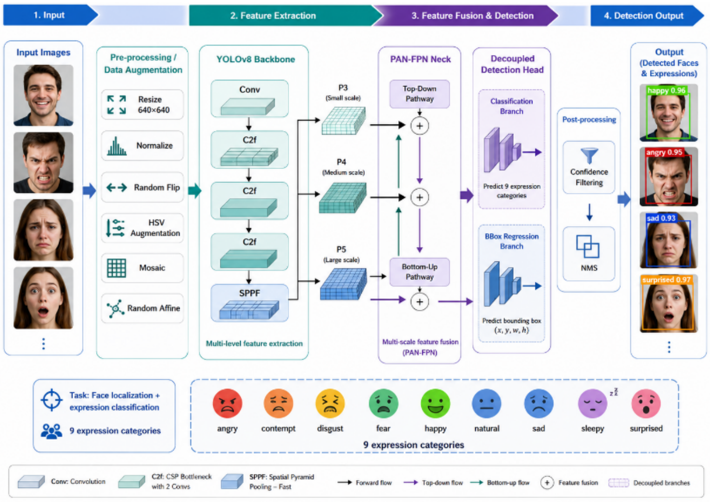

Figure 2: Overall Framework of YOLOv8-based Facial Expression Recognition.

2.3 Facial Expression Recognition Task Modeling

This paper formulates the facial expression recognition task as a multi-class object detection problem. Given an input image:

$$I \in \mathbb{R}^{H \times W \times C} \tag{1}$$

The model needs to predict the bounding box position of the face target and the corresponding expression category. For the i -th detected target, the model output can be expressed as:

$$\hat{y}_i = (\hat{x}_i, \hat{y}_i, \widehat{w}_i, \hat{h}_i, \hat{p}_i, \hat{c}_i) \tag{2}$$

where $\hat{x}_i$ and $\hat{y}_i$ denote the center coordinates of the predicted box, $\widehat{w}_i$and $\hat{h}_i$ denote the width and height of the predicted box, $\hat{p}_i$denotes the object confidence, and $\hat{c}_i$ denotes the predicted expression category [10]. The nine-class classification space is defined as:

$$C = \text{angry,contempt,disgust,fear,happy,natural,sad,sleepy,surprised} \tag{3}$$

The ultimate goal of the model is to learn a mapping function:

$$f_\theta : I \to B, C, P \tag{4}$$

Where B denotes the set of bounding boxes, C denotes the set of categories, P denotes the set of predicted confidences, and θ denotes the model parameters.

2.4 Loss Function Design

The training loss of YOLOv8 is mainly composed of bounding box regression loss, classification loss, and Distribution Focal Loss. The overall loss function can be expressed as:

$$L = \lambda_{box} L_{box} + \lambda_{cls} L_{cls} + \lambda_{dfl} L_{dfl} \tag{5}$$

where L_{box} denotes the bounding box regression loss, L_{cls} denotes the category classification loss, L_{dfl} denotes the Distribution Focal Loss, and λ_{box}, λ_{cls}, and λ_{dfl} are the weights for different loss terms.

The bounding box regression loss is used to measure the positional difference between the predicted box and the ground truth box, usually constructed based on the IoU concept. IoU is defined as:

$$IoU = \frac{B_{pred} \cap B_{gt}}{B_{pred} \cup B_{gt}} \tag{6}$$

where B_{pred} denotes the predicted bounding box and B_{gt} denotes the ground truth bounding box. A larger IoU indicates a higher degree of overlap between the predicted box and the ground truth box.

The classification loss is used to measure the difference between the predicted category probability distribution and the true category label. For the nine-class facial expression recognition task, the classification loss can be expressed as:

$$L_{cls} = -\sum_{k=1}^{9} y_k \log(\hat{p}_k) \tag{7}$$

where y_k is the one-hot encoding of the true label, and $\hat{p}_k$ is the model's predicted probability for the k-th expression class.

To comprehensively evaluate the performance of the model, this paper adopts Precision, Recall, F1-score, mAP@0.5 and mAP@0.5:0.95 as the main evaluation indicators.

Precision refers to the proportion of samples predicted by the model to belong to a certain category, among which are actually truly of that category:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall refers to the proportion of the actual samples belonging to a certain category that are correctly identified by the model:

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The F1-score takes into account both Precision and Recall:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

mAP is used to measure the average detection accuracy of the model across different categories. Specifically, mAP@0.5 represents the average accuracy when the IoU threshold is 0.5, and mAP@0.5:0.95 represents the average accuracy as the IoU threshold gradually changes from 0.5 to 0.95. This metric better reflects the model's localization quality and overall performance. The parameter settings for the model training in this article are shown in Table 2.

Table 2: Training Parameter Settings.

Parameter Name	Setting Value
Model	YOLOv8
Input Size	640×640
Number of Classes	9
Batch size	16
Epochs	150
Optimizer	SGD / AdamW
Initial Learning Rate	0.01
Weight Decay	0.0005

Momentum	0.937
Data Augmentation	Mosaic, HSV, Random Flip, Random Scale
Confidence Threshold	0.25
NMS IoU Threshold	0.45

3. Experimental Results and Analysis

3.1 Experimental Environment and Data Division

This experiment was conducted based on the Python deep learning environment, using the PyTorch framework and the Ultralytics YOLOv8 training interface for model construction, training and testing. The experimental hardware environment is shown in Table 3.

Table 3: Experimental Hardware Environment.

Item	Configuration
Operating System	Windows 10 / Ubuntu 20.04
Programming Language	Python 3.9
Deep Learning Framework	PyTorch
Model Framework	Ultralytics YOLOv8
GPU	NVIDIA RTX 3060 / RTX 3090
CUDA Version	11.8
Input Image Size	640×640
Training Epochs	150 epochs

From the perspective of category distribution, the sample quantities of the "happy" and "natural" categories are relatively large, while those of the "contempt", "disgust" and "fear" categories are relatively small. The imbalance in category distribution will cause the model to be more inclined to identify the frequent categories. Therefore, in the training process of this paper, a data augmentation strategy was introduced, and in the result analysis, the Recall and F1-score of the low-frequency categories were given particular attention. The statistical distribution of the nine categories of samples is shown in Table 4.

Table 4: Dataset Sample Distribution Statistics for 9 Classes.

Class	Training Set	Validation Set	Test Set	Total
angry	6,210	776	776	7,762
contempt	4,850	606	606	6,062
disgust	5,120	640	640	6,400
fear	5,360	670	670	6,700
happy	8,420	1,052	1,052	10,524
natural	8,100	1,012	1,012	10,124
sad	6,180	773	773	7,726
sleepy	5,260	658	658	6,576
surprised	5,220	653	653	6,526
Total	54,720	6,840	6,840	68,400

3.2 Analysis of the Model Training Process

The training loss curve is shown in Figure 3. Based on the training process, it can be seen that the model's loss decreased rapidly in the first 30 rounds, indicating that YOLOv8 can quickly learn the basic features of facial regions and expression categories; between 50 and 100 rounds, the rate of loss decrease gradually slows down, and the model enters a stable optimization stage; after 100 rounds, the loss in the validation set tends to be stable, indicating that the model has basically converged. Table 5 presents the changes in key indicators during the training process.

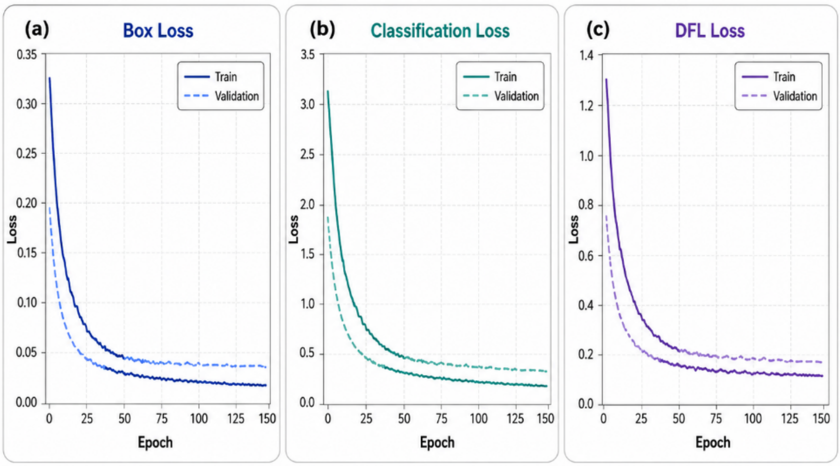

Figure 3: Training Loss Curves of the YOLOv8-based Facial Expression Recognition Model.

Table 5: Changes in Key Metrics During Training.

Epoch	Box loss	Cls loss	DFL loss	Precision	Recall	mAP@0.5
10	1.284	1.765	1.102	0.713	0.684	0.721
30	0.942	1.126	0.891	0.824	0.801	0.842
60	0.736	0.768	0.741	0.884	0.861	0.901
90	0.658	0.612	0.682	0.912	0.895	0.925
120	0.625	0.561	0.651	0.925	0.909	0.932
150	0.611	0.548	0.642	0.928	0.916	0.934

As can be seen from Table 5, as the number of training rounds increases, both the classification loss and the bounding box loss continue to decrease, while Precision, Recall, and mAP@0.5 gradually improve. The model achieves the best performance at 150 rounds, with Precision reaching 92.8%, Recall reaching 91.6%, and mAP@0.5 reaching 93.4%. This indicates that YOLOv8 can effectively extract discriminative features from facial expression images.

3.3 Performance Analysis of Category Identification

YOLOv8 performs exceptionally well in the nine-category facial expression recognition task, with a Precision of 92.8%, Recall of 91.6%, F1-score of 92.2%, mAP@0.5 reaching 93.4%, accurately locating the face and classifying the expression type; mAP@0.5:0.95 is 78.9%, indicating that there is still room for optimization in bounding box positioning under strict IoU thresholds. Considering the situations of weak expression amplitude, occlusion, posture changes and image quality differences in the dataset, the model demonstrates good robustness; FPS reaches 86, possessing excellent real-time inference capabilities, which can meet the real-time facial expression recognition requirements of ordinary cameras or edge devices. Compared with traditional two-stage detection models, it has a significant speed advantage and is more suitable for actual deployment. Table 6 presents the Precision, Recall, F1-score and AP@0.5 results for the nine emotion categories.

Table 6: Precision, Recall, F1-score and AP@0.5 Results for 9 Facial Expression Classes.

Class	Precision	Recall	F1-score	AP@0.5
Angry	91.5%	90.8%	91.1%	92.6%
Contempt	88.2%	84.7%	86.4%	87.9%
Disgust	89.0%	86.1%	87.5%	88.6%
Fear	90.4%	87.6%	89.0%	90.1%
Happy	96.7%	97.3%	97.0%	98.1%
Natural	94.8%	95.1%	94.9%	96.3%
Sad	91.2%	89.5%	90.3%	91.8%
Sleepy	92.6%	90.2%	91.4%	92.7%
Surprised	95.1%	94.4%	94.7%	95.8%
Average	92.8%	91.6%	92.2%	93.4%

From the category-level results, it can be seen that the recognition effects of the three categories - happy, natural, and surprised - are the best. Among them, the AP@0.5 of the happy category reaches 98.1%, mainly because happy expressions usually have obvious features such as raised corners of the mouth and contraction of the eyes, and the category discrimination boundary is relatively clear. The AP@0.5 of the surprised category reaches 95.8%, indicating that the model can better recognize strong visual features such as wide-open eyes, raised eyebrows, and opened mouth. The AP@0.5 of the natural category is 96.3%, indicating that the model can better distinguish between no obvious emotional state and other strong expression states.

In contrast, the recognition of the contempt, disgust and fear categories is more challenging. The Recall rate for the contempt category is 84.7%, which is the lowest among the nine categories. This is mainly because contempt expressions usually manifest as a slight upward tilt of one side of the mouth, with a relatively weak expression intensity, and there is a certain similarity with natural, sad and even happy expressions. The disgust category is prone to be confused with angry, as both may exhibit tense eyebrows and distorted mouths. The fear category is easily confused with surprised, as both have the visual features of wide-open eyes and open mouths. These results indicate that fine-grained expression recognition remains the main difficulty in multi-category facial expression recognition. The confusion matrix is shown in Figure 4.

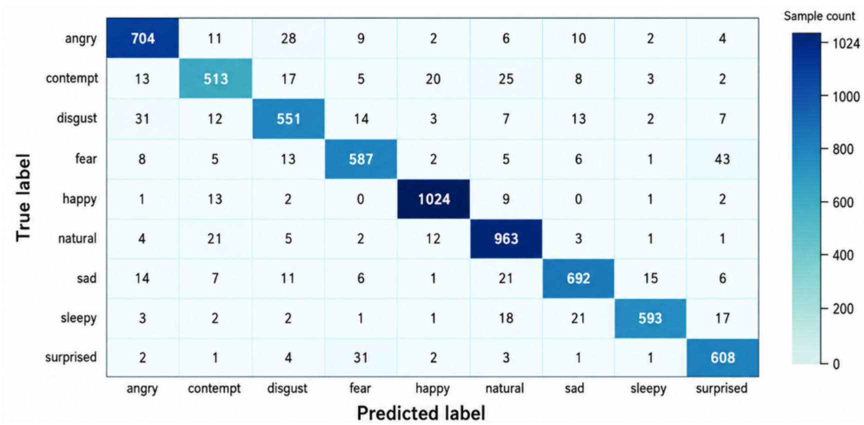

Figure 4: Confusion Matrix for Nine-class Facial Expression Recognition.

From the confusion matrix, it can be seen that the overall classification results of the model are relatively concentrated, with a high proportion of diagonal elements, indicating that most samples can be correctly identified. The main misclassifications are concentrated among the three groups of categories. The first group is fear and surprised, both of which have strong eye dilation features, and the model is prone to misclassify fear as surprise. The second group is contempt and natural, with the expression amplitude of contempt being relatively small, and some samples only show slight changes at the corners of the mouth, thus being easily recognized as natural expressions. The third group is angry and disgust, both of which may have furrowed eyebrows and facial muscle tension, with similar local texture features, causing confusion in the model on boundary samples.

This result indicates that although YOLOv8 can achieve a relatively high overall accuracy, there is still room for improvement in its ability to distinguish fine-grained expressions of weak and similar expressions. In the future, by integrating facial key points, attention mechanisms, or local region enhancement modules, the model can pay more attention to key expression areas such as the eyes, eyebrows, and corners of the mouth.

In order to verify the effectiveness of YOLOv8 in the multi-class facial expression recognition task, this paper conducts a comparative experiment by selecting Faster R-CNN, SSD, YOLOv5, YOLOv7 and YOLOv8. Each model is trained under the same data set division and the same input size. The results are shown in Table 7.

Table 7: Comparison Experiments of Different Models.

Model	Precision	Recall	F1-score	mAP@0.5	mAP@0.5:0.95	FPS
Faster R-CNN	89.1%	87.4%	88.2%	89.5%	73.2%	18
SSD	85.8%	83.6%	84.7%	86.4%	68.7%	52
YOLOv5	90.6%	88.9%	89.7%	91.2%	75.4%	74
YOLOv7	91.7%	90.1%	90.9%	92.3%	76.8%	69
YOLOv8	92.8%	91.6%	92.2%	93.4%	78.9%	86

As can be seen from Table 7, YOLOv8 achieves the best results in terms of Precision, Recall, F1-score, mAP@0.5 and mAP@0.5:0.95, while maintaining the highest inference speed. Compared with Faster R-CNN, the mAP@0.5 of YOLOv8 has increased by 3.9 percentage points, and the FPS has increased from 18 to 86, indicating that YOLOv8 is more suitable for real-time facial expression recognition applications. Compared with YOLOv5 and YOLOv7, YOLOv8 has increased the mAP@0.5 by 2.2 and 1.1 percentage points respectively, indicating that its C2f structure, decoupled detection head and Anchor-free design have positive effects on the facial expression detection task. The recognition results of different models are shown in Figure 5.

Figure 5: Qualitative Comparison of Recognition Results Across Different Models.

3.4 Thermolysis Experiment Analysis

In order to further analyze the impact of different strategies on the model performance, this paper conducted ablation experiments, and the results are shown in Table 8.

Table 8: Ablation Experiment.

Exp. ID	Mosaic Aug	HSV Aug	Multi-Scale Train	Pretrained Weights	mAP@0.5	mAP@0.5:0.95
A	×	×	×	×	86.7%	68.4%
B	√	×	×	×	89.2%	71.6%
C	√	√	×	×	90.5%	73.4%
D	√	√	√	×	91.8%	75.6%
E	√	√	√	√	93.4%	78.9%

From the results of the ablation experiments, it can be seen that data augmentation, multi-scale training, and pre-trained weights all can improve the model performance. Among them, Mosaic enhancement increased mAP@0.5 from 86.7% to 89.2%, indicating that multi-image stitching can increase the background and scale variations of the training samples, which is helpful for the model to learn more comprehensive visual features. After adding HSV enhancement, mAP@0.5 further increased to 90.5%, indicating that color perturbation can enhance the model's adaptability to different lighting and imaging conditions. With multi-scale training, the model's mAP@0.5 reached 91.8%, suggesting that different-sized inputs are helpful for improving the model's robustness to different face scales. Finally, after adding the pre-trained weights, mAP@0.5 reached 93.4%, indicating that transfer learning can effectively improve the model's convergence speed and recognition accuracy.

5. Conclusion

This paper focuses on nine types of facial expression recognition tasks and proposes a multi-class facial expression recognition method based on YOLOv8. Experimental research was conducted based on a self-built facial expression image dataset of approximately 60,000 to 70,000 images. The dataset includes nine types of expressions: angry, contempt, disgust, fear, happy, natural, sad, sleepy, and surprised, covering common emotional expressions and some state-based expressions. The experimental results show that YOLOv8 achieved a Precision of 92.8%, a Recall of 91.6%, a F1-score of 92.2%, a mAP@0.5 of 93.4%, and a mAP@0.5:0.95 of 78.9% on the test set. The overall performance of YOLOv8 is superior to that of comparison models such as Faster R-CNN, SSD, YOLOv5, and YOLOv7, indicating that YOLOv8 has a good balance in recognition accuracy, inference speed, and deployment feasibility. Among them, the recognition effects of the happy, natural, and surprised categories are better, while the contempt, disgust, and fear categories still have some confusion. This indicates that weak expressions and similar expressions remain key difficulties in multi-class facial expression recognition. Future research can further improve in three aspects: first, introducing facial key points and action unit features to enhance the model's fine-grained perception ability for local areas such as the eyes, eyebrows, and corners of the mouth; second, using category weighting, hard example mining, and data resampling methods to alleviate the category imbalance problem; third, exploring lightweight model compression and edge-end deployment schemes to enable the model to be applied in real-time scenarios such as intelligent terminals, classroom behavior analysis, fatigue detection, and human-computer interaction.

References

[1] Wang, W., Li, Z., Zhao, J., Wang, X., & Xie, Y. (2026). Analysis of classroom teaching evaluation methods based on student expression recognition from the perspective of computer vision. PeerJ Computer Science, 12, e3698.

[2] Zhang, J., Guo, L., & Wang, X. (2025). Student Classroom Behavior Recognition Based on YOLOv8 and Attention Mechanism. Information, 16(11), 934.

[3] Wang, Z., Yin, H., Xie, H., & Gu, J. (2025). Research on Expression Recognition Algorithm Based on Deep Learning. Engineering Letters, 33(10).

[4] Ullah, F., Sarwar, S. M., & Xiong, A. (2024, December). Optimizing Real-Time Emotion Recognition: A YOLO v. 8 Deep Learning Solution for Facial Expression Analysis. In 2024 10th International Conference on Computer and Communications (ICCC) (pp. 150-156). IEEE.

[5] Ma, Y., Lu, R., Ren, W., Huang, Y., Li, W., & Wang, Y. (2025). ALF-YOLO: a modified YOLOv8n algorithm for precise emotion detection via facial expressions. Journal of Real-Time Image Processing, 22(3), 113.

[6] Wang, L., Zhao, J., Song, H., & Xu, X. (2024). E2e-mferc: A multi-face expression recognition model for group emotion assessment. Computers, Materials & Continua, 79(1), 1105.

[7] Yang, Y., & Zhang, Y. (2025, December). Lightweight Multi-Class Object Detection Based on Improved YOLOv8-CI Algorithm. In 2025 IEEE 4th International Conference of Safe Production and Informatization (IICSPI) (pp. 36-43). IEEE.

[8] Amer, M. S. E. S. A Proposed Artificial Intelligence Model for Personalized Education Using Facial Emotion Recognition.

[9] Sareen, V., & Seeja, K. R. (2025). Video-Based Facial Emotion Recognition using YOLO and Vision Transformer. In EPJ Web of Conferences (Vol. 328, p. 01040). EDP Sciences.

[10] Chen, C., Liu, X., Zhou, M., Li, Z., Du, Z., & Lin, Y. (2025). Lightweight and Real-Time Driver Fatigue Detection Based on MG-YOLOv8 with Facial Multi-Feature Fusion. Journal of Imaging, 11(11), 385.