



## Emotion Recognition in Consumers Based on Deep Learning and Image Processing: Applications in Advertising

Liangping Sun<sup>ID</sup>, Wenting Song<sup>ID</sup>, Jie Han<sup>\*ID</sup>, Yang Li<sup>ID</sup>

Business School, Qingdao University of Technology, Qingdao 266520, China

Corresponding Author Email: [hanjie@qut.edu.cn](mailto:hanjie@qut.edu.cn)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420223>

### ABSTRACT

**Received:** 9 October 2024

**Revised:** 27 February 2025

**Accepted:** 10 March 2025

**Available online:** 30 April 2025

#### Keywords:

*consumer emotion recognition, deep learning, image processing, micro-expressions, advertising, emotion recognition model*

With the continuous advancement of deep learning and image processing technologies, consumer emotion recognition has emerged as a significant area of research in advertising and marketing. Emotional responses from consumers play a crucial role in optimizing advertising effectiveness and marketing strategies. Among these, micro-expressions—subtle and involuntary facial movements—offer rich emotional cues that can enhance understanding of consumer sentiment. However, existing studies predominantly focus on conventional facial expressions or single-dimensional emotion classification, lacking in-depth exploration and accurate detection of micro-expressions. Additionally, current approaches often overlook individual differences and the dynamic nature of emotional changes, resulting in limited accuracy and real-time performance. Effectively leveraging deep learning and image processing for precise emotion recognition thus presents a critical challenge in modern advertising. Traditional methods—based on facial expressions, speech, or physiological signals—face various limitations in practical applications. Facial expression-based models are sensitive to individual variations and rely heavily on the quality of facial feature extraction. Although speech and physiological signal-based techniques can offer valuable emotional insights, constraints in data acquisition and processing hinder their effectiveness in recognizing complex emotional states. This study aims to enhance the precision and real-time capability of consumer emotion recognition by utilizing deep learning and image processing techniques. The key research contributions include: (1) proposing an improved preprocessing method for micro-expression images to enhance emotional feature extraction; (2) designing a deep learning model tailored for micro-expression recognition to optimize emotion classification accuracy; and (3) developing adaptive advertising strategies based on emotion recognition results to maximize advertising impact. The findings of this research provide novel technological and theoretical support for precision marketing and consumer emotion analysis.

## 1. INTRODUCTION

With the rapid development of artificial intelligence technology, deep learning and image processing technologies have been widely applied in multiple fields [1-4]. Among them, consumer emotion recognition, as an emerging research direction [5, 6], has gradually attracted widespread attention from both academia and industry. Emotion is an important driving force in consumer decision-making [7], and through emotion recognition technology, especially image processing technology targeting micro-expressions [8-10], it is possible to effectively reveal consumers' true emotional responses when facing advertisements and products. Therefore, consumer emotion recognition research based on deep learning and image processing not only provides accurate data support for advertising and marketing but also provides theoretical basis and practical guidance for brand owners to optimize advertising effectiveness, improve consumer satisfaction, and enhance consumer loyalty.

Related studies have shown that the application of emotion

recognition technology has already penetrated into many fields such as advertising, education, and healthcare [11-14]. By recognizing consumers' emotional responses, businesses can more accurately adjust marketing strategies and improve advertising delivery effectiveness. However, although emotion recognition technology has great application potential in advertising, existing studies mostly focus on single-dimensional emotion recognition, lacking in-depth mining of subtle emotional signals such as micro-expressions [15, 16]. In addition, most studies have certain limitations in dealing with consumer emotions, often ignoring individual differences and diversity in emotional expressions, leading to inaccurate emotion recognition results [5, 6]. Therefore, studying the depth and accuracy of consumer emotion recognition, especially how to extract effective emotional features from micro-expressions, has become a research hotspot at present.

At present, there are several research methods in the field of emotion recognition, including emotion recognition models based on facial expressions, emotion analysis methods based on voice, and emotion measurement techniques based on

physiological data. However, these methods all have certain shortcomings. For example, the emotion recognition model based on facial expressions proposed in literature [17] usually relies on the extraction of facial feature points, but due to differences in facial expressions among individuals, misrecognition is likely to occur; while the emotion analysis method based on voice and physiological data proposed in literature [18], although capable of providing certain emotional clues, is limited by the accuracy of data acquisition and the diversity of subjects, and its recognition accuracy still has considerable room for improvement. In addition, existing studies often lack a comprehensive emotion analysis framework and cannot fully consider the dynamic and diverse nature of emotional changes.

This paper mainly studies consumer emotion recognition technology based on deep learning and image processing, and explores its application in advertising. Specifically, this paper conducts research in three parts: firstly, an improved image preprocessing method is proposed for the micro-expression image preprocessing problem in consumer emotion recognition, in order to improve the extraction effect of micro-expression features; secondly, a deep learning-based micro-expression recognition model is designed to optimize emotion classification accuracy and real-time performance; finally, based on the consumer emotion recognition results, advertising marketing strategy adjustment methods are studied, and how to adjust advertising content and delivery strategies according to consumer emotion changes is proposed, in order to maximize advertising marketing effect. This research not only enriches the theoretical system of consumer emotion recognition but also provides a new technical path for precision marketing in the advertising field, with important practical value and social significance.

## 2. MICRO-EXPRESSION IMAGE PREPROCESSING FOR CONSUMER EMOTION RECOGNITION

Since consumers in advertising marketing scenarios are usually in a natural state, their facial expressions are extremely subtle and short-lived, often obscured by background interference, lighting changes, or camera angle effects. Therefore, to ensure that the emotion recognition model focuses on real and effective facial information, the first step is face detection to remove background information unrelated to the face. In this study, a 68-key-point face detection algorithm model is used, which can quickly and accurately detect the position of the face and the coordinates of key points in the frame sequence. This detection method enables subsequent processing to be based on accurate facial structure operations, thereby improving the effectiveness of emotional feature extraction and helping to enhance the accuracy of emotion recognition of consumers in a natural state.

Furthermore, since consumers may show slight head shaking or posture changes unconsciously while watching advertisements, the faces in micro-expression image sequences may shift positions between different frames, affecting the model's judgment of temporal emotional changes. Especially under the premise that the amplitude of micro-expression movement itself is extremely small, if not processed, these non-expression changes in facial displacement may mask the real emotional variation features, interfering with model learning. Therefore, after detecting the face and its key points, this study introduces face alignment

operations to spatially align facial images and eliminate noise caused by inter-frame shaking. Face alignment can provide a stable reference structure for subsequent micro-expression dynamic feature analysis, effectively enhancing the model's ability to perceive small emotional changes. The adopted face alignment method is based on the principle of facial structural symmetry, selecting point 28 (nasal root) and point 31 (nasal tip) among facial key points to determine the facial central axis. Then, using the nasal tip point as the center of rotation, the image is rotated by a certain angle to align the central axis with the vertical line of the image, thereby achieving facial standardization. This method not only enhances the consistency of facial structure in the image but also reduces the impact of individual differences in angle and posture on the recognition result. Assuming the coordinates of the 31st and 28th key points on the face are  $(a_{31}, b_{31})$  and  $(a_{28}, b_{28})$  respectively, the angle  $\phi$  between the line connecting the two points and the vertical line is calculated by the following formula:

$$\phi = \text{ARCTAN} \left( \frac{a_{31} - a_{28}}{b_{28} - b_{31}} \right) \quad (1)$$

In the field of advertising marketing, consumers' micro-expressions are usually very subtle and change rapidly, so extracting effective facial features is key to improving recognition accuracy. To ensure that the model focuses on effective facial information, face cropping is an indispensable step in the preprocessing process. Based on the detected facial key points, such as the coordinates of the inner eye corners and nasal tip, a face-only image is cropped to effectively remove background information and irrelevant interference. This paper adopts a rectangular box cropping method, determining the cropping region size based on eye corners and nasal tip position to ensure that only the most representative facial area is retained in each image. Specifically, the coordinates of the two inner eye corners are denoted as  $r_1(a_2, b_2)$ ,  $r_2(a_3, b_3)$ , and the nasal tip coordinate is denoted as  $v(a_1, b_1)$ . These three points are used as markers for the cropping region. The face image is cropped using a rectangle with width  $Q=3|a_2-b_3|$  and height  $|G=3|b_1-(b_2+b_3)/2|$ . This operation maximally reduces noise interference from the background environment in micro-expression recognition and ensures the model focuses on facial detail changes.

Since each person's facial structure and shape are different, the cropped image sizes may vary, which brings certain challenges to the input requirements of the subsequent neural network model. To ensure that the neural network can handle input data of consistent size, each cropped face image needs to be size-normalized. In this study, a size of  $256 \times 256$  pixels is adopted as the standard output size. This operation not only ensures uniformity when the image is input into the neural network but also avoids calculation bias caused by different sizes. Moreover, uniform-sized images are conducive to batch processing and improve the training efficiency of the model. The consistency of image size after normalization helps the model learn facial expression patterns.

To enhance the visibility of consumers' subtle micro-expression changes, the Learning-based Video Motion Magnification (LVMM) algorithm is introduced into the micro-expression image preprocessing process. Through the LVMM algorithm, temporal dimensional changes in the image sequence are magnified, thus effectively enlarging the

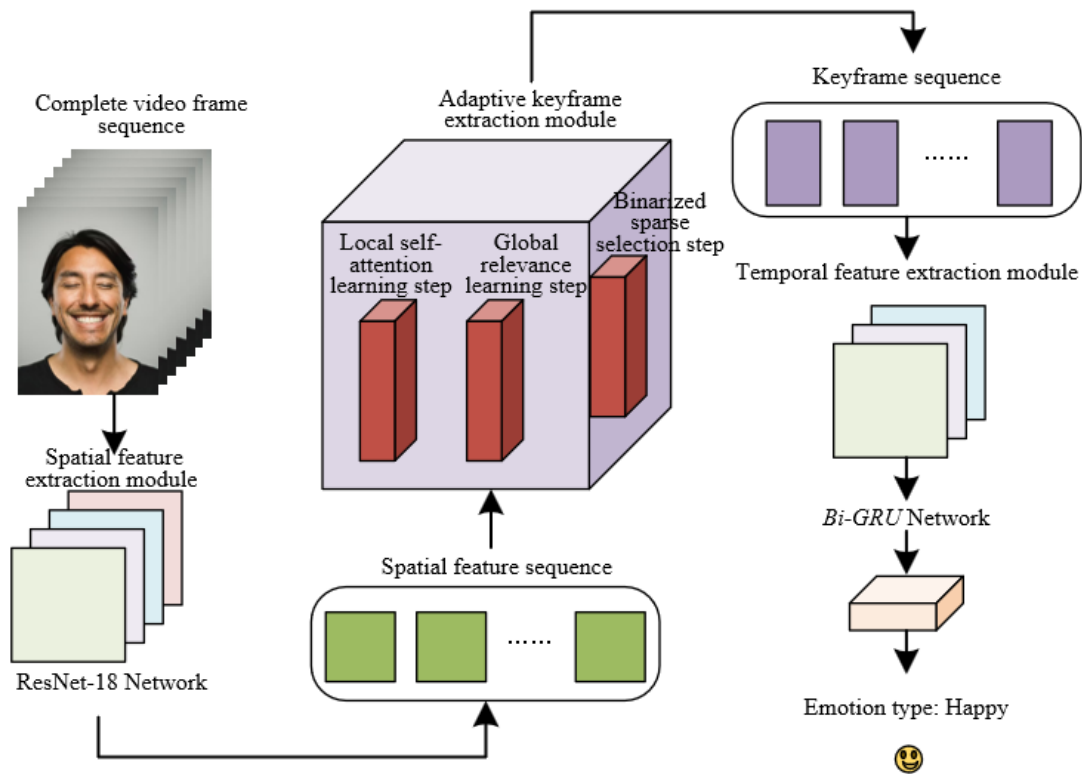
dynamic changes of facial micro-expressions, making it easier for the model to extract distinguishable features. Compared with traditional filtering techniques, this method does not require manually designed filters and can automatically learn more discriminative features through deep learning. In addition, motion magnification operations can highlight slight facial movement changes without introducing too much noise, thereby improving the model's sensitivity to consumer emotional responses. Therefore, motion magnification can significantly improve the accuracy of consumer emotion recognition, especially when consumers face advertising content, where micro-expression changes are often very subtle and happen instantaneously.

To further enhance the diversity of training data and address the problem of insufficient samples in the micro-expression dataset, this study expands the dataset by applying multiple magnification factor  $\alpha$  to the video for motion magnification. Different  $\alpha$  values adjust the magnification degree to obtain different effects, where smaller  $\alpha$  values can effectively avoid excessive noise amplification and ensure the data quality in the training set. In the experiment,  $\alpha$  values in the range of 2 to 5 were selected to ensure that the magnification does not introduce too much distortion. During dataset augmentation, the video frames are also rotated within a  $\pm 15^\circ$  range with a  $5^\circ$  increment to further increase data diversity. After such processing, the size of the training dataset will reach 35 times the original data, significantly improving the generalization ability of the model.

### 3. DESIGN OF MICRO-EXPRESSION RECOGNITION MODEL FOR CONSUMER EMOTION RECOGNITION

In application scenarios aimed at consumer emotion recognition, the recognition of micro-expressions requires

processing a large number of video frame sequences and accurately extracting the subtle emotional responses of consumers when faced with advertisements, product displays, or other marketing content. To achieve this goal, this paper proposes an adaptive key frame extraction model, which fully considers the characteristics of consumer emotion recognition. Through precise feature extraction and classification strategies, the model ensures effective capture of micro-expression changes in consumers. The model is divided into three modules: spatial feature extraction module, adaptive key frame extraction module, and temporal feature extraction module. Among them, the spatial feature extraction module processes the complete video frame sequence input and extracts the spatial features of each frame. These features reflect the basic structure and key details of facial expressions, providing strong support for subsequent emotion recognition. The temporal feature extraction module further processes the key frame sequence from the adaptive key frame extraction module to mine the temporal dimension information hidden in the frame sequence. This module can utilize the fusion of spatio-temporal features to fully exploit the temporal information in the video frame sequence and accurately identify the dynamic changes in consumer emotions. Consumers' emotional changes in advertising scenarios are usually instantaneous and subtle. Therefore, how to effectively capture and understand these emotional fluctuations is the key to model design. By fusing spatial and temporal features, the model ensures high sensitivity to consumer emotional changes. Through this structural design, the model can not only accurately extract micro facial expression changes related to emotions, but also accurately predict consumers' emotional responses to advertising content, providing reliable emotional data analysis support for marketing and advertisement optimization. Figure 1 shows the structural diagram of the micro-expression recognition model for consumer emotion recognition.

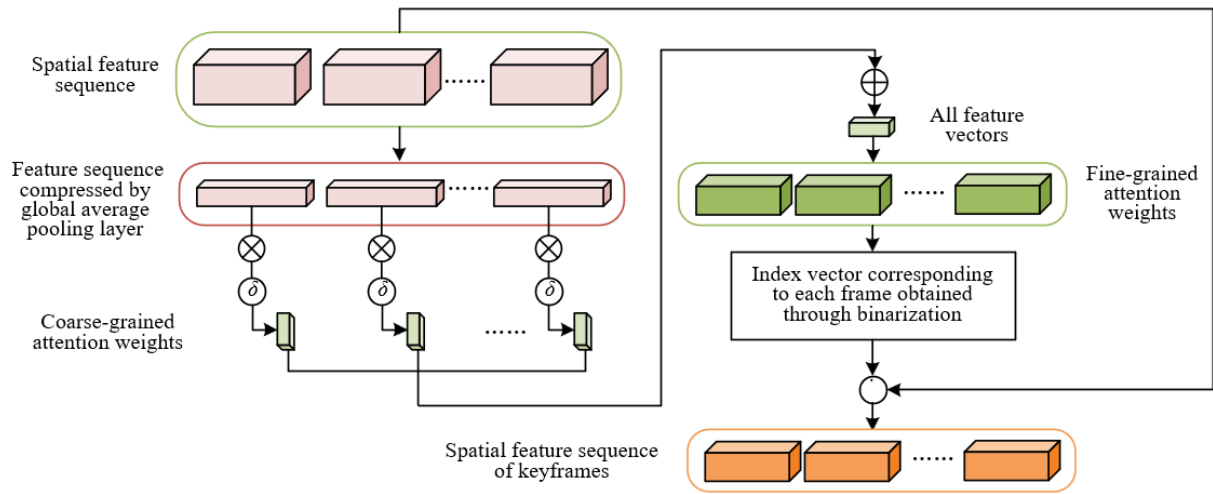


**Figure 1.** Structural diagram of micro-expression recognition model for consumer emotion recognition

In order to accurately capture subtle emotional changes in micro-expressions, this paper adopts the ResNet-18 network in the spatial feature extraction module. ResNet-18 has fewer parameters and a simplified structure, which allows it to effectively reduce the risk of overfitting and lower computational load and system resource consumption when facing relatively small micro-expression datasets. By iterating through each frame in the video sequence, ResNet-18 can extract identifiable facial spatial features for each frame, providing basic data for subsequent emotion recognition tasks. Since consumer emotional changes are often instantaneous, accurate extraction of spatial features from each frame is essential. To overcome the problem of insufficient micro-expression dataset samples, this study adopts a pre-training and fine-tuning strategy during training. Specifically, ResNet-18 is first pre-trained on the large-scale general image dataset ImageNet, and the trained network parameters are applied to the micro-expression recognition task. The purpose of this process is to transfer the feature extraction capability accumulated by ResNet in handling general image classification tasks to facial micro-expression recognition tasks, so that the model can still maintain good performance

under limited data conditions.

Efficiently and accurately extracting representative key frames from video not only helps reduce the interference of redundant information, but also improves the model's sensitivity to consumer emotional fluctuations. The adaptive key frame extraction module proposed in this paper, through three main steps, aims to adaptively extract key frames most related to consumer emotional changes from the video frame sequence, thereby providing accurate and concise data input for subsequent emotion recognition tasks. This module can be understood as learning a function mapping relationship:  $d_{AN}: D \rightarrow D^{KE}$ , where the input is the spatial feature sequence  $D$ . To reduce parameters and computational load while retaining the main features of the image during dimensionality reduction and compression, this paper sets a global average pooling layer within the module. After processing through this layer, a pooled feature sequence  $D'$  can be generated. The model compresses each feature map  $O_s$  in the spatial feature sequence  $D$  into a feature vector  $O$  containing more compact spatial feature information. Figure 2 shows the schematic diagram of the adaptive key frame extraction module.



**Figure 2.** Schematic diagram of adaptive key frame extraction module

Firstly, this paper introduces a local self-attention mechanism to assign a coarse-grained attention weight  $\beta_s$  to each frame feature vector  $O_s$ . The key purpose of this step is to evaluate the importance of each frame in the entire video sequence based on its spatial feature information. In consumer emotion recognition scenarios, certain moments of facial micro-expressions may more sensitively reflect emotional fluctuations, such as consumers' instantaneous reactions to specific scenes or product displays in advertisements. Through the local self-attention mechanism, the model can assign higher weights to frames that potentially reflect consumer emotional changes according to each frame's feature vector, thereby ensuring that the extracted key frames can reflect emotional fluctuations to the greatest extent. The coarse-grained attention weight of the compressed feature vector  $O_s$  for each frame can be calculated through a fully connected layer with a Sigmoid activation function. Assuming the Sigmoid activation function is denoted by  $\delta$  and the trainable weight matrix of the layer is denoted by  $Q^s$ , the calculation formula is:

$$\beta_s = \delta(Q^s \bar{O}_s) \quad (2)$$

Based on the coarse-grained attention weights obtained in the first step, by aggregating all feature vectors and their corresponding attention weights, a global feature vector  $\bar{O}$  can be obtained. The calculation formula for the global feature vector is:

$$\bar{O} = \sum_{s=1}^S \beta_s \bar{O}_s \quad (3)$$

Subsequently, by calculating the cosine similarity between each feature vector and the global feature vector, the fine-grained weight of each frame feature vector is learned. This process is called global correlation learning. In consumer emotion recognition tasks, emotional responses are often influenced by a series of external stimuli, such as a certain plot in the advertisement or a specific product display. Therefore, global correlation learning can help the model identify frames most related to emotional fluctuations throughout the video sequence. Through this step, the model can not only capture emotional information in a single frame image but also identify the important position of key frames on the time axis by comparing the correlation between frames, thereby

improving the accuracy of emotion recognition. Let  $\|\cdot\|_2$  represent the  $L2$  norm, the process of calculating the fine-grained weight  $\alpha_s$  of each feature vector  $O_s$  based on cosine similarity is expressed as follows:

$$\alpha_s = \frac{\bar{O}_s \times \tilde{O}}{\|\bar{O}_s\|_2 \times \|\tilde{O}\|_2} \quad (4)$$

$$\|a\|_2 = \sqrt{\sum_{u=1}^v a_u^2} \quad (5)$$

Furthermore, fine-grained weights are used for global sparsity filtering, and a binarized sparse vector  $Y$  is generated. This step uses fine-grained weights  $\alpha_s$  to screen out the most representative key frames, to reduce unnecessary computational burden and improve classification efficiency. The generation of the binarized index vector  $Y$  is done by calculating the fine-grained weight of each frame and converting it into binary form according to certain rules. Specifically, the model selectively retains the frames most critical to consumer emotion recognition based on the magnitude of these weights. In advertising and marketing scenarios, reducing redundant frames and extracting the most representative key frames can significantly improve the model's operational efficiency, especially in real-time analysis and large-scale data processing, significantly reducing computational cost and latency. The specific calculation formula is:

$$Y_s = \begin{cases} 1, \alpha_s > \frac{1}{S} \sum_{s=1}^S \alpha_s \\ 0, \alpha_s \leq \frac{1}{S} \sum_{s=1}^S \alpha_s \end{cases} \quad (6)$$

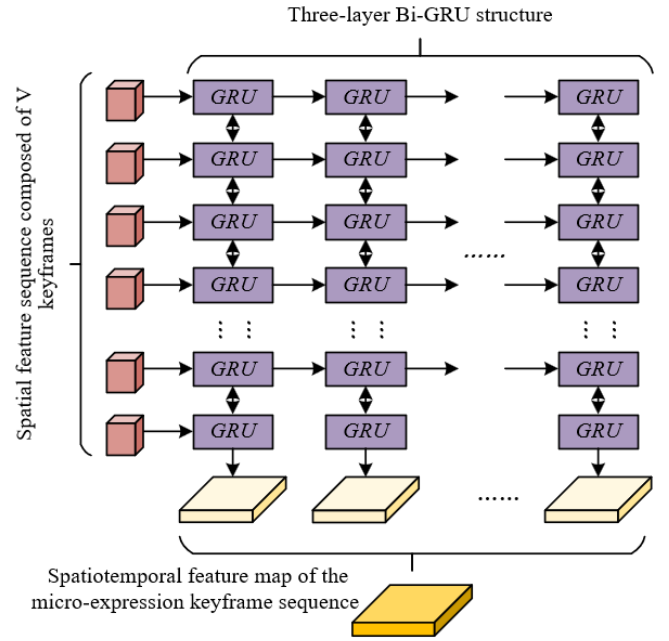
After  $Y$  is calculated, each video key frame index is marked as 1. Assuming the tensor product is represented by  $\otimes$ , the spatial feature sequence  $D^{KE}$  composed of key frames can be obtained through the following formula:

$$D^{KE} = D \otimes Y \quad (7)$$

Fluctuations in consumer emotions, especially in advertising and marketing scenarios, are usually manifested as rapid and short-lived facial micro-expression changes. These changes not only involve the expression of emotions at a single moment but also involve dynamic changes across multiple time points. Therefore, temporal feature extraction can effectively capture the temporal dimension information of emotional fluctuations, thereby improving the model's ability to recognize micro-expressions. To address this challenge, this chapter introduces a Bidirectional Gated Recurrent Unit (Bi-GRU) as the temporal feature extraction module, aiming to capture the time-series characteristics of consumers' emotional responses through a recurrent structure. Figure 3 shows the schematic diagram of the temporal feature extraction module.

In the model, the first step of introducing the temporal feature extraction module is to construct a spatial feature sequence containing only keyframes based on the adaptive keyframe extraction module, denoted as  $D^{KE} = \{O_{j1}, O_{j2}, \dots, O_{jV}\}$ , and input it into the Bi-GRU network for temporal

feature extraction and spatiotemporal feature fusion. The core goal of this step is to extract representative keyframes from the video sequence and pass the spatial feature sequence of these keyframes to the Bi-GRU network. Considering that micro-expression changes in consumer emotion recognition scenarios are often short and subtle, traditional inter-frame feature extraction methods are difficult to capture sufficient temporal information. However, by adaptively selecting keyframes, redundant data can be effectively reduced, computational efficiency can be improved, and important emotional information can be retained. Therefore, the keyframe extraction module can efficiently capture consumers' subtle emotional fluctuations while watching advertisements, product introductions, etc., thus providing accurate input for the temporal feature extraction module.



**Figure 3.** Schematic diagram of the temporal feature extraction module

Each layer structure of the Bi-GRU network operates on a feature pixel point sequence of size  $V \times Z$  and extracts temporal features from the same pixel point of all keyframes. Specifically, in this study, we use a three-layer Bi-GRU network structure to capture dynamic changes in the temporal dimension while recursively processing each frame image and to avoid information loss or gradient vanishing caused by too many layers. In consumer emotion recognition applications, consumers' slight facial expression changes in response to advertisements, promotional activities, etc., reflect their emotional reactions. Through the recursive processing of the Bi-GRU network frame by frame, the model can accurately capture this temporal information, thereby determining consumers' emotional attitudes toward different products or services. The temporal feature extraction module can improve the model's sensitivity to consumer emotional fluctuations by effectively processing the temporal information of all frames and make emotion recognition more accurate. For each layer structure of the network, at the spatial position  $(u, k)$  of the feature pixel points of all keyframes, a sequential input sequence along the time dimension is given:

$$D_{(u,k)}^{KE} = \{O_{j_1}^{(u,k)}, O_{j_2}^{(u,k)}, \dots, O_{j_V}^{(u,k)}\} \quad (8)$$



Assuming that the intermediate layer feature matrices are represented by  $c_v$  and  $g_v$ , the output  $g_v$  at the feature pixel point of the  $v$ -th keyframe is calculated as follows:

$$g_v = (1 - c_v) * g_{v-1} + c_v * \tilde{g}_v \quad (9)$$

Assuming the stacking operation of different feature vectors is represented by  $[]$ , the element-wise multiplication operation by broadcasting of matrices is represented by  $*$ , the  $SI$  activation function is represented by  $\delta$ , the  $\tanh$  activation function is represented by  $\tanh$ , the output feature of the previous keyframe is represented by  $g_{v-1}$ , and the trainable weight matrices of the Bi-GRU network are represented by  $Q_c$ ,  $Q_{g-}$ , and  $Q_e$ . The intermediate layer feature matrices  $c_v$  and  $g_v$  are calculated as follows:

$$c_v = \delta \left( Q_c \cdot \left[ g_{v-1}, O_{j_v}^{(u,k)} \right] \right) \quad (10)$$

$$\tilde{g}_v = \tanh \left( Q_g \cdot \left[ \delta \left( Q_e \cdot \left[ g_{v-1}, O_{j_v}^{(u,k)} \right] \right) * g_{v-1}, O_{j_v}^{(u,k)} \right] \right) \quad (11)$$

The outputs of the forward GRU and backward GRU of the Bi-GRU network in each layer structure are concatenated to generate the spatiotemporal feature map  $H$  of the micro-expression keyframe sequence, where  $Z$  represents the number of channels in the spatiotemporal feature matrix output by each layer of the Bi-GRU network. In consumer emotion recognition application scenarios, the spatiotemporal feature map obtained in this way not only contains information in the time dimension but also retains the details of the original spatial features. Therefore, the concatenated spatiotemporal feature map can effectively fuse temporal and spatial features, enabling the model to pay attention to both the dynamic changes of emotional expressions and the spatial distribution of facial expressions.

#### 4. ADVERTISING AND MARKETING STRATEGY ADJUSTMENT BASED ON CONSUMER EMOTION RECOGNITION RESULTS

In this paper, the core goal of emotion recognition is to accurately capture consumers' emotional responses while watching advertisements, thereby providing a scientific basis for optimizing advertising and marketing strategies. The emotion recognition results can be used as a decision support tool to adjust the content, form, and dissemination methods of advertisements, so as to improve the effectiveness of advertisements and consumer engagement.

Consumer emotion recognition results can provide data support for the adjustment of advertising content. If emotion recognition shows that most consumers display negative emotional responses while watching an advertisement, advertisers can further adjust the plot or visual elements in the advertisement by analyzing the consumers' emotional changes. For example, if a certain scene in the advertisement makes the audience feel uncomfortable or confused, advertisers can modify the presentation of that scene, change the way the product is displayed, or strengthen the emotional resonance points in the advertisement. Emotion recognition results can also reveal which elements in the advertisement are most likely to trigger positive emotions. Advertisers can increase the proportion of these elements or adjust the

advertising rhythm to enhance viewers' positive emotional responses. Figure 4 shows the micro-expressions corresponding to different emotions.



**Figure 4.** Micro-expressions corresponding to different emotions

Different consumer groups may exhibit differences in emotional responses. Emotion recognition technology can provide a basis for precision marketing. By analyzing emotional data, advertisers can also determine which groups generate strong emotional reactions under certain advertisement elements, and then implement customized advertisement delivery strategies. For example, for young consumer groups, if emotion recognition shows that they show positive emotional responses to humorous advertising elements, advertisers can push more creative and humorous advertisements to this group. For older consumer groups, more sentimental and emotional content may be needed to trigger resonance. Therefore, differentiated advertisement delivery based on emotional responses can significantly improve the targeting and effectiveness of advertisements.

Emotion recognition results can also help advertisers make more precise adjustments in terms of advertisement timing and frequency. If emotion recognition shows that consumers' emotional fluctuations are greater during a certain time period, it may mean that the audience's emotions are more sensitive or emotional responses are stronger during this time. Advertisers can choose to increase the advertisement delivery frequency during these periods to maximize emotional resonance opportunities. Conversely, if the emotional response is relatively flat, the effect of advertisement delivery may not be as expected. At this time, advertisers can consider adjusting the advertisement content or selecting more appropriate delivery timing to avoid resource waste.

Emotion recognition results can also help advertisers evaluate the effectiveness of advertisements and provide feedback for subsequent optimization. By monitoring consumers' emotional changes in advertisements in real time, advertisers can quickly assess the impact of the advertisement, understand which advertisement elements successfully triggered the audience's emotional response, and which elements failed to elicit the expected effect. For example, if consumers' emotions drop significantly during the final part of the advertisement, it may indicate that the ending of the advertisement fails to effectively attract the audience's attention. Advertisers can adjust the ending part of the advertisement based on this information to ensure that the audience maintains a positive emotional response throughout the viewing process. In this way, advertisers can continuously

optimize the advertisement content and form a dynamic adjustment mechanism to make advertising and marketing strategies more accurate and efficient.

Through consumer emotion recognition technology based on deep learning, advertisers can obtain real-time data about consumers' emotional responses and make fine adjustments to advertising and marketing strategies based on this data. This process can not only improve the appeal of advertisements and audience engagement, but also help advertisers gain greater commercial value in a competitive market. By continuously optimizing aspects such as advertisement content, dissemination methods, and delivery timing, advertisers can better meet consumer needs, enhance brand image, and maximize the effectiveness of advertising and marketing.

5. EXPERIMENTAL RESULTS AND ANALYSIS

From the confusion matrix data in Figure 5, the model's performance in recognizing consumers' micro-expressions shows certain characteristics. Among the various emotions, the correct prediction rate of "excited" is the highest, reaching 0.84; the correct prediction rate of "satisfied" is 0.81; "confused" is 0.77; "dissatisfied" is 0.78; and "indifferent" is relatively low, at 0.69. At the same time, there are some confusion phenomena between certain emotion categories. For example, the probability of "dissatisfied" being predicted as "indifferent" is 0.81, and the probability of "indifferent" being predicted as "dissatisfied" also reaches 0.81, indicating that these two types of emotions are relatively easy to confuse in model recognition. In addition, the probability of "indifferent" being predicted as "confused" is 0.13, and the probability of "confused" being predicted as "indifferent" is 0.05, showing a certain degree of overlap. According to the experimental results, the model is effective in recognizing consumers' micro-expressions, but there are deficiencies in distinguishing certain emotion categories. This is due to the fact that the micro-expression features of these two types of emotions are relatively similar, making it difficult for the model to accurately distinguish them during feature extraction and classification.

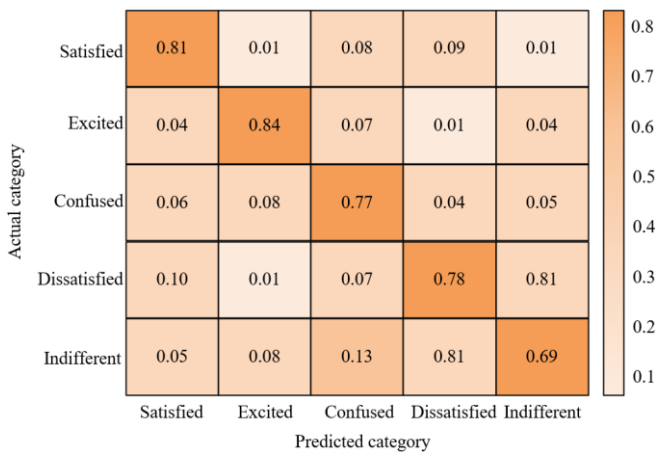


Figure 5. Confusion matrix of consumer micro-expression recognition by the proposed model

To improve model performance, it is possible to further optimize image preprocessing methods for the fine features of easily confused emotions to enhance the accuracy of feature extraction. Meanwhile, in the design of deep learning models,

the learning ability of the model for feature differences between easily confused categories can be strengthened to improve the accuracy of emotion classification. This is crucial for advertising and marketing strategy adjustment, as more accurate emotion recognition can ensure better alignment between advertisement content and delivery strategies and consumers' emotional changes, thereby maximizing the effectiveness of advertising and marketing.

Table 1. Experimental verification of the effectiveness of video motion magnification and data augmentation operations

Operation Combination	Recognition Accuracy (%)
Not using the method proposed in this paper	72.31
Using the method proposed in this paper	76.58

From the experimental results in Table 1, it can be seen that the recognition accuracy of micro-expression is 72.31% without using the video motion magnification and data augmentation method proposed in this paper; while after introducing the method, the recognition accuracy increases to 76.58%, with an improvement of 4.27 percentage points. This significant performance improvement indicates that the preprocessing strategy proposed in this paper has a positive effect on improving the overall performance of the micro-expression recognition model. Especially in application scenarios for consumer emotion recognition, this method can more effectively highlight subtle motion changes in micro-expressions and enhance the diversity of training samples, thereby enhancing the model's discriminative ability for micro-expressions. This improvement benefits from the video motion magnification technology enhancing the details of micro-expressions, making it easier for the model to capture subtle facial reactions of consumers when viewing products, advertisements, or service information. In addition, data augmentation operations improve the generalization ability of the model by simulating various expression changes, reducing its dependence on specific postures, lighting conditions, or groups of people, and thus providing stronger robustness in complex and variable real consumer usage scenarios.

Table 2. Experimental verification of the effectiveness of the adaptive key frame extraction step

Different Algorithm Models	Recognition Accuracy (%)
Removing the adaptive key frame extraction step	74.21
Method proposed in this paper	76.23

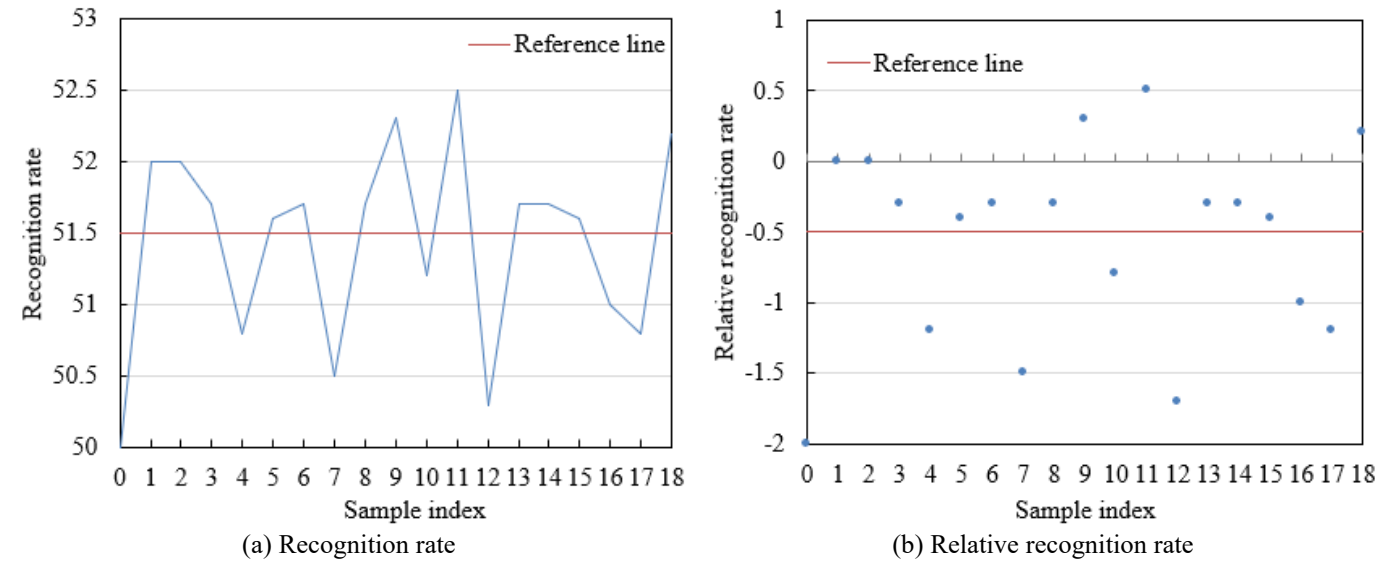
From the experimental results in Table 2, it can be seen that the recognition accuracy is 74.21% after removing the adaptive key frame extraction step, while using the adaptive key frame extraction method proposed in this paper increases the recognition accuracy to 76.23%, with an improvement of 2.02 percentage points. This shows that the adaptive key frame extraction step proposed in this paper has a significant effect on improving the accuracy of micro-expression recognition tasks. The adaptive key frame extraction technology can automatically select the most representative frames in a video sequence, thereby reducing redundant data and retaining important emotional change information, making the subsequent temporal feature extraction more accurate.

Analyzing this result, it can be inferred that the adaptive key frame extraction method effectively reduces unnecessary computational overhead by selecting frames with significant features, while better capturing key facial changes during emotional fluctuations of consumers. In consumer emotion

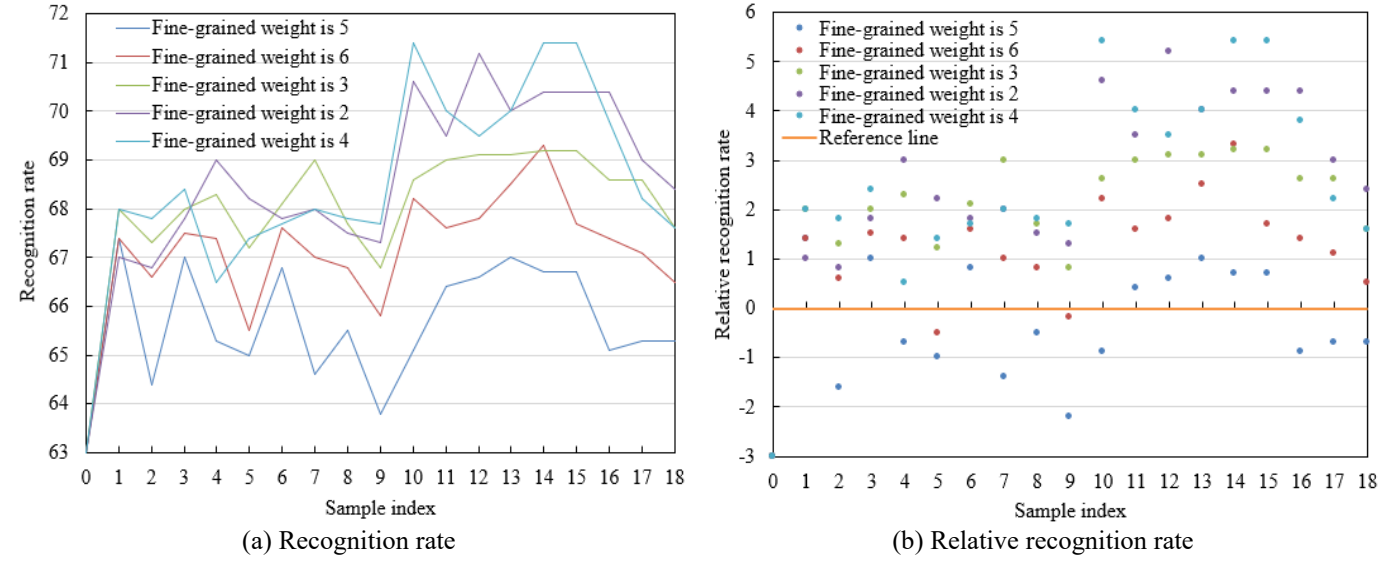
recognition application scenarios, micro-expressions are often transient and subtle. By adaptively selecting key frames, the method can avoid the interference of a large number of irrelevant frames that may affect the effectiveness of emotion recognition.

**Table 3.** Accuracy comparison of different models

Algorithm	Training Set		Validation Set		Test Set	
	<i>F1</i>	<i>UAR</i>	<i>F1</i>	<i>UAR</i>	<i>F1</i>	<i>UAR</i>
Non-linear <i>SVM</i>	0.7251	0.7652	0.5326	0.6231	0.4215	0.4216
<i>VGGNet</i>	0.8856	0.8541	0.6785	0.6542	0.5326	0.5268
<i>DenseNet</i>	0.8654	0.8569	0.6741	0.7158	0.6452	0.6741
<i>LSTM</i>	0.8213	0.8423	0.6652	0.6452	0.6231	0.7123
<i>AT-Net</i>	0.7895	0.7652	0.5412	0.5326	0.4895	0.4785
<i>SE - Net</i>	0.8542	0.8412	0.6523	0.6659	0.5756	0.5526
<i>CBAM</i>	0.7123	0.7126	0.5789	0.5785	0.6146	0.5869
Proposed Model	0.8236	0.8326	0.7125	0.7125	0.6723	0.7123



**Figure 6.** Line chart of consumer micro-expression recognition rate before image preprocessing and point chart of relative recognition rate



**Figure 7.** Line chart of consumer micro-expression recognition rate after image preprocessing and point chart of relative recognition rate

From the experimental results in Table 3, it can be seen that the proposed model shows relatively high performance in

terms of *F1* and *UAR* values on the training set, validation set, and test set. On the training set, the *F1* value of the proposed



model is 0.8236 and the UAR is 0.8326; on the validation set, the F1 value is 0.7125 and the UAR is 0.7125; on the test set, the F1 value is 0.6723 and the UAR is 0.7123. These results indicate that the proposed model maintains relatively stable performance across all datasets. Especially on the training set and validation set, the F1 and UAR values are close, showing strong generalization ability. Compared with other models, the proposed model performs better particularly on the test set, where it shows superior results in both F1 and UAR indicators compared to models such as Non-linear SVM, VGGNet, and AT-Net.

By comparing the performance of other models, it can be found that VGGNet and DenseNet perform well on the training and validation sets, but their F1 and UAR values drop on the test set, indicating a risk of overfitting. In contrast, the results of the proposed model on the training, validation, and test sets are more balanced and do not show signs of overfitting. Specifically, the proposed model achieves an F1 value of 0.6723 and a UAR of 0.7123 on the test set, which remains relatively high compared with other advanced models, proving its effectiveness in consumer emotion recognition tasks. In conclusion, the proposed model demonstrates strong robustness in micro-expression recognition tasks, adapts to variations across datasets, and has high practical value, especially in real-world applications such as consumer emotion analysis, where it can effectively recognize micro-expression details and provide more accurate emotion judgment.

Observing the line chart of consumer micro-expression recognition rate before image preprocessing shown in Figure 6, the recognition rate of sample numbers 1 to 18 fluctuates around 50, with the reference line at 51.5, and the fluctuation amplitude is large. In the relative recognition rate point chart, most of the points are below -0.5, showing poor stability and significant deviation. Looking at Figure 7, which shows the line chart of recognition rate after image preprocessing, the recognition rate under different fine-grained weights 5, 6, 3, 2, 4 is mostly above 63, significantly improved compared to before preprocessing. In the relative recognition rate point chart, the distribution of points is closer to the reference line, with some points above 0, indicating a significant improvement in stability. The experimental results show that image preprocessing has a significant impact on consumer micro-expression recognition. Before preprocessing, the recognition rate is low and fluctuates greatly, and the feature extraction effect is poor. After preprocessing, with different fine-grained weights applied, the recognition rate greatly improves and the stability increases, indicating that the improved image preprocessing method effectively optimizes the micro-expression feature extraction and enhances the recognition model performance. This provides a more reliable foundation for subsequent adjustments to advertising strategies based on emotional recognition, enabling more precise content and placement strategy optimization according to consumer emotional changes, thereby maximizing advertising marketing effectiveness.

Further analysis of Figure 7: The first chart is the consumer micro-expression recognition rate line chart, with the horizontal axis representing sample numbers and the vertical axis representing the recognition rate. Different color lines represent different fine-grained weights: blue (weight 5), red (weight 6), green (weight 3), purple (weight 2), and cyan (weight 4). As seen in the chart, the recognition rate fluctuates significantly under different weights, such as weight 5 (blue)

having a low recognition rate initially, followed by a fluctuating upward trend; weight 6 shows large fluctuations overall and a downward trend later. The second chart is the relative recognition rate point chart, with the vertical axis ranging from -3 to 6, and the orange reference line at 0. The points of different colors correspond to different weights, with weight 5's points distributed around the 0 line, indicating both positive and negative relative recognition rates; some points for certain weights are concentrated above the 0 line, showing relatively good recognition performance for some samples. From the recognition rate line chart, it can be seen that different fine-grained weights have a significant impact on micro-expression recognition rates, and the model's performance varies across different samples, suggesting that the improved preprocessing and deep learning model still have room for optimization in terms of stability. The relative recognition rate point chart further indicates that the recognition performance of each weight is superior to the reference in some samples, but inferior in others, reflecting an uneven adaptability of the model to different samples. This suggests that further research can focus on samples with large fluctuations, optimizing model robustness to improve recognition accuracy and stability, thus providing more precise support for advertising marketing strategy adjustments and maximizing marketing effectiveness.

## 6. CONCLUSION

This paper delved into the consumer emotion recognition technology based on deep learning and image processing and successfully applied it to advertising marketing strategies. In the first part of the study, this paper proposed an improved image preprocessing method aimed at enhancing micro-expression feature extraction, thereby improving emotion recognition accuracy. Through techniques such as adaptive keyframe extraction and video motion amplification, this paper effectively enhanced the prominence of key features in micro-expression images, addressing the limitations of traditional methods in extracting subtle facial expression changes. In the second part, this paper designed and optimized a deep learning-based micro-expression recognition model, using a Bi-GRU network to extract and fuse temporal features, achieving high recognition accuracy and improving emotion classification real-time performance. Through experimental comparisons with various models, the proposed model outperformed traditional methods on multiple datasets, demonstrating its outstanding performance in consumer emotion recognition tasks. Finally, based on emotion recognition results, this paper further explored methods for adjusting advertising marketing strategies, proposing a framework for dynamically adjusting advertising content and placement strategies based on consumer emotion changes, providing a new technological path for advertising marketing.

In summary, this research not only enriched the theoretical system of consumer emotion recognition but also provided an innovative and precise marketing method for the advertising industry, with significant practical value and social implications. However, this paper also has certain limitations, such as the scale and diversity of the dataset, which may affect the model's generalization ability in different scenarios, and the real-time performance of the model still needs further optimization, especially in terms of efficiency when handling large-scale data. Future research can further explore more

diverse emotion recognition datasets and combine more efficient algorithms, such as reinforcement learning or adaptive algorithms, to improve model computational efficiency and real-time response capabilities. Moreover, with the continuous development of advertising technology, integrating emotion recognition results with more complex consumer behavior data for cross-platform personalized advertising recommendations is also an important direction for future research.

## REFERENCES

- [1] Ahmed, Y.A.E., Yue, B., Gu, Z., Yang, J. (2023). An overview: Big data analysis by deep learning and image processing. *International Journal of Quantum Information*, 21(07): 2340009. <https://doi.org/10.1142/S0219749923400099>
- [2] dos Santos, C.F.G.D., Arrais, R.R., da Silva, J.V.S.D., da Silva, M.H.M.D., et al. (2025). ISP meets deep learning: A survey on deep learning methods for image signal processing. *ACM Computing Surveys*, 57(5): 127. <https://doi.org/10.1145/3708516>
- [3] Archana, R., Jeevaraj, P.E. (2024). Deep learning models for digital image processing: A review. *Artificial Intelligence Review*, 57(1): 11. <https://doi.org/10.1007/s10462-023-10631-z>
- [4] Kawulok, M., Benecki, P., Piechaczek, S., Hrynczenko, K., Kostrzewa, D., Nalepa, J. (2019). Deep learning for multiple-image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 17(6): 1062-1066. <https://doi.org/10.1109/LGRS.2019.2940483>
- [5] Alshouha, B., Serrano-Guerrero, J., Elizondo, D., Romero, F.P., Olivas, J.A. (2024). What is the consumer attitude toward healthcare services? A transfer learning approach for detecting emotions from consumer feedback. *Journal of Universal Computer Science*, 30(1): 3-24. <https://doi.org/10.3897/jucs.104093>
- [6] Sharma, A., Kumar, A. (2023). DREAM: Deep learning-based recognition of emotions from multiple affective modalities using consumer-grade body sensors and video cameras. *IEEE Transactions on Consumer Electronics*, 70(1): 1434-1442. <https://doi.org/10.1109/TCE.2023.3325317>
- [7] Benisha, S., Mirnalinee, T.T. (2023). Human facial emotion recognition using deep neural networks. *The International Arab Journal of Information Technology*, 20(3): 303-309.
- [8] Borza, D., Danescu, R., Itu, R., Darabant, A.S. (2017). High-speed video system for micro-expression detection and recognition. *Sensors*, 17(12): 2913. <https://doi.org/10.3390/s17122913>
- [9] Liong, S.T., See, J., Wong, K., Phan, R.C.W. (2018). Less is more: Micro-expression recognition from video using apex frame. *Signal Processing: Image Communication*, 62, 82-92. <https://doi.org/10.1016/j.image.2017.11.006>
- [10] Lalitha, S.D., Thyagarajan, K.K. (2019). Micro-facial expression recognition based on deep-rooted learning algorithm. *International Journal of Computational Intelligence Systems*, 12(2): 903-913. <https://doi.org/10.2991/ijcis.d.190801.001>
- [11] Kaya, Ü., Akay, D., Ayan, S.Ş. (2024). EEG-based emotion recognition in neuromarketing using fuzzy linguistic summarization. *IEEE Transactions on Fuzzy Systems*, 32(8): 4248-4259. <https://doi.org/10.1109/TFUZZ.2024.3392495>
- [12] Mazhar, A., Bailey, C.S. (2025). Emotion-specific recognition biases and how they relate to emotion-specific recognition accuracy, family and child demographic factors, and social behaviour. *Cognition and Emotion*, 39(2): 320-338. <https://doi.org/10.1080/02699931.2024.2408652>
- [13] Mualem, O., Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education*, 33(4): 413-425. <https://doi.org/10.1177/0255761415584292>
- [14] Cebula, K.R., Wishart, J.G., Willis, D.S., Pitcairn, T.K. (2017). Emotion recognition in children with Down syndrome: Influence of emotion label and expression intensity. *American Journal on intellectual and developmental disabilities*, 122(2): 138-155. <https://doi.org/10.1352/1944-7558-122.2.138>
- [15] Lindell, A.K. (2013). The silent social/emotional signals in left and right cheek poses: A literature review. *Laterality: Asymmetries of Body, Brain and Cognition*, 18(5): 612-624. <https://doi.org/10.1080/1357650X.2012.737330>
- [16] Veeranki, Y.R., Ganapathy, N., Swaminathan, R. (2022). Analysis of fluctuation patterns in emotional states using electrodermal activity signals and improved symbolic aggregate approximation. *Fluctuation and Noise Letters*, 21(2): 2250013. <https://doi.org/10.1142/S0219477522500134>
- [17] Rathee, N., Ganotra, D. (2017). Modelling facial features for emotion recognition and synthesis. *IETE Journal of Research*, 63(6): 845-852.
- [18] Fong, S., Lan, K., Wong, R. (2013). Classifying human voices by using hybrid SFX time-series preprocessing and ensemble feature selection. *BioMed research international*, 2013(1): 720834. <https://doi.org/10.1155/2013/720834>