Micro-Expression Recognition by Using Convolutional Neural Network

[1]*Jie-Yu He and* [2]*Chi-Chou Kao*

Department of Computer Science and Information Engineering,
National University of Tainan, Tainan, Taiwan
E-mail: [1]ellen90031616@gmail.com, [2]cckao@mail.nutn.edu.tw

ABSTRACT

Micro-expressions (MEs) are fleeting, involuntary facial expressions that last only a fraction of a second, often revealing a person's true emotions. Detecting and analyzing these MEs presents a significant challenge due to their brief and subtle nature. In this study, we propose an approach that leverages convolutional neural networks (CNNs) to train models on widely used micro-expression datasets, including CASME II, SAMM, and SMIC. These datasets provide a rich source of annotated micro-expression data, which allows for the development of more accurate recognition systems. To further enhance the system's performance, we incorporate rule-based corrections to address common misclassifications that occur during the recognition process. Our experimental results demonstrate that this combined approach of CNNs with rule-based post-processing yields substantial improvements in both accuracy and efficiency, compared to traditional methods. These advancements hold great promise for practical applications in various fields, such as psychological assessments, forensic investigations, security monitoring, and human-computer interaction, where understanding and interpreting subtle emotional cues is crucial for effective decision-making.

Keywords: *Micro-expression recognition, Deep learning, Convolutional neural networks, Emotion analysis.*

1. INTRODUCTION

Facial expression recognition plays a critical role in human-computer interaction by enabling machines to better understand human emotions and intentions. However, several challenges persist, including the need for large datasets, significant computational resources, and variations in model performance across different environments. This research proposes a novel approach that combines Emotion Classification Convolutional Neural Network (ECCNN) with an Emotion Classification Regular Library (ECRL) to enhance the accuracy and efficiency of recognizing seven emotions—Neutral, Happy, Surprise, Angry, Fear, Disgust, and Sad—especially in scenarios with limited sample data.

Building on the work of Dwijayanti et al. [1], this study addresses some of the ongoing challenges in facial expression recognition. While deep learning, particularly Convolutional Neural Networks (CNNs), has made significant advancements, CNNs still face issues such as the need for large datasets and high computational power, which can hinder real-time applications. Datasets such as SMIC (Li et al. [2]), CASME II (Yan et al. [3]), and SAMM (Davison et al. [4]) have been key to advancing research in micro-expression recognition, providing controlled environments to capture subtle facial movements, as initially described by Shi et al. [5]. Despite these advancements, challenges remain in accurately recognizing fewer common emotions like "Disgust" and optimizing models for faster processing speeds, as highlighted by Hammal [6].

Traditional facial expression recognition methods struggle significantly with micro-expression recognition due to the brief and subtle nature of these expressions. Conventional techniques often fail to capture these fleeting features accurately, leading to frequent misclassifications, particularly for rare emotions. Furthermore, these methods are computationally intensive, resulting in slower processing speeds that are not suitable for real-time applications. Therefore, improving both accuracy and processing efficiency is crucial for the continued progress of this field.

To overcome these limitations, this paper proposes a hybrid approach that combines an ECCNN-based emotion classification model with a rule-based system (ECRL). Building on the work of Han et al. [7], this method aims to address the shortcomings of traditional techniques, improving both the accuracy and efficiency of emotion recognition. Specifically, the proposed approach increases recognition speed by 4-5 milliseconds and boosts accuracy by 5-7%, meeting the critical requirements of both speed and precision for real-time emotion recognition systems.

The rest of the paper is organized as follows: the next section presents the proposed architecture, followed by an experimental evaluation in Section 3. Finally, Section 4 concludes the study with a summary of findings and future directions.

2. PROPOSED ARCHITECTURE

The proposed architecture combines a Convolutional Neural Network (CNN) with a rule-based system to improve the accuracy and efficiency of recognizing subtle and fleeting facial expressions, particularly for real-time applications. The system operates in six stages: First, image preprocessing converts color images to grayscale, reduces complexity, and enhances facial feature visibility through contrast adjustment and edge sharpening. Next, face detection uses the Dlib algorithm [8] to precisely locate and isolate facial regions for focused analysis. Facial image preprocessing then resizes, normalizes, and converts these regions to RGB format, ensuring standardized input for the CNN. The feature extraction stage leverages the Emotion Classification Convolutional Neural Network (ECCNN) to process the images through multiple layers—convolutional, pooling, local region, dropout, and fully connected layers—to capture detailed facial features necessary for emotion classification. Following this, the rule-based emotion classification step applies the Emotion Classification Regular Library (ECRL) to correct potential misclassifications, enhancing the system's robustness, particularly for less common emotions. Finally, result integration combines the CNN outputs with rule-based corrections using a decision strategy to optimize emotion classification. This hybrid approach boosts both the accuracy and efficiency of micro-expression recognition, making it well-suited for real-time applications where speed and precision are crucial.

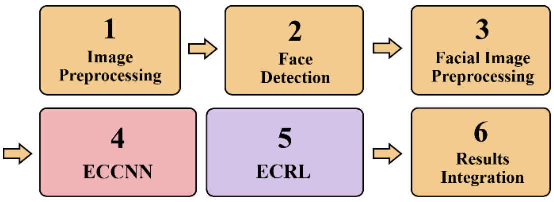

Fig. 1. Presents the overall system research process flowchart.

The Emotion Classification Convolutional Neural Network (ECCNN), shown in Figure 2, is the core component of the proposed system and plays a crucial role in addressing the challenges of micro-expression recognition. The ECCNN is designed to effectively detect and extract facial features associated with various emotions through several key components.

The first of these components is the convolutional layers, which are responsible for detecting various features within facial images, ranging from basic edges and textures to more complex patterns tied to specific emotions [9]. The architecture includes multiple convolutional layers, each followed by a pooling layer to reduce dimensionality while retaining essential features. The primary operation of each convolutional layer is defined by the following equation:

$$\boldsymbol{O(i,j)} = \sum_{m=0}^{F-1}\sum_{n=0}^{F-1} \boldsymbol{I(i+m,\ j+n)\cdot K(m,n)} \quad (1)$$

where $O(i, j)$ is the output feature map at position (i, j), I is the input image, K is the convolution kernel of size $F\times F$, and m, n are the indices over the kernel dimensions. This operation captures essential features such as edges and textures within the facial image.

Max pooling [10] is used to downsample the feature maps, effectively reducing computational complexity and mitigating the risk of overfitting. Pooling layers play a key role in maintaining the spatial hierarchy of features while highlighting the most significant ones. The reduction in the dimensionality of the feature maps is achieved through the following max pooling operation:

$$\boldsymbol{O_{pool}(i,j) = max\begin{Bmatrix} I(2i,2j), I(2i+1,2j), \\ I(2i,2j+1), I(2i+1,2j+1) \end{Bmatrix}} \quad (2)$$

where O_{pool} is the pooled output, and the operation reduces the size of the feature map while retaining the most prominent features. As shown in Figure 3,this operation involves selecting the maximum value within each window of the feature map, thus preserving the most significant features while reducing the overall size of the feature map.

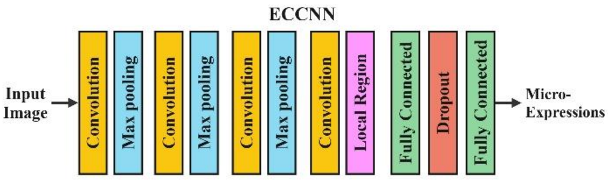

Fig. 2. Illustrates the architecture of the convolutional neural network (CNN) used for emotion classification.

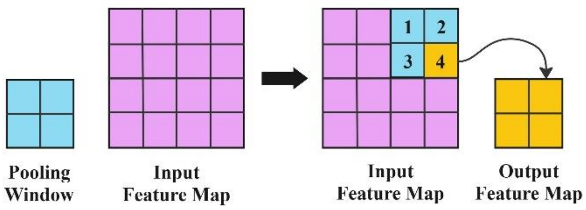

Fig. 3. Illustrates the schematic diagram of the Max Pooling operation.

While pooling layers are effective for reducing image dimensionality, they inevitably lead to some loss of information. This drawback is particularly significant in micro-expression recognition, where the critical features are often subtle and fine-grained. The downsampling process can discard these minute details, resulting in the model missing important information that is crucial for accurate classification. Consequently, each additional pooling layer can reduce recognition accuracy by approximately 2-4%, as finer features are ignored. However, this trade-off results in improved recognition speed, with each extra pooling layer reducing processing time by roughly 10 milliseconds.

To address this issue, the system incorporates a Local Region Layer, which partitions the feature map into smaller, localized regions and applies independent convolution operations within each region [11]. The main objective of this layer is to capture subtle, localized variations in facial features, enhancing the model's ability to detect fine-scale changes in micro-expressions. By focusing on these smaller regions, the Local Region Layer improves the accuracy of micro-expression recognition, helping to retain the delicate details that pooling layers might overlook.

The decision to divide the feature map into a 3×3 grid is informed by the natural distribution of key facial features, such as the eyes, nose, and mouth, which roughly divide the face into three distinct regions. This grid configuration allows the model to more effectively capture localized variations in these critical areas. Furthermore, setting the grid size to 3×3 strikes a balance between preserving essential facial details and minimizing computational load, ensuring efficient processing without sacrificing model performance.

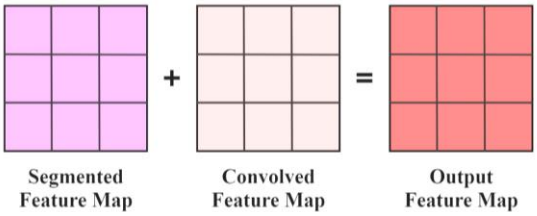

Fig. 4. Illustrates the schematic diagram of the residual connection.

Additionally, a residual connection within this layer, as illustrated in Figure 4, helps preserve important information from the original input, enhancing the model's ability to recognize complex emotional expressions. The residual connection combines the original input with the output of the localized convolution operations, ensuring that crucial features are maintained and effectively integrated into the final recognition process.

$$y = x + f(x) \tag{3}$$

where *x* represents the original input feature map, and *f(x)* represents the output after applying the convolution operation to each small region, and *y* is the resulting output feature map. This residual connection ensures that the network captures fine-grained details without losing the original information, thereby improving the accuracy of micro-expression recognition.

To reduce overfitting and enhance generalization, a dropout layer is incorporated into the model. Dropout randomly deactivates a fraction of neurons during training, forcing the network to learn more distributed and generalized features [12]. This helps the model better generalize to new, unseen data and mitigates the risk of overfitting. During training, each neuron is retained with a probability *p*, and the outputs are scaled by $\frac{1}{1-p}$ during inference. The dropout operation can be mathematically expressed as:

$$y = \frac{1}{1-p}(m \odot x) \tag{4}$$

where *m* is a binary mask vector with values drawn from a Bernoulli distribution with probability $1-p$, and $\odot$ denotes the element-wise multiplication. This technique forces the network to learn more robust features, reducing its dependency on specific neurons.

After feature extraction, the fully connected layers process the high-level features to create a feature vector that represents the input image [13]. This vector is then used for emotion classification. The feature maps are flattened into a one-dimensional vector, which is subsequently transformed using:

$$y = W \cdot x + b \tag{5}$$

where *y* is the output feature vector, *W* is the weight matrix, *x* is the input feature vector, and *b* is the bias vector. This transformation enables the network to combine the learned features and produce a final classification. In the proposed architecture, the final layer of the Convolutional Neural Network (CNN) employs the LogSoftmax function to convert the output logits into log-probabilities. The LogSoftmax function is mathematically defined as:

$$\boldsymbol{LogSoftmax}(\boldsymbol{z_i}) = \boldsymbol{log}\left(\frac{\boldsymbol{e^{z_j}}}{\boldsymbol{\sum_{j=1}^{N} e^{z_j}}}\right) \tag{6}$$

where z_i represents the logit corresponding to class *i*. This function is employed to ensure numerical stability by operating in the logarithmic domain, which mitigates potential issues such as overflow that may arise with the standard Softmax function. The log-probabilities produced by the LogSoftmax function are particularly advantageous when combined with the Negative Log Likelihood Loss (NLLLoss), as they align directly with the requirements of this loss function. The training of the proposed model is driven by the Negative Log Likelihood Loss (NLLLoss) function, which is ideal for classification tasks where the outputs are in the form of log-probabilities. The NLLLoss is defined as:

$$\boldsymbol{NLLLoss} = -\boldsymbol{\sum_{i=1}^{N} y_i \cdot LogSoftmax(z_i)} \tag{7}$$

where y_i denotes the true class label, and z_i represents the predicted log-probability for the corresponding class. This loss function penalizes the divergence between predicted log-probabilities and actual labels, encouraging higher probabilities for correct classes. The combination of LogSoftmax and NLLLoss ensures stable, well-calibrated outputs, leading to more accurate predictions.

To enhance the accuracy of the proposed micro-expression recognition system, the Emotion Classification Regular Library (ECRL) is integrated with the CNN. ECRL utilizes a rule-based approach based on facial landmark analysis [14], as illustrated in Figure 5, to refine the predictions of the CNN and correct misclassifications, particularly for less common emotions such as "Disgust." These rules are grounded in the pioneering work of Paul Ekman, a leading expert on human facial expressions. By leveraging Ekman's "Facial Action Coding System (FACS)," which offers a systematic framework for categorizing facial movements, we developed rules aimed at improving the recognition system's accuracy. Additionally, the rules are informed by empirical studies and datasets, including SMIC, CASME II, and SAMM, and focus on key geometric features like Eyebrow Relative Position Change, Eyebrow Width Ratio, and Mouth Curvature—critical factors for precise emotion classification.

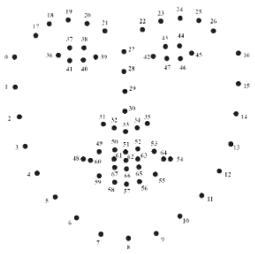

Fig. 5. Facial landmark localization.

The relative movement of the eyebrows is calculated by considering both the eyebrow width and height, normalized according to the overall dimensions of the face. This relationship is expressed as:

$$\textbf{Eyebrow Relative Position Change} = \frac{\textbf{Eyebrow Width and Height}}{\textbf{Face Width and Height}} \tag{8}$$

A greater change in the ratio indicates a higher likelihood of emotions like surprise and fear. The eyebrow width ratio is calculated as the ratio of the total eyebrow width to the overall width of the face:

$$\textbf{Eyebrow Width Ratio} = \frac{\textbf{Total Eyebrow Width}}{\textbf{Face Width}} \tag{9}$$

A smaller change in the ratio is more strongly associated with emotions such as Disgust and Sadness. The curvature of the mouth is an important feature for differentiating between positive and negative emotions. It is calculated as:

$$\textbf{Mouth Curvature} = \frac{\textbf{Mouth Height}}{|\textbf{Mouth Width}|} \tag{10}$$

A positive mouth curvature indicates an upward curve, typically associated with emotions like "Happy," while a negative curvature suggests a downward curve, often linked to emotions such as "Sad." During prediction, ECRL assesses the CNN's classification using predefined rules and makes adjustments when discrepancies are detected based on facial landmark movements. This process refines the results, improving the reliability of the system by combining the CNN's feature extraction capabilities with targeted corrections, ultimately leading to more accurate micro-expression recognition.

To enhance the accuracy of the proposed micro-expression recognition system, the predictions from the Emotion Classification Convolutional Neural Network (ECCNN) and the Emotion Classification Regular Library (ECRL) are combined. As illustrated in Figure 6, this integration ensures more robust final emotion recognition by applying different decision strategies based on the predicted emotion type.

Fig. 6. Emotion recognition system integration flowchart.

For positive emotions such as "Neutral," "Happy," or "Surprise," the system directly accepts the prediction generated by the ECCNN without further evaluation by the ECRL. This approach takes advantage of the ECCNN's high accuracy and efficiency in recognizing these emotions, thus avoiding unnecessary computational overhead. For negative emotions like "Angry," "Fear," "Disgust," or "Sad," the system initiates a secondary evaluation with the ECRL to verify the classification accuracy. If both the ECCNN and ECRL produce the same result, this classification is confirmed as the final output. In cases of discrepancies between the ECCNN and ECRL predictions, the system prioritizes the ECCNN's prediction, as it has been empirically shown to provide higher accuracy in most cases.

In a special case involving the emotion "Disgust," ECRL takes precedence over the ECCNN due to the limited representation of "Disgust" in the ECCNN training data, which increases the risk of misclassification. This ensures more reliable recognition, as ECRL has demonstrated superior performance for this particular emotion. By applying these strategies, the system achieves high recognition accuracy with efficient processing. The combination of the CNN's feature extraction capabilities and the ECRL's rule-based refinement offers a robust solution for recognizing complex emotions, enhancing the system's adaptability and accuracy, especially in challenging or ambiguous cases.

3. EXPERIMENTAL RESULTS

We conducted a series of ablation experiments on the SMIC, CASME II, and SAMM datasets to assess the impact of modifying the number of pooling layers in the ECCNN model. The results, detailed in Table 1, reveal a clear trade-off between recognition time and accuracy. Specifically, increasing the number of pooling layers led to a reduction in recognition time, measured in milliseconds, as expected due to the downsampling effect. However, this also resulted in a decrease in accuracy, measured as a percentage, indicating that while pooling layers reduce computational load and accelerate the recognition process, they also cause a loss of critical, fine-grained features. This loss of information, particularly for subtle facial expressions, diminishes the model's ability to accurately classify emotions, especially those involving small variations in facial features. These results highlight the importance of balancing the number of pooling layers in the network to achieve optimal performance—too few pooling layers can increase computational demand, while too many can sacrifice accuracy by discarding important details.

To further investigate the model's performance, we incorporated a Local Region Layer to focus on capturing localized variations in facial features. This layer was designed to address the limitations introduced by pooling layers, which tend to lose finer details by pooling information over larger regions. The results from the experiments, shown in Table 2, demonstrate a significant improvement in classification accuracy after adding the Local Region Layer. The layer's ability to focus on smaller, localized facial regions allowed the model to capture subtle changes in facial expressions, such as small eyebrow movements or slight mouth curvature, which are often critical for accurately recognizing micro-expressions. This enhancement in accuracy underscores the importance of preserving local facial information, which is particularly vital in the context of micro-expression recognition, where emotions are conveyed through very subtle and quick facial changes.

While the inclusion of the Local Region Layer led to a slight increase in recognition time—due to the additional computational steps required for processing localized regions—the improvement in accuracy justifies its inclusion in the model. In fact, the substantial gain in accuracy indicates that the benefit of capturing detailed, localized features outweighs the small increase in processing time. These findings suggest that for micro-expression recognition, it is essential to not only optimize the computational efficiency of the model but also ensure that subtle, localized features are preserved and effectively utilized for classification. The combination of the ECCNN with the Local Region Layer represents a more robust approach, balancing both speed and precision, which is crucial for real-time applications where both accuracy and processing speed are essential.

Table 1. Impact of different numbers of pooling layers on ECCNN's recognition time and accuracy.

Dataset / Architecture	SMIC		CASME II		SAMM	
	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)
ECCNN (Without Local Region Layer)	55	68	52.7	69	54	66
ECCNN	60.5	74	62	75	64.3	71

Table 2. Comparison of recognition time and accuracy with the inclusion of the local region layer.

Dataset / Architecture and Number of Pooling Layers	SMIC		CASME II		SAMM	
	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)
ECCNN (One Pooling Layer)	80.3	78	87.5	80	85	78
ECCNN (Two Pooling Layers)	70.4	76	75	78	75	74
ECCNN (Three Pooling Layers)	60.5	74	62	75	64.3	71

The proposed ECCNN model, integrated with the Emotion Classification Regular Library (ECRL), was subjected to a thorough evaluation across multiple well-established datasets, including SMIC, CASME II, and SAMM. The results, as shown in Table 3, demonstrate that the ECCNN + ECRL model significantly outperforms other state-of-the-art models, such as DRML [15], AlexNet [16], ConvNet [17], and LCN [18], in both recognition accuracy and processing time. These datasets represent a broad spectrum of facial expressions and micro-expressions, providing a comprehensive testbed for evaluating model performance. Across all datasets, the ECCNN + ECRL combination consistently achieved superior accuracy, along with faster recognition times, underlining its effectiveness in both real-time and high-accuracy applications.

In particular, the integration of the ECRL with the ECCNN contributed to significant improvements in emotion recognition, especially for subtle and less commonly recognized emotions. For instance, the model showed remarkable performance in identifying the "Disgust" emotion, which is often challenging due to its subtlety and underrepresentation in training data. The addition of the ECRL allowed the model to leverage facial landmark-based rules that refined the ECCNN's predictions, particularly for these harder-to-detect emotions. This made the model more robust, ensuring accurate predictions even with limited or imbalanced data, which is often a challenge in real-world applications of emotion recognition.

An important aspect of our findings is that the integration of ECRL only resulted in a minimal increase in recognition time—roughly 0.5 to 1 millisecond. This trade-off is negligible compared to the substantial gains in accuracy, particularly for underrepresented or more complex emotions like "Disgust" and "Fear." The ability to improve recognition performance with such a small computational cost makes the model highly efficient, ensuring it remains practical for real-time systems that demand both speed and precision.

Table 3 further highlights the superiority of the ECCNN + ECRL model, particularly in scenarios involving data imbalance. Traditional models often struggle with data imbalance, where some emotions are underrepresented in training datasets. However, by incorporating rule-based corrections from the ECRL, the ECCNN model was able to mitigate this issue, showing improved accuracy even in cases where certain emotions were less frequently represented in the training data. The rule-based approach of the ECRL helped refine predictions by compensating for the data imbalance, ensuring that the model did not disproportionately favor more common emotions at the expense of rarer ones.

In summary, the ECCNN + ECRL model not only outperforms existing models in terms of recognition accuracy and speed but also demonstrates its robustness in handling challenges such as limited data, data imbalance, and the recognition of subtle emotions. This combination of deep learning and rule-based refinement offers a powerful, adaptable solution for real-time micro-expression recognition, making it highly suitable for applications in fields like psychological analysis, forensic investigations, and human-computer interaction.

Table 3. Comparison of recognition time and accuracy between ECCNN + ECRL and other models.

Dataset / Architecture	SMIC		CASME II		SAMM	
	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)	Recognition Time (ms)	Accuracy (%)
DRML	104.7	68	105	70	106	64
AlexNet	65	61	67.3	63	69	60
ConvNet	70.4	59	72	60	72.5	55
LCN	302	63	300	65	303.7	60
ECCNN	**60**	**71**	**61**	**71**	**63.5**	**69**
ECCNN + ECRL	**60.5**	**74**	**62**	**75**	**64.3**	**71**

4. CONCLUSION

In this paper, we introduced a micro-expression recognition system that combines a Convolutional Neural Network (ECCNN) with a rule-based system (ECRL) to address the challenges of detecting subtle, fleeting facial expressions. Evaluations on the SMIC, CASME II, and SAMM datasets demonstrated significant improvements in both recognition accuracy and processing efficiency. The ECCNN+ECRL model outperformed state-of-the-art methods like DRML, AlexNet, and ConvNet in terms of classification accuracy and speed, showcasing the effectiveness of our hybrid approach.

A key innovation in this work is the use of local region layers and pooling layers. The local region layer captures fine-grained facial movements crucial for micro-expression recognition, while pooling layers reduce computational load, ensuring fast, accurate real-time performance. Additionally, the integration of ECRL enhanced the system's ability to accurately recognize "Disgust"—a challenging emotion with limited training data—by applying facial landmark-based rules to refine ECCNN predictions, improving the model's overall reliability and accuracy.

ACKNOWLEDGEMENT

The paper is supported by Ministry of Science and Technology, the grant number: NSTC 112-2813-C-024-014-E.

REFERENCES

[1] S. Dwijayanti, R. R. Abdillah, H. Hikmarika, Hermawati, Z. Husin, B. Y. Suprapto, "Facial Expression Recognition and Face Recognition Using a Convolutional Neural Network," in Proc. ISRITI, Indralaya, Indonesia, 2020, pp. 621-626.

[2] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietik€ainen, “A spontaneous micro-expression database: Inducement, collection and baseline,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.,2013, pp. 1-6.

[3] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, “CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., 2013, pp. 1-7.

[4] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, “SAMM: A Spontaneous Micro-Facial Movement Dataset,” IEEE Trans. Affective Comput., vol. 9, no. 1, pp. 116-129, Mar. 2018.

[5] S. Shi, G. An, Q. Ruan, "Multi-Scale Region with Local Relationship Learning for Facial Action Unit Detection," in Proc. ICSP, 2020, pp. 252-253.

[6] Z. Hammal, "Facial features segmentation, analysis and recognition of facial expressions by the Transferable Belief Model," in Proc. 2006 Int. Conf. on Development and Learning and Epigenetic Robotics, Lausanne, Switzerland, 2006, pp. 82-88.

[7] G. Han, M. Duan, L. Liu, "Person Re-Identification by Deep MAX Pooling Network," in Proc. 2017 10th Int. Congr. Image Signal Process., BioMedical Eng. Informatics (CISP-BMEI 2017), Nanjing, China, 2017, pp. 1-6.

[8] S. Sharma, S. Karthikeyan, and S. Kumar, “FAREC - CNN Based Efficient Face Recognition Technique using Dlib,” in Proc. ICACCCT, Chennai, India, 2016, pp. 192-195.

[9] J. Zou, X. Cao, S. Zhang, B. Ge, "A Facial Expression Recognition Based on Improved Convolutional Neural Network," in Proc. IEEE ICIASE, 2019, pp. 301-304.

[10] Y. More, K. Dumbre, B. Shiragapur, "Horizontal Max Pooling a Novel Approach for Noise Reduction in Max Pooling for Better Feature Detection," in Proc. ESCI, Pune, India, Mar. 2023, pp. 20-23.

[11] S. Cho, J. Lee, "Learning Local Attention With Guidance Map for Pose Robust Facial Expression Recognition," IEEE Access, vol. 10, pp. 85929-85932, Aug. 2022.

[12] M. Yusoff, A. Salim, and H. Arof, "Effect of Dropout Layer on Classical Regression Problems," in Proc. 2017 IEEE Int. Conf. on Artificial Intelligence and Data Processing, Malatya, Turkey, 2017, pp. 1-4.

[13] Z. Xu, Y. Lu, W. He, "Revisit Linear Transformation for Image Privacy in Machine Learning," in Proc. 5th IEEE Int. Conf. TPS-ISA, 2023, pp. 156-160.

[14] Z Hammal, Segmentation des traits du visage, analyse et reconnaissance dexpressions faciales par le Modèle de Croyance Transférable. PhD Thesis in Cognitive Science, University of Joseph Fourier, 2006.

[15] M. Chen, P. Wang, B. Bennamoun, H. Liu, R. Zhao, and M. El-Sallam, "Deep region and multi-label learning for facial action unit detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Honolulu, HI, USA, 2017, pp. 3391-3399.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 2012, pp. 1097-1105.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[18] A. Coates, H. Lee, and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," in Proc. Int. Conf. Mach. Learn., Bellevue, WA, USA, 2011, pp. 215-223.