Micro-Expression Recognition Base on Optical Flow Features and Improved MobileNetV2

Wei Xu[1#], Hao Zheng[2#*], Zhongxue Yang[2, 3], and Yingjie Yang[4]
[1] College of Computer Science and Information Engineering, Guangxi Normal University
Guilin, 541004 China
[e-mail: hsuwei@stu.gxnu.edu.cn]
[2] School of Information and Engineering, Nanjing XiaoZhuang University
Nanjing, 211171 China
[e-mail: zhh710@163.com]
[3] Nanjing University Of Aeronautics And Astronautics
Nanjing, 211106 China
[e-mail: young-sir@vip.sina.com]
[4] Centre for Computational Intelligence, De Montfort University
Leicester, Leicestershire UK
[e-mail: yyang@dmu.ac.uk]
[*]Corresponding author: Hao Zheng

Received February 23, 2021; revised April 18, 2021; accepted May 16, 2021; published June 30, 2021

Abstract

When a person tries to conceal emotions, real emotions will manifest themselves in the form of micro-expressions. Research on facial micro-expression recognition is still extremely challenging in the field of pattern recognition. This is because it is difficult to implement the best feature extraction method to cope with micro-expressions with small changes and short duration. Most methods are based on hand-crafted features to extract subtle facial movements. In this study, we introduce a method that incorporates optical flow and deep learning. First, we take out the onset frame and the apex frame from each video sequence. Then, the motion features between these two frames are extracted using the optical flow method. Finally, the features are inputted into an improved MobileNetV2 model, where SVM is applied to classify expressions. In order to evaluate the effectiveness of the method, we conduct experiments on the public spontaneous micro-expression database CASME II. Under the condition of applying the leave-one-subject-out cross-validation method, the recognition accuracy rate reaches 53.01%, and the F-score reaches 0.5231. The results show that the proposed method can significantly improve the micro-expression recognition performance.

Keywords: Micro-expression, MobileNetV2, optical flow, SVM

Wei Xu and Hao Zheng contributed equally to this work. This work has been funded by National Natural Science Foundation of China (Grant No. 61976118).

http://doi.org/10.3837/tiis.2021.06.002 **ISSN : 1976-7277**

1. Introduction

Micro-expressions are the facial muscle movements that people involuntarily reveal when they try to hide their true emotions, and are inherited by human beings through a long period of evolutionary inheritance. Due to the hidden nature of micro-expressions and the tendency to confuse them with macro-expressions, micro-expressions went unnoticed until 1966, when Haggard and Isaacs [1] first discovered their existence. Subsequently, in 1969, Ekman and Friesen [2] analyzed a video interview with a depressed patient, Mary, and found that the patient was generally optimistic when asked by the doctor, but when the doctor asked Mary about her future life plans, a painful expression appeared on Mary's face that lasted for only two frames of about 1/12th of a second. Short expressions are defined as micro-expressions. As a true expression of a person's internal emotions, micro-expressions can reveal a person's true emotions and have such characteristics as short duration, low intensity, and difficult to induce [3], making them one of the most reliable physiological characteristics in the fields of psychology and emotion analysis. However, micro-expression recognition remains extremely difficult for trained individuals. Likewise, in the computer field, the current research methods in micro-expression recognition are mainly concentrated in the traditional machine learning field, and the recognition rate of the existing methods is still quite a long way from the macro-expression recognition, so they cannot reach the practical application requirements, so it is a very challenging and extraordinary task to research the micro-expression algorithm to improve its recognition accuracy.

Current research work mostly uses micro-expression video sequences for recognition. This operation is not only computationally complex and requires long computational processing time, but it is also difficult to obtain complete micro-expression video sequences for classification in practical application scenarios, which has certain limitations. In response to this problem, we propose a new approach. The main work of this study is as follows:

1) Selection of the onset frame and the apex frame from micro-expression sequence and applying TV-L1 optical flow techniques to extract the facial muscle movement features form them.

2) Proposal of an improved MobileNetV2 [4] that uses SVM as the classifier to improve micro-expression recognition.

3) Comprehensive evaluation of the proposed method [5] is performed on CASME II database to validate its effectiveness and generalizability.

2. Related Work

Micro-expression recognition algorithms are mainly divided into five categories: feature description algorithms, feature transformation algorithms, frequency domain algorithms, optical flow algorithms, and deep learning algorithms.

The feature description algorithm achieves the feature representation of micro-expressions by describing the facial muscle movement characteristics and texture characteristics of the

micro-expressions, and reduces the influence of noise, illumination and other factors on the basis of local binary pattern (LBP) to improve the accuracy of feature description. Zhao et al. [6] proposed a local binary mode (LBP-TOP) of three orthogonal planes, extracting LBP features from the video XY plane, XT plane, and YT plane respectively, and stitching them together to obtain the final feature expression. Wang et al. [7] proposed the Local binary pattern with six intersection points (LBP-SIP), which uses the unique different points on the three intersection lines on three orthogonal planes to calculate the space-time pattern, which improves the efficiency of feature extraction by reducing the dimension of features to, and its processing speed is 2.8 times that of LBP-TOP.

The feature transformation algorithm treats the micro-expression sequence as a tensor and performs matrix transformation to remove redundant information and add information such as color, space and time. Wang et al. [8] proposed tensor independent color space (Tensor independent color space, TICS). Since the three channel components of an RGB-encoded image are highly correlated, the features extracted from it may not bring the improvement effect. In another work [9], two color spaces, CIELab and CIELuv, were tried to improve the effect of recognition.

The frequency domain algorithm treats the micro-expression sequence as a time domain signal and converts it to the frequency domain by Fourier transform, Gabor transform, etc. to extract features such as amplitude and phase information, which can effectively extract local features such as corner points and edge information of facial contours, mainly by Riesz wavelet [10] and Gabor transform [11][12].

The optical flow algorithm mainly analyzes the changes of micro-expression sequences in the optical flow field, extracts the relative motion of two adjacent frames of pixels from the pixel's perspective, captures the subtle changes of the face, and reduces the effects of head motion and light changes on the features. Xu et al. [13] proposed a facial dynamics map (FDM) to calculate the dense optical flow field, which effectively removes the errors caused by facial translation, but its bottleneck is that the computational volume is too large and the number of feature dimensions is high. Liu et al. [14] proposed the Main directional mean optical flow feature (MDMO) to locate the key points of the face using the DRMF model [15], and to segment the face into 36 non-overlapping action unit-based regions of interest when extracting features, and on each of them Extract the main direction. The resulting vector can be modeled using a support vector machine.

The deep learning algorithms use features autonomously learned by neural networks to obtain high-level semantic information, which is significantly distinct from traditional algorithms [16-18]. Deep learning has been employed in various fields with great results [19-24]. Patel et al. [25] first explored deep learning for micro-expression recognition tasks, and proposed a selective deep feature for micro-expression recognition to remove redundant features. Peng et al [26] proposed Dual temporal scale convolutional neural network (DTSCNN), which uses different network streams to adapt the micro-expression video sequences with different frame rates to avoid the overfitting problem. At the same time, the optical stream sequences are added to the shallow network, which allows the network to obtain high level features. Verma [27] et al. proposed the Lateral Accretive Hybrid Network (LEARNet) to capture the micro-features of facial expressions, adding accretion layers (AL)

to the network.

3. Proposed Algorithm

The proposed algorithm contains the following three steps:

1. Preprocessing: to fetch onset frames and apex frames from micro-expression video sequences, which are then face aligned and cropped.

2. Optical flow features elicitation: to extracting the motion characteristics caused by facial muscles when a person shows micro-expressions.

3. Classification using Improved MobileNetV2: to improve the expressiveness of the optical flow features extracted in the second step, it allows further learning of relevant spatiotemporal context information.

A simplified flowchart of the proposed method is presented in **Fig. 1**. The following subsections will introduce each step in detail one by one.

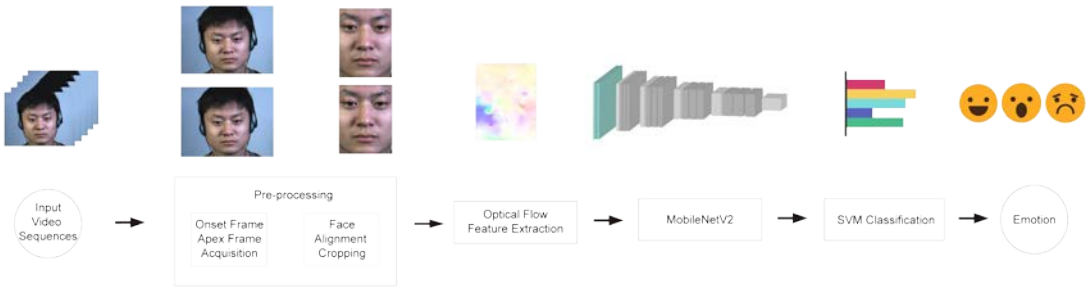

Fig. 1. Simplified flowchart

3.1 Preprocessing

Before face alignment and cropping, the onset frame and apex frame must be positioned first. The position of the corresponding frame has been marked in the CASME II database, and the corresponding marked frame in the data set can be directly taken out. In the image pre-processing process, OpenCV and Dlib libraries are used to detect and annotate 68 key features of the face. Rectangles are drawn to crop onset and apex frames based on the detected key points. Since there are various postures and angles in the face images collected by the camera, the face alignment is used to make them more suitable for micro-expression recognition. The left inner-eye coordinate (x_1, y_1) and the right inner-eye coordinate (x_2, y_2) are selected among the marked feature key points, and the horizontal face angle θ and the human-eye distance *dist* are calculated as shown in (1). Taking the midpoint of the face-eye distance as the center of rotation (x_0, y_0), the coordinates of the rotation are calculated as shown in (2). The preprocessing display diagram is shown in **Fig. 2**.

$$\begin{cases} \theta = \arctan[(y_2 - y_1)/(x_2 - x_1)] \\ dist = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \end{cases} \tag{1}$$

$$\begin{cases} x' = (x - x_0)\cos\theta + (y - y_0)(-\sin\theta) + x_0 \\ y' = (x - x_0)\sin\theta + (y - y_0)\cos\theta + y_0 \end{cases} \tag{2}$$

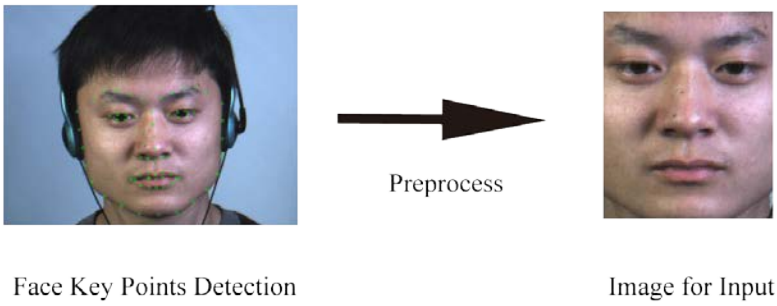

Fig. 2. Preprocessing display diagram

3.2 Optical Flow Features Elicitation

The optical flow method is usually based on two premise assumptions:

1. The illumination energy of the front and rear frames remains unchanged, that is, the brightness and all channel values are constant;

2. The movement of the same pixel between adjacent frames is small.

The micro-expression change range is low, the relative displacement between frames is small, and there are similar motions between adjacent pixels; moreover, the micro-expression duration is very short, the interval from the onset frame to the apex frame is only 1/25~1/5 seconds, the brightness of the image between frames basically does not change, so the properties of micro-expression dictate that its video sequences satisfy exactly the above two basic assumptions [28].

Suppose the light energy of a point in the first frame is expressed as $I(x, y, t)$, where (x, y) is the pixel coordinate, t is time, and after time dt, the point has moved dx, dy distance. The intensity of the two adjacent frames is achieved as:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{3}$$

By performing Taylor expansion on (3), it transforms into the form:

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt \tag{4}$$

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt = 0 \tag{5}$$

By dividing both sides of the equations by *dt*:

$$I_x u + I_y v + I_t = 0 \tag{6}$$

$$I_x = \frac{\partial I}{\partial x}; I_y = \frac{\partial I}{\partial y} dy \tag{7}$$

$$u = \frac{dx}{dt}; v = \frac{dy}{dt} \tag{8}$$

where *u* and *v* are the horizontal and vertical component of the optical flow field respectively. The constant brightness assumption only yields a constraint equation that is not sufficient to solve for the displacements *u* and *v* in the *x* and *y* directions, which is the pathological problem of solving for optical flow fields. Therefore, additional constraints on the displacement field vectors are required to solve the optical flow field model. Horn et al. [29] add a smoothness constraint to the optical flow constraint, which assumes that the velocity of the object's motion is locally smooth in most cases. In particular, when the target is in a rigid motion without deformation, each adjacent pixel should have the same velocity, i.e. the rate of change of velocity of the adjacent points is 0. Thus, the solution of the optical flow field is transformed into an energy generalized minimal value problem:

$$\min_{u,v} \left\{ E_{\text{H-S}} = \int_\Omega \left(|\nabla u|^2 + |\nabla v|^2 \right) d\Omega + \lambda \int_\Omega \left(I_x u + I_y v + I_t \right)^2 d\Omega \right\} \tag{9}$$

The first term penalizes large variations in the optical flow field in order to obtain a smooth velocity field. The second term is the data term, i.e. the basic optical flow constraint, which assumes constant grey values before and after the motion of the corresponding point. λ is the regularization parameter, which is a weighting parameter associated with the regular and data terms. Since the regular term of the H-S model adopts smoothness constraints, the discontinuity of the displacement field cannot be maintained, which will cause severe blurring and loss of important information during the image evolution. To overcome the disadvantages of the H-S model, Pock et al. [30] proposed a TV-L1 optical flow model based on a variational approach, and improved the optical flow constraint by introducing a variable *w*. The modeling of light variation by *w* was obtained:

$$\min_{u,v,w} \left\{ E_{\text{TV-L1}} = \int_\Omega \left(|\nabla u| + |\nabla v| + |\nabla w| \right) d\Omega + \lambda \int_\Omega |\rho(u,v,w)| d\Omega \right\} \tag{10}$$

where $\rho(u,v,w) = I_x(u-u_0) + I_y(v-v_0) + I_t + \beta w$. The parameter β is a weighting factor for the light change term. Although the changes are small compared with the H-S model, the registration accuracy has been greatly improved. First, the total variation regular term keeps the discontinuity of the displacement field and protects the edge information from being blurred during the diffusion process. This regular term after replacement is the same as the regular term of the famous ROF (Rudin, Osher, Fatemi) model [31], and the ROF model has a good denoising effect while keeping the edge information from being blurred, precisely because of the Non-quadratic regular term. Secondly, the data items are less sensitive to changes in luminance than the H-S model using a robust L1 parametrization [32].

In order to be able to improve the robustness of the method as much as possible, the optical flow technique adopted for the method in this paper is TV-L1 [33] method described above.

Fig. 3 shows some typical visualizations of optical flow features, which are calculated by the TVL1 method for various emotion samples. The approximate locations of key points on a face, such as the eyes, eyebrows, mouth, nose, cheeks, etc., can be identified in most visualizations through human eye observation. The facial muscle movements of 'happiness' emotion are mainly concentrated in the cheek area, 'disgust' emotion in the eyebrow area, 'repression' emotion in the mouth area, 'surprise' emotion in the eye area, and 'others' emotions in the forehead area. However, not all samples followed the above pattern.

Fig. 3. Visualization of optical flow features

3.3 Classification Using Improved MobileNetV2

MobileNetV2 is Sandler's proposed neural network architecture for facial attribute detection. It has been trained and evaluated on Google's internal database [34] and can be applied to mobile devices. Its main innovations are the introduction of inverted residuals and linear bottlenecks, and the realization of recent results balancing inference time and performance for generic benchmarks such as ImageNet [35], COCO [36] and VOC [37]. The MobileNetV2 network structure is shown in **Table 1**. Compared to most operational approaches our training approach is simple and straightforward: We do not perform a sophisticated alignment on the network architecture, just replace the classifiers.

Table 1. MobileNetV2 network structure: l represents the expansion coefficient of the intermediate convolution channel; o represents the number of output channels; r represents the layer repeated several times; s represents the stride of the convolution; b represents the bottleneck

Input	Operator	l	o	r	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	c	1	16	1	1
$112^2 \times 16$	b	6	24	2	2
$56^2 \times 24$	b	6	32	3	2
$28^2 \times 32$	b	6	64	4	2
$14^2 \times 64$	b	6	96	3	1
$14^2 \times 96$	b	6	160	3	2
$7^2 \times 160$	b	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1

$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

MobileNetV2 uses the SoftMax classifier. SoftMax regression is an improvement on Logistic regression, which is used for biclass problems, and SoftMax is used for multiclass problems. When a given sample is inputted, it will output a value between 0 and 1, which represents the probability that the input sample belongs to that class. For expression image **x**, the probability of its class *j* is

$$p(y^{(i)} = j \mid \mathbf{x}^{(i)}, \theta) = \frac{e^{\theta_j^T \mathbf{x}^{(i)}}}{\sum_{i=1}^{k} e^{\theta_j^T \mathbf{x}^{(i)}}} \tag{11}$$

where $p(y^{(i)} = j \mid \mathbf{x}^{(i)}, \theta)$ is the probability that expression image **x** corresponds to each expression category *j*, and θ is the parameter to be fitted. Ultimately, the category with the highest probability value is the final result of the neural network's predictive classification. The characteristics of micro-expressions determine that the distinction between classes of different micro-expressions is not high. In this situation, since Softmax layer minimizes cross-entropy from the global point of view, it will probably lead to misclassification once there are negative samples with large differences. So it is not appropriate to use Softmax classifier for micro-expression recognition. To solve this problem, SVM is used as the classifier. SVM tries to find the maximum margin between data points of different categories. It has better differentiability, and the regularization term penalizes the wrongly scored data more strongly, with strong generalization capabilities, thus facilitating the differentiation of micro-expression features. Its Hinge loss function is defined as:

$$\text{loss}(\mathbf{x},\text{y}) = \frac{1}{N} \sum_{i=1, i \neq y}^{N} \max(0, (\text{margin} - \mathbf{x}_y + \mathbf{x}_i)^p) \tag{12}$$

where $1 \leq y \leq N$ denotes the label.

4. Experiments

4.1 Database

The CASME II micro-expression database provided by Fu Xiaolan was used in the experiment, including 255 micro-expression video samples. The sampling rate was 200 frames/s, and the resolution was 280×340 pixels. 26 subjects were collected from Asia. Micro expressions were classified into seven categories: disgust (63 samples), happiness (32 samples), surprise (28 samples), repression (27 samples), sadness (4 samples), fear (2 samples) and others (99). Because the number of samples of sadness and fear is too small, this experiment only uses 249 samples from the other five classes, and the composition distribution is shown in **Fig. 4**. It can be seen that the number of ‘others’ samples accounts for a large part.

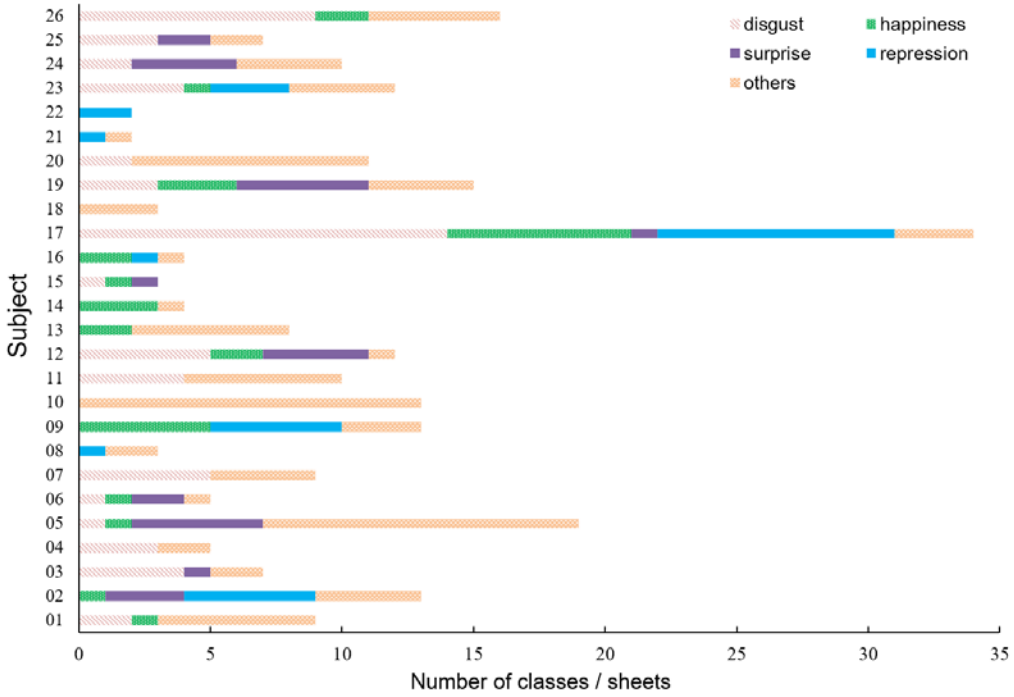

Fig. 4. Emotion distribution of CASME II database

4.2 Experiment Settings

In these experiments, OpenCV and Dlib libraries are used in the preprocessing part, and Pytorch framework is used to build the neural network structure in the deep learning part. At the same time, NVDIA CUDA framework 10.1 and cuDNN 8.0.3 library are used in this experiment. The CPU of the hardware platform used for these experiments is Intel Core i9-9900X, the memory is 64 GB, the video card model is NVDIA Geforce RTX 2080TI * 2, and the hard disk is SAMSUNG 970 EVO PLUS 1TB.

In this paper, Leave-One-Subject-Out Cross-Validation (LOSOCV) method is adopted. Each collected subject is taken as the test dataset, and 26 training and testing processes are conducted on the CASME II database. During the validation process, all the micro-expression samples will be operated by the classifier once. This validation method is mostly used in current research work, with the advantage that the maximum possible number of samples are used for training in each iteration, making full use of the data, sampling is deterministic. At the same time, the disadvantage that this validation method increases the computational overhead is reduced to a very low level because of the small volume of the micro-expression dataset. In addition, there is some variation in the micro-expressions of each collected subject, and using all the micro-expression samples of a single collected subject as a test set can better reflect the generalization ability of the method. In this section, the overall accuracy of the database is determined to evaluate the performance of the method, and the calculation formula of the accuracy is defined as:

$$\text{accuracy}=\frac{\text{Total number of correctly identified samples}}{\text{Total number of database samples}} \tag{13}$$

To further measure the performance of the classifier comprehensively, an additional evaluation criterion F1-score was added to the experiment, defined as:

$$\text{F1-score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{14}$$

$$\text{Precison}=\frac{TP}{TP+FP} \tag{15}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{16}$$

where *TP* indicates true positive, *FN* indicates false negative, and *FP* indicates false positive.

4.3 Results and Discussion

In order to investigate whether the TVL1 optical flow method of extracting motion features is effective for micro-expression recognition, we conducted four sets of experiments using apex frames, u, v, and u + v (refer to (8)) as inputs, and classifiers for the method proposed in this paper, respectively. The results of the four sets of experiments are shown in **Table 2**. The extracted optical flow features are more representative than apex frames, the recognition rate is increased by 18.47%, and F1-score is increased by 0.2847. The effect of using only one direction of the optical flow feature is slightly better than using an apex frame and lower than using a full optical flow. This is due to the incomplete amount of information contained in a single directional optical flow feature. At the same time, it also proves that TV-L1 optical flow can effectively reflect the characteristics of micro-expression.

To validate the effectiveness of the SVM classifier, a fair comparison is established, which selects two approaches (i.e., MobileNetV2 + Softmax and MobileNetV2 + SVM approach) and the experimental configuration of each comparison method is set to be similar. In **Table 3**, it can be seen that the MobileNetV2 network using the SVM classifier has a 14.46% improvement in accuracy and a 0.1871 improvement in F1-score. The lift from replacing Softmax with SVM is huge.

Table 4 shows the experimental results of the various methods on CASME II database, and the algorithm in this paper improves in terms of accuracy compared to other micro-expression recognition methods and outperforms most algorithms in terms of F1-score. But overall the accuracy and F1-score are still not very high. This is due to the unbalanced distribution of emotions, which is shown in **Fig. 4**. In traditional deep learning, adequate model prediction of results requires that the database sample size is sufficient and that the training dataset and test dataset conform to the same distribution. Clearly the present database does not satisfy the conditions.

Table 2. Experimental results of different inputs on CASME II database in terms of recognition accuracy and F1-score

Methods	Accuracy (%)	F1-score
apex frame	34.54	0.2384
u	38.96	0.3151
v	42.17	0.3612
u + v	53.01	0.5231

Table 3. Experimental results of different classifiers on CASME II database in terms of recognition accuracy and F1-score

Methods	Accuracy (%)	F1-score
MobileNetV2 + Softmax	38.55	0.3360
MobileNetV2 + SVM	53.01	0.5231

Table 4. Comparison between our method and some existing methods on CASME II database

Methods	Accuracy (%)	F1-score	Class
LBP [6]	39.68	0.3589	5
OSW [38]	41.70	0.3820	5
FDM [10]	41.96	0.2972	5
MRW [13]	46.15	0.4307	5
LBP-SIP [7]	43.32	0.3976	5
MDMO [11]	44.25	0.4416	5
Sparse Sampling [39]	49.00	0.5100	5
Ours	**53.01**	**0.5231**	**5**

Table 5. Confusion matrix obtained by our method on CASME II database

	Happiness	Disgust	Repression	Surprise	Others
Happiness	37.50	6.25	9.38	6.25	40.63
Disgust	6.35	41.27	3.17	3.17	46.03
Repression	25.93	7.41	33.33	7.41	25.93
Surprise	3.57	7.14	3.57	64.29	21.43
Others	6.06	20.20	5.05	1.01	67.68

For a more detailed analysis to the recognition performance of the five classes of micro-expression, confusion matrix is calculated, as shown in **Table 5**. In general, confusion matrix is a visualization tool used in machine learning to analyze the prediction results of classification models. It depicts the relationship between the true attributes of the sample data and the classification prediction result class in the form of a matrix. It can be seen that 'others' emotion has the highest accuracy compared to the remaining emotions. This is due to the fact that 'others' emotion has the highest proportion in the entire CASME II database (refer to **Fig. 4**). Surprise emotion have a sample size just above repression, but they have the second recognition rate. The recognition rate for the remaining three emotions is moderate and needs improvement. In particular, the disgust emotion scored low despite its high sample size. This may be due to the fact that it is easy to confuse with 'others' emotion.

5. Conclusion and Future Work

In this study, an improved micro-expression recognition method based on optical flow features and MobileNetV2 in combination with SVM is proposed. Our experimental evaluation on CASME II database showed that the accuracy and F1-score are 53.01% and 0.5231. It can distinguish micro-expression classes more precisely than existing methods. In fact, micro-expression only involves local areas of the face, and there are some irrelevant muscle actions when the face produces micro-expression. Therefore, the feature vectors extracted using the global approach for faces uniformly contain more redundant information, which can reduce the expressiveness of the vectors and hence the recognition effect. Excluding areas that have nothing to do with micro-expressions is a major challenge. Since the current micro-expression database is extremely small, macro-expressions and micro-expressions have some commonalities, so learning a priori knowledge from the field of macro-expression recognition is also worth studying. In addition, efforts can be focused on solving the problem of imbalance of sample data in the micro-expression database, so that various existing methods can present better recognition results.

References

[1] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," *Methods of research in psychotherapy*, pp. 154–165, 1966. Article (CrossRef Link)

[2] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88-106, 1969. Article (CrossRef Link)

[3] X. Shen, Q. Wu, and X. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *Journal of Zhejiang University Science B*, vol. 13, no. 3, pp. 221–230, Mar. 2012. Article (CrossRef Link)

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018. Article (CrossRef Link)

[5] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, and X. Fu, "CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation," *PLoS ONE*, vol. 9, no. 1, p. e86041, Jan. 2014. Article (CrossRef Link)

[6] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915-928, June 2007. Article (CrossRef Link)
[7] Y. Wang, J. See, R. C. W. Phan, and Y. H. Oh, "LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Micro-expression Recognition," in *Proc. of Asian conference on computer vision*, Springer, Cham, pp. 525-537, 2014. Article (CrossRef Link)
[8] S. Wang, W. Yan, X. Li, G. Zhao and X. Fu, "Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space," in *Proc. of 2014 22nd International Conference on Pattern Recognition*, pp. 4678-4683, 2014. Article (CrossRef Link)
[9] S. J. Wang, W. J. Yan, X. Li, G. Zhao, C. G. Zhou, X. Fu, M. Yang, and J. Tao, "Micro-Expression Recognition Using Color Spaces," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6034–6047, Dec. 2015. Article (CrossRef Link)
[10] Y. H. Oh, A. C. Le Ngo, J. See, S. T. Liong, R. C. W. Phan and H. C. Ling, "Monogenic Riesz wavelet representation for micro-expression recognition," in *Proc. of 2015 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1237-1241, 2015. Article (CrossRef Link)
[11] Q. Wu, X. Shen, and X. Fu, "The machine knows what you are hiding: an automatic micro-expression recognition system," in *Proc. of international conference on affective computing and intelligent Interaction*, Springer, Berlin, Heidelberg, pp. 152-162, 2011. Article (CrossRef Link)
[12] P. Zhang, X. Ben, R. Yan, C. Wu, and C. Guo, "Micro-expression recognition system," *Optik*, vol. 127, no. 3, pp. 1395–1400, Feb. 2016. Article (CrossRef Link)
[13] F. Xu, J. Zhang and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 254-267, 1 April-June 2017. Article (CrossRef Link)
[14] Y. Liu, J. Zhang, W. Yan, S. Wang, G. Zhao, and X. Fu, "A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition," *IEEE Transactions on Affective Computing*, vol. 7, no. 4, pp. 299-310, 1 Oct.-Dec. 2016. Article (CrossRef Link)
[15] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Proc. of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013. Article (CrossRef Link)
[16] G. Ravikanth, K. V. N. Sunitha, and B. E. Reddy, "Location related signals with satellite image fusion method using visual image integration method," *Computer Systems Science and Engineering*, vol. 35, no.5, pp. 385–393, 2020. Article (CrossRef Link)
[17] J. Li, Y. Lv, B. Ma, M. Yang, C. Wang and Y. Zheng, "Video source identification algorithm based on 3D geometric transformation," *Computer Systems Science and Engineering*, vol. 35, no.6, pp.513–521, 2020. Article (CrossRef Link)
[18] M. B. Nejad, and M. E. Shiri, "A new enhanced learning approach to automatic image classification based on SALP swarm algorithm," *Computer Systems Science and Engineering*, vol. 34, no.2, pp. 91–100, 2019. Article (CrossRef Link)
[19] M. Duan, K. Li, K. Li, and Q. Tian, "A Novel Multi-task Tensor Correlation Neural Network for Facial Attribute Prediction," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 1, pp. 1–22, Feb. 2021. Article (CrossRef Link)
[20] W. Fang, F. Zhang, Y. Ding, and J. Sheng, "A new sequential image prediction method based on LSTM and DCGAN," *Computers, Materials & Continua*, vol. 64, no. 1, pp. 217–231, 2020. Article (CrossRef Link)
[21] W. Fang, L. Pang and W. N. Yi, "Survey on the application of deep reinforcement learning in image processing," *Journal on Artificial Intelligence*, vol. 2, no. 1, pp. 39-58, 2020. Article (CrossRef Link)
[22] X. Wu, C. Luo, Q. Zhang, J. Zhou, H. Yang, and Y. Li, "Text detection and recognition for natural scene images using deep convolutional neural networks," *Computers, Materials & Continua*, vol. 61, no. 1, pp. 289–300, 2019. Article (CrossRef Link)
[23] O. B. Sezer and A. M. Ozbayoglu, "Financial trading model with stock bar chart image time series with deep convolutional neural networks," *Intelligent Automation & Soft Computing*, vol. 26, no.2, pp. 323–334, 2020. Article (CrossRef Link)

[24] K. Zhu, N. Zhang, Q. Zhang, S. Ying and X. Wang, "Software defect prediction based on non-linear manifold learning and hybrid deep learning techniques," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1467–1486, 2020. Article (CrossRef Link)
[25] D. Patel, X. Hong, and G. Zhao, "Selective deep features for micro-expression recognition," in *Proc. of 2016 23rd international conference on pattern recognition*, pp. 2258-2263, 2016. Article (CrossRef Link)
[26] M. Peng, C. Wang, T. Chen, G. Liu, and X. Fu, "Dual temporal scale convolutional neural network for micro-expression recognition," *Frontiers in psychology*, vol. 8, Oct. 2017. Article (CrossRef Link)
[27] M. Verma, S. K. Vipparthi, G. Singh, and S. Murala, "LEARNet: Dynamic imaging network for micro expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 1618-1627, 2020. Article (CrossRef Link)
[28] Y. S. Gan, S. T. Liong, W. C. Yau, Y. C. Huang, and L. K. Tan, "Off-apexnet on micro-expression recognition system," *Signal Processing: Image Communication*, vol. 74, pp. 129-139, 2019. Article (CrossRef Link)
[29] B. Horn, and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, 17(1-3), pp. 185-203, 1981. Article (CrossRef Link)
[30] T. Pock, M. Urschler, C. Zach, R. Beichel, and H. Bischof, "A Duality Based Algorithm for TV-L1-Optical-Flow Image Registration," in *Proc. of International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Berlin, Heidelberg, pp. 511-518, 2007. Article (CrossRef Link)
[31] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259-268, 1992. Article (CrossRef Link)
[32] G. Zhang, X. Sun, J. Liu, and J. Chu, "Research on TV-L1 Optical Flow Model for Image Registration Based on Fractional-order Difierentiation," *Acta Automatica Sinca*, vol. 43, no. 12, pp. 2213-2224, 2017. Article (CrossRef Link)
[33] C. Zach, T. Pock, and H. Bischof, "A Duality Based Approach for Realtime TV-L1 Optical Flow," in *Proc. of Joint pattern recognition symposium*, Springer, Berlin, Heidelberg, pp. 214-223, 2007. Article (CrossRef Link)
[34] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, Apr. 2015. Article (CrossRef Link)
[36] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Proc. of European conference on computer vision*, Springer, Cham, pp. 740-755, 2014. Article (CrossRef Link)
[37] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jun. 2015. Article (CrossRef Link)
[38] S. T. Liong, J. See, R. C. W. Phan, A. C. Le Ngo, Y. H. Oh, and K. Wong, "Subtle expression recognition using optical strain weighted features," in *Proc. of Asian conference on computer vision*, pp. 644-657, 2014. Article (CrossRef Link)
[39] A. C. Le Ngo, J. See and R. C. W. Phan, "Sparsity in Dynamics of Spontaneous Subtle Emotions: Analysis and Application," *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 396-411, 1 July-Sept. 2017. Article (CrossRef Link)

Wei Xu received the B.S. degree in software engineering from NanJing XiaoZhuang University, Nanjing, China, in 2018. He is currently pursuing the M.S. degree in software engineering at Guangxi Normal University, College of Computer Science and Information Engineering, Guilin, China. His research interests include artificial intelligence and image processing.

Hao Zheng received the B.S. degree from South East University in 1998, the M.S. degree from Nanjing University Posts and Telecommunications in 2005, and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology in 2013. He is currently a professor at NanJing XiaoZhuang University. His research interests include pattern recognition, image processing, face recognition, and computer vision.

Zhongxue Yang received the B.S. degree from Nanjing Normal University and the M.S. degree from Nanjing University. He is currently a professor at NanJing XiaoZhuang University. His research interests include pattern recognition and artificial intelligence.

Yingjie Yang received the B.S., M.S. and Ph.D. degrees in engineering from Northeastern University (China) in 1987, 1990, and 1994, respectively. He was awarded his Ph.D. degree in computer science at Loughborough University (UK) in 2008. He is currently a Professor of Computational Intelligence at the Institute of Artificial Intelligence of De Mont fort University. His research interests include the representation and modeling of various uncertainties, and the application of computational intelligence to real world problems.