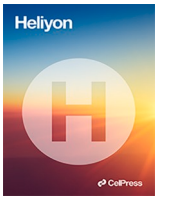
Heliyon 9 (2023) e20508

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



S3DCN-OLSR: A shallow 3D CNN method for online learning state recognition

Jing Bai [a], Xiaohong Yang [a], Qi Li [a], Jinxiong Zhao [b,c,*], Sensen Guo [c,**]

[a] Northwest Normal University, Lanzhou, 730070, China
[b] State Grid Gansu Electric Power Research Institute, Lanzhou, 730070, China
[c] School of Cybersecurity, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China

ARTICLE INFO

Keywords:
S3DC-OLSR model
Online learning
Shallow 3D CNN
Micro-expression

ABSTRACT

The repeated recurrence of COVID-19 has significantly disrupted learning for students in face-to-face instructional settings. While moving from offline to online instruction has proven to be one of the best solutions, classroom management and capturing students' learning states have emerged as important challenges with the increasing popularity of online instruction. To address these challenges, in this paper we propose an online learning status recognition method based on shallow 3D convolution (S3DC-OLSR) for online students, to identify students' online learning states by analysing their micro-expressions. Specifically, we first use the data augmentation method proposed in this paper to decompose the students' online video file into three features: horizontal component of optical flow, vertical component of optical flow and optical amplitude. Next, the students' online learning status is recognised by feeding the processed data into a shallow 3D convolution neural network. To test the performance of our method, we conduct extensive experiments on the CASME II and SMIC datasets, and the results indicate that our method outperforms the other state-of-the-art methods considered in terms of recognition accuracy, UF1 and UAR, which demonstrates the superiority of our method in identifying students' online learning states.

1. Introduction

The unpredictability of COVID-19 has significantly disrupted normal life, especially as it pertains to in-school instruction. With the rapid development of information technology, it is necessary to conduct online teaching activities on network platforms to ensure the sustainability and effectiveness of student learning, as online learning has become the 'new normal'. Online learning affords greater convenience and flexibility than classroom-based learning, allowing for instruction to be conducted at any time and anywhere.

While online teaching offers convenience, it also raises many issues [1]. First, unlike traditional teaching activities, online teaching does not allow instructors to provide timely feedback on students' learning status over the course of the learning process (before, during and after class). In traditional face-to-face settings, teachers can directly connect with students without barriers of distance, and students' learning states are constantly observable through their facial micro-expressions. When conducting online teaching on network platforms, teachers cannot pay attention to the online learning status of each student in real time due to the constraints of the

* Corresponding author. State Grid Gansu Electric Power Research Institute, Lanzhou, 730070, China.
** Corresponding author.
E-mail addresses: jxzhao1229@163.com (J. Zhao), guosensen@mail.nwpu.edu.cn (S. Guo).

https://doi.org/10.1016/j.heliyon.2023.e20508
Received 11 January 2023; Received in revised form 19 September 2023; Accepted 27 September 2023
Available online 11 October 2023

teaching platform and software. Second, teachers find it more difficult to supervise the entire process of students' online learning. Online learning sites are often less supervised and less focused than traditional teaching settings. When teaching online, it is difficult for teachers to give reminders to students within the limited classroom time. Third, online teaching is not conducive to improving students' enthusiasm for learning. Due to the lack of mutual encouragement and competition, students' enthusiasm and initiative for learning are lower when learning online.

Discerning subtle changes in students' micro-expressions is crucial to capture their learning states accurately and reliably. Micro-expression recognition methods can be divided into two main approaches: handcraft-based methods and deep learning-based methods [2]. Handcraft-based methods involve designing and extracting features from facial expressions using handcrafted rules and techniques. These features may include information about the texture, colour and shape of specific facial regions. Machine learning algorithms, such as support vector machines, are then used to classify the extracted features into different micro-expressions [3,4]. Wang et al. [5] proposed a method of recognising facial micro-expressions using colour space, while Huang et al. [6] introduced a method of distinguishing micro-expressions using spatio-temporal completed local quantised patterns. Handcraft-based methods exhibit a certain level of accuracy in recognising micro-expressions, and their high interpretability stems from the fact that the designed features can be understood and explained. However, they rely heavily on human-designed features and may not capture subtle variations in micro-expressions. Their ability to use large-scale datasets is also limited [7].

In contrast, deep learning-based methods have made substantial progress in micro-expression recognition. Deep learning models, especially convolutional neural networks (CNNs) and recurrent neural networks, can automatically learn expression representations from raw data [8]. Zhou et al. [8] predicted micro-expressions by extracting and merging salient and discriminative features of specific expressions. Zhao et al. [2] further improved the performance of deep learning-based micro-expression recognition methods by incorporating attention mechanisms. Instead of handcrafted features, deep learning-based methods learn high-level features from large-scale datasets through an end-to-end training process. This helps to capture the fine details and temporal dynamics of micro-expressions, leading to improved recognition performance. Although machine learning-based methods have shown significant advantages in micro-expression recognition tasks, the short duration of micro-expressions is a challenge for accurate detection, and current research datasets suffer from problems such as small sample size and unbalanced distribution. These challenges collectively contribute to the performance gap of deep learning-based micro-expression recognition methods in real-world applications [7].

Related studies have demonstrated that the performance of micro-expression recognition models can be further enhanced by combining handcrafted features and machine learning methods [9]. For example, Gan et al. [10] found that the apex frame in videos represented the highest intensity facial movements in all frames and that the optical flow signal effectively reflected facial expression changes. Based on this observation, the authors proposed a micro-expression recognition framework that combined handcrafted features with deep neural networks, achieving good classification results. Building on this framework, Liong et al. [11] suggested using a shallow network to extract detailed information from micro-expressions, thereby improving the model's detection capability. Although the aforementioned methods provide satisfactory classification results, the computation of complex optical strains requires substantial computational power and time. To shorten the computational and training time of micro-expression recognition models, this paper proposes a method of extracting detailed information about micro-expressions from three dimensions: horizontal component of optical flow, vertical component of optical flow and optical amplitude. In addition, micro-expression recognition predominantly deals with video data; thus, 3D convolution, unlike 2D convolution, can effectively model the temporal dimension of video data and extract contextual information, enhancing the model's expressive power [12,13]. In conclusion, this study adopts a combined approach of handcrafted features and machine learning by first extracting features from micro-expression videos from three dimensions (i.e., horizontal component of optical flow, vertical component of optical flow and optical amplitude) and then using a 3D CNN to explore the detailed features of the three-dimensional data, enabling accurate micro-expression recognition. The experimental results demonstrate that the proposed micro-expression recognition method outperforms existing baseline algorithms in two representative datasets.

The main contributions of this paper are as follows.

(1) This is the first method to organically combine a 3D CNN network with students' online learning status recognition, thus addressing the difficulty of capturing students' learning status.
(2) This paper constructs a small and shallow 3D CNN network with excellent performance. The computational power required for network performance is low, and the model can detect very rich micro-expression features.
(3) A data enhancement method is proposed to decompose two micro-expression datasets into three features: horizontal component of optical flow, vertical component of optical flow and optical amplitude.

2. Related research

Many scholars have conducted research on online learning. Ma et al. [14] provided a basic but detailed overview of online learning, predicted five macro development trends in online learning from a multi-focus perspective and emphasised the importance of recognising over online learning. Seeking to improve the contributions of online learners, Cui et al. [15] used the generalised linear model to conduct an in-depth analysis of online learner characteristics. Their research results showed that the fragmented channels had no obvious effect on the improvement of student knowledge. Zhan et al. [16] conducted an in-depth investigation into online peer feedback, demonstrating that the pathway from specific peer online feedback design elements to specific learning influences had not yet been established. Analysing online learning behaviour, Leng et al. [17] adopted a content analysis method to re-encode all content. Mu et al. [18] redefined deep learning in online settings according to its focus in the field of teaching, and presented its characteristics

and representation framework. Wu et al. [19] proposed a multi-level approach to developing students' information literacy. Shen et al. [20] proposed a comprehensive and systematic behaviour and evaluation model for online learning that provides an effective basis for the process of online learning. Li et al. [21] conducted in-depth research on the factors responsible for the emergence of MOOCs as leaders in digitisation and remoteness. Aiming to improve teachers' online teaching ability, Xu et al. [22] constructed a model of their online teaching from seven dimensions. From the perspective of teachers' teaching ability, this model provides a concrete implementation path to improve the quality and efficiency of students' online learning. Cai et al. [23] proposed a highly effective implementation path from different perspectives in view of students' online learning feedback. Schnaubert et al. [24] conducted a detailed study of group consciousness with the aid of computer tools.

3. Methodology

The work reported in this paper includes data processing, model design and model training, as shown in Fig. 1. For data processing, we focus mainly on redesigning the CASME II and SMIC datasets and adjusting the complex optical strain to the optical amplitude. For the design of the model, we propose the S3DC-OLSR model. Model training mainly involves using the CASME II-TD and SMIC datasets to train the S3DC-OLSR model.

3.1. Data processing module

High computing resources are still relatively expensive considering the large amount of computing resources required by most users. How to achieve good training results with limited computing power has been the focus of many studies [25]. The method proposed by Liong et al. [11] increased the richness of the data. However, calculating the complex optical strain requires significant computing power and a lot of time. To shorten the calculation time and model training time, we extract the features of the CASME II and SMIC datasets and decompose them into three features to further increase data richness: horizontal component of optical flow, vertical component of optical flow and optical amplitude. Finally, the three features are combined to form a new three-channel dataset, which is used as the model input. The micro-expression video sequence is defined as Equation (1):

$$V = [v_1, v_2, v_3, \cdots v_n] \tag{1}$$

n represents the total number of videos. The definition of the i-th segment video is displayed in Equation (2):

$$v_i = \{f_{ij} | i = 1, 2, 3, \cdots n; j = F_1, F_2, F_3, \cdots, F_i\} \tag{2}$$

$_i$ is the total number of video frames in the i-th video sequence. Each video contains a start frame, a vertex frame and an end frame, denoted by $f_{i,1}$, $f_{i,m}$ and $f_{i,Fi}$, respectively. The owning relationship of the vertex frame, end frame, and start frame is displayed in Equation (3):

$$f_{i,m} \in f_{i,1}, f_{i,2}, f_{i,3}, \cdots, f_{i,F_i} \tag{3}$$

Optical flow is calculated by the starting frame and vertex frame. The applied calculation method is shown in Equation (4):

$$O_i = \{(u(x,y), v(x,y)) | x = 1, 2, 3 \cdots, X, y = 1, 2, 3 \cdots, Y\} \tag{4}$$

where X and Y represent the width and height of the video frame respectively. u and v represent the horizontal and vertical components of optical flow, respectively. The optical amplitude can be calculated according to Equation (5) as follows:

$$\gamma = \sqrt{x^2 + y^2} \tag{5}$$

where $\overrightarrow{u} = [u, v]^T$ is the displacement vector.

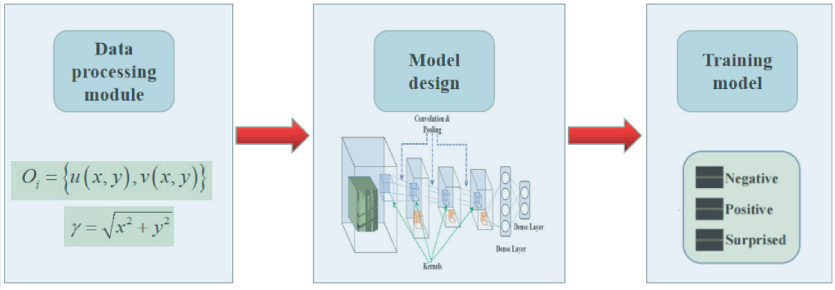


Fig. 1. Process diagram of this study.

3.2. S3DC-OLSR model design

Considering that 2D CNN networks can only represent data spatially, 3D CNN networks are more suitable for spatial–temporal feature learning. A 3D CNN network can use 3D convolution and 3D pooling to fully characterise and model the data's spatio-temporal information [12,13]. As shown in Fig. 2, the S3DC-OLSR model proposed in this paper first extracts the horizontal component of optical flow, the vertical component of optical flow and optical amplitude features from micro-expression data; then a multilayer 3D CNN is used to extract high-dimensional features from these data; finally, the model uses a fully connected layer to map the results into three classes, namely positive, negative and surprised. In the above process, the S3DC-OLSR model constructs multi-frame images adjacent to the horizontal component of optical flow, vertical component of optical flow and optical amplitude into a three-dimensional space-time cube as the model input. The specific calculation process of the j-th feature of position (x,y,z) in the i-th layer is shown in Equation (6):

$$v_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \tag{6}$$

where $\tanh(\cdot)$ represents the equation of the double tangent positive curve, b_{ij} represents the deviation of the feature mapping, m represents the mapping from the previous layer to the current layer, ω_{ijm}^{pqr} represents the value of the corresponding kernel function at (p,q,r), and P_i, Q_i and R_i represent the height, width and depth of the kernel, respectively.

In this paper, the micro-expression dataset is decomposed into three features: horizontal component of optical flow, vertical component of optical flow and optical amplitude. The S3DC-OLSR model better fits the temporal and spatial information characteristics of the CASME II-TD and SMIC datasets. Furthermore, because of the advantages afforded by its shallow structure, it can be trained under the premise of non-harsh operating environment indicators such as computing power. The S3DC model contains a total of five layers in addition to the input and output layers. There are two convolutions in the third layer, C3a and C3b. These convolutions are intended to improve the abstraction of image features and the richness of expression, which helps to improve the accuracy of micro-expression recognition. The input sizes of the corresponding convolution layers are successively 28*28*3, 14*14, 7*7 and 7*7 from front to back. In the model design, batch normalisation is added to the first and second layers respectively. This allows for the input of the S3DC-OLSR model to maintain the same distribution in the training process. The specific calculation process is shown in Equations 7–10:

$$\mu_\beta = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$\sigma_\beta^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_\beta\right)^2 \tag{8}$$

$$\widehat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \varepsilon}} \tag{9}$$

$$y_i = \gamma\widehat{x}_i + \beta \tag{10}$$

where μ represents the mean value of all inputs, σ represents the variance of all inputs, ε represents a constant, and γ and β are the learning parameters of the model.

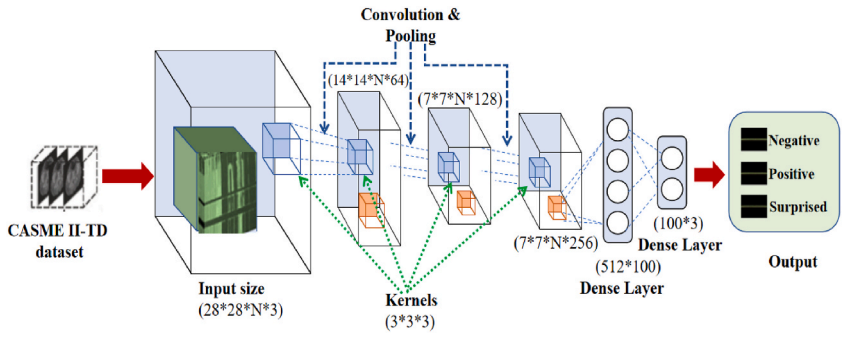


Fig. 2. S3DC-OLSR model structure.

4. Experiment and discussion

4.1. Setup

(1) Datasets

The objective of this study is to accurately recognise the listening state of students in online classes. Due to privacy concerns and legal regulations, it is difficult to obtain real online student lecture videos for practical application in research. Therefore, this study primarily relies on two publicly available micro-expression datasets that resemble the classroom setting to conduct relevant experiments. Furthermore, as the degree of students' understanding of the content presented by the teacher in class is broadly classified as positive, negative and surprised, the two datasets are divided into the three categories mentioned above. The datasets used in this study are presented below.

- **CASME II-TD** (Chinese Academy of Sciences Micro-Expression II): CASME II-TD is a dataset of micro-expression videos for studying facial expressions and emotions. It consists of 246 sequences from 30 participants, covering various emotions. The dataset includes time-delayed frames to analyse the temporal dynamics of micro-expressions. It also provides detailed annotations for developing automated recognition algorithms. CASME II-TD is useful for research on micro-expression analysis and emotion recognition.
- **SMIC** (Selective Magnetic Induction Capture): The SMIC dataset is a collection of high-quality facial expression videos captured using magnetic induction technology. It includes 100 videos covering six emotions. This dataset is useful for developing facial expression recognition algorithms and studying emotions.

(2) Baseline models

To validate the effectiveness of the proposed model, this section mainly compares the performance of the S3DC-OLSR model with that of other micro-expression recognition methods that combine handcrafted features and machine learning, as well as methods that use 3D convolution for micro-expression recognition. To this end, six state-of-the-art micro-expression recognition algorithms, namely OFF-ApexNet [10], MECapsuleNet [25], STSTNet [11], MERSiamC3D [8], ME-PLAN [2] and FeatRef [26], are used as baseline models. A brief description of each algorithm is provided below.

- **OFF-ApexNet**: This approach recognises micro-expressions by combining handcrafted features (i.e., optical flow-derived components) and a fully data-driven architecture (i.e., CNNs).
- **MECapsuleNet**: In this approach, the author first detects and pre-processes the apex frames of the micro-expression sequences, and then a transfer learning mechanism is used to train the detection network.
- **STSTNet**: In this approach, the author designs a shallow network to extract high-level features of micro-expressions from images of optical strain, horizontal and vertical optical flow.
- **MERSiamC3D**: In this approach, the author first constructs the keyframe sequence to summarise the original micro-expression video, and then conducts micro-expression recognition using prior learning and target learning.
- **ME-PLAN**: In this approach, the author proposes a prototype learning framework with local attention to learn the specific micro-expression features of the prototype through expression-related knowledge transfer and episodic training.
- **FeatRef**: This approach seeks to obtain discriminative and salient features for specific expressions and predicts micro-expressions by merging expression-specific features.

(3) Parameter settings

The input sizes of the corresponding convolution layers are 28*28*3, 14*14, 7*7 and 7*7 from front to back. In model design, batch normalisation is added to the first and second layers. This allows the input of the S3DC-OLSR model to maintain the same distribution in the training process.

(4) Evaluation index

Table 1
Performance of the proposed S3DC-OLSR model.

Method	CASME II		SMIC	
	UF1	URA	UF1	URA
OFF-ApexNet	0.8764	0.8764	0.6817	0.6695
MECapsuleNet	0.6175	0.6420	0.5203	0.4649
STSTNet	0.8382	0.8382	0.6801	0.6801
MERSiamC3D	0.8818	0.8763	**0.7356**	0.7598
ME-PLAN	0.8632	0.8778	0.7127	0.7256
FeatRef	0.8915	0.8873	0.7011	0.7083
S3DC-OLSR (Ours)	**0.9776**	**0.9905**	0.7106	**0.7613**

The evaluation indexes used in the experiment are accuracy (Acc), UF1 and UAR [11,27]. Acc is a measure of the number of correct results predicted by the model as a percentage of the total sample. UAR represents the average recall rate for each type of data sample.

4.2. Performance of the S3DC-OLSR

To validate the effectiveness of the proposed S3DC-OLSR model using the CASME II and SMIC datasets, this section compares the model's performance with that of other state-of-the-art algorithms, namely OFF-ApexNet, MeCapsuleNet, STSTNet, MERSiamC3D, ME-PLAN and FeatRef. The evaluation metrics considered here are UF1 and UAR. The experimental results are shown in Table 1.

As shown in Tables 1 and in the comparative experiments using the CASME II dataset, the MECapsuleNet model shows the worst classification performance, with UF1 and UAR scores below 0.65. The performance of the other baseline algorithms is superior to that of the MECapsuleNet model, with UF1 and UAR scores ranging from 0.8 to 0.9. Among them, the FeatRef algorithm performs the best. In comparison, the UF1 and UAR scores of the proposed S3DC-OLSR model are about 8 % and 10 % higher, respectively, than those of the FeatRef algorithm. In the comparative experiments based on the SMIC dataset, although the UF1 score of the proposed S3DC-OLSR model is slightly lower than that of the MERSiamC3D algorithm, the UAR score of the proposed model is higher than that of the other algorithms. Overall, the S3DC-OLSR model achieves state-of-the-art classification performance using the SMIC dataset.

In summary, the S3DC-OLSR model proposed in this paper has two main advantages. First, by extracting the horizontal component of optical flow, vertical component of optical flow and optical amplitude from the original dataset, it can filter out redundant information and obtain salient micro-expression features. Second, by using a 3D CNN, the model can extract dynamic information from the data, thereby improving its ability to understand and analyse video data. The experimental results also indicate that the proposed S3DC-OLSR model significantly outperforms other state-of-the-art models in micro-expression recognition tasks.

4.3. ROC and PR curves of the proposed method

To further evaluate the performance of the proposed S3DC-OLSR model, this section evaluates its receiver operating characteristic (ROC) and precision-recall (PR) curve metrics using the CASME II and SMIC datasets. The experimental results are shown in Figs. 3 and 4, respectively.

As shown in Fig. 3a and b, in the evaluation experiments using the CASME II dataset, the S3DC-OLSR model has AUC values of 0.97 or above for all emotion categories and its AP values are all above 0.83. These results indicate that the S3DC-OLSR model performs well in recognising the micro-expressions of individuals using the CASME II dataset.

As shown in Fig. 4a and b, in the evaluation experiments based on the SMIC dataset, the overall performance of the S3DC-OLSR model in classifying different emotion categories is slightly inferior to its performance using the CASME II dataset. However, its results for recognising the micro-expressions of individuals using the SMIC dataset are still satisfactory.

4.4. Comparison of the structure of the S3DC-OLSR model

As can be seen from the network structure in Table 2, the network depth of the SS3DC-OLSR model is one layer higher than that of the STSTNet method. When the number of adjacent frames in the S3DC-OLSR model is 3, 4, 5 and 6, the time spent on each iteration is 41, 39, 41 and 42 s, respectively. In terms of execution time, the training time of the S3DC-OLSR model corresponding to each iteration is more than 6 times that of STSTNet. Compared with the S3DC-OLSR model, the depth, number of parameters and input size of the VGG16 model are about 5 times, 13 times and 2 times greater, respectively, which indicates that the depth of the S3DC-OLSR model is indeed relatively shallow. Compared with the OFF-ApexNet model, the depth of the S3DC-OLSR model is 2 layers less than that of the OFF-ApexNet model, but the number of parameters of the S3DC-OLSR model is about 3.8 times that of the OFF-ApexNet model. Overall, the S3DC-OLSR model is characterized by shallow depth, moderate parameters and reasonable training time for each iteration.

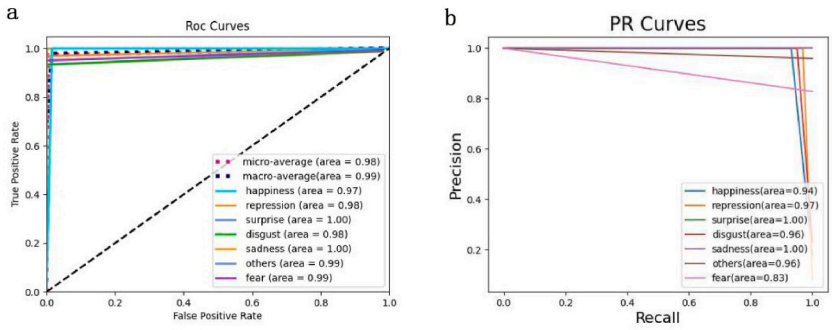


Fig. 3. Performance of the S3DC-OLSR using the CASME II dataset.

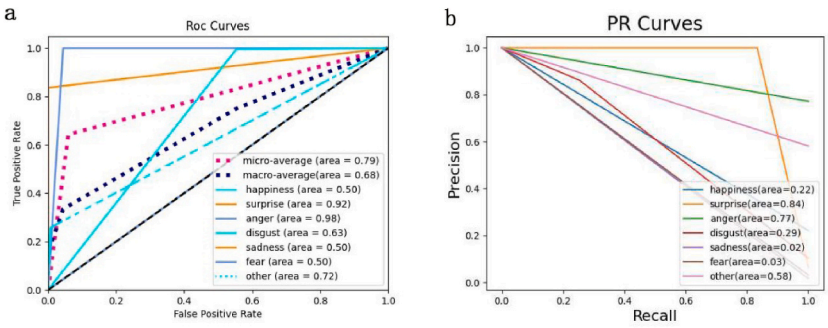


Fig. 4. Performance of the S3DC-OLSR using the SMIC dataset.

Table 2
Structural comparison of the S3DC-OLSR model and other algorithms.

Serial number	Method	Depth	Number of parameters (millions)	Input size	Execution time (seconds)
1	S3DC-OLSR	3	10.6	28*28*N*3	41(3)/39(4)/41(5)/42(6)
2	VGG16 [9]	16	138	224*224*3	95
3	OFF-ApexNet [10]	5	2.77	28*28*2	6
4	STSTNet [11]	2	0.00167	28*28*3	6

4.5. Ablation studies

As mentioned, this study uses a 3D CNN to extract high-dimensional features from the three handcrafted features. This enables accurate recognition of micro-expression changes in video data. Among the dimensions, adjacent frames are a crucial parameter that affects the performance of 3D CNNs. Therefore, this section primarily investigates the influence of this parameter on the experimental results. Specifically, this subsection sets the adjacent frame parameter of the S3DC-OLSR model to 3, 4 and 5, and the classification performance of the above models on the CASME II dataset is recorded. The experimental results are presented in Table 3.

As indicated by the experimental results, among the three experimental scenarios, the S3DC-OLSR model achieves the highest accuracy in micro-expression recognition when the adjacent frame parameter is set to 4. This suggests that an appropriate value for this parameter has a significant impact on the detection performance of the S3DC-OLSR model. Therefore, in this paper, the adjacent frame of the S3DC-OLSR model is set to 4.

5. Conclusion

In this paper, we first propose to organically combine a 3D CNN network with students' online learning status recognition, and to build a shallow 3D CNN network with excellent performance, low computational power requirements and strong learning abilities. In addition, a data enhancement method is proposed to decompose the dataset into three features, i.e., horizontal component of optical flow, vertical component of optical flow and optical amplitude, which together effectively improve the model performance. The proposed S3DC-OLSR model is superior to the other state-of-the-art approaches considered in terms of UF1 and URA indexes. As the S3DC-OLSR model can accurately identify students' online learning status, it will help teachers to analyse students' learning status. As online teaching activities become the 'new normal', the method proposed in this paper will provide a new regulatory tool. We will delve into how online learning can be transformed to help students adjust to online learning. In the future, we plan to select three classes in the same grade (each class has about 40 students), capture the status photos of each class 5 times, and then use the proposed model to make real-time discrimination. Finally, the performance of the five times is given a comprehensive evaluation.

Author contribution statement

Jing Bai: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Xiaohong Yang: Conceived and designed the experiments.

Qi Li: Performed the experiments.

Jinxiong Zhao: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Data availability statement

The authors do not have permission to share data.

Table 3
The impact of adjacent frames on classification performance.

Adjacent Frames	Acc	UF1	URA
3	0.9750	0.9504	0.9700
4	**0.9821**	**0.9776**	**0.9905**
5	0.9605	0.8788	0.9804

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the InnovationFoundation for Doctor Dissertation of Northwestern Polytechnical University (Grant No. CX2022065), Youth Lift Talent Project of Gansu Science and Technology Association(Grant No. GXH202220530-10), National Key R&D Program of China (Grant No. 2020AAA0107700), and InnovationFoundation for Doctor Dissertation of Northwestern Polytechnical University (CX2022065). Professional English language editing support provided by AsiaEdit (asiaedit.com).

References

[1] Z. Lu, X. Liu, X. Cao, Thinking and practice of online teaching of College Mathematics[J], Education Modernization 56 (2021) 14–17.
[2] S. Zhao, H. Tang, S. Liu, et al., ME-PLAN: a deep prototypical learning with local attention network for dynamic micro-expression recognition[J], Neural Network. 153 (2022) 427–443.
[3] X. Li, X. Hong, A. Moilanen, et al., Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods [J], IEEE Transactions on Affective Computing 9 (4) (2017) 563–577.
[4] Y.J. Liu, J.K. Zhang, W.J. Yan, et al., A main directional mean optical flow feature for spontaneous micro-expression recognition[J], IEEE Transactions on Affective Computing 7 (4) (2015) 299–310.
[5] S.J. Wang, W.J. Yan, X. Li, et al., Micro-expression recognition using color spaces[J], IEEE Trans. Image Process. 24 (12) (2015) 6034–6047.
[6] X. Huang, G. Zhao, X. Hong, et al., Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J], Neurocomputing 175 (2016) 564–578.
[7] H.X. Xie, L. Lo, H.H. Shuai, et al., An Overview of Facial Micro-expression Analysis: Data, Methodology and challenge[J], IEEE Transactions on Affective Computing, 2022.
[8] S. Zhao, H. Tao, Y. Zhang, et al., A two-stage 3D CNN based learning method for spontaneous micro-expression recognition[J], Neurocomputing 448 (2021) 276–289.
[9] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition[C], Proceedings of the Conference on Computer Vision and Pattern Recognition 1409 (2014) 1556.
[10] Y.S. Gan, S.T. Liong, W.C. Yau, et al., OFF-ApexNet on micro-expression recognition system[J], Signal Process. Image Commun. 74 (2019) 129–139.
[11] S.T. Liong, Y.S. Gan, J. See, et al., Shallow triple stream three-dimensional CNN (STSTNet) for micro-expression recognition[C], Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE (2019) 1–5.
[12] S. Mittal, A survey of accelerator architectures for 3D convolution neural networks[J], J. Syst. Architect. 115 (2021), 102041.
[13] L. Zhou, X. Shao, Q. Mao, A survey of micro-expression recognition[J], Image Vis Comput. 105 (2021), 104043.
[14] B. Mark, C. Eamonn, T. Enda, J. Xiao, Five major trends shaping online learning: a multifocal view of possible futures[J], Distance Education in China 6 (2022) 21–35.
[15] B. Mark, C. Eamonn, T. Enda, J. Xiao, Five major trends shaping online learning: a multifocal view of possible futures[J], Distance Education in China 6 (2022) 21–35.
[16] Y. Zhan, Z.H. Wan, D. Sun, Online formative peer feedback in Chinese contexts at the tertiary Level: a critical review on its design, impacts and influencing factors[J], Comput. Educ. 176 (2022), 104341.
[17] J. Leng, Xi Lu, J. Song, Exploring university students' online knowledge construction behavioral patterns and sequential model[J], Distance Education in China 1 (2022) 85–91.
[18] S. Mu, X.J. Wang, Participation and engagement: deep learning in online learning[J], Distance Education in China 2 (2019) 17–25.
[19] D. Wu, C. Zhou, Y. Li, M. Chen, Factors associated with teachers' competence to develop students' information literacy: a multilevel approach[J], Comput. Educ. 176 (2022), 104360.
[20] X. Shen, J. Wu, Y. Zhang, Y. Li, Y. Ma, Towards an evaluation model of online learning behavior and learning effectiveness for MOOCAP learners[J], Distance Education in China 7 (2019) 38–46.
[21] L. Li, J. Johnson, W. Aarhus, D. Shah, Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: what makes a hit[J], Comput. Educ. 176 (2022), 104354.
[22] P. Xu, Y. Sun, The construction of teachers' online teaching ability based on quality standards of online courses and its improvement path[J], China Educational Technology 6 (2022) 89–95.
[23] M. Cai, W. Guo, Y. Lou, How to implement effective feedback during online learning: online feedback discussion based on self-regulated learning theory[J], China Educational Technology 10 (2020) 82–88.
[24] L. Schnaubert, D. Bodemer, Group awareness and regulation in computer-supported collaborative learning[J], International Journal of Computer-Supported Collaborative Learning 17 (1) (2022) 11–38.
[25] N. Van Quang, J. Chun, T. Tokuyama, CapsuleNet for micro-expression recognition[C], Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE (2019) 1–7.
[26] L. Zhou, Q. Mao, X. Huang, et al., Feature refinement: an expression-specific feature learning and fusion method for micro-expression recognition[J], Pattern Recogn. 122 (2022), 108275.
[27] J. Zhao, S. Guo, D. Mu, DouBiGRU-A: software defect detection algorithm based on attention mechanism and double BiGRU[J], Comput. Secur. 111 (2021), 102459.