



Facial expression recognition method with multi-label distribution learning for non-verbal behavior understanding in the classroom

Tingting Liu^{a,b}, Jixin Wang^b, Bing Yang^{a,*}, Xuan Wang^c

^a School of Education, Hubei University, 368 Youyi Road, Wuhan 430062, Hubei, China

^b Collaborative Innovation Centre for Information Technology and Balanced Development of K-12 Education, Central China Normal University, 152 Luoyu Road, Wuhan, Hubei, China

^c National Engineering Research Center for E-Learning, Central China Normal University, 152 Luoyu Road, Wuhan, Hubei, China



ARTICLE INFO

Keywords:

Facial expression recognition
Learning behavior analysis
Cauchy distribution
Machine learning
Classroom teaching
K-12 education

ABSTRACT

The automatical recognition of human facial expression has attracted attention in the field of computer vision and machine learning. Previous works on this topic set many constraints, such as the impact caused by restricted scenarios and low image quality. To address those problems, we propose a new infrared facial expression recognition method with multi-label distribution learning for understanding non-verbal behaviors in the classroom. Specifically, we first compute the feature similarities of seven basic facial expressions to describe the relationship among the adjacent expression images. Then, the similarity values are fitted by a Cauchy distribution function. Furthermore, we construct a new deep network with Cauchy distribution-based label learning (CDLLNet), instead of the conventional single expression labels. By these revised labels, one infrared facial expression can contribute to the learning of neighboring expression labels, as well as its real expression label. The performance of proposed network is evaluated on two facial expression datasets: Oulu-CASIA and CK+. Several qualitative and quantitative experimental results verify that the CDLLNet network can achieve robust results and significantly outperforms the existing state-of-the-art facial expression algorithms.

1. Introduction

Human facial expression recognition (FER) under active infrared (IR) imaging is an important way to understand the attention of human action [1–7]. It has attracted much attention in the fields of computer vision and image understanding [8,9], such as educational technology [10–12], robot vision [11,13–15], human-computer interaction [16,17], computer vision [18–21], and intelligent control [13,22–24]. In the classroom, teachers' use of expression can be considered as a common form of nonverbal communication with the students [25–27]. The facial expression includes important cues for emotional states and intentions. However, complex background and irrelevant facial regions (such as the hat and the glasses) create problems when utilizing the conventional facial expression recognition methods (see Fig. 1).

Over the past decades, infrared FER technology has made great progress with expression datasets and classification algorithms [8,28–40]. The imaging system can obtain the clean images without the visible light, under active near-infrared (NIR) imaging (780–1100 nm) [41–43]. The facial expression recognition task major includes three

steps, such as face detection, feature extraction, and expression recognition. For the first step, many face detectors are developed to locate the faces in the crowd, which includes the CNN-MT and Dlib [8]. The captured face images can be pre-processed for subsequent steps. In the second step, lots of algorithms are proposed to extract the facial geometrical features, which is caused by various human emotions and appearances. The last step is to conduct classification for different facial expressions. In sum, the FER methods can be classified into three groups, hand-crafted feature (HCF) based method, auxiliary information-based FER method (AIB), and deep learning based FER method (DLB).

For the HCF method, the hand-crafted features are extracted by different mathematic operators, which includes the Gabor wavelet coefficients [44–46], histogram of gradients (HoG) [47–49], fractal geometry [20,50], and local binary pattern [51]. However, only the texture-based local features can be revealed by those operators. To overcome this problem, some works developed the key points marking method to extract the global facial features, such as the mouths, eyebrows, and noses. Then, the local features and global features can also be combined and extracted to enrich the facial feature representation. For

* Corresponding author.

E-mail address: yangbing.cn88@gmail.com (B. Yang).

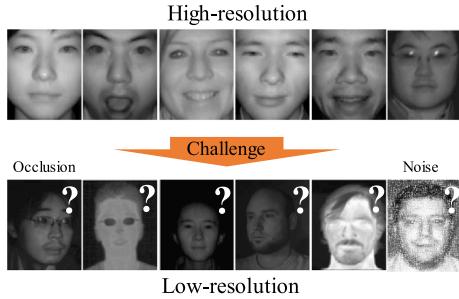


Fig. 1. The occlusion, head pose variation and random noise problems exist in infrared facial expression images. The first row denotes the high-resolution facial images and second row presents the low-resolution samples.

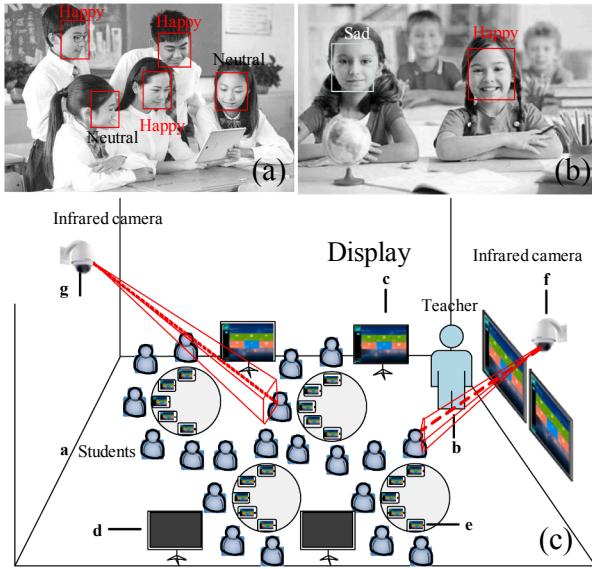


Fig. 2. Facial expression recognition in the classroom teaching. (a) and (b) expression recognition results. (c) Advanced digital classroom.

the DLB methods [4,23,38,52], a novel CNN-based FER method is proposed by McDuff et al. [53,54] to describe the IR facial expression images. Multi-poses and multi-scales facial features can be learned by this network. In [55–57], Kahou et al. proposed a novel deep neural network to reveal the facial expression features by the modality information fusion. The proposed model obtains impressive results and wins the champion of FER2013 competition.

In our previous work [2,11,25,58], we proposed a hand gesture estimation method to capture the instruction behavior in the classroom. A non-linear neural network is proposed with four convolution layers to extract the features of the infrared hand gesture images. In fact, the facial expression also belongs to one type of instruction behaviors in the classroom (see Fig. 2). Unlike the well illuminated facial images, infrared facial images perform better in the dark environment. The

active near-infrared illumination provides the stable imaging conditions to capture the facial expression images. The images can be considered as several categories, such as the anger, fear, happy, neutral, surprise, sadness, and disgust. However, the occlusion, head pose variation and random noise problems exist in infrared facial expression images. To address those challenges, we propose an effective infrared facial expression recognition method with multi-label distribution learning. The expression labels are constructed as a Cauchy distribution to overcome the ambiguous facial expression problem. Furthermore, the KL divergence is introduced to measure the predicted and ground-truth distributions, which is robust for the low-quality infrared facial images. Overall, the major contributions of this paper can be summarized as follows,

- (1) We propose a Cauchy distribution label learning network (CDLLNet) to learn the facial feature in the IR expression images. The method adopts GoogleNet, which can leverage covariance pooling layer to capture second-order image features.
- (2) A Cauchy distribution-based label is constructed for each expression, which can assist the CDLLNet to judge the expression image kind. The distribution labels are designed according to the similarity values of expression image features.
- (3) Experimental results on several public datasets demonstrate that the proposed CDLLNet model obtains the performance 86.71% on Oulu-CASIA, and 83.14% on CK+, which are better than the state-of-the-art approaches.

The rest of this article is organized as follows. Section 2 computes the similarities and corrections between different facial expressions and propose a Cauchy distribution-based label. Then, we describe the methodology of the proposed method and the optimization by the mini-batch gradient descent algorithm. Section 3 illustrates the experimental results on several public datasets. Finally, we conclude this study in Section 4.

2. The architecture of CDLLNet

We propose a novel multi-label learning network for FER. The expression image labels are constructed as the Cauchy distribution. Since each expression has its own similar expressions, the labels of each expression are different. Finally, we introduce the alternate training strategy for fully training the CDLLNet and illustrate the inference strategy in details.

2.1. Overview of CDLLNet model

The architecture of the proposed network includes three layers, such as convolutional layer, covariance pooling layer, and output layer. The pipeline of the CDLLNet network is shown in Fig. 3.

2.2. Cauchy distribution-based expression labels

The first step of the label distribution learning is the label distribution construction. The performance of facial expression recognition will

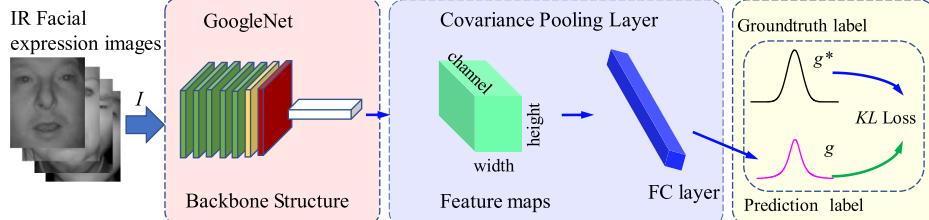


Fig. 3. The framework of the CDLLNet architecture. It has three main components: GoogleNet backbone structure, covariance pooling layer and KL loss.

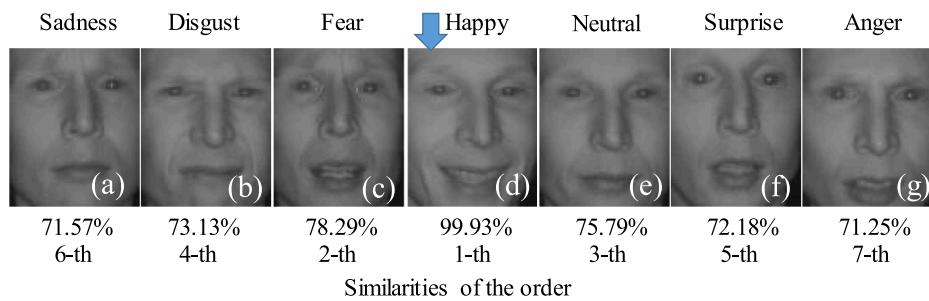


Fig. 4. Similarity between various expression (%). Given a facial expression image, the relevant expression labels on the single-label ground-true are sorted by Hump sort algorithm.

be raised markedly with an accurate expression label construction. The cosine similarity [59–61] is introduced to measure the feature similarity of two facial expression images. Given two different infrared image Img_1 and Img_2 , the similarities of expression features are computed as follows.

Firstly, a pre-trained neural network (GoogleNet) is introduced to extract the major feature of the infrared expression. In the last full-connection layer, the output vector can represent the distinguishing feature in NIR facial image. To calculate the similarities of two facial expressions, two vectors can be achieved from the pre-trained neural network.

$$SIM(Img_1, Img_2) = \frac{FEA(Img_1) \cdot FEA(Img_2)}{\|FEA(Img_1)\| \times \|FEA(Img_2)\|} \quad (1)$$

where $FEA(Img)$ is the vector of the last full-connection layer. Since the ambiguity exists in each NIR facial expression, the similarities values between expression Img_1 and other expressions are not equal. The expression feature vector is predicted by the network and decoded into a specific expression value in the prediction step. Furthermore, each value in feature vector could be considered as a probability value, which denotes the possibility of facial expression image in different facial categories. Thus, the similarities values between all the facial categories are leveraged to construct the expression label distributions. For the given facial expression image (Neutral), there are six similarity values, which are computed by Eq. (1). Hump sort algorithm is introduced to sort all the similarity values (probability values). Taking the happy expression as an example, the expression order can be illustrated as,

$$\begin{aligned} O(I_{Neu}) &= [p_6(\Lambda(I_{Sur})), p_4(\Lambda(I_{Fea})), p_2(\Lambda(I_{Ang}))p_1(\Lambda(I_{Neu})), \\ &\quad p_3(\Lambda(I_{Sad})), p_5(\Lambda(I_{Hap})), p_7(\Lambda(I_{Dis}))] \\ &= [\text{Sur, Fea, Ang, Neu, Sad, Hap, Dis}] \end{aligned} \quad (2)$$

where the symbol $O(I_{Neu})$ denotes the sorting relevant expression labels on the happy expression. Other expressions can be sorted by the similar manner. In Fig. 5, we plot the probability values distribution for each fixed label distribution.

In this paper, the essential idea of the soft label distribution construction is to find the most similar expression for the given expression (Neutral) in the rest six expressions. The tiny difference between the similar expressions can be distinguished by the constructed label distributions. To measure the distribution accurately, the famous Cauchy distribution [15,62–64] is employed to fit the soft label distribution. Each label in \mathbf{g} can be formulated as follows,

$$g(x, x_0, \gamma) = \frac{1}{\pi} \left(\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right) \quad (3)$$

where n denotes the degrees of freedom, and $\Gamma(\cdot)$ means the Gamma function. The hyper-parameter n plays an important role in controlling the shape of the distribution. The smaller the n value is set, the wider the distribution achieves. With different degrees of freedom value, the

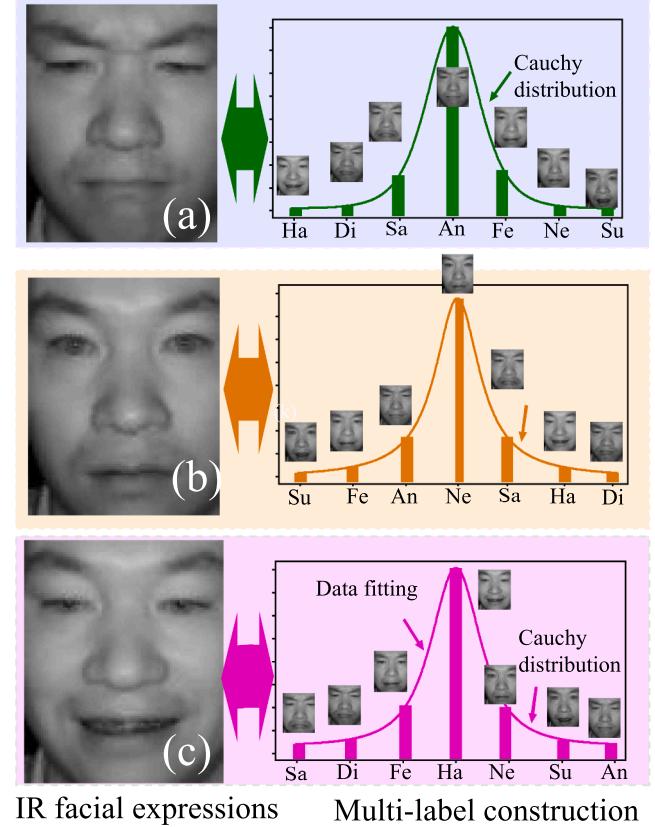


Fig. 5. Soft labels construction for all the infrared facial expression categories. (a) Happy expression. (b) Anger expression. (c) Neutral expression. The second column denotes the constructed labels, which are fitted by the Cauchy distribution function.

Cauchy distributions are shown in Fig. 4. It is easy to see that the Gauss normal distribution could be considered as a special case of the Cauchy distribution when n approaches $+\infty$.

2.3. Multi-label distribution modeling

The aim of the soft label distribution learning model is to learn a mapping function $f(y|I): I \rightarrow \mathcal{D}$. Given the facial image Img , the goal of the proposed model is to seek the network parameters that could predict a distribution \mathbf{g} which approaches the ground-truth expression distribution \mathbf{g}_i . Then, the Kullback–Leibler divergence is introduced to measure the predicted and ground-truth distributions,

$$L_{KL} = \sum_{j=1}^c f_j \ln \frac{g_j}{f_j} \quad (4)$$

where d denotes the categories number or the dimension of distributions. The symbols f_j and g_j denote the j -th element of the predicted distribution and ground-truth distribution respectively. The f_j is normalized as

$$f_j = \frac{\exp(s_j)}{\sum \exp(s_k)} \quad (5)$$

where s_j is the output score of the last fully connected layer for j -th class and its vectored form for all classes denotes as s . The network is selected as the pre-trained network GoogleNet. To avoid the over-fitting problem, the model parameters is also regularized by the L_2 norm. Thus, the final loss function of the CDLLNet model is formulated as,

$$L(\theta) = \sum_{j=1}^c f_j \ln \frac{g_j}{f_j} + \alpha \|\theta\|^2 \quad (6)$$

where α means the regularization parameter, which can balance the fidelity term and two regularization terms. The proposed model is named as the soft label distribution learning (CDLLNet) for infrared facial expression recognition in human-computer interaction.

2.4. Optimization algorithm

Since the non-convex term is existed in the loss function, it is very difficult to optimize the JF divergence term. To address this issue, the mini-batch gradient descent algorithm is introduced to efficiently optimize this loss. The minimization detail is illustrated as follows. The loss function can be expanded as

$$L(\theta) = \sum_{j=1}^c [-f_j \ln g_j + f_j \ln f_j] + \alpha \|\theta\|^2 \quad (7)$$

After some manipulation, the first term can be safely dropped out. Thus, the soft label distribution-based model (7) can equal as the minimization of the loss function,

$$L(\theta) = \sum_{j=1}^c [-f_j \ln g_j] + \alpha \|\theta\|^2 \quad (8)$$

The equation will be optimized in the Caffe [65] implement framework. The back propagation can be carried out when the gradient value is achieved. The update of the model parameters can be executed orderly in the convolutional neural network framework. The equation could be easily vectorized and computed for the training batch input.

3. Experiments and analysis

3.1. Experimental preliminary

(1) **Comparing methods:** Five models are chosen below in our comparative experiments:

- **LBP-FER [31]:** This method extracts the IR facial features from three orthogonal planes by local binary patterns. For the near infrared video cases, it is also extended with the image frame information.

Table 1

Comparison of recognition accuracy on the Oulu-CASIA database by the proposed method and other comparing methods.

Algorithms	Image selection	Categories	Accuracy (%)
LBP-FER [31]	Last three images	Seven	70.12%
E3DNet [29]	Last image	Seven	79.21%
MA-FER [66]	Last three images	Seven	84.56%
DBMNet [35]	Last three images	Seven	85.36%
Ours (CDLLNet)	Last three images	Seven	86.28%

- **E3DNet [29]:** A 3D structure network is proposed to extract the spatial feature and temporal feature from the NIR facial expressions. The local facial feature and global feature are fused in the developed network by multiple streams.

- **MA-FER [66]:** To improve the performance of the expression recognition, this method introduces the visual attention mechanisms to the facial feature extraction process. Two important steps are included in the developed network, such as the expression recognition sub-net and region-aware sub-net.

- **DBMNet [35]:** Aiming to solve the problems of ambiguity and continuity in the natural expressions, this method introduces the manifold learning to the proposed network. The discriminative features can preserve the local affinity of expression feature and manifold structures of expression label.

- **CDLLNet:** The Cauchy distribution is introduced to construct the multi-label in the IR facial expressions. To measure the ground-truth label and predicted label, this method introduces the KL divergence.

- (2) **Evaluation index:** Many indexes have been proposed to evaluate the FER rate. In this paper, we focus on the error between predicted label and ground-truth one. The recognition rate can determine the performance of the comparing methods on capturing the essential feature of NIR facial expression images. Therefore, the prediction accuracy is measured by two famous indexes[67,68], such as root mean square error ($RMSE$), and mean accuracy (ACC). The $RMSE$ and ACC indexes are formulated as

$$RMSE = \sqrt{\frac{1}{|R_{label}|} \sum_{k=1}^M (g - \hat{g})^2}, \quad (9)$$

and

$$ACC = \frac{1}{m} \sum_{k=1}^M acc_k, \quad (10)$$

where acc_k denotes the accuracy in k -th validation, and m denotes the cross validation numbers. The larger the ACC value achieves, the better performance of the developed network.

- (3) **Datasets:** To validate the proposed network, we introduce two facial expression datasets in this article. One dataset is about the NIR facial images, and the other consists of RGB images.

- (a) **Oulu-CASIA Database [31]:** The infrared database is captured by the visible and near-infrared cameras. The facial images sequences are selected from the infrared videos. All the facial expression images can be classified into six types, such as the disgust, anger, happy, sadness, surprise and fear expression. Each expression includes 480 sequence images and is labeled with number and expression text. In this experiment, we choose four image frames as the network input in each expression sequences.

- (b) **CK+ dataset:** This dataset is collected in lab environment, which contains 327 video sequences from 123 persons. Seven basic expressions are included in this dataset. Each sequence displays the expression from the neutral to the peak frame. Only last three frames of each sequence are chosen for each input.

Both the Oulu-CASIA and CK+ datasets provide the groundtruth labels. Firstly, we construct the soft labels according to the Cauchy distribution, and then crop and resize the facial expression images to the same resolution. All the parameters in the comparing methods are adjusted the best ones according to their papers. To gain the unbiased and objective results, the 20–80% test-train settings are adopted and

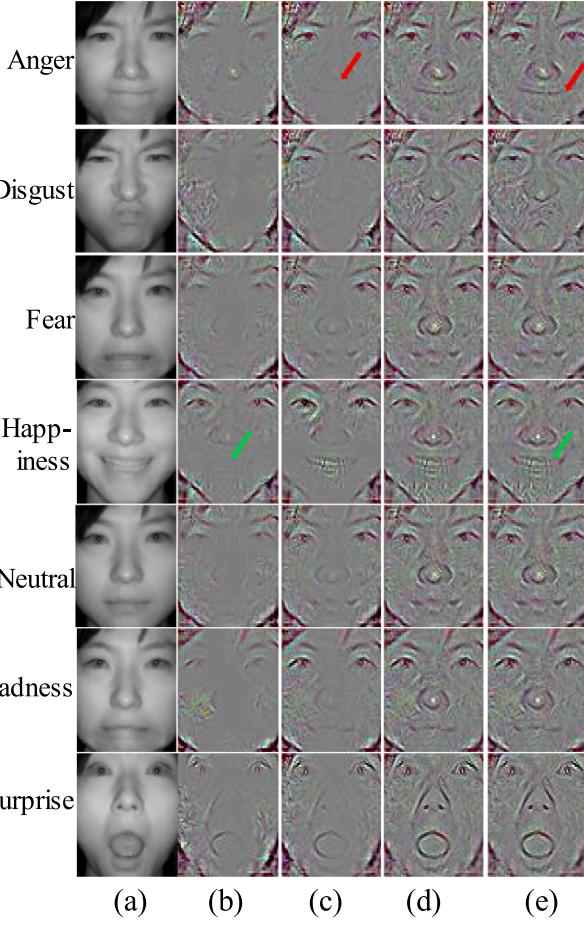


Fig. 6. Visualization results among three compared models on the Oulu_CASIA dataset. (a) Original IR facial expression images. (b) E3DNet method. (c) MA-FER method. (d) DBMNet method. (e) Proposed CDLLNet.

sixfold cross-validation is utilized.

3.2. Accuracy analysis of CDLLNet

To validate the performance of CDLLNet, several methods are introduced for the comparison experiments. In Table 1, we plot the indexes of ACC. In Fig. 6, the visualization results is presented with the three compared models on the Oulu_CASIA dataset. Generally speaking, the traditional models have a worse performance than the deep learning models. The CDLLNet and DBMNet demonstrate great improvement, compared with the baseline model LBP-FER.

In Table 1, the results of total FER recognition ratios are presented among CDLLNet model and four comparing methods. The MA-FER is leveraged as the backbone, and fine-tuned on the Oulu_CASIA dataset. With the basic CNN frameworks, only 84.56% recognition accuracy is achieved. When the soft label distribution is adopted in the CDLLNet model, the average recognition ratio can be raised 1.72%, comparing with the CNN-based method. The DBMNet method achieves an accuracy of 85.36%, which outperforms the MA-FER fine-tuning method. Since DBMNet adopts the manifold embedding learning to extract the discriminative features for blended expressions. Combining with the multi-label learning technology, this method can preserve the manifold structure of emotion labels.

Zhao et al. [31] proposed a dynamic FER method with local binary patterns feature descriptor (LBP-FER) for the infrared expression sequences. This method has the advantage of illumination invariant for the strong, dark and weak lighting cases. It obtains an accuracy of 70.12%, making an ensemble of classification result among non CNN based

Table 2

Performance comparison by LBP-FER, E3DNet, MA-FER, DBMNet, and CDLLNet methods on the CK+ database.

Algorithms	Categories	Accuracy
LBP-FER [31]	Seven	71.25%
E3DNet [29]	Seven	80.36%
MA-FER [66]	Seven	83.28%
DBMNet [35]	Seven	85.27%
Ours (CDLLNet)	Seven	87.42%

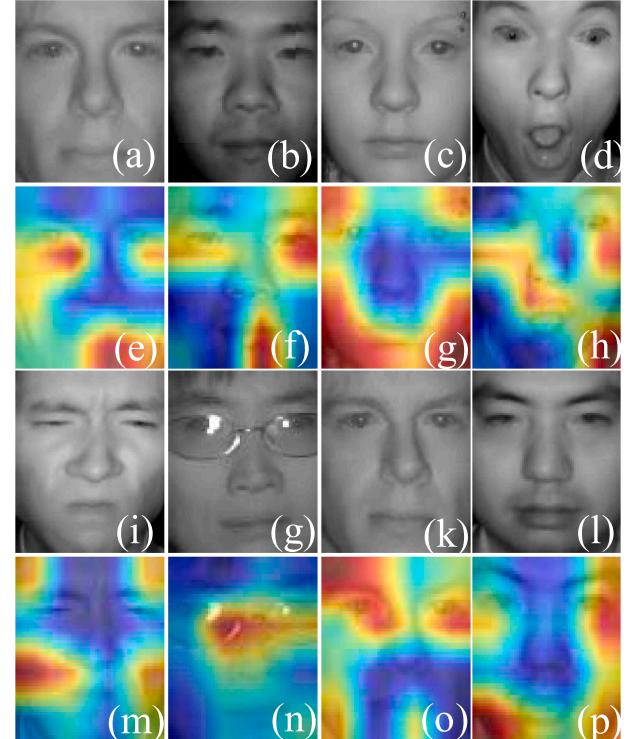


Fig. 7. Class activations from different infrared facial expressions on Oulu_CASIA dataset. Only four expressions are chosen to be visually displayed.

method. Hoever, the proposed CDLLNet method achieves 86.28% recognition ratio, which is much better than the LBP-FER method. Furthermore, the CDLLNet model obtains significant performance then the state-of-the-art models [35,66,29]. In Table 2, the proposed CDLLNet method achieves 87.42% accuracy on the CK+ dataset.

3.3. Visualization for understanding

To reveal the essential features for different infrared expressions, the grad-cam maps tool [69] is leveraged to demonstrate the important feature region in the facial images. The baseline methods DBMNet is also introduced to compare with the proposed CDLLNet method. The visualization results are illustrated in Fig. 6. The first row images are the four original infrared facial expressions. From the second to four row, it shows the results of proposed CDLLNet network and DBMNet methods, respectively. In Fig. 6(b), the learned features are located at the eyes, mouth and nose regions, which are very significant in the facial images. In Fig. 7(c) and (d), the learned features are relatively messy and unclear. Some important facial components are missed, which are shown by the red arrows (mouth) and green arrows (nose). As can be seen, the proposed CDLLNet model learns more accurate features than the other two comparing methods.

In Fig. 7, it shows the class activations of different persons by the proposed model. Only four expressions from the Oulu-CASIA dataset are

Table 3

Confusion matrix by the DBMNet model on the Oulu-CASIA database.

		Predicted label						
		SU	FE	DI	HA	SA	AN	NE
True Cauchy label	SU	87%	3.0%	6.0%	0.0%	2.0%	1.0%	0.0%
	FE	6.0%	82%	3.0%	7.0%	2.0%	1.0%	0.0%
	DI	2.0%	4.0%	77%	1.0%	6.0%	10%	0.0%
	HA	0.0%	1.0%	1.0%	93%	1.0%	0.0%	2.0%
	SA	0.0%	6.0%	4.0%	1.0%	81%	9.0%	4.0%
	AN	0.0%	1.0%	9.0%	2.0%	7.0%	80%	7.0%
	NE	0.0%	0.0%	0.0%	0.0%	6.0%	8.0%	82%

Table 4

Confusion matrix by the developed CDLLNet method on the Oulu-CASIA database.

		Predicted label						
		SU	FE	DI	HA	SA	AN	NE
True Cauchy label	SU	93%	2.0%	5.0%	0.0%	0.0%	1.0%	0.0%
	FE	4.0%	84%	2.0%	6.0%	0.0%	1.0%	0.0%
	DI	2.0%	3.0%	81%	1.0%	0.0%	2.0%	0.0%
	HA	1.0%	1.0%	1.0%	95%	4.0%	0.0%	0.0%
	SA	0.0%	4.0%	4.0%	1.0%	83%	7.0%	7.0%
	AN	0.0%	0.0%	6.0%	2.0%	6.0%	86%	6.0%
	NE	0.0%	0.0%	0.0%	0.0%	8.0%	85%	

chosen for the visualization. Since the infrared images are less-textured and illumination insensitive, many facial details, such as eyebrow and tooth, are blurry and unsharp. Thus, the discriminative features avoid these regions. Also, with different person facial expressions, the expression features appear in the different regions.

3.4. Convergence and computation complexity

In Table 3, the facial expression classification result of the DBMNet is demonstrated by the confusion matrix. For the proposed CDLLNet method (average 86.71% in Table 4), it obtains the 95% recognition ratio on happy expression, since this type of expression can be distinguished easily. For the sadness and disgust expressions, only 81% accuracies ratios are obtained. The result is lower than other type of

expressions. Sadness is very similar to the surprise expression, which is often misrecognized as the surprise. The same phenomenon occurs on the disgust and neutral expressions. However, the recognition accuracies of the proposed CDLLNet method are better than those of the state-of-the-art DBMNet method. It is caused by the fact that the Cauchy distribution label can distinguish the most similar expressions.

3.5. Multi-person case in the classroom

Afterward, the proposed CDLLNet is executed on the multi-person case in the teaching environment. When the students are learning in the classroom, the crowd expression images are captured and recognized by the proposed method. The FER results are shown in Fig. 8. It can be observed that all the student faces can be detected and expressions are well recognized. The faces are marked by the different size boxes. The expression labels are texted around the boxes. The recognized expressions are very natural in the real scenarios. Four images in Fig. 8 demonstrate that the Cauchy distribution-based multi-label learning can work well in the proposed CDLLNet method.

4. Conclusion

In this paper, we develop a new facial expression recognition method under infrared imaging. The loss function includes a L_2 regression loss, which is combined with Cauchy distribution label learning. We address the expression ambiguity observed in the IR expression images. Firstly, we compute the similarities values of different facial expressions with the same subject. Then, the similarities values are employed to construct the expression multi-label by the Cauchy distribution function. The expression multi-label used the proposed network can not only learn the features of its own real expression image, but also learn its neighboring expression images. The performance of proposed network is evaluated on two facial expression datasets: Oulu-CASIA and CK+. Several qualitative and quantitative experimental results verify that the CDLLNet network can achieve robust results and significantly outperforms the existing state-of-the-art facial expression algorithms. In future, we will introduce the graph neural network [38,70,71] and attention mechanism [38] for the infrared expression recognition tasks.



Fig. 8. Facial expression recognition results for the crowd in the classroom.

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.

Acknowledgements

The authors sincerely thank anonymous reviewers for their constructive comments, which helped improve this article. This work was supported in part by the National Natural Science Foundation of China under Grant 62005092, 62011530436, and 61875068, the China's ministry level project under Grant MCM20160409, and the National Science Foundation under Grant 18YJA880090.

References

- [1] Y. Liu, X. Zhang, Y. Lin, H. Wang, Facial expression recognition via deep action units graph network based on psychological mechanism, *IEEE Trans. Cogn. Dev. Syst.* 12 (2020) 311–322.
- [2] J. Wang, T. Liu, X. Wang, Human hand gesture recognition with convolutional neural networks for K-12 double-teachers instruction mode classroom, *Infrared Phys. Technol.* 111 (2020) 103464.
- [3] H. Liu, Z. Zhang, J. Sun, S. Liu, Blind spectral deconvolution algorithm for Raman spectrum with Poisson noise, *Photon. Res.* 2 (2014) 168–171.
- [4] F. Zhang, T. Zhang, Q. Mao, C. Xu, Geometry Guided Pose-Invariant Facial Expression Recognition, *IEEE Trans. Image Process.* 29 (2020) 4445–4460.
- [5] H. Liu, Z. Zhang, S. Liu, J. Shu, T. Liu, L. Yan, T. Zhang, Spectral blind deconvolution with differential entropy regularization for infrared spectrum, *Infrared Phys. Technol.* 71 (2015) 481–491.
- [6] Z. Huang, L. Chen, Y. Zhang, Z. Yu, H. Fang, T. Zhang, Robust contact-point detection from pantograph-catenary infrared images by employing horizontal-vertical enhancement operator, *Infrared Phys. Technol.* 101 (2019) 146–155.
- [7] T. Liu, H. Liu, Z. Chen, A.M. Lesgold, Fast blind instrument function estimation method for industrial infrared spectrometers, *IEEE Trans. Ind. Inf.* 14 (2018) 5268–5277.
- [8] S. Li, W. Deng, Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, *IEEE Trans. Image Process.* 28 (2019) 356–370.
- [9] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu, J. Jiang, Infrared and visible image fusion via detail preserving adversarial learning, *Inf. Fusion* 54 (2020) 85–98.
- [10] L. Xu, J. Chen, Y. Gan, Head pose estimation with soft labels using regularized convolutional neural network, *Neurocomputing* 337 (2019) 339–353.
- [11] T. Liu, H. Liu, Y. Li, Z. Chen, Z. Zhang, S. Liu, Flexible FTIR spectral imaging enhancement for industrial robot infrared vision sensing, *IEEE Trans. Ind. Inf.* 16 (2020) 544–554.
- [12] Z. Zhang, Z. Li, H. Liu, T. Cao, S. Liu, Data-driven online learning engagement detection via facial expression and mouse behavior recognition technology, *J. Educ. Comput. Res.* 58 (2020) 63–83.
- [13] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, G. Yang, cGAN based facial expression recognition for human-robot interaction, *IEEE Access* 7 (2019) 9848–9859.
- [14] M. Liu, Y. Li, H. Liu, 3D gaze estimation for head-mounted eye tracking system with auto-calibration method, *IEEE Access* 8 (2020) 104207–104215.
- [15] T. Liu, Y. Li, H. Liu, Z. Zhang, S. Liu, RISIR: Rapid infrared spectral imaging restoration model for industrial material detection in intelligent video systems, *IEEE Trans. Ind. Inf.* (2020), <https://doi.org/10.1109/TII.2019.2930463>.
- [16] J. Ma, J. Zhao, J. Jiang, H. Zhou, X. Guo, Locality preserving matching, *Int. J. Comput. Vision* 127 (2019) 512–531.
- [17] H. Liu, S. Liu, Z. Zhang, J. Sun, J. Shu, Adaptive total variation-based spectral deconvolution with the split Bregman method, *Appl. Opt.* 53 (2014) 8240–8248.
- [18] T. Liu, Z. Chen, S. Liu, Z. Zhang, J. Shu, Blind image restoration with sparse prior regularization for passive millimeter-wave images, *J. Visual Commun. Image Represent.* 40 (Part A) (2016) 58–66.
- [19] T. Huang, H. Liu, Z. Zhang, S. Liu, T. Liu, X. Shen, T. Zhang, J. Zhang, Blind deconvolution using the similarity of multiscales regularization for infrared spectrum, *Meas. Sci. Technol.* 26 (2015) 115502.
- [20] H. Liu, M. Zhou, Z. Zhang, J. Shu, T. Liu, T. Zhang, Multi-order blind deconvolution algorithm with adaptive Tikhonov regularization for infrared spectroscopic data, *Infrared Phys. Technol.* 71 (2015) 63–69.
- [21] Z. Huang, Y. Zhang, Q. Li, X. Li, T. Zhang, N. Sang, H. Hong, Joint analysis and weighted synthesis sparsity priors for simultaneous denoising and destriping optical remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 58 (2020) 6958–6982.
- [22] T. Liu, Z. Chen, Z. Zhang, H. Liu, D. Kong, Computer-assisted teaching environment for digital signal recognition course learning, in: International Symposium on Educational Technology (ISET) 2018, 2018, pp. 240–244.
- [23] Y. Yan, Y. Huang, S. Chen, C. Shen, H. Wang, Joint deep learning of facial expression synthesis and recognition, *IEEE Trans. Multimedia* 22 (2020) 2792–2807.
- [24] T. Liu, Z. Chen, H. Liu, Z. Zhang, Y. Chen, Multi-modal hand gesture designing in multi-screen touchable teaching system for human-computer interaction, in: 2018 the 2nd International Conference on Advances in Image Processing, Chengdu, Sichuan, 2018, 2018, pp. 100–109.
- [25] T. Liu, Z. Chen, H. Liu, Z. Zhang, FTIR spectral imaging enhancement for teacher's facial expressions recognition in the intelligent learning environment, *Infrared Phys. Technol.* 93 (2018) 213–222.
- [26] Z. Huang, Y. Zhang, Q. Li, Z. Li, T. Zhang, N. Sang, S. Xiong, Unidirectional variation and deep CNN denoiser priors for simultaneously destriping and denoising optical remote sensing images, *Int. J. Remote Sens.* 40 (2019) 5737–5748.
- [27] T. Liu, H. Liu, Z. Chen, Y. Chen, S. Wang, Z. Liu, H. Zhang, FBRDLR: Fast blind reconstruction approach with dictionary learning regularization for infrared microscopy spectra, *Infrared Phys. Technol.* 90 (2018) 101–109.
- [28] H. Zheng, X. Geng, D. Tao, Z. Jin, A multi-task model for simultaneous face identification and facial expression recognition, *Neurocomputing* 171 (2016) 515–523.
- [29] Z. Wu, T. Chen, Y. Chen, Z. Zhang, G. Liu, NIRExpNet: Three-stream 3D convolutional neural network for near infrared facial expression recognition, *Appl. Sci.* 7 (2017) 1184.
- [30] C.-M. Kuo, S.-H. Lai, M. Sarkis, A compact deep learning model for robust facial expression recognition, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2018, 2018, pp. 2202–2208.
- [31] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. Pietikäinen, Facial expression recognition from near-infrared videos, *Image Vis. Comput.* 29 (2011) 607–619.
- [32] K. Wang, X. Peng, J. Yang, S. Lu, Y. Qiao, Suppressing uncertainties for large-scale facial expression recognition, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020, 2020, pp. 6896–6905.
- [33] J. Lee, S. Kim, S. Kim, K. Sohn, Multi-modal recurrent attention networks for facial expression recognition, *IEEE Trans. Image Process.* 29 (2020) 6977–6991.
- [34] Y. Chen, Z. Zhang, L. Zhong, T. Chen, J. Chen, Y. Yu, Three-stream convolutional neural network with squeeze-and-excitation block for near-infrared facial expression recognition, *Electronics* 8 (2019) 385.
- [35] S. Li, W. Deng, Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning, *Int. J. Comput. Vision* 127 (2019) 884–906.
- [36] Y. Zhong, A. Maki, Regularizing CNN transfer learning with randomised regression, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020, 2020, pp. 13634–13643.
- [37] L. Yan, H. Liu, S. Zhong, H. Fang, Semi-blind spectral deconvolution with adaptive Tikhonov regularization, *Appl. Spectrosc.* 66 (2012) 1334–1346.
- [38] Y. Li, J. Zeng, S. Shan, X. Chen, Occlusion aware facial expression recognition using CNN with attention mechanism, *IEEE Trans. Image Process.* 28 (2019) 2439–2450.
- [39] H. Liu, Z. Zhang, S. Liu, T. Liu, L. Yan, T. Zhang, Richardson-Lucy blind deconvolution of spectroscopic data with wavelet regularization, *Appl. Opt.* 54 (2015) 1770–1775.
- [40] Z. Zhang, C. Lai, H. Liu, Y.-F. Li, Infrared facial expression recognition via Gaussian-based label distribution learning in the dark illumination environment for human emotion detection, *Neurocomputing* 409 (2020) 341–350.
- [41] H. Liu, L. Yan, Y. Chang, H. Fang, T. Zhang, Spectral deconvolution and feature extraction with robust adaptive Tikhonov regularization, *IEEE Trans. Instrum. Meas.* 62 (2013) 315–327.
- [42] L. Yan, H. Liu, L. Chen, H. Fang, Y. Chang, T. Zhang, Parametric semi-blind deconvolution algorithm with Huber–Markov regularization for passive millimeter-wave images, *J. Mod. Opt.* 60 (2013) 970–982.
- [43] H. Liu, Z. Zhang, S. Liu, J. Shu, T. Liu, T. Zhang, Blind spectrum reconstruction algorithm with L0-sparse representation, *Meas. Sci. Technol.* 26 (2015), 085502 (085501–085507).
- [44] Y. Zhang, J. Ye, D. Niu, P. Cao, Facial expression recognition based on Gabor wavelet transformation and elastic templates matching, *Int. J. Image Graph.* 06 (2006) 125–138.
- [45] H. Liu, Z. Zhang, S. Liu, L. Yan, T. Liu, T. Zhang, Joint Baseline-Correction and Denoising for Raman Spectra, *Appl. Spectrosc.* 69 (2015) 1013–1022.
- [46] Z. Huang, Y. Zhang, Q. Li, T. Zhang, N. Sang, Spatially adaptive denoising for X-ray cardiovascular angiogram images, *Biomed. Signal Process. Control* 40 (2018) 131–139.
- [47] M. Nazir, Z. Jan, M. Sajjad, Facial expression recognition using histogram of oriented gradients based transformed features, *Cluster Comput.* 21 (2018) 539–548.
- [48] H. Liu, T. Zhang, L. Yan, H. Fang, Y. Chang, A MAP-based algorithm for spectroscopic semi-blind deconvolution, *Analyst* 137 (2012) 3862–3873.
- [49] Z. Huang, Y. Zhang, Q. Li, T. Zhang, N. Sang, H. Hong, Progressive dual-domain filter for enhancing and denoising optical remote-sensing images, *IEEE Geosci. Remote Sens. Lett.* 15 (2018) 759–763.
- [50] Y. Ding, Q. Zhao, B. Li, X. Yuan, Facial expression recognition from image sequence based on LBP and Taylor expansion, *IEEE Access* 5 (2017) 19409–19419.
- [51] W.-L. Chao, J.-J. Ding, J.-Z. Liu, Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection, *Signal Process.* 117 (2015) 1–10.
- [52] H. Liu, S. Liu, T. Huang, Z. Zhang, Y. Hu, T. Zhang, Infrared spectrum blind deconvolution algorithm via learned dictionaries and sparse representation, *Appl. Opt.* 55 (2016) 2813–2818.
- [53] D. McDuff, e.R. Kaliouby, Applications of automated facial coding in media measurement, *IEEE Trans. Affective Comput.* 8 (2017) 148–160.

- [54] H. Liu, L. Yan, T. Huang, S. Liu, Z. Zhang, Blind spectral signal deconvolution with sparsity regularization: an iteratively reweighted least-squares solution, *Circ. Syst. Signal Process.* 36 (2017) 435–446.
- [55] S.E. Kahou, C. Pal, X. Bouthillier, Combining modality specific deep neural networks for emotion recognition in video, in: Proceedings of the 15th ACM on international conference on multimodal interaction, 2013, pp. 543–550.
- [56] H. Liu, Y. Li, Z. Zhang, S. Liu, T. Liu, Blind Poissonian reconstruction algorithm via curvelet regularization for an FTIR spectrometer, *Opt. Express* 26 (2018) 22837–22856.
- [57] Z. Huang, H. Fang, Q. Li, Z. Li, T. Zhang, N. Sang, Y. Li, Optical remote sensing image enhancement with weak structure preservation via spatially adaptive gamma correction, *Infrared Phys. Technol.* 94 (2018) 38–47.
- [58] T. Liu, Z. Chen, X. Wang, Automatic instructional pointing gesture recognition by machine learning in the intelligent learning environment, in: Proceedings of the 2019 4th International Conference on Distance Education and Learning, ACM, Shanghai, China, 2019, pp. 153–157.
- [59] L. Zheng, K. Jia, T. Bi, Y. Fang, Z. Yang, Cosine similarity based line protection for large scale wind farms, *IEEE Trans. Ind. Electron.* (2020) 1, <https://doi.org/10.1109/TIE.2020.2998756>.
- [60] H. Liu, H. Nie, Z. Zhang, Y.-F. Li, Anisotropic angle distribution learning for head pose estimation, *Neurocomputing* (2020), <https://doi.org/10.1016/j.neucom.2020.09.068>.
- [61] Z. Huang, L. Huang, Q. Li, T. Zhang, N. Sang, Framelet regularization for uneven intensity correction of color images with illumination and reflectance estimation, *Neurocomputing* 314 (2018) 154–168.
- [62] P. Zhou, Y. Lv, H. Wang, T. Chai, Data-driven robust RVFLNs modeling of a blast furnace iron-making process using cauchy distribution weighted M-estimation, *IEEE Trans. Ind. Electron.* 64 (2017) 7141–7151.
- [63] M. Liu, Y.-F. Li, H. Liu, Towards robust auto-calibration for head-mounted gaze tracking systems, in: IEEE International Conference on Mechatronics and Automation (ICMA) 2020, 2020, pp. 588–593.
- [64] Z. Huang, Q. Li, T. Zhang, N. Sang, H. Hong, Iterative weighted sparse representation for X-ray cardiovascular angiogram image denoising over learned dictionary, *IET Image Proc.* 12 (2018) 254–261.
- [65] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe, Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, Association for Computing Machinery, Orlando, Florida, USA, 2014, pp. 675–678.
- [66] Y. Gan, J. Chen, Z. Yang, L. Xu, Multiple attention network for facial expression recognition, *IEEE Access* 8 (2020) 7383–7393.
- [67] H. Liu, Y.-F. Li, D. Su, Z. Zhang, S. Liu, DISR: Deep infrared spectral restoration algorithm for robot sensing and intelligent visual tracking systems, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2019, 2019, pp. 8012–8017.
- [68] T. Liu, H. Liu, Y. Li, Z. Zhang, S. Liu, Efficient blind signal reconstruction with wavelet transforms regularization for educational robot infrared vision sensing, *IEEE/ASME Trans. Mechatron.* 24 (2019) 384–394.
- [69] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: IEEE International Conference on Computer Vision (ICCV) 2017, 2017, pp. 618–626.
- [70] Z. Zhang, H. Liu, J. Shu, H. Nie, N. Xiong, On automatic recommender algorithm with regularized convolutional neural network and IR technology in the self-regulated learning process, *Infrared Phys. Technol.* 105 (2020) 103211.
- [71] Z. Zhang, Z. Li, H. Liu, N.N. Xiong, Multi-scale Dynamic Convolutional Network for Knowledge Graph Embedding, *IEEE Trans. Knowl. Data Eng.* (2020), <https://doi.org/10.1109/TKDE.2020.3005952>.