

Tackling Micro-Expression Data Shortage via Dataset Alignment and Active Learning

Xianye Ben, *Senior Member, IEEE*, Chen Gong, *Member, IEEE*, Tianhuan Huang, Chuanye Li, Rui Yan and Yujun Li, *Member, IEEE*

Abstract—The research on micro-expression recognition has been drawing great attention in recent years, because of its great potential in the lie detection, clinical diagnosis, and national security. Amongst many challenges, data shortage stands out as it directly prevents an accurate training of micro-expression recognition algorithm. In this work, we present our approach within a dataset alignment and active learning (DAAL) framework. DAAL effectively queries minimum examples to label, as well as transfers features from micro-expression dataset to macro-expression dataset. Specifically, the features from micro-expression dataset are mapped to the macro-expression dataset with a translator, so that the classifier trained in macro-expression dataset can be adjusted and adapted to boost the classification performance on the micro-expression dataset. Besides, the most informative examples in the micro-expression dataset are selected through active learning in an iterative way, which effectively improves the classification ability of the model. Comprehensive experiments on CASME, CASME II, SAMM and SMIC databases firmly demonstrate that the proposed DAAL outperforms previous works by a large margin on micro-expression recognition task.

Index Terms—Dataset alignment, active learning, micro-expression recognition.

I. INTRODUCTION

MICRO-EXPRESSION is a facial action with an extremely short duration that can reveal the real emotion a person tries to hide. Thus, micro-expression has great potential in lie detection, clinical diagnosis, national security, and other fields. Compared with ordinary dynamic facial expressions, micro-expressions have much lower intensity and shorter duration, which makes the automatic recognition very difficult.

Currently, micro-expression recognition methods can be summarized into three categories. The first category is based on traditional feature extractors [1]. Up to now, various feature descriptors have been proposed to capture low-intensity micro-expression features. Zong et al. [2] proposed a feature

descriptor of spatiotemporal hierarchy to enhance the ability of describing micro-expressions, and a Kernel Group Sparse Learning model was introduced to deal with micro-expression features with hierarchical structure. This method has achieved some promising recognition results on both CASME II and SMIC database. The second category is based on deep learning [3]–[5]. For example, Li et al. [6] applied a 3D flow-based CNN model to video-based micro-expression recognition. The deep learning features can better represent the fine motion flow generated by subtle facial motion, which greatly enhances the recognition performance. In the same year, Li et al. [7] proposed a deep convolutional network model based on a single APEX frame. Through extracting the APEX frame of the micro-expression video sequences and fine-tuning the VGG-Face model, this algorithm achieves the accuracy of 63%. Compared with the naked-eye recognized results, the methods of these two categories have achieved significant improvements. However, due to the limited micro-expression examples, the performances of these methods are all limited fundamentally. The third category is based on transfer learning [8]–[10]. Its essential idea is to transfer the useful information from the source domain to the target domain to assist the recognition tasks in target domain. Based on transfer learning, Peng et al. [11] proposed a convolutional neural network to recognize the micro-expressions and obtained some promising results. Transfer learning is a good assistance to micro-expression recognition, but they still need a certain number of labeled examples so as to train effective models. Thus, how to effectively recognize micro-expressions in case of extremely small size of labeled examples is still a serious challenge.

In this paper, we innovatively propose the dataset alignment and active learning (DAAL) framework for micro-expressions recognition with an extremely small sample size. Dataset alignment aims to acquire useful knowledge from the macro-expression dataset to assist learning tasks in the micro-expression dataset. The problem of small sample size of micro-expression recognition is solved by dataset alignment, by which a translator can be learned from the micro-expression feature dataset to the macro-expression feature dataset. Active learning can train an effective classifier with minimum examples by selecting a few but high-quality examples. In the active learning stage, the uncertainty criterion is firstly adopted to establish an initial active learning model in the micro-expression dataset. Subsequently, the most informative examples in the micro-expression dataset are selected from a candidate set in an iterative way, and added into the training set after manual labeling, which can gradually improve the

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515010186, in part by the Natural Science Foundation of China under Grant 61971468 and Grant 61973162, in part by the Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) under Grant 2019JZZY010119, and in part by the Fundamental Research Funds for the Central Universities (Nos: 2022JC017, 30920032202, 30921013114). (Corresponding authors: Xianye Ben, Yujun Li)

X. Ben, T. Huang, C. Li, Y. Li are with the School of Information Science and Engineering, Shandong University, Qingdao 266237, China (e-mail: benxianye@gmail.com; liyujun@sdu.edu.cn).

C. Gong is with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, School of Computer Science and Engineering, Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: chen.gong@njust.edu.cn).

R. Yan is with Amazon, Bellevue, WA, 98004, USA (email: raymondino.yan@gmail.com).

classification model. Note that the dataset alignment designed in this paper is different from transfer learning, as transfer learning is not applicable in our problem. If we transfer the data from macro-expression space into micro-expression space with transfer learning, the parameters of classifier would be initialized with the micro-expression data. However, the micro-expression examples can be used is very limited, which makes the initial training impossible.

The main technical contributions of this paper can be summarized as follows:

- (1) This paper proposes a new Dataset Alignment and Active Learning (DAAL) framework to solve the problem of insufficient micro-expression examples. By using the uncertainty criterion in iterative query process, high-quality micro-expression examples are selected to build a good micro-expression recognition model.
- (2) The micro-expression data is translated and together with macro-expression data for the establishment of the initial DAAL model in this paper. This translation process for achieving dataset alignment between the micro- and macro-expression datasets is designed, which can further improve the recognition performance.

The rest of this paper is organized as follows. In Section II, the state-of-the-art research of micro-expression recognition and the related active learning are introduced in detail. The proposed DAAL framework and its optimization solution are given in Section III. Then, the experimental validation and performance evaluation are presented in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

A. Micro-expression Recognition

The traditional automatic micro-expression recognition method is mainly composed of feature extraction and classification. To accomplish feature extraction, Pfister et al. [12] extended the features of the Local Binary Pattern (LBP) by adding time series and achieved the dynamic micro-expressions analysis. Zhao et al. [13] used three orthogonal plane local binary patterns (LBP-TOP) to describe the micro-expression video clips. Experimental results show that the LBP-TOP feature is effective for micro-expression recognition tasks. Based on Zhao et al. [13], Ruiz-Hernandez et al. [14] employed re-parameterization of second order Gaussian jet to boost LBP-TOP. In order to better describe micro-expressions, Wang et al. [15] proposed a new time-space descriptor with six intersection points named LBP-SIP. LBP-SIP can effectively reduce the redundant information in LBP-TOP, thereby improving the recognition efficiency. Huang et al. developed a variety of spatiotemporal descriptors, such as SpatioTemporal LBP with Integrated Projection (STLBP-IP) [16] and completed local quantized pattern-TOP (CLQP-TOP) [17]. In addition to the above spatial-temporal based descriptors, optical flow-based features have also been studied. Xu et al. [18] proposed a Facial Dynamics Map (FDM) to analyze the dynamic characteristics of micro-expressions. Liu

et al. [19] proposed a simple and effective Main Directional Mean Optical flow (MDMO) to extract micro-expression features. MDMO is a normalized statistical feature based on the region of interest. It fully considers temporal and spatial information and performs well. After extracting features, the traditional classifiers such as SVM and K nearest neighbor (KNN) classifier [20] are adopted to conduct micro-expression recognition. Although the above methods based on feature extraction and classification can recognize micro-expressions to some extent, their recognition ability is far from ideal due to the small sample size and poor example quality.

In recent years, deep learning has also been applied to micro-expression recognition. Kim et al. [21] proposed a deep learning framework for micro-expression recognition which consists of a Convolutional Neural Network (CNN) and a Long-term Short-term Memory (LSTM) recursive network. In this framework, a representative emotion frame in each micro-expression video clip is first selected to train the CNN. Then the CNN features of each image frame in the video segment are extracted to train the LSTM network, which ultimately implementing micro-expression recognition. Takalkar et al. [22] proposed a CNN-based model that uses two micro-expression databases for data enhancement and generates a wide range of synthetic image datasets. This algorithm solves the small sample size problem of micro-expression recognition and obtains a high accuracy.

The micro-expression recognition method based on transfer learning has appeared only a few years ago, and the establishments of them are commonly based on singular value decomposition (SVD) [23], coupled source domain targetized [24], coupled metric learning [25], and transductive transfer regression (TTRM) [26]. Jia et al. [23] proposed a linear reconstruction of speech features to generate micro-expression features using SVD. Zhu et al. [24] proposed a coupled source-domain target recognition algorithm based on updated tag vectors, which transfers rich speech data to micro-expression and enhances the micro-expression recognition ability of the method. Ben et al. [25] used a coupled metric learning method to obtain a common subspace of micro-expression examples and trained a more accurate micro-expression recognition classifier. Zong et al. [26] paid attention to micro-expression examples from another database, of which feature distribution gap between the source and target domains is smaller than [23], [24], and [25]. Sun et al. [8] proposed a knowledge transfer technique that distills and transfers knowledge from action unit for micro-expression recognition. The knowledge from a pre-trained deep teacher neural network is distilled and transferred to a shallow student neural network. Liu et al. [9] proposed a neural micro-expression recognizer to solve micro-expressions recognition tasks with small datasets. The part-based model and two domain adaptation techniques were their main contributions. Xia et al. [10] proposed two Expression Identity Disentangle Network, named MicroNet and MacroNet, as the feature extractor to disentangle expression-related features for micro- and macro-expression samples and then fixed the MacroNet which is used to guide the fine-tuning of MicroNet from both label and feature space so that the MicroNet can efficiently capture the shared features of micro-

expression and macro-expression samples. These methods outperform the state-of-the-art non-transferring learning methods in the same period. Different from References [8]–[10], the proposed method aims to deal with the micro-expression recognition based on semi-supervised learning. Considering that there are extremely few labeled data in the training dataset, the proposed method does not adopt deep learning methods, but adopts an active learning method of making uses of the unlabeled data in the training dataset, meanwhile aligns these data with the data in another domain, and uses the classifier in another domain to predict and recognize micro-expressions.

B. Active Learning

Active learning [27], [28] is an effective method to solve small sample size problems. In the active learning process, the design of the query function is the most important. The two commonly used strategies are uncertainty criterion and diversity criterion. The uncertainty criterion is to find an example with high uncertainty, that is, to select an example with abundant information to adjust the training model. Diversity criterion refers to selecting from the most uncertain low-redundancy examples. Active learning algorithms based on uncertainty criterion can generally be divided into three categories.

The first category is based on the querying committee active learning, which considers the uncertainty of the example with the greatest disagreement between the learner committees. Tuia et al. [29] proposed Entropy-based Query by Bagging (EQB) to reduce computational complexity and the search time of hypothesis space. Copa et al. [30] improved the EQB algorithm to further avoid oversampling. Methods in this category have the advantage of high adaptability for all kinds of models, but require training a certain number of classifiers and have a high computational complexity.

The second category is based on the marginal classifier, which determines the uncertainty of the examples by measuring the distance from the candidate object to the marginal classifier. Support Vector Machine (SVM) [31] is a good basic algorithm of active learning. Schohn et al. [32] made full use of the geometric properties of SVM and proposed marginal sampling. The main strategy of marginal sampling is to select the examples that are closest to the classification hyperplane. To extend to multi-class cases, Demir et al. [33] proposed multi-level uncertain sampling that selects the most uncertain examples based on confidence. Since the uncertainty is learned with both distribution information and the kernel weights, it can capture the various complex data structure effectively and the query examples can improve the generalization of the classifier more significantly.

The third category is based on the posterior probability. These algorithms determine the uncertain region by analyzing the posterior probability for example selection. Thus, they rely on the posterior probability of the examples and show the satisfactory computation speed. Rajan et al. [34] proposed a new active learning algorithm, of which the key idea is to select examples with the greatest change in posterior probability, leading to great compatibility with posterior probability based classifiers.

Currently, researchers are combining active learning with deep learning. Since training a deep neural network requires a lot of labeled data, Wang et al [35] proposed an active learning algorithm based on deep learning (AL-DL), which can be applied to stacked RBMs or stacked self-encoder models. The disadvantage is that the AL-DL algorithm treats the deep training model and active learning as two separate processes. In order to effectively combine active learning with deep learning, Ranganathan [36] integrated active learning criteria into the loss function of the Deep Belief Network (DBN). The algorithm calculates the cross-entropy loss and entropy loss on the labeled examples simultaneously, aiming to form a joint loss function to optimize the DBN for better classification results.

It should be noted that the proposed DAAL is novel, although uncertainty criterion has been fully explored in other fields. However, compared to the prior active learning based uncertainty, the differences are mainly on three aspects: (1) In terms of motivation: The motivation from the prior active learning based on uncertainty is to select a few unlabeled examples to query their label by its uncertainty prediction since those samples with large uncertainty have greater probability of being more informative and representative. We are motivated by the uncertainty prediction. The novelty of the proposed method lies in the extension of such unlabeled-examples selection for querying their labels onto the cases that the number of labeled samples is extremely small in a target domain (micro-expressions) but with sufficient labeled samples of a relevant source domain (macro-expressions). (2) In terms of objective function: for instance, active learning by querying informative and representative examples (QUIRE) [43] is one of famous active learning methods based on uncertainty. QUIRE minimizes an evaluation function depending on uncertainty based on the labelled and unlabeled data, and simply provides a systematic way for selecting samples in the only one domain. However, the proposed DAAL minimizes classification error of examples from source and target domains and an uncertainty evaluation function of the features of all candidate examples translated from the target domain into the source domain. (3) In terms of application: This paper targets on the micro-expression recognition problem with the help of sufficient macro-expression data and a few unlabelled and labelled micro-expression data. Such heterogeneous data cannot be handled only by the uncertainty criterion. Besides uncertainty, the newly proposed DAAL alleviates the feature variance between micro-expressions and macro-expressions through mapping the micro-expression features from the target domain into the source domain by a translator. This means DAAL simultaneously alleviates feature bias between micro-expressions and macro-expressions and incorporates querying and manually labelling examples from the candidate unlabelled micro-expression examples by uncertainty criterion. The strategy assures the performance improvements on micro-expression recognition under data shortage. Additional performance comparisons between active learning methods including QUIRE and the proposed DAAL are included in Sections IV. F.

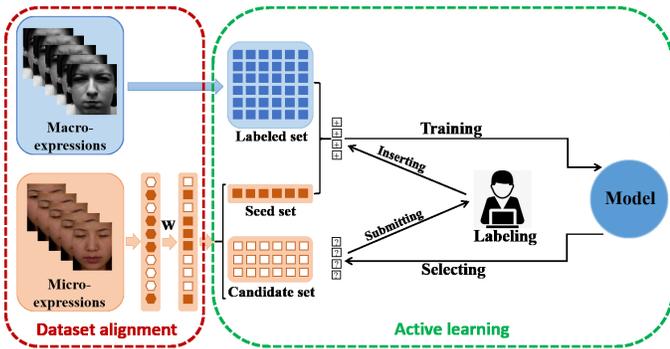


Fig. 1. Pipeline of the proposed DAAL framework. Colored blocks and white blocks represent labeled and unlabeled examples, respectively.

C. Active Transfer Learning

Recently, there are also some works that combine active learning with transfer learning to actively decide which examples can be transferred from source domain and target domain. For instance, Deng et al. [37] exploited stacked sparse auto encoder to transfer the knowledge from source domain to target domain, and utilized two active learning strategies for example selection, namely: 1) selecting a small number of the most informative examples from the target domain, and 2) removing other bad examples which are incompatible with the target distribution in the source domain. Lin et al. [38] proposed a deep-mapping-based heterogeneous transfer learning model via querying salient examples, and sought the correlation between source and target domains by using canonical correlation analysis layer by layer. However, negative similarities between source and target domains may lead to negative transfer problem. In order to eliminate this, Peng et al. [39] minimized the maximum mean discrepancy by introducing the orthogonal projection matrix and the weight coefficient vector, and proposed an information diversity term to select the informative and discriminative subsets from the source domain. In this paper, we aim to recognize micro-expressions, which are in the target domain, while the macro-expressions are from the source domain. The proposed translator in this paper can map the micro-expression examples to the macro-expression space. In addition, the classifier is trained by the macro- and micro-expression datasets, in which most of the examples are macro-expressions and only a few of them are micro-expressions. Therefore, the proposed method is different from active transfer learning.

III. THE PROPOSED METHODOLOGY

This section details the framework of Dataset Alignment and Active Learning (DAAL) and then describes the related optimization process.

A. Problem Description

Currently, the features extracted from micro-expression sequences by hand-crafted feature descriptors usually have very low discriminability, especially when the number of examples in current micro-expression datasets is too small. Therefore, DAAL framework is proposed to address such small sample problem. Some notations and descriptions are summarized in

Table I. The notations $\mathbf{x}_s^i \in \mathbb{R}^{d_s}$ is the i -th example from the macro-expression dataset, where d_s is feature dimension of this dataset. $\mathbf{x}_l^i \in \mathbb{R}^{d_t}$ and $\mathbf{x}_u^i \in \mathbb{R}^{d_t}$ are the i -th labeled and the i -th unlabeled micro-expression examples, respectively, where d_t is feature dimension of these datasets. Here $\mathcal{L}_s = \{(\mathbf{x}_s^i, \mathbf{y}_s^i)\}_{i=1}^{n_s}$ is a labeled training macro-expression dataset, where \mathbf{y}_s^i is the class label of \mathbf{x}_s^i , and n_s is the number of labeled examples in the macro-expression dataset. Besides, $\mathcal{L}_l = \{(\mathbf{x}_l^i, \mathbf{y}_l^i)\}_{i=1}^{n_l}$ is a labeled training micro-expression dataset, where \mathbf{y}_l^i is the label of \mathbf{x}_l^i , and n_l is the number of labeled examples in the micro-expression dataset. The micro-expression candidate dataset is defined as $\mathcal{L}_u = \{\mathbf{x}_u^i\}_{i=1}^{n_u}$ with unknown labels and $\mathbf{x}_u^i \in \mathbb{R}^{d_t}$, where n_u is the number of unlabeled examples in the micro-expression dataset.

Active learning relies on the knowledge of \mathcal{L}_l to train an initial classifier, then optimizes the performance of the classifier by iteratively selecting the most critical examples from \mathcal{L}_u . After that, these critical examples are manually labeled and then added to the training set. Thus, the quality of the initial classifier is critical to the overall performance of the algorithm. However, the sample size of \mathcal{L}_l is usually very small with little information, which makes it difficult to train an ideal classifier for micro-expression recognition. Therefore, we hope to improve the effect of active learning by introducing dataset alignment. Through dataset alignment, rich relevant information is linked together between the macro- and the micro-expression datasets, and thus the active learning aided model training can be carried out with the assistance of \mathcal{L}_s . Based on the above analysis, we aim to establish a joint learning framework of active learning and dataset alignment to improve the effect of micro-expression recognition.

The whole pipeline of the proposed algorithm is presented in Fig. 1. Through a translator W , the micro-expression examples are mapped into the macro-expression space, and an initial classification model is trained with the labelled macro-expression examples and limited labelled micro-expression examples in a seed set. Subsequently, the most informative examples in the micro-expression dataset are selected from a candidate set, which are added into the training set after manual labeling to further incrementally train the classification model. The above process iterates so that the final model with good classification performance can be obtained.

B. Dataset Alignment

Any classifier model f^* trained on labeled examples can be written as

$$f^* = \arg \min_{f \in \mathcal{F}} \left(\lambda J(f) + \sum_{i=1}^{n_l} \ell(\mathbf{y}_l^i, f(\mathbf{x}_l^i)) \right), \quad (1)$$

where \mathcal{F} is a hypothesis space; $J(f)$ represents a regularization term to constrain the complexity of the classification model f which is a functionality defined in the hypothesis space [40]; λ is a penalty coefficient on $J(f)$; and $\ell(\mathbf{y}_l^i, f(\mathbf{x}_l^i))$ denotes a loss function with $f(\mathbf{x}_l^i)$ being the output of f on \mathbf{x}_l^i .

Dataset alignment aims to acquire useful knowledge from the macro-expression dataset to assist the learning tasks in the

TABLE I
LIST OF IMPORTANT MATHEMATICAL NOTATIONS

Symbol	Description	Symbol	Description	Symbol	Description
\mathbf{x}_s^i	The i -th example from the macro-expression dataset.	\mathbf{x}_l^i	The i -th labeled example from the micro-expression dataset.	\mathbf{x}_u^i	The i -th unlabeled example from the micro-expression dataset.
\mathcal{L}_s	The labeled macro-expression dataset.	\mathcal{L}_l	The labeled micro-expression dataset.	\mathcal{L}_u	The unlabeled micro-expression candidate dataset.
\mathbf{y}_s^i	The class label of \mathbf{x}_s^i .	\mathbf{y}_l^i	The class label of \mathbf{x}_l^i .		
n_s	The number of labeled examples in the macro-expression dataset.	n_l	The number of labeled examples in the micro-expression dataset.	n_u	The number of unlabeled examples in the micro-expression dataset.
d_s	The feature dimension of the macro-expression dataset.	d_t	The feature dimension of the micro-expression dataset.		

micro-expression dataset. In order to ensure the consistency of data distribution between micro- and macro-expression datasets, we adopt a translator \mathbf{W} by learning a feature mapping from the micro-expression space to the macro-expression space. The common subspace of the two datasets can be obtained, which is helpful for the active learning step that will be introduced in Section III-C. Specifically, the dataset alignment is achieved by solving

$$\arg \min_{f, \mathbf{W}} \left(\begin{aligned} &\lambda J(f) + \sum_{i=1}^{n_s} (y_s^i - f(\mathbf{x}_s^i))^2 \\ &+ \sum_{i=1}^{n_l} (y_l^i - f(\mathbf{W}^\top \mathbf{x}_l^i))^2 + R(\mathbf{W}) \end{aligned} \right), \quad (2)$$

where $R(\mathbf{W})$ is a regularization term. By utilizing the translator matrix \mathbf{W} , any micro-expression example \mathbf{x}_l^i is mapped into the macro-expression space as $\mathbf{W}^\top \mathbf{x}_l^i$, so that the classification information of macro-expression dataset can be shared for micro-expression recognition.

The formulation (2) is to solve the binary classification problem, but the micro-expression recognition is a multi-classification problem. Therefore, we extend this model to a multi-class framework. The multi-class task can be divided into C binary-class tasks (C is the number of class), which are determined by classifiers f_1, f_2, \dots, f_C and the labels of examples can be defined as vectors $\mathbf{y}_s^i = (y_s^{i1}, y_s^{i2}, \dots, y_s^{iC})^\top$, $\mathbf{y}_l^i = (y_l^{i1}, y_l^{i2}, \dots, y_l^{iC})^\top$, where

$$y_s^{ic} = \begin{cases} 1, & \text{if } \mathbf{x}_s^i \text{ is a positive sample of the } c\text{-th classifier} \\ -1, & \text{otherwise} \end{cases}$$

and

$$y_l^{ic} = \begin{cases} 1, & \text{if } \mathbf{x}_l^i \text{ is a positive sample of the } c\text{-th classifier} \\ -1, & \text{otherwise} \end{cases}.$$

Therefore, a multi-class dataset alignment framework can be written as

$$\arg \min_{Q, f, \mathbf{W}} \sum_{c=1}^C \left(J(f_c) + \sum_{i=1}^{n_s} (y_s^{ic} - f_c(\mathbf{x}_s^i))^2 + \sum_{i=1}^{n_l} (y_l^{ic} - f_c(\mathbf{W}^\top \mathbf{x}_l^i))^2 \right) + R(\mathbf{W}), \quad (3)$$

where $\sum_{i=1}^{n_s} (y_s^{ic} - f_c(\mathbf{x}_s^i))^2$ represents the structure risk of labeled examples from the macro-expression dataset, and $\sum_{i=1}^{n_l} (y_l^{ic} - f_c(\mathbf{W}^\top \mathbf{x}_l^i))^2$ represents the structure risk of f_c on the labeled micro-expression examples which are translated by \mathbf{W} .

C. Active Learning

In the active learning, the criterion for picking up the critical examples from the candidate set is the core problem [41]. In this paper, we introduce structure risk minimization into active learning [42], and use the uncertainty criterion based on the decision boundary to evaluate the examples.

Based on the uncertainty criterion [32], the examples closest to the decision boundary are

$$Q^* = \arg \min_{\mathbf{x} \in Q} |f^*(\mathbf{x})|, \quad (4)$$

where Q^* and Q are respectively an optimal subset and a candidate set for example selection in \mathcal{L}_u . Then, the selected examples are added into the training set to optimize the classifier performance. Therefore, Eq. (1) can be further formulated as

$$\{f^*, Q^*\} = \arg \min_{Q, f} \left(\begin{aligned} &\lambda J(f) + \sum_{i=1}^{n_l} \ell(y_l^i, f(\mathbf{x}_l^i)) \\ &+ \sum_{\mathbf{x} \in Q} \ell(\hat{y}, f(\mathbf{x})) \end{aligned} \right), \quad (5)$$

where $\hat{y} \in \{-1, 1\}$ is pseudo label of \mathbf{x} . The maximum possible regularized risk after querying the samples in Q can be written as

$$\max_{\hat{y}: \forall \mathbf{x} \in Q} \min_{Q, f} \left(\begin{aligned} &\lambda J(f) + \sum_{i=1}^{n_l} \ell(y_l^i, f(\mathbf{x}_l^i)) \\ &+ \sum_{\mathbf{x} \in Q} \ell(\hat{y}, f(\mathbf{x})) \end{aligned} \right). \quad (6)$$

Fixing Q and f , we minimize the worst-case risk introduced by the query samples to solve Eq.(6) w.r.t. \hat{y} . The worst case should be that the pseudo labels are given by wrong labels, i. e.

$$\hat{y} = -\text{sign}(f(\mathbf{x})). \quad (7)$$

Then the related risk terms are represented by

$$\min_{Q, f} \left(\begin{aligned} &\lambda J(f) + \sum_{i=1}^{n_l} \ell(y_l^i, f(\mathbf{x}_l^i)) \\ &+ \sum_{\mathbf{x} \in Q} (f(\mathbf{x})^2 + 2|f(\mathbf{x})| + 1) \end{aligned} \right), \quad (8)$$

which is still an upper bound of the true risk. In order to simplify the computation, we choose the square loss and remove the constant term, and then the following Eq.(9) can be derived, which is

$$\{f^*, Q^*\} = \arg \min_{Q, f} \left(\begin{aligned} &\lambda J(f) + \sum_{i=1}^{n_l} (y_l^i - f(\mathbf{W}^\top \mathbf{x}_l^i))^2 \\ &+ \sum_{\mathbf{x} \in Q} (f(\mathbf{x})^2 + 2|f(\mathbf{x})|) \end{aligned} \right). \quad (9)$$

Based on the above analysis, we construct a multi-class framework based on Eq. (9) as

$$\arg \min_{Q, f, \mathbf{W}} \sum_{c=1}^C \left(J(f_c) + \sum_{i=1}^{n_l} (y_l^{ic} - f_c(\mathbf{W}^\top \mathbf{x}_l^i))^2 + \sum_{\mathbf{x}_i \in Q} (f_c(\mathbf{W}^\top \mathbf{x}_i)^2 + 2|f_c(\mathbf{W}^\top \mathbf{x}_i)|) \right), \quad (10)$$

where $\sum_{\mathbf{x}_i \in Q} (f_c(\mathbf{W}^\top \mathbf{x}_i)^2 + 2|f_c(\mathbf{W}^\top \mathbf{x}_i)|)$ is the uncertainty evaluation of all candidate examples, which can be viewed as a function of distances between the examples and their classification boundaries [32] [43].

D. The Integrated Dataset Alignment and Active Learning (DAAL) Framework

In order to balance the penalties on the classification error of macro- and micro-expression examples, two nonnegative parameters C_s and C_t are introduced to the DAAL framework based on Eq. (3) and Eq. (10) as

$$\arg \min_{Q, f, \mathbf{W}} \sum_{c=1}^C \left\{ J(f_c) + \frac{C_s}{2} \left(\sum_{i=1}^{n_s} (y_s^{ic} - f_c(\mathbf{x}_s^i))^2 \right) + \frac{C_t}{2} \left(\sum_{i=1}^{n_l} (y_l^{ic} - f_c(\mathbf{W}^\top \mathbf{x}_l^i))^2 + \sum_{\mathbf{x}_i \in Q} (f_c(\mathbf{W}^\top \mathbf{x}_i)^2 + 2|f_c(\mathbf{W}^\top \mathbf{x}_i)|) \right) \right\} + R(\mathbf{W}). \quad (11)$$

With the assistance of active learning, this model can select micro-expression examples from multiple categories with the richest information. Also, the effective supervised information from the macro-expression dataset can be used to make up for the insufficient labeled micro-expression data in the initial stage of active learning.

For the classifier $f_c = \mathbf{w}_c^\top x + b_c$ with \mathbf{w}_c being the parameter and b_c being the offset, then the optimization problem is formulated as

$$\arg \min_{\{\mathbf{w}_c, b_c\}, \mathbf{W}, Q} \sum_{c=1}^C \left\{ \frac{1}{2} \left\| \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix} \right\|_2^2 + \frac{C_s}{2} \sum_{i=1}^{n_s} \left(y_s^{ic} - \begin{pmatrix} \mathbf{x}_s^i \\ 1 \end{pmatrix}^\top \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix} \right)^2 + \frac{C_t}{2} \left[\sum_{i=1}^{n_l} \left(y_l^{ic} - \begin{pmatrix} \mathbf{x}_l^i \\ 1 \end{pmatrix}^\top \mathbf{W} \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix} \right)^2 + \sum_{\mathbf{x}_i \in Q} \left(\left(\begin{pmatrix} \mathbf{x}_i^i \\ 1 \end{pmatrix}^\top \mathbf{W} \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix} \right)^2 + 2 \left| \begin{pmatrix} \mathbf{x}_i^i \\ 1 \end{pmatrix}^\top \mathbf{W} \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix} \right| \right] \right\} + R(\mathbf{W}). \quad (12)$$

To simplify Eq. (12), we denote $\mathbf{w}_c \triangleq \begin{pmatrix} \mathbf{w}_c \\ b_c \end{pmatrix}$, $\mathbf{x}_l^i \triangleq \begin{pmatrix} \mathbf{x}_l^i \\ 1 \end{pmatrix}$, $\mathbf{x}_s^i \triangleq \begin{pmatrix} \mathbf{x}_s^i \\ 1 \end{pmatrix}$ and use the Ridge regularization $R(\mathbf{W}) = \frac{1}{2} \|\mathbf{W}\|_F^2$, so the final objective function becomes

$$\arg \min_{\{\mathbf{w}_c\}, \mathbf{W}, \sum q_i=1} \sum_{c=1}^C \left\{ \frac{1}{2} \|\mathbf{w}_c\|_2^2 + \frac{C_s}{2} \|\mathbf{Y}_s^c - \mathbf{X}_s^\top \mathbf{w}_c\|_2^2 + \frac{C_t}{2} \left[\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W} \mathbf{w}_c\|_2^2 + \left(\sum_{i=1}^{n_u} q_i \left(\left((\mathbf{x}_u^i)^\top \mathbf{W} \mathbf{w}_c \right)^2 + 2 \left| (\mathbf{x}_u^i)^\top \mathbf{W} \mathbf{w}_c \right| \right) \right) \right] \right\} + \frac{1}{2} \|\mathbf{W}\|_F^2, \quad (13)$$

where $\mathbf{X}_s = (\mathbf{x}_s^1, \mathbf{x}_s^2, \dots, \mathbf{x}_s^{n_s})$ and $\mathbf{X}_l = (\mathbf{x}_l^1, \mathbf{x}_l^2, \dots, \mathbf{x}_l^{n_l})$ are the macro- and micro-expression datasets, respectively; $\mathbf{Y}_s^c = (\mathbf{y}_s^{1c}, \mathbf{y}_s^{2c}, \dots, \mathbf{y}_s^{n_s c})^\top$, $\mathbf{Y}_l^c = (\mathbf{y}_l^{1c}, \mathbf{y}_l^{2c}, \dots, \mathbf{y}_l^{n_l c})^\top$; and $q_i \in \{0, 1\}$, $i = 1, 2, \dots, n_u$ are used to indicate which example is queried in \mathcal{L}_u .

E. Optimization

Eq. (13) is a non-convex optimization problem and can be solved by an alternative optimization procedure. Each \mathbf{w}_c , \mathbf{W} and q_i can be solved by fixing the rest parameters as constants.

1) Computing \mathbf{w}_c , for fixed \mathbf{W} and q_i :

We first introduce an auxiliary vector $\mathbf{z} = \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c$, where \mathbf{X}_q is assumed as a \mathcal{L}_u subset corresponding to the data matrix determined by q_i . The augmented Lagrangian function of Eq. (13) is then given by

$$L_\rho(\mathbf{w}_c, \mathbf{z}, \boldsymbol{\lambda}) = \frac{1}{2} \|\mathbf{w}_c\|_2^2 + \frac{C_s}{2} \|\mathbf{Y}_s^c - \mathbf{X}_s^\top \mathbf{w}_c\|_2^2 + \frac{C_t}{2} \left(\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W} \mathbf{w}_c\|_2^2 + \|\mathbf{z}\|_2^2 + 2\|\mathbf{z}\|_1 \right) + \boldsymbol{\lambda}^\top (\mathbf{z} - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c) + \frac{\rho}{2} \|\mathbf{z} - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c\|_2^2, \quad (14)$$

where $\boldsymbol{\lambda}$ is a Lagrangian multiplier and $\rho > 0$ is a constraint penalty parameter.

The variables \mathbf{w}_c , \mathbf{z} , $\boldsymbol{\lambda}$ in Eq. (14) can be solved separately with the basic Gauss-Seidel iteration method. At the iteration $k+1$ ($k \geq 0$), we have

$$\begin{cases} \mathbf{w}_c^{k+1} = \arg \min_{\mathbf{w}_c} L_\rho(\mathbf{w}_c, \mathbf{z}^k, \boldsymbol{\lambda}^k) \\ \mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} L_\rho(\mathbf{w}_c^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k) \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho (\mathbf{z}^{k+1} - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c^{k+1}) \end{cases}. \quad (15)$$

Then, the details related to the solutions of sub-problems Eq. (15) are presented below. The sub-problem on \mathbf{w}_c^{k+1} becomes a typical convex problem and can be effectively solved by simple mathematical manipulations, namely

$$\mathbf{w}_c^{k+1} = \mathbf{A}^{-1} \mathbf{r}, \quad (16)$$

where $\mathbf{A} = \mathbf{I} + C_s \mathbf{X}_s \mathbf{X}_s^\top + C_t \mathbf{W}^\top \mathbf{X}_l \mathbf{X}_l^\top \mathbf{W} + \rho \mathbf{W}^\top \mathbf{X}_q \mathbf{X}_q^\top \mathbf{W}$ with \mathbf{I} being an identity matrix, and $\mathbf{r} = C_s \mathbf{X}_s \mathbf{Y}_s^c + C_t \mathbf{W}^\top \mathbf{X}_l \mathbf{Y}_s^c + \mathbf{W}^\top \mathbf{X}_q \boldsymbol{\lambda}^k + \rho \mathbf{W}^\top \mathbf{X}_q \mathbf{z}^k$.

As for the sub-problem regarding \mathbf{z}^{k+1} in Eq. (15), we have

$$\begin{aligned} \mathbf{z}^{k+1} &= \arg \min_{\mathbf{z}} \frac{C_t}{2} \left(\|\mathbf{z}\|_2^2 + 2\|\mathbf{z}\|_1 \right) + (\boldsymbol{\lambda}^k)^\top \mathbf{z} \\ &\quad + \frac{\rho}{2} \|\mathbf{z} - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c^{k+1}\|_2^2 \\ &= \arg \min_{\mathbf{z}} \frac{1}{2} \|\mathbf{z} - \mathbf{v}\|_2^2 + \eta \|\mathbf{z}\|_1, \end{aligned} \quad (17)$$

where $\mathbf{v} = \frac{\frac{\rho}{C_t} \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c^{k+1} - \frac{\boldsymbol{\lambda}^k}{C_t}}{\frac{\rho}{C_t} + 1}$ and $\eta = \frac{1}{C_t + 1}$. Eq. (17) is a typical sparse representation problem and can be easily solved via [44].

2) Computing \mathbf{W} , for fixed \mathbf{w}_c and q_i :

Similar to \mathbf{w}_c , by introducing $\mathbf{z}_c = \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c$, the augmented Lagrangian function with respect to \mathbf{W} is presented as

$$\begin{aligned} L_\rho(\mathbf{W}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_C) \\ = \sum_{c=1}^C \left(\frac{C_t}{2} \left(\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W} \mathbf{w}_c\|_2^2 + \|\mathbf{z}_c\|_2^2 + 2\|\mathbf{z}_c\|_1 \right) \right. \\ \left. + \boldsymbol{\lambda}_c^\top (\mathbf{z}_c - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c) + \frac{\rho}{2} \|\mathbf{z}_c - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c\|_2^2 \right) \\ + \frac{1}{2} \|\mathbf{W}\|_F^2, \end{aligned} \quad (18)$$

where \mathbf{z}_c , $\boldsymbol{\lambda}_c$ and \mathbf{w}_c for $c = 1, 2, \dots, C$ correspond to the parameters related to class c . The variables \mathbf{W} , $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C$, $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_C$ in Eq. (18) can also be solved separately. At iteration $k+1$ ($k \geq 0$), we have

$$\begin{cases} \mathbf{W}^{k+1} = \arg \min_{\mathbf{W}} L_\rho(\mathbf{W}, \mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_C^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \dots, \boldsymbol{\lambda}_C^k) \\ \mathbf{z}_c^{k+1} = \arg \min_{\mathbf{z}_c} \frac{C_t}{2} \left(\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W}^{k+1} \mathbf{w}_c\|_2^2 + \|\mathbf{z}_c\|_2^2 \right. \\ \quad \left. + 2\|\mathbf{z}_c\|_1 \right) + (\boldsymbol{\lambda}_c^k)^\top (\mathbf{z}_c - \mathbf{X}_q^\top \mathbf{W}^{k+1} \mathbf{w}_c) \\ \quad + \frac{\rho}{2} \|\mathbf{z}_c - \mathbf{X}_q^\top \mathbf{W}^{k+1} \mathbf{w}_c\|_2^2 \\ \boldsymbol{\lambda}_c^{k+1} = \boldsymbol{\lambda}_c^k + \rho (\mathbf{z}_c^{k+1} - \mathbf{X}_q^\top \mathbf{W}^{k+1} \mathbf{w}_c). \end{cases} \quad (19)$$

The sub-problem about \mathbf{W}^{k+1} becomes a typical convex optimization problem as

$$\begin{aligned} \mathbf{W}^{k+1} \\ = \arg \min_{\mathbf{W}} L_\rho(\mathbf{W}, \mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_C^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \dots, \boldsymbol{\lambda}_C^k) \\ = \arg \min_{\mathbf{W}} \frac{1}{2} \|\mathbf{W}\|_F^2 + \sum_{c=1}^C \left(\frac{C_t}{2} \left(\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W} \mathbf{w}_c\|_2^2 \right) - \right. \\ \left. \boldsymbol{\lambda}_c^\top \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c + \frac{\rho}{2} \|\mathbf{z}_c - \mathbf{X}_q^\top \mathbf{W} \mathbf{w}_c\|_2^2 \right), \end{aligned} \quad (20)$$

which can be solved by gradient descent algorithm.

3) Computing q_i , for fixed \mathbf{w}_c and \mathbf{W} :

When \mathbf{w}_c and \mathbf{W} are fixed, Eq. (13) can be converted into solving the sub-problem of q_i . The formula for solving q_i can be written as

$$\arg \min_{\sum q_i=1} \sum_{i=1}^{n_u} q_i \left(\sum_{c=1}^C \left((\mathbf{x}_u^i)^\top \mathbf{W} \mathbf{w}_c \right)^2 + 2 \left| (\mathbf{x}_u^i)^\top \mathbf{W} \mathbf{w}_c \right| \right). \quad (21)$$

Algorithm 1 Solution for Eq. (13)

Input:

- The training macro-expression dataset: $\mathcal{L}_s = \{(\mathbf{x}_s^i, \mathbf{y}_s^i)\}_{i=1}^{n_s}$.
- The training micro-expression dataset: $\mathcal{L}_l = \{(\mathbf{x}_l^i, \mathbf{y}_l^i)\}_{i=1}^{n_l}$.
- The candidate micro-expression dataset: $\mathcal{L}_u = \{\mathbf{x}_u^i\}_{i=1}^{n_u}$.
- Two nonnegative trade-off parameters C_s and C_t .

Output:

\mathbf{w}_c ($c = 1, \dots, C$), \mathbf{W} , \mathbf{q}_i

Algorithm steps:

- 1: Initialize: $q_i^{(0)} = 0$, $\varepsilon = 10^{-3}$, maxIter = 50, $k = 0$.
- 2: Calculate $\mathbf{w}_c^{(0)}$ for $c = 1, \dots, C$ by solving the problem: $\min_{\mathbf{w}_c} \frac{1}{2} \|\mathbf{w}_c\|_2^2 + \frac{C_s}{2} \|\mathbf{Y}_s^c - \mathbf{X}_s^\top \mathbf{w}_c\|_2^2$
- 3: Calculate $\mathbf{W}^{(0)}$ before selecting examples from the candidate set by solving the problem: $\min_{\mathbf{W}} \frac{1}{2} \|\mathbf{W}\|_F^2 + \sum_{c=1}^C \left(\frac{C_t}{2} \left(\|\mathbf{Y}_l^c - \mathbf{X}_l^\top \mathbf{W} \mathbf{w}_c\|_2^2 \right) \right)$
- 4: **while** not converged or $k \leq \text{maxIter}$ **do**
- 5: Update $q_i^{(k+1)}$ according to Eq. (21) with $\mathbf{W}^{(k)}$ and $\mathbf{w}_c^{(k)}$ fixed;
- 6: **for** $c = 1 : C$ **do**
- 7: Initialize: $\mathbf{w}_c^{(k+1)} = 0$, $\mathbf{z}^{(k+1)} = 0$, $\boldsymbol{\lambda}^{(k+1)} = 0$,
- 8: **while** not converged **do**
- 9: Update $\mathbf{w}_c^{(k+1)}$ according to Eq. (16) with $\mathbf{W}^{(k)}$ and $q_i^{(k+1)}$ fixed;
- 10: **end while**
- 11: **end for**
- 12: Initialize: $\mathbf{z}_c^{(k+1)} = 0$ for $c = 1, \dots, C$, $\boldsymbol{\lambda}^{(k+1)} = 0$, $\rho = 1$,
- 13: **while** not converged **do**
- 14: Update $\mathbf{W}^{(k+1)}$ according to Eq. (20) with $\mathbf{w}_c^{(k+1)}$ and $q_i^{(k+1)}$ fixed;
- 15: **end while**
- 16: $k = k + 1$
- 17: **end while**

In Eq. (21), assuming that q_i is relaxed to the continuous space $[0, 1]$, then linear programming can be used in its solution. While the values of q_i for $i = 1, \dots, n_u$ should be either 1 or 0, so the obtained q_i for $i = 1, \dots, n_u$ are rearranged, and the maximum value is set to 1 and the others are zeros.

Based on the above analysis, the whole procedure of our solutions is presented in **Algorithm 1**.

IV. EXPERIMENTS

This section presents the experimental results on four micro-expression databases, evaluates the impact of the penalty coefficients of DAAL on the performance, and also conducts some ablation studies on the proposed DAAL method.

A. Databases

CK+ [45], RAF-DB [46], AffectNet [47], CASME [48], CASME II [49], SAMM [50] and SMIC [51] databases are used for algorithm evaluation. Fig. 2 shows the snapshots of seven macro-/ micro-expression databases.

CK+ includes 593 image sequences from 123 subjects. In addition, 327 out of the 593 sequences have emotion labels. CK+ database is commonly used for facial expression recognition and contains 7 categories of human emotions, such as happiness, surprise, fear, sadness, disgust, anger and contempt.



Fig. 2. Snapshots of seven macro-/ micro-expression databases. The rectangle shows where the face muscles change. The facial features of all the databases are properly aligned.

RAF-DB database contains single and mixed expressions of 29,672 images from the real world. We use six categories (Surprise, Fear, Disgust, Happiness, Sadness, and Anger) in the single expression subset. In order to meet the requirements of data alignment between micro-expression dataset (collected in laboratory environment) and macro-expression dataset, all the macro-expression samples are carefully screened. Those images with non-frontal head, occlusion or pollution and even too bright or too dark lighting are removed from the dataset, and namely, samples with better shooting conditions are left. Finally, 6248 macro-expression samples are used for experiments.

AffectNet database contains 420,299 images captured in the real world. It includes 11 categories of annotated emotions, such as Neutral, Happiness, Sadness, Surprise, Fear, Disgust, Anger, Contempt, None, Uncertain, No-face, and 7 categories (Happiness, Sadness, Surprise, Fear, Disgust, Anger, and Contempt) are selected. We perform the same pre-processing as the operations on RAF-DB, and 69,027 samples leaves for experiments.

CASME database consists of 195 micro-expressions examples from 20 subjects. The examples were taken at 640×480 under 60fps and contain seven emotions: happiness, surprise, fear, sadness, disgust, repression and tension. CASME II is an extension of CASME, which consists of 247 micro-expression segments captured from 26 subjects, using high speed camera with 200fps to capture the time-varying process of micro-expressions and relieved the influence of illumination variation. Compared with CASME, examples in CASME II are larger in image size, therefore can provide more detailed information of facial muscle movement. To simplify the emotion classification based on CASME, CASME II divides the micro-expressions into happiness, surprise, fear, sadness, disgust, repression, and others.

SAMP database contains 159 micro-expressions from 29 subjects at 200 fps. Consistent with CASME, SAMP also contains seven emotions: anger, contempt, happiness, surprise, fear, sadness and disgust. It uses similar procedures like CASME II but has a higher image resolution and employs an array of LEDs to avoid flickering. Because of the creators with professional rating skills, these recorded expressions are obtained from stricter lab situations and labeled more accurately.

SMIC database contains 164 spontaneous micro-expressions from 16 subjects. It is recorded by 100 fps high-speed cameras. The participants are requested to watch high emotional video clips with punishment and threat in an interrogation room to induce the participants' micro-expressions, so as to establish the database. The emotion categories in SMIC are positive, negative and surprise.

B. Experimental Settings

In order to verify the effectiveness of DAAL, all experiments are carried out on two databases, including a macro-expression database and a micro-expression database. We choose CK+, RAF-DB and AffectNet as the macro-expression databases and the other four as the micro-expression database for experiments, forming twelve dataset combinations which are respectively denoted as CK+&CASME, CK+&CASME II, CK+&SAMP, CK+&SMIC, RAF-DB&CASME, RAF-DB&CASME II, RAF-DB&SAMP, RAF-DB&SMIC, AffectNet&CASME, AffectNet&CASME II, AffectNet&SAMP and AffectNet&SMIC.

To achieve dataset alignment and use a unified evaluation metric, for CK+&CASME, CK+&CASME II, CK+&SAMP, RAF-DB&CASME, RAF-DB&CASME II, RAF-DB&SAMP, AffectNet&CASME, AffectNet&CASME II, AffectNet&SAMP, we choose three categories of examples: happiness, surprise and disgust, which are common included in these four databases and have larger example numbers. Since the proposed method is a semi-supervised learning method, and the experimental dataset needs to be divided into the training and test datasets. Moreover, the training set contains labeled and unlabeled samples. The samples from the categories of fear and sadness are too few to train in CK+, CASME, CASME II, SAMP, therefore, we do not choose the samples of these two categories. For CK+&SMIC, RAF-DB&SMIC, AffectNet&SMIC, we merge seven categories in CK+, RAF-DB and AffectNet into three classes to keep the consistency with SMIC. The happy micro-expressions are classified into 'Positive' class as they indicate good emotions of subjects. In contrast, the disgust, sadness, fear, contempt and anger micro-expressions are classified into 'Negative' class as they are usually considered as bad emotions. For above twelve dataset combinations, LBP can explicitly reflect a statistical distribution of local binary patterns, in addition, it is suitable to describe the feature of macro-expression which may be an image. While a micro-expression is an image sequence, and MDMO has been shown its best feature representation ability in the field of micro-expression recognition [52]. Therefore, LBP and MDMO are respectively used to extract the features of macro-expressions and micro-expressions. Although LBP and MDMO have different characteristics, we assume that the features of macro-expression and micro-expression can be translated by a linear translator \mathbf{W} , because this model is relatively simple. The proposed model takes advantage of a linear translator \mathbf{W} to transform the two features of LBP and MDMO into the same feature space.

The implementation details of DAAL is as follows: for each database combination, all examples of the macro-expression database are used for training. The micro-expression database is firstly divided into two parts, each kind of example is

randomly divided, one-third as the test set, two-thirds as the 'training set' (Note: the subjects in the test set and the 'training set' are not crossed). Then, the 'training set' is divided into two parts again. From the 'training set', K examples are selected as the seed set for each kind of emotion and the rest constitutes the candidate pool. The value of K is selected to be 20% of the number of micro-expression samples, therefore, the values of K are selected as 1, 2, 1, 3 on CASME, CASMEII, SAMM and SMIC respectively combined with any of the three macro-expression databases (CK+, RAF-DB and AffectNet). Finally, the micro-expression database is divided into three parts: test set, seed set and candidate pool. All macro-expression examples and seed set together constitute the final training set. Linear classifier and cross validation are used to determine the penalty coefficients C_s and C_t , which are selected from the grid $\{10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$ and all the experimental results are reported under optimal parameters.

DAAL is an iterative process, in each iteration the most valuable query example from the candidate pool is selected and added into the seed set after manual tagging. The classifier is subsequently trained with the updated seed set. The accuracy (noted as Acc. for short), unweighted average recall (UAR) and F1-score [26] on the test set are used as the evaluation criterion of the proposed method. Accuracy is the fraction of correct classifications. UAR can measure the direct unweighted average multi-class recall performance. F1-score is the harmonic mean of the precision and recall. Besides, there is another evaluation metric called weighted average recall (WAR), measuring the weighted average multiclass recall performance, which is equal to accuracy. Since the test set, seed set, and candidate pool are randomly obtained, we repeat each experiment 20 times independently and the averaged accuracy, the averaged unweighted average recall (UAR) and the averaged F1-score are calculated as the final result.

C. Impact of C_s and C_t

We evaluate the impact of the penalty coefficients C_s and C_t on the performance of DAAL using CK+&CASME II. Note that C_s and C_t are parameters to balance the dataset alignment and active learning in the whole DAAL framework, so they are important to improve the micro-expression recognition performance. To study their impact, we first fix $C_t=10^3$ and select C_s from $\{10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$. The number of query examples is set to $\{1, 2, 3, 4, 5\}$, respectively. Subsequently, we fix $C_s=10^2$ and select C_t from $\{10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$. The number of query examples is also set as $\{1, 2, 3, 4, 5\}$, respectively. We report the accuracies w.r.t. the change of the number of query examples for the proposed DAAL in Fig. 3, from which we can see that $C_s=10^2$ is a turning point with fixed numbers of query examples and the accuracy reaches the peak on CK+&CASME II. When C_s is selected from $\{10^{-1}, 1, 10, 10^2\}$, there is a positive correlation between the accuracy and C_s , while an opposite tendency can be witnessed when C_s is greater than 10^2 . It is consistent with the structural risk minimization theory. C_s is a penalty parameter on classification error of macro-expression examples. If C_s is small, the classifier cannot fit macro-expression examples well and

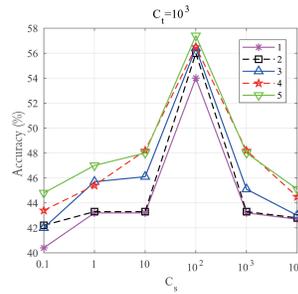


Fig. 3. The impact of C_s on the performance of the proposed DAAL on CK+&CASME II.

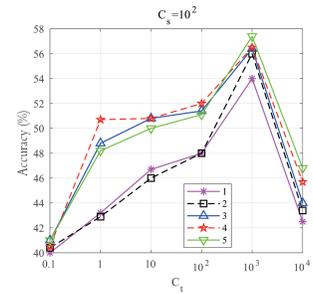


Fig. 4. The impact of C_t on the performance of the proposed DAAL on CK+&CASME II.

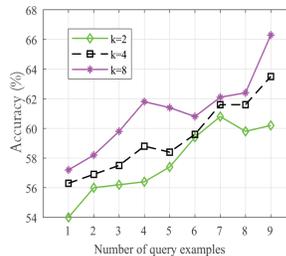


Fig. 5. The impact of K on the performance of the proposed DAAL on CK+&CASME II.

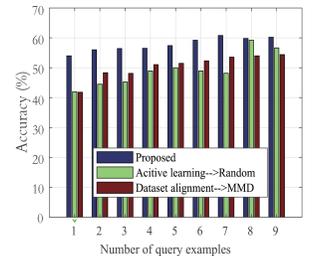


Fig. 6. Results of ablation study on the CK+&CASME II

less useful information is extracted from the macro-expression dataset, which will decrease the effect of dataset alignment. If C_s is too large, the classifier may over-fit the macro-expression examples, which causes poor dataset alignment performance on the heterogeneous data of macro-expression and micro-expression. On the other hand, the competition exists between dataset alignment and active learning in our DAAL framework. If the small number of labeled micro-expression examples on which active learning relies are underestimated, the classifier may have opposite evaluation of the example information in candidate pool, causing bad effect on active learning. Similarly, as shown in Fig. 4, $C_t=10^3$ is a turning point with the fixed number of query examples and accuracy reaches the peak on CK+&CASME II datasets. There is a positive correlation between the accuracy and C_t when C_t is smaller than 10^3 . Once C_t is larger than 10^3 , the accuracy decreases with the increase of C_t . C_t is a penalty parameter on the classification error of the labeled micro-expression examples. With the increasing of C_t , the objective function may pay more attention to labeled micro-expression examples, which brings more instructive information to the model and improves the effect of active learning. However, when C_t increases continuously, dataset alignment may compete against active learning, which will decrease the effect of active learning.

D. Impact of K

In order to verify the influence of the parameter K , we conduct experiments on CK+&CASME II, and the experimental results are shown in the Fig. 5. From the Fig. 5, we can find that when K is equal to 2, 4, 8, the highest accuracies on CK+&CASME II can reach 60.8%, 63.5%, and 66.3%, respectively. As the parameter K increases, the accuracy of

the proposed method on CK+&CASME II also increases. This result is reasonable since the larger K leads to more supervised information.

E. Ablation Study

In order to verify the performance of the proposed method, we conduct ablation experiments. Replacing the dataset alignment terms of the proposed method, we minimize the Maximum Mean Discrepancy (MMD) [26], which is a distribution distance measure, to eliminate the domain divergence between macro-expressions and micro-expressions. In accordance with Section IV-C, we choose CK+&CASME II for the analysis of ablation study. We repeat each above-mentioned classification experiments under each number of query examples ranging from 1 to 9 for 20 times. The average accuracies of CK+&CASME II are reported in Fig. 6. It can be seen that the proposed method after replacing dataset alignment terms with MMD achieves the accuracy of 59.2%, which is 1.6% lower than the accuracy of the proposed method. It indicates that the proposed dataset alignment between the micro-expression and the macro-expression datasets plays an important role in improving the recognition performance.

Similarly, in order to verify the performance of active learning in the proposed method, we replace active learning terms with random selections, and also repeat the above-mentioned experiments under each number of query examples ranging from 1 to 9 for 20 times. The experimental results are shown in Fig. 6, and it can be observed that the average accuracy on CK+&CASME II is 54.4%, which is 6.4% lower than that of the original proposed model. Therefore, it proves that the active learning terms make a valuable contribution towards raising the micro-expression recognition performance.

F. Comparison of DAAL with Existing Active Learning Methods

In order to verify the performance of the proposed method, we compared it with four active learning methods including random selection method and other three state-of-the-art active learning methods such as USDM [53], Marginal Probability (MP) [54], QUIRE [43]. In addition, we also compared it with a transfer learning method which is a combination of Auxiliary Set Selection Model and Transductive Transfer Regression Model and denoted as (ASSM+TTRM) [26]. USDM follows information quantity criterion. It estimates the probability of examples in candidate pool by building the on-graph wandering model of seed set and candidate pool so as to calculate the information entropy of examples. This method is mainly based on effectively setting the graph-related parameters. MP is based on matching marginal probability distribution. It can reduce marginal probability differences between seed set and candidate pool, and select examples accord with representative criterion. QUIRE is based on the minimum maximization framework which can select both informative and representative examples. ASSM selects an optimal set of micro-expression examples from the target domain, and TTRM bridges the feature distribution gap between the micro- and macro-expression and target domains.

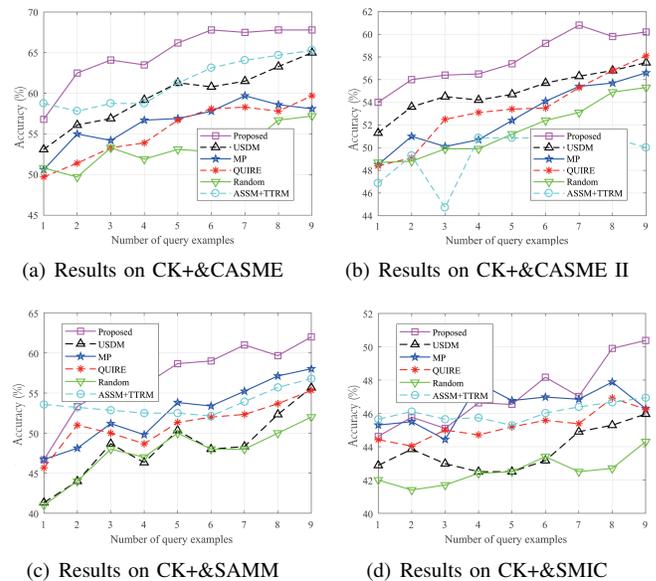


Fig. 7. Comparison results (%) with other active learning methods.

We carry out the comparative experiments on CK+&CASME, CK+&CASME II, CK+&SAMM and CK+&SMIC respectively. Fig. 7 shows the recognition results (%) w.r.t. different algorithms under different numbers of query examples. As can be seen from Fig. 7, the proposed method obtains the highest accuracy of 67.8% on CK+&CASME when the number of query examples is 6, 8, 9, respectively, and the highest accuracy of 60.8% on CK+&CASME II when the number of query examples is 7. On CK+&SAMM and CK+&SMIC, the highest accuracies are 62% and 50.38% respectively when 9 query examples are selected. From Fig. 7 we can see that when the same example size is selected from the candidate pool, on CK+&CASME, CK+&CASME and CK+&SAMM, the proposed method can always obtain relatively high accuracy, which benefits from both the dataset alignment and active learning, especially the dataset alignment. Dataset alignment allows our method to have access to macro-expression data, the proposed method to transfer more information from the macro-expression dataset. USDM, MP and QUIRE are all methods based on active learning. Among them, USDM and MP are active learning methods with category maximization constraints, which can select multiple dissimilar examples at the same time. USDM and MP can avoid duplicate data while increasing the number of examples, but they (including QUIRE) all lack of dataset alignment. Compared with DAAL, they cannot use information from the macro-expression database, so their accuracies are lower than DAAL. On the contrary, ASSM+TTRM, although it includes the dataset alignment in the example selection phase, but its example selection strategy only considers the feature distribution gap between micro- and macro expression and target domains, and does not evaluate the effectiveness of information carried by the example itself, so its accuracy is also lower than that of DAAL. Compared with other three database combinations, the experimental results of all methods on CK+&SMIC shown in Fig. 7(d) are not ideal. On CK+&SMIC, we solve

TABLE II
COMPARISON RESULTS (%) WITH OTHER ADVANCED MICRO-EXPRESSION RECOGNITION METHODS('-' INDICATES NONE.)

Methods	Auxiliary dataset	CASME			CASME II			SAMM			SMIC		
		Acc.	UAR	F1-score	Acc.	UAR	F1-score	Acc.	UAR	F1-score	Acc.	UAR	F1-score
LBP-TOP [13]	-	35.5	34.8	29.5	31.6	36.2	28.4	38.2	41.9	33.1	39.5	39.1	36.8
LBP-SIP [15]	-	35.4	33.4	24.4	29.1	40.7	28.4	38.6	39.2	32.5	39.6	40.2	37.7
FDM [18]	-	34.6	42.6	33.3	35.5	39.6	34.4	32.3	38.5	31.9	36.4	40.5	35.1
MDMO [19]	-	41.7	39.9	28.2	42.1	38.8	33.5	42.1	48.3	39.5	44.8	42.0	37.8
Sparse MDMO [55]	-	38.9	45.3	37.1	48.1	46.6	42.6	44.2	32.3	24.6	46.1	46.8	42.7
ASSM+TTRM [26]	CK+	65.3	40.5	52.0	50.9	46.2	47.9	56.8	46.7	39.8	46.9	47.2	42.1
	RAF-DB	52.9	44.1	33.6	44.6	46.2	41.1	47.9	44.1	42.1	39.1	38.1	37.1
	AffectNet	54.1	52.7	41.7	46.7	44.1	42.1	50.7	44.6	39.9	43.0	41.5	40.0
Deep Transfer [11]	CK+	50.0	53.5	49.6	52.5	45.7	43.0	53.3	51.3	48.0	46.3	46.2	45.4
RCN-X [56]	-	63.9	54.0	49.6	58.8	53.0	47.2	51.9	46.4	41.4	45.2	41.5	33.7
TS-AUCNN [8]	CK+	52.8	39.9	37.7	44.0	31.7	29.6	48.0	32.5	21.6	-	-	-
MEGC [9]	CK+	58.3	54.1	49.2	50.3	44.0	42.2	50.0	51.9	48.5	46.8	42.5	41.9
EIDNet+MTMNet [10]	CK+	59.6	41.3	38.2	56.7	43.3	42.3	50.6	48.5	43.7	47.1	41.7	39.5
G-TCN [57]	-	51.7	46.8	45.3	50.8	44.8	45.0	44.9	41.6	41.5	43.6	40.1	40.6
AU-GACN [58]	-	34.3	38.1	33.2	38.7	39.2	37.9	35.2	40.1	32.5	-	-	-
The proposed DAAL	CK+	67.8	59.3	56.3	60.8	63.2	59.2	62.0	56.1	57.0	50.4	51.3	50.4
	RAF-DB	60.3	56.7	51.7	52.4	44.9	45.0	56.3	48.5	48.9	45.2	45.2	44.8
	AffectNet	62.5	56.7	52.1	53.6	48.2	48.0	60.0	51.6	51.6	47.1	46.3	46.3

a rough-classification problem. We forcibly map the seven categories of examples in CK+ to a reduced label subspace and only use a very small number of examples for training, which undoubtedly is a great challenge. But our method still achieves the highest accuracy of 50.4%, which strongly demonstrates the effectiveness of our method again.

G. Comparison with Other Advanced Micro-expression Recognition Methods

To further demonstrate the superiority of DAAL, we conducted comprehensive experiments on CK+&CASME, CK+&CASME II, CK+&SAMM, CK+&SMIC, RAF-DB&CASME, RAF-DB&CASME II, RAF-DB&SAMM, RAF-DB&SMIC, AffectNet&CASME, AffectNet&CASME II, AffectNet&SAMM, AffectNet&SMIC these twelve dataset combinations. We compare it with some typical micro-expression recognition methods like LBP-TOP [13], LBP-SIP [15], FDM [18], MDMO [19], Sparse MDMO [55], and ASSM+TTRM [26], and advanced methods based on deep convolutional neural network like Deep Transfer [11], RCN-X [56], TS-AUCNN [8], MEGC [9], EIFNet+MTMNet [10], G-TCN [57] and AU-GACN [58] in Table II. Worthy of note is that, TS-AUCNN needs the support of action unit labels of micro-expression and macro-expression data; unfortunately, SMIC does not provide them. Therefore, the performance of TS-AUCNN cannot be evaluated on CK+&SMIC. Besides, AffectNet and RAF-DB also does not offer us AU labels, so the experiments conducted by the AffectNet and RAF-DB as auxiliary datasets cannot be done. MEGC needs to extract optical flow feature from the macro-expression data, and Deep Transfer uses 3DCNN to extract macro-expression features. Namely, only video frame sequences for macro-expression data can be used for MEGC and Deep Transfer. EIDNet+MTMNet generates more samples with the help of identity labels of macro-expression data. However, AffectNet and RAF-DB does not contain identity information. Therefore, MEGC, Deep Transfer and EIDNet+MTMNet are not suitable to conduct experiments under the dataset combination with

AffectNet or RAF-DB as the auxiliary set. In our proposed method, active learning can continue until all samples in the candidate set are sampled, and we can also set the query period to be a certain positive integer which is less than the number of samples in the candidate. On each dataset combination, we query 1~9 samples, and Table II reports the highest accuracy among these 9 test results. Deep Transfer [11] follows a route of transfer learning and alleviates the problem of small sample size of micro-expression database. It takes ResNet10 as the backbone and improves the accuracy of the model in the micro-expression database by pre-training on the macro-expression database. RCN-X is a recurrent convolution network integrating three modules including wide expansion, short connection and attention unit. We conduct experiments on the six derived RCN models and reported the best recognition results. For the convenience of expression, we denote the structures corresponding to the best results as RCN-X uniformly. For a fair comparison, the training data (including labeled data and unlabeled candidate set) is the same, and the test sets are set the same for all the methods. As Table II shows, we can find that the proposed DAAL shows obvious advantages when compared with other micro-expression recognition methods. When CK+ is used as an auxiliary dataset, the accuracy, UAR and F1-score of the proposed DAAL are at least 2.5%, 5.2% and 4.3% higher than the best result among other methods on the CASME, respectively. Moreover, the recognition accuracy of the proposed DAAL exceeds all the other machine learning methods by a large gap over 9.9%, and higher than deep methods by 2% on CASME II. Meanwhile, compared with both machine learning and deep learning methods, the proposed DAAL has a significant advantage in UAR and F1-score, which are 10.2% and 11.3% higher than the second best method, respectively. We also have obtained similar experimental results on SAMM and SMIC databases. The recognition accuracy of the proposed method on SAMM and SMIC are 62.0% and 50.4% respectively, which are 5.2% and 3.5% higher than the second best methods. In addition,

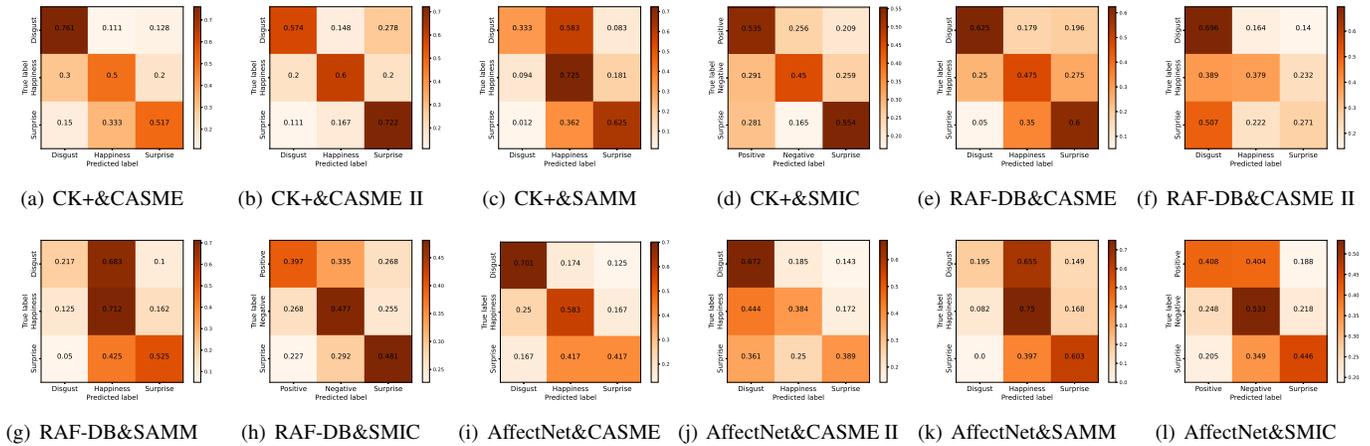


Fig. 8. Class-wise accuracy results in the form of confusion matrices for all dataset combinations.

the proposed DAAL has achieved better UAR and F1-score. We also conduct experiments on a larger auxiliary dataset such as RAF-DB, AffectNet, and the recognition accuracies of the proposed method on the four micro-expression datasets are 60.3%, 52.4%, 56.3%, 45.2% and 62.5%, 53.6%, 60.0%, 47.1%, respectively, which are lower than the accuracies of the auxiliary dataset of CK+. Similarly, when the auxiliary dataset is RAF-DB or AffectNet, the values of UAR and F1-Score are also lower than those when the auxiliary set is CK+. This is because both RAF-DB and AffectNet contain facial expression images of real world with variable lighting, poses and occlusions in the wild, accordingly, it is difficult for them to achieve full alignment with micro-expression images captured under the controllable laboratory environment in the strict sense. It can be seen from Table II that the proposed method has advantages compared with TS-AUCNN, MEGC, EIFNet+MTMNet, G-TCN and AU-GACN, especially when the auxiliary dataset is CK+.

The class-wise accuracy results in the form of confusion matrices for all dataset combinations are shown in Fig. 8. From Fig. 8, we can find that on the CASME database, when the auxiliary set is CK+, RAF-DB or AffectNet, the recognition performance of the three categories is good, especially the disgust category. On CASME II, when CK+ is the auxiliary set, the recognition performance of each category is the best, and the accuracies of the three categories are 57.4%, 60.0% and 72.2%. On the SAMM database and SMIC database, the recognition performance of each category is not as good as that of CASME and CASME II. This can be attributed to the small sample size and sample imbalance of the database. Moreover, sample imbalance is the reason for this phenomenon that a simple average for all three categories is sometimes lower than the Acc. value reported in Table II by up to about 7% accuracy. This also confirms once again that we should evaluate the performance with accuracy (equal to WAR), UAR and class-wise accuracy (confusion matrices). The experimental results on these four micro-expression databases show that DAAL which integrates dataset alignment and active learning has full advantages compared with other methods, and has good generalization ability. After adding the examples selected through active learning into the seed set, the recognition result is greatly improved, which also indicates that the proposed

DAAL is very effective and conducive to selecting the best examples to help the classifier training. In addition, it should be noted that the proposed DAAL can achieve excellent recognition results through tagging a small size of micro-expression examples, which can effectively solve the small sample or even extremely small sample problem of micro-expression recognition.

On the other hand, different feature extraction methods have great difference in their recognition results due to their different abilities of expressing micro-expression data. Different feature extraction methods can represent the characteristics of data from different aspects. Table II also shows that the feature extraction has a great influence on the recognition performance.

V. CONCLUSION

In this paper, we propose a novel dataset alignment and active learning (DAAL) framework for micro-expression recognition. An important advantage of the proposed method is learning a feature translator matrix from the micro-expression dataset to the macro-expression dataset. Thus, the common features of the two datasets can be obtained. With the assistance of active learning, this method can select micro-expression examples from multiple categories with the richest information. Extensive experimental results on CASME, CASME II, SAMM and SMIC show that the proposed DAAL can achieve superior performance to other existing approaches. In the future, we will further explore different feature extraction methods to boost the recognition performance. Moreover, the problem of unbalanced sample categories in micro-expression recognition will also be investigated to further improve the robustness of the algorithm. In addition, recent studies [59] have shown that spatial-temporal saliency are very successful in computer vision, so it is worth trying to deep dataset alignment and active learning problem with spatial-temporal saliency under extremely small size of training example.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Hong-Xia Xie and Prof. Wen-Huang Cheng from the National Chiao Tung University for helping us provide the results of the AU-GACN for a fair comparison with the proposed method.

REFERENCES

[1] B. Allaert, I. M. Bilasco, and C. Djeraba, "Micro and macro facial expression recognition using advanced local motion patterns," *IEEE Trans. Affective Comput.*, vol. 13, no. 1, pp. 147–158, 2022.

[2] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning from hierarchical spatiotemporal descriptors for micro-expression recognition," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3160–3172, 2018.

[3] N. Van Quang, J. Chun, and T. Tokuyama, "CapsuleNet for micro-expression recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2019, pp. 1–7.

[4] M. Verma, S. K. Vipparthi, G. Singh, and S. Murala, "LEARNet: Dynamic imaging network for micro expression recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 1618–1627, 2020.

[5] Z. Xia, X. Hong, X. Gao, X. Feng, and G. Zhao, "Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 626–640, 2020.

[6] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3D flow convolutional neural network," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1331–1339, 2019.

[7] Y. Li, X. Huang, and G. Zhao, "Can micro-expression be recognized based on single Apex frame?" in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 3094–3098.

[8] B. Sun, S. Cao, D. Li, J. He, and L. Yu, "Dynamic micro-expression recognition using knowledge distillation," *IEEE Trans. Affective Comput.*, vol. 13, no. 2, pp. 1037–1043, 2022.

[9] Y. Liu, H. Du, L. Zheng, and T. Gedeon, "A neural micro-expression recognizer," in *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*, 2019, pp. 14–18.

[10] B. Xia, W. Wang, S. Wang, and E. Chen, "Learning from macro-expression: a micro-expression recognition framework," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2936–2944.

[11] M. Peng, Z. Wu, Z. Zhang, and T. Chen, "From macro to micro expression recognition: Deep learning on small datasets using transfer learning," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2018, pp. 657–661.

[12] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1449–1456.

[13] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, 2007.

[14] J. A. Ruiz-Hernandez and M. Pietikäinen, "Encoding local binary patterns using the re-parametrization of the second order Gaussian Jet," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2013, pp. 1–6.

[15] Y. Wang, J. See, W. Phan, and Y. H. Oh, "LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition," in *ACCV*, 2015, pp. 525–537.

[16] X. Huang, S. Wang, G. Zhao, and M. Pietikäinen, "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 1–9.

[17] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns," *Neurocomputing*, vol. 175, pp. 564–578, 2016.

[18] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affective Comput.*, vol. 8, no. 2, pp. 254–267, 2017.

[19] 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 Trans. Affective Comput.*, vol. 7, no. 4, pp. 299–310, 2016.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley, 2001.

[21] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *ACM Multimedia*, 2016, pp. 382–386.

[22] M. Takalkar and M. Xu, "Image based facial micro-expression recognition using deep learning on small datasets," in *Proc. IEEE Int. Conf. Digital Image Computing: Techniques and Applications*, 2017, pp. 1–7.

[23] X. Jia, X. Ben, H. Yuan, K. Kpalma, and W. Meng, "Macro-to-micro transformation model for micro-expression recognition," *Journal of Computational Science*, vol. 25, pp. 289–297, 2017.

[24] X. Zhu, X. Ben, S. Liu, R. Yan, and W. Meng, "Coupled source domain targetized with updating tag vectors for micro-expression recognition," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3105–3124, 2018.

[25] X. Ben, X. Jia, R. Yan, X. Zhang, and W. Meng, "Learning effective binary descriptors for micro-expression recognition transferred by macro-information," *Pattern Recognition Letters*, vol. 107, pp. 50–58, 2017.

[26] Y. Zong, W. Zheng, Z. Cui, G. Zhao, and B. Hu, "Toward bridging microexpressions from different domains," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 5047–5060, 2020.

[27] Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, and D. Xu, "Image classification by cross-media active learning with privileged information," *IEEE Trans. on Multimedia*, vol. 18, no. 12, pp. 2494–2502, 2016.

[28] J. Wu, S. Zhao, V. S. Sheng, J. Zhang, C. Ye, P. Zhao, and Z. Cui, "Weak-labeled active learning with conditional label dependence for multilabel image classification," *IEEE Trans. on Multimedia*, vol. 19, no. 6, pp. 1156–1169, 2017.

[29] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, 2009.

[30] L. Copa, D. Tuia, M. Volpi, and M. Kanevski, "Unbiased query-by-bagging active learning for VHR image classification," *Remote Sensing*, vol. 7830, pp. 176–183, 2010.

[31] B. Li, Y. Zhou, R. Xiao, J. Wang, X. Ben, K. Kpalma, and H. Zhou, "Unsupervised cross-database micro-expression recognition based on distribution adaptation," *Multimedia Systems*, vol. 28, no. 3, pp. 1099–1116, 2022.

[32] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *International Conference on Machine Learning*, 2000, pp. 839–846.

[33] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, 2011.

[34] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, 2008.

[35] D. Wang and Y. Shang, "A new active labeling method for deep learning," in *International Joint Conference on Neural Networks*, 2014, pp. 112–119.

[36] H. Ranganathan, H. Venkateswara, S. Chakraborty, and S. Panchanathan, "Deep active learning for image classification," in *International Conference on Image Processing*, 2017, pp. 3934–3938.

[37] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectral-spatial feature learning model for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1741–1754, 2019.

[38] J. Lin, L. Zhao, S. Li, R. K. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4048–4062, 2018.

[39] Z. Peng, W. Zhang, N. Han, X. Fang, P. Kang, and L. Teng, "Active transfer learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1022–1036, 2020.

[40] L. Gerhardt, "Pattern recognition and machine learning," *IEEE Transactions on Automatic Control*, vol. 19, no. 4, pp. 461–462, 1974.

[41] W. N. Hsu and H. T. Lin, "Active learning by learning," in *29th AAAI Conference on Artificial Intelligence*, 2015, pp. 2659–2665.

[42] B. Settles, "Active learning literature survey," *University of Wisconsin-Madison*, 2010.

[43] S. Huang, R. Jin, and Z. Zhou, "Active learning by querying informative and representative examples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1936–1949, 2014.

[44] "Proximal toolbox," http://spams-devel.gforge.inria.fr/doc/html/doc_spams006.html#sec25.

[45] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog. Workshops*, 2010, pp. 94–101.

[46] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 356–370, 2019.

[47] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affective Comput.*, vol. 10, no. 1, pp. 18–31, 2019.

[48] W. Yan, Q. Wu, Y. Liu, S. Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog. Workshops*, 2013, pp. 1–7.

[49] W. Yan, X. Li, S. Wang, G. Zhao, Y. Liu, Y. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLOS ONE*, vol. 9, no. 1, p. e86041, 2014.

[50] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A spontaneous micro-facial movement dataset," *IEEE Trans. Affective Comput.*, vol. 9, no. 1, pp. 116–129, 2018.

[51] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. IEEE Int. Conf. Workshops Automatic Face Gesture Recog.*, 2013, pp. 1–6.

[52] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, and Y.-J. Liu, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021. [Online]. Available: <https://doi.org/10.1109/TPAMI.2021.3067464>

[53] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 113–127, 2015.

[54] R. Chattopadhyay, Z. Wang, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, "Batch mode active sampling based on marginal probability distribution matching," *ACM Trans. Knowledge Discovery from Data*, vol. 7, no. 3, pp. 1–25, 2013.

[55] Y.-J. Liu, B.-J. Li, and Y.-K. Lai, "Sparse MDMO: Learning a discriminative feature for micro-expression recognition," *IEEE Trans. Affective Comput.*, vol. 12, no. 1, pp. 254–261, 2021.

[56] Z. Xia, W. Peng, H. Q. Khor, X. Feng, and G. Zhao, "Revealing the invisible with model and data shrinking for composite-database micro-expression recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 8590–8605, 2020.

[57] L. Lei, J. Li, T. Chen, and S. Li, "A novel Graph-TCN with a graph structured representation for micro-expression recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2237–2245.

[58] H.-X. Xie, L. Lo, H.-H. Shuai, and W.-H. Cheng, "AU-assisted graph attention convolutional network for micro-expression recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2871–2880.

[59] T. Huang, X. Ben, C. Gong, B. Zhang, R. Yan, and Q. Wu, "Enhanced spatial-temporal salience for cross-view gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, 2022. [Online]. Available: <https://doi.org/10.1109/TCSVT.2022.3175959>



Tianhuan Huang received the B.E. degree in Electronic Information Engineering from School of Physical Science and Technology, Nanjing normal university, Nanjing, China, in 2018. She is currently a Ph.D candidate with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her current research interests include computer vision and machine learning.



Chuanye Li received the M.S. degree in School of Information Science and Engineering, Shandong University, Jinan, China, in 2018. His research interests include micro-expression spotting/detection and recognition.



Rui Yan graduated from Rensselaer Polytechnic Institute majoring in Computer Science with a Ph.D. degree. He is now an applied scientist at Amazon. His research interests include knowledge graph, machine learning and pattern recognition.



Yujun Li received PhD degree from Harbin Institute of Technology in 2001. He is currently working as a Full Professor with the School of Information Science and Engineering, Shandong University, Qingdao, China. His research interests include computer vision, natural language processing and sentiment analysis.



Xianye Ben received the Ph.D. degree in pattern recognition and intelligent system from the College of Automation, Harbin Engineering University, Harbin, China, in 2010. She is currently working as a Full Professor with the School of Information Science and Engineering, Shandong University, Qingdao, China. She has authored or coauthored more than 100 papers in major journals and conferences, such as IEEE T-PAMI, IEEE T-IP, IEEE T-CSVT, IEEE T-MM, PR, CVPR, etc. Her current research interests include pattern recognition and

image processing. She received the Excellent Doctorial Dissertation awarded by Harbin Engineering University. She was also enrolled by the Qilu Young Scholars Program of Shandong University.



Chen Gong received his B.E. degree from East China University of Science and Technology (ECUST) in 2010, and dual doctoral degree from Shanghai Jiao Tong University (SJTU) and University of Technology Sydney (UTS) in 2016 and 2017, under the supervision of Prof. Jie Yang and Prof. Dacheng Tao, respectively. Currently, he is a full professor in the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interests mainly include machine learning, data mining, and learning-based vision problems.

He has published more than 50 technical papers at prominent journals and conferences such as IEEE T-NNLS, IEEE T-IP, IEEE T-CYB, IEEE T-CSVT, IEEE T-MM, IEEE T-ITS, CVPR, AAAI, IJCAI, ICDM, etc. He received the Excellent Doctorial Dissertation awarded by Shanghai Jiao Tong University (SJTU) and Chinese Association for Artificial Intelligence (CAAI). He was also enrolled by the Summit of the Six Top Talents Program of Jiangsu Province, China, and the Lift Program for Young Talents of China Association for Science and Technology.