Image and Vision Computing xxx (2020) xxx

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A survey of micro-expression recognition

Ling Zhou [a,1], Xiuyan Shao [b,1], Qirong Mao [c,*]

[a] School of Computer Science and Communication Engineering, Jiangsu University, China
[b] Southeast University, China
[c] School of Computer Science and Communication Engineering, Jiangsu Engineering Research Center, Big Data Ubiquitous Perception and Intelligent Agriculture Applications, Jiangsu University, China

ARTICLE INFO

Article history:
Received 13 August 2020
Accepted 2 October 2020
Available online xxxx

Keywords:
Micro-expression recognition
Deep learning
Micro-expression datasets
Survey

ABSTRACT

The limited capacity to recognize micro-expressions with subtle and rapid motion changes is a long-standing problem that presents a unique challenge for expression recognition systems and even for humans. The problem regarding micro-expression is less covered by research when compared to macro-expression. Nevertheless, micro-expression recognition (MER) is imperative to exploit the full potential of expression recognition for real-world applications. Recent MER systems generally focus on three important issues: overfitting caused by a lack of sufficient training data, the imbalanced distribution of samples, and robust features for improving the accuracy of recognition. In this paper, we provide a comprehensive survey on MER, including datasets and algorithms that provide insights into these intrinsic problems. First, we introduce the available datasets that are widely used in the literature. We then describe the pre-processing in the standard pipeline of an MER system. For the state of the art in MER, we divide the existing novel algorithms into 6 different tasks according to the type of classes and evaluation protocols. Detailed experiment settings and competitive performances for those 6 tasks are summarized in this section. Finally, we review the remaining challenges and corresponding opportunities in this field as well as future directions for the design of robust MER systems.

1. Introduction

Micro-expression is a very brief and involuntary form of facial expressions occurring when people want to hide one's true emotions, usually lasts between 0.04 to 0.2 s and occurs in specific regions of the face [1,2]. Those brief micro-expressions are cumbersome to be recognized because of their subtleness and brevity. Automatic micro-expression analysis is a computer vision task that has been extensively studied for several years. There are two sub-task involved in automatic micro-expression analysis: micro-expression spotting and micro-expression recognition (MER) [3]. In our survey, we limit our discussion on the sub-task of MER.

Fig. 1 demonstrates the general pipeline of an MER system. Databases are the most basic guarantee for effective MER. MER aims to classify the micro-expressions into correct categories. There are two category types in publicly micro-expression databases [4–8], *i.e.*, emotion classes, and objective classes. After samples pre-processing in micro-expression databases, how to exploring a set of robust and discriminative features is the core issue and challenge in MER. Feature representation in MER can be divided into two main categories according to the method of obtaining the features: handcrafted features (*i.e.*, [9–14]) and learning features (*i.e.*, [15–23]).

Besides the micro-expression classes and feature types, there also three common evaluation protocols in MER which divide MER approaches into three categories: MER evaluated on the Sole Database Evaluation protocol (SDE), MER evaluated on Composite Database Evaluation (CDE) [24] protocol, and MER evaluated on Holdout-database Evaluation (HDE) [25–27] protocol. According to which type of classes the MER method focuses on and which evaluation protocol is used, the approaches of MER can be divided into 6 tasks as shown in Fig. 2. Exhaustive surveys on automatic micro-expression recognition have been published in recent years [3,28,29]. These surveys have established a set of standard algorithmic pipelines for MER. However, they focused mainly on Task 1 (emotion classes based and evaluated on SDE protocol), while other tasks were ignored. We make systematic research on MER tasks based on those 6 tasks in MER including the regrouped database information and evaluation metrics in the different MER tasks. We aim to give a newcomer to this filed an overview of the systematic framework and the state-of-the-art results of different tasks.

Based on the pipeline of an MER system as shown in Fig. 1 and the 6 tasks in MER as shown in Fig. 2, the rest of this paper is organized as follows. Publicly micro-expression databases are introduced in Section 2. Section 3 identifies some main steps required for an MER system in

* Corresponding author.
E-mail addresses: 2111808003@stmail.ujs.edu.cn (L. Zhou), xiuyan_shao@seu.edu.cn (X. Shao), mao_qr@ujs.edu.cn (Q. Mao).
[1] The authors have contributed equally to this work and are co-first authors.

https://doi.org/10.1016/j.imavis.2020.104043
0262-8856/

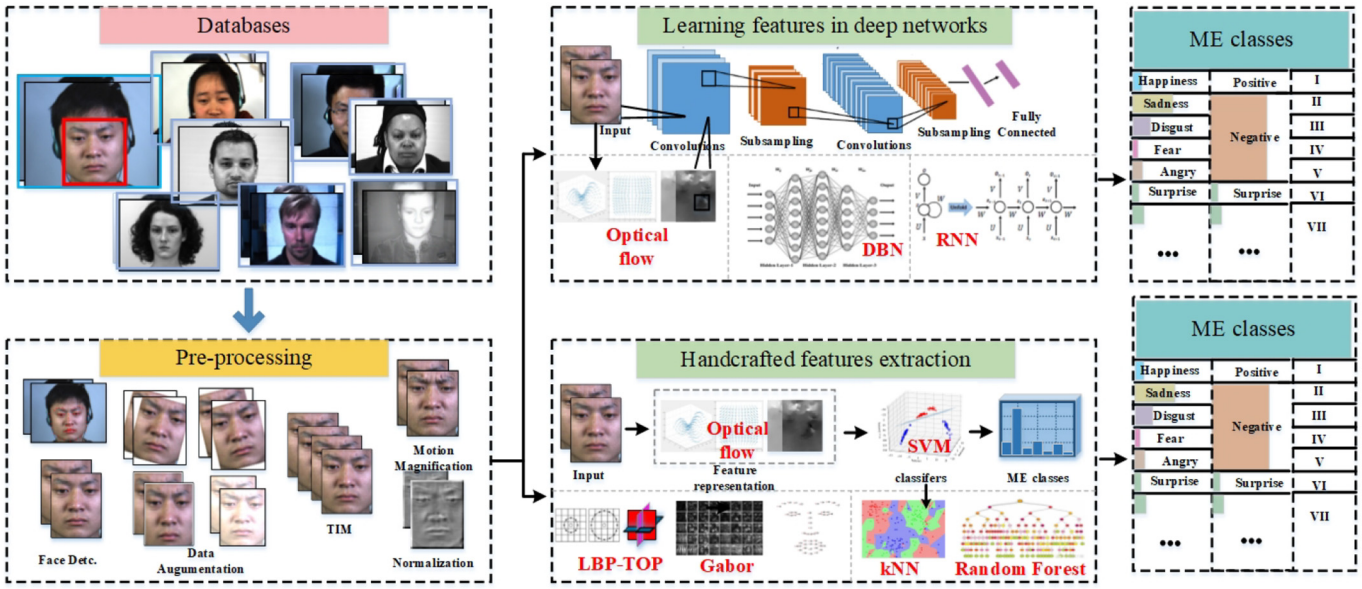

Fig. 1. The general pipeline of micro-expression recognition systems.

the pre-processing for all the tasks. Section 4 gives a detailed description of the 6 tasks, including the re-organized databases in each task, evaluation metrics, and the state-of-the-art methods of each task. We also give a comprehensive results comparison according to the different tasks. Section 5 discusses some of the challenges and opportunities in this field and identifies potential future directions.

2. Micro-expression databases

Nowadays, research on micro-expression recognition focuses the experiments on spontaneous expressions. CASME II [6], SMIC [5] and SAMM [7] are the three most commonly used databases thanks to their relatively sufficient micro-expression samples, rigorous record environment, and label process. MEVIEW [30] is the unique publicly micro-expression database connected in-the-wild conditions. Table 1 provides an overview of these datasets.

2.1. SMIC [5]

SMIC contains three sub-databases: SMIC-HS (recorded by a high-speed camera of 100 *fps*), SMIC-VIS (by a normal visual camera of 25 *fps*), and SMIC-NIR (by a near-infrared camera). There are 164 samples from 16 subjects, 71 samples from 8 subjects, and 71 samples from 8 subjects in SMIC-HS, SMIC-VIS and SMIC-NIR, respectively. The samples in three sub-databases are annotated as *Negative*, *Positive*, and *Surprise*. The sample resolution is 640 × 480 *pixels* and the facial area is around 190 × 230 *pixels*.

2.2. CASME [4]

The CASME database is laboratory-controlled and includes 195 samples from 35 subjects. Those sequences are labeled with 8 emotion classes, including *Amusement*, *Sadness*, *Disgust*, *Surprise*, *Contempt*, *Fear*, *Repression*,and *Tense*. All samples were also coded with the onset, apex and offset frames, with AUs marked. The sample resolution is 1280 × 720 *pixels* or 640 × 480 *pixels*, with the facial area around 190 × 230 *pixels*.

2.3. CASME II [6]

The CASME II database is an improved version of CASME which offers larger face size (around 280 × 340 *pixels* on facial area) and higher temporal resolution at 200 fps. It contains two versions: the first one includes 247 samples of 5 micro-expression classes (*Happiness*, *Surprise*,

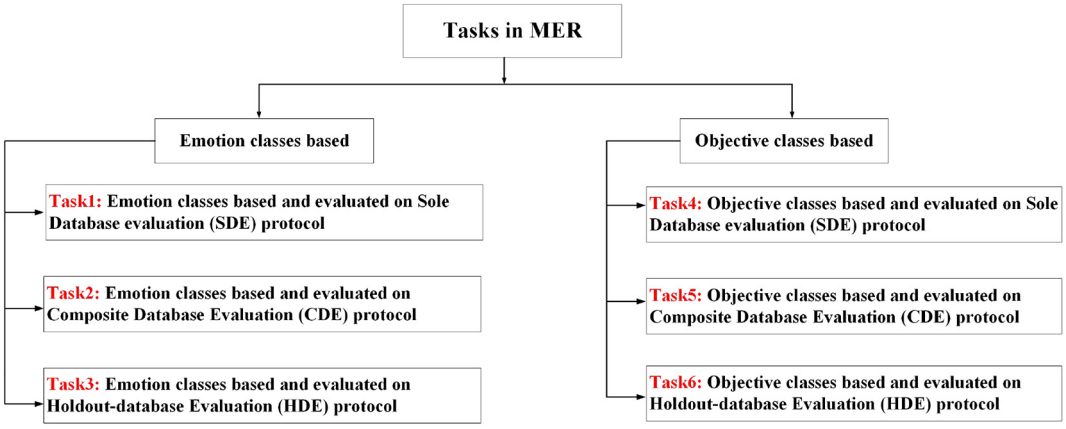

Fig. 2. Six tasks in MER.

Table 1
An overview of the spontaneous micro-expression datasets.

Database	Subversion	Samples	Subjects (Avg. age)	Ethn.	Fps	FACS Code	Sample Resolution	Face Resolution	Condition	Emot. Class	Obje. Class	Frame tag
SMIC [5]	SMIC-HS	164	16	3	100	N	640 * 480	190 * 230	Lab	3	–	–
	SMIC-VIS	71	8	3	25							
	SMIC-NIR	71	8	3	25							
	Website: http://www.cse.oulu.fi/SMICDatabase											
CASME [4]	–	195	35 (22.03)	1	60	Y	1280 * 720 640 * 480	150 * 190	Lab	8	–	Onset/Apex Offset
	Website: http://fu.psych.ac.cn/CASME/casme-en.php											
CASME II [6]	First	247	26 (22.03)	1	200	Y	640 * 480	280 * 340	Lab	5	–	Onset/Apex Offset
	Second	255								7	7	
	Website: http://fu.psych.ac.cn/CASME/casme2-en.php											
CAS(ME)2) [8]	PartA	87 (micro&macro)	22 (22.59)	1	30	Y	640 * 480	–	Lab	4	–	Onset/Apex Offset
	PartB	300 (57 micro)										
	Website: http://fu.psych.ac.cn/CASME/cas(me)2-en.php											
SAMM [7]	–	159	32 (33.24)	13	200	Y	2040 * 1080	400 * 400	Lab	7	7	Onset/Apex/Offset
	Website: http://www2.docm.mmu.ac.uk/STAFF/M.Yap/dataset.php											
MEVIEW [30]	–	31	16	–	25	N	–	–	Web	6	–	Onset/Offset
	Website: http://cmp.felk.cvut.cz/~cechj/ME/											

"-" = Not Report; "Avg. age" = the average age of the subjects; "Ethn." = Ethnicities; "Emot." = Emotion; "Obje." = Objective.

Disgust, *Repression*, and *Others*), and the second has 255 samples of 7 classes (*Happiness*, *Surprise*, *Disgust*, *Sadness*, *Fear*, *Repression*, and *Others*). All the samples are gathered from 26 subjects. Based on the labeled AUs, samples in the second of CASME II were relabeled into 7 objective classes: I–VII [31]. Some researchers also regrouped those emotion classes into three categories such as *Positive*, *Negative*, *Surprise*, or four categories which includes *Others* besides the three regrouped classes.

2.4. CAS(ME)2 [8]

The CAS(ME)2 is a laboratory-controlled database that contains both micro-expression and macro-expression in long videos. All the samples are recorded with 30 fps from 22 participants. Two parts are contained in CAS(ME)2. Part A contains 87 long videos that include spontaneous macro-expressions and micro-expressions, which is mainly used for micro-expression spotting task. Part B includes 300 cropped spontaneous macro-expression samples and 57 micro-expression samples, which can be used for micro-expression recognition. All the samples are encoded with AUs and 4 emotion classes are involved in the database: *Positive*, *Negative*, *Surprise* and *Others*.

2.5. SAMM [7]

The SAMM database provides 159 micro-expression instances from 32 participants at 200 *fps*. Seven classes of micro-expressions are captured including *Happiness*, *Surprise*, *Disgust*, *Repression*, *Angry*, *Fear*, *Contempt*, and *Others*. The samples are with the onset, apex and offset frames labeled and AUs encoded. Seven objective classes also provides based on the AUs. The resolution of the samples are 2040 × 1088 *pixels* and the resolution of facial area is around 400 × 400 *pixels*.

Table 2
Summary of different types of face detection/alignment methods that are widely used in MER models.

Method	Points	Speed	Used in
ASM [32]	68	Fair	[19,33–41]
AAM [42]	68	Fair	[43,44]
CLM [45]	68	Fair	[46]
DRMF [47]	66	Fast	[11,48]
TCDCN [49]	5	Fast	[50]
Face++ [51]	68	Fast	[7,31,52,53]

2.6. MEVIEW [30]

The MEVIEW is the first and unique micro-expression database which collected from the website. All the samples are mostly poker game videos downloaded from YouTube, including micro-expression and macro-expression. The database contains 31 samples from 16 subjects, which can be applied for the micro-expression spotting task. There is no uniform resolution for the sample.

3. Pre-processing

In this section, we describe the pre-processing step in MER as shown in Fig. 1. Pre-processing in MER mainly includes face detection and aline, frame normalization, motion magnification, and data augmentation. Among those pre-processing, face detection is the first and absolutely indispensable step.

3.1. Face detection and alignment

Face detection and alignment is a traditional pre-processing step in many face related recognition tasks. This step is crucial because it can remove background and non-face areas, then reduce the variation in face scale and in-plane rotation. Table 2 lists facial landmark detection algorithms widely-used in MER. The earliest well-known framework for face detection and alignment are the Active Shape Model (ASM) [32] and the Active Appearance Model (AAM) [42]. Recently, discriminative response

Table 3
The evaluation metrics and cross-validation protocols used in different tasks of MER.

Task	Detail description	Evaluation metric	Cross-validation protocol
Task 1	Emotion classes based, evaluated on SDE protocol	Acc/UF1	LOSO/LOSV/K-fold
Task 2	Emotion classes based, evaluated on CDE protocol	UF1/UAR	LOSO
Task 3	Emotion classes based, evaluated on HDE protocol	Acc/UF1	2-fold
Task 4	Objective classes based, evaluated on SDE protocol	Acc/UF1	LOSO/LOSV
Task 5	Objective classes based, evaluated on CDE protocol	UF1/WF1	LOSO
Task 6	Objective classes based, evaluated on HDE protocol	WAR/UAR	2-fold

map fitting (DRMF) [47], Constrained Local Model (CLM) [45], Face++ [51] and Tasks-Constrained Deep Convolutional Network (TCDCN) [49] were popularly used. In general, cascaded regression methods such as TCDCN and Multi-task Cascaded Convolutional Networks (MTCNN) [54] have become the most popular and state-of-the-art methods for face alignment as its high speed and accuracy. Besides, High-resolution networks (HRNets) [55] is also a well-known method for efficient landmark detection and face alignment, which would be popular used for face detection and alignment for MER tasks.

3.2. Frame normalization

Frame normalization means selecting meaningful frames or frames in the entire micro-expression sequence for MER. Many researchers in MER [33,34,37,53,56–58] applied the Temporal interpolation method (TIM) [59] is to uniformly align the input samples into the same number of frames. Given a micro-expression video, TIM can map high-dimensional visual features extracted from each frame onto a low-dimensional continuous curve determined by a set of trigonometric functions embedded within a path graph, and project the entire curve back into the original domain to get the interpolated video. Besides the TIM, Sparsity-Promoting Dynamic Mode Decomposition (DMDSP) [60] is reported in [61,62] to have the ability to improve recognition performance when interpolating a small number of frames. But with a larger number of frames, it would degrade the recognition performance because of over interpolation.

3.3. Motion magnification

By enlarging the motions, the motion magnification techniques allow us to see small motions previously invisible to the naked eyes, such as subtle motion changes of micro-expressions. Although it is not a vital stage in pre-processing of MER, some literature [16,57,63–68] reported that with proper motions magnification such as Eulerian Video Magnification (EVM) [69] and Global Lagrangian Motion Magnification (GLMM) [68], higher MER accuracy results can be achieved. EVM is the most popular method used in MER for motions magnification, which decomposes video frames into representations that facilitate manipulation of motions, without requiring explicit tracking. It consists of three stages: decomposing frames into an alternative representation, manipulating the representation, and reconstructing the manipulated representation to magnified frames. In the MER methods, Wang et al. [67], Li et al. [57] and Yu et al. [66] extracted handcrafted features after using EVM to magnify the sequences, while Liu et al. [16], Wang et al. [63] and Xia et al. [64,65] fed the samples into the deep learning models for better recognition accuracy. GLMM is an improved version of EVM, which was utilized by Le Ngo et al. [68] in MER to improve the model performance.

3.4. Data augmentation

Data argumentation is mainly used for training deep learning networks of MER as deep neural networks require sufficient training data to ensure the generalization of a given recognition task. In [18], 150 augmented sequences from each training sequence were generated. In detail, each training video sequence was flipped horizontally, rotated between a range angles with fixed increment, translated along with designed groups of pixels in x and y axis, and scaled with scaling different factors. Peng et al. [19] cut samples at different places, *i.e.*, up, down, left, right, center and upper left, upper right, lower left, and lower right part of the frame to enrich the training samples. Xia et al. [21] trained model with 50 times of original samples by proposing two multi-scale data augmentation strategies to enrich training samples. Samples in [22] are rotated between [−45°,45°] with the increment of 15°, enhanced by histogram equalization, and then all generated images are flipped horizontally to get enriched samples for training. Yu et al. [70] used GAN [71] to generate samples to tackle the overfilling issue. Extended Cohn-Kanade dataset (CK+) [72] were employed to enlarge the data size in [16], and CK+, Oulu-CASIA NIR&VIS facial expression [73]], Jaffe [74], and MUGFE [75] were combined as a single dataset to train the MER model in [76]. Since the classes of micro-expression databases are not always the same as the macro-expression database, a main pre-processing is needed before utilizing the macro-expression dataset to enrich the training data in MER, *i.e.,* regrouping the macro-expression data into the same classes as micro-expression.

4. Tasks of MER

AS shown in Fig. 2, there are 6 tasks in MER. In this section, we first briefly introduce the evaluation metrics used in MER tasks, then we summarize the state-of-the-art methods for MER and give detailed performance comparisons on these methods.

4.1. Evaluation metrics and cross-validation protocols

The metrics and cross-validation protocols used in different tasks of MER are listed in Table 3. Four metrics are introduced in the tasks to measure the performance of different approaches, *i.e.*, unweighted average recall (UAR), weighted average recall (WAR/Acc), Unweighted F1 score (UF1), and Weighted F1 score (WF1). Those four metrics (WAR/Acc, UAR, UF1, WF1) mentioned above are calculated as follows:

$$WAR = \frac{\sum_{c=1}^{C} TP_c}{N} \tag{1}$$

$$UAR = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{N_c} \tag{2}$$

$$UF1 = \frac{1}{C}\sum_{c=1}^{C}\frac{2\cdot TP_c}{2\cdot TP_c + FP_c + FN_c} \tag{3}$$

$$WF1 = \sum_{c=1}^{C}\frac{N_c}{N}\frac{2\cdot TP_c}{2\cdot TP_c + FP_c + FN_c} \tag{4}$$

where C is the number of classes, $c \le C$. N_c is the number of samples in the ground truth of the $c-$th class, and N is the total samples. TP_c, FP_c, and FN_c are the true positives, false positives, and false negatives of the $c-$th class, respectively.

As mentioned in Section 1, there are mainly three cross-validation protocols for MER, *i.e.,* LOSO, LOSV and K-fold. As shown in Table 3, LOSO protocols are used in Task 1, Task 2, Task 4, and Task 5, LOSV are mainly applied for Task 1 and Task 4, and K-fold protocol also can be found in Task 1. For the HDE task such as Task 3 and Task 6, 2-fold protocol is usually introduced.

4.2. The state-of-the-art in MER tasks

In this subsection, we review the state-of-the-art methods in different tasks of MER. Features are the most important part of a MER system, which can be categorized into handcrafted features and learning features. Handcrafted feature extraction approach mostly relies on the manually designed extractor, which needs professional knowledge and complex parameter adjustment process. In the meanwhile, each method suffers from poor generalization ability and robustness. Furthermore, due to the limited representation ability, engineered features may hardly handle the challenge of nonlinear feature warping caused by complicated situations, *e.g.* under different environments. Since 2015, with the well-designed network architecture *i.e.*, Inception [90], Resnet [91], Alexnet [92], and LSTM [93], studies in various fields have begun to transfer to deep learning methods. In this section, we summarize the performance of representative methods for the 6 tasks in MER, including handcrafted features and learning features.

Table 4
The detailed databases information used in Task 1. "#Class" = number of classes; "#S" = number of samples.

Database	#Classes (#S)	Detail information
SMIC-HS [5]	3 (164)	Positive (51), Negative (70), Sunrise(43)
CASME [4]	4 (97)	Positive (5), Negative (48), Surprise (14), Others (30)
	4 (150)	Repression (29), Disgust (40), Surprise (18), Tense (63)
	4 (167)	Positive (9), Negative (48), Surprise (15), Others (95)
	4 (171)	Disgust (44), Surprise (20), Repression (38), Tense (69)
	4 (180)	Disgust (−), Surprise (−), Repression (−), Tense (−)
CASME II [6]	5 (136)	Happiness (19), Surprise (16), Disgust (21), Repression (12), Tense (68)
	5 (236)	Positive (31), Negative (65), Surprise (21), Others (119)
	5 (246)	Repression (27), Happiness (32), Surprise (25), Disgust (63), Others (99)
	5 (247)	Repression (27), Happiness (32), Surprise (25), Disgust (64), Others (99)
	4 (256)	Negative (73)[a], Positive (32), Surprise (25), Others (126)[b]

[a] Negative class of CASME II: Disgust, Sadness and Fear.
[b] Others class of CASME II: Repression and Others.

4.2.1. Emotion classes based MER tasks

Section 2 gives a detailed description of the publicly micro-expression databases. In many specific tasks of MER, to enable the different databases to be used together, a common reduced set of classes are used, with appropriate mappings from the original classes. Table 4, Table 5, and Table 7 list the detail information of databases used in the emotion classed based tasks of Task 1, Task 2 and Task 3, respectively.

4.2.1.1. Task 1: Emotion classes based and evaluated on SDE protocol. Most handcrafted features based methods are originally designed for Task 1 (emotion classes based and evaluated on SDE protocol) as shown in Table 10. The handcrafted features used in MER also can be divided into two categories: appearance-based features and Geometric-based features. Local Binary Pattern from Three Orthogonal Planes (LBP-TOP) [9] is the most widely used appearance-based feature for micro-expression recognition. Due to its low computational complexity, many LBP-TOP variants has been proposed, *e.g.* LBP-TOP on TICS [77], LBP from three Mean Orthogonal Planes (LBP-MOP) [94], Spatiotemporal Completed Local Quantized Patterns (STCLQP) [34], hierarchical spatiotemporal descriptors [12], LBP with Six Intersection Points (LBP-SIP), Spatiotemporal LBP with integral projection (STLBP-IP) [10], discriminative spatiotemporal LBP with revisited integral projection (DiSTLBP-RIP) [58] and others [38,39]. Besides the LBP family, 3D Histograms of Oriented Gradients (3DHOG) [95,96] is another typical appearance-based feature which focuses on counting occurrences of gradient orientation in localized portions of the image sequence. Different from appearance-based features, geometric-based features aim to represent micro-expression samples by the aspect of face geometry, *e.g.* shapes, and location of facial landmarks, including the Delaunay-based Temporal Coding Model (DTCM) [43], Main Directional Mean Optical Flow (MDMO) [11], Facial Dynamics Map (FDM) [13], and Bi-Weighted Oriented Optical Flow (Bi-WOOF) [14].

Besides the handcrafted features for Task 1, many learning feature based methods have been proposed [18,21,22,56,63,97,98]. Kim et al. [18] proposed a feature representation for the spatial information at different temporal states, which was based on the Long Short-Term Memory (LSTM) [93] recurrent neural network. In [56], Li et al. claimed Three-stream 3D flow convolutional neural network [56] and evaluated on three micro-expression databases (SMIC, CASME and CASME II). It consisted of three data stream sub-networks that each sub-network with dependent input data, *i.e.,* grayscale frame, the vertical and horizontal optical flow. To exploring the semantic relationships between AUs and emotion classes, Lo et al. [89] unsized GCN to MER, it is also the first work that incorporating GCN into MER research. Compared with other deep methods, the better performance in [21,63] indicate that EVM can make sense for higher recognition accuracy, but the magnification factor of EVM is not easy to handle, which is mainly based on extensive experiments. In detail, Spatiotemporal Recurrent Convolutional Networks (STRCN) is proposed by Xia et al. for SMIC, CASME II, and SAMM databases. Besides the tailor-designed network for MER, proposing an effective data augmentation method to void the overfitting and designing a balanced loss to tackle the data imbalance issue are the important contributions of [21]. Xia et al. also proposed an end-to-end model for MER, although the performance was not as good as the multi-stage model.

4.2.1.2. Task 2: Emotion classes based and evaluated on CDE protocol. Task 2 was launched by MEGC 2019, with the baselines of LBP-TOP and Bi-WOOF. It has attracted a greater attention since 2019 and many literatures are proposed [15–17,65,70,99–101] as shown in Table 11.

As a deep learning feature based on CDE protocol, Peng et al. [19] leveraged Dual Temporal Scale Convolutional Neural Network (DTSCNN) for MER. The DTSCNN was the first work in MER that utilized a shallow two-stream neural network with inputs of optical-flow sequences, which shed light on important future MER research for exploring shallow networks for data-limited MER. Then a series of shallow networks

Table 6
The sub-tasks used in Task 3.

Type	sub-Task	Source Database	Target Database
Type-I	Exp.1: H→V	SMIC-HS	SMIC-VIS
	Exp.2: V→H	SMIC-VIS	SMIC-HS
	Exp.3: H→N	SMIC-HS	SMIC-NIR
	Exp.4: N→H	SMIC-NIR	SMIC-HS
	Exp.5: V→N	SMIC-VIS	SMIC-NIR
	Exp.6: N→V	SMIC-NIR	SMIC-VIS
Type-II	Exp.7: C→H	CASME II	SMIC-HS
	Exp.8: H→C	SMIC-HS	CASME II
	Exp.9: C→V	CASME II	SMIC-VIS
	Exp.10: V→C	SMIC-VIS	CASME II
	Exp.11: C→N	CASME II	SMIC-NIR
	Exp.12: N→C	SMIC-NIR	CASME II

Table 5
The detailed databases information used in Task 2. "#Class" = number of classes; "#S" = number of samples.

Database	#Classes (#S)	Detail information
SMIC-HS [5]	3 (164)	Negative (70), Positive (51), Surprise (43)
CASME II [6]	3 (145)	Negative(88)[a], Positive (32), Surprise (25)
SAMM [7]	3 (133)	Negative (92)[b], Positive (26), Surprise (15)
Composite	3 (442)	Negative (250), Positive (109), Surprise (83)

[a] Negative class of CASME II: Disgust and Repression.
[b] Negative class of SAMM: Anger, Contempt, Disgust, Fear and Sadness.

Table 7
The detailed databases information used in Task3. "#Class" = number of classes; "#S" = number of samples.

Database	#Classes (#S)	Detail information
SMIC-HS [5]	3 (164)	Negative (70), Positive (51), Surprise (43)
SMIC-NIR [5]	3 (71)	Negative (23), Positive (28), Surprise (20)
SMIC-VIS [5]	3 (71)	Negative (23), Positive (28), Surprise (20)
CASME II [6]	3 (130)	Negative (73)[a], Positive (32), Surprise (25)

[a] Negative class of CASME II: Disgust, Sadness and Fear.

Table 8
Relationship between action unit and objective classes I–V.

Class	Action Units
I	AU6, AU12, AU6 + AU12, AU6 + AU7 + AU12, AU7 + AU12
II	AU1 + AU2, AU5, AU25, AU1 + AU2 + AU25, AU25 + AU26, AU5 + AU24
III	A23, AU4, AU4 + AU7, AU4 + AU5, AU4 + AU5 + AU7, AU17 + AU24, AU4 + AU6 + AU7, AU4 + AU38
IV	AU10, AU9, AU4 + AU9, AU4 + AU40, AU4 + AU5 + AU40, AU4 + AU7 + AU9, AU4 + AU9 + AU17, AU4 + AU7 + AU10, AU4 + AU5 + AU7 + AU9, AU7 + AU10
V	AU1, AU15, AU1 + AU4, AU6 + AU15, AU15 + AU17
VI	AU1 + AU2 + AU4, AU20
VII	Others

Table 9
The detailed databases information used in Task4–6.

Database	Objective Class							
	I	II	III	IV	V	VI	VII	Total
CASME II [6]	25	15	99	26	20	1	69	255
SAMM [7]	24	13	20	8	3	7	84	159
Composite	49	28	119	34	23	8	153	415

were proposed for MER [15–17,65,70,99–101]. More specifically, Optical Flow Feature from Apex frame Network (OFF-ApexNet) [99] and its modified version [100] extracted optical flow features from the onset and apex frames of each video, then learned features representation by feeding horizontal and vertical components of optical flows into a two-stream CNN network. Since the CNN network was shallow, it reduces the over-fitting caused by the scarcity of data in the micro-expression databases. Expression Magnification and Reduction (EMR) with adversarial training [16] was a part-based deep neural network approach with adversarial training and expression magnification. With the special data augmentation strategy of expression magnification and reduction, EMR won the first place in MEGC 2019. Shallow Triple Stream Three-dimensional CNN (STSTNet) [15] is an extended version of OFF-ApexNet. Dual-Inception [101] was achieved by feeding the optical flow features extracted from the onset and mid-position frames into a designed two-stream Inception network. With data augmentation methods, Quang et al. [17] applied Capsule Networks (CapsuleNet) based on the apex frames to MER.

Besides these shallow networks mentioned above, some deep depth networks were also applied for MER, such as [23,65,70]. Specifically, motivated by attention strategy and transfer learning mechanism, Zhou et al. [23] utilized Resnet with the input of apex frames for MER, while the performance was not satisfactory. Xia et al. [65] designed a recurrent convolutional network (RCN) to improve recognition performance by exploring the shallower-architecture and lower-resolution input data. Yu et al. [70] proposed a novel Identity-aware and Capsule-Enhanced Generative Adversarial Network (ICE-GAN) model to improve MER performance in an end-to-end way. The samples generated by GAN avoid overfitting in the training model.

Table 10
Task 1: Performances of representative methods for emotion classes based micro-expression recognition on SDE protocol. "Pre-prop." = Pre-processing; "DA" = data augmentation.

Paper	Method	Pre-	Cross	SMIC-HS			CASME II			CASME		
Year		proc.	validation	#Class (#S.)	Acc	UF1	#Class (#S.)	Acc	UF1	#Class (#S.)	Acc	UF1
Handcrafted features												
2013 [5]	LBP-TOP	TIM	LOSO	3(164)	48.78	–	–	–	–	–	–	–
2014 [77]	LBP-TOP on TICS	–	LOVO	–	–	–	5 (136)	61.77	–	4 (97)	61.86	–
2014 [37]	DLSTD	ASM TIM	LOVO	3 (164)	68.29	–	5 (246)	63.41	–	–	–	–
2014 [37]	LBP-SIP	–	LOSO	3 (164)	44.51	0.4492	5 (246)	46.56	0.4451	–	–	–
2014 [37]	LBP-SIP	–	LOVO	–	–	–	5 (246)	67.21	–	–	–	–
2014 [78]	OS	–	LOSO	3 (164)	53.56	–	–	–	–	–	–	–
2014 [79]	weighted OS	–	LOSO	3 (164)	53.66	–	–	–	–	–	–	–
2014 [80]	STM	TIM	LOSO	3 (164)	44.34	0.4731	5 (246)	43.78	0.3337	–	–	–
2015 [38]	LBP-MOP	–	LOSO	3 (164)	50.61	–	5 (246)	44.13	–	–	–	–
2015 [38]	LBP-MOP	–	LOVO	3 (164)	60.98	–	5 (246)	66.80	–	–	–	–
2015 [81]	LBP-TOP	Adaptive MM	LOSO	–	–	–	5 (247)	51.91	–	–	–	–
2015 [81]	LBP-TOP	Adaptive MM	LOVO	–	–	–	5 (247)	69.63	–	–	–	–
2015 [10]	STLBP-IP	–	LOSO	3(164)	57.93	–	5 (247)	59.51	–	–	–	–
2016 [34]	STCLQP	–	LOSO	3 (164)	64.02	0.6381	5 (247)	58.39	0.5836	4 (171)	57.31	0.5000
2016 [11]	MDMO	DRMF	LOSO	–	–	–	4 (236)	67.37	–	4 (167)	68.86	–
2016 [14]	Bi-WOOF	–	LOSO	3 (164)	62.20	0.6200	5 (247)	57.89	0.6100	–	–	–
2016 [82]	LBP-TOP	EVM	LOSO	–	–	–	5 (247)	–	0.5100	–	–	–
2017 [61]	LBP-TOP	DMDSP	LOSO	3 (164)	58.00	0.6000	5 (247)	49.00	0.5100	–	–	–
2017 [83]	Bi-WOOF + Phase	–	LOSO	3 (164)	68.29	0.6700	5 (247)	62.55	0.6150	–	–	–
2017 [13]	FDM	ASM	LOSO	3 (164)	64.02	–	5 (247)	45.93	0.4053	4 (171)	46.14	0.4912
2017 [84]	OF Maps	–	LOSO	–	–	–	5 (246)	65.35	–	–	–	–
2017 [85]	MMFL	TIM	LOSO	3 (164)	63.15	–	5 (246)	59.81	–	–	–	–
2017 [86]	STRBP	TIM	LOSO	3 (164)	60.98	–	5 (247)	64.37	–	–	–	–
2017 [58]	STLBP-RIP	TIM	LOSO	3 (164)	60.98	–	5 (247)	64.37	–	–	–	–
2017 [58]	DiSTLBP-RIP	TIM	LOSO	3(164)	63.41	–	5 (247)	64.78	–	4 (171)	64.33	–
2018 [57]	HIGO	TIM EVM	LOSO	3 (164)	67.21	–	5 (247)	68.29	–	–	–	–
2018 [87]	Hierarchical STLBP-IP	–	LOSO	3 (164)	60.78	0.6126	5 (247)	63.83	0.6110	–	–	–
2018 [88]	FMBH	–	LOSO	3 (164)	64.02	–	5 (246)	62.60	–	4 (150)	52.00	–
Learning features												
2016 [18]	CNN + LSTM	DA	LOSO	–	–	–	5 (246)	60.98	–	–	–	–
2018 [56]	3DFCNN	–	LOSO	3 (164)	55.49	–	5 (247)	59.11	–	4 (180)	55.44	–
2020 [89]	MER-GCN	–	LOSO	–	–	–	5 (247)	58.82	–	–	–	–
2020 [89]	MER-GCN	–	k-fold	–	–	–	5 (247)	42.71	–	–	–	–
2020 [63]	EM-C3D	EVM	LOSO	–	–	–	4 (255)	69.761	–	–	–	–
2020 [21]	STRCN	DA EVM	LOSO	3 (164)	72.30	0.6950	4 (255)	80.30	0.7470	–	–	–
2020 [21]	STRCN	DA EVM	LOVO	3 (164)	74.90	0.7100	4 (255)	83.30	0.8070	–	–	–

Table 11
Task 2: Performances of representative methods for emotion classes based micro-expression recognition on CDE protocol. All the results are obtained by the LOSO cross-validation.

Paper	Method	Composite		SMIC-HS		CASME II		SAMM	
Year		UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
Handcrafted features									
2007 [9][a]	LBP-TOP	0.5882	0.5785	0.2000	0.5280	0.7026	0.7429	0.3954	0.4102
2016 [14][a]	Bi-WOOF	0.6296	0.6227	0.5727	0.5829	0.7805	0.8026	0.5211	0.5139
Learning features									
2019 [17]	CapsuleNet	0.6520	0.6506	0.5820	0.5877	0.7068	0.7018	0.6209	0.5989
2019 [99]	OFF-ApexNet	0.7196	0.7096	0.6817	0.6695	0.8764	0.8681	0.5409	0.5392
2019 [100]	modified OFF-ApexNet	0.7301	0.6964	–	–	–	–	–	–
2019 [101]	Dual-Inception	0.7322	0.7278	0.6645	0.6726	0.8621	0.8560	0.5868	0.5663
2019 [15]	STSTNet	0.7353	0.7605	0.6801	0.7013	0.8382	0.8686	0.6588	0.6810
2019 [16]	EMR	0.7885	0.7824	0.7461	0.7530	0.8293	0.8209	0.7754	0.7152
2020 [65]	RCN	0.7052	0.7164	0.5980	0.5991	0.8087	0.8563	0.6771	0.6976
2020 [70]	ICE-GAN	0.8450	0.8410	0.7900	0.7910	0.8760	0.8680	0.8550	0.8230

[a] Means that the emotion classed based MER results were not reported in the originally literatures, but implemented by See et al. in [24].

4.2.1.3. Task 3: Emotion classes based and evaluated on HDE protocol. There are 12 sub-experiments in Task 3. According to whether the source and target databases involve the same subjects, 12 sub-experiments are further grouped into TYPE-I and TYPE-II. The detailed setup is depicted in Table 6. Each source to target sub-experiment of Task 3 is denoted by $Exp.\ i : S \rightarrow T$, where *Expi* is the number of this sub-experiment, S and T are the source and target databases, respectively.

Task 3 is first introduced by Zong et al. [25], then a benchmark of Tasks was released in [27]. All the features are implemented by Zong et al. [27] and aimed to given a benchmark for MER researcher for cross-database MER. There are two features specifically designed for Task 3: Target Sample Re-Generator (TSRG) [25], Feature Space with unchanged Target domain (DRFS-T) [12] and Domain Regeneration in the Label Space (DRLS) [12]. TSRG aimed to learn a sample regenerator for the target micro-expression samples and reduce the feature distribution gap between the source and target micro-expression databases. DRFS-T and DRLS were established in [12], which were the improved version of TSRG.

Since Zong et al. have established a benchmark for Task 3, we just cite the results from [27] in Table 12 and Table 13. The parameter settings for comparative algorithms in the two tables are described as followed:

- For LBP-TOP [9], the uniform pattern is employed for LBP coding. The neighboring radius R and the number of the neighboring points P are two importance parameters. The experiment considers four groups of R and P, *i.e.*, R1P4, R1P8, and R3P8.
- For LBP-SIP [77], the only one parameter of the neighboring radius R is set as 1 and 3, respectively.
- For LPQ-TOP [102], the size of the local window in each dimension is set as a default value of [5,5,5], and the factor for correlation model *decorr* is set as [0.1, 0.1] and [0, 0], respectively.
- For HOG-TOP [57] and HIGO-TOP [57], there is only one parameter of them, *i.e.*, the number of bins p. Both in HOG-TOP and HIGH-TOP, p is set as 4 and 8, respectively.
- For three dimensional convolutional neural network (C3D) [103], the micro-expression features are extracted from the last two fully connected layers from the pre-trained Sport-1 M [104] and UCF101 [105] model.

4.2.2. Objective classes based MER tasks

As shown in Table 8, there are 7 objective classes in CASME II and SAMM databases, *i.e.*, class I–VII. The number of each class in the two databases are listed in Table 9. Task 4, Task 5, and Task 6 are based on the objective classes: Task 4 is evaluated on SDE protocol, aims to identify 5 or 7 objective classes in a solo database; Task 5 is similar to Task 2, while

Table 12
Task 3 (Type-II): Performances (UF1/Acc) of representative methods for emotion classes based micro-expression recognition on HDE protocol.

Feature	Exp.1: H→V	Exp.2: V→H	Exp.3: H→N	Exp.4: N→H	Exp.5: V→N	Exp.6: N→V	Average
Handcrafted features							
LBP-TOP(R3P8) [9]	0.8002/80.28	0.5421/54.27	0.5455/53.52	0.4878/54.88	0.6186/63.38	0.6078/63.38	0.6003/61.62
LBP-TOP(R1P4) [9]	0.7185/71.83	0.3366/40.24	0.4969/49.30	0.3457/40.24	0.5480/57.75	0.5085/59.15	0.4924/53.32
LBP-TOP(R1P8) [9]	0.8561/85.92	0.5329/53.66	0.5164/57.75	0.3246/35.37	0.5124/57.75	0.4481/50.70	0.5318/56.86
LBP-TOP(R3P4) [9]	0.4656/49.30	0.4122/45.12	0.3682/40.85	0.3396/40.85	0.5069/59.15	0.5144/60.56	0.4345/49.31
LBP-SIP(R1) [77]	0.6290/63.38	0.3447/40.85	0.3249/33.80	0.3490/42.07	0.5477/60.56	0.5509/60.56	0.4577/50.20
LBP-SIP(R3) [77]	0.8574/85.92	0.4886/50.00	0.4977/54.93	0.4038/42.68	0.5444/59.15	0.3994/46.48	0.5319/56.53
LPQ-TOP(decorr = 0.1) [102]	0.9455/94.37	0.5523/54.88	0.5456/61.97	0.4729/47.56	0.5416/57.75	0.6365/66.20	0.6157/63.79
LPQ-TOP(decorr = 0) [102]	0.7711/77.46	0.4726/48.78	0.6771/67.61	0.4701/48.17	0.7076/71.83	0.6963/70.42	0.6325/64.05
HOG-TOP(p = 4) [57]	0.7068/71.83	0.5649/57.32	0.6977/70.42	0.2830/29.27	0.4569/49.30	0.3218/36.62	0.4554/48.47
HOG-TOP(p = 8) [57]	0.7364/74.65	0.5526/56.10	0.3990/46.48	0.2941/32.32	0.4137/46.48	0.3245/38.03	0.4453/49.01
HIGO-TOP(p = 4) [57]	0.7933/80.28	0.4775/50.61	0.4023/47.89	0.3445/35.98	0.5000/53.52	0.3747/40.85	0.4821/51.52
HIGO-TOP(p = 8) [57]	0.8445/84.51	0.5186/53.66	0.4793/54.93	0.4322/43.90	0.5054/54.93	0.4056/46.48	0.5309/56.40
TSRG [25]	0.8869/88.73	0.5652/56.71	0.6484/64.79	0.5770/57.93	0.7056/70.42	0.8116/81.69	0.6991/70.05
DRFS-T [12]	0.8643/85.92	0.5767/57.32	0.7179/71.83	0.6163/61.59	0.7286/73.24	0.7732/77.46	0.7128/71.23
DRLS [12]	0.8604/85.92	0.6120/60.98	0.6599/66.20	0.5599/55.49	0.6620/69.01	0.5771/61.97	0.6552/66.60
Learning features							
C3D-FC1 (Sports1M) [103]	0.1577/30.99	0.2188/23.78	0.1667/30.99	0.3119/34.15	0.3802/49.30	0.3032/36.62	0.2564/34.31
C3D-FC2 (Sports1M) [103]	0.2555/36.62	0.2974/29.27	0.2804/33.80	0.3239/36.59	0.4518/47.89	0.3620/38.03	0.3285/37.03
C3D-FC1 (UCF101) [103]	0.3803/46.48	0.3134/34.76	0.3697/47.89	0.3440/34.76	0.3916/47.89	0.2433/29.58	0.3404/40.23
C3D-FC2 (UCF101) [103]	0.4162/46.48	0.2842/32.32	0.3053/42.25	0.2531/28.05	0.3937/47.89	0.2489/32.39	0.3169/38.23

Table 13
Task 3 (Type-II): Performances (UF1/Acc)of representative methods for emotion classes based micro-expression recognition on HDE protocol.

Feature	Exp.7: C→H	Exp.8: H→C	Exp.9: C→V	Exp.10: V→C	Exp.11: C→N	Exp.12: N→C	Average
Handcrafted features							
LBP-TOP(R3P8) [9]	0.3697/45.12	0.3245/48.46	0.4701/50.70	0.5367/53.08	0.5295/52.11	0.2368/23.85	0.4112/45.55
LBP-TOP(R1P4) [9]	0.3358/44.51	0.3260/47.69	0.2111/35.21	0.1902/26.92	0.3810/43.66	0.24.92/2692	0.2823/37.49
LBP-TOP(R1P8) [9]	0.3680/43.90	0.3339/54.62	0.4624/49.30	0.5880/57.69	0.3000/33.80	0.1927/23.08	0.3742/43.73
LBP-TOP(R3P4) [9]	0.3117/43.90	0.3436/44.62	0.2723/39.44	0.2356/28.46	0.3818/48.30	0.2332/25.38	0.2964/38.52
LBP-SIP(R1) [77]	0.3580/45.12	0.3039/44.62	0.2537/38.03	0.1991/26.92	0.3610/46.48	0.2194/26.92	0.2825/38.02
LBP-SIP(R3) [77]	0.3772/42.68	0.3742/56.15	0.5846/59.15	0.6065/60.00	0.3469/35.21	0.2790/27.69	0.4279/46.81
LPQ-TOP(decorr = 0.1) [102]	0.3060/42.07	0.3852/48.46	0.2525/33.80	0.4866/47.69	0.3020/35.21	0.2094/23.85	0.3236/38.51
LPQ-TOP(decorr = 0) [102]	0.2368/43.90	0.2890/51.54	0.2531/38.03	0.3947/40.77	0.2369/35.21	0.4008/41.54	0.3019/41.83
HOG-TOP(p = 4) [57]	0.3156/3476	0.3502/47.69	0.3266/35.21	0.4658/49.20	0.3219/35.21	0.2163/27.46	0.3327/37.91
HOG-TOP(p = 8) [57]	0.3992/43.90	0.4154/52.31	0.4403/45.07	0.4678/47.69	0.4107/40.85	0.1390/20.77	0.3787/41.77
HIGO-TOP(p = 4) [57]	0.2945/3963	0.3420/53.85	0.3236/40.85	0.5590/55.38	0.2887/29.58	0.2668/31.54	0.3458/41.81
HIGO-TOP(p = 8) [57]	0.2978/41.46	0.3609/50.00	0.3679/43.66	0.5699/54.62	0.3395/33.80	0.1743/22.31	0.3517/40.98
TSRG [25]	0.5042/51.83	0.5171/60.77	0.5935/59.15	0.6208/63.08	0.5624/56.34	0.4105/46.15	0.5348/56.22
DRFS-T [12]	0.4524/46.95	0.5460/60.00	0.6217/63.38	0.6762/68.46	0.5369/56.34	0.4653/50.77	0.5498/57.65
DRLS [12]	0.4924/53.05	0.5267/59.23	0.5757/57.75	0.5942/60.00	0.4885/49.83	0.3838/42.37	0.5102/53.71
Learning features							
C3D-FC1 (Sports1M) [103]	0.1994/42.68	0.2394/56.15	0.1631/32.39	0.1075/19.23	0.1631/3239	0.2397/5615	0.1854/3983
C3D-FC2 (Sports1M) [103]	0.1994/42.68	0.1317/24.62	0.1631/32.39	0.1075/19.23	0.1631/32.39	0.2397/5615	0.1674/34.58
C3D-FC1 (UCF101) [103]	0.1581/31.10	0.1075/19.23	0.1886/39.44	0.1075/19.23	0.1886/39.44	0.2397/5615	0.1650/34.10
C3D-FC2 (UCF101) [103]	0.1994/42.68	0.1705/19.23	0.1631/32.39	0.1075/19.23	0.1631/32.39	0.1075/19.23	0.1414/27.53

based on objective classes from SAMM and CASME II. Task 6 is the same as the cross-database evaluation, which means training and testing samples are from different databases for each fold. Task 5 and Task 6 are the tasks launched by MEGC 2018 [26], which focus on the objective classes I–V.

Three handcrafted features were implemented in [31] as the baselines for objective class-based MER which were originally designed for emotion classes based MER, *i.e.*, LBP-TOP, HOOF, and 3DHOG. Compared with the number of literature for emotion classes based MER, objective classes based MER receives loss attention [31,53,76].

Two deep learning features for objective classes based MER are proposed by Peng et al. [76] and Khor et al. [53]. To better represent the subtle changes in micro-expression, Khor et al. adopted Enriched Long-term Recurrent Convolutional Network (ELRCN) [53] based on optical flow features. It contained the channel-wise for spatial enrichment and the feature-wise for temporal enrichment predicted the micro-expression by passing the feature vector through LSTM. The good performance in [53] is partly contributed to the applying of motion information of optical flow. Optical flow encodes the motion of an object in vectorized notations, indicating the direction and intensity of the motion or flow of image pixels. The horizontal and vertical readily portray the subtle changes exhibited by micro-expressions. [53] is training on a deep learning network with a large number of learnable parameters. Without data augmentation, the performance of [53] is not as good as [76], which received the best recognition result in MEGC 2018.

To alleviate the overfitting issue, Peng et al. [76] adopted pre-trained Resnet10 [106] as a backbone and introduced the transfer learning strategy to improve MER performance. Specifically, RestNet10 was trained on ImageNet [92] then was fine-tuned on some public macro-expression databases, and finally fine-tuned on the CASMEII and SAMM databases by using apex frames. The delicate pre-training process on four macro-expression databases was the key to improving the performance in [76]. Since [76] was based on the apex frame, the domain gap inevitably existed in Task 5 and Task 6, where the training and testing samples are from two different micro-expression databases. Under this setting, the training and testing samples would have different feature distributions and hence the performance may decrease. Thus, how to incorporate domain adaption strategy into the cross-database MER, is also a significant issue in MER that need to be tackled.

Table 14, Table 15 and Table 16 show the comparison result of existing method for objective classes based MER which evaluated on SDE, CDE and HDE protocols, respectively.

5. Challenges and opportunities

5.1. Micro-expression dataset

As the low-intensity characteristic of micro-expression, many researchers have committed to employing deep learning technologies for MER. Given that MER is a data-driven task and that training a sufficiently

Table 14
Task 4: Performances of representative methods for objective classes based micro-expression recognition on SDE protocol. All the results are obtained by the LOSO cross-validation.

Paper	Method	#Class	CASME II		SAMM	
Year			Acc	UF1	Acc	UF1
Handcrafted features						
2007 [9][a]	LBP-TOP	I–V	67.80	0.51	44.70	0.35
		I–VI	67.94	0.51	45.89	0.31
		I–VII	61.92	0.35	54.93	0.39
2009 [95][a]	HOG3D	I–V	69.53	0.51	34.16	0.22
		I–VI	69.87	0.51	36.39	0.19
		I–VII	61.33	0.51	63.93	0.44
2016 [11][a]	HOOF	I–V	69.64	0.56	42.17	0.33
		I–VI	73.52	0.60	40.89	0.27
		I–VII	76.60	0.55	60.06	0.48
Learning features						
[53]	ELRCN	I–V	52.44	0.500	–	–

[a] Means that the objective classed based MER results were not reported in the originally literatures, but implemented by Davison et al. in [31].

Table 15
Task 5: Performances of representative methods for objective classes based micro-expression recognition on CDE protocol.

Paper	Method	UF1	WF1
Year			
Handcrafted features			
2007 [9][a]	LBP-TOP	0.400	0.524
2009 [95][a]	3DHOG	0.271	0.436
2016 [11][a]	HOOF	0.404	0.527
2018 [31]	–	0.454	0.579
Learning features			
2018 [53]	ELRCN	0.393	0.523
2018 [76]	–	0.639	0.733

[a] Means that the objective classed based MER results were not reported in the originally literatures, but implemented by Davison et al. in [31].

Table 16
Task 6: Performances of representative methods for objective classes based micro-expression recognition on HDE protocol.

Paper	Method	CASME II→SAMM		SAMM→CASME II		Avg.	
Year		WAR	UAR	WAR	UAR	WAR	UAR
Handcrafted features							
2007 [9][a]	LBP-TOP	0.338	0.327	0.232	0.316	0.285	0.322
2009 [95][a]	3DHOG	0.353	0.269	0.373	0.187	0.363	0.228
2016 [11][a]	HOOF	0.441	0.349	0.265	0.346	0.353	0.348
Learning features							
2018 [53]	ELRCN	0.485	0.382	0.384	0.322	0.435	0.352
2018 [76]	–	0.544	0.440	0.578	0.337	0.561	0.389

[a] means that the objective classed based MER results were not reported in the original literatures, but implemented by Davison et al. in [31].

deep network to capture subtle micro-expression related deformations requires a large amount of training data, the major challenge that deep MER systems face is the lack of training data in terms of both quantity and quality. Because people of different age ranges, cultures and genders display and interpret micro-expression in different ways, an ideal micro-expression dataset is expected to include abundant samples with larger age ranges and multi-cultures, which would facilitate the reliability of the database and the system of MER. On the other hand, accurately annotating of micro-expression data is an obvious impediment to the construction of micro-expression datasets. Further exploration and verification are needed for labeling micro-expressions with self-report and AUs, or only according to one of them.

5.2. End-to-end MER system

nother major issue that requiring consideration is that there are few end-to-end models for MER. Since the subtle changes in micro-expression, deep learning networks in MER were fed with optical flow features that were extracted in the pro-processing stage. Using the apex frame to represent each micro-expression sample or computing optical flow from the onset and apex frames is considered as an effective method that has been verified in many deep learning literature [15,16,21,23,40,99,101]. These methods always extracted exact optical flows using traditional handcrafted methods [107–110] where optical flows are needed to be pre-computed and stored on disk, leading to multi-stage MER approaches. With the development of deep learning flow methods [111–114], it is worth considering to incorporate flow learning network into the deep learning MER systems. With the end-to-end training, the parameters of the flow network can be further fine-tuned, and the MER performance can be improved by the learned richer and task-specific patterns beyond exact optical flow.

5.3. Multimodal MER

Last but not the least, human expressive behaviors in realistic applications involve encoding from different perspectives, and the micro-expression is only one modality. As pure micro-expression recognition based on visible face samples can not achieve promising results, incorporating with other models into a high-level framework can provide complementary information and further enhance the robustness. For example, the fusion of other modalities, such as audio, heart rate, and physiological data, is becoming a promising research direction due to the large complementarity for micro-expression.

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

This work was supported in part by the National Natural Science Foundation of China under Grant 61672267, Grant U1836220, in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX19 1616, Qing Lan Talent Program of Jiangsu Province, Jiangsu Engineering Research Center of big data ubiquitous perception and intelligent agriculture applications, and the National Natural Science Foundation of China under Grant 72001040, Zhishan Youth Scholar Program of Southeast University.

References

[1] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage, (Revised Edition) WW Norton & Company, 2009.
[2] P. Ekman, W.V. Friesen, Nonverbal leakage and clues to deception? Psychiatry 32 (1) (1969) 88–106.
[3] Y. Oh, J. See, A.C.L. Ngo, R.C. Phan, V.M. Baskaran, A survey of automatic facial micro-expression analysis: databases, methods and challenges, Front. Psychol. 9 (2018) 1128–1140.
[4] W. Yan, Q. Wu, Y. Liu, S. Wang, X. Fu, CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces, Proceedings of the 2013 International Conference and Workshops on Automatic Face and Gesture Recognition (FG) 2013, pp. 1–7.
[5] X. Li, T. Pfister, X. Huang, G. Zhao, M. Pietikäinen, A spontaneous micro-expression database: inducement, collection and baseline, Proceedings of the 2013 International Conference and Workshops on Automatic Face and Gesture Recognition (FG) 2013, pp. 1–6.
[6] W. Yan, X. Li, S. Wang, G. Zhao, Y. Liu, Y. Chen, X. Fu, Casme ii: an improved spontaneous micro-expression database and the baseline evaluation, PLoS One 9 (1) (2014) 1–8.
[7] A.K. Davison, C. Lansley, N. Costen, K. Tan, M.H. Yap, SAMM: a spontaneous micro-facial movement dataset, IEEE Trans. Affect. Comput. 9 (1) (2018) 116–129.
[8] F. Qu, S. Wang, W. Yan, H. Li, S. Wu, X. Fu, Cas(me)^2: a database for spontaneous macro-expression and micro-expression spotting and recognition, IEEE Trans. Affect. Comput. 9 (4) (2018) 424–436.
[9] G. Zhao, M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 915–928.
[10] X. Huang, S. Wang, G. Zhao, M. Pietikäinen, Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection, Proceedings of the 2015 International Conference on Computer Vision Workshop (ICCV) 2015, pp. 1–9.
[11] Y. Liu, J. Zhang, W. Yan, S. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Trans. Affect. Comput. 7 (4) (2016) 299–310.
[12] Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui, G. Zhao, Domain regeneration for cross-database micro-expression recognition, IEEE Trans. Image Process. 27 (5) (2018) 2484–2498.
[13] F. Xu, J. Zhang, J.Z. Wang, Microexpression identification and categorization using a facial dynamics map, IEEE Trans. Affect. Comput. 8 (2) (2017) 254–267.
[14] S. Liong, J. See, K. Wong, R.C. Phan, Less is more: micro-expression recognition from video using apex frame, Signal Process. Image Commun. 62 (2018) 82–92.
[15] S. Liong, Y.S. Gan, J. See, H. Khor, Y. Huang, Shallow triple stream three-dimensional CNN (ststnet) for micro-expression recognition, Proceedings of the 2019 IEEE International Conference on Automatic Face & Gesture Recognition (FG) 2019, pp. 1–5.
[16] Y. Liu, H. Du, L. Zheng, T. Gedeon, A neural micro-expression recognizer, Proceedings of the 2019 International Conference on Automatic Face & Gesture Recognition (FG) 2019, pp. 1–4.
[17] N.V. Quang, J. Chun, T. Tokuyama, Capsulenet for micro-expression recognition, Proceedings of the 2019 IEEE International Conference on Automatic Face & Gesture Recognition (FG) 2019, pp. 1–7.
[18] D.H. Kim, W.J. Baddar, Y.M. Ro, Micro-expression recognition with expression-state constrained spatio-temporal feature representations, Proceedings of the 2016 ACM Conference on Multimedia (ACM MM) 2016, pp. 382–386.
[19] M. Peng, C. Wang, T. Chen, G. Liu, X. Fu, Dual temporal scale convolutional neural network for micro-expression recognition, Front. Psychol. 8 (2017) 1745.
[20] S. Nag, A.K. Bhunia, A. Konwer, P.P. Roy, Facial micro-expression spotting and recognition using time contrasted feature with visual memory, Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP) 2019, pp. 2022–2026.
[21] Z. Xia, X. Hong, X. Gao, X. Feng, G. Zhao, Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions, IEEE Trans. Multimed. 22 (3) (2020) 626–640.
[22] M. Verma, S.K. Vipparthi, G. Singh, S. Murala, Learnet: dynamic imaging network for micro expression recognition, IEEE Trans. Image Process. 29 (2020) 1618–1627.
[23] L. Zhou, Q. Mao, L. Xue, Cross-database micro-expression recognition: a style aggregated and attention transfer approach, Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops, (ICME) 2019, pp. 102–107.

[24] J. See, M.H. Yap, J. Li, X. Hong, S. Wang, MEGC 2019 - the second facial micro-expressions grand challenge, Proceedings of the 2019 International Conference on Automatic Face & Gesture Recognition (FG) 2019, pp. 1–5.
[25] Y. Zong, X. Huang, W. Zheng, Z. Cui, G. Zhao, Learning a target sample re-generator for cross-database micro-expression recognition, Proceedings of the 2017 ACM International Conference on Multimedia (ACM MM) 2017, pp. 872–880.
[26] M.H. Yap, J. See, X. Hong, S. Wang, Facial micro-expressions grand challenge 2018 summary, Proceedings of the 2018 International Conference on Automatic Face & Gesture Recognition (FG) 2018, pp. 675–678.
[27] Y. Zong, W. Zheng, X. Hong, C. Tang, Z. Cui, G. Zhao, Cross-database micro-expression recognition: a benchmark, Proceedings of the 2019 on International Conference on Multimedia Retrieval (ICMR) 2019, pp. 354–363.
[28] M.A. Takalkar, M. Xu, Q. Wu, Z. Chaczko, A survey: facial micro-expression recognition, Multimed. Tools Appl. 77 (15) (2018) 19301–19325.
[29] K.M. Goh, C.H. Ng, L.L. Lim, U.U. Sheikh, Micro-expression recognition: an updated review of current trends, challenges and solutions, Vis. Comput. 36 (3) (2020) 445–468.
[30] J.M. Petr Husak, Jan Cech, Spotting facial micro-expressions?in the wild? Proceedings of the 2017 Computer Vision Winter Workshop, Pattern Recognition and Image Processing Group (PRIP) and PRIP Club, 2017.
[31] A.K. Davison, W. Merghani, M.H. Yap, Objective classes for micro-facial expression recognition, J. Imaging 4 (10) (2018) 119.
[32] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59.
[33] T. Pfister, X. Li, G. Zhao, Recognising spontaneous facial micro-expressions, Proceedings of the 2011 International Conference on Computer Vision (ICCV) 2011, pp. 1449–1456.
[34] X. Huang, G. Zhao, X. Hong, W. Zheng, M. Pietikäinen, Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns, Neurocomputing 175 (2016) 564–578.
[35] Q. Wu, X. Shen, X. Fu, The machine knows what you are hiding: an automatic micro-expression recognition system, Proceedings of the 2011 International Conference on Affective Computing and Intelligent Interaction (ACII) 2011, pp. 152–162.
[36] Y. Wang, J. See, R.C. Phan, Y. Oh, LBP with six intersection points: reducing redundant information in LBP-TOP for micro-expression recognition, Proceedings of the 2014 Asian Conference on Computer Vision (ACCV) 2014, pp. 525–537.
[37] S. Wang, W. Yan, G. Zhao, X. Fu, C. Zhou, Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features, Proceedings of the 2014 European Conference on Computer Vision Workshops (ECCV) 2014, pp. 325–338.
[38] Y. Wang, J. See, R.C.-W. Phan, Y.-H. Oh, Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition, PLoS One 10 (5) (2015) 1–20.
[39] S. Wang, W. Yan, T. Sun, G. Zhao, X. Fu, Sparse tensor canonical correlation analysis for micro-expression recognition, Neurocomputing 214 (2016) 218–232.
[40] Y. Li, X. Huang, G. Zhao, Can micro-expression be recognized based on single apex frame? Proceedings of the 2018 International Conference on Image Processing (ICIP) 2018, pp. 3094–3098.
[41] D.K. Jain, Z. Zhang, K. Huang, Random walk-based feature learning for micro-expression recognition, Pattern Recogn. Lett. 115 (2018) 92–100.
[42] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.
[43] Z. Lu, Z. Luo, H. Zheng, J. Chen, W. Li, A delaunay-based temporal coding model for micro-expression recognition, Proceedings of the 2014 Asian Conference on Computer Vision Workshops (ACCV) 2014, pp. 698–711.
[44] C.A. Duque, O. Alata, R. Emonet, H. Konik, A. Legrand, Mean oriented riesz features for micro expression classification, Pattern Recogn. Lett. 135 (2020) 382–389.
[45] D. Cristinacce, T.F. Cootes, Feature detection and tracking with constrained local models, Proceedings of the 2006 British Machine Vision Conference (BMVC) 2006, pp. 929–938.
[46] W. Yan, S. Wang, Y. Chen, G. Zhao, X. Fu, Quantifying micro-expressions with constraint local model and local binary pattern, in: L. Agapito, M.M. Bronstein, C. Rother (Eds.),Proceedings of the 2014 European Conference on Computer Vision Workshops (ECCVW) 2014, pp. 296–305.
[47] A. Asthana, S. Zafeiriou, S. Cheng, M. Pantic, Robust discriminative response map fitting with constrained local models, Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013, pp. 3444–3451.
[48] S. Wang, B. Li, Y. Liu, W. Yan, X. Ou, X. Huang, F. Xu, X. Fu, Micro-expression recognition with small sample size by transferring long-term convolutional neural network, Neurocomputing 312 (2018) 251–262.
[49] Z. Zhang, P. Luo, C.C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, Proceedings of the 2014 European Conference on Computer Vision (ECCV), Vol. 8694 of Lecture Notes in Computer Science, Springer 2014, pp. 94–108.
[50] K. Li, Y. Zong, B. Song, J. Zhu, J. Shi, W. Zheng, L. Zhao, Three-stream convolutional neural network for micro-expression recognition, Aust. J. Intell. Inf. Process. Syst. 15 (3) (2019) 41–48.
[51] E. Zhou, H. Fan, Z. Cao, Y. Jiang, Q. Yin, Extensive facial landmark localization with coarse-to-fine convolutional network cascade, Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops (ICCV) 2013, pp. 386–391.
[52] A.K. Davison, M.H. Yap, N. Costen, K. Tan, C. Lansley, D. Leightley, Micro-facial movements: an investigation on spatio-temporal descriptors, Proceedings of the 2014 European Conference Computer Vision Workshops (ECCV) 2014, pp. 111–123.
[53] H. Khor, J. See, R.C. Phan, W. Lin, Enriched long-term recurrent convolutional network for facial micro-expression recognition, Proceedings of the 2018 International Conference on Automatic Face & Gesture Recognition (FG) 2018, pp. 667–674.
[54] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Sig. Process. Lett. 23 (10) (2016) 1499–1503.
[55] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, J. Wang, High-resolution representations for labeling pixels and regions, CoRR (2019) abs/1904.04514.
[56] J. Li, Y. Wang, J. See, W. Liu, Micro-expression recognition based on 3d flow convolutional neural network, Pattern. Anal. Applic. 22(4) (2018) 1331–1339.
[57] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, M. Pietikäinen, Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods, IEEE Trans. Affect. Comput. 9 (4) (2018) 563–577.
[58] X. Huang, S. Wang, X. Liu, G. Zhao, X. Feng, M. Pietikäinen, Discriminative spatio-temporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition, IEEE Trans. Affect. Comput. 10 (1) (2019) 32–47.
[59] Z. Zhou, G. Zhao, M. Pietikäinen, Towards a practical lipreading system, Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society 2011, pp. 137–144.
[60] J. Annoni, P. Seiler, M.R. Jovanovic, Sparsity-promoting dynamic mode decomposition for systems with inputs, Proceedings of the 2016 IEEE Conference on Decision and Control (CDC) 2016, pp. 6506–6511.
[61] A.C.L. Ngo, J. See, R.C. Phan, Sparsity in dynamics of spontaneous subtle emotions: analysis and application, IEEE Trans. Affect. Comput. 8 (3) (2017) 396–411.
[62] A.C.L. Ngo, S. Liong, J. See, R.C. Phan, Are subtle expressions too sparse to recognize? Proceedings of the 2015 IEEE International Conference on Digital Signal Processing (DSP) 2015, pp. 1246–1250.
[63] Y. Wang, H. Ma, X. Xing, Z. Pan, Eulerian motion based 3dcnn architecture for facial micro-expression recognition, Proceedings of the International Conference on MultiMedia Modeling (MMM), Vol. 11961 of Lecture Notes in Computer Science, Springer 2020, pp. 266–277.
[64] Z. Xia, H. Liang, X. Hong, X. Feng, Cross-database micro-expression recognition with deep convolutional networks, Proceedings of the 2019 International Conference on Biometric Engineering and Applications (ICBEA) 2019, pp. 56–60.
[65] Z. Xia, W. Peng, H. Khor, X. Feng, G. Zhao, Revealing the invisible with model and data shrinking for composite-database micro-expression recognition, CoRR (2020) abs/2006.09674.
[66] Y. Yu, H. Duan, M. Yu, Spatiotemporal features selection for spontaneous micro-expression recognition, J. Intell. Fuzzy Syst. 35 (4) (2018) 4773–4784.
[67] Y. Wang, J. See, Y. Oh, R.C. Phan, Y. Rahulamathavan, H. Ling, S. Tan, X. Li, Effective recognition of facial micro-expressions with video motion magnification, Multimed. Tools Appl. 76 (20) (2017) 21665–21690.
[68] A.C.L. Ngo, A. Johnston, R.C. Phan, J. See, Micro-expression motion magnification: Global lagrangian vs. local eulerian approaches, Proceedings of the 2018 IEEE International Conference on Automatic Face & Gesture Recognition (FG) 2018, pp. 650–656.
[69] H. Wu, M. Rubinstein, E. Shih, J.V. Guttag, F. Durand, W.T. Freeman, Eulerian video magnification for revealing subtle changes in the world, ACM Trans. Graph. 31 (4) (2012) (65:1–65:8).
[70] J. Yu, C. Zhang, Y. Song, W. Cai, ICE-GAN: identity-aware and capsule-enhanced GAN for micro-expression recognition and synthesis, CoRR (2020) abs/2005.04370.
[71] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial networks, CoRR (2014) abs/1406.2661.
[72] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression, Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops 2010, pp. 94–101.
[73] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. Pietikäinen, Facial expression recognition from near-infrared videos, Image Vis. Comput. 29 (9) (2011) 607–619.
[74] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with gabor wavelets, Proceedings of the 1998 International Conference on Face & Gesture Recognition (FG) 1998, pp. 200–205.
[75] N. Aifanti, C. Papachristou, A. Delopoulos, The MUG facial expression database, Proceedings of the 2010 International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 2010, pp. 1–4.
[76] M. Peng, Z. Wu, Z. Zhang, T. Chen, From macro to micro expression recognition: deep learning on small datasets using transfer learning, Proceedings of the 2018 International Conference on Automatic Face & Gesture Recognition (FG) 2018, pp. 657–661.
[77] S. Wang, W. Yan, X. Li, G. Zhao, X. Fu, Micro-expression recognition using dynamic textures on tensor independent color space, Proceedings of the 2014 International Conference on Pattern Recognition (ICPR) 2014, pp. 4678–4683.
[78] S. Liong, R.C. Phan, J. See, Y. Oh, K. Wong, Optical strain based recognition of subtle emotions, Proceedings of the 2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2014, pp. 180–184.
[79] S. Liong, J. See, R.C. Phan, A.C.L. Ngo, Y. Oh, K. Wong, Subtle expression recognition using optical strain weighted features, Proceedings of the 2014 Asian Conference on Computer Vision (ACCV) 2014, pp. 644–657.
[80] A.C.L. Ngo, R.C. Phan, J. See, Spontaneous subtle expression recognition: imbalanced databases and solutions, Proceedings of the 2014 Asian Conference on Computer Vision (ACCV) 2014, pp. 33–48.

[81] S.Y. Park, S. Lee, Y.M. Ro, Subtle facial expression recognition using adaptive magnification of discriminative facial motion, Proceedings of the 2015 ACM Conference on Multimedia Conference (ACM MM) 2015, pp. 911–914.
[82] A.C. Le Ngo, Y. Oh, R.C. Phan, J. See, Eulerian emotion magnification for subtle expression recognition, Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016, pp. 1243–1247.
[83] S. Liong, K. Wong, Micro-expression recognition using apex frame with phase information, Proceedings of the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) 2017, pp. 534–537.
[84] B. Allaert, I.M. Bilasco, C. Djeraba, Consistent optical flow maps for full and micro facial expression recognition, Proceedings of the 2017 International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP) 2017, pp. 235–242.
[85] J. He, J. Hu, X. Lu, W. Zheng, Multi-task mid-level feature learning for micro-expression recognition, Pattern Recogn. 66 (2017) 44–52.
[86] X. Huang, G. Zhao, Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern, Proceedings of the 2017 International Conference on Frontiers and Advances in Data Science (FADS) 2017, pp. 159–164.
[87] Y. Zong, X. Huang, W. Zheng, Z. Cui, G. Zhao, Learning from hierarchical spatiotemporal descriptors for micro-expression recognition, IEEE Trans. Multimed. 20 (11) (2018) 3160–3172.
[88] H. Lu, K. Kpalma, J. Ronsin, Motion descriptors for micro-expression recognition, Signal Process. Image Commun. 67 (2018) 108–117.
[89] L. Lo, H. Xie, H. Shuai, W. Cheng, MER-GCN: micro expression recognition based on relation modeling with graph convolutional network, CoRR (2020) abs/2004.08915.
[90] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the 2015 International Conference on Computer Vision and Pattern Recognition (CVPR) 2015, pp. 1–9.
[91] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, Proceedings of the 2016 European Conference Computer Vision (ECCV) 2016, pp. 630–645.
[92] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M.S. Bernstein, A.C. Berg, F. Li, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[93] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[94] S. Wang, W. Yan, X. Li, G. Zhao, C. Zhou, X. Fu, M. Yang, J. Tao, Micro-expression recognition using color spaces, IEEE Trans. Image Process. 24 (12) (2015) 6034–6047.
[95] S. Polikovsky, Y. Kameda, Y. Ohta, Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor, Proceedings of the 2009 International Conference on Imaging for Crime Detection and Prevention (ICDP) 2009, pp. 1–6.
[96] S. Polikovsky, Y. Kameda, Y. Ohta, Facial micro-expression detection in hi-speed video based on facial action coding system (FACS), IEICE Trans. Inf. Syst. 96-D (1) (2013) 81–92.
[97] S.P.T. Reddy, S.T. Karri, S.R. Dubey, S. Mukherjee, Spontaneous facial micro-expression recognition using 3d spatiotemporal convolutional neural networks, CoRR (2019) abs/1904.01390.
[98] M. Verma, S.K. Vipparthi, G. Singh, Non-linearities improve originet based on active imaging for micro expression recognition, CoRR (2020) abs/2005.07991.
[99] Y.S. Gan, S. Liong, W. Yau, Y. Huang, T.L. Ken, Off-apexnet on micro-expression recognition system, Signal Process. Image Commun. 74 (2019) 129–139.
[100] S. Liong, Y.S. Gan, J. See, H. Khor, A shallow triple stream three-dimensional CNN (ststnet) for micro-expression recognition system, CoRR (2019) abs/1902.03634.
[101] L. Zhou, Q. Mao, L. Xue, Dual-inception network for cross-database micro-expression recognition, Proceedings of the 2019 IEEE International Conference on Automatic Face & Gesture Recognition (FG) 2019, pp. 1–5.
[102] J. Päivärinta, E. Rahtu, J. Heikkilä, Volume local phase quantization for blur-insensitive dynamic texture classification, Proceedings of 2011 Scandinavian Conference on Image Analysis (SCIA) 2011, pp. 360–369.
[103] D. Tran, L.D. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3d convolutional networks, Proceedings of the 2015 International Conference on Computer Vision (ICCV) 2015, pp. 4489–4497.
[104] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F. Li, Large-scale video classification with convolutional neural networks, Proceedings of the 2014 International Conference on Computer Vision and Pattern Recognition (CVPR) 2014, pp. 1725–1732.
[105] K. Soomro, A.R. Zamir, M. Shah, UCF101: A dataset of 101 human actions classes from videos in the wild, CoRR (2012) abs/1212.0402.
[106] M. Simon, E. Rodner, J. Denzler, Imagenet pre-trained models with batch normalization, CoRR (2016) abs/1612.01452.
[107] C. Zach, T. Pock, H. Bischof, A duality based approach for realtime tv-L1 optical flow, Proceedings of the 2007 Pattern Recognition, DAGM Symposium 2007, pp. 214–223.
[108] T. Senst, V. Eiselein, T. Sikora, Robust local optical flow for feature tracking, IEEE Trans. Circuits Syst. Video Techn. 22 (9) (2012) 1377–1387.
[109] G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in: J. Bigün, T. Gustavsson (Eds.),Proceedings of the 2003 Scandinavian Conference Image Analysis (SCIA), Vol. 2749 of Lecture Notes in Computer Science 2003, pp. 363–370.
[110] J.L. Barron, D.J. Fleet, S.S. Beauchemin, T.A. Burkitt, Performance of optical flow techniques, Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1992, pp. 236–242.
[111] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, Flownet 2.0: evolution of optical flow estimation with deep networks, Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017, pp. 1647–1655.
[112] D. Sun, X. Yang, M. Liu, J. Kautz, Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume, Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018, pp. 8934–8943.
[113] L. Fan, W. Huang, C. Gan, S. Ermon, B. Gong, J. Huang, End-to-end learning of motion representation for video understanding, Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018, pp. 6016–6025.
[114] A.J. Piergiovanni, M.S. Ryoo, Representation flow for action recognition, Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2019, pp. 9945–9953.