Review

Review of Automatic Microexpression Recognition in the Past Decade

Liangfei Zhang * and Ognjen Arandjelović

School of Computer Science, University of St Andrews, St Andrews KY16 9SX, UK; oa7@st-andrews.ac.uk
* Correspondence: lz36@st-andrews.ac.uk

Abstract: Facial expressions provide important information concerning one's emotional state. Unlike regular facial expressions, microexpressions are particular kinds of small quick facial movements, which generally last only 0.05 to 0.2 s. They reflect individuals' subjective emotions and real psychological states more accurately than regular expressions which can be acted. However, the small range and short duration of facial movements when microexpressions happen make them challenging to recognize both by humans and machines alike. In the past decade, automatic microexpression recognition has attracted the attention of researchers in psychology, computer science, and security, amongst others. In addition, a number of specialized microexpression databases have been collected and made publicly available. The purpose of this article is to provide a comprehensive overview of the current state of the art automatic facial microexpression recognition work. To be specific, the features and learning methods used in automatic microexpression recognition, the existing microexpression data sets, the major outstanding challenges, and possible future development directions are all discussed.

Keywords: affective computing; microexpression recognition; emotion recognition; microexpression database; video feature extraction; deep learning

1. Introduction

Facial expressions are important for interpersonal communication [1], in no small part because they are key in understanding people's mental state and emotions [2]. Different from 'conventional' facial expressions, which can be consciously controlled, microexpressions are effected by short-lasting, unconscious contraction of facial muscles under psychological inhibition, see Figure 1. As such they can be used as a means of inferring a person's emotions even if there is an attempt to conceal them.


Figure 1. Microexpressions (**top**) and macroexpressions (**bottom**) are qualitatively identical but the former are involuntary and much shorter in expression than the latter.

The concept of a microexpression was first introduced by Haggard et al. [3] in 1966. Following on this work, Ekman et al. [4] reported a case study on the topic, thus providing the first evidence in support of the idea. If the occurrence of microexpressions is detected, and the corresponding emotional associations understood, the true sentiments of individuals could be accurately identified even when there is an attempt at concealment [5], thus improving lie detection rates. For example, during psychological diagnostic testing, when a patient presents a microexpression of joy, depending on the context, it may mean that they are successful in passing the test. A patient's microexpression of fear may indicate a fear of betraying something they wish to keep secret. When the patient exhibits a microexpression of surprise, it may indicate that they have not considered the relevant question or do not understand something. Therefore, microexpressions can help us understand the true emotions of individuals and provide important clues for lie detection. In addition, microexpressions have high reliability and potential value in emotion-related tasks, such as communication negotiation [6] and teaching evaluation [7].

In recent years, research on microexpressions has been attracting an increasing amount of attention in the scholastic community. As illustrated by the plot in Figure 2, following the publication of two open-source microexpression databases in 2013, the number of published articles pertaining to microexpressions has increased every year. Since 2018, the Micro-Expression Grand Challenge (MEGC) workshop has been organized as part of the IEEE International Conference on Automatic Face and Gesture Recognition, contributing to the popularization of the topic in the computer vision and machine learning community.
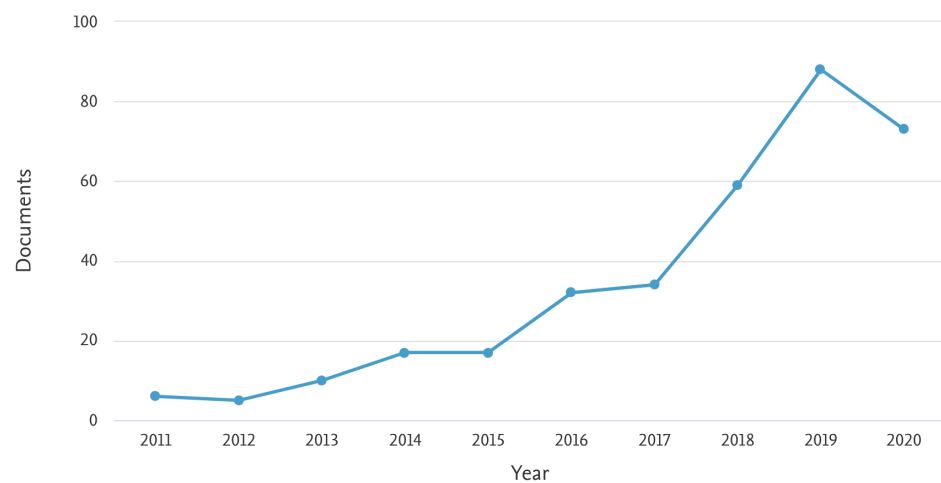
Figure 2. The number of microexpression recognition publications from 2011 to 2020 (Data Source: Scopus).

Within the realm of microexpression research, there are several related but nonetheless distinct research directions which have emerged over the years. These include the differentiation of macro- and microexpressions, the identification of specific microexpressions over a period of observed facial movement (referred to as microexpression detection or spotting), and the inference of emotions revealed by microexpressions. The last of these is the most commonly addressed challenge and is often referred to as Micro-Expression Recognition (MER). The task is to recognize emotions expressed in a sequence of faces known to be microexpressions. In recent years, many researchers have begun to use computer vision technology for automatic MER, which significantly improves the feasibility of microexpression applications. The use of computer technology for MER has unique advantages. Even the very fastest facial movements can be captured by high-speed cameras and processed by computers. In addition, when an efficient and stable model can be trained, computers

are able to process large scale MER tasks with low cost, greatly exceeding the efficiency of manual recognition of microexpressions by professionals.

The present article reviews the fundamentals behind facial MER, summarizes the key techniques and data sets used in related research, discusses the most prominent outstanding problems in the field, and lays out possible future research directions. There are several aspects of our work which set it apart from the existing reviews related to microexpressions [8–11]. Firstly, herein we specifically focus on microexpression recognition, rather than e.g., spotting. Thus we provide a broader overview of methods in this specific realm, from those involving 'classical' manually engineered features to the newly emerging deep learning based approaches. We also present the most comprehensive and up-to-date review of relevant data sets available to researchers and an in-depth discussion of evaluation approaches and data acquisition protocols. Lastly, we offer a new and different take on open research questions.

## 2. Features Used in Microexpression Recognition

In recent years, research on MER has increased considerably, leading to the development of a variety of different, specialized features. Popular examples of such features include 3D Histograms of Oriented Gradients (3DHOG) as the simplest extension of the 'traditional' HOG features, subsequently succeeded by more nuanced extensions such as Local Binary Pattern-Three Orthogonal Planes (LBP-TOP), Histograms of Oriented Optical Flow (HOOF), and their variations. Since 2016, the application of deep learning in MER has been increasing and it can be expected to continue to proliferate, thus becoming the main methodology in MER research in the future.

2.1. 3D Histograms of Oriented Gradients (3DHOG)

Polikovsky et al. [12] proposed the use of a 3D gradient feature to describe local spatiotemporal dynamics of the face, see Figure 3. Following the segmentation of a face into 12 regions according to the Facial Action Coding System (FACS) [13], each region corresponding to an independent facial muscle complex, and the appearance normalization of individual regions, Polikovsky et al. obtain 12 separate spatiotemporal blocks. The magnitudes of gradient projections along each of the three canonical directions are then used to construct histograms across different regions, which are used as features. The authors assume that each frame of the microexpression image sequence involves only one action unit (AU), which represents one specific activated facial muscle complex in FACS, and this unit can be used as an annotation of the image. The k-means algorithm is used for clustering in the gradient histogram feature space in all frames of microexpression image sequences, and the number of clusters is set to the number of action units that have appeared in all ME samples. The action unit corresponding to the greatest number of features is regarded as the real label of each cluster.

The feature extraction method of this work is relatively simple and is an extension of the plane gradient histogram. The model construction adopts a more complicated process, which can be regarded as a k-nearest neighbour model constructed by the k-means algorithm. It is robust to the correctness of the labels and insensitive to a small number of false annotations.

The main limitation of this work lies in the aforementioned assumption that only a single action unit is active in each frame, which is overly restrictive in practice.
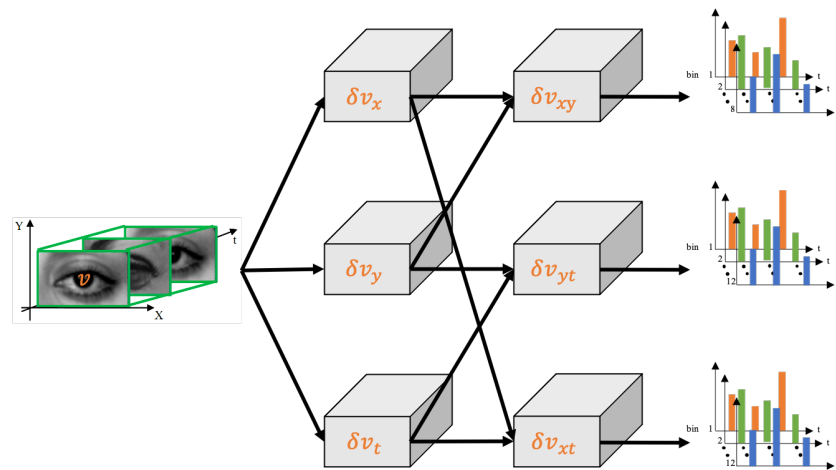
Figure 3. Conceptual summary of the descriptor extraction process for a facial cube using 3DHOG.

2.2. Local Binary Pattern-Three Orthogonal Planes (LBP-TOP)

A local binary pattern (LBP) is a descriptor originally proposed to describe local appearance in an image. The key idea behind it is that the relative brightness of neighbouring pixels can be used to describe local appearance in a geometrically and photometrically robust manner [14–16]. The basic LBP feature extractor relies on two free parameters, call them R and P. Uniformly sampling P points on the circumference of a circle with the radius R centred at a pixel, and taking their brightness relative to the centre pixel (brighter than, or not—one bit of information) allows the neighbourhood to be characterized by a P-bit number.

In recognition of microexpressions, in order to encode the spatiotemporal co-occurrence pattern, the LBP-TOP (Local Binary Pattern on Three Orthogonal Plane) [17] is used to extract the LBP features separately for the XY, XT, and YT planes in image sequences; see Figure 4. Neighbourhood sampling is now performed over a circle in the purely spatial plane and over ellipses in the spatiotemporal planes.
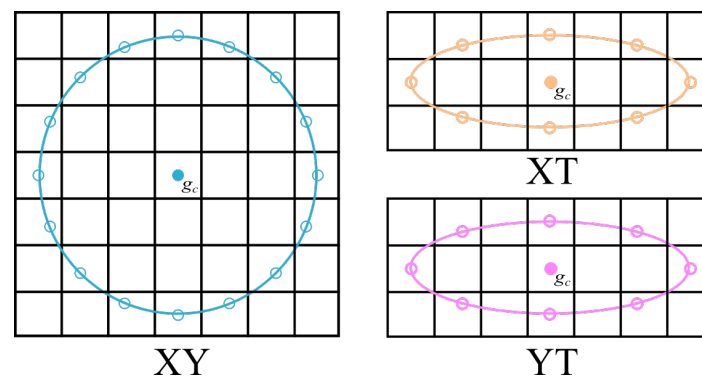
$$LBP(M,R) = \sum_{m=0}^{M-1} s(g_m - g_c)2^m$$

Figure 4. Detailed sampling for LBP-TOP with $RX = RY = R = 3$, $RT = R = 1$, $PXY = M = 16$, and $PXT = PYT = M = 8$.

Pfister et al. [18] made one of the earliest attempts to recognize microexpressions automatically. Their method, in which LBP-TOP is used for the feature extraction, has been highly influential in the field and much follow-up work drew inspiration from it; see Figure 5. Pfister et al. first use a 68-point Active Shape Model (ASM) [19] to locate the key points of the face. Based on the key points obtained, the deformation relationship between the first facial frame of each sequence and the model facial frame is calculated using the Local Weighted Mean (LWM) [20] method. A geometric transformation is then

applied to each frame of the sequence so as to normalize for small pose variation and coarse expression changes. In order to account for differences in the number of frames between different input sequences, Temporal Interpolation Model (TIM) is used to temporally interpolate between frames, thus normalizing sequence length to a specific count. LBP-TOP features are extracted from thus normalized sequences. Finally, Support Vector Machine (SVM), Random Forest (RF), and Multiple Kernel Learning (MKL) methods are used for classification. Wang et al. [21] expressed the microexpression sequence and its LBP features by tensor and performed a sparse tensor canonical correlation analysis on the tensor to learn the relationship between the microexpression sequence itself and its LBP features. The simple nearest neighbour algorithm is used for classification. In experiments, the authors demonstrate the superiority of their approach over the original LBP-TOP method of Pfister et al.
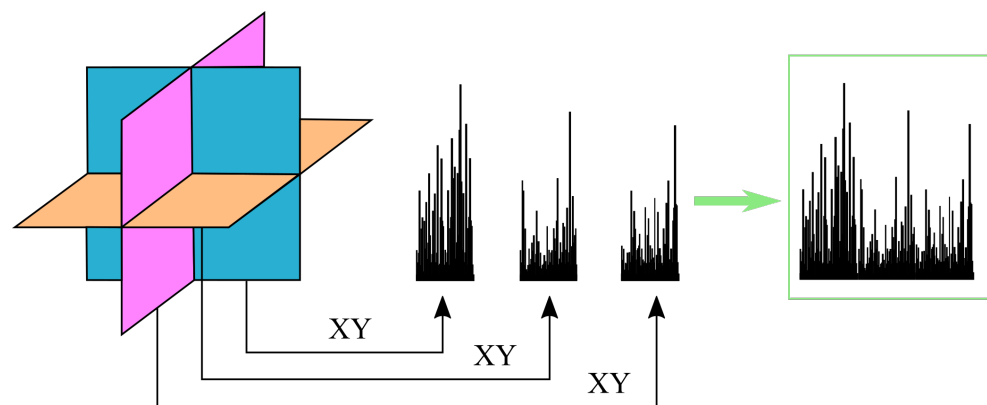
Figure 5. Conceptual illustration of the Histogram of Concatenated Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) feature.

Local Binary Pattern with Six Intersection Points (LBP-SIP) [22] extends LBP features for MER in a different manner. The main improvement of the work of Wang et al. is to reduce the feature dimension to improve feature extraction. Compared with LBP-TOP, it reduces information redundancy, thus providing a more compact representation. Experimental evidence suggests that its extraction is nearly three times faster than that of LBP-TOP. Specifically, in the same experimental environment, the average LBP-TOP extraction time of the CASME II (Chinese Academy of Sciences Micro-Expression II) database is 18.289 s, and the LBP-SIP extraction time is 15.888 s. Furthermore, in the context of the use of the descriptors for recognition, the LBP-TOP based microexpression recognition takes 0.584 s per sequence, in contrast to only 0.208 for LBP-SIP based.

The Centralized Binary Pattern (CBP) [23] descriptor is another variation on the conceptual theme set out by LBP. In broad terms, it is computed in a similar way to LBP. However, unlike in the case of LBP, CBP compares the central pixel of an area with a pair of neighbours, see Figure 6. Therefore, the corresponding binary code length is about half of that of LBP, with a lower dimensionality of the corresponding histogram. Indeed, the key advantage of CBP compared to LBP is that it produces lower dimensionality features. Hence, Guo et al. [24] employ the CBP-TOP operator in place of LBP-TOP, with an Extreme Learning Machine (ELM) for classification, and experimentally demonstrate that performance improvement is indeed effected by their approach.

$$CBP(M,R) = \sum_{m=0}^{(M/2)-1} s\left(g_m - g_{m+(M/2)}\right) 2^m + s\left(g_c - \frac{1}{M+1}\left(\sum_{m=0}^{M-1} g_m + g_c\right)\right) 2^{M/2}$$
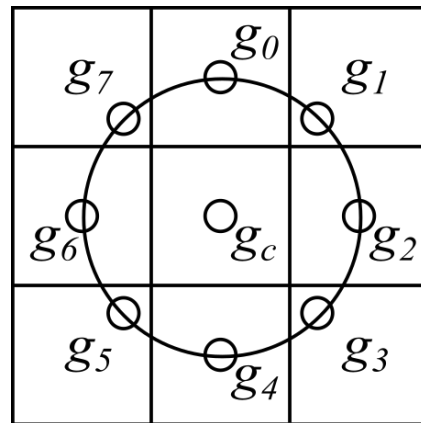
Figure 6. Detailed sampling for CBP operator with $RX = RY = R = 1$, $PXY = M = 8$.

In addition to standard texture features, some researchers have also considered the use of colour on micromovement extraction (colour has indeed been shown to be important in face analysis more generally [25]). If the usual RGB space that the original face image data is represented in, is adopted for the extraction of the aforementioned local appearance features (such as the commonly used LBP-TOP), the three channels result in redundant information, failing to effect improvement over greyscale. Hence, Wang et al. [26] considered this problem and instead proposed the use of Tensor Independent Colour Space (TICS). In another work [27], the researchers tried to use CIELab and CIELuv colour spaces, which have already demonstrated success in applications needing human skin detection [28]. Their experiments showed that the transformation of colour space can effect an improvement in recognition.

2.3. Histogram of Oriented Optical Flow (HOOF)

One of the influential works which does not follow the common theme of using LBP-like local features is that of Liu et al. [29] which uses a different local measure, namely optical flow. The authors extract the main motion direction in the video sequence and calculate the average optical flow characteristics in the partial facial blocks. Hence, they introduce the Main Directional Mean Optical flow feature (MDMO). Firstly, the face key point of each frame is located by using the Discriminative Response Map Fitting (DRMF) model [30]. Then the optical flow field of each frame relative to the succeeding frame is used to find an affine transformation matrix which corrects for pose change. The transformation matrix makes the difference of facial landmarks in each frame from the first frame minimal. The authors then calculate the average of the most similar motion vectors of the optical flow field in each region as the motion characteristic of the region. Specifically, they calculate the HOOF (histogram of oriented optical flow) feature [31] in each region and quantize all optical flow direction vectors to eight intervals to obtain a histogram of the aforementioned directions. The resulting histogram features are finally fed into a support vector machine, trained to classify microexpressions.

Following in spirit but unlike the work of Liu et al. [29], Xu et al. [32] used the optical flow field as the key low level feature to describe the pattern of microexpression movement using the facial dynamics map (FDM); see Figure 7. The FDM better reflects intricate local motion patterns characteristic of microexpressions and has the appeal of being beneficial in interpretability by virtue of its useful visualization. Nevertheless, the uniform and indeed major disadvantage of HOOF methods lies in their high computational cost, which makes them unsuitable for real-time, large-scale MER.

| Frames | FDM (Optical Flow Estimation) | Bins | Histogram |
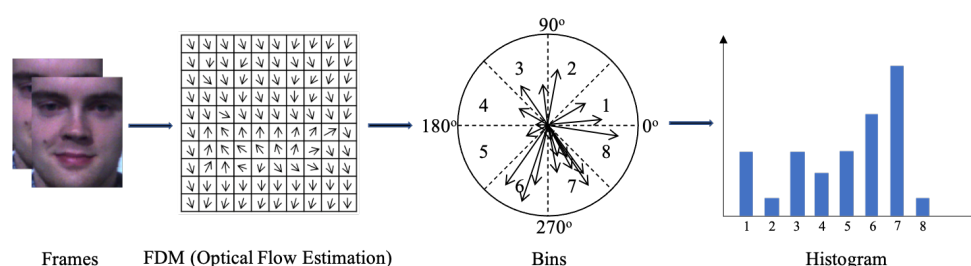
Figure 7. One-dimensional histogram of FDM from optical flow estimation.

2.4. Deep Learning

Although deep learning techniques and deep neural networks are widely used in other face related recognition tasks, they are still novel in the field of MER research. As shown in Figure 8, deep learning in the real of microexpression analysis started around 2016. However, the annual number of publications shows an exponential increase in the following years.
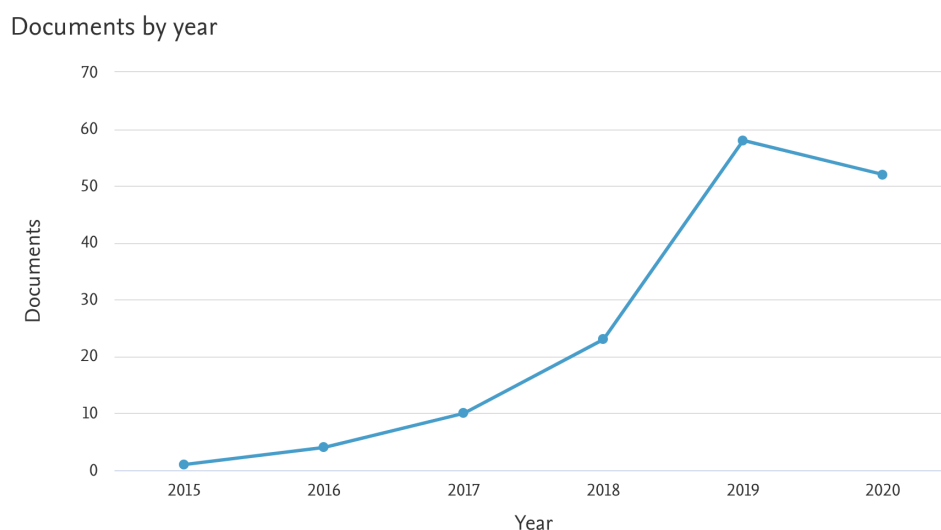
Figure 8. The number of microexpression recognition publications using deep learning in the period 2016–2020 (source: Scopus).

Kim et al. [33] use deep learning and introduce a feature representation based on expression states—Convolutional Neural Networks (CNN) are employed for encoding different expression states (start, start to apex, apex, apex to end, and end). Several objective functions are optimised during spatial learning to improve expression class separability. The encoded features are then processed by a Long Short-Term Memory (LSTM) network to learn features related to time scales. Interestingly, their approach failed to demonstrate an improvement over more old-fashioned, hand-crafted feature based methods, merely performing on par with them. While these results need to be taken with a degree of caution due to the limited scale of empirical testing and low data diversity (a single data set, CASME II, discussed shortly, was used), they suggest that the opportunity for innovation in the sphere of deep learning in the context of microexpressions is wide open.

Peng et al. [34] also adopt a deep learning paradigm, while making use of ideas previously shown to be successful in the realm of conventional methods, by using a sequence of optical flow data as input. To overcome the limitation imposed by the availability of training data, their Dual Time Scale Convolutional Neural Network (DTSCNN) comprises a shallow neural network for MER and only four layers for the convolutional and pooling stages. On a data set formed by merging CASME and CASME II, using four different mi-

croexpression classes—namely negative, positive, surprise, and other—DTSCNN achieved higher accuracy than the competing methods: STCLQP [35], MDMO [29], and FDM [32].

Khor et al. [36] proposed an Enhanced Long-term Recursive Convolutional Network (ELRCN) for microexpression recognition, which uses the architecture previously described by Donahue et al. [37] to characterize small facial changes. The ELRCN model includes a deep spatial feature extractor and a time extractor. These two variants of network are enriching the spatial dimension by input channel superposition and the time dimension by depth feature superposition. Experimental evidence suggests that spatial and time modules play different roles within this framework and that they are highly interdependent in effecting accurate performance. The experiments were performed on the usual data sets, with the appealing modification that training and test were performed on data sets with different provenances. Namely, while training training was done on CASME II, testing was performed on SAMM.

2.5. Closing Remarks

To summarize this section, in the realm of conventional computer vision approaches to microexpression recognition and analysis, there is a broad similarity between different approaches described in the literature, all of them being based on appearance based local (in time or space) features. In general, simple spatial LBP-TOP features (and similar variations) perform better than spatiotemporal 3DHOG and HOOF, when high-resolution images are used. However, when image resolution is low, the reverse is observed. This observation is consistent with what one might expect from theory. Namely, since LBP-TOP features rely on local spatial information, the loss of spatial information effected by lowering resolution negatively affects their performance. In contrast, HOOF and 3DHOG also strongly depend on temporal variation. Thus, interframe information is less affected, though not unaffected, by changes in image resolution.

Contrasting conventional computer vision approaches are emergent deep learning methods. Though a number of different microexpression recognition algorithms based on deep learning have now been described in the literature, the performance of this umbrella of methods is yet to demonstrate its value in this field.

Finally, for completeness, we include a detailed summary of a comprehensive list of different conventional and deep learning approaches in Table A1, including many minor variations on the themes directly surveyed in this section and which do not offer sufficient novelty to warrant being discussed in detail.

3. Microexpression Databases

A consideration of data used to assess different solutions put forward in the literature is of major importance in every subfield of modern computer vision. Arguably, considering the relative youth of the field, this consideration is particularly important in the realm of microexpression recognition. Standardization of data is crucial in facilitating fair comparison of methods, and its breadth and quality key to understanding how well different methods work, what limitations they have, and what direction new research should follow.

Some of the most widely used microexpression related databases include USF-HD [38], Polikovsky Data-set [12], York Deception Detection Test (YorkDDT) [39], Chinese Academy of Sciences Micro-Expressions (CASME) [40], Spontaneous Micro-Expression Corpus (SMIC) [41], Chinese Academy of Sciences Micro-Expression II (CASME II) [42], Spontaneous Actions and Micro-Movements (SAMM) [43], and Chinese Academy of Sciences Spontaneous Macro-Expressions and Micro-Expressions (CAS(ME)^2) [44]. The nature and purpose of these data sets varies substantially, in some cases subtly, in others less so. In particular, the first three databases are older and proprietary and contain video sequences with nonspontaneous microexpression exhibition. The USF-HD is used to evaluate methods which aim to distinguish between macroexpressions and microexpressions. Different yet, the Polikovsky data set was collected for assessing keyframe detection in the context of microexpressions, whereas the York DDT is specifically aimed at lie detection.

For the acquisition of data for nonspontaneous databases, participants are required to watch the video or image data of the microexpressions and try to imitate them. Therefore, this data should be used with due caution and not assumed to represent the strict ground truth. Therefore, only open-source spontaneous microexpression databases will be discussed here. These exhibit significant differences between them, and their particularities are important to appreciate so that the findings in the current literature can be interpreted properly and future experiments can be designed appropriately.

3.1. Open-Source Spontaneous Microexpression Databases

Recall that the duration of a microexpression is usually only 1/25 to 1/5 of a second. In contrast, the frame rate of a regular camera is 25 frames per second. Therefore, if conventional imaging equipment is used, only a small number of frames capturing a microexpression is obtained, which makes any subsequent analysis difficult. Nevertheless, considering the ubiquity of such standardized imaging equipment, some data sets such as SMIC-VIS and SMIC-NIR (see Section 3.1.2), do contain sequences with precisely this frame rate. On the other hand, in order to facilitate more accurate and nuanced microexpression analysis, most microexpression data sets in widespread use in the existing academic literature use high-speed cameras for image acquisition. For example, SMIC uses a 100 fps camera and CASME uses a 60 fps one (see Sections 3.1.2 and 3.1.1 respectively), in order to gather more temporally fine grained information. The highest frame rate in the existing literature is the SAMM and CASME II (see Sections 3.1.4 and 3.1.3 respectively), which both use a high-speed camera at the rate of 200 frames per second.

3.1.1. CASME

The Chinese Academy of Sciences Micro-Expressions (CASME) [40] data set contains 195 sequences of spontaneously exhibited microexpressions. The database is divided into two parts, referred to as Part A and Part B. The resolution of images in Part A is 640×480 pixels, and they were acquired indoors, with two obliquely positioned LED lights used to illuminate faces. Part B images have the resolution of 1280×720 pixels and were acquired under natural light. Microexpressions in CASME are categorized as expressing one of the following: amusement, sadness, disgust, surprise, contempt, fear, repression, or tension; see Figure 9. Considering that some emotions are more difficult to excite than others in a laboratory setting, the number of examples across the aforementioned classes is unevenly distributed.

| Happiness | Sadness | Disgust | Surprise | Tense | Repression |

Figure 9. Examples of frames from sequences in the Chinese Academy of Sciences Micro-Expressions (CASME) data set [40].

3.1.2. SMIC

The Spontaneous Micro-Expression Corpus (SMIC) [41] contains videos of 20 participants, exhibiting 164 spontaneously produced microexpressions. What most prominently distinguishes SMIC from other microexpression data sets is the inclusion of multiple imaging modalities. The first part of the data set contains videos acquired in the visible spectrum using a 100-fps high-speed (HS) camera. The second part also contains videos acquired in the visible spectrum but at the lower frame rate of 25 fps. Lastly, videos in the near-infrared (NIR) spectrum are included (n.b., only of 10 out of 16 individuals in the database).

Hence, sometimes reference is made not to SMIC as a whole but to its constituents, namely SMIC-HS, SMIC-VIS, and SMIC-NIR respectively; see Figure 10.

Figure 10. Examples of frames from sequences in the three subsets of the Spontaneous Micro-expression Corpus (SMIC), namely SMIC-HS, SMIC-VIS, and SMIC-NIR [41].

3.1.3. CASME II

The Chinese Academy of Sciences Micro-Expression II (CASME II) [42] data set is a large collection of spontaneously produced microexpressions, containing 247 video sequences of 26 Asian participants with the average age of approximately 22 years; see Figure 11. The data was captured under uniform illumination, without a strobe. In contrast to CASME, the emotional category labels in CASME II are much broader—namely, happiness, sadness, disgust, surprise, and 'others'—thus making the trade-off between class representation and balance, and emotional nuance, in the opposite direction.
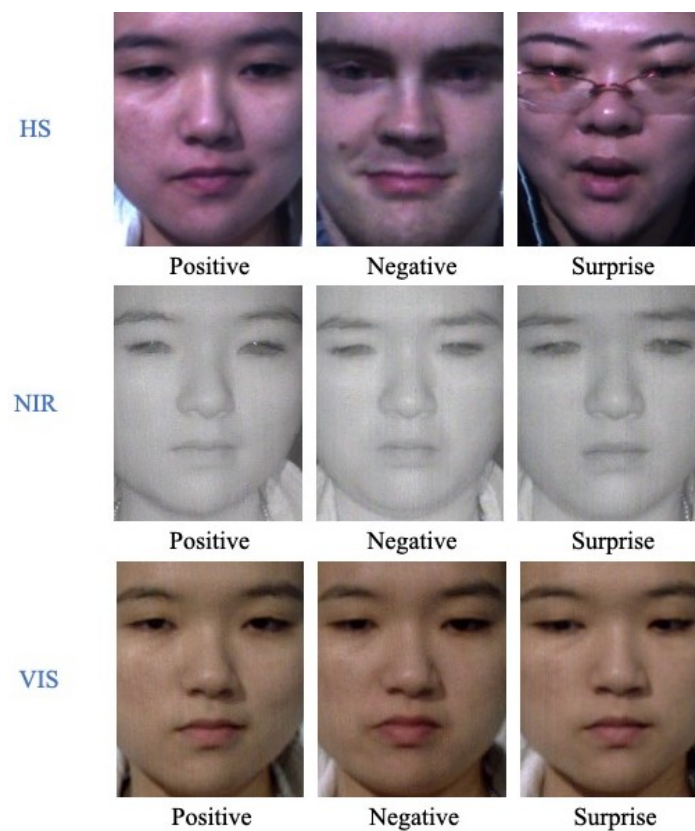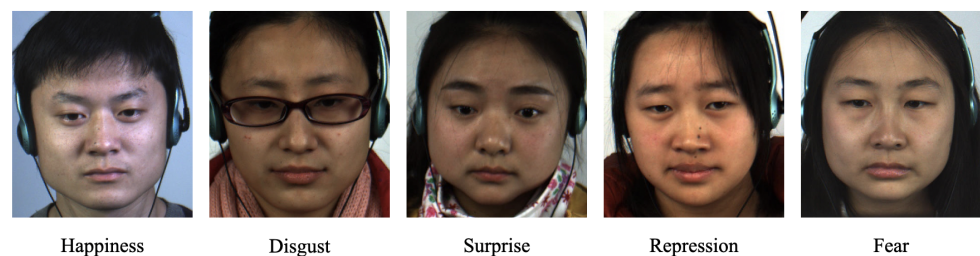

Figure 11. Examples of frames from sequences in the Chinese Academy of Sciences Micro-Expression II (CASME II) data set [42].

3.1.4. SAMM

The Spontaneous Actions and Micro-Movement (SAMM) [43] data set is the newest addition to the choice of microexpression related databases freely available to researchers;

see Figure 12. It contains 159 microexpressions, spontaneously produced in response to visual stimulus, of 32 gender balanced participants with the average age of approximately 33 years. Being the most recently acquired data set, in addition to the standard categorized imagery, SAMM contains a series of annotations which have emerged as being of potential use from previous research. In particular, associated with each video sequence are the indexes of the frame when the relevant microexpression starts and ends and the index of the so-called vertex frame (frame when the greatest temporal change in appearance is observed). In addition to being categorized as expressing contempt, disgust, fear, anger, sadness, happiness, or surprise, each video sequence in the data set also contains a list of FACS action units (AU) engaged during the expression.


Figure 12. Examples of images from the SAMM data set [43].

3.1.5. CAS(ME)^2

Like several other corpora described previously, the Chinese Academy of Sciences Spontaneous Macro-Expressions and Micro-Expressions (CAS(ME)^2) [44] data set is also heterogeneous in nature. The first part of this corpus, referred to as Part A, contains 87 long videos, which contain both macroexpressions and microexpressions. The second part of CAS(ME)^2, Part B, contains 303 separate short videos, each lasting only for the duration that an expression (be it a macroexpression or a microexpression) is exhibited. The numbers of macroexpression and microexpression samples are 250 and 53 respectively. In all cases, in comparison with most other data sets, the expressions are rather coarsely classified as positive, negative, surprised, or 'other'.

3.2. Data Collection and Methods for Systematic Microexpression Evocation

One difficulty in the process of collecting microexpression video sequence corpora lies in the difficulty of inciting microexpressions in a reliable and uniform manner. A common approach adopted in the published literature consists of presenting participants with emotional content (usually short clips or movies) which is expected to rouse their emotions, while at the same time asking them to disguise their emotions and maintain a neutral facial expression. A typical data acquisition setup is diagrammatically shown in Figure 13.
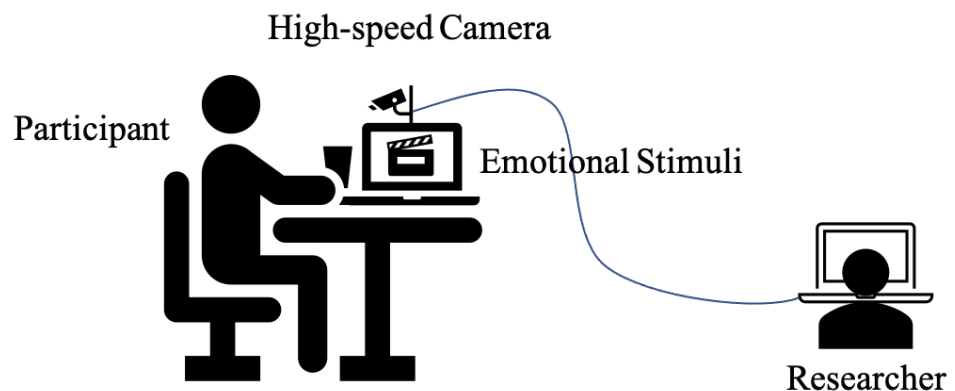
Figure 13. For the purposes of data collection, participants watch an emotional video while their faces are imaged by high-speed camera.

When the aforementioned data collection protocol is considered with some care, it is straightforward to see that a number of practical problems present themselves. Firstly, in some instances, the assumption that the content presented to the participants will elicit sufficient emotion may be invalidated. Thus, no meaningful microexpression may be present in a video sequence of a person's face (e.g., in SMIC out of 20 individuals who participated in the recording sessions, only 16 exhibited sufficiently well expressed microexpressions). This problem can be partially ameliorated by ensuring that the stimuli are strong enough, though this must be done with due consideration of possible ethical issues. On the complementary side, so to speak, considering that microexpressions are involuntarily expressed, it is important to suppress as much as possible any conscious confound. In other words, there must exist sufficient incentive to encourage participants to conceal their true feelings.

3.2.1. CAS Data Acquisition Protocol

During data collection of CASME [40], CASME II [42], and CAS(ME)^2 [44], participants were asked to watch different emotional videos while maintaining a neutral facial expression. As explained before, the intention is to incite involuntary microexpressions, rather than have them acted, which results in data which is not realistic. During the collection process, the participants were required to remain expressionless and not move their body, thus removing any need for body or head pose normalization. Lastly, as a means of encouraging participants to conceal their emotions, they were offered the potential of a monetary award. Specifically, the award was paid out if a participant successfully managed to hide their emotion from the researcher supervising the process (the researcher was unaware of the video content).

3.2.2. SMIC Data Acquisition Protocol

Much like in the case of three CAS data sets, in the process of data collecting for SMIC [41], the participants were shown emotional videos and asked to attempt to conceal their reactions, and a researcher, unaware of the video content watched, was asked to guess the participants' emotions. Unlike for CASME [40], CASME II [42], and CAS(ME)^2 [44] when participants were incentivized by a reward for success (in hiding their emotions), now participants were disincentivized by a 'punishment'—namely, unsuccessful participants had to fill in a lengthy questionnaire.

3.2.3. SAMM Data Acquisition Protocol

Highlighting the point we made previously—the need to understand well the nuanced differences between different microexpression data sets—the data acquisition protocol employed in collecting the SAMM data set is different still from all of the described thus

far. Firstly, all participants were asked to fill out a questionnaire before the actual imaging session. The aim of this was to allow the researchers to personalize emotional stimuli (e.g., exploiting specific individual's fears, likes and dislikes, etc.). Additionally, and again in contrast with e.g., CAS data sets, in order to make the participants more relaxed and less affected by their knowledge that they are partaking in an experiment, the participants were filmed without any supervision of or oversight by the researchers.

3.3. Publicly Available Current Micro-Expression Data Sets—A Recap

At present, the amount of microexpression databases is minimal, and especially there are only five spontaneous microexpression databases. The number of microexpression samples contained in each database is only about 200. When the currently available spontaneous microexpression databases are considered, each of them can be seen to offer some kind of advantage over the others; nevertheless, the amount of data in any of them does not meet the requirements of the traditional deep learning algorithms. SAMM and CASME II have the highest frame rate of 200 fps, and SAMM has the highest resolution. The SMIC database contains both high speed camera samples as well as samples suitable as training data for a model used in a typical non-high-speed camera environment. CAS(ME)^2 contains not only the FACS information and emotion labels associated with individual microexpressions but also can be used to distinguish between macro- and microexpressions. The specific comparison of each database is shown in the table above. Therefore, combining different databases in an experiment may be an approach method for the training of MER models at present.

4. Outstanding Challenges and Future Work

The study of microexpression is still in its early stages, rather than a mature research field. Thus, unsurprisingly, there remain many challenges. This section briefly outlines the current challenges and possible promising future research directions.

4.1. Action Unit Detection

Action unit detection is an essential subtask in conventional facial expression recognition, that is, in macroexpression recognition. Considering the qualitative equivalence of micro and macroexpressions, the use of information on action unit engagement for the analysis of the former naturally imposes itself. However, the quantitative difference between the two types of expressions, that is the lesser extent of activation of action unit during microexpressions, makes action unit activation detection much more difficult in the case of microexpressions. Yet, research to date strongly suggests that information on action units' involvement in an expression is highly useful in the detection of emotions. It is also information that is readily interpretable by humans. Further research in this direction certainly seems very promising both in terms of the achievement of the coarse end goal itself, that is microexpression recognition, as well as in terms of interdisciplinary significance and furthering our understanding of microexpressions.

4.2. Data and Its Limitations

As we saw in the previous section, a major practical obstacle limiting research on microexpressions concerns the availability, quality, and standardization of data used by researchers. One of the fundamental issues stems from the fact that repeatable and uniform stimulation of spontaneous microexpressions is challenging. In research to date, participants are usually exposed to emotional videos which are then expected to rouse participants' emotions but which the participants are asked to attempt to conceal. Since in some instances emotional arousal fails, many recordings end up being useless as they contain no microexpression exhibition—this is one of the reasons why both the number of microexpression corpora is small and why each of the data sets contain relatively few class examples.

Another practical difficulty, pervasive in data intensive applications, concerns the encoding or labeling of data, which is very time consuming and laborious. The process requires a trained and skilled labeler, repeated examination of participants' recordings (often in slow motion), and the marking of the microexpression onset, peak, and termination. Thus, in addition to the process being laborious and slow, it is also inexact, with interlabeled variability being an issue. Closely connected to this problem is the fact that there is no uniform and widely accepted standard for the classification of microexpressions. Therefore, the labeling approaches adopted for different databases are different (with similar microexpressions treated as different depending on the data set used), posing challenges to the understanding of the performance of the state of the art, relative performances of different methods, and their advantages and disadvantages. There is no doubt that further work in this area is badly needed and that contributions to standardization would benefit the field enormously.

4.3. Real-Time Microexpression Recognition

In the realm of microexpression analysis, the tasks of microexpression classification and the mapping of the corresponding class clusters onto the space of emotions are certainly the most widely addressed ones in the literature and arguably the most important ones. In some practical applications, it is desirable to be able to do this in real-time. Considering that the duration of a microexpression is very short, from 1/25 s to 1/5 of a second, it is clear that this is a major computational challenge, especially when the application is required to run on embedded or mobile devices. Although workarounds are possible in principle, e.g., by offloading computation to more powerful servers, this may not always be possible and new potential bottlenecks emerge due to the need to transmit large amounts of data. Given the lack of attention to the problems associated with computational efficiency in the existing literature and the aforementioned need for microexpression analysis in real time, this direction of research also offers a range of opportunities for valuable future contributions.

4.4. Standardization of Performance Metrics

Another practical issue of great importance, which is also a consequence of the field's relative youth, concerns performance metrics used to evaluate proposed methods. While there has already been some discussion on this topic in the literature, there remains much room for improvement in the realm of standardization of the entire evaluation process.

In the existing MER literature, the most commonly used cross-validation method employed when only a single data set is used, is Leave-One-Subject-Out (LOSO). In LOSO, a single subject's data is withheld and used as a validation data set, while all remaining subjects' data is used for training. The overall performance of a method is then assessed by aggregating the results of all different possible iterations of the process, i.e., of all subjects being withheld in turn. For cross-database evaluations, one or more databases are used as the training set and a different corpus (not contributing to the training set) for validation.

Since most microexpression data sets are unbalanced in terms of class representation (as discussed previously), the use of F1-score is widely adopted and indeed advised. Using what may appear as the most natural assessment metric, the average accuracy, likely results in bias towards classes with larger numbers of examples, thus overestimating the method's performance. In addition, because all currently available data sets are rather small in the total count of videos of expressions, it is highly desirable to perform evaluation using multiple corpora; this is also advisable due to different characteristics of different corpora (e.g., the participants' ethnicities, as illustrated in Table 1). For cross-database evaluations, it is recommended to use the unweighted average recall rate (UAR) and the weighted average recall rate (WAR). WAR is defined as the number of correctly classified samples divided by the total number of samples, whereas UAR is computed as the average of per-class accuracies by category (and is hence invariant to the number of data samples in each class).

Table 1. Summary of Spontaneous Microexpression databases.

Database	CASME [40]	SMIC [41]			CASME II [42]	SAMM [43]	CAS(ME)² [44]
		HS	VIS	NIR			
Microexpressions	195	164	71	71	247	159	57
Participants	35	20	10	10	35	32	22
FPS	60	100	25	25	200	200	30
Ethnicities	1	3			1	13	1
Average Age	22.03	N/A			22.03	33.24	22.59
Resolution	640 × 480 1280 × 720	640 × 480			640 × 480	2040 × 1088	640 × 480
Facial Resolution	150 × 190	190 × 230			280 × 340	400 × 400	N/A
Emotion Classes	**8** Happiness Sadness Disgust Surprise Contempt Fear Repression Tense	**3** Positive Negative Surprise			**5** Happiness Disgust Surprise Repression Others	**7** Contempt Disgust Fear Anger Sadness Happiness Surprise	**4** Positive Negative Surprise Others

5. Conclusions

In this article we provided an up-to-date summary of published work of microexpression recognition, a comparative overview of publicly available microexpression data sets, and a discussion of methodological issues of interest to researchers on automated microexpression analysis. We also sought to elucidate some of the most prominent challenges in the field and illuminate promising future research directions. To recap, there is an immediate need for the development of more standardized, reliable, and repeatable protocols for microexpression data collection, and the establishment of universal protocols for the evaluation of algorithms in the field. On the technical front, the detection of action unit engagement and the development of more task-specific deep learning based approaches appear as the most promising research directions at this point in time. Lastly, it bears emphasizing that all of the aforementioned challenges demand collaborative, interdisciplinary efforts, involving expertise in computer science, psychology, and physiology.

Abbreviations

The following abbreviations are used in this manuscript:

MER	Microexpression Recognition
MEGC	Microexpression Grand Challenge
HOG	Histograms of Oriented Gradients
LBP-TOP	Local Binary Pattern-Three Orthogonal Planes
HOOF	Histograms of Oriented Optical Flow

FACS	Facial Action Coding System
LBP-SIP	Local Binary Pattern with Six Intersection Points
CBP	Centralized Binary Pattern
SMIC	Spontaneous Microexpression Corpus
CASME	Chinese Academy of Sciences Micro-Expressions
CASME II	Chinese Academy of Sciences Micro-Expression II
SAMM	Spontaneous Actions and Micromovements
CAS(ME)^2	Chinese Academy of Sciences Spontaneous Macro-Expressions and Micro-Expressions
LOSO	Leave One Subject Out

Appendix A

Table A1. Partial Summary of Microexpression Recognition Work on Spontaneous Databases from 2011 to 2020.

Paper	Feature	Method	Database	Best Result
2011 Pfister et al. [18]	Hand-crafted	LBP-TOP	Earlier version of SMIC	Acc: 71.4%
2013 Li et al. [41]	Hand-crafted	LBP-TOP	SMIC	Acc: 52.11% (VIS)
2014 Guo et al. [45]	Hand-crafted	LBP-TOP	SMIC	Acc: 65.83%
2014 Wang et al. [26]	Hand-crafted	TICS	CASME CASME II	Acc: 61.85% Acc: 58.53%
2014 Wang et al. [46]	Hand-crafted	DTSA	CASME	Acc: 46.90%
2014 Yan et al. [42]	Hand-crafted	LBP-TOP	CASME II	Acc: 63.41%
2015 Huang et al. [47]	Hand-crafted	STLBP-IP	SMIC CASME II	Acc: 57.93% Acc: 59.51%
2015 Huang et al. [35]	Hand-crafted	STCLQP	SMIC CASME CASME II	Acc: 64.02% Acc: 57.31% Acc: 58.39%
2015 Le et al. [48]	Hand-crafted	DMDSP+LBP-TOP	CASME II	F1-score: 0.52
2015 Le et al. [49]	Hand-crafted	LBP-TOP+STM	SMIC CASME II	Acc: 44.34% Acc: 43.78%
2015 Liong et al. [50]	Hand-crafted	OSW-LBP-TOP	SMIC CASME II	Acc: 57.54% Acc: 66.40%
2015 Lu et al. [51]	Hand-crafted	DTCM	SMIC CASME CASME II	Acc: 82.86% Acc: 64.95% Acc: 64.19%
2015 Wang et al. [27]	Hand-crafted	TICS, CIELuv and CIELab	CASME CASME II	Acc: 61.86% Acc: 62.30%
2015 Wang et al. [52]	Hand-crafted	LBP-SIP and LBP-MOP	CASME	Acc: 66.8%
2016 Ben et al. [53]	Hand-crafted	MMPTR	CASME	Acc: 80.2%
2016 Chen et al. [54]	Hand-crafted	3DHOG	CASME II	Acc: 86.67%
2016 Kim et al. [33]	Deep Learning	CNN+LSTM	CASME II	Acc: 60.98%
2016 Liong et al. [55]	Hand-crafted	Optical Strain	SMIC CASME II	Acc: 52.44% Acc: 63.41%
2016 Liu et al. [29]	Hand-crafted	MDMO	SMIC CASME CASME II	Acc: 80% Acc: 68.86% Acc: 67.37%
2016 Oh et al. [56]	Hand-crafted	I2D	SMIC CASME II	F1-score: 0.44 F1-score: 0.41
2016 Talukder et al. [57]	Hand-crafted	LBP-TOP	SMIC	Acc: 62% (NIR)
2016 Wang et al. [21]	Hand-crafted	STCCA	CASME CASME II	Acc: 41.20% Acc: 38.39%

Table A1. *Cont.*

Paper	Feature	Method	Database	Best Result
2016 Zheng et al. [58]	Hand-crafted	LBP-TOP, HOOF	CASME CASME II	Acc: 69.04% Acc: 63.25%
2017 Happy and Routray [59]	Hand-crafted	FHOFO	SMIC	F1-score: 0.5243
			CASME CASME II	F1-score: 0.5489 F1-score: 0.5248
2017 Liong et al. [60]	Hand-crafted	Bi-WOOF	SMIC CASME II	Acc: 53.52% (VIS) F1-score: 0.59
2017 Peng et al. [34]	Deep Learning	DTSCNN	CASMEI/II	Acc: 66.67%
2017 Wang et al. [61]	Hand-crafted	LBP-TOP	CASME II	Acc: 75.30%
2017 Zhang et al. [62]	Hand-crafted	LBP-TOP	CASME II	Acc: 62.50%
2017 Zong et al. [63]	Hand-crafted	LBP-TOP,TSRG	CASME II and SMIC	UAR: 0.6015
2018 Ben et al. [64]	Hand-crafted	HWP-TOP	CASME II	Acc: 86.8%
2018 Hu et al. [65]	Hand-crafted	LGBP-TOP and CNN	SMIC CASME II	Acc: 65.1% Acc: 66.2%
2018 Khor et al. [36]	Deep Learning	ELRCN	CASME II SAMM	F1-score: 0.5 F1-score: 0.409
2018 Li et al. [66]	Hand-crafted	HIGO	SMIC CASME II	Acc: 68.29 (HS) Acc: 67.21
2018 Liong et al. [67]	Hand-crafted	Bi-WOOF	SMIC CASME II	F1-score: 0.62 (HS) F1-score: 0.61
2018 Su et al. [68]	Hand-crafted	DS-OMMA	CASME II CAS(ME)^2	F1-score: 0.7236 F1-score: 0.7367
2018 Zhu et al. [69]	Hand-crafted	LBP-TOP and OF	CASME II	Acc: 53.3%
2018 Zong et al. [70]	Hand-crafted	STLBP-IP	CASME II	Acc: 63.97%
2019 Gan et al. [71]	Deep Learning	OFF-ApexNet	SMIC CASME II SAMM	Acc: 67.6% Acc: 88.28% Acc: 69.18%
2019 Huang et al. [72]	Hand-crafted	DiSTLBP-RIP	SMIC CASME CASME II	Acc: 63.41% Acc: 64.33% Acc: 64.78%
2019 Li et al. [73]	Deep Learning	3D-FCNN	SMIC CASME CASME II	Acc: 55.49% Acc: 54.44% Acc: 59.11%
2019 Liong et al. [74]	Deep Learning	STSTNet	SMIC, CASME II and SAMM	UF1: 0.7353 and UAR: 0.7605
2019 Liu et al. [75]	Deep Learning	EMR	SMIC, CASME II and SAMM	UF1: 0.7885 and UAR: 0.7824
2019 Peng et al. [76]	Hand-crafted	HIGO-TOP, ME-Booster	SMIC CASME II	Acc: 68.90% (HS) Acc: 70.85%
2019 Peng et al. [77]	Deep Learning	Apex-Time Network	SMIC CASME II SAMM	UF1: 0.497 and UAR: 0.489 UF1: 0.523 and UAR: 0.501 UF1: 0.429 and UAR: 0.427
2019 Van Quang et al. [78]	Deep Learning	CapsuleNet	SMIC, CASME II and SAMM	UF1: 0.6520 and UAR: 0.6506
2019 Xia et al. [79]	Deep Learning	MER-RCNN	SMIC CASME CASME II	Acc: 57.1% Acc: 63.2% Acc: 65.8%
2019 Zhao and Xu [80]	Hand-crafted	NMPs	SMIC CASME II	Acc: 69.37% Acc: 72.08%
2019 Zhou et al. [81]	Deep Learning	Dual-Inception	SMIC, CASME II and SAMM	UF1: 0.7322 and UAR: 0.7278

Table A1. *Cont.*

Paper	Feature	Method	Database	Best Result
2020 Wang et al. [82]	Deep Learning	ResNet, Micro-Attention	SMIC CASME II SAMM	Acc:49.4% Acc:65.9% Acc: 48.5%
2020 Xie et al. [83]	Deep Learning	AU-GACN	CASME II SAMM	Acc:49.2% Acc: 48.9%

References

1.   Ekman, P.; Oster, H. Facial Expressions of Emotion. *Annu. Rev. Psychol.* **1979**, *30*, 527–554. [CrossRef]
2.   Dimberg, U.; Thunberg, M.; Elmehed, K. Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* **2000**, *11*, 86–89. [CrossRef] [PubMed]
3.   Haggard, E.A.; Isaacs, K.S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In *Methods of Research in Psychotherapy*; Springer: Berlin/Heidelberg, Germany, 1966.
4.   Ekman, P.; Friesen, W.V. Nonverbal Leakage and Clues to Deception. *Psychiatry* **1969**, *32*, 88–106. [CrossRef] [PubMed]
5.   Ekman, P. Lie Catching and Microexpressions. In *The Philosophy of Deception*; Oxford University Press: New York, USA, 2011.
6.   Morris, M.W.; Keltner, D. How emotions work: The social functions of emotional expression in negotiations. *Res. Organ. Behav.* **2000**, *22*, 1–50. [CrossRef]
7.   Whitehill, J.; Serpell, Z.; Lin, Y.C.; Foster, A.; Movellan, J.R. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Trans. Affect. Comput.* **2014**, *5*, 86–98. [CrossRef]
8.   Oh, Y.H.; See, J.; Le Ngo, A.C.; Phan, R.C.W.; Baskaran, V.M. A survey of automatic facial micro-expression analysis: Databases, methods, and challenges. *Front. Psychol.* **2018**, *9*, 1128. [CrossRef]
9.   Goh, K.M.; Ng, C.H.; Lim, L.L.; Sheikh, U.U. Micro-expression recognition: An updated review of current trends, challenges and solutions. *Vis. Comput.* **2020**, *36*, 445–468. [CrossRef]
10. Rani, M.; Rathee, N. Microexpression Analysis: A Review. In Proceedings of the 3rd International Conference on Computing Informatics and Networks: ICCIN 2020, Delhi, India, 29–30 July 2020; Springer Nature: Berlin/Heidelberg, Germany, 2021; p. 125.
11. Pan, H.; Xie, L.; Wang, Z.; Liu, B.; Yang, M.; Tao, J. Review of micro-expression spotting and recognition in video sequences. *Virtual Real. Intell. Hardw.* **2021**, *3*, 1–17. [CrossRef]
12. Polikovsky, S.; Kameda, Y.; Ohta, Y. *Facial Micro-Expressions Recognition Using High Speed Camera and 3D-Gradient Descriptor*; IET Seminar Digest; IET: London, UK, 2009.
13. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1971.
14. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]
15. Fan, J.; Arandjelović, O. Employing domain specific discriminative information to address inherent limitations of the LBP descriptor in face recognition. In Proceedings of the International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–7.
16. Karsten, J.; Arandjelović, O. Automatic vertebrae localization from CT scans using volumetric descriptors. In Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, Jeju, Korea, 11–15 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 576–579.
17. Zhao, G.; Pietikäinen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [CrossRef]
18. Pfister, T.; Li, X.; Zhao, G.; Pietikäinen, M. Recognising spontaneous facial micro-expressions. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
19. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active shape models-their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59. [CrossRef]
20. Goshtasby, A. Image registration by local approximation methods. *Image Vis. Comput.* **1988**, *6*, 255–261. [CrossRef]
21. Wang, S.J.; Yan, W.J.; Sun, T.; Zhao, G.; Fu, X. Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing* **2016**, *214*, 218–232. [CrossRef]
22. Wang, Y.; See, J.; Raphael, R.; Oh, Y.H. LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2015.
23. Fu, X.; Wei, W. Centralized binary patterns embedded with image euclidean distance for facial expression recognition. In Proceedings of the 4th International Conference on Natural Computation, ICNC 2008, Jinan, China, 18–20 October 2008.
24. Guo, Y.; Xue, C.; Wang, Y.; Yu, M. Micro-expression recognition based on CBP-TOP feature with ELM. *Optik* **2015**, *126*, 4446–4451. [CrossRef]
25. Arandjelović, O. Colour invariants under a non-linear photometric camera model and their application to face recognition from video. *Pattern Recognit.* **2012**, *45*, 2499–2509. [CrossRef]

26. Wang, S.J.; Yan, W.J.; Li, X.; Zhao, G.; Fu, X. Micro-expression recognition using dynamic textures on tensor independent color space. In Proceedings of the International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014.

27. Wang, S.J.; Yan, W.J.; Li, X.; Zhao, G.; Zhou, C.G.; Fu, X.; Yang, M.; Tao, J. Micro-Expression Recognition Using Color Spaces. *IEEE Trans. Image Process.* **2015**, *24*, 6034–6047. [CrossRef]

28. Zhang, L.; Arandjelović, O.; Dewar, S.; Astell, A.; Doherty, G.; Ellis, M. Quantification of advanced dementia patients' engagement in therapeutic sessions: An automatic video based approach using computer vision and machine learning. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 5785–5788.

29. Liu, Y.J.; Zhang, J.K.; Yan, W.J.; Wang, S.J.; Zhao, G.; Fu, X. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition. *IEEE Trans. Affect. Comput.* **2016**, *7*, 299–310. [CrossRef]

30. Asthana, A.; Zafeiriou, S.; Cheng, S.; Pantic, M. Robust discriminative response map fitting with constrained local models. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.

31. Chaudhry, R.; Ravichandran, A.; Hager, G.; Vidal, R. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, Miami, FL, USA, 20–25 June 2009.

32. Xu, F.; Zhang, J.; Wang, J.Z. Microexpression Identification and Categorization Using a Facial Dynamics Map. *IEEE Trans. Affect. Comput.* **2017**, *8*, 254–267. [CrossRef]

33. Kim, D.H.; Baddar, W.J.; Ro, Y.M. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In Proceedings of the MM 2016—Proceedings of the 2016 ACM Multimedia Conference, Amsterdam, The Netherlands, 15–19 October 2016.

34. Peng, M.; Wang, C.; Chen, T.; Liu, G.; Fu, X. Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* **2017**, *8*, 1745. [CrossRef]

35. Huang, X.; Zhao, G.; Hong, X.; Zheng, W.; Pietikäinen, M. Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns. *Neurocomputing* **2015**, *175*, 564–578. [CrossRef]

36. Khor, H.Q.; See, J.; Phan, R.C.W.; Lin, W. Enriched long-term recurrent convolutional network for facial micro-expression recognition. In Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Xi'an, China, 15–19 May 2018.

37. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Darrell, T.; Saenko, K. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

38. Shreve, M.; Godavarthy, S.; Goldgof, D.; Sarkar, S. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In Proceedings of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011, Santa Barbara, CA, USA, 21–25 March 2011.

39. Warren, G.; Schertler, E.; Bull, P. Detecting deception from emotional and unemotional cues. *J. Nonverbal Behav.* **2009**, *33*, 59–56. [CrossRef]

40. Yan, W.J.; Wu, Q.; Liu, Y.J.; Wang, S.J.; Fu, X. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013, Shanghai, China, 22–26 April 2013.

41. Li, X.; Pfister, T.; Huang, X.; Zhao, G.; Pietikainen, M. A Spontaneous Micro-expression Database: Inducement, collection and baseline. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013, Shanghai, China, 22–26 April 2013.

42. Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* **2014**, *9*, e86041. [CrossRef]

43. Davison, A.K.; Lansley, C.; Costen, N.; Tan, K.; Yap, M.H. SAMM: A Spontaneous Micro-Facial Movement Dataset. *IEEE Trans. Affect. Comput.* **2018**, *9*, 116–129. [CrossRef]

44. Qu, F.; Wang, S.J.; Yan, W.J.; Li, H.; Wu, S.; Fu, X. CAS(ME)^2: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. *IEEE Trans. Affect. Comput.* **2018**, *9*, 424–436. [CrossRef]

45. Guo, Y.; Tian, Y.; Gao, X.; Zhang, X. Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method. In Proceedings of the International Joint Conference on Neural Networks, Beijing, China, 6–11 July 2014.

46. Wang, S.J.; Chen, H.L.; Yan, W.J.; Chen, Y.H.; Fu, X. Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. *Neural Process. Lett.* **2014**, *39*, 25–43. [CrossRef]

47. Huang, X.; Wang, S.J.; Zhao, G.; Piteikainen, M. Facial Micro-Expression Recognition Using Spatiotemporal Local Binary Pattern with Integral Projection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.

48. Le Ngo, A.C.; Liong, S.T.; See, J.; Phan, R.C.W. Are subtle expressions too sparse to recognize? In Proceedings of the International Conference on Digital Signal Processing, DSP, Singapore, 21–24 July 2015.

49. Le Ngo, A.C.; Phan, R.C.W.; See, J. Spontaneous subtle expression recognition: Imbalanced databases and solutions. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2015.

50. Liong, S.T.; See, J.; Phan, R.C.; Le Ngo, A.C.; Oh, Y.H.; Wong, K.S. Subtle expression recognition using optical strain weighted features. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2015.

51. Lu, Z.; Luo, Z.; Zheng, H.; Chen, J.; Li, W. A Delaunay-based temporal coding model for micro-expression recognition. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2015.

52. Wang, Y.; See, J.; Phan, R.C.; Oh, Y.H. Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. *PLoS ONE* **2015**, *10*, e0124674. [CrossRef] [PubMed]

53. Ben, X.; Zhang, P.; Yan, R.; Yang, M.; Ge, G. Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation. *Neural Comput. Appl.* **2016**, *27*, 2629–2646. [CrossRef]

54. Chen, M.; Ma, H.T.; Li, J.; Wang, H. Emotion recognition using fixed length micro-expressions sequence and weighting method. In Proceedings of the 2016 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2016, Angkor Wat, Cambodia, 6–10 June 2016.

55. Liong, S.T.; See, J.; Phan, R.C.; Oh, Y.H.; Cat Le Ngo, A.; Wong, K.S.; Tan, S.W. Spontaneous subtle expression detection and recognition based on facial strain. *Signal Process. Image Commun.* **2016**, *47*, 170–182. [CrossRef]

56. Oh, Y.H.; Le Ngo, A.C.; Phari, R.C.; See, J.; Ling, H.C. Intrinsic two-dimensional local structures for micro-expression recognition. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20–25 March 2016.

57. Talukder, B.M.; Chowdhury, B.; Howlader, T.; Rahman, S.M. Intelligent recognition of spontaneous expression using motion magnification of spatio-temporal data. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2016.

58. Zheng, H.; Geng, X.; Yang, Z. A relaxed K-SVD algorithm for spontaneous micro-expression recognition. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2016.

59. Happy, S.L.; Routray, A. Fuzzy Histogram of Optical Flow Orientations for Micro-expression Recognition. *IEEE Trans. Affect. Comput.* **2017**, *10*, 394–406. [CrossRef]

60. Liong, S.T.; See, J.; Wong, K.; Phan, R.C.W. Automatic micro-expression recognition from long video using a single spotted apex. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2017.

61. Wang, Y.; See, J.; Oh, Y.H.; Phan, R.C.; Rahulamathavan, Y.; Ling, H.C.; Tan, S.W.; Li, X. Effective recognition of facial micro-expressions with video motion magnification. *Multimed. Tools Appl.* **2017**, *76*, 21665–21690. [CrossRef]

62. Zhang, S.; Feng, B.; Chen, Z.; Huang, X. Micro-expression recognition by aggregating local spatio-temporal patterns. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2017.

63. Zong, Y.; Huang, X.; Zheng, W.; Cui, Z.; Zhao, G. Learning a target sample re-generator for cross-database micro-expression recognition. In Proceedings of the MM 2017—Proceedings of the 2017 ACM Multimedia Conference, Mountain View, CA, USA, 23–27 October 2017.

64. Ben, X.; Jia, X.; Yan, R.; Zhang, X.; Meng, W. Learning effective binary descriptors for micro-expression recognition transferred by macro-information. *Pattern Recognit. Lett.* **2018**, *107*, 50–58. [CrossRef]

65. Hu, C.; Jiang, D.; Zou, H.; Zuo, X.; Shu, Y. Multi-task Micro-expression Recognition Combining Deep and Handcrafted Features. In Proceedings of the International Conference on Pattern Recognition, Beijing, China, 20–24 August 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA , 2018; pp. 946–951.

66. Li, X.; Hong, X.; Moilanen, A.; Huang, X.; Pfister, T.; Zhao, G.; Pietikainen, M. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput.* **2018**, *9*, 563–577. [CrossRef]

67. Liong, S.T.; See, J.; Wong, K.S.; Phan, R.C. Less is more: Micro-expression recognition from video using apex frame. *Signal Process. Image Commun.* **2018**, *62*, 82–92. [CrossRef]

68. Su, W.; Wang, Y.; Su, F.; Zhao, Z. Micro-expression recognition based on the spatio-temporal feature. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2018, San Diego, CA, USA, 23–27 July 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018.

69. Zhu, X.; Ben, X.; Liu, S.; Yan, R.; Meng, W. Coupled source domain targetized with updating tag vectors for micro-expression recognition. *Multimed. Tools Appl.* **2018**, *77*, 3105–3124. [CrossRef]

70. Zong, Y.; Huang, X.; Zheng, W.; Cui, Z.; Zhao, G. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Trans. Multimed.* **2018**, *20*, 3160–3172. [CrossRef]

71. Gan, Y.S.; Liong, S.T.; Yau, W.C.; Huang, Y.C.; Tan, L.K. OFF-ApexNet on micro-expression recognition system. *Signal Process. Image Commun.* **2019**, *74*, 129–139. [CrossRef]

72. Huang, X.; Wang, S.J.; Liu, X.; Zhao, G.; Feng, X.; Pietikainen, M. Discriminative Spatiotemporal Local Binary Pattern with Revisited Integral Projection for Spontaneous Facial Micro-Expression Recognition. *IEEE Trans. Affect. Comput.* **2019**, *10*, 32–47. [CrossRef]
73. Li, J.; Wang, Y.; See, J.; Liu, W. Micro-expression recognition based on 3D flow convolutional neural network. *Pattern Anal. Appl.* **2019**, *22*, 1331–1339. [CrossRef]
74. Liong, S.T.; Gan, Y.S.; See, J.; Khor, H.Q.; Huang, Y.C. Shallow triple stream three-dimensional CNN (STSTNet) for micro-expression recognition. In Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Lille, France, 14–18 May 2019.
75. Liu, Y.; Du, H.; Zheng, L.; Gedeon, T. A neural micro-expression recognizer. In Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Lille, France, 14–18 May 2019.
76. Peng, W.; Hong, X.; Xu, Y.; Zhao, G. A boost in revealing subtle facial expressions: A consolidated Eulerian framework. In Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Lille, France, 14–18 May 2019.
77. Peng, M.; Shi, Y.; Wang, C.; Zhou, X.; Bi, T.; Chen, T. A novel apex-time network for cross-dataset micro-expression recognition. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; IEEE: Piscataway, NJ, USA, 2019.
78. Van Quang, N.; Chun, J.; Tokuyama, T. CapsuleNet for micro-expression recognition. In Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Lille, France, 14–18 May 2019.
79. Xia, Z.; Feng, X.; Hong, X.; Zhao, G. Spontaneous facial micro-expression recognition via deep convolutional network. In Proceedings of the 2018 8th International Conference on Image Processing Theory, Tools and Applications, IPTA 2018, Xi'an, China, 7–10 November 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019.
80. Zhao, Y.; Xu, J. An improved micro-expression recognition method based on necessary morphological patches. *Symmetry* **2019**, *11*, 497. [CrossRef]
81. Zhou, L.; Mao, Q.; Xue, L. Dual-inception network for cross-database micro-expression recognition. In Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Lille, France, 14–18 May 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019.
82. Wang, C.; Peng, M.; Bi, T.; Chen, T. Micro-attention for micro-expression recognition. *Neurocomputing* **2020**, *410*, 354–362. [CrossRef]
83. Xie, H.X.; Lo, L.; Shuai, H.H.; Cheng, W.H. AU-assisted Graph Attention Convolutional Network for Micro-Expression Recognition. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020.