



Detecting Micro-expression Intensity Changes from Videos Based on Hybrid Deep CNN

Selvarajah Thuseethan^(✉), Sutharshan Rajasegarar, and John Yearwood

Deakin University, Geelong, VIC 3220, Australia
{tselvarajah, srajas, john.yearwood}@deakin.edu.au

Abstract. Facial micro-expressions, which usually last only for a fraction of a second, are challenging to detect by the human eye or machine. They are useful for understanding the genuine emotional state of a human face, and have various applications in education, medical, surveillance and legal sectors. Existing works on micro-expressions are focused on binary classification of the micro-expressions. However, detecting the micro-expression intensity changes over the spanning time, i.e., the micro-expression profiling, is not addressed in the literature. In this paper, we present a novel deep Convolutional Neural Network (CNN) based hybrid framework for micro-expression intensity change detection together with an image pre-processing technique. The two components of our hybrid framework, namely a micro-expression stage classifier, and an intensity estimator, are designed using a 3D and 2D shallow deep CNNs respectively. Moreover, we propose a fusion mechanism to improve the micro-expression intensity classification accuracy. Evaluation using the recent benchmark micro-expression datasets; CASME, CASME II and SAMM, demonstrates that our hybrid framework can accurately classify the various intensity levels of each micro-expression. Further, comparison with the state-of-the-art methods reveals the superiority of our hybrid approach in classifying the micro-expressions accurately.

Keywords: Micro-expression intensity ·
Convolutional Neural Networks · Hybrid framework ·
Fusion mechanism

1 Introduction

Facial expressions provide affluent information in understanding the emotional states of the human face. Facial expressions are broadly categorised into macro and micro expressions based on the spanning time and the strength of muscle movements. The macro-expression, also known as the regular expression, is easy to recognise under real-time settings through naked eyes since it continues for a considerable sprint of time. Unlike macro-expression, the micro-expression

provokes involuntarily as a rapid and brief facial expression. Hence, it is difficult to detect micro-expressions accurately and spontaneously by observers (e.g., another human). Presently, only the highly trained professionals can characterise the micro-expressions spontaneously, albeit with low accuracies. However, micro-expression is one of the most significant features for revealing the genuine emotions of a person, and therefore it is important to be analysed conscientiously. In the past, a widely used approach for analysing the micro expressions is through video processing techniques. Many micro-expression detection methods in the literature are confined, since they do not focus on systematic micro-expression profiling, such as detecting micro-expression intensity changes using spatial or temporal information.

Incorporating micro-expression is not only limited to traditional applications, such as lie or threatening behaviour detection, but also can be utilised in detecting positive behaviours, such as the level of confidence [2]. Micro-expressions are useful in many domains, such as educational, medical and legal sectors. For example, teachers can use the detected changes in micro-expressions to identify the students' stress, confidence level and other significant behavioural changes that happen in a classroom environment. In the medical field, doctors and psychologists can reveal the concealed emotions of their patients to provide further assistance. Micro-expression profiling plays a vital role during police investigations to recognise the abnormal behaviours, such as lies. For example, the intensity classification of micro-expressions can be used to analyse the onset of changes of a trouble maker in a public place. More importantly, the low level micro-expressions can also be detected by some intelligent algorithms, therefore, it is hard to conceal the actual emotion even for an expert human being. However, there are only a limited work exist in the literature that focus on comprehensive micro-expression profiling. This is due to the involuntariness and the shorter operating duration nature of the micro-expressions.

In this paper, we propose a novel hybrid micro-expression profiling framework that detects the *intensity changes* of micro-expression spontaneously from a video sequence. To the best of our knowledge, no previous work exist to detect the intensity changes of micro-expression on the fly. In our work, we use both spatial and temporal features, and a deep Convolutional Neural Network (CNN) based hybrid architecture, along with a fusion mechanism, to detect the intensity changes from videos. Further, we introduce a comprehensive image pre-processing technique to improve the micro-expression profiling, along with the major aspiration of this paper. In summary, the novelty and contributions of this research work are three-fold:

- We propose an efficient deep CNN based hybrid micro-expression intensity change detector to detect the fine-grained changes in a micro-expression. Our hybrid framework first classifies the micro-expression changes into three *stages*, namely *formation*, *peak* and *release*, using temporal information. Then, it detects the *frame level intensity changes* using spatial information.
- We introduce a fusion mechanism to enhance the automatic micro-expression intensity change detection accuracy.

- We demonstrate that our proposed hybrid framework is capable of accurately detecting the intensity changes through a cross-subject evaluation on three recent benchmark datasets. Moreover, we present a comparison results, before and after the fusion operation, to demonstrates the improvements achieved in the classification accuracy by the fusion process. Further, we compare our hybrid CNN based approach with the state-of-the-art algorithms and demonstrates its superiority in classifying the micro-expressions accurately.

The rest of this paper is organized as follows. Section 2 presents a review of the related work on micro-expression detection. We explain our proposed hybrid framework in Sect. 3. Experiment results and discussions are provided in Sect. 4. Finally, we conclude with suggestions for future work in Sect. 5.

2 Related Work

Here, we discuss the general history of micro-expressions, and provide a survey on the different micro-expression detection approaches used in the past. A recent review by Merghani et al. [8] provides an extensive summary of the literature on micro-expressions analysis. However, no work has been identified on the micro-expression intensity detection in the recent literature.

Two psychologists, namely Ekman and Friesen, have first discovered the micro-expression in 1969 while investigating a medical case, in which a subject tried to conceal the emotion sad nicely masked with a smile on his face [2]. Since then, numerous social psychologists, such as Ekman [2] and Gottman [3], have studied the micro-expressions extensively. Most of the techniques associated with the micro-expression are primarily focussed on detecting basic emotions, also known as prototypical or universal emotions, to provide Facial Action Coding System (FACS) based facial action unit annotations. Although the micro-expressions are hard to observe spontaneously, they are empirically inspected through psychological studies. Further, training programs (e.g., Ekman’s Micro Expressions Training Tools¹) have been created to enhance the support for micro-expression analysis.

In computer vision, only a handful of work exists on micro-expression analysis due to the high complexity of micro-expressions. Many current research focused on hand engineered features for analysis, such as 3D Histograms of Oriented Gradients (3DHOG) [1], Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) and variations [1, 11, 14], and Histogram of Oriented Optical Flow (HOOF) [14]. For computer vision tasks, nowadays, deep CNN has become a mainstream feature extractor. As a reflection of this, deep CNN based micro-expression detection research have emerged over the past few years, and two noteworthy works on this are [5] and [9]. In the past, researchers have used standard classifiers to classify the micro-expressions, such as support vector machines [15], long short-term memory (LSTM) [5] and dual temporal scale CNN (DTSCNN) [9]. In this

¹ <https://www.paulekman.com/micro-expressions-training-tools/>.

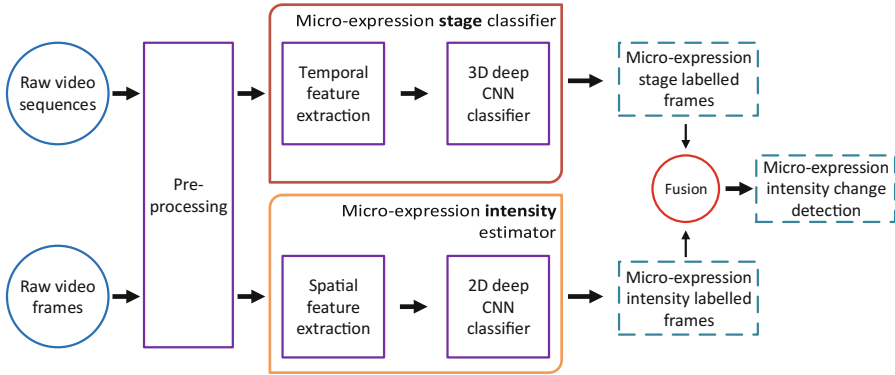


Fig. 1. Micro-expression intensity change detection framework.

paper, we focus on deep learning based feature extractors and classifiers for micro expression intensity detection.

In 2016, Kim et al. [5] adopted CNN for spontaneous micro-expression recognition from video sequences using spatiotemporal features. They encoded the spatial features from representative expression-state frames (i.e., onset, offset and apex) using CNN. Next, the temporal features are analysed after the learned spatial features are passed through an LSTM recurrent network. The overall accuracy achieved was 60.98% on CASME II [11], which is slightly better than the existing state-of-the-art approaches. Peng et al. [9] presented another significant work in micro-expression recognition using a two stream shallow network (4 layers only), namely DTSCNN. They used a selected set of images from the merged CASME I [12] and CASME II [11] datasets to train their model, which comprised four classes: positive, negative, surprise and others. Optical flow sequences, instead of raw sequences, are used to reduce the complexity of the classifier, and they achieved an overall accuracy of 66.67%.

In summary, works in [1, 4–6, 9, 11, 14, 15] have only addressed the recognition (binary classification) problem of the micro-expressions. None of the existing works focused on detecting the *intensity changes* of micro-expressions. In this paper, we propose a novel hybrid framework for detecting the micro-expression intensity changes from video. Moreover, a significant problem faced by the deep learning based methods is the lack of larger datasets for training, contributing to the lower overall detection accuracy. Hence, we also present a carefully crafted hybrid framework, which includes a pre-processing mechanism, that prepares a comprehensive training set for use with the deep learning based micro-expression analysis. Next, we present our proposed framework.

3 The Proposed Hybrid Framework

Figure 1 illustrates the overall architecture of our proposed hybrid micro-expression intensity change detection framework. The framework primarily com-

prises a pre-processing unit, two separate deep CNN components and a fusion component. In the pre-processing phase, we augment and normalise the training set to use with the subsequent deep learning hybrid framework. In the deep CNN component, a *micro-expression stage classifier* classifies the three stages (formation, peak and release) of micro-expression changes using temporal feature information. The next component, namely the *micro-expression intensity estimator*, estimates the frame-wise micro-emotion intensities using spatial feature information. Finally, the fusion mechanism updates the intensity predictions incorporating the stage classifier predictions. Below, we explain each of the components of our hybrid framework in detail.

3.1 Pre-processing

In the pre-processing process, we convert the input video frames to grayscale, as the first step, in order to reduce the cross-database discrepancy between the video frames. Our descriptor comprises two significant pre-processing phases namely (a) data augmentation or synthetic sample generation and (b) normalisation. In the data augmentation step, we generate a set of synthesised frames in large amounts to increase the number of video frames, especially for training purposes using a deep learning model, which often require larger dataset. Adopting the approach of [10], we added random noise, using a 2D Gaussian distribution, in the eye centre and nose regions of the face to produce the synthetic frames. We used the individual frames to train the micro-expression intensity estimator, and the whole sequence to train the micro-expression stage classifier, which classifies the frames into three stages; *formation* (start of the micro-expression), *peak* (highest intensity level of the micro-expression) and *release* (end of the micro-expression).

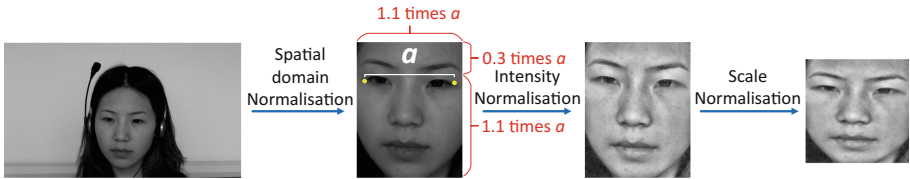


Fig. 2. Pre-processing steps: *Spatial domain normalisation* is performed based on the distance “ a ” between the active appearance model (AAM) facial feature points 37 and 46. The *intensity* and *scale* normalisations are applied subsequently.

We then perform a series of normalisation process in a sequel, as illustrated in Fig. 2. First, in the spatial normalisation step, a region of interest (ROI) is selected to eliminate the insignificant areas of the video frames for feature extraction, and each video frame is cropped accordingly. In this work, we eliminate not only the background information, but also some parts of the face, such as ears, chin and forehead, which do not reflect any micro-expression related information. The cropping of the facial region is performed based on the distance, indicated

as a , between the active appearance model (AAM) facial feature points 37 and 46, as shown in the second image of Fig. 2. Second, we apply an intensity normalisation process, using the Contrast Limited Adaptive Equalization (CLAHE) [16] method, on each video frame to reduce the variation in the feature vector. An advantage of the CLAHE is that it redistributes the histogram part which exceeds the clip limit between all histogram bins, rather than just eliminating it. In this work, we use a Rayleigh distribution with clip limit of 0.01 and α value of 1. Third, in the scale normalisation step, we down-sampled the video frames to reduce the size to 128×128 pixels, using linear interpolation. Scale normalisation enables the same facial feature points of different video frames to co-occur approximately in the same location. Fourth, as the final step of the pre-processing, we performed the temporal normalisation to the input dataset (both training and testing datasets) of the micro-expression stage classifier. The input dataset is normalised to eight frames, and used as input to classify them as one of the micro-expression stages, i.e., formation, release or peak.

The pre-processed frames are used as input to our deep CNN components, as illustrated in Fig. 1, which we discuss in detail next.

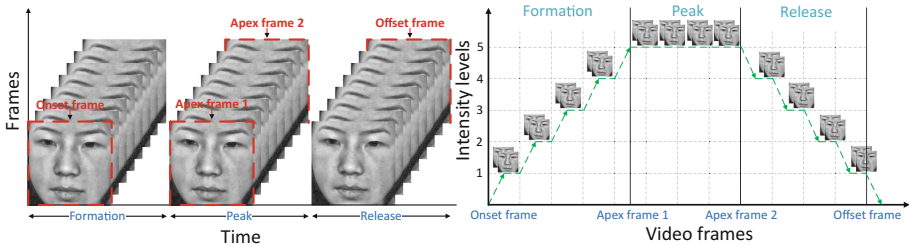


Fig. 3. The three stages of a spontaneously recognised micro-expression in a video sequence. The green dotted line demonstrates the intensity changes of the micro-expression throughout the sequence. (Color figure online)

3.2 Deep CNN Components

Figure 3 shows the ground-truth annotation of a micro-expression video sequence. In here, the onset and offset frames are the first and the last recognised frames of a particular micro-expression. The maximum intensity of the micro-expression (i.e., micro-expression intensity level 5 in our work) is identified from the frames between the apex frame 1 and the apex frame 2. The three stages, from onset frame to apex frame 1, apex frame 1 to apex frame 2 and apex frame 2 to offset frame, are named as *formation*, *peak* and *release* respectively. The micro-expression intensity shows a linearly increasing trend (levels 1 to 4) in the formation stage, and reaches the maximum (level 5) in the peak stage. After remaining unchanged for a while during the peak stage, the intensity drops linearly, during the release stage, to reach the minimum (level 1) again. Using the behaviour observed here, we perform the micro-expression stage detection, the frame-wise intensity estimation and a fusion process as explained below.

Micro-expression Stage Classifier. Micro-expression stage classifier is an action classifier, which classifies three stages of a micro-expression, called formation, peak and release. We use a shallow (i.e., three-layer) 3-dimensional convolutional network as the stage classifier. Further, two fully-connected (FC) layers are assembled, as the FC layer is known for learning a non-linear function in a computationally efficient way from the high-level feature space. We use a large-size convolutional kernel (i.e., input frames) while maintaining a fixed temporal dimension, which is not fine-tuned during the experiment. Unlike 2-dimensional CNNs, where each video frame is considered as an object, 3-dimensional CNNs treat the whole video sequence as an object for classification. As indicated in the figure, the size of the input image is $128 \times 128 \times 1 \times 8$ (*width* \times *height* \times *channel* \times *frames*). Each convolution layer is followed by a dropout layer, in which the dropout probability is set and tuned during the experiment. Subsequently, a pooling layer is assembled, after each dropout layer, which uses max-pooling with a kernel size of $1 \times 2 \times 2$.

We use the same values for learning parameters, such as momentum, learning rate and decay, to train the hybrid framework on CASME [12], CASME II [11] and SAMM [1] datasets. Stochastic gradient decent with momentum (SGDM) optimiser was applied with the parameters momentum and learning rate set to 0.95 and 0.001 respectively. The classification output of this stage classifier forms one of the inputs of the subsequent fusion process to estimate the micro-expression intensity changes, as shown in Algorithm 1.

Micro-expression Intensity Estimator. In the literature, the intensities of action units (AUs) were used to estimate the intensities of *macro* emotions [7]. However, unlike macro expressions, micro-expression intensity estimation using AU intensity is challenging due to its rapid and feeble nature. Here, we introduce a novel frame-wise micro-expression intensity estimator using *relative intensity differences* between the frames of a video sequence. Zhao et al. [13] used a relative intensity based method to estimate the intensities of the macro-emotions. Inspired by this work, to assign the frame-wise intensities of micro-expressions in a video sequence, we use the following Eq. (1) to obtain the intensity level of j^{th} frame I_j in a video sequence. The maximum and minimum intensities of micro-expression are denoted as $I_h = 5$ and $I_l = 1$, where the maximum intensity is attained at the peak micro-expression stage, and the minimum is attained at the formation and release micro-expression stages.

$$I_j = \left\lfloor \frac{j}{a_1} (I_h - I_l) \right\rfloor \psi_{1 \leq j < a_1} + I_h \psi_{a_1 \leq j \leq a_2} + \left\lfloor \frac{n-j}{n-a_2} (I_h - I_l) \right\rfloor \psi_{a_2 \leq j \leq n} \quad (1)$$

In Eq. (1), ψ is the indicator function applied on the frame number j , where $j = 1 \cdots n$, and n is the number of frames in a video sequence. a_1 and a_2 are the apex frame 1 and apex frame 2, respectively. The acquired micro-expression intensity curve for a video sequence is shown in Fig. 3.

For micro-expression intensity estimation, we propose a 2-dimensional deep CNN architecture, and the structure is similar to the one proposed for the micro-

expression stage classifier, but with the replacement of all the 3-dimensional convolution layers with 2-dimensional convolution layers. The input layer of the micro-expression intensity estimator takes images with size $128 \times 128 \times 1$ (*width* \times *height* \times *channel*) as input. The output size of the last fully connected layer is set to 5, representing the five-level micro-expression intensity values/levels. We set the learning environmental settings of the micro-expression intensity change detector to the same values as used for the micro-expression stage classifier.

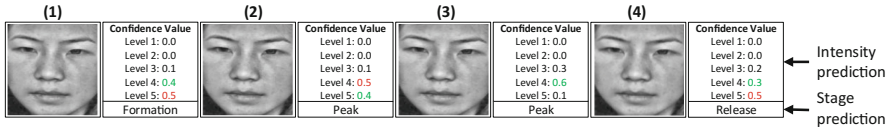


Fig. 4. For each frame, the predicted stage (formation, peak or release), and the confidence score for each intensity level, are shown. The red value indicate the highest confidence score and the corresponding intensity level of that frame. The green value indicates the confidence score and the corresponding intensity level that is assigned for that frame after the fusion process. (Color figure online)

3.3 Fusion Mechanism

The fusion process takes each frame and its intensity level obtained from the *micro-expression intensity estimator*, and adjusts its intensity level to match with the correct stage's (obtained from the *micro-expression stage classifier*; *formation*, *peak* and *release*) intensity level as described below. Figure 4 shows an example of this process.

- For the frames in the *formation and release stages*, we update the frames predicted with the highest intensity (i.e., level 5) by the intensity prediction of the second highest confidence score (e.g., frames (1) and (4) in Fig. 4). This adjustment is done because these frames are identified as belonging to either the formation stage or the release stage. Hence, their intensity levels are expected to be less than the maximum intensity level. We use the intensity level of the second highest confidence score to decide this.
- For the frames in the *peak stage*, we update the frames with predicted intensity levels 1, 2, 3 and 4 to intensity level 5 (e.g., frame (2)), if the second highest confidence value obtained is for level 5; otherwise we keep the intensity level of the highest confidence score (e.g., frame (3)). This adjustment is done because these frames are identified as belonging to the peak stage. Hence, their intensity levels are expected to be at the maximum level. We use the intensity level of the second highest confidence score to decide this.

Algorithm 1 explains the steps involved in the fusion process in detail.

Algorithm 1. Fusion process: S_i and F_i are the outputs of stage detector and intensity change detector components. Γ_i returns the intensity level for i^{th} confidence score rank. I_h indicates the highest intensity level. $\Omega(I)$ returns the confidence value of level I

```

1: procedure FUSION-PROCESS( $S_i, F_i$ )
2:   For all  $i$ ,  $S_i \in \{'Formation', 'Release'\}$ 
3:   if  $\Gamma_1(F_i) \neq I_h$  then
4:      $I_i = \Gamma_1(F_i)$ 
5:   else
6:      $I_i = \Gamma_2(F_i)$ 
7:   end if
8:   For all  $i$ ,  $S_i \in \{'Peak'\}$ 
9:   if  $\Gamma_1(F_i) = I_h$  or  $(\Gamma_1(F_i) \neq I_h \text{ and } \Omega(I_h) \geq \Omega(\Gamma_2(F_i)))$  then
10:     $I_i = I_h$ 
11:  else
12:     $I_i = \Gamma(I_i)$ 
13:  end if
14:  Output:  $I$ 
15: end procedure
    
```

4 Results and Analysis

In this section, we evaluate the proposed hybrid framework under divergent environments. The results are reported using classification accuracy measure obtained via leave-one-subject out cross validation procedure.

4.1 Micro-expression Intensity Change Detection

We present the results and analysis obtained using our proposed hybrid architecture, evaluated on three recent spontaneous micro-expression benchmark datasets CASME [12], CASME II [11] and SAMM [1]. CASME [12] dataset comprises 195 samples including eight micro-emotions namely, amusement, sadness, disgust, surprise, contempt, fear, repression and tense. The second dataset, CASME II [11], is an improved version of CASME [12], which comprises five distinct micro-expression classes: happiness, surprise, disgust, repression and others. SAMM [1] is another recent dataset with 159 micro-expression samples based on seven basic emotions; contempt, disgust, fear, anger, sadness, happiness, and surprise.

The confusion matrices shown in Fig. 5 presents the detection accuracies of the micro-expression stage classifier, where the average prediction results obtained are 89.37%, 93.2% and 93.6% on CASME [12], CASME II [11] and SAMM [1], respectively. We use this prediction result as the input for the fusion process to refine the micro-expression intensity change detection in videos. Furthermore, we evaluated the prediction accuracy of the micro-expression stage classifier *before* and *after* the *pre-processing* of datasets, and it clearly demonstrates a 10% increase in the classification rate on average. Although, the temporal normalisation is included in both of the experiments to maintain the uniformity of the fed CNN input, the above evaluation results are avoided for brevity.

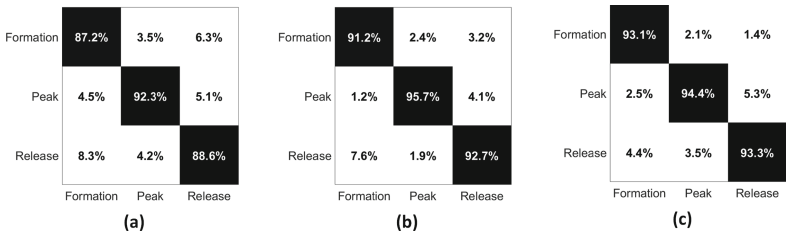


Fig. 5. Confusion matrix for the leave-one-subject-out classification of micro-expression stage classifier on (a) CASME [12], (b) CASME II [11] and (c) SAMM [1].

Table 1 shows the micro-expression intensity change classification results, adopting the entire validation methodology as described before. In order to validate the significance of the fusion process, we also carried out the experiment and recorded the prediction accuracies *before* and *after* the *fusion process*. In validation, before the fusion process, the intensity estimation is performed without the stage classifier. In validation, after the fusion process, the stage classifier is used to improve the output of the micro-expression intensity estimator. After the fusion process, the results demonstrate that the highest micro-expression intensity detection achieved is 76% with the SAMM [1] dataset, while the lowest is reported for the CASME [12] dataset, with 71%. As it can be seen, after the fusion process, we have a substantial increase ($>7\%$ better than the accuracy of before fusion) in the intensity prediction accuracy of any emotion, regardless of the micro-expression dataset. More precisely, the micro-expression intensity detection prediction results are more accurate when in combination with a stage classifier and the fusion process. For example, the Fig. 6 illustrates the intensity updation operation for a micro-expression video sequence after applying the fusion process, using confident scores (in this case, the micro-expression stages are correctly classified). In this, frames 9, 10 and 15 are adjusted to the correct intensity levels by the fusion process. This explicitly improves the classification accuracy from 62.5% to 75.0%.

Table 1. Micro-expression intensity change detection results before and after the fusion process. The pair of numbers represent the classification accuracies in the form of (before fusion, after fusion).

Micro-expressions	Datasets		
	CASME [12]	CASME II [11]	SAMM [1]
Amusement	(0.61, 0.68)	—	—
Anger	—	—	(0.69, 0.74)
Contempt	(0.57, 0.59)	—	(0.55, 0.60)
Disgust	(0.72, 0.79)	(0.79, 0.84)	(0.77, 0.81)
Fear	(0.77, 0.81)	—	(0.71, 0.79)
Happy	—	(0.73, 0.81)	(0.66, 0.71)
Others	—	(0.47, 0.52)	—
Repression	(0.52, 0.59)	(0.55, 0.59)	—
Sad	(0.81, 0.85)	—	(0.79, 0.84)
Surprise	(0.74, 0.77)	(0.81, 0.87)	(0.76, 0.83)
Tense	(0.58, 0.61)	—	—
Average rate	(0.67, 0.71)	(0.67, 0.73)	(0.70, 0.76)

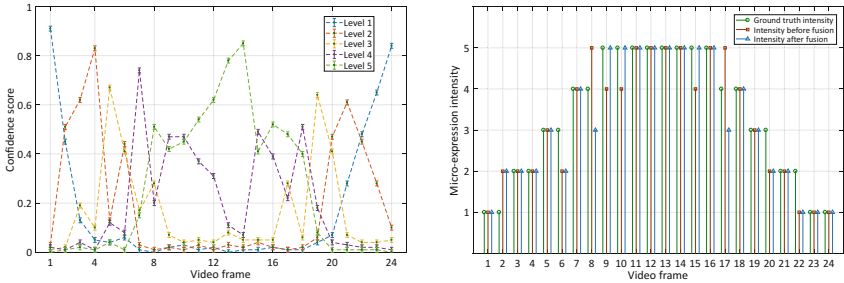


Fig. 6. Fusion process for a sequence in CASME [12] with emotion *happy*. Left figure illustrates frame-wise confidence scores and the right figure presents the frame-wise intensity labelling before and after the fusion process

4.2 Comparison with Existing Works

As discussed in the related work section, the closely related works to ours either targeted intensity estimation of the *macro*-expressions [13] or a binary classification (i.e., existence or not) of the micro-expression [1, 4–6, 11, 14]. In order to perform a fair comparison of our framework with the existing works, we need to convert the intensity level (five levels) output of our framework to a binary form. Hence, we combined all the micro-expression intensity levels from 1 to 5 and considered them as the indicator of the existence of the micro-expression, while the level 0 is considered as the indicator of non-existence of the micro-expression. We compared these works using the CASME [12], CASME II [11] and SAMM

[1] datasets. Table 2 provides the comparison results. It can be observed that the deep learning based approaches, both CNN based approach of [5] and our approach have outperformed the hand-crafted feature-based approaches. Further, our scheme showed the highest accuracy among all; 77%, 82% and 91% on CASME [12], CASME II [11] and SAMM [1] respectively. This demonstrates that our hybrid architecture and the fusion mechanism helped to attain higher accuracy. Moreover, our proposed framework is capable of further classifying the various *intensity* levels within the micro-expression as demonstrated in the previous section.

Table 2. Comparison of micro-expression recognition methods against the proposed hybrid framework with leave-one-subject-out cross-validation on CASME [12], CASME II [11] and SAMM [1]. ‡ and § indicate the performance measures recognition accuracy and F-measure respectively.

Methods	Datasets		
	CASME [12]	CASME II [11]	SAMM [1]
LBP-TOP [1, 11, 14]	0.69 [‡]	0.39 [§]	0.55 [§]
HOOF [14]	–	0.59 [‡]	–
3DHOG [1]	–	–	0.89 [‡]
Bi-WOOF [6]	–	0.61 [§]	–
FHOFO [4]	0.55 [§]	0.52 [§]	–
CNN [5]	–	0.61 [‡]	–
Ours	(0.77 [‡] , 0.61 [§])	(0.82 [‡] , 0.68 [§])	(0.91 [‡] , 0.70 [§])

5 Conclusion

In this paper, we proposed a novel hybrid micro-expression intensity change detection framework, consisting of three components, namely, micro-expression stage classifier, micro-expression intensity estimator and a fusion mechanism. The hybrid framework is built on a combination of 3D and 2D deep CNN based architectures. The fusion mechanism is integrated to further enhance the classification efficiency. Moreover, we proposed a comprehensive preprocessing technique to cater the requirements of deep networks; a large amount of training data and reduced complexity. Experiments were carried out on three benchmark micro-expression datasets CASME [12], CASME II [11] and SAMM [1]. We demonstrated that our hybrid framework is capable of accurately detecting the intensity changes in the micro-expression. Further, we compared the accuracy before and after the fusion mechanism, and shown the improvement achieved in the accuracy of intensity change detection. In the future, we aim to apply this in combination with the macro-expression intensity changes for fast detection of onset of emotions.

References

1. Davison, A.K., Lansley, C., Costen, N., Tan, K., Yap, M.H.: SAMM: a spontaneous micro-facial movement dataset. *IEEE Trans. Affect. Comput.* **9**(1), 116–129 (2018)
2. Ekman, P.: Lie catching and microexpressions. *Philos. Decept.* **1**, 5 (2009)
3. Gottman, J.M., Levenson, R.W.: A two-factor model for predicting when a couple will divorce: exploratory analyses using 14-year longitudinal data. *Fam. Process* **41**(1), 83–96 (2002)
4. Happy, S.L., Routray, A.: Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Trans. Affect. Comput.* (2017). <https://doi.org/10.1109/TAFFC.2017.2723386>
5. Kim, D.H., Baddar, W.J., Ro, Y.M.: Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In: *Proceedings of the 2016 ACM on Multimedia Conference*, pp. 382–386. ACM (2016)
6. Liong, S.T., See, J., Wong, K., Phan, R.C.W.: Less is more: micro-expression recognition from video using apex frame. *Signal Process.: Image Commun.* **62**, 82–92 (2018)
7. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Matthews, I.: Painful data: the UNBC-McMaster shoulder pain expression archive database. In: *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011)*, pp. 57–64. IEEE (2011)
8. Merghani, W., Davison, A.K., Yap, M.H.: A review on facial micro-expressions analysis: datasets, features and metrics. *arXiv preprint [arXiv:1805.02397](https://arxiv.org/abs/1805.02397)* (2018)
9. Peng, M., Wang, C., Chen, T., Liu, G., Fu, X.: Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* **8**, 1745 (2017)
10. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: *Null*, p. 958. IEEE (2003)
11. Yan, W.J., et al.: CASME II: an improved spontaneous micro-expression database and the baseline evaluation. *PloS One* **9**(1), e86041 (2014)
12. 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: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–7. IEEE (2013)
13. Zhao, R., Gan, Q., Wang, S., Ji, Q.: Facial expression intensity estimation using ordinal information. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3466–3474 (2016)
14. Zheng, H., Geng, X., Yang, Z.: A relaxed K-SVD algorithm for spontaneous micro-expression recognition. In: *Booth, R., Zhang, M.-L. (eds.) PRICAI 2016. LNCS (LNAI), vol. 9810, pp. 692–699. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42911-3_58*
15. 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.* **77**(3), 3105–3124 (2018)
16. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: *Graphics Gems*, pp. 474–485 (1994)