Image based Facial Micro-Expression Recognition using Deep Learning on Small Datasets

Madhumita A. Takalkar[1], Min Xu[2]
[1,2]School of Electrical and Data Engineering
University of Technology Sydney
15 Broadway, Ultimo, NSW 2007, Australia
[1]madhumita.a.takalkar@student.uts.edu.au, [2]min.xu@uts.edu.au

***Abstract*—Facial micro-expression refers to split-second muscle changes in the face, indicating that a person is either consciously or unconsciously suppressing their true emotions and even mental health. Therefore, micro-expression recognition attracts increasing research efforts in both fields of psychology and computer vision. Existing research on micro-expression recognition has mainly used hand-crafted features, for example, Local Binary Pattern-Three Orthogonal Planes (LBP-TOP), Gabor filter and optical flow. Recently, Deep Convolutional neural systems have demonstrated a high degree effectiveness for difficult face recognition tasks. This paper explores the possible use of deep learning for micro-expression recognition. To develop a reliable deep neural network extensive training sets are required with a huge number of labeled image samples. However, micro-expression recognition is a challenging task due to the repressed facial appearance and short duration, which results in the lack of training data. In this paper, we propose to generate extensive training datasets of synthetic images using data augmentation on CASME and CASME II databases. Then, these datasets are combined to tune a satisfactory CNN-based micro-expression recognizer. Experimental results demonstrate the effectiveness of the proposed CNN approach in image based micro-expression recognition and present comparable results with the best-related works.**

Keywords— Micro-expression recognition; deep learning; Convolutional Neural Network (CNN); small training data; data augmentation

I. INTRODUCTION

Breaking down facial micro-expression is a relatively new research theme with much-developing interests as of late. The principle explanation behind naturally distinguishing the micro-expression is that micro- expression is a critical emotional sign for different real-life applications. Micro-expression is a form of non-verbal communication that unconsciously reveals the true sentiments of a person. As compared to macro-expression, micro-expression has three basic qualities: short duration, subtle movement, and trouble about concealing [32]. Micro-expressions are exhibited subtly and typically occur very briefly, at a length of about 1/5 to 1/25 of a second [7], and they usually occur in several parts of the face where most people do not realise [18]. Similar to macro-expressions, micro-expressions can be grouped into six basic expressions: happy, surprise, anger, sad, disgust, and fear.

It is not straightforward to recognise the genuine emotion shown on one's face. Thus recognising micro-expressions is beneficial in our daily life as we can read if someone is attempting to cover his/her feeling or trying to deceive you. To solve this challenging problem, Ekman developed the Micro Expression Training Tool (METT) which can help people to detect micro-expression [3]. Some research found that it was still hard for a trained human to detect facial micro-expression. In a psychological experiment [4] conducted using METT micro-expression training dataset, the average micro-expression recognition rate was 50%. To solve these problems, the advanced computer vision technology can achieve higher detection and classification performance and solve this issue.

Deep learning is a group of machine learning methods biologically inspired from the structure of the brain. Convolutional Neural Networks (CNN) learns hierarchical features from multiple labelled images. It has been used for various applications including recognition of objects, actions, and places, face recognition [9, 24] and face expression recognition [20].

The deep convolutional neural network has recently yielded excellent performance in a wide variety of image classification tasks. The careful design for mapping of local to global feature learning with convolution, pooling, and layered architecture renders excellent visual representation ability, making it an effective tool for facial micro-expression recognition. In this paper, we concentrate on the task of the image based static facial micro-expression recognition on CASME, CASME II and both combined (CASME+2) with deep Convolutional Neural Networks (CNN). The input for recognition is a raw image; which is then pre-processed and given as input to CNN to predict the facial micro-expression label which should be one of the labels: Disgust, Fear, Happiness, Neutral, Sadness, and Surprise.

Deep learning requires a lot of data but for micro-expression recognition, there is not enough data: there are three spontaneous micro-expression datasets which altogether contain 748 videos. It is also not possible to collect millions of labelled training images from the internet for micro-expressions. Considering these limitations, we decided to convert the labelled micro-expression videos into frames and use these labelled frames for training the CNN model for image based micro-expression recognition. The frames are extracted every 0.2 seconds (or 200 milliseconds) from the video. The

number of frames per video depends on the length of the video. For example, for a video of length 2 seconds, 157 frames could be extracted. All the images extracted from a video will have the same label as the video. Even though we have extracted the frames from the videos, the number of image samples is not enough to train the deep neural network model. This leads us to the idea of combining the widely used micro-expression datasets to form a large dataset of training samples.

Our significant research contributions can be summarised as follows: 1. We propose a CNN architecture that achieves satisfactory recognition accuracy on micro-expression images; 2. We also aim to present a novel thought to combine the widely used micro-expression databases CASME and CASME II to increase the number of samples for training CNN model.

The structure of the paper is as follows: Section II reviews related work; Section III discusses the existing publicly available and widely used datasets for micro-expressions; Section IV introduces our proposed micro-expression recognition model; Section V presents the experimental setup and results, and the conclusions are drawn in Section VI.

II. Related Work

Studies in psychology demonstrate that facial features of expression are located around mouth, nose and eyes and their locations are essential for categorising facial micro-expressions. In this part, we review some existing micro-expression analysing approaches as well as the deep learning techniques. Most work on micro-expressions in the field of Computer Vision has mainly focused on micro-expression recognition [10, 19, 22, 23, 28, 29]. Micro-expression recognition task is defined as recognising the emotional label of well-segmented video containing micro-expression from start to end. Micro-expression analysing methods can be separated into two major groups, the local methods and the holistic methods. Initial research work on micro-expression recognition has been conducted on posed micro-expression. According to the Action Units [5], the local methods partition the face area into some subregions. The reported results were carried out within each face subregions. Wu et al. [28] extracted features using Gabor filters and used Support Vector Machine (SVM) to recognise them. Polikovsky et al. [23] proposed to utilise Active Shape Model (ASM) model to detect facial landmarks which used to segment face area into twelve subregions. In each subregion, 3D-gradient orientation descriptor was extracted as a descriptor of facial muscle movement. Finally, the micro-expression video was divided into onset, apex and offset stages. The holistic methods handled the whole face for analysing the micro-expression category. Pfister et al. [22] proposed a spatiotemporal holistic feature for recognising micro-expression. More specifically, the algorithm detected 68 landmarks in the first sequence by using ASM model. Then, these landmarks were utilised to align face for alleviating the issue of head movements. Temporal Interpolation Model (TIM) was used to each sequence so that the lengths of all the video were normalised [22]. Furthermore, LBP-TOP features were used to extract the feature of micro-expression and SVM, Multiple Kernel Learning (MKL), Random Forest (RF) were used to perform classification. Huang et al. [10] proposed SpatioTemporal Completed Local Quantization Patterns (STCLQ) feature for recognition. Liu et al. [19] proposed Main Directional Mean Optical Flow feature and used SVM for classification. Feng et al. [29] calculated principal optical flow direction resulting Facial Dynamics Map with SVM as a classifier. From the literature review, it was found that the main method for facial landmark localisation is ASM.

Recently, Deep Learning [13] has become very effective image analysis approach, such as image classification, semantic segmentation, object detection and image super-resolution. Compare to traditional hand-engineered features, such as Local Binary Pattern (LBP) [31] and Histogram of Gradients (HoG) [1, 16]; a deep convolutional neural network consists of multiple layers which can automatically learn hierarchies visual features directly from the raw image pixels. In the research [12], Kim et al. introduced deep learning features for micro-expression recognition. They proposed a new method consisting of two sections. First, the spatial features of micro-expressions at different expression states (onset, onset to apex transition, apex, apex to offset transition, and offset) are converted using convolutional neural networks (CNN). Next, the learned spatial features with expression-state constraints are transferred to learn temporal features of micro-expression. The temporal feature learning converts the temporal characteristics of the different states of the micro-expression using long short-term memory (LSTM) recurrent neural networks (RNNs) [12]. The time scale dependent information that resides along the video sequences is consequently learned by using LSTM.

Recent advances in micro-expression recognition focus on recognising more spontaneous facial micro-expressions. The Chinese Academy of Sciences Micro-Expression (CASME) [32] and CASME II [30] were collected to mimic more spontaneous scenarios and contain seven basic micro-expression categories. Both these datasets contain video clips which were recorded in a controlled environment. The idea is that videos, although not truly spontaneous, at least provide facial micro-expressions in a much more natural and versatile way than posed datasets. With the introduction of deep learning methods, a wide range of image classification work gives the finest performance. The existing works focus on spatio-temporal features for micro-expression recognition whereas we concentrate on the task of image-based static facial micro-expression recognition on CASME and CASME II with deep CNNs.

III. Databases

There are three publicly available spontaneous micro-expression databases for recognition task: spontaneous micro-expression dataset (SMIC) [15], the Chinese Academy of Sciences Micro-Expression (CASME) [32]and CASME II [30]. These datasets have recorded micro-expression faces in frontal view. Table I lists the key features of existing micro-expression databases.

In order to evaluate the proposed architecture, we use two of these known databases, CASME and CASME II, as well as present an additional database which is an aggregation of CASME and CASME II, we will refer it as CASME+2 in our

TABLE I
DETAILS OF EXISTING SPONTANEOUS MICRO-EXPRESSION DATABASES

Dataset		Frame rate (Fps)	Subjects	Samples	Emotion class
SMIC	HS	100	20	164	3 (Positive, Negative, Surprise)
	VIR	25	10	71	
	NIS	25	10	71	
CASME		60	35	195	8 (Contempt, Disgust, Fear, Happiness, Regression, Sadness, Surprise, Tense)
CASME II		200	35	247	7 (Disgust, Fear, Happiness, Others, Regression, Sadness, Surprise, Tense)

article.

IV. PROPOSED METHOD

The conventional pipeline consists of four stages: 1) face detection; 2) pre-processing; 3) feature extraction; and 4) classification. Fig. 1 shows the basic block diagram of the micro-expression recognition. In this section, we describe the whole structure of our micro-expression recognition. We explain the method to increase the number of samples in the dataset using data augmentation in Section A. Section B discusses the preparation of the training, testing and validation datasets for training and validating the developed CNN model. Section C explains about the method applied for face detection and pre-processing, i.e. steps 1 and 2 of the block diagram. Section D represents steps 3, and 4 of the block diagram where the deep network does the feature extraction and classification. We developed CNN with variable depths to evaluate the performance of our model for facial micro-expression recognition.

A. Data Augmentation

The absence of large training datasets is a crucial bottleneck that keeps the utilisation of profound (deep) learning techniques in such cases, as the models will overfit drastically when utilising small training datasets. To address this issue, a large number of strategies have been proposed: fine-tuning models trained from other large public datasets (e.g. ImageNet [2]), using the big synthetic training datasets explored by some authors [8, 12, 14].

There are two critical points of interest of utilising synthetic data (i) one can produce the same number of training samples as required, and (ii) it permits explicit control over the unwanted factors. Data augmentation is a technique that is commonly used to reduce the scarcity problem. It is a set of label-preserving transforms that introduce some new instances without collecting the new data. In this, the existing training images are transformed without affecting the semantic class label. Examples of such transformations are horizontal/vertical mirroring [17], cropping, small rotations, etc. Flipping and

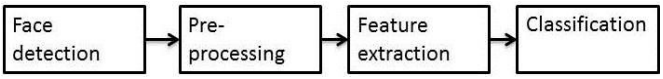

Fig. 1. General block diagram of micro-expression recognition

TABLE II
MICRO-EXPRESSION DATABASE AFTER DATA AUGMENTATION

Database	Original	After augmentation
CASME	26,423	52,846
CASME II	46,416	92,832
CASME+2	69,520	1,39,040

mirroring images vertically or horizontally producing two samples of each is a commonly used data augmentation technique for face recognition. In our evaluations, both original and vertically mirrored images are used for training. Table II demonstrates the data augmentation process has doubled the number of samples.

B. Data preparation

Training and testing of the model have carried on images from CASME, CASME II and CASME+2 databases which comprise of human faces; each labelled with one of 5 emotion categories: disgust, fear, happiness, sadness, and surprise. We have also taken into consideration the 6th category as 'neutral'. The given images are divided into two different sets which are training and testing sets. For data augmentation, mirrored images were generated by flipping images vertically. The training set comprises of 80% of the total images in the synthetic database and remaining 20% of the images are further divided as the testing set (10%) and validation set (10%).

C. Face detection and Pre-processing

We note that all the images are pre-processed so that they form a bounding box around the face region. Using the raw images from the CASME and CASME II database, we implemented the DLib face detector in OpenCV to detect and crop the face region from the raw image. The cropped face is then processed for head pose correction by computing the angle between the eye centroids and later applying the affine transformation. The transformed image is again passed to DLib face detector to crop and save the more accurate face region.

Fig. 2. shows the face detection and pre-processing steps. The face region is detected from the raw image frame given as the input. The Fig. 2. shows the red bounding box around the detected face region and the cropped face image. Since the frames are extracted from the videos, there is slight head movement observed in some of the videos. In the pre-processing step, the head pose correction method is applied to the cropped face. The head pose is aligned with the use of affine transformation, and the face region is cropped again to ensure that only the face region is given as input to the CNN model.

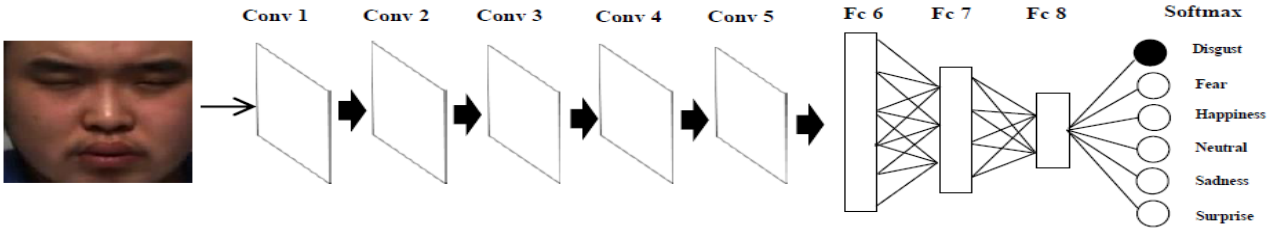

Fig. 3.The CNN architecture of our deep convolutional neural network

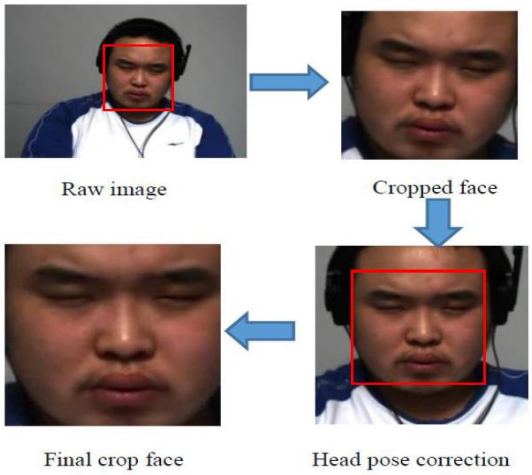

Fig. 2. Face detection and Pre-processing raw face images

D. *Convolutional Neural Network (CNN) framework*

Fig. 3. describes a basic framework for facial micro-expression recognition using the deep convolutional neural network. The pre-processed and cropped face image is passed as input to CNN model where the image has to pass through different layers of CNN: 1) Convolutional; 2) Rectified Linear Unit (ReLU); 3) Pooling or Sub sampling, and 4) Classification (Fully Connected Layer).

The primary purpose of the Convolutional step is to extract features from the input image. It maintains the spatial relationship between pixels by learning image features using small squares of input data and creates a feature map. ReLU is a non-linear operation. ReLU is a component wise operation (applied per pixel) and replaces all negative pixel values in the feature map by zero. Convolution is a linearity process which is element wise matrix multiplication and addition, so we represent non-linearity by presenting a non-linear function like ReLU. Spatial Pooling (also called subsampling or downsampling) reduces the dimensionality of each feature map but retains the most important information. In case of Max Pooling, the largest element from the rectified feature map within that window is taken.

Unitedly these layers select the useful features from the images, embed non-linearity in the network and reduce feature dimension while intending to make the features to some degree equivariant to scale and translation. The output of the third pooling layer acts as an input to the Fully Connected Layer. The Fully Connected Layer is a conventional Multi Layer Perceptron that uses a Softmax initiation function in the output layer (different classifiers like SVM can likewise be utilised,

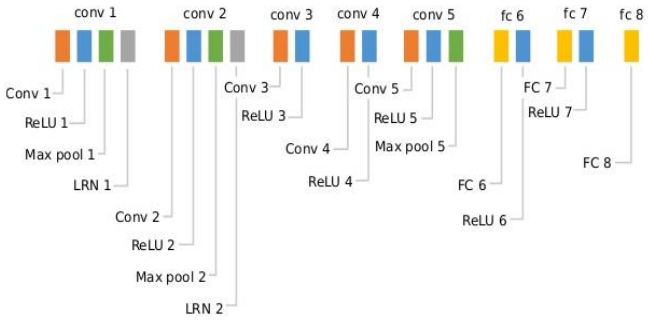

Fig. 4. Deep Network configuration

however we are utilising Softmax). The output of the convolutional and pooling layers constitute high-level features of the input image. The purpose of the Fully Connected layer is to utilise these features for classifying the input image into several classes based on the training dataset.

Putting it all together, the Convolution and Pooling layers act as Feature Extractors from the input image while the Fully Connected later serves as a Classifier.

The network contains five convolutional layers where the Conv1 layer has 11x11 filters and rest with 3x3 filters, three max pooling layers of kernel size 3x3, stride 2 and three fully connected layers (Fig. 4). The fully connected layers contain dropout, a mechanism for randomization which reduces the risk of the network over fitting. The Rectified Linear Unit (ReLU) was used for activation function. The system operates in two main phases: Training and Testing. During training, the system receives a training data comprising of images of faces with their respective micro-expression label. To ensure that the training performance is not affected by order of presentation of the examples, a few images are separated as a validation set. During testing, the images in the validation set are fed to the network which outputs the predicted micro-expression label using the final network weights learned during training. The base learning rate (base_lr) for the model is set to 0.001, stepsize parameter is set to 10,000, and maximum iterations (max_iter) are 100,000. The batch size of the images is 50 images per batch for training. We also changed the learning policy parameter (lr_policy) value to “step” where the learning rate will drop at every step size. The rest of the parameter settings were used the default. We used a baseline Softmax classifier.

The VGGFace is a network trained on a very large-scale face images dataset (2.6M images, 2.6k people) for the task of face recognition available from Visual Geometry Group at the University of Oxford [21]. Since the dataset was trained for a

similar application but on a much larger dataset than ours, we tried fine-tuning the model for micro-expressions.

V. Experiment and Results

The proposed micro-expression recognition verifies the effectiveness by conducting experiments on the CASME, CASME II and CASME+2 datasets. All the images in the databases are pre-processed and flipped vertically to increase the number of samples. The new synthetic database is then divided into two groups as Training and Testing. Each image has been categorized as: 0 = Disgust, 1 = Fear, 2 = Happiness, 3 = Neutral, 4 = Sadness, and 5 = Surprise.

We implemented the deep convolutional neural networks based on the Caffe [11] (a fast open framework for deep learning and computer vision) and took 10-12 hours to train this network. The models were trained for 100,000 iterations on CASME and 41,000 iterations on CASME II and CASME+2. The learning rate is changed from 0.001 to 0.0001 when the training iteration reaches 10,000. At each round of iterations in model training, the layer parameters of the network are updated based on the loss. We set a maximum number of iterations, and when the training time reaches the number, we obtain a trained model, which is essentially the parameter of all the filters. We then save the model so we can use the model to predict a micro-expression of images.

The input is given from the Validation sets which are the raw face images collected from the original databases. For each experiment, a corresponding Validation set is used depending on the training database. The confusion matrixes for each experiment are as shown in the Figures 5, 6 and 7.

The recognition accuracy results for the three databases used are summarised in Table III. From the table, we can observe that the recognition accuracy improves as the number

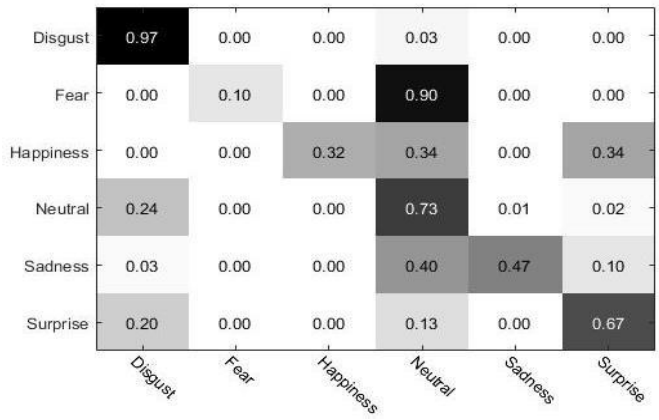

Fig. 5. Confusion matrix on CASME database

TABLE III
Micro-Expression Recognition Accuracy with Different Databases

Database	CASME	CASME II	CASME+2
Accuracy	74.25%	75.57%	**78.02%**

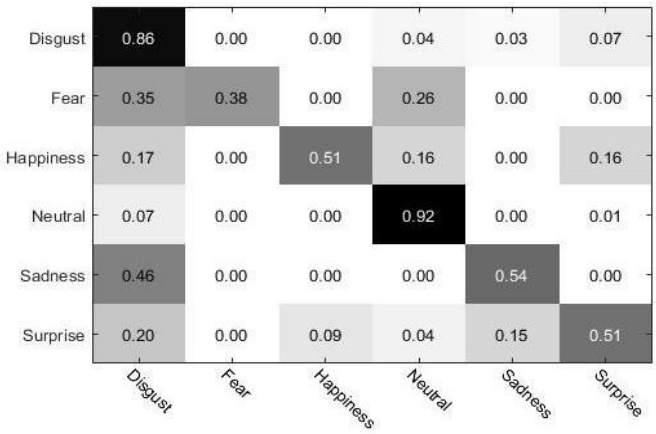

Fig. 6. Confusion matrix on CASME II database

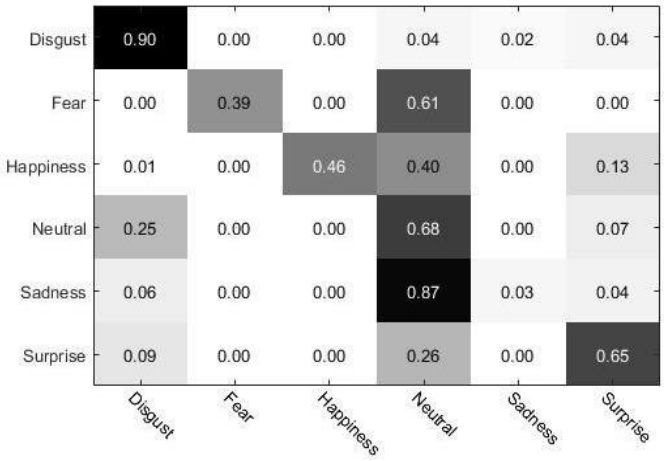

Fig. 7. Confusion matrix on CASME + CASME II database

of training samples increases. It can be interpreted from the results in Table II that the larger the training dataset is, the better the recognition accuracy is achieved.

Tables IV and V lists the recognition accuracy of using our method and of the state-of-the-art methods in CASME dataset and CASME II dataset respectively. Most of the existing methods are video based, which more or less take advantage of videos temporal information. Our method is image based, which applies CNN on image frames extracted from videos. The tables testify that proposed CNN method exhibits satisfying micro-expression recognition accuracy. These results also demonstrate that image based micro-expression recognition delivers identical results as video based approaches.

Due to larger number samples in CASME II dataset as compared to CASME dataset, the researchers in [12] opted to demonstrate deep learning results on video clips from CASME II dataset. In our research, we have applied data augmentation technique to increase the number of samples. Therefore, our experiments use both CASME and CASME II datasets to showcase the effectiveness of proposed CNN method.

TABLE IV
MICRO-EXPRESSION RECOGNITION ON CASME DATABASE

Method	Accuracy
LBP-TOP+ELM [6]	73.82%
MDMO+SVM [19]	68.86%
LBP-TOP+SVM [25]	61.85%
Proposed CNN method	**74.25%**

TABLE V
MICRO-EXPRESSION RECOGNITION ON CASME II DATABASE

Method	Accuracy
LBP-TOP+SVM [26]	75.30%
MDMO+SVM [19]	67.37%
CNN+LSTM [12]	60.98%
Proposed CNN method	**75.57%**

In [6, 12, 19, 25, 26], the researchers have considered the temporal factor from the video for recognition of micro-expressions which have contributed an additional feature in the calculations. In case of our method, we tried to eliminate the temporal factor and simply exhibit the image based micro-expression recognition approach.

A. Difficulties with certain expressions

We observe that some classes appear to be "harder" to train in the sense that, (1) none of our models were able to score highly on them, and (2) our model predictions for those classes tend to fluctuate depending on our training scheme and architecture. This appears to be the case for classes such as "fear", "happiness", and "sadness". There might be two reasons discussed as follows. First, these classes tend to have much fewer training samples compared to classes such as "disgust", "neutral", and "surprise", making it harder to train the CNN to recognise them. Second, these micro-expressions can be very nuanced, making it difficult even for humans to agree on their correct labelling [27].

We suspect that the inherent difficulty in assigning labels to some of the samples may have caused them to be "mislabelled", thereby affecting the models that were trained on them. Lastly, we note that all except one model (CASME II) were unable to predict a sufficient number of samples for label "fear" correctly. The reason for this could be an imbalance in the training datasets. The imbalance in the number of training samples for each class of micro-expressions most likely caused our models to overfit the micro-expressions with more samples (e.g. "disgust") at the expense of this class. Furthermore, the expression of fear is very subtle, which means that it will be hard for our CNN models to discover features to robustly distinguish this micro-expression from other similarly nuanced expressions such as sad, happiness, and surprise. This can be verified by examining the confusion matrices in Figure 3, 4 and 5, which indicates that the "disgust" class is often the highest scoring class. The classes "fear", "neutral", "happiness", "sadness", and "surprise" are often mistaken for each other by our models.

The combination of these two factors makes it even harder to train models to predict this micro-expression accurately. The above observations highlight the difficulty in training CNNs using a small unbalanced dataset with classes that are not visually distinctive.

VI. CONCLUSION

We have shown that it is possible to obtain a significant improvement in accuracy over the baseline results for micro-expression classification using CNNs pre-trained model utilised for the task for face recognition and fine-tuning it on face micro-expression databases. The experiments also conclude that the small sizes of the datasets do not favour them for being used for training CNNs. However, CNNs trained on sufficiently large face micro-expression datasets also can be used to obtain better results than the baseline without using the data augmentation technique to increase the size of the dataset artificially. This suggests that if we were to exploit deep neural networks such as CNN for face micro-expression recognition to achieve the significant gains seen in other domains, then having bigger datasets is crucial. This is where we implanted the idea of combining the two databases CASME and CASME II to form a larger database. Image based face expression recognition is a popular research topic, but we demonstrated through our experiments that image based micro-expression recognition could also yield acceptable accuracy.

Lastly, we also noted the inherent difficulty in assigning correct labels to faces depicting some of the more nuanced micro-expressions and how that can affect the performance of our models.

REFERENCES

[1] G. K. Chavali, S. K. N. Bhavaraju, T. Adusumilli, and V. Puripanda, "Micro-Expression Extraction For Lie Detection Using Eulerian Video (Motion and Color) Magnication," 2014.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255.

[3] P. Ekman, *Micro Expressions Training Tool*: Emotionsrevealed. com, 2003.

[4] J. Endres and A. Laidlaw, "Micro-expression recognition training in medical students: a pilot study," *BMC Medical Education*, vol. 9, p. 47, 2009.

[5] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, 1978.

[6] Y. Guo, C. Xue, Y. Wang, and M. Yu, "Micro-expression recognition based on CBP-TOP feature with ELM," *Optik-International Journal for Light and Electron Optics*, vol. 126, pp. 4446-4451, 2015.

[7] C. House and R. Meyer, "Preprocessing and Descriptor Features for Facial Micro-Expression Recognition," 2015.

[8] G. Hu, X. Peng, Y. Yang, T. Hospedales, and J. Verbeek, "Frankenstein: Learning deep face representations using small data," *arXiv preprint arXiv:1603.06470*, 2016.

[9] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, *et al.*, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in

Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 142-150.

[10] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns," *Neurocomputing*, vol. 175, pp. 564-578, 2016.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675-678.

[12] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations," in *Proceedings of the 2016 ACM on Multimedia Conference*, 2016, pp. 382-386.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.

[14] W. Li, M. Li, Z. Su, and Z. Zhu, "A deep-learning approach to facial expression recognition with candid images," in *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, 2015, pp. 279-282.

[15] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, 2013, pp. 1-6.

[16] X. Li, H. Xiaopeng, A. Moilanen, X. Huang, T. Pfister, G. Zhao, *et al.*, "Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods," *IEEE Transactions on Affective Computing*, 2017.

[17] X. Li, J. Yu, and S. Zhan, "Spontaneous facial micro-expression detection based on deep learning," in *Signal Processing (ICSP), 2016 IEEE 13th International Conference on*, 2016, pp. 1130-1134.

[18] S.-T. Liong, J. See, K. Wong, A. C. Le Ngo, Y.-H. Oh, and R. Phan, "Automatic apex frame spotting in micro-expression database," in *Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on*, 2015, pp. 665-669.

[19] Y. J. Liu, J. K. Zhang, W. J. Yan, S. J. Wang, G. Zhao, and X. Fu, "A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1-1, 2015.

[20] Y. Lv, Z. Feng, and C. Xu, "Facial expression recognition via deep learning," in *Smart Computing (SMARTCOMP), 2014 International Conference on*, 2014, pp. 303-308.

[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *BMVC*, 2015, p. 6.

[22] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *2011 International Conference on Computer Vision*, 2011, pp. 1449-1456.

[23] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on*, 2009, pp. 1-6.

[24] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708.

[25] S. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, "Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space," in *ICPR*, 2014, pp. 4678-4683.

[26] Y. Wang, J. See, Y.-H. Oh, R. C.-W. Phan, Y. Rahulamathavan, H.-C. Ling, *et al.*, "Effective recognition of facial micro-expressions with video motion magnification," *Multimedia Tools and Applications*, pp. 1-26, 2016.

[27] S. C. Widen, J. A. Russell, and A. Brooks, "Anger and disgust: Discrete or overlapping categories," in *2004 APS Annual Convention, Boston College, Chicago, IL*, 2004.

[28] Q. Wu, X. Shen, and X. Fu, "The machine knows what you are hiding: an automatic micro-expression recognition system," in *International Conference on Affective Computing and Intelligent Interaction*, 2011, pp. 152-162.

[29] F. Xu, J. Zhang, and J. Wang, "Microexpression Identification and Categorization using a Facial Dynamics Map," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1-1, 2016.

[30] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, *et al.*, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, p. e86041, 2014.

[31] W.-J. Yan, S.-J. Wang, Y.-H. Chen, G. Zhao, and X. Fu, "Quantifying micro-expressions with constraint local model and local binary pattern," in *Workshop at the European Conference on Computer Vision*, 2014, pp. 296-305.

[32] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, 2013, pp. 1-7.