Microe for Responsible Software Systems with Emotional Intelligence

Patricia Kirabo Nakalembe
Department of Computer Science
Makerere university
Kampala, Uganda
nakalembe.patricia@students.mak.ac.ug

Calvin Kirabo
Department of Computer Science
Makerere university
Kampala, Uganda
kirabocalvin@students.mak.ac.ug

Edwin Micheal Kibuuka
Department of Computer Science
Makerere university
Kampala, Uganda
michaeledwin.kibuuka@students.mak.ac.ug

Derrick Ahumuza
Department of Computer Sciencey
Makerere university
Kampala, Uganda
ahumuza.derrick@students.mak.ac.ug

Nakayiza Hellen
Department of Computer and Information Sciences
Muni university
Arua, Uganda
h.nakayiza@muni.ac.ug

Ggaliwango Marvin
Department of Computer Science
Makerere university
Kampala, Uganda
ggaliwango.marvin@mak.ac.ug

***Abstract*—Micro expressions are facial expressions that occur for a short period of time, usually less than a second. These expressions usually go unnoticed but represent the true expression that is felt by the person because they are involuntary. There have been recent attempts to make datasets that contain micro expressions in an attempt to carry out accurate Facial Emotion Recognition. However, many of these datasets have had various limitations, ranging from low diversity of the dataset which introduced bias into the emotion detection models, low quality images, a limited number of images that resulted into low accuracy of the model, etc. In this paper, we introduce a dataset called Microe dataset which contains images that were collected from various sources to ensure a high diversity and more features that are going to be discussed in sections below. Experiments were also carried out to evaluate this dataset and the results that were acquired have been included in this paper.**

***Index Terms*—Responsible Software Systems, Micro expressions, Emotion recognition, Dataset, Facial expressions**

I. Introduction

Micro expressions are one of two categories of facial expressions; the other being macro expressions. Unlike macro expressions which are voluntary and last longer, micro expressions are known to be involuntary and subtle facial movements which usually last less than a second [1], [2]. It is important to recognize these subtle facial changes as they usually depict the instant emotion that is felt by a person when encountering a given situation since these expressions cannot be suppressed. Detection of these expressions can be important in aspects such as; lie detection, criminal interrogation, autism, depression analysis, psychoanalysis, etc [3]. After constant evolution in the field of emotion detection, many researchers have attempted to accurately identify these micro expressions. However, despite their importance, detection of these expressions is challenging and therefore different approaches have been proposed to assist in this task. Due to their ability to process input at multiple layers, Deep Learning methods have been adopted by the majority of researchers to understand the complex patterns in micro expression detection with the most common one being Convolutional Neural Networks as used by [4], [5], [6], [7] and [8]. Micro Expression (ME) datasets i.e. datasets that consist of images that can be used for micro expression detection are at the centre of the emotion recognition process as they determine the accuracy of the emotion recognition model. If these datasets are carefully made to include all the different aspects that may make the model more accurate in the real-world scenario e.g. correct labels, adequate number of images, etc, the datasets are considered to be properly established and can be used for different models. Different researchers have used various techniques to come up with these datasets some of which include; using stimuli e.g. videos, discussing emotional or personal topics [9], asking people to represent various emotions, inducing stress or pressure on people and observing them [9], etc. All these various techniques have been adopted to trigger emotions in people and in turn capture micro expressions. However, despite these various attempts to create micro expression datasets, there have been multiple gaps which have in turn affected the accuracy of models which are trained using these datasets.[10] states “small dataset size” as one of the main constraints in the existing ME datasets, for example SMIC dataset which consists of only 164 samples[9] , CASME which consists of only 195 samples[11] and CASME II [12] which only consists of 247 samples, and this leads to few training samples for the model. [11] further states unnatural micro expressions as one of the problems in the existing ME datasets. This is due to the environment setup requiring participants to force the expressions and this removes the entire meaning of micro expressions since they are supposed to be involuntary. [12] also states poor video quality as a problem faced since micro expressions are low in intensity and for proper detection, require very high-quality images and/or videos. [13] highlights another issue concerned with ME datasets that is known as imbalances between different micro expression classes i.e. Some classes having way more samples than others. Lastly, [14] talks about the inconsistency in labelling used in Micro Expression datasets which leads to inaccurate results. In conclusion, our dataset has been created in an attempt to address the gaps in the existing ME datasets, some of which have been stated above.

II. DATASET DESCRIPTION

The Microe dataset is a collection of images of size 48x48 which represent the different emotions we intend to detect. It contains a total of 41000 observations. For simplicity, the dataset was organized into 8 separate categories i.e. angry, contempt, disgust, fear, happy, neutral, sad and surprised. It should also be noted that some classes have fewer samples of data as compared to others for example contempt, which has the least number of samples.

This dataset is going to be used to train our model to help it learn the key points to look out for in each micro expression. This will help the model to learn the different micro expressions and in turn be used to identify these expressions in later stages of our project. It is made up of images from different sources such as Kaggle, public datasets, various dataset repositories, etc. that were made accessible to everyone by their owners. These were collected and combined in a period of approximately 30 days.

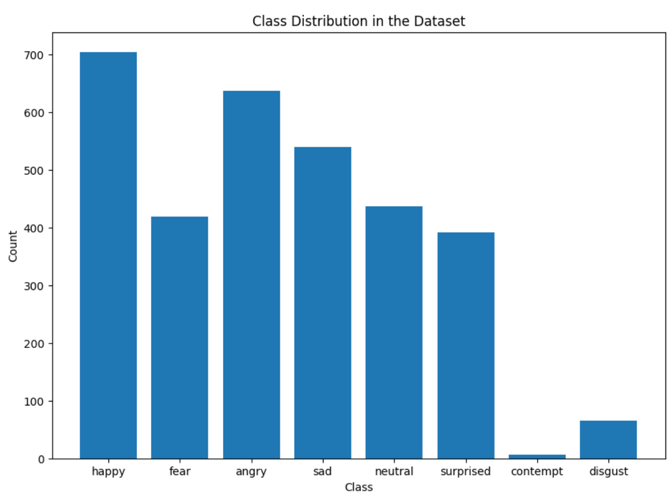

Fig. 1. Class distribution in the dataset.

From the graph above we see that: The x-axis represents the classes we used for the dataset and the y-axis represents the ranges of images in those particular classes. In general, this graph provides a visual analysis of the balance of images in the dataset.

Using this visualisation, Class-specific patterns can be identified by examining mean images. For example If mean images are visually distinct, it suggests proper labeling and assignment of the images in their given classes. On the other hand, similar mean images could indicate potential mislabeling or inconsistencies which can aid researchers and data scientists in their subsequent analysis.

Fig. 2. Sample of images in different classes

The figure above represents a small sample of images that are contained in the given classes. The faces represent individuals from different races and this provides a representation of how diverse our dataset is. Through this inclusivity, we see that each face possesses distinctive features that the model learns from and which makes it more applicable in real world scenarios

III. METHODOLOGY

Our data was gathered through secondary data acquisition, which means we looked at data that was previously gathered and utilized by other users on Kaggle, the Computer Vision and Pattern Recognition (CVPR) dataset repository, ADE20K, ImageNet, VisualData, Open Access Series of Imaging Studies (OASIS), Google Dataset Search, AWS Public Datasets, Visual Genome, and COCO (Common Objects in Context) and then selected the data that was relevant to our work. This was then combined to form this single dataset.

IV. DATA PROCESSING AND VALIDATION

To prepare our collected images for analysis, a couple of steps were taken and these include:

A. Cleaning

We combined several datasets into a single, consistent dataset and assigned each image to a certain category in order to clean our data. To make things easier for the model, we resized every image and gave them a consistent size i.e. 48x48.

B. Transformation

To lessen the possibility of bias, the data was altered so that: Every image was turned to grayscale. Additionally, indexes are assigned to the images, which can be utilized to load a specific image. Image normalization was done so that the value of each pixel is divided by 255.0. It is crucial to do this normalizing phase to guarantee that the picture values are

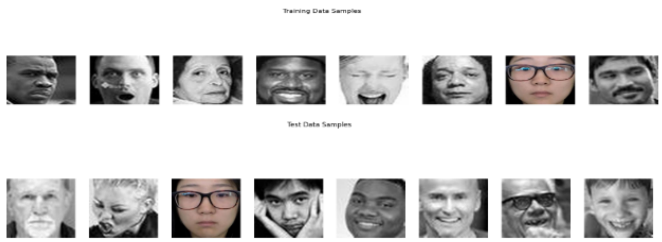

Fig. 3. Train and test samples

comparable and that the model does not become dependent on the image brightness.

The figure above contains sample images from the train dataset after the splitting randomly from the 8 classes of emotions and sample images from the test dataset after the splitting randomly from the 8 classes of emotions. As we can observe, majority of the images have been converted into grayscale to remove potential bias.

C. Handling Missing Data

Since we made sure that every picture was categorised correctly and there were 41000 photos overall, there was never a case of missing data.

D. Data Augmentation

We have partitioned our dataset so that 20/100 is test data and 80/100 is train data. Parts of the 80/100 of our training data undergo modifications that can assist our model be trained on the various input data changes it may encounter, thus increasing its accuracy. Among the adjustments performed to our training data are the following: blurring, rotating the photos to specific degrees, flipping the images horizontally, and other enhancements. The final dataset used to train our model is created by combining the altered datasets with the original dataset

E. Data Labelling

Prior to labelling our data, we had to choose the labels—that is, the emotions we wanted our model to learn—for our data. Anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise are a few of these. We did not utilize any labelling tools because our dataset was derived from pre-existing datasets. Instead, we aggregated photographs from other datasets that belonged to the same category and labelled them together. We repeated this process for each of the eight labels until each label had an image associated with it.

F. Quality Assurance

To guarantee precision and uniformity in the labelling procedure, we employed several methods such as double-checking each label and having different team members review the labels and photos. This made it easier for us to verify that the labels on the photos were accurate.

V. EXPERIMENTS

In this section we show the various experiments that we carried out to determine the effectiveness of our dataset in image classification. Some of these experiments were carried out using subsets of the dataset according to the resources that we had within our means. This section has the primary goal of evaluating the dataset's suitability for training and evaluating image recognition models.

A. statistical analysis of the dataset

In this section we analyze different parts of the dataset from which we can gain valuable insights.

TABLE I
TABLE SHOWING IMAGE DIMENSIONS

Training Image Dimensions Statistics	Test Image Dimensions Statistics
Mean Height: 224.0	Mean Height: 224.0
Median Height: 224.0	Median Height: 224.0
Standard Deviation of Height: 0.0	Standard Deviation of Height: 0.0
Min Height: 224	Min Height: 224
Max Height: 224	Max Height: 224
Mean Width: 224.0	Mean Width: 224.0
Median Width: 224.0.0	Median Width: 224.0
Standard Deviation of Width: 0.0	Standard Deviation of Width: 0.0
Min Width: 224	Min Width: 224
Max Width: 224	Max Width: 224

From the table above: The dataset set is run through a transformer which resizes the images to enforce uniformity and data loaders which store the transformed images are generated hence the images are prepared and passed through the models when standardized to 224x224 images hence height and width the same .This therefore leads to the mean being the same and there being no deviation

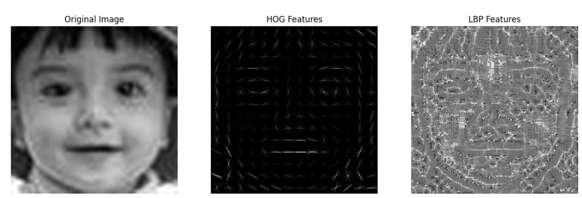

Fig. 4. Image showing HOG and LBP features

HOG(Histogram of Gradients) and LBP(Local Binary Pattern) are feature extraction techniques used in computer vision for tasks like image classification and object detection. While statistical analysis can provide general insights into your image dataset(e.g size distribution, color histograms), HOG and LBP offer a more specific way to analyze content of your images. This is so because they can capture informative Features, Quantitative Representation, Dimensional Reduction. Our dataset comprises of images which worked well with these techniques in helping in classification a seen above in a sample image of an original and its HOG and LBP version with features clearly seen.

Similarly the dataset images also respond perfectly to the edge detection which can be further used for evaluation and feature selection as shown in the figure above.

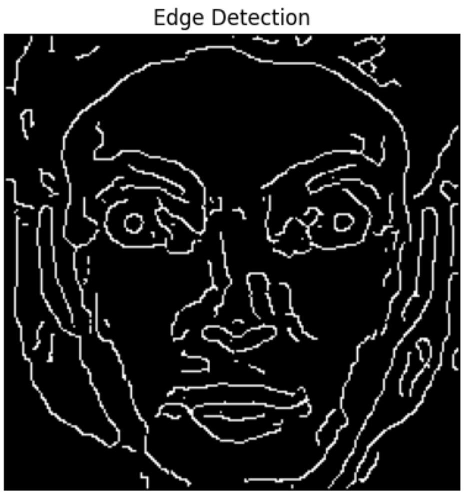

Fig. 5. Image showing Edge detection

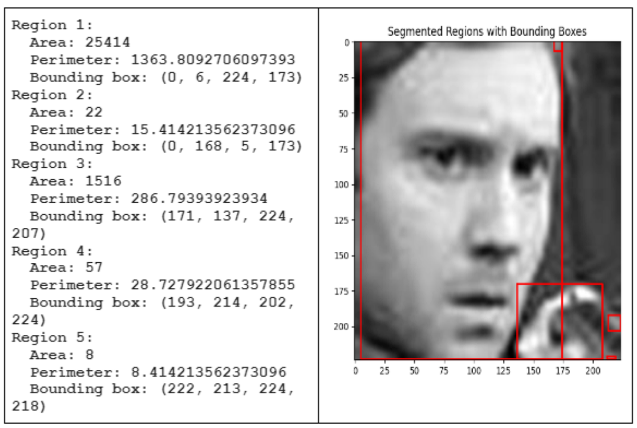

Fig. 6. Image region properties of the given image.

In regards to the figure above:

We performed various experiments and analysis on the dataset images on is the example below where the image was divided into 5 region and each part is as follows: area: represents the total number of pixels within the region a larger value corresponds to a larger connected region Perimeter: This represents the total length of the boundary surrounding the region, in pixels Bounding box: This defines the rectangular area that completely encloses the region. It's specified by the top-left corner coordinates(X.Y0 and the bottom-right corner coordinates(X,Y).

Color space statistics

Fig. 7. Dominant colors in the dataset

The figure below is proof that most of the images of the dataset are already gray scale hence the black bar when you try to plot color dominance. Hence little work is needed in converting them to gray scale before using them in most of the models

Overall: These statistics strongly suggest that your image dataset is dominated by gray scale images or images with very limited color variations. The average colors are close to mid-gray, and the standard deviation and percentiles indicate a narrow range of color intensities across all the three channels(red,green, blue)

B. AlexNet

AlexNet is a convolutional neural network (CNN) architecture that significantly advanced the field of computer vision and deep learning. The model has 8 layers i.e. Input Layer, 4 Convolutional Layers, Flattening, Fully Connected Layers, Output Layer through which the images of 8 emotional classes are passed for training. To get started, the dataset was split into training, validation and testing sets using a common split ration where training took 0.7, validation took 0.2 and testing took 0.1. The images were also resized to a fixed size appropriate to Alex Net's input layer i.e. 227x277 pixels.

1) Training AlexNet: The AlexNet is evaluated on the validation set after training on the training split to look for overfitting. Early stopping was implemented so that training may be stopped if validation performance peaked or decreased.

Below is the training and validation graph :

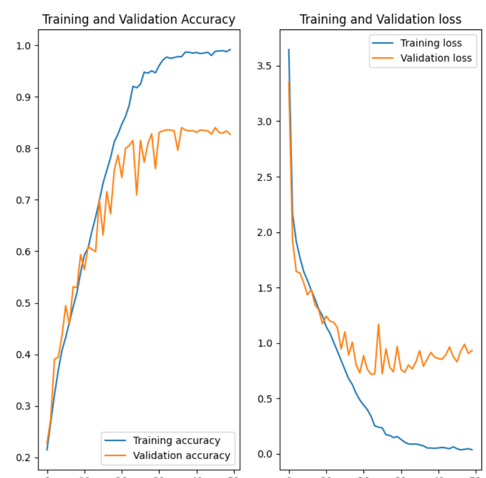

Fig. 8. Training and validation graph

From the figure above:

a) Training Accuracy: The Model is efficiently learning from the training data because the training accuracy curve is rising throughout the course of the epochs. b) Validation Accuracy: Despite increasing more slowly than the accuracy curve, the validation curve is also increasing. This indicates the presence of some overfitting, in which the model is picking up patterns from the training set that may not transfer well to unseen or new data. c) Training loss: The training loss is decreasing as expected as the model improves. d) Validation Loss: Likewise, the validation loss curve is decreasing but it starts to deviate slightly from the training loss curve around epoch 20, which further suggests overfitting.

Below are some predictions made by the model:

Fig. 9. Sample predictions made by the model

The model predicts the emotion class of the images from the validation set of the dataset. Figure above shows a sample of the images, their actual labels, and the labels predicted by the model, along with the corresponding confidence percentages. This information can be used to evaluate the performance of the model and identify any areas where it may need improvement.

TABLE II
TABLE SHOWING THE CLASSIFICATION REPORT OF THE ALEX MODEL

	precision	recall	f1-score	support
anger	0.81	0.79	0.80	58
contempt	0.78	0.64	0.70	11
disgust	0.83	0.65	0.73	23
fear	0.78	0.93	0.84	41
happy	0.91	0.98	0.94	95
neutral	0.82	0.88	0.85	32
sad	0.83	0.74	0.78	46
surprise	0.90	0.84	0.87	55
Accuracy			0.85	361
Macro avg	0.83	0.80	0.81	361
Weighted avg	0.85	0.85	0.85	361

The table above shows:

The classification report of the Alex model provides a comprehensive evaluation of the model's performance on the validation set. The report includes various metrics such as precision, recall, f1-score, and support for each emotion class. Precision measures the proportion of images correctly classified as a particular emotion among all images predicted as that emotion. Recall measures the proportion of images correctly classified as a particular emotion among all images with that emotion in the validation set. F1-score is the harmonic mean of precision and recall, providing a balanced view of both metrics. Support indicates the number of images in the validation set with a particular emotion.

Overall, the Alex model achieves an accuracy of 83/100 on the validation set, indicating good performance in classifying emotions from images. Among the emotion classes, "happy" has the highest precision (0.91) and recall (0.98), demonstrating the model's ability to accurately identify happy images. "Contempt" has the lowest recall (0.64), suggesting that the model may struggle to identify images expressing contempt. The model's performance is relatively balanced across different emotion classes, as indicated by the similar values of precision, recall, and f1-score for most classes.

VI. RESULTS AND DISCUSSION

The AlexNet model performs with an accuracy of 83/100 on the dataset. This therefore means that the dataset is suitable as it provides a good representation of the real-world in regards to classification of images and it helps the model learn the different facial features used in tasks of emotion recognition

VII. FAIR COMPLIANCE

The FAIR principles are guidelines that are used to address the challenges of data sharing and use in research. Our dataset complies with these principles in the following ways;

Findable: Our dataset is to be made available under the Creative Commons License to make it easy for people to find it and once this happens, metadata related with the dataset will also be publicly available. This metadata includes information such as; title of the dataset, the creators, the date of publishing, keywords, and a concise description. The dataset will also have a DOI (Digital Object Identifier) link for easy citation. **A**ccessible: When available, our dataset will be under an open access licence i.e. Creative Commons License to make it easy for researchers. The dataset is also in a commonly used format i.e. "csv" which will make it easy for researchers to download it. Alternatively, if only a few images in the dataset are required, the images are in a common image format and they will be easily accessible. Furthermore, the images are indexed for better access.

Interoperable: Due to the use of commonly used formats, our dataset will be compatible with various platforms as required by the researcher. The dataset also works with popular libraries and tools for data analysis, including: Python libraries (Pandas, OpenCV, Dlib), machine learning frameworks (TensorFlow, PyTorch) and visualisation tools (Matplotlib, Seaborn)

Reusable: Since the Creative Commons license is to be used, researchers will be able to understand the terms of reuse and modification of the dataset. All the various image sources will also be available, some of which include Kaggle, public datasets, etc. Problems like outliers, inconsistent labelling, and missing values have been taken care of to guarantee the quality of your data. Basic data cleansing and validation has also been done and this will be advantageous to researchers when the dataset is shared.

The dataset metadata also includes information such as; image format, image size, label definitions, image sources, etc. It also includes the total number of images and how they are distributed in the 8 classes. If possible, the dataset is to be hosted in a public and open repository and clear access procedures will be provided there. The labels used i.e. Angry, contempt, disgust, fear, happy, neutral, sad and surprised are widely used and they align with existing emotion classification standards and the dataset is compatible with popular machine learning libraries such as TensorFlow. Creative Commons licence will allow users to download, alter, and distribute the information in accordance with certain specifications. By adhering to these points, the Microe dataset can be reused in various micro-expression recognition studies for example constructing and assessing machine learning models for the classification of micro expressions, comparing the effectiveness of various algorithms for the identification of micro expressions, creating fresh methods to address the imbalance of classes in micro-expression datasets, examining the efficiency of various data augmentation methods for the identification of micro expressions and investigating the connection between various micro-expressions and particular facial traits.

VIII. Applications and Use Cases

Our dataset, consisting of various facial images, can be utilised in areas or fields that utilise facial expressions to enhance the accuracy of their operations. Some of these fields where our dataset can be used include: online learning systems, lie detectors, clinical psychology, pain assessment, psychiatry, etc. Due to the large diversity in the dataset, elimination of bias, proper and careful data processing carried out by the team, we can safely conclude that our dataset is accurate and consistent which makes it more suitable than majority of previously created datasets as it addresses the gaps faced in those datasets. Therefore, when utilised, this dataset presents potential advancement in emotion recognition studies in the form of higher accuracy.

IX. Review and Perspective

As previously indicated, different researchers have come up with a number of datasets with the main aim of micro expression detection. These researchers have used different approaches to come up with these datasets with some of them using images and others using videos. However, despite all these efforts, the majority of these have not been successful due to the gaps in them which limit model accuracy. As seen earlier, datasets such as SMIC, CASME and CASME(II) have been seen to show some of the gaps, such as limited dataset size, limited or no diversity, poor quality of images or videos, etc.

In comparison to these datasets, our dataset aims to address some of these gaps by providing a very large dataset size with a large range of diversity in terms of gender, ethnicity, age, etc. This makes our dataset more suitable for use in this field as it provides potential for higher accuracy of the model. Emotion recognition, specifically facial micro expression studies have a lot of potential for improvement that when exploited can further develop the emotion recognition field and include it in a number of other fields to yield better results. With the daily advancement in technology, there is potential for high resolution cameras to improve quality of images and videos, integration of emotion recognition with wearable devices, use of automation in emotion recognition, better environments for micro expression collection, etc.

Our dataset can help facilitate the future research by providing potential for various experiments to be carried out on the dataset using various approaches which can help provide deeper insights, different viewpoints and laying the groundwork for the creation of a more dependable, ethical, and accurate emotion identification system that can be applied in different fields such as healthcare, security, marketing, etc.

X. Data Accessibility

Since our dataset isn't currently publicly accessible, if someone outside of our team finds a way to access it and utilises it, they must give us credit for being the dataset's proprietors. However, using the previously stated collection methods, one should be in position to come up with a similar dataset if desired.

Furthermore, in regards to licensing, the dataset will be made available in the near future for usage under Creative Commons Licences, which permit anyone to use it as long as they give us credit for being the original creators. Once made available, the dataset citation information will be shared including the creators, dataset name, date of publication, etc. A DOI link will also be provided to make citation easier for users.

XI. Conclusion

Based on the experiments carried out on the dataset and the results we attained, we can conclude that the dataset has a high potential to be used in emotion recognition tasks. The high accuracy attained by the AlexNet model when it was trained on the dataset represents unique qualities in the dataset for example a rich data collection and focusing specifically on micro expressions. The dataset also adheres to the FAIR principles i.e. it is Findable, Accessible, Interoperable and Reusable which makes it more efficient for research. Future work on the dataset can explore its effectiveness with other deep learning models which can then be applied in fields such as security, Human Computer Interaction (HCI) and psychology.

Acknowledgment

We would like to acknowledge each of the creators or contributors who put in time and effort to collect and test the dataset to ensure that it meets the desired standards. Special acknowledgement also goes out to the CoCIS Seedt grant funding for believing in our team and providing all the necessary support that enabled us reach this far.

REFERENCES

[1] Y. Fang, D. Deng, L. Wu, F. Jumelle and B. Shi, "RMES: Real-Time Micro-Expression Spotting Using Phase From Riesz Pyramid," 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 2023, pp. 222-227, doi: 10.1109/ICME55011.2023.00046. keywords: Smoothing methods;Pipelines;Noise reduction;Real-time systems;Noise measurement;Feeds;Time complexity;Facial expression recognition;emotion detection;image pyramid;CNN,

[2] X. Fan, X. Chen, M. Jiang, A. R. Shahid and H. Yan, "SelfME: Self-Supervised Motion Learning for Micro-Expression Recognition," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 13834-13843, doi: 10.1109/CVPR52729.2023.01329. keywords: Computer vision;Face recognition;Pipelines;Self-supervised learning;Benchmark testing;Feature extraction;Transformers;Humans: Face;body;pose;gesture;movement,

[3] M. Verma, P. Lubal, S. K. Vipparthi and M. Abdel-Mottaleb, "RNAS-MER: A Refined Neural Architecture Search with Hybrid Spatiotemporal Operations for Micro-Expression Recognition," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 4759-4768, doi: 10.1109/WACV56688.2023.00475. keywords: Training;Computer vision;Convolution;Microprocessors;Stacking;Computer architecture;Benchmark testing;Algorithms: Image recognition and understanding (object detection;categorization;segmentation);Biometrics;face;gesture;body pose;Video recognition and understanding (tracking;action recognition;etc.),

[4] S. Indolia, S. Nigam and R. Singh, "An Optimized Convolution Neural Network Framework for Facial Expression Recognition," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 93-98, doi: 10.1109/ICIIP53038.2021.9702639. keywords: Deep learning;Training;Emotion recognition;Image recognition;Convolution;Face recognition;Neural networks;Facial expression recognition;deep learning;hyperparameter tuning;data augmentation;convolutional neural network,

[5] A. Halim et al., "Facial Expressions Analysis To Evaluate The Level Of Students' Understanding," 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 2023, pp. 424-429, doi: 10.1109/IMSA58542.2023.10217489. keywords: Human computer interaction;Ethics;Computer vision;Emotion recognition;Face recognition;Focusing;Iron;ME;FE;FER;FEA;FER2013;AffectNet-8;ResNet50-CBAM;CNN;DCNN-CBAM,

[6] A. J. Rakesh Kumar and B. Bhanu, "Micro-Expression Classification based on Landmark Relations with Graph Attention Convolutional Network," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 1511-1520, doi: 10.1109/CVPRW53098.2021.00167. keywords: Computer vision;Conferences;Feature extraction;Facial muscles;Pattern recognition;Optical flow,

[7] Y. Fang, D. Deng, L. Wu, F. Jumelle and B. Shi, "RMES: Real-Time Micro-Expression Spotting Using Phase From Riesz Pyramid," 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 2023, pp. 222-227, doi: 10.1109/ICME55011.2023.00046. keywords: Smoothing methods;Pipelines;Noise reduction;Real-time systems;Noise measurement;Feeds;Time complexity;Facial expression recognition;emotion detection;image pyramid;CNN,

[8] E. Pei, M. C. Oveneke, Y. Zhao, D. Jiang and H. Sahli, "Monocular 3D Facial Expression Features for Continuous Affect Recognition," in IEEE Transactions on Multimedia, vol. 23, pp. 3540-3550, 2021, doi: 10.1109/TMM.2020.3026894. keywords: Face recognition;Three-dimensional displays;Feature extraction;Solid modeling;Emotion recognition;Optical imaging;Strain;3D morphable model;3D scene flow;continuous affect recognition,

[9] X. Li, T. Pfister, X. Huang, G. Zhao and M. Pietikäinen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 2013, pp. 1-6, doi: 10.1109/FG.2013.6553717. keywords: Databases;Cameras;Films;Motion pictures;YouTube;Feature extraction;Face recognition,

[10] X. Li, et al.,"4DME: A Spontaneous 4D Micro-Expression Dataset With Multimodalities" in IEEE Transactions on Affective Computing, vol. 14, no. 04, pp. 3031-3047, 2023, doi: 10.1109/TAFFC.2022.3182342keywords: videos;face recognition;labeling;gold;three-dimensional displays;task analysis;computer vision,

[11] Yan, Wen-Jing , Wu, Qi , Liu, Yong-Jin , Wang, Su-Jing , Fu, Xiaolan. (2013). CASME Database: a dataset of spontaneous micro-expressions collected from neutralized faces. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013. 10.1109/FG.2013.6553799.

[12] [12] Yan WJ, Li X, Wang SJ, Zhao G, Liu YJ, Chen YH, Fu X. CASME II: an improved spontaneous micro-expression database and the baseline evaluation. PLoS One. 2014 Jan 27;9(1):e86041. doi: 10.1371/journal.pone.0086041. PMID: 24475068; PMCID: PMC3903513.

[13] [13] Y. Wang, H. Shi and R. Wang, "Action Decouple Multi-Tasking for Micro-Expression Recognition," in IEEE Access, vol. 11, pp. 82978-82988, 2023, doi: 10.1109/ACCESS.2023.3301950.keywords: Feature extraction;Task analysis;Deep learning;Optical flow;Face recognition;Multitasking;Convolutional neural networks;Micro-expression recognition;Contrastive learning;Multi-task learning;Decouple method,

[14] [14] Y. Dai and L. Feng, "Cross-Domain Few-Shot Micro-Expression Recognition Incorporating Action Units," in IEEE Access, vol. 9, pp. 142071-142083, 2021, doi: 10.1109/ACCESS.2021.3120542.keywords: Feature extraction;Face recognition;Gold;Task analysis;Image recognition;Optical imaging;Emotion recognition;Cross-domain few-shot learning;facial action unit;micro-expression recognition;multimodal feature fusion,