Survey on emotion detection with partially occluded facial images

Sutharshan Rajasegarar, *Member, IEEE*, Nimantha Weerasooriya, and Selvarajah Thuseethan, *student, IEEE*

***Abstract*—Facial expressions can be considered as the best way to express emotions among other human expressions, voices, and body gestures in daily communication. Recognizing facial emotions is crucial in numerous areas such as human-computer interaction, healthcare, security, teaching, social media, etc. the majority of completed studies in facial emotion recognition are based on frontal and non-occluded faces. Recognizing emotions from occluded faces and non-frontal faces is a challenging and decisive task in the emotion detection area. It can be found considerable types of face occlusions diminish the capability of facial emotion detection. This survey paper is an overview of recent facial emotion recognition studies conducted with occluded faces. In this survey, we have categorized some recent studies into three groups, facial emotion detection from partially occluded faces, groups and multiple faces and masked faces. The purpose of this survey is to provide an encyclopedic review of existing researches and identify the existing challenges in facial emotion recognition from occluded faces. Moreover, we are presenting some challenges that need to be addressed in the future at the end of this survey.**

***Index Terms*—Facial emotion recognition, occluded faces, group emotion, crowd emotion.**

I. Introduction

In daily human communication, people express their emotions through verbal and non-verbal expressions. Facial expressions and body gestures can be considered as non-verbal expressions and voices can be considered as an example of verbal expressions. According to Mehrabian, [1] the capability to convey emotions through facial expression is higher than verbal expressions and body gestures. In his research, it has stated that the facial expressions, body gestures, and vocal expressions have 55%, 38%, and 7% ability of emotion conveyance respectively.

In facial expressions, the human emotions represent on the face by facial changes. Identifying the emotion by these facial changes is vital for numerous fields nowadays. Facial emotion recognition is used in the health care sector to diagnose various illnesses. For instance, in [2], facial emotion recognition has been used to diagnose the mental health of college students. Moreover, research for predicting Alexithymia has been done by using facial emotion recognition in [3]. During the Covid-19 pandemic, the education sector has moved to virtual platforms to continue teaching students without gathering at schools. The research [4] presents a facial emotion recognition system to help teachers to change their teaching techniques according to students' emotions.

Sutharshan Rajasegarar, Nimantha Weerasooriya, Selvarajah Thuseethan are with the School of Information Technology, Deakin University, Geelong, Australia. Emails: srajas@deakin.edu.au, nimanthadw@gmail.com, tselvarajah@deakin.edu.au

Emotion recognition is more challenging when faces are occluded in various ways. The survey, [5] published in 2019, has mentioned that complicated conditions such as occlusion and pose-variation will be the main obstacles to facial emotion recognition in the future. The researchers are concerned about this challenge and they have implemented some methods to minimize the impact of the occlusions. The survey [6], has been done to compare the existing methods of facial expression recognition under changing emotions. Motion-based, model-based, muscles-based, and Hybrid Approaches were analyzed and compared in this survey. The main objective of the survey [7] is to analyze facial emotion recognition under partial occlusion.

II. Masked faces

Due to the Covid -19, it is a must to wear a face mask in public to avoid spreading the virus among society. In facial emotion detection, lower face expressions and upper face expressions play an important role. According to [8], it is clear that the mouth area is the most important part of a human face to recognize basic facial expressions chiefly the happy emotion. Moreover, [9], states that the eye area is the most important part when identifying complex mental states. Especially the changes of the mouth according to the emotion help to identify the emotion efficiently. Covering the lower face with a face mask has become a huge challenge in emotion recognition by facial expressions. Some existing researches have been done to determine the impact of wearing face masks on emotion recognition by human experiments.

The bubble technique has been utilized in [8], to distinguish the best part of the face that can use for facial expression recognition. In this experiment, the pictures and videos of 10 Caucasian actors have been used to create the stimuli. Each picture and video presents one of eight emotions (anger, disgust, fear, happy, pain, sad, surprise, and neutral). 41 Caucasian people have participated to recognize emotions from pictures. The pictures consisted of above mentioned eight emotions and they were displayed for 500ms for emotion recognition. The video consisted of 15 frames starting from the neutral expression to the peak of one of eight facial expressions. The duration of the video is 500ms. 59 different people have participated to recognize emotions from videos. By using the bubble technique, it has helped to randomly sample the visual data contained in pictures and videos. After the stimuli were displayed for 500ms, The participants were instructed to press the corresponding key of the keyboard to those 8 emotions. The ROI analysis has been done to find the most important cue of a human face for facial emotion recognition from the experiment results. First, they have done ROI analysis and calculated the average z-score values for the mouth area and eye area and the ratio of those values has been calculated (mouth/eye). The ratio values for the classification of video frames and static images were 99.6% and 100% respectively. Therefore, the researchers have concluded that the mouth area is more informative for facial emotion recognition.

In [10], the researchers have identified the importance of improving existing facial emotion recognition technology on masked faces. The absence of the dataset of face images waring a face mask was the major challenge to improve the effectiveness of the existing facial emotion recognition methods. As a solution to this problem, they have introduced a method to automatically add face masks to existing facial expression recognition datasets. The publicly available datasets have selected and faces in the wild (LFW) dataset has labeled. next the KDDI dataset has constructed. Both datasets were annotated according to three types of facial expressions (positive, negative, and neutral) and obtained the LFW-FER dataset and KDDI-FER dataset. Then the M-LFW-FER and M-KDDI-FER datasets have obtained by putting masks with the proposed automatic warning face masks method. M-FER-T real-world masked dataset was constructed for model testing. VGG19 and MobileNet models were trained using LFW-FER, KDDI-FER, M-LFW-FER, and M-KDDI-FER datasets. Then the trained models were tested with the M-FER-T dataset. According to the test results, the emotion recognition accuracy on the M-FER-T dataset was insufficient. The limitation of this experiment is the insufficiency of training images. Recognition ability can be improved by increasing the size of the training dataset.

Another experiment has been done in [11], to recognize emotions from masked faces. In this research that tests the effect of masks and sunglasses, three experiments have been designed. familiar face matching, unfamiliar face matching, and emotion categorization. The researchers have used the NimStim face database for the emotion categorization experiment. That database contains images of 18 identities (9 males and 9 females) displaying angry, disgust, fear, happy, neutral, sad, and surprise emotions. The masks and sunglasses have been added to those images with emotions using photoshop. The participants for this experiment were selected from three groups. 100 participants from the unpracticed control group who may or may not have done face processing experiments earlier, 100 participants from the practiced control group who have participated in the previous face processing experiments, and 100 from the super-recognizers group who likely have been randomly invited to a comparable number of face processing projects. The masked face images with 6 emotions were presented one by one to the participants for 1000ms. The participants have chosen the emotion of the masked face images and the results have been recorded. According to the emotion recognition accuracies, the participants have recognized the neutral emotion better than other emotions. The neutral emotion recognition accuracies of three participant groups (unpracticed control, practiced control, and super-recognizers) were 82.33%, 85%, and 88.67% respectively. The least accuracy was recorded for the emotion disgust. The emotion recognition accuracies of the above three groups were 28.67%, 34%, and 36.33% respectively. The overall emotion recognition accuracies of three participant groups (unpracticed control, practiced control, and super-recognizers) were 68.89%, 72.46%, and 76.95% respectively. The results of this study are based on still images. Additional cues will be available when recognising emotions in live viewing environments.A deep neural network can be used for better accuracy on emotion recognition.

The research [12], has been done to determine the effects of face masks on emotion recognition accuracy and social judgments. For this experiment, they have selected 191 participants living in Germany. The composition of the participants is 52.9% female and 47.1% males. The FACES database which have portrait photographs of German people has been used. This experiment has conducted online with 38 trials. The participants have have randomly assigned into two group. Those two groups had two different conditions. In control condition, participants have shown original face images and in the masked condition, they have shown face images covered with face masks. In this experiment 2 practice trials have given to them and the results of the rest 36 trials have been recorded. First a portrait photograph of a face has shown to a participant for two seconds. Next the photograph has disappeared and 9 emotion options (Neutral, Proud, Fearful, Surprised, Amused, Sad, Happy, Angry, and Disgusted) were displayed for selection. The correct emotion choices has been coded as 1 while incorrect emotion choices has been recorded by 0. The overall emotion recognition accuracy for unmasked faces was 69.9% and the overall emotion recognition accuracy of masked faces was 48.9%.

Moreover, another research [13] has been conducted to examine the effect of face masks on facial emotion recognition and FER on masked and unmasked faces using people with autistic traits. The researchers have done three experiments in this research. In these three experiments, the participants were expected to complete three tasks, facial expression identification, selecting the level of confidence of their choice, and selecting the intensity of the emotion.

In experiment 1, 420 undergraduate students from the University of British Columbia have been selected as participants. 96 images of eight males and eight females expressing anger, disgust, fear, happiness, sadness, neutral were used for Facial expression recognition from the FACES database [14], and another set of images was created by covering the faces of original images with a facemask. Above mentioned three tasks were completed on masked and unmasked image datasets in 192 trials. The results of this experiment show that the ability of facial expression recognition is higher in unmasked faces than masked faces. Furthermore, the researchers were concluded that the expressions of masked male faces (M = 0.87) are more identifiable than masked female faces (M = 0.81). In terms of emotions, anger, disgust, fear, and naturally were the most recognizable emotions of unmasked male faces, and happiness and sadness were more identifiable in unmasked female faces. The participants were more confidant in recognizing anger and fear of male faces than female faces with masks. On the other hand, the confidence of recognizing the sadness of masked female faces than masked male faces.

Experiment 2 has done to test the results of experiment 1. The participants were changed in experiment 2. Participants were selected from a general population other than university students in experiment 2. The dataset used in experiment 2 is a subset of the dataset used in experiment 1. The number of trials was reduced to 144 in this experiment. 72 stimuli from 12 young faces (6 males and 6 females) with six emotions were used from Faces Database. Another set of 72 images was created by adding a face mask to the original image same as experiment 1. However, the results of experiment 2 were similar to the results of experiment 1. The expressions of masked male faces (M = 0.67) are more identifiable than masked female faces (M = 0.59).

The final experiment of this research has been done to examine the effect of the face mask on FER for high autistic trait people. The same stimuli used in experiment 1 have been used in this experiment. 142 participants were recruited from the human subject pool at the University of British Columbia. They have divided into two groups. All participants had to complete AQ-10 [15] (10 item autism spectrum Quotient). Based on their AQ-10 score, they have decided into two groups. 71 participants were selected to the high scores group which the score is 6 or higher and 71 for the low score group which the score is 5 or lower.

According to the experimental results, it can be seen that the facial emotion recognition accuracy on masked faces is low (M = 0.86) of the participants who endorsed more autistic traits (high AQ-10 scores) while the participants who endorsed more autistic traits (low AQ-10 scores) scored high facial emotion recognition accuracy (M = 0.88). Furthermore, it has been found that the participants who had high scores in AQ-10, were better on facial emotion recognition in male faces than female faces. The confidant of recognizing emotions fear, neutral, and sad has reduced from unmasked faces to masked faces in both high and low scores. However, both low score and high score participants have performed well on recognizing the happy emotion of unmasked faces. There was a significant drop in the confidence in recognizing emotions fear, neutral, and sad of masked faces in high scores than low scores. researchers have found no interaction between AQ-10 scores and masks in facial expression recognition.

The main objective of [16] was to investigate the effect of the face mask on facial emotion recognition, trust attribution, and re-identification of faces. The researchers have designed this experiment in two sections. Firstly, the emotion recognition and trust attribution tasks have been done from the face images with four emotional expressions Fear, Sadness, Happiness, and neutral. Secondly, the re-identification task has been done. The facial emotion recognition task has been done in three facial conditions, faces covering the entire mouth region with a standard face mask (SM), faces covering the mouth region with a transparent facemask which makes the mouth region visible (TM), and faces without masks (NM). 122 participants were used for this experiment. The stimuli were taken from the Karolinska Directed Emotional Faces (KDEF) database [17] for the emotion recognition task. 40 frontal face images were selected with four emotions happiness, fear, sad, and neutral. To find the impact of standard masks and transparent masks on emotion recognition, standard masks, and transparent masks were graphically added to unmasked face images . In this experiment, participants had to observe the randomized series of 48 face images which was clustered in 4 blocks of 3 mask conditions, NM, SM, and TM. all faces were expressing one of above mentioned 4 facial expressions. After the observation, they had to decide the emotions of that stimuli.

Participants have given 1 point for every correct prediction and 0 for the wrong prediction. Based on these test results, the researchers have done three analyses. The first analysis was done to find the effect of three mask conditions on emotion recognition. According to the analysis, the ability of emotion recognition was significantly reduced in SM conditions (Mdn = 0.81) than in other mask conditions TM (Mdn = 0.93) and NM (Mdn = 0.93). Secondly, the researchers have analyzed the effect of face masks on different emotions. based on analysis findings, it can be seen that there is a significant effect on happiness, fear, and sadness emotions for all mask conditions (SM, TM, and NM) while neutral emotion faces do not have an effect of masks. The final analysis of emotion recognition in this experiment was to find the most affected emotions by mask conditions. This analysis shows that there is no effect of NM on recognizing all emotions. In TM mask condition, the ability to recognize happiness is better than other emotions fear, sadness, and neutral. On the other hand, recognizing neutral is better than the other three emotions in SM mask condition. In this experiment, emotion recognition has done only for four emotions (Neutral, Happiness, Sadness, Fear).

The impact of wareing face masks on facial emotion readability was tested in another experiment in [18]. In this experiment, the researchers have used the FACES database [14] for selecting face stimuli. Frontal face images of 12 people (6 males and 6 females) who are in three different age groups were selected. To create the dataset for this experiment, 6 images were taken from each person that belongs to 6 emotional states (angry, disgusted, fearful, happy, neutral, and sad) and created a masked emotion face dataset by adding a face mask for those images by photoshop. The full dataset was contained 144 images (72 unmasked images and 72 masked images). A set of 41 people have participated in this experiment. SoSciSurvey online platform [19] was used to run this experiment. Each participant had to select the emotions of the face stimuli displayed one by one. Moreover, they had to select their confidence level of deciding the emotion from 1 (very unconfident) to 7 (very confident).

According to the experiment results, the accuracies of recognizing emotions, angry, disgusted, fearful, happy, neutral, and sad in unmasked face images were 83.7%, 93.9%, 92.5%, 98.8%, 93.1%, and 76.0%, respectively. Moreover, the accuracies for those emotions for masked face images were 69.5%, 43.7%, 93.5%, 74.2%, 93.1%, and 62.6% respectively. The accuracy of recognizing emotions in unmasked faces was greater than 83% for all emotions except sad. According to the confusion matrix, the emotion sad was confused with disgust in 20.3% of cases. In masked faces, emotion recognition accuracy was reduced for all emotions except fearful and neutral emotions. The emotion sad was confused with disgusted and neutral. Moreover, the emotion angry was confused with disgusted, neutral, and sad.

The main objective of the [20] research was to determine whether and to what extent the face mask affects emotion recognition among healthcare students. An experiment has been done using healthcare students of the University of Milano-Bicocca, Italy, to measure the ability of emotion recognition in masked and unmasked face images. Voluntary and anonymous participation was taken place with the approval of the ethical committee of the University of Milan-Bicocca. 1572 students have been invited to participate in this survey via an email with the link to the study survey. However, 208 students (115 medical and 93 nursing) have completed the survey before the deadline. In this research, Only four emotions fear, happiness, sadness, and anger were tested.

The researchers have used the Diagnostic Analysis of Nonverbal Accuracy 2-Adult Faces (DANVA2-AF) [21] in this experiment. It contains 24 images of adult faces for four emotions: fear, happiness, sadness, and anger, in three emotion intensities. A light blue surgical mask was digitally added to each image and created a modified version of DANVA2-AF. Participants were randomly assigned to masked or unmasked versions of DANVA2-AF, and they were asked to select the emotion of each image presented for 2 seconds. Higher errors were recorded in facial expression recognition for masked faces (9.6 ± 2.21) than unmasked faces (4.96 ± 2.78). The errors in predicting the masked images emotions happiness, sadness, and anger were higher than unmasked images for all emotion intensity stages. However, emotion fear has not shown a significant difference in recognizing emotions for masked and unmasked face images.

An online experiment has been done in the research [22], in order to determine the impact of wearing face masks on emotion recognition. 200 participants have attended this online experiment from all over the United Kindom. The composition of participants was 140 females and 60 males. Stimuli for this experiment were taken from the ADFES dataset [23]. This dataset consisted of 48 different videos for two genders, three actors, two mask conditions, two emotions (happy and sad), and two scenes (store background and park background). However, only 24 randomly selected videos were presented to each participant to avoid fatigue. Those videos were eddied, and a surgical face mask was added to each video using augmented virtual reality software. The goals of this study are to find whether face masks have an impact on emotion recognition, whether face masks reduce interpersonal closeness, and whether face masks reduce associative facial mimicry.

The complete task has taken about 15 minutes. Participants were asked to turn on their webcam during the experiment, and their facial activity was recorded with their permission while they saw 24 short videos. Then they had to rate the target emotion using the 7-point scale of emotions (happiness, sadness, fear, anger, disgust, and surprise) at the end of each video. The participants have asked to rate the emotion intensity as well. The accuracy of participants' prediction for masked and unmasked conditions was 99.7% and 99.8%, respectively, as only the emotions happy and sad shown in this experiment. In further analysis, the researchers have found that the target intensity of the expression is significantly lower in both happy and sad faces. According to these experimental results, they have concluded that wearing a mask reduces the emotion signals. The experiment was done by testing only two emotions, sad and happy.

Research [24] has been done to understand the impact of wearing face masks on identifying facial identity, emotion, age, and gender. Most of the experiments on facial recognition were conducted to obtain only the prediction accuracy. However, in this research, the researchers have designed the experiment to find the prediction accuracy and the recognition speed. They have designed four experiments to determine the impact of facemasks on identifying facial identity, emotion, age, and gender. 116 total participants were recruited from Ariel University, and they were divided into four groups for four different experiments. 30, 30, 24, and 32 participants have been allocated for these four experiments. The most crucial part of this experiment is the selection of stimuli. In most existing experiments, the researchers edited the emotion face images and added a mask. However, in this experiment, the researchers have photographed the front view of masked and unmasked faces of real people and created a real masked and unmasked face image database. The photographs were taken from 16 people of four groups, four young females, four young males, four old females, and four old males. The photographs of natural and sad emotions were taken in masked and unmasked conditions. These 64 images were converted into gray-scale images using photo editing software. Then, 32 images of the best 8 identities (4 male and 4 female) out of 16 identities were selected from those 64 images. Then the researchers have added inverted images of each image to the dataset.

In the emotion recognition experiment, the participants had to classify the images into two emotions, neutral and angry, and the experiment had six blocks. In block 1, there were upright face images with half of the masked faces and half of unmasked faces. Block 2, block 3, block 4, and block 5 consisted of upright face images of masked faces, upright images of unmasked faces, inverted images of masked faces, and inverted images of unmasked faces, respectively. The inverted face images with half of the masked faces and half of the unmasked faces were in block 6. in this research, the researchers have used two experimental designs, mixed design and blocked design. The mixed design was created using blocks 1 and 6, while blocks 2-5 were in the blocked design. These experiments were run in Macromedia Authorware software [25], and there were a total of 1920 trials in this experiment. According to the results of the mixed design, it can be seen that the masks reduced the emotion recognition accuracy. Moreover, wearing masks was not impacted emotion recognition accuracy in blocked design. However, the emotion recognition speed was reduced in both blocked and mixed designs.

All of the above researches have conducted human experiments to determine the impact of face masks on emotion recognition. However, an automated system to recognize emotions arousal and valence from masked faces using a deep neural network has been introduced in [26]. In this experiment, researchers have implemented a model of the FaceChannel [27] to recognize emotions from masked faces. The implemented FaceChannel has 10 convolutional layers, and each layer block has followed by 4 pooling layers. In the last convolutional layer, they have applied a shunting inhibitory layer [28] to boost the flexibility of emotion recognition, and the activation function was ReLU. Furthermore, they have implemented a fully connected ReLU-based hidden layer after the convolutional layers. The hidden layer is followed by two output layers that implement linear activation. In this experiment, AffectNet [29], and MaskedAffectNet datasets were used for model training. The MaskedAffectNet dataset was created by adding face masks artificially to the face images of the AffectNet dataset. Used parameter exploration was based on a Three-Parzen Exploration method [30] and selected one layer, 256 units per layer, 0.05 as the learning rate, and SGD as the optimizer for the fully connected layer.

Four experiments have been designed in this research. They have trained the FeceChannel on 1: AffectNet from scratch 2: MaskedAffectNet from scratch, 3: trained on AffectNet, then fine-tuned on MaskedAffectNet only the last convolutional layer 4: trained on AffectNet, then fine-tuned on MaskedAffectNet all layers. They have used concordance correlation coefficient (CCC) [31] as the primary metric to measure the performance. The overall results of these experiments show that the model performance was decreased when the model was trained with AffectNet and evaluated with MaskedAffectNet. However, when the model was trained with MaskedAffectNet and evaluated with AffectNet, the performance was not decreased considerably. The best model performance of recognizing emotions can be seen when the network is fully retrained with AffectNet, fine-tuned with MaskedAffectNet, and evaluated on MaskedAffectNet. The CCC values were 0.53 and 0.45 for arousal and valence, respectively.

III. Occluded faces

Face occlusion can happen in various ways other than face masks. For instance, objects, shadows, hair, different face poses, Etc. There can be an impact in face occlusion on facial expression recognition. Some existing research can be found to investigate the emotion recognition ability when face occlusion occurs. The progressive transfer learning approach [32] was used to recognize the facial emotions of the face images with pose variations and face occlusions in [33]. In this research, they have used four different existing face image datasets. Two were used for model training, while the other two were used for model testing. JAFFE [34] and CK+ [35] [36] databases have used for model training while VT-KFER [37] and, W300 [38] databases were used for model testing. The JAFFE [34] database contains 213 posed images with six basic and neutral expressions of frontal faces in greyscale. Total number of 4001 greyscale and RGB frontal images of 6 basic expressions were in CK+ [35] [36] database. In VT-KFE [37] there are 11619 total frontal and non-frontal (left and right) face images of 6 basic expressions of 3 different intensities. All images of this database are in RGB format. A total of 240 images of 6 basic expressions were selected among 600 images (300 indoor and 300 outdoor) from W300 [38] database. This database contains images in various illumination conditions, poses, occlusion, and face sizes.

In this experiment, the researchers have applied the progressive transfer learning approach [32] in three stages. First, they have transferred the knowledge from AlexNet [39] to a new Concolusional neural network. Then the new network was fine-tuned on JAFFE [34] database. The knowledge gained from AlexNet [39]and JAFFE [34] was transferred to a new network, and that network was trained on CK+ [35] [36] database. Furthermore, the resultant network was fine-tuned and tested on VT-KFE [37] and W300 [38] databases. This experiment's test results were compared with the two existing experimental results. The proposed Multi-stage Progressive Transfer Learning has performed better than existing Transfer learning [34] and Progressive Transfer Learning [32] approaches. They have tested the final model, and other existing models on frontal face images, non-frontal images, frontal and non-frontal faces of VT-KFE [37] database, and 300W [38] database. In each testing, the proposed method, Multi-stage Progressive Transfer Learning, obtained a higher emotion recognition accuracy than the other two existing methods. Three face image sets of VT-KFER, Only Frontal, Only Non-frontal and both frontal and non-frontal have achieved 84.2%, 62.3% and 77.9% respectively. Faces-in-the-wild images in 300W dataset has achieved 61.7% emotion recognition accuracy.

When recognizing facial expressions from occluded faces, it is crucial to identify occluded regions and focus on the unblocked most informative regions. In [40], Patch-Gated Convolution Neural Network (PG-CNN) has been introduced to do this task. PG-CNN can automatically identify the occlusions and non-occluded face regions to extract the features of the expressions. The researchers have designed this experiment to evaluate PG-CNN and compare this method with other state-of-the-art FER methods. In this experiment, both in-the- wild datasets (RAF-DB [41] and AffectNet [29]) and in-the- lab datasets(CK+ [35] [36], MMI [42], and Oulu-CASIA [43]) were used for the evaluation. The items that have a high chance to be occlusion objects, such as beer, bread, wall, hand, hair, hat, book, cabinet, computer, cup, etc., have been collected from RAF-DB [41].

The implementation was done by adopting VGG-16 [44] as the base for PG-CNN [45]. The proposed method, PG-CNN, was compared with DLP-CNN [46]. The experiment reported the overall recognition accuracy of seven emotions (six basic emotions + neutral) for occluded and non-occluded images. The ratio of all datasets and their modification with synthesized facial occlusions was 1:1. All these networks were trained with six different occlusion types, non-occlusion, three occlusion object sizes (8*8, 16*16,24*24), eyes occluded, and mouth occluded. All experiment results showed that the proposed Patch-Gated CNN had performed better than other tested state-of-the-art methods. The best accuracy was achieved when train the model on AffectNet dataset [29]. The test accuracies for CK+ occluded, MMI and Oulu-CASIA datasets were 86.27%, 63.94% and 54.18% respectively.

Moreover, identifying the blocked areas of the face and paying attention to un-blocked regions have been done by Convolution Neural Network with attention mechanism (ACNN) in [47]. This research has proposed two versions of ACNN Patch Based ACNN (pACNN) and Global-Local Based ACNN (gACNN). The pACNN focuses on local discriminative and representative patches, while the gACNN focuses more on essential information on faces. In this experiment, RAF-DB [41] ,AffectNet [29], SFEW [48], CK+ [35] [36], MMI [42], and Oulu-CASIA [43] databases were used. The researchers have collected manually synthesized occluded images from RAF-DB [41] dataset, same as [40]. Furthermore, they collected images, annotated real facial occlusions to those face images, and created a dataset FED-RO [47]. This database consisted of 400 face images of six basic and neutral expressions. ACNNs were implemented using Caffe deep learning framework [45] and adopted VGG-16 [44] as the backbone network. This research reports all datasets' facial emotion recognition performances and the overall accuracy of seven emotion categories on both occluded and non-occluded face images. Then the proposed methods were compared with two other state-of-art methods, WLS-RF [49], and RGBT [50].Proposed methods, pACNN, and gACNN, were given higher emotion recognition accuracy in all occlusion types, non-occlusion, three occlusion object sizes (8*8, 16*16,24*24), eyes occluded, and mouth occluded. The best accuracy was achieved when train the model on AffectNet dataset [29]. The test accuracies for CK+ occluded, MMI, Oulu-CASIA and SFEW datasets were 88.17%, 65.48%, 55.42% and 51.72% respectively.

Furthermore, it has been used a CNN based on the attentional convolutional network in [51] to focus on important parts of the face on emotion recognition. In the proposed model of this research, it has added an attention mechanism through a spatial transformer network. It helps the framework to focus on special regions of the faces. The researchers have trained the proposed model on four existing facial expression databases and recorded the facial emotion recognition accuracies. Then the facial emotion recognition accuracy of each dataset on the proposed model was compared with the results of other existing studies. The datasets used in this experiment are FER2013 [52], CK+ [35] [36], JAFFE [34] and FERG [53]. According to the experiment results, the proposed model has performed well for three databases than other existing methods. For databases, JAFFE [34], CK+ [35] [36] and FERG [53], the proposed model has achieved the highest facial expression recognition accuracy which was 92.8%, 98% and 99.3% respectively. However for FER2013 [52] database it has achieved 70.02% accuracy which was less than the accuracy of an existing method, 75%.

An Occlusion-Adaptive Deep Network (OADN) was proposed in [54] to recognize facial expressions of occluded faces in-the-wild. This proposed network consisted of two branches, the landmark-guided attention branch and the facial region branch, to maximize facial recognition ability. The OADN was implemented by removing the average pooling layer and the fully connected layer of ResNet50 [55]. The landmark guided attention branch was used to extract information from non-occluded regions, while the facial region branch was used to learn region-based classifiers. The proposed network was validated with two large in-the-wild facial expression datasets, RAF-DB [41], and AffectNet [29]. Furthermore, the network was validated on three recent real-world occluded facial expression datasets, FED-RO [47], Occlusion-AffectNet [56], and Occlusion- FERPlus [56]. Two different test sets from RAF-DB [41], and AffectNet [29] datasets were used for the model testing. Then test results of the proposed method were compared with existing methods. The proposed model has achieved 87.16% and 61.89% test accuracies on RAF-DB [41] and AffectNet [29] test sets respectively. The test accuracy of FED-RO [47] dataset was 68.11%. The proposed network (OADN) has outperformed existing methods.

Recognizing facial expressions from the faces which are occluded in the upper area has been done in the following experiments. The experiment [57] proposed a CNN-based approach to recognizing the facial expressions of face images waring VR headsets. This approach focuses on paying more attention to the lower part of the face and recognizing emotion accurately.

The researchers have focused on recognizing facial expressions from faces while wearing a virtual reality headset. When a person is wearing a virtual reality headset, the upper section of the face, including the eyes, forehead, and eyebrows, is occluded. To overcome the problem of the inexistence of face images occluded with virtual reality headsets, the researchers have applied a VR patch to all train images by masking the upper region of the face. The experiment [57] proposed a CNN-based approach to recognizing the facial expressions of face images waring VR headsets. This approach focuses on paying more attention to the lower part of the face and recognizing emotion accurately. This experiment has designed in two stages. The pre-trained models VGG-face [58] and [59] have used in this research. In the first stage of this experiment, these two models were fine-tuned on non-occluded full faces. Then the models were fine-tuned on the faces in which the upper face is occluded as the second stage of this experiment. FER+ [60] and AffectNet [29] databases were used to train the models. The experiment results of various trining and testing methods were recorderd and compared with other exiting methods. The experiment was designed in three parts, training and testing on full faces, training on full faces and testing on lower-half faces, and training and testing on lower-half faces. According to the results of these experiments, training on Affectnet [29] database showed lower accuracy than training on FER+ [60] database. Training on full faces and testing on lower-half faces showed lowest emotion recognition accuracies compaired to other two testing methods. Training and testing on lower-half faces recorded emotion recognition accuracies, 49.23%, 82.28% on AffectNet [29] and FER+ [60] respectively.

Furthermore, in [61], researchers have used a geometric model for a realistic occlusion when masking the train images. The datasets used in this research was FER+ [60], RAF-DB [41] and AffectNet [29]. The transfer learning method was used in this experiment and fine-tuned the pre-trained models VGG-face [58] and ResNet50 [55] on images occluded by applying a VR patch. These two pre-trained models were fine-tuned on images of databases FER+ [60], RAF-DB [41] and AffectNet [29] with VR patch respectively. Then the test results were compared with the results of [57]. The results of training the model from scratch were also recorded for comparison. The highest emotion recognition accuracy of this experiment achieved by the VGG-face [58] model pre-trained on VGG-face database and fine-tuned on FER+ [60] database. the accuracy was 79.98%. However, this accuracy is less than the highest accuracy of VGG-face [58] in [57] which was 82.28%.

IV. Group and Multiple faces

Emotion recognition of the group and multiple faces is another most important area of facial emotion recognition. When multiple faces have appeared in a single frame or an image of a group of people, emotion recognition has to be done on all faces. The research [62] has been done to recognize the emotions of the face images taken from a movie. These images contains single faces and multiple faces of people in a one frame. The main objective of this research is to generate an emotional summary of the faces of characters in a movie. The experiment has been designed into four main segments. First, the researchers have proposed an entropy-based shots segmentation technique to create the frames of scenes and categorize those shots into informative and noninformative. Then only the informative frames were selected for the experiment. The second stage is the Saliency Extraction and Face Detection. This method was used to discard the noninformative shots and to detect the faces from selected frames. The facial expression recognition using transfer learning was the third step of this study. In this stage, the researchers have used the pre-trained model, ResNet50 [55] for the implementation. Two sets of experiments were performed for facial exoression recognition stage. The model trained on KDEF dataset [17] evaluated and compared with other models as the first experiment. The second experiment is a subjective evaluation of five Hollywood movies of different genres.

The VGG-face [58] was used to train the model from scratch, and the KDEF [17] datasets was used for transfer learning. The test results shows that the model ResNet50 [55] (128x128) has achieved the highest accuracy. It was 93.65%. According to the results of movie summarization, it can be seen that the proposed method had better performance with compared to other movie summarization techniques.

Moreover, in [63], deep learning techniques have been used to recognize human sentiment and activity in disaster situations using images on social media. The main objective of this research is to recognize human emotions and activities of multi persons in an occluded environment. The researchers have conducted a crowd survey to annotate a disaster-related dataset with human sentiments and activity responses related to the sentiment. That dataset was created with images of 1200 participants, and the total number of images was 3995. All of these images were collected from social networks, Twitter, Google Images, Flicker, etc. The disasters, floods, earthquakes, tornadoes, tsunamis, etc. have been used as the search tags when the images are collected. The total images were annotated by presenting those images to participants of this annotating phase. During this analysis, 10,000 different answers from 2300 different participants were collected. 2300 participants were selected from multiple ages, gender groups, and 25 different countries. The annotating study was done by providing a questionnaire to participants.

Among other experiments in this research, the emotion recognition task was done using the transfer learning technique. Dense Net [64], AlexNet [39], Inception Net [65], VGGNet [66], and ResNet [67] pre-trained models used and fine-tuned on annotated large dataset. The dataset was divided into 70%, 10%, 20% train, validation, and test dataset ratio respectively. The facial expression recognition has done in three emotion classes, negative, neutral, and positive. According to the test accuracies of these pre-trained models, VGGNet. The accuracy was 92.88%.

V. Future direction

Multimodal approach to improve the detection (face + voice + text)

References

[1] A. Mehrabian, "Nonverbal communication," 2017.

[2] N. Liu, H. Liu, and H. Liu, "Mental health diagnosis of college students based on facial recognition and neural network," *Journal of Intelligent Fuzzy Systems*, vol. 40, no. 4, pp. 7061–7072, 2021.

[3] N. Farhoumandi, S. Mollaey, S. Heysieattalab, M. Zarean, and R. Eyvazpour, "Facial emotion recognition predicts alexithymia using machine learning," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021.

[4] D. Yang, A. Alsadoon, P. Prasad, A. Singh, and A. Elchouemi, "An emotion recognition model based on facial recognition in virtual learning environment," *Procedia Computer Science*, vol. 125, pp. 2–10, 2018.

[5] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, p. 1189, 2019.

[6] M. Murtaza, M. Sharif, M. Raza, and J. Shah, "Analysis of face recognition under varying facial expression: A survey," *International Arab Journal of Information Technology*, vol. 10, pp. 378–388, 09 2013.

[7] L. Zhang, B. Verma, D. Tjondronegoro, and V. Chandran, "Facial expression analysis under partial occlusion," *ACM Computing Surveys*, vol. 51, no. 2, pp. 1–49, 2019.

[8] C. Blais, C. Roy, D. Fiset, M. Arguin, and F. Gosselin, "The eyes are not the window to basic emotions," *Neuropsychologia*, vol. 50, no. 12, pp. 2830–2838, 2012.

[9] S. Baron-Cohen, S. Wheelwright, Jolliffe, and Therese, "Is there a "language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome," *Visual Cognition*, vol. 4, no. 3, pp. 311–331, 1997.

[10] B. Yang, J. Wu, and G. Hattori, "Facial expression recognition with the advent of face masks," *19th International Conference on Mobile and Ubiquitous Multimedia*, 2020.

[11] E. Noyes, J. P. Davis, N. Petrov, K. L. H. Gray, and K. L. Ritchie, "The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers," *Royal Society Open Science*, vol. 8, no. 3, 2021.

[12] F. Grundmann, K. Epstude, and S. Scheibe, "Face masks reduce emotion-recognition accuracy and perceived closeness," *PLOS ONE*, vol. 16, no. 4, p. e0249792, 2021.

[13] F. Pazhoohi, L. Forby, and A. Kingstone, "Facial masks affect emotion recognition in the general population and individuals with autistic traits," *PLOS ONE*, vol. 16, no. 9, p. e0257740, 2021.

[14] N. C. Ebner, M. Riediger, and U. Lindenberger, "Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior Research Methods*, vol. 42, no. 1, pp. 351–362, 2010.

[15] C. Allison, B. Auyeung, and S. Baron-Cohen, "Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls," 2021.

[16] M. Marini, A. Ansani, F. Paglieri, F. Caruana, and M. Viola, "The impact of facemasks on emotion recognition, trust attribution and re-identification," *Scientific Reports*, vol. 11, no. 1, 2021.

[17] D. Lundqvist, A. Flykt, and A. Öhman, "Karolinska directed emotional faces," *PsycTESTS Dataset*, 1998.

[18] C.-C. Carbon, "Wearing face masks strongly confuses counterparts in reading emotions," *Frontiers in Psychology*, vol. 11, 2020.

[19] 2022. [Online]. Available: https://www.soscisurvey.de/

[20] M. Bani, S. Russo, S. Ardenghi, G. Rampoldi, V. Wickline, S. Nowicki, and M. G. Strepparava, "Behind the mask: Emotion recognition in healthcare students," *Medical Science Educator*, vol. 31, no. 4, pp. 1273–1277, 2021.

[21] S. Nowicki and J. Carton, "The measurement of emotional intensity from facial expressions," *The Journal of Social Psychology*, vol. 133, no. 5, pp. 749–750, 1993.

[22] T. Kastendieck, S. Zillmer, and U. Hess, "(un)mask yourself! effects of face masks on facial mimicry and emotion perception during the covid-19 pandemic," *Cognition and Emotion*, vol. 36, no. 1, pp. 59–69, 2021.

[23] J. van der Schalk, S. T. Hawk, A. H. Fischer, and B. Doosje, "Moving faces, looking places: Validation of the amsterdam dynamic facial expression set (adfes)." *Emotion*, vol. 11, no. 4, pp. 907–920, 2011.

[24] D. Fitousi, N. Rotschild, C. Pnini, and O. Azizi, "Understanding the impact of face masks on the processing of facial identity, emotion, age, and gender," *Frontiers in Psychology*, vol. 12, 2021.

[25] H. Marler and A. Ditton, ""i'm smiling back at you": Exploring the impact of mask wearing on communication in healthcare," *International Journal of Language Communication Disorders*, vol. 56, no. 1, pp. 205–214, 2020.

[26] P. Barros and A. Sciutti, "I only have eyes for you: The impact of masks on convolutional-based facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021, pp. 1226–1231.

[27] P. Barros, N. Churamani, and A. Sciutti, "The facechannel: A fast and furious deep neural network for facial expression recognition," *SN Computer Science*, vol. 1, no. 6, 2020.

[28] Y. Frégnac, C. Monier, F. Chavane, P. Baudot, and L. Graham, "Shunting inhibition, a silent step in visual cortical computation," *Journal of Physiology-Paris*, vol. 97, no. 4-6, pp. 441–451, 2003.

[29] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.

[30] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011.

[31] L. I.-K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, no. 1, p. 255, 1989.

[32] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015.

[33] S. F. Aly and A. L. Abbott, "Facial emotion recognition with varying poses and/or partial occlusion using multi-stage progressive transfer learning," in *Scandinavian Conference on Image Analysis*. Springer, 2019, pp. 101–112.

[34] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*.

[35] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010.

[36] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*.

[37] S. Aly, A. Trubanova, L. Abbott, S. White, and A. Youssef, "Vt-kfer: A kinect-based rgbd+time dataset for spontaneous and non-spontaneous facial expression recognition," *2015 International Conference on Biometrics (ICB)*, 2015.

[38] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: database and results," *Image and Vision Computing*, vol. 47, pp. 3–18, 2016.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[40] Y. Li, J. Zeng, S. Shan, and X. Chen, "Patch-gated cnn for occlusion-aware facial expression recognition," *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018.

[41] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[42] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," *2005 IEEE International Conference on Multimedia and Expo*.

[43] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, 2011.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

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

[46] L. Zhao, X. Li, Y. Zhuang, and J. Wang, "Deeply-learned part-aligned representations for person re-identification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[47] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using cnn with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2019.

[48] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011.

[49] A. Dapogny, K. Bailly, and S. Dubuisson, "Confidence-weighted local expression predictions for occlusion handling in expression recognition and action unit detection," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 255–271, 2017.

[50] L. Zhang, D. Tjondronegoro, and V. Chandran, "Random gabor based templates for facial expression recognition in images with facial occlusion," *Neurocomputing*, vol. 145, pp. 451–464, 2014.

[51] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, 2021.

[52] I. J. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. e. a. Lee, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015.

[53] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," *Computer Vision – ACCV 2016*, pp. 136–153, 2017.

[54] H. Ding, P. Zhou, and R. Chellappa, "Occlusion-adaptive deep network for robust facial expression recognition," *2020 IEEE International Joint Conference on Biometrics (IJCB)*, 2020.

[55] B. Bozorgtabar, M. S. Rad, H. Kemal Ekenel, and J.-P. Thiran, "Using photorealistic face synthesis and domain adaptation to improve facial expression analysis," *2019 14th IEEE International Conference on Automatic Face amp; Gesture Recognition (FG 2019)*, 2019.

[56] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4057–4069, 2020.

[57] M.-I. Georgescu and R. T. Ionescu, "Recognizing facial expressions of occluded faces using convolutional neural networks," *Communications in Computer and Information Science*, pp. 645–653, 2019.

[58] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," *Procedings of the British Machine Vision Conference 2015*, 2015.

[59] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014. [Online]. Available: https://arxiv.org/abs/1405.3531

[60] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016.

[61] B. Houshmand and N. Mefraz Khan, "Facial expression recognition under partial occlusion from virtual reality headsets based on transfer learning," *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, 2020.

[62] I. Ul Haq, A. Ullah, K. Muhammad, M. Y. Lee, and S. W. Baik, "Personalized movie summarization using deep cnn-assisted facial expression recognition," *Complexity*, vol. 2019, pp. 1–10, 2019.

[63] A. M. Sadiq, H. Ahn, and Y. B. Choi, "Human sentiment and activity recognition in disaster situations using social media images based on deep learning," *Sensors*, vol. 20, no. 24, p. 7115, 2020.

[64] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," 04 2014.

[65] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[67] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.