Compound Facial Expression Recognition Using Learning Machine Classifier*

*

1st Abanoub Wagdy Shafik
FGSSR (of Aff.)
CAIRO UNIVERSITY (of Aff.)
Cairo, Egypt
Student Id : 202100596

2th Ammar Mohamed
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Abstract Data mining and machine learning are nowadays among the most known topics in research and they are used to analyze several probabilities of the characterization of databases. we are going to classify different expressions of faces and implement them in real time on the streaming data, in this process of building an end to end pipeline, we are going to dealing with image datasets, performing data processing and augmentation as and when required, creating and training a convolutional neural network. facial expression recognition is a technology which uses biometric markers to detect emotions in human faces. More precisely, this technology is a sentiment analysis tool and is able to automatically detect six basic universal expressions happiness, sadness, anger, surprise, fear and disgust. Understanding facial expressions is very important because facial expressions can be nonverbal communication cues that play an important role in interpersonal relations.These cues complement speech by helping the listener to interpret the intended meaning of the spoken. This paper provides a brief review of researches in the field of deep learning and an experiment on the identification/detection of compound emotion categories, i.e. combination of two basic emotions, for example: Happily Surprised, Sadly Surprised, etc**

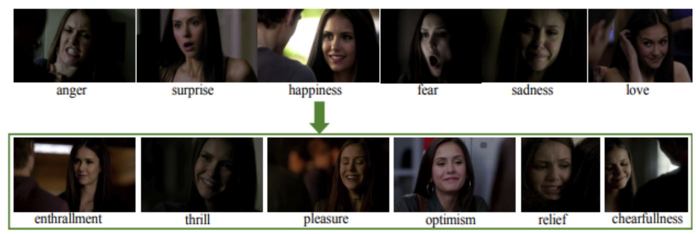
***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

FACIAL expression plays a vital role in revealing a person's internal thoughts and feelings. Over the past few decades, facial expression recognition (FER) has been a hot issue in the field of computer vision and human-computer interaction. In general, FER methods follow two lines of research: categorical methods and continuous methods (also called dimensional methods) [39][37], which map the given resources (images or videos) to discrete classes and describe affective state of resources in dimensional space formulated as valence, arousal and power, respectively. Though continuous methods may represent a wider range of facial expressions, they are less competitive than categorical methods due to the LaTeX.

(SVM) [22], K Nearest Neighbbor (KNN) [23] have been applied successfully to classify facial expression. These methods require a tedious manual tuning of parameters to hand-design an effective feature. As an alternative, recently, research

Identify applicable funding agency here. If none, delete this.



studies show that features can be learned automatically using deep learning techniques [24] [25]. Recent works [15] [16] [17] show also that the traditional approaches do not have the capability to, first, detect faces from images captured in a spontaneous uncontrolled manner, then, extract features and finally recognize facial expression. However, the deep learning techniques, which are based on the neural network, is able to extract undefined features from the training database, then the network can extract features that generalize well to images captured under different acquisition conditions on which the network was not trained. To achieve high performance, deep learning methods require huge amounts of training data. Most recently, different large databases [18] [24] [26] are available which allow to train neural network in different fields like: multi-view face detection [27] [28] and FER [15] [17]. In this paper, we are interested in recognizing compound facial expressions using CFEE database. In order to overcome the problem of hand-designing an effective feature, we present and evaluate the Highway Convolutional Neural Network architecture which, to our best knowledge, is used for the first time in FER field. The rest of this paper is structured as follows. Section 2 reviews related works about compound/basic facial expressions using traditional and deep learning techniques. Section 3 presents our proposed methodology using CNNs architecture. Section 4 describes the experimental method and results on the CFEE database. Finally, we conclude the work in Section 5

II. Related Work

Automatic facial recognition system is typically composed of three main modules. The first module is to detect and register facial region in the images. The second module consists of extracting and representing facial changes caused by facial expressions. The final module takes as input the previously extracted feature vectors for performing the classification task using machine learning techniques. There is two different approaches relied on handcrafted features: methods based on appearance features and methods based on geometric features. Appearance features describe the texture of the skin such as wrinkles and furrows. The most commonly used approaches for facial expression analysis include Gabor Wavelet [29], Haar-like features [30], LBP [13] and its variants [31]. Shan et al. [32] applied LBP to represent facial expressions. They showed that LBP features are more discriminating and effective than those of Gabor. Local Ternary Pattern (LTP) [33] with an additional level of discrimination and ternary codes were introduced to address the limitations of LBP and to combat nonuniform noise. Geometric features describe the shape of the facial components (i.e. the mouth, eyes, eyebrows, and nose) and their location (i.e. the corners of the eyes, the corners of the mouth, etc.). Approaches using geometric features rely mainly on the landmarks position such as visual information or the geometric displacement of the landmarks, or shape parameterization of the facial component [34] [35]. The aforementioned traditional approaches have shown their effectiveness on different facial expression databases such as CK+ [36], Oulu-CASIA [37] and FEED [38], where the images representing the six basic emotions (i.e. happiness, surprise, fear, anger, disgust, sadness) captured in the "lab-controlled" manner. However, they are not most effective when it comes to recognize facial expression in the wild where faces often express compound emotions or when images are captured in uncontrolled manner (poor illumination, low resolution, etc.). RAF-DB [18] and SFEW [39] are the most used databases representing the real-world facial expressions. To overcome these difficult situations, recently, several studies have used the deep learning techniques and have shown that these techniques are more robust than traditional ones [18] [25]. Among the most commonly used architectures, Convolutional Neural Network (CNN) is a deep neural network architecture which attracts researchers for studying vision and deep learning [18] [25] [40]. Li and Deng [25] propose a Deep Bi-Manifold CNN (DBMCNN), to learn the discriminative feature for multi-label expressions by jointly preserving the local affinity of deep features and the manifold structures of emotion labels. Li et al [18] propose a Deep Locality-Preserving CNN (DLP-CNN) to address the ambiguity and multi-modality of real-world facial expressions by adding a new supervised layer on the basic architecture of Deep Convolutional Neural Network (DCNN). They show that the proposed DLP-CNN outperforms the handcrafted features and deep learning based methods for the expression recognition in the wild. The authors in [41] proposed approach called Boosted Deep Belief Network (BDBN). Their approach performs the three stages of learning (feature learning, feature selection and classifier construction) iteratively in a unified system. The approach is based only on the six basic emotions using "lab-controlled" databases CK+ and JAFFE [42]. A new loss function called cluster loss is proposed in [40] to perform deep features compact. This approach is based on deep CNN model and achieve significant improvement when combining data augmentation and cluster loss using realworld six basic emotions from RAF-DB. In [43], Mollahosseini et al. propose a deep neural network architecture which consists of two convolutional layers each followed by max pooling and then four Inception layers. Their study addresses the FER problem across real-world and lab-controlled databases. In [6], Egede et al. encode shape and appearance information in both types of handcrafted and deep learning feature extraction. For handcrafted features, HOG and a number of metrics were extracted from 49 facial dots to represent, respectively, appearance characteristics and geometric characteristics. Then, based on CNN, features are learned from a combination of pixels from the original image (appearance) and binary masks (face shape). Yong et al. [44] propose a Patch-Gated Convolution Neutral Network (PG-CNN) which detects automatically the occluded region of the face and concentrates on the most discriminative un-occluded regions. PGCNN devises an intermediate feature map into several patches according to the positions of related facial landmarks to define the possible facial region of interest. This approach is performed on two different in-the-wild databases RAF-DB and AffectNet [45]. Compound Facial Expression Recognition Based on Highway CNN

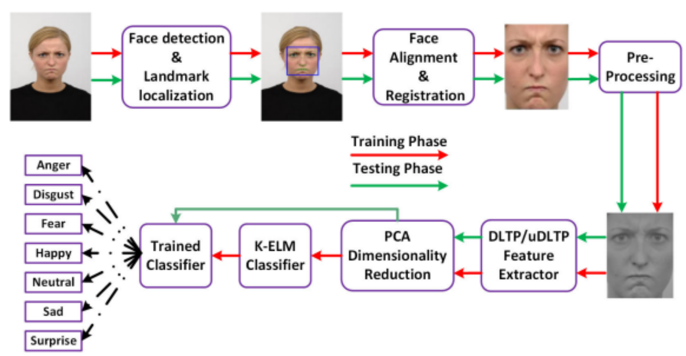


Fig. 1. Algorithmic pipeline of the proposed facial expression recognition system"

using the SVNN classifier into one of the seven facial expressions. Several techniques developed for the classification of facial expressions have also employed feature selection techniques. By reducing the dimensions of the features, these techniques achieve the fast classification of expressions besides improving their classification accuracy. Ghosh et al. [16] proposed a new feature selection (FS) algorithm based on Late Hill Climbing and Memetic Algorithm (MA) (LHCMA). The LHCMA FS algorithm achieved superior performance than the popular FS algorithms when tested with several facial descriptors, namely the LBP, Histogram of Oriented Gradients (HOG), etc. In other related work, Saha et al. [40] introduced the supervised filter harmony search algorithm (SFHSA) for FS in the FER task. The SFHSA algorithm use cosine similarity to remove similar features from feature vectors and minimal-redundancy maximal-relevance (mRMR) to determine the feasibility of the optimal feature subsets using Pearson's correlation coefficient (PCC). Their designed SFHSA algorithm, when tested with five state-of-the-art feature descriptors using the RaF and JAFFE datasets, achieved a notable improvement in the recognition accuracy. The FER technique introduced by Shanthi and Nickolas [41] has fused facial features extracted using the LBP and Local Neighborhood Encoded Pattern (LNEP). The chi-square statistical analysis is used to select the most relevant features from the original high-dimensional feature vectors and is classified using the SVM classifier. Siddiqi et al. [42] introduced a system for FER that uses the wavelet transform for feature extraction, a new robust step-wise linear discriminant analysis (SWLDA) feature selection algorithm, and a hidden Markov model (HMM) classifier. Given the facial images, their designed FER system first detects faces using a novel unsupervised technique based on the active contour (AC) model. The FER pipeline proposed by Kumar and Rajagopal [43] has used normalized minimal feature vectors and semi-supervised Twin Support Vector Machine (TWSVM) learning. Li and Wen [44], proposed a sample awareness-based personalized (SAP) FER method that uses the Bayesian learning method to select the optimal classifier from the global perspective and then used the selected classifier to identify the emotional class of each test sample. The authors in [45] proposed a novel sparse modified Marginal Fisher analysis (SMMFA) for the FER task. SMMFA efficiently reduces the dimension of the facial features and thus helps in extracting discriminant features for FER. In another work, Li et al. [46] proposed a novel FER scheme that uses a dynamic ensemble pruning method called graph-based dynamic ensemble pruning (GDEP) for the recognition of facial expression in static facial images. Since their inception, the deep learning techniques have proved their efficacy in solving several computer vision problems like image classification, object detection, speech recognition, etc. Therefore, in the last few years, many works have been proposed for FER in static images using deep learning [20], [47]–[56]. The deep learning algorithms are datadependent algorithms, whose performance linearly increases with the amount of the dataset. Therefore, on small-scale FER datasets like the CK+, RaF, JAFFE, KDEF, etc., the traditional machine-learning-based FER methods outperform the deep-learning-based FER methods. On the contrary, on largescale FER datasets, like RAF-DB [57], FER2013 [58], and AffectNet [59], the deep learning CNN models have outperformed the traditional machine-learning-based FER methods [60].

III. Methodology

Past analysis on facial emotion expression recognition chiefly targeted on six fundamental categories: happy, surprised, fearful, sad, angry, and disgusted. In any case, there are numerous advanced and more expounded facial expressions humans do, built from the mix of distinctive fundamental one, that begun to attract a lot of consideration from the past few years within the computer vision and machine learning communities, i.e., the supposed compound emotions. In our experimentation, we use a database that contains facial expression images of compound emotions.

A. Compound Facial Expressions of Emotion

3.1 Compound Facial Expressions of Emotion Compound Facial Expressions of Emotion (CFEE) [4] is a database of facial expression images, standard and non-commercial. 230 subjects (130 women, mean age 23 years) were enrolled from the university region and gotten a little budgetary remunerate for their cooperation. Most ethnicities and races have been included, Caucasian, Asian, African-American and Hispanic individuals are represented in the database. Facial occlusions were minimized without glasses or hair. Individuals who required remedial lenses were in contact. Men were inquired to shave their faces as cleanly as conceivable. They were also welcomed to discover their brow to completely appear their eyebrows. The CFEE database has a total number of 5,060 images; 1610 of fundamental emotions and 3450 of compound ones expressed by 230 subjects, with each subject have frontal direction. Hence, the categories of emotions depicted in this database can be classified into two bunches. The primary gather contains six basic emotions; "happiness", "sadness", "fear", "anger", "surprise", and "disgust" (see Fig. 2 (1-6)). The second bunch relates to compound emotions. Here, compound means that the emotion category is built as a combination of two basic emotion categories. Fig.2 (7-18) shows the 12 compound emotions most commonly communicated by humans

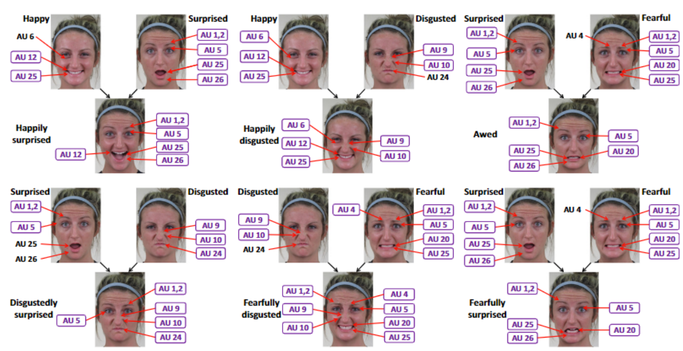


Fig. 2. Shown here are the AUs of six compound facial expressions of emotion. The AUs of the basic emotions are combined as shown to produce thecompound category. The AUs of the basic expressions kept to produce the compound emotion are marked with a bounding box. These relationships definethe subordinate classes of each category and their interrelatedness. In turn, these results define possible confusion of the compound emotion categories bytheir subordinates and vice versa.

B. Data Preprocessing

face detection is the key issue in FER. Excessive background data that is uncorrelated to expression acknowledgment exists in a facial image, indeed when the image chosen from a benchmarking facial expression dataset. Hence, a robust FER system depends on the precision of the outcomes of face detection. OpenCV Library already contains numerous pre-trained classifiers for the face, eyes, smile, etc. Within the considered study, we utilize Haar cascade detection of frontal faces in OpenCV Library to identify the faces of CFEE database, then we resize the detected faces to 454x454 pixels and finally, we convert the color images into grayscale images

C. EVALUATION PROCEDURES

For performance validation, the proposed FER pipeline has used two evaluation procedures viz the cross-validation [29], and cross-dataset [11]. Moreover, for a fair comparison with the existing works, the performance has been evaluated using four metrics (recognition accuracy, precision, recall, and F1-score). The following section provides details of different evaluation procedures and metrics.

- CROSS-VALIDATION

The pattern recognition tasks usually use K-fold crossvalidation (CV) to measure the performance of a classifier in two scenarios: (a) available data is not sufficient, and (b) distribution of the dataset into training and test set is not known. In these scenarios, K-fold CV is performed by randomly dividing the data roughly into K equal parts. For each fold, a classifier is trained on the (K-1) data parts and tested on the remaining. Afterward, the test accuracy obtained on each fold of the 10-fold CV is summed up and divided by K to get the average accuracy. Since the dataset is divided randomly in the 10-fold CV, its multiple runs each time give different average accuracy. Therefore, as suggested by Holder and Tapamo [29], the proposed FER testing protocol has utilized ten runs of 10-fold CV and uses their mean accuracy as the final measure of the performanc

D. Result and discussion

Database. If we are to build a database that can be successfullyused in computer vision and machine learning experiments aswell as cognitive science and neuroscience studies, data collec-tion must adhere to strict protocols. Because little is knownabout compound emotions, our goal is to minimize effects due tolighting, pose, and subtleness of the expression. All other vari-ables should, however, vary to guarantee proper analysis. An excellent style manual for science writers is.

Sample images of the 22 categories in the database: (A) neutral, (B) happy, (C) sad, (D) fearful, (E) angry, (F) surprised, (G) disgusted, (H) happilysurprised, (I) happily disgusted, (J) sadly fearful, (K) sadly angry, (L) sadly surprised, (M) sadly disgusted, (N) fearfully angry, (O) fearfully surprised, (P) fearfullydisgusted, (Q) angrily surprised, (R) angrily disgusted, (S) disgustedly surprised, (T) appalled, (U) hatred, and (V) awed.Du et al.PNAS—Published online March 31, 2014—E1455COMPUTER SCIENCESPHYSIOLOGYPNAS PLUS

Action Units Analysis. In their seminal work, Ekman and Friesen(12) defined a coding system that makes for a clear,

Fig. 3. Sample images

compactrepresentation of the muscle activation of a facial expression.Their Facial Action Coding System (FACS) is given by a set ofaction units (AUs). Each AU codes the fundamental actionsof individual or groups of muscles typically seen while pro-ducing facial expressions of emotion. For example, AU 4 definesthe contraction of two muscles resulting in the lowering of theeyebrows (with the emphasis being in the inner section). ThisAU is typically observed in expressions of sadness, fear, andanger (7).We FACS coded all of the images in our database. The con-sistently active AUs, present in more than 70 of the subjects ineach of the emotion categories, are shown in Table 1. Typicalintersubject variabilities are given in brackets; these correspondto AUs seen in some but not all individuals, with the percentagesnext to them representing the proportion of subjects that use thisAU when expressing this emotion.As expected, the AU analysis of the six basic emotions in ourdatabase is consistent with that given in ref. 12. The only smalldifference is in some of the observed intersubject variabilitygiven in parentheses—i.e., AUs that some but not all subjectsused when expressing one of the basic emotion categories; this isto be expected because our database incorporates a much largerset of subjects than the one in ref. 12. Also, all of the subjects wehave FACS coded showed their teeth when expressing happiness(AU 25), and this was not the case in ref. 12. Moreover, only halfof our subjects used AU 6 (cheek raiser) when expressing sad-ness, which suggests a small relevance of this AU as other studieshave previously suggested (13–15). Similarly, most of our sub-jects did not include AU 27 (mouth stretch) in fear, which seemsto be active only when this expression is exaggerated.Table 1 also lists the AUs for each of the compound emotioncategories. Note that the AUs of the subordinate categories areused to form the compound category unless there is a conflict.For example, lip presser (AU 24) may be used to express disgustwhile lips part (AU 25) is used in joy. When producing the facialexpression of happily disgusted, it is impossible to keep both. Inthis case, AU 24 is dropped. Fig. 2 shows this and five otherexamples (further illustrated in Table 1). The underlined AUs ofa compound emotion are present in both of their subordinatecategories. An asterisk indicates the AU does not occur in eitherof the basic categories and is, hence, novel to the compoundemotion. We did not find any such AU consistently used by mostsubjects; nevertheless, a few subjects did incorporate them, e.g.,AU 25 (lips part) in sadly disgusted. Additional examples aregiven inFig. S1, where we include a figure with the subordinaterelations for the nine remaining compound facial expressionsof emotion.

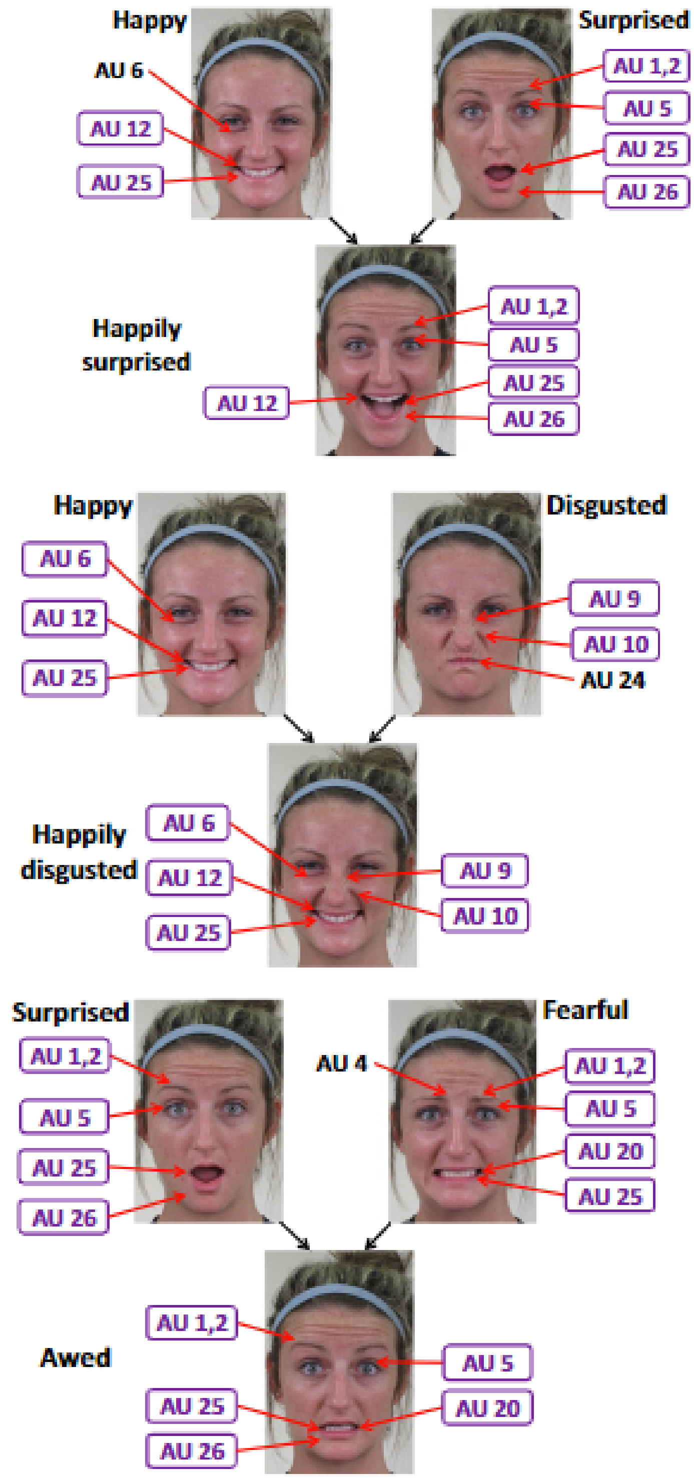
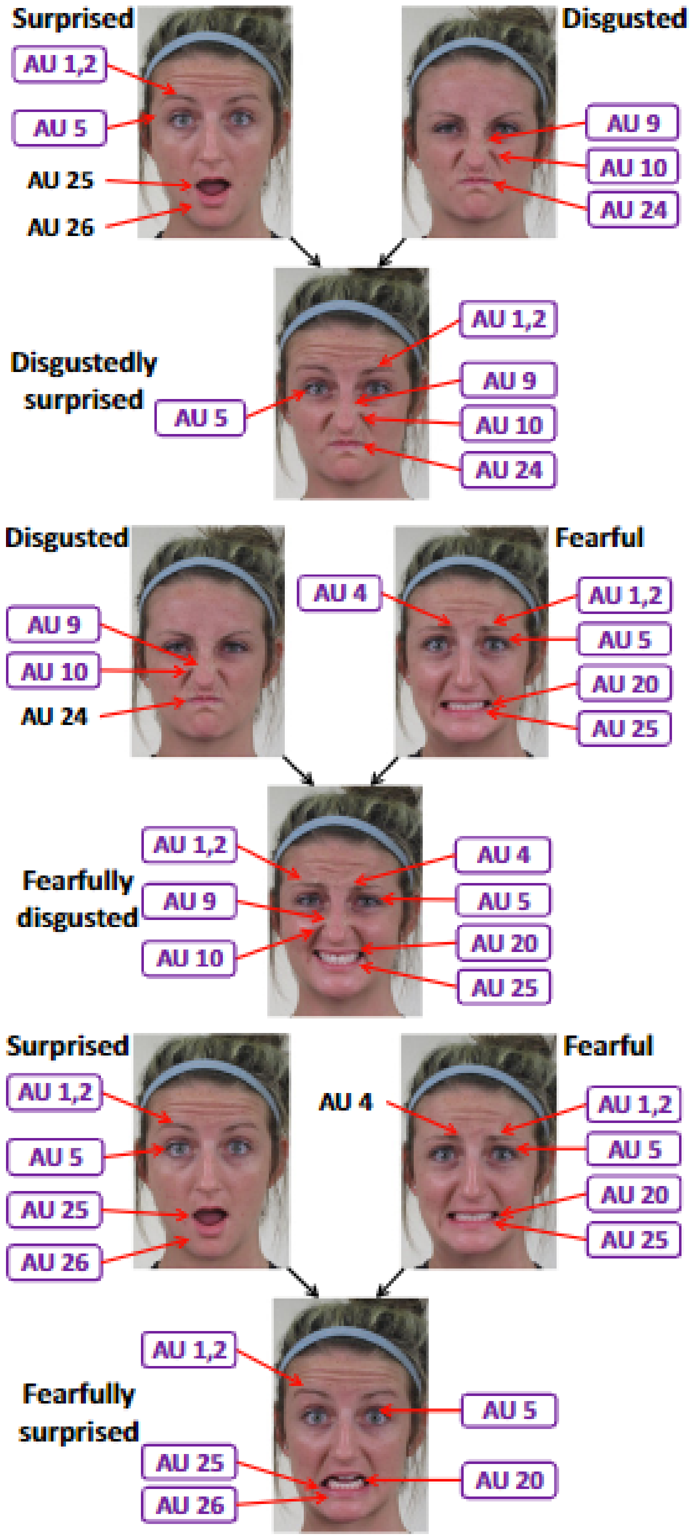


Table 1. Prototypical AUs observed in each basic and compound emotion category

Category	Prototypical (and variant AUs)
Happy	12, 25 [6 (51%)]
Sad	4, 15 [1 (60%), 6 (50%), 11 (26%), 17 (67%)]
Fearful	1, 4, 20, 25 [2 (57%), 5 (63%), 26 (33%)]
Angry	4, 7, 24 [10 (26%), 17 (52%), 23 (29%)]
Surprised	1, 2, 25, 26 [5 (66%)]
Disgusted	9, 10, 17 [4 (31%), 24 (26%)]
Happily surprised	1, 2, 12, 25 [5 (64%), 26 (67%)]
Happily disgusted	10, 12, 25 [4 (32%), 6 (61%), 9 (59%)]
Sadly fearful	1, 4, 20, 25 [2 (46%), 5 (24%), 6 (34%), 15 (30%)]
Sadly angry	4, 15 [6 (26%), 7 (48%), 11 (20%), 17 (50%)]
Sadly surprised	1, 4, 25, 26 [2 (27%), 6 (31%)]
Sadly disgusted	4, 10 [1 (49%), 6 (61%), 9 (20%), 11 (35%), 15 (54%), 17 (47%), 25 (43%)*]
Fearfully angry	4, 20, 25 [5 (40%), 7 (39%), 10 (30%), 11 (33%)*]
Fearfully surprised	1, 2, 5, 20, 25 [4 (47%), 10 (35%)*, 11 (22%)*, 26 (51%)]
Fearfully disgusted	1, 4, 10, 20, 25 [2 (64%), 5 (50%), 6 (26%)*, 9 (28%), 15 (33%)*]
Angrily surprised	4, 25, 26 [5 (35%), 7 (50%), 10 (34%)]
Angrily disgusted	4, 10, 17 [7 (60%), 9 (57%), 24 (36%)]
Disgustedly surprised	1, 2, 5, 10 [4 (45%), 9 (37%), 17 (66%), 24 (33%)]
Appalled	4, 10, [6 (25%)*, 9 (56%), 17 (67%), 24 (36%)]
Hatred	4, 10, [7 (57%), 9 (27%), 17 (63%), 24 (37%)]
Awed	1, 2, 5, 25, [4 (21%), 20 (62%), 26 (56%)]

Fig. 4. AUs used by a subset of the subjects are shown in brackets with the percentage of the subjects using this lesscommon AU in parentheses. The underlined AUs listed in the compound emotions are present in both their basiccategories. An asterisk (*) indicates the AU does not appear in either of the two subordinate categories.

We note obvious and unexpected production similarities betweensome compound expressions. Not surprisingly, the prototypical AUsof hatred and appalled are the same, because they are both varia-tions of angrily disgusted that can only be detected by the strengthin the activation of their AUs. More interestingly, there is a no-ticeable difference in over half the subjects who use AU 7 (eyelidtightener) when expressing hate. Also interesting is the differencebetween the expression of these two categories and that of angrilydisgusted, where AU 17 (chin raiser) is prototypical. These differ-ences make the three facial expressions distinct from one another.The facial expression of sadly angry does not include anyprototypical AU unique to anger, although its image seems to express anger quite clearly (Fig. 1K). Similarly, sadly fearful doesnot include any prototypical AU unique to sadness, but its imageis distinct from that of fear (Fig. 1DandJ)



Shown here are the AUs of six compound facial expressions of emotion. The AUs of the basic emotions are combined as shown to produce thecompound category. The AUs of the basic expressions kept to produce the compound emotion are marked with a bounding box. These relationships definethe subordinate classes of each category and their interrelatedness. In turn, these results define possible confusion of the compound emotion categories bytheir subordinates and vice versa. In order to analyze the statistics of emotion categories misclassification, we generated the confusion matrix of diverse emotion categories recognition (see Table 2). We first analyzed confusion matrix and we found that the average accuracy of compound emotion was 47.73It can be seen that the highway CNN architecture performed well on some particular emotion categories (e.g. angry, surprised, happy and happily disgusted) and others are challenging to recognize. Particularly when the compound emotions are confounded with the fundamental emotions that compose them, such as "happily surprised" category confused with "happy" and "fearfully surprise with an error rates of 13" and 9, respectively. "Happily disgusted" was recognized by a classification rate of 85 and it confused with happiness by 8. "sadly angry" category is confused with "angry" and "appaled" with an error rates of 64 and 18, respectively, (appal is the feeling of disgust and anger with the emphasis being on disgust). Furthermore, the expression of "sadly angry" (9) is very subtle, and some categories are often mistaken for each other such as: "Sad, sadly angry and angry", "fearful, fearfully angry and sadly fearful" and "Appalled, hatred and angry". Which means that it will be difficult for a highway CNN model to discover features to distinguish this emotion from basic categories that compose them. This may happen due to the complexity of the task. Frequently, the basic emotions are easier to recognize than compound emotion categories. That is makes them as a very challenging task. Not that, sometimes they were able to recognize the compound emotion but failed to recognize the basic one (or vice-versa).

Predicted Label (columns) / True Label (rows)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
a	.71	0	0	0	.24	.02	0	0	0	0	0	0	0	.03	0	0	0	0	0	0	0	0
b	0	.88	.01	.09	0	.01	0	0	0	0	0	0	.01	0	0	0	0	0	0	0	0	0
c	0	.13	.71	.01	0	.03	0	0	0	0	0	.09	.01	0	0	0	0	0	0	0	0	.02
d	0	.08	0	.85	0	0	0	0	.01	0	0	0	.03	0	0	0	0	0	.01	.02	0	0
e	.07	0	0	0	.49	.02	0	0	.02	0	0	0	0	.35	0	0	0	.01	.01	.03	0	0
f	.02	0	0	.01	.04	.71	0	.03	.03	0	.03	.03	.04	.03	.01	0	.01	0	0	0	0	.01
g	0	0	0	0	.03	0	.09	0	.03	0	0	0	0	.64	0	0	0	0	.02	.18	.01	0
h	0	0	0	0	0	.12	0	.45	0	0	.03	.01	0	0	.30	0	.07	0	0	0	0	.02
i	0	0	0	.03	.03	.04	0	0	.29	0	.04	0	.03	.16	.01	0	0	0	.10	.25	.02	0
j	.04	.01	0	.01	.02	.40	0	.03	0	.14	.02	.14	.07	.01	0	0	.02	.01	0	0	0	.08
k	0	0	0	.04	.01	.24	0	.03	.07	0	.32	0	.09	.08	.09	0	0	0	0	.03	0	0
l	.02	.01	.04	.01	0	.15	0	.04	0	0	0	.54	.01	0	0	0	.09	.01	0	0	0	.08
m	0	0	0	.09	.02	.24	0	0	.05	0	.05	.05	.41	.01	.01	0	0	.04	.01	.02	0	0
n	0	0	0	0	.01	0	0	0	.01	0	0	0	0	.95	0	0	0	0	0	.03	0	0
o	0	0	0	0	0	.02	0	.10	0	0	.08	0	0	0	.79	0	.01	0	0	0	0	0
p	0	0	0	0	0	0	0	0	.01	0	0	0	0	.30	0	.17	0	0	.14	.38	0	0
q	0	0	0	0	0	0	0	0	.01	0	0	.03	0	.01	0	0	.94	0	0	0	0	.01
r	.02	0	0	.01	.08	.02	0	0	.02	.01	0	.01	0	.07	0	0	0	.60	.12	.04	0	0
s	0	0	0	0	.01	0	0	0	.03	0	0	0	.02	.16	0	0	0	0	.20	.57	.01	0
t	0	0	0	0	0	0	0	0	.01	0	0	0	0	.20	0	0	0	0	.05	.74	0	0
u	0	0	0	0	.01	.01	0	0	.02	0	.02	0	0	.52	0	0	0	.01	.01	.30	.10	0
v	.02	0	0	0	.01	.08	0	.05	0	.01	.01	.17	0	0	0	0	.24	.02	0	0	0	.39

0-.9 | .10-.19 | .20-.29 | .30-.39 | .40-.49 | .50-.69 | Recognized emotion categories

E. Conclusion

In this paper, we have proposed a compound facial expression. recognition system based on highway CNN architecture. The whole system is composed of image preprocessing stage, feature extraction and expression classification stage. Experiment is performed on the CFEE database, which represents compound facial expression. The model implemented uses the specialized capabilities of the highway networks by building convolutional and max pooling layers before training the fully connected layers functions along with batch normalization functions to deliver the yield. The Highway CNN architecture is used for the first time in FER field. In the experimental results presented, the highway networks provides better performance on four emotion "angry", "surprised", "happy" and "happily disgusted". Also, it provides a poor performances on the other emotions. This is due to the fact that CFEE database contains small data which do not allow the network to learn features and make correct classifications of the input data. As future work, we want to test this approach in others largest databases like RAF-DB and iCV-MEFED

F. ACKNOWLEDGMENT

This work has been supported by the grant of National Center for Scientific and Technical Research (CNRST- Morocco): No. 757UIT for the first author and it has used the facial expression from the Compound Facial Expressions of Emotion Database.

G. REFERENCES

[1] T.-H.-S. Li, P.-H. Kuo, T.-N. Tsai, and P.-C. Luan, "CNN and LSTM based facial expression analysis model for a humanoid robot," IEEE Access, vol. 7, pp. 93998–94011, 2019. [2] M. R. Jeong and B. C. Ko, "Driver's facial expression recognition in realtime for safe driving," Sensors, vol. 18, no. 12, p. 4270, Dec. 2018. [3] M. S. Bouzakraoui, A. Sadiq, and N. Enneya, "A customer emotion recognition through facial expression using POEM descriptor and SVM classifier," in Proc. 2nd Int. Conf. Big Data, Cloud Appl., Mar. 2017, pp. 1–6. [4] Y. S. Su, H. Y. Suen, and K. E. Hung, "Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews," J. Real-Time Image Process., vol. 19, pp. 1011–1021, Jan. 2021. [5] M. Alhussein, "Automatic facial emotion recognition using weber local descriptor for e-Healthcare system," Cluster Comput., vol. 19, no. 1, pp. 99–108, Mar. 2016. [6] T. Ashwin and R. M. R. Guddeti, "Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks," Educ. Inf. Technol., vol. 25, pp. 1–29, Mar. 2019. [7] S. Saurav, S. Singh, R. Saini, and M. Yadav, "Facial expression recognition using improved adaptive local ternary pattern," in Proc. 3rd Int. Conf. Comput. Vis. Image Process. Singapore: Springer, 2020, pp. 39–52. [8] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D 2D CNN LSTM networks," Biomed. Signal Process. Control, vol. 47, pp. 312–323, Jan. 2019. [9] E. Avots, T. Sapiński, M. Bachmann, and D. Kamińska, "Audiovisual emotion recognition in wild," Mach. Vis. Appl., vol. 30, no. 5, pp. 975–985, 2019. [10] S. Oh, J.-Y. Lee, and D. K. Kim, "The design of CNN architectures for optimal six basic emotion classification using multiple physiological signals," Sensors, vol. 20, no. 3, p. 866, Feb. 2020. [11] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image Vis. Comput., vol. 27, no. 6, pp. 803–816, 2009. [12] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local directionbased robust features and deep belief network," IEEE Access, vol. 5, pp. 4525–4536, 2017. [13] J. Chen, R. Xu, and L. Liu, "Deep peak-neutral difference feature for facial expression recognition," Multimedia Tools Appl., vol. 77, no. 22, pp. 29871–29887, Nov. 2018. [14] I. M. Revina and W. R. S. Emmanuel, "Face expression recognition with the optimization based multi-SVNN classifier and the modified LDP features," J. Vis. Commun. Image Represent., vol. 62, pp. 43–55, Jul. 2019. [15] M. Nazir, Z. Jan, and M. Sajjad, "Facial expression recognition using histogram of oriented gradients based transformed features," Cluster Comput., vol. 21, no. 1, pp. 539–548, Mar. 2018. [16] M. Ghosh, T. Kundu, D. Ghosh, and R. Sarkar, "Feature selection for facial emotion recognition using late hill-climbing based memetic algorithm," Multimedia Tools Appl., vol. 78, no. 18, pp. 25753–25779, Sep. 2019. [17] S. Zhou, G. Feng, and J. Xie, "Facial expression recognition based on classification tree," in Proc. Chin. Conf. Biometric Recognit. Cham, Switzerland: Springer, 2014, pp. 128–135. [18] H. Boughrara, M. Chtourou, C. B. Amar, and L. Chen, "Facial expression recognition based on a mlp neural network using constructive training algorithm," Multimedia Tools Appl., vol. 75, no. 2, pp. 709–731, 2016. [19] T. Li, C. Du, T. Naren, Z. Chen, S. Liu, J. Zhou, and X. Xu, "Using feature points and angles between them to recognise facial expression by a neural network approach," IET Image Process., vol. 12, no. 11, pp. 1951–1955, Nov. 2018. [20] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," Pattern Recognit. Lett., vol. 120, pp. 69–74, Apr. 2019. [21] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," Sensors, vol. 17, no. 4, p. 712, Mar. 2017. [22] K. Bahreini, W. van der Vegt, and W. Westera, "A fuzzy logic approach to reliable real-time recognition of facial emotions," Multimedia Tools Appl., vol. 78, no. 14, pp. 18943–18966, Jul. 2019. [23] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," IEEE Trans. Image Process., vol. 28, no. 5, pp. 2439–2450, May 2018. [24] A. R. Rivera, R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1740–1752, May 2012. [25] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE Trans. Image Process., vol. 19, no. 6, pp. 1635–1650, Jun. 2010. [26] T. Jabid, M. H. Kabir, and O. Chae, "Facial expression recognition using local directional pattern (LDP)," in Proc. IEEE Int. Conf. Image Process., Sep. 2010, pp. 1605–1608. [27] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, "Local directional ternary pattern for facial expression recognition," IEEE Trans. Image Process., vol. 26, no. 12, pp. 6006–6018, Dec. 2017. [28] F. Ahmed and E. Hossain, "Automated facial expression recognition using gradient-based ternary texture patterns," Chin. J. Eng., vol. 2013, pp. 1–8, Dec. 2013. [29] R. P. Holder and J. R. Tapamo, "Improved gradient local ternary patterns for facial expression recognition," EURASIP J. Image Video Process., vol. 2017, no. 1, p. 42, Dec. 2017. [30] M. Guo, X. Hou, Y. Ma, and X. Wu, "Facial expression recognition using ELBP based on covariance matrix transform in KLT," Multimedia Tools Appl., vol. 76, no. 2, pp. 2995–3010, 2017. [31] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W.

Gao, “WLD: A robust local image descriptor,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1705–1720, Sep. 2010. [32] S. A. Khan, A. Hussain, and M. Usman, “Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features,” Multimedia Tools Appl., vol. 77, no. 1, pp. 1133–1165, Jan. 2018. [33] A. Mahmood, S. Hussain, K. Iqbal, and W. S. Elkilani, “Recognition of facial expressions under varying conditions using dual-feature fusion,” Math. Problems Eng., vol. 2019, pp. 1–12, Aug. 2019. [34] S. L. Happy and A. Routray, “Automatic facial expression recognition using features of salient facial patches,” IEEE Trans. Affective Comput., vol. 6, no. 1, pp. 1–12, Jan. 2014. [35] S. Kherchaoui and A. Houacine, “Facial expression identification using gradient local phase,” Multimedia Tools Appl., vol. 78, no. 12, pp. 16843–16859, Jun. 2019.