A novel database of children's spontaneous facial expressions (LIRIS-CSE)

Rizwan Ahmed Khan, Arthur Crenn, Alexandre Meyer, Saida Bouakaz

A novel database of children's spontaneous facial expressions (LIRIS-CSE)

Rizwan Ahmed khan[a,b,*], Crenn Arthur[a], Alexandre Meyer[a], Saida Bouakaz[a]

[a] *Faculty of IT, Barrett Hodgson University, Karachi, Pakistan.*
[b] *LIRIS, Université Claude Bernard Lyon1, France.*

Abstract

Computing environment is moving towards human-centered designs instead of computer centered designs and human's tend to communicate wealth of information through affective states or expressions. Traditional Human Computer Interaction (HCI) based systems ignores bulk of information communicated through those affective states and just caters for users intentional input. Generally, for evaluating and benchmarking different facial expression analysis algorithms, standardized databases are needed to enable a meaningful comparison. In the absence of comparative tests on such standardized databases it is difficult to find relative strengths and weaknesses of different facial expression recognition algorithms. In this article we present a novel video database for Children's Spontaneous facial Expressions (LIRIS-CSE). Proposed video database contains six basic spontaneous facial expressions shown by 12 ethnically diverse children between the ages of 6 and 12 years with mean age of 7.3 years. To the best of our knowledge, this database is first of its kind as it records and shows spontaneous facial expressions of children. Previously there were few database of children expressions and all of them show posed or exaggerated expressions which are different from spontaneous or natural expressions. Thus, this database will be a milestone for human behavior researchers. This database will be a excellent resource for vision community for benchmarking and comparing results. In this article, we have also proposed framework for automatic expression recognition based on convolutional neural network (CNN) architecture with transfer learning approach. Proposed architecture achieved average classification accuracy of 75% on our proposed database i.e. LIRIS-CSE.

Keywords: `Facial expressions database, spontaneous expressions, convolutional neural network, expression recognition, transfer learning.`

*Corresponding author

1. Introduction

Computing paradigm has shifted from computer-centered computing to human-centered computing [1]. This paradigm shift has created tremendous opportunity for computer vision research community to propose solution to existing problems and invent ingenious applications and products which were not though of before. One of the most important property of human-centered computing interfaces is the ability of machines to understand and react to social and affective or emotional signals [2, 3].

Mostly humans express their emotion via facial channel, also known as facial expressions [3]. Humans are blessed with the amazing ability to recognize facial expression robustly in real-time but for machines it still is a difficult task to decode facial expressions. Variability in pose, illumination and the way people show expressions across cultures are some of the parameters that make this task more difficult [4].

Another problem that hinders the development of such system for real world applications is the lack of databases with natural displays of expressions [5]. There are number of publicly available benchmark databases with posed displays of the six basic emotions [6] i.e. happiness, anger, disgust, fear, surprise and sadness, exist but there is no equivalent of this for spontaneous / natural basic emotions. While, it has been proved that spontaneous facial expressions differ substantially from posed expressions [7].

Another issue with most of publicly available databases is absence of children in recorded videos or images. Research community has put lot of efforts to built databases of emotional videos or images but almost all of them contain adult emotional faces [8, 9]. By excluding children's stimuli in publicly available databases vision research community not only restricted itself to application catering only adults but also produced limited study for the interpretation of expressions developmentally [10].

2. Publicly available databases of children's emotional stimuli

To the best of our knowledge there are only three publicly available databases that contains children emotional stimuli / images. They are:

1. The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS) [11]
2. The Dartmouth Database of Childrens Faces [12]
3. The Child Affective Facial Expression (CAFE) [10]

The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS) [11] database has 482 emotional frames containing expressions of "fear", "anger", "happy" and "sad" with two gaze conditions: direct and averted gaze. Children that posed for this database were between 10 and 17 years of age. The databases is validated by 20 adult raters.

The Dartmouth Database of children Faces [12] contains emotional images (six basic emotions) of 40 male and 40 female Caucasian children between the

Figure 1: **Six universal expressions**: first row show example images from the Dartmouth database [12]. Second row show emotional images from the Child Affective Facial Expression (CAFE) [10], while last row show emotional images from the movie clip of **our proposed database, LIRIS-CSE**. First column corresponds to expression of "happiness", second column corresponds to expression of "surprise", third column corresponds to expression of "sadness", fourth column corresponds to expression of "anger", fifth column corresponds to expression of "fear" and last column corresponds to expression of "disgust". Expressions in our proposed database are spontaneous and natural and can easily be differentiated from posed / exaggerated expressions of the other two database.

ages of 6 and 16 years. All facial images in the database were assessed by at least human 20 raters for facial expression identifiability and intensity. Expression of happy was most accurately identified while fear was least accurately identified by human raters. Human raters correctly classified 94.3% of the happy faces while expression of fear was correctly identified in 49.08% of the images, least identifiable by human raters. On average human raters correctly identified expression in 79.7% of the images. Refer Figure 1 for examples images from the database.

The Child Affective Facial Expression (CAFE) database [10] is composed of 1192 emotional images (six basic emotions and neutral) of 2 to 8 years old children. Children that posed for this database were ethnically and racially diverse. Refer Figure 1 for examples frames from the database.

2.1. Weaknesses of publicly available databases of children's emotional stimuli

Although above describe children expression databases are diverse in terms of pose, camera angles and illumination but have following drawbacks:

1. Above mentioned databases contain posed expressions and as mentioned before that spontaneous or natural facial expressions differ substantially from posed expressions as they exhibit real emotion whereas, posed expressions are fake and disguise inner feelings [7, 13].

2. All of these databases contains only static images / mug shots with expression at peak intensity. According to study conducted by psychologist Bassili [14] it was concluded that facial muscle motion/movement is fundamental to the recognition of facial expressions. He also concluded that human can robustly recognize expressions from video clip than by just looking at mug shot.
3. All of the above mentioned databases for children facial expressions present few hundred to maximum 1200 static frames. In current era where requirement of amount of data for learning a concept / computational model (deep learning [15]) has increased exponentially, only few hundred static images are not enough.

Figure 2: **Example of variations in illumination condition and background**. Clips in column 1, 3 and 4 were recorded in home condition while image in column 2 was recorded in lab / classroom environment.

Generally, for evaluating and benchmarking different facial expression analysis algorithms, standardized databases are needed to enable a meaningful comparison. In the absence of comparative tests on such standardized databases it is difficult to find relative strengths and weaknesses of different facial expression recognition algorithms. Thus, it is utmost important to develop natural / spontaneous emotional database contains children movie clip / dynamic images. This will allow research community to built robust system for children's natural facial expression recognition. Thus, our *contributions in this study are two-fold*:

1. We are presenting a novel emotional database (LIRIS-CSE) that contains 208 movie clip / dynamic images of 12 ethnically diverse children showing spontaneous expressions in two environments, i.e. 1) lab / classroom environment 2) home environment (refer Figure 2).
2. We have also proposed a framework for automatic facial expression recognition based on convolutional neural network (CNN) architecture with transfer learning approach. Proposed architecture achieved average classification accuracy of 75% on our proposed database.

3. Novelty of proposed database (LIRIS-CSE)

To overcome above mentioned drawbacks of databases of children's facial expression (refer Section 2.1), we are presenting a novel emotional database that contains movie clip / dynamic images of 12 ethnically diverse children. This

Figure 3: **Example of expression transition**. First row shows example of expression of "Disgust". Second row shows expression of "Sadness", while third row corresponds to expression of "Happiness".

unique database contains spontaneous / natural facial expression of children in diverse settings (refer Figure 2 to see variations in recording scenarios) showing six universal or prototypic emotional expressions ("happiness", "sadness", "anger", "surprise", "disgust" and "fear") [16, 17]. Children are recorded in constraint free environment (no restriction on head movement, no restriction on hands movement, free sitting setting, no restriction of any sort) while they watched specially built / selected stimuli. This constraint free environment allowed us to record spontaneous / natural expression of children as they occur. The database has been validated by 22 human raters. Details of recording parameters are presented in Table 2. In comparison with above mentioned databases for children facial expressions that have only few hundred images, our database (LIRIS-CSE) contains 26 thousand (26,000) frames of emotional data, refer Section 4.3 for details.

The spontaneity of recorded expressions can easily be observed in Figure 1. Expressions in our proposed database are spontaneous and natural and can easily be differentiated from posed / exaggerated expressions of the other two databases. Figure 3 shows facial muscle motion / transition for different spontaneous expressions.

3.1. Participants

In total 12 (five male and seven female children) ethnically diverse children between the ages of 6 and 12 years with mean age of 7.3 years participated in our database recording session. 60 % of recordings are done in classroom / lab environment, while 40% of the clips in the database are recorded in home

conditions. Recording children in two different environments has been done to have different background and illumination conditions in the recorded database. Refer Figure 2 for example images with different backgrounds and illumination conditions.

4. Database acquisition details

First step for the creation of proposed spontaneous expressions database was the selection of visual stimuli that can induce emotions in children. Considering ethical reasons and young age of children we carefully selected stimuli and removed any stimuli that can have long term negative impact on the children. Due to these ethical reasons we did not include emotion inducer clips for the negative expression of "anger" and selected very few clips to induce emotion of "fear" and "sadness". The same has been practiced before by Valstar et. al [18]. Due to this very reason the proposed database contains more emotional clips of expressions of "happiness" and "surprise". Although there were no emotion inducer clips for the expression of "anger" but still database contains very few clips where children show expression of "anger" (refer Figure 1) due to the fact that young children use expressions of "disgust" and "anger" interchangeably [19].

4.1. Emotion inducing stimuli

We either selected only animated cartoon / movies or small video clips of kids doing funny actions to stimuli list. The reasons for selecting videos to induce emotions in children are as follows:

1. All the selected videos for inducing emotions contains audio as well. Video stimuli along with audio gives immersive experience, thus is powerful emotion inducer [20].

2. Video stimuli provides more engaging experience then static images, restricting undesirable head movement.

3. Video stimuli can evoke emotions for a longer duration. This helped us in recording and spotting children facial expressions.

List of stimuli selected as emotion inducers are presented in Table 1. Total running length of selected stimuli is 17 minutes and 35 seconds. One of the consideration for not selecting more stimuli is to prevent children's lose of interest or disengagement over time [21].

4.2. Recording setup

Inspired by Li et al. [20], we setup high speed webcam, mounted at the top of laptop with speaker output, at a distance of 50 cm. As explained above, The audio output enhanced the visual experience of a child, thus helping us induce emotions robustly. Recording setup is illustrated in Figure 4. As mentioned in

Sr.No:	Induced Expression	Source	Clip Name	Time
1	Disgust	YouTube	Babies Eating Lemons for the First Time Compilation 2013	42 Sec
2	Disgust	YouTube	On a Plane with Mr Bean (Kid puke)	50 Sec
3	Fear and surprise	YouTube	Ghoul Friend - A Mickey Mouse Cartoon - Disney Shows	50 Sec
4	Fear	YouTube	Mickey Mouse - The Mad Doctor - 1933	57 Sec
5	Fear & surprise	Film	How To Train Your Dragon (Monster dragon suddenly appears and kills small dragon)	121 Sec
6	Fear	Film	How To Train Your Dragon (Monster dragon throwing fire)	65 Sec
7	Fear	YouTube	Les Trois Petits Cochons	104 Sec
8	Happy	YouTube	Best Babies Laughing Video Compilation 2014 (three clips)	59 Sec
9	Happy	YouTube	Tom And Jerry Cartoon Trap Happy	81 Sec
10	Happy, surprise & fear	YouTube	Donald Duck- Lion Around 1950	40 Sec
11	Happily surprised	YouTube	Bip Bip et Coyote - Tired and feathered	44 Sec
12	Happily surprised	YouTube	Donald Duck - Happy Camping	53 Sec
13	Sad	YouTube	Fox and the Hound - Sad scene	57 Sec
14	Sad	YouTube	Crying Anime Crying Spree 3	14 Sec
15	Sad	YouTube	Bulldog and Kitten Snuggling	29 Sec
16	Surprise	Film	Ice Age- Scrat's Continental Crack-Up	32 Sec
17	Surprise & happy	Film	Ice Age (4-5) Movie CLIP - Ice Slide (2002)	111 Sec
18	Happy	YouTube	bikes funny (3)	03 Sec
19	Happy	YouTube	bikes funny	06 Sec
20	Happy	YouTube	The Pink Panther in 'Pink Blue Plate	37 Sec
Total running length of stimuli = 17 minutes and 35 Seconds				

Table 1: Stimuli used to induce spontaneous expression

Section 3.1, children were recorded in two different environments i.e. classroom / lab environment and home environment. Details of recording parameters are

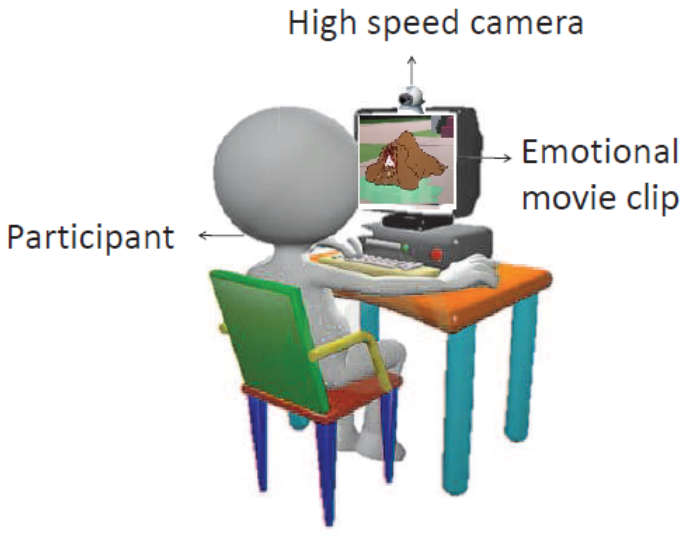

Figure 4: **Database recording setup**. Children watch emotions inducing stimuli while high speed webcam records their video focusing on face. Figure inspired by [20].

Subject	**Recording Environment**	**FPS**	**Video Resolution**
S1-S7	Classroom	25	800 * 600
S8-S10	Home	25	720 * 480
S11-S12	Home	25	1920 * 1080

Table 2: Database videos recording parameters

presented in Table 2.

4.3. Video segmentation

After recording video for each child we carefully examined the recoding and removed any unnecessary recorded part, usually at the beginning and at the end of video recording. As the video (with audio) stimuli that children watched was the combination of different emotional videos (refer Section 4.1 for the details of visual stimuli), our recorded video contained whole spectrum of expressions in one single video. We then manually segmented one single video recording of each child into segments of small video chunks / clips such that each video clip show one pronounced expression. Refer Figure 3 to see results after segmentation process. It can be observed from the referred figure that each small video clip contains neutral expression at the beginning, then shows onset of an expression, and finishes when expression is visible at its peak along with some frames after peak expression frame. Total number of small video clips, each containing specific expression, present in our database (LIRIS-CSE) are 208. Total running length of segmented clips in presented database (videos having children facial

expressions) is seventeen minutes and twenty-four seconds. That makes total of around 26 thousand (26,000) frames of emotional data, considering recording is done at 25 frames / second, refer Table 2.

Figure 5: **Blended Expressions**. Example images that show co-occurrence of more than one expressions in a single clip of the database. First row present frames from a clip that show occurrence of expressions of "sadness" and "anger". Similarly, second row shows co-occurrence of expressions of "surprise" and "happiness".

There are seventeen (17) video clips present in this database that have two labels, for example "happily surprised", " Fear surprise" etc. This is due to the fact that for young children different expressions co-occur / blended expressions [22, 19] or a visual stimuli was so immersive that transition from one expression to another expression was not pronounced. Refer Figure 5 to see example images from segmented video segments that show blended expressions.

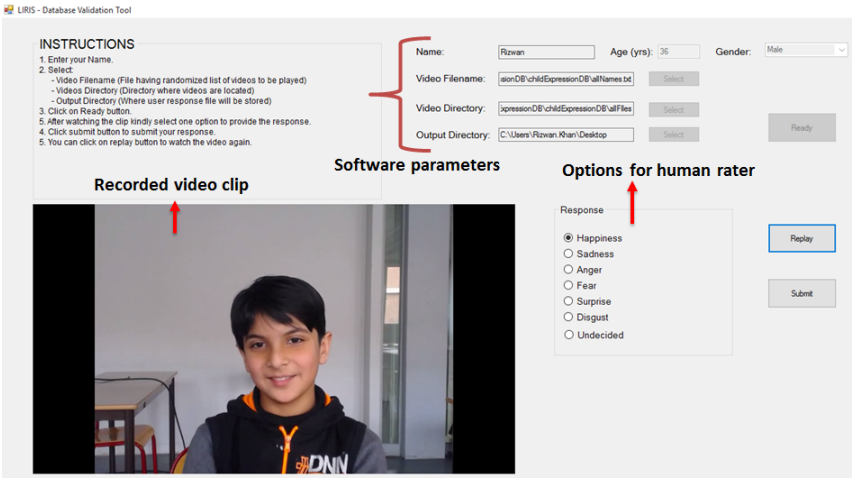

Figure 6: **Validation tool.** Screen capture of validation tool used for collecting ratings / labels from human evaluators for each video clip.

4.4. Database validation

The database has been validated by 22 human raters / evaluators between the ages of 18 and 40 years with mean age of 26.8 years. 50% of database raters / evaluators were in the age bracket of [18 - 25] years and rest of 50% were in the age bracket of [26 - 40] years. Human evaluators who were in the age bracket of [18 - 25] years were university students and other group of evaluators were university faculty members. Human raters / evaluators were briefed about the experiment before they started validating the database.

For database validation purpose we built software that played segmented video (in random order) and records human evaluator choice of expression label. The screen capture of the software is presented in Figure 6. If required, evaluator can play video multiple times before recoding their choice for any specific video.

In summary, instructions given to human raters / evaluators were following:

1. Watch carefully each segmented video and select expression that is shown in the played segmented video.

2. If played video did not show any visible / pronounced expression, selected an option of "undecided". Each video can be played multiple times without any upper bound on number of times video to be played.

3. Once expression label / response is submitted for a played segmented video then this label can not be edited.

Figure 7: **Confusion Marix**. Rows: induced intended expressions (average). Columns: expression label given by human raters (average). Diagonal represents agreement between induced intended expressions and expression label given by human raters, with darker colors representing greater agreement.

4.4.1. Validation data analysis

After validation data collection, we performed statistical analysis on the gathered data and calculated confusion matrix. Refer Figure 7 to see calculated

confusion matrix. Rows in the referred confusion matrix show induced intended expressions (average%), while columns show expression label given by human raters / evaluators (average %). Diagonal values represent agreement between induced intended expressions and expression label given by human evaluators, with darker colors representing greater agreement.

As per calculated results expression of "happy" was most correctly spotted by evaluators, with average accuracy of 73.2%. On the other hand expression of " fear" was least correctly identified by evaluators, with average accuracy of 40.1%. These results are consistent with results from [12, 16]. We did not include expression of "anger" in analysis as there is only one movie clip with anger expression in the database.

Expression of "fear" which is least identified, is often perceptually mixed with expressions of "surprise" and "disgust". As mentioned above, this is due to the fact that for young children different expressions co-occur (blended expressions) [22, 19] or a visual stimuli was so immersive that transition from one expression to another expression was not pronounced. Refer Figure 5 to see example images from segmented video segments that show blended expressions.

Overall average accuracy of human evaluators / raters is 55.3%. As per study published by Matsumoto et al. [23] human's usually can spot expressions correctly 50% of the time and the easiest expression for human's to identify are "happy" and "surprise". These results conforms well with the results that we obtained from human evaluators as expression of "happy" was most correctly identified while average accuracy of human evaluators raters is also around 50% (55.3% to be exact).

5. Database availability

The novel database of Children's Spontaneous Expressions (LIRIS-CSE) is available for research purposes only. It can be downloaded by researcher / lab after signing End User License Agreement (EULA). Website to download LIRIS-CSE database is: https://childrenfacialexpression.projet.liris.cnrs.fr/.

6. Automatic recognition of affect, a transfer learning based approach

In order to provide benchmark machine learning / automatic classification of expression results on our database (LIRIS-CSE), we have done experiment based on transfer learning paradigm. Usually machine learning algorithms make prediction on data that is similar to what algorithm is trained on; training and test data are drawn from same distribution. On the contrary transfer learning allows distributions used in training and testing to be different [24]. We used transfer learning approach in our experiment due to following facts:

1. We wanted to benefit from deep learning model that has achieved high accuracy on recognition tasks that take image as input i.e. ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [25], and is available for research purposes.

2. It requires large database to train deep learning model from the scratch [26]. In a given scenario, we can not train deep learning model from the very beginning.

6.1. Convolutional Neural Network (CNN)

Researchers have been successful in developing models that can recognize affect robustly [4, 27, 28]. Recently, most of successful models are based on deep learning approach [29, 30, 31], specifically on Convolutional Neural Network (CNN) architecture [15]. CNNs are class of deep, feed forward neural networks that have shown robust results for applications involving visual input i.e image / object recognition [32], face expression analysis [31], semantic scene analysis / semantic segmentation [32, 33], gender classification [34] etc.

The architecture of CNN was first proposed by LeCun [15]. It is a multi-stage or multi-layer architecture. This essentially means there are multiple stages in CNN for feature extraction. Every stage in the network has an input and output which is composed of arrays known as feature maps. Every output feature map consists of patterns or features extracted on locations of the input feature map. Every stage is made up of layers after which classification takes place [35, 36, 37]. Generally, these layers are:

1. Convolution layer: This layer makes use of filters, which are convolved with the image, producing activation or feature maps.
2. Feature Pooling layer: This layer is inserted to reduce the size of the image representation, to make the computation efficient. The number of parameters is also reduced which in turn controls over-fitting.
3. Classification layer: This is the fully connected layer. This layer computes the probability / score learned classes from the extracted features from convolution layer in the preceding steps.

6.2. VGGNet architecture and transfer learning

Since 2012, deep Convolution Networks (ConvNets) have become a focus of computer vision scientists. Various architectures were proposed to achieve higher accuracy for a given tasks. For example, best submissions to ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [25], [39, 40] proposed to use a smaller receptive window size / smaller filter size and smaller stride in the first convolutional layer. Generally, ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has served as a platform for advancements in deep visual recognition architectures.

The best proposed ConvNets architectures for ILSVRC 2014 competition were GoogleNet (a.k.a. Inception V1) from Google [41] and VGGNet by Simonyan and Zisserman [38]. GoogleNet contains 1 x 1 Convolution at the middle of the network and global average pooling was used at the end of the network instead of using fully connected layers, refer Section 6.1 for discussion on different types of layers. VGGNet consists of 16 convolutional layers (VGG16). It is one of the most appealing framework because of its uniform architecture, refer

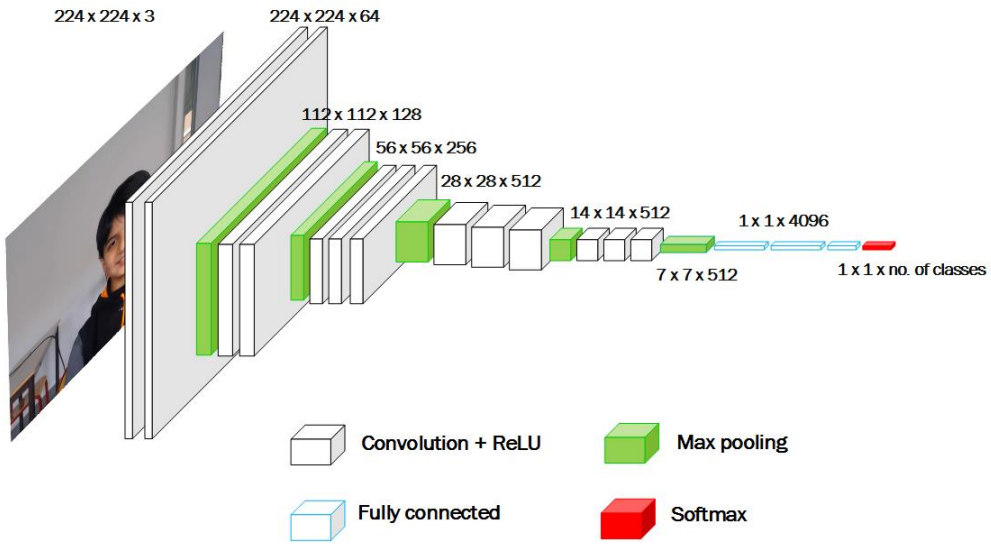

Figure 8: An illustration of VGG16 architecture [38].

Figure 8. It's pre-trained model is freely available for research purpose, thus making a good choice for transfer learning.

VGG16 architecture (refer Figure 8) takes image of 224 x 224 with the receptive field size of 3 x 3. The convolution stride is 1 pixel and padding is 1 (for receptive field of 3 x 3). There are two fully connected layers with 4096 units each, the last layer is a softmax classification layer with x units (representing x classes / x classes to recognize) and the activation function is the rectified linear unit (ReLU). The only downside of VGG16 architecture is its huge number of trainable parameters. VGG16 consists of 138 million parameters.

6.3. Experimental framework and results

As discussed earlier, CNN requires large database to learn concept [42, 26], making it impractical for different applications. This bottleneck is usually avoided using transfer learning technique [43]. Transfer learning is a machine learning approach that focuses on ability to apply relevant knowledge from previous learning experiences to a different but related problem. We have used transfer learning approach to built framework for expression recognition using our proposed database (LIRIS-CSE) as the size of our database is not sufficiently large to robustly train all layers of CNN from the very beginning. We used pre-trained VGG model (VGG16, a 16 layered architecture) [38], which is a deep convolutional network trained for object recognition [32]. It is developed and trained by Oxford University's Visual Geometry Group (VGG) and shown to achieve robust performance on the ImageNet dataset [44]for object recognition. Refer Section 6.2 for discussion on VGG16.

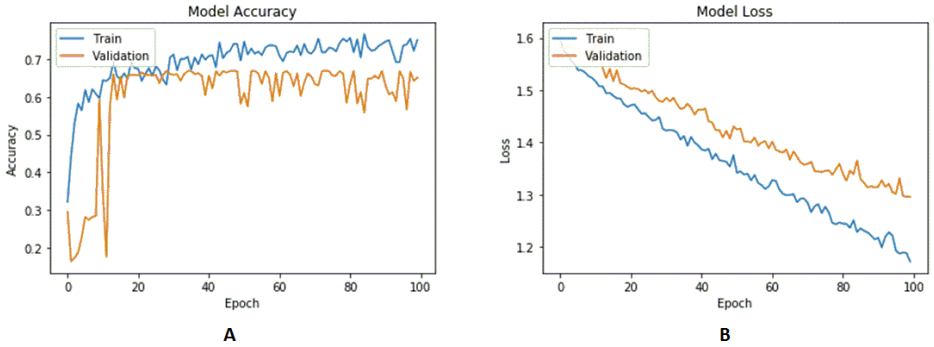

Figure 9: CNN model learning: (A) Training accuracy vs Validation accuracy (B) Training loss vs Validation loss.

We replaced last fully connected layer of VGG16 pre-trained model with dense layer having five outputs. This makes 5005 trainable parameters. Number of output of last dense layer corresponds to number of classes to be recognized, in our experiment we learned concept of five classes i.e. five expression to be recognized (out of six universal expression, expression of "anger" was not included in this experiment as there are few (one) clip(s) for "anger", for explanation see Section 4). We trained last dense layer with images (frames from videos) from our proposed database using softmax activation function and ADAM optimizer [45].

Our proposed database consists of video but for this experiment we extracted frames from videos and fed them to above described ConvNet architecture. We used 80% of frames for training and 10% of frames for validation process. With above mentioned parameters, proposed CNN achieved average expression accuracy of 75% on our proposed database (five expressions). Model accuracy and loss curves are shown in Figure 9.

7. Conclusion

In this article we presented novel database for Children's Spontaneous Expressions (LIRIS-CSE). The database contains six universal spontaneous expression shown by 12 ethnically diverse children between the ages of 6 and 12 years with mean age of 7.3 years. There were five male and seven female children. 60% of recordings were done in classroom / lab environment and 40% of the clips in the database were recorded in home conditions.

The LIRIS-CSE database contains 208 small video clips (on average each clip is 5 seconds long), with each clip containing one specific expression. Clips have neutral expression / face at the beginning of clip, then it show onset of an expression, and finishes when expression is visible at its peak along with some frames after peak expression frame. The database has been validated by 22 human raters / evaluators between the ages of 18 and 40 years.

To the best of our knowledge, this database is first of its kind as it records and shows six (rather five as expression of "anger" is spotted/recorded only once) universal spontaneous expressions of children. Previously there were few image databases of children expressions and all of them show posed or exaggerated expressions which are different from spontaneous or natural expressions. Thus, this database will be a milestone for human behavior researchers. This database will be a excellent resource for vision community for benchmarking and comparing results.

For benchmarking automatic recognition of expression we have also provided result using Convolutional Neural Network (CNN) architecture with transfer learning approach. Proposed approach obtained average expression accuracy of 75% on our proposed database, LIRIS-CSE (five expressions).

References

References

[1] M. Pantic, A. Pentland, A. Nijholt, T. Huang, Human computing and machine understanding of human behavior: A survey, 2006.

[2] M. Pantic, Machine analysis of facial behaviour: naturalistic and dynamic behaviour, Philosophical Transactions of the Royal Society B: Biological Sciences, 364 (1535) (2009) 3505–3513.

[3] B. C. Ko, A brief review of facial emotion recognition based on visual information, Sensors 18 (2) (2018) 2–20. doi:https://doi.org/10.3390/s18020401.

[4] R. A. Khan, A. Meyer, H. Konik, S. Bouakaz, Framework for reliable, real-time facial expression recognition for low resolution images, Pattern Recognition Letters 34 (10) (2013) 1159 – 1168. doi:https://doi.org/10.1016/j.patrec.2013.03.022.
URL http://www.sciencedirect.com/science/article/pii/S0167865513001268

[5] M. Valstar, M. Pantic, Induced disgust, happiness and surprise: an addition to the MMI facial expression database, in: International Language Resources and Evaluation Conference, 2010.

[6] P. Ekman, Universals and cultural differences in facial expressions of emotion, in: Nebraska Symposium on Motivation, Lincoln University of Nebraska Press, 1971, pp. 207–283.

[7] B. M. S., L. G., B. B., S. T. J., M. J. R., A prototype for automatic recognition of spontaneous facial actions, in: Advances in Neural Information Processing Systems, 2002.

[8] L. P., C. J. F., K. T., S. J., A. Z., M. I., The extended cohn-kande dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression., in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2010.

[9] M. Pantic, M. F. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: IEEE International Conference on Multimedia and Expo, 2005.

[10] L. V., T. C., The child affective facial expression (CAFE) set: validity and reliability from untrained adults, Frontiers in Psychology 5 (1532). doi:http://doi.org/10.3389/fpsyg.2014.01532.

[11] H. L. Egger, D. S. Pine, E. Nelson, E. Leibenluft, M. Ernst, K. E. Towbin, A. Angold, The nimh child emotional faces picture set (NIMH-ChEFS): a new set of children's facial emotion stimuli, International Journal of Methods in Psychiatric Research (2011) 145–156.

[12] K. A. Dalrymple, J. Gomez, B. Duchaine, The dartmouth database of childrens faces: Acquisition and validation of a new face stimulus set, PLOS ONE 8 (11) (2013) 1–7. doi:10.1371/journal.pone.0079131.
URL https://doi.org/10.1371/journal.pone.0079131

[13] Q. Gan, S. Nie, S. Wang, Q. Ji, Differentiating between posed and spontaneous expressions with latent regression bayesian network, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[14] J. N. Bassili, Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face, Journal of Personality and Social Psychology 37 (1979) 2049–58.

[15] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: IEEE International Symposium on Circuits and Systems, 2010.

[16] P. Ekman, W. V. Friesen, Pictures of facial affect, 1976.

[17] P. Ekman, Facial expression of emotion, Psychologist 48 (1993) 384–392.

[18] M. F. Valstar, M. Pantic, Induced disgust, happiness and surprise: an addition to the mmi facial expression database, in: Proceedings of Int'l Conf. Language Resources and Evaluation, Workshop on EMOTION, Malta, 2010, pp. 65–70.

[19] W. SC, R. JA., Children's recognition of disgust in others, Psychological Bulletin 139 (2013) 271–299.

[20] X. Li, T. Pfister, X. Huang, G. Zhao, M. Pietikinen, A spontaneous micro-expression database: Inducement, collection and baseline, in: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1–6. doi:10.1109/FG.2013.6553717.

[21] R. A. Khan, A. Meyer, H. Konik, S. Bouakaz, Exploring human visual system: study to aid the development of automatic facial expression recognition framework, in: Computer Vision and Pattern Recognition Workshop, 2012.

[22] M. W. Sullivan, M. Lewis, Emotional expressions of young infants and children: A practitioner's primer, Infants & Young Children (2003) 120–142.

[23] D. Matsumoto, H. S. Hwang, Reading facial expressions of emotion, Tech. rep., American Psychological Association (APA) , Psychological Science Agenda (2011).

[24] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345–1359. doi:10.1109/TKDE.2009.191.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision (IJCV) 115 (3) (2015) 211–252. doi:10.1007/s11263-015-0816-y.

[26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[27] R. A. Khan, A. Meyer, S. Bouakaz, Automatic affect analysis: From children to adults, in: Advances in Visual Computing, Springer International Publishing, 2015, pp. 304–313.

[28] R. A. Khan, A. Meyer, H. Konik, S. Bouakaz, Saliency-based framework for facial expression recognition, Frontiers of Computer Science 13 (2019) 183–198.

[29] H.-W. Ng, N. V. Dung, V. Vassilios, W. Stefan, Deep learning for emotion recognition on small datasets using transfer learning, in: International Conference on Multimodal Interaction, ICMI '15, ACM, New York, NY, USA, 2015, pp. 443–449. doi:10.1145/2818346.2830593.
URL http://doi.acm.org/10.1145/2818346.2830593

[30] D. Hamester, P. Barros, S. Wermter, Face expression recognition with a 2-channel convolutional neural network, in: International Joint Conference on Neural Networks, 2015.

[31] L. Chen, M. Zhou, W. Su, M. Wu, J. She, K. Hirota, Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction, Information Sciences 428 (2018) 49–61.

[32] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (6) (2018) 1452–1464. doi:10.1109/TPAMI.2017.2723009.

[33] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (4) (2018) 834–848. doi:10.1109/TPAMI.2017.2699184.

[34] O. Arriaga, M. Valdenegro-Toro, P. Plger, Real-time convolutional neural networks for emotion and gender classification, CoRR arXiv:1710.07557.

[35] I. Hadji, R. P. Wildes, What do we understand about convolutional networks?, CoRR abs/1803.08834.

[36] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 818–833.

[37] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, S. Nasrin, B. C. V. Esesn, A. A. S. Awwal, V. K. Asari, The history began from alexnet: A comprehensive survey on deep learning approaches, CoRR arXiv:1803.01164.

[38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Arxiv, 2014.

[39] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional neural networks, in: European Conference on Computer Vision, 2014.

[40] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization, and detection using convolutional networks, in: International Conference on Learning Representations, 2014.

[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. doi:10.1109/CVPR.2015.7298594.

[42] F. Zhou, B. Wu, Z. Li, Deep meta-learning: Learning to learn in the concept space, CoRR arXiv:1802.03596.

[43] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345–1359.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Computer Vision and Pattern Recognition, 2009.

[45] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, CoRR arXiv:1412.6980.