Old Dominion University

ODU Digital Commons

Electrical & Computer Engineering Faculty Publications

Electrical & Computer Engineering

2019

Transfer Learning Approach to Multiclass Classification of Child Facial Expressions

Megan A. Witherow
Old Dominion University, mwith010@odu.edu

Manar D. Samad

Khan M. Iftekharuddin
Old Dominion University, kiftekha@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs

Part of the Child Psychology Commons, Electrical and Computer Engineering Commons, and the Graphics and Human Computer Interfaces Commons

Original Publication Citation
Megan A. Witherow, Manar D. Samad, and Khan M. Iftekharuddin "Transfer learning approach to multiclass classification of child facial expressions", Proc. SPIE 11139, Applications of Machine Learning, 1113911 (6 September 2019); https://doi.org/10.1117/12.2530397.

This Conference Paper is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Transfer learning approach to multiclass classification of child facial expressions

Megan A. Witherow, Manar D. Samad, Khan M. Iftekharuddin

Megan A. Witherow, Manar D. Samad, Khan M. Iftekharuddin, "Transfer learning approach to multiclass classification of child facial expressions," Proc. SPIE 11139, Applications of Machine Learning, 1113911 (6 September 2019); doi: 10.1117/12.2530397

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

SPIE.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Transfer Learning Approach to Multiclass Classification of Child Facial Expressions

Megan A. Witherow[a], Manar D. Samad[b], Khan M. Iftekharuddin*[a]

[a]Vision Lab, Dept. of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA 23529; [b]Dept. of Computer Science, Tennessee State University, Nashville, TN, USA 37209

ABSTRACT

The classification of facial expression has been extensively studied using adult facial images which are not appropriate ground truths for classifying facial expressions in children. The state-of-the-art deep learning approaches have been successful in the classification of facial expressions in adults. A deep learning model may be better able to learn the subtle but important features underlying child facial expressions and improve upon the performance of traditional machine learning and feature extraction methods. However, unlike adult data, only a limited number of ground truth images exist for training and validating models for child facial expression classification and there is a dearth of literature in child facial expression analysis. Recent advances in transfer learning methods have enabled the use of deep learning architectures, trained on adult facial expression images, to be tuned for classifying child facial expressions with limited training samples. The network will learn generic facial expression patterns from adult expressions which can be fine-tuned to capture representative features of child facial expressions. This work proposes a transfer learning approach for multi-class classification of the seven prototypical expressions including the 'neutral' expression in children using a recently published child facial expression data set. This work holds promise to facilitate the development of technologies that focus on children and monitoring of children throughout their developmental stages to detect early symptoms related to developmental disorders, such as Autism Spectrum Disorder (ASD).

Keywords: Transfer Learning, Deep Learning, Facial Expression Classification, Child Facial Expressions

1. INTRODUCTION

An essential canvas for human communication, the face and facial features provide a medium for conveying signals of identity and emotion. The study of facial expressions appears in a wide range of research and applications including human-computer interactions (HCI), psychology, and behavioral science. Facial expressions may contain psychophysical information as useful markers for certain neurodevelopmental, emotional and behavioral, and psychiatric disorders. The discrimination and production of facial expressions by children may give valuable insights into their emotional and developmental state. The analysis of facial expression for child development understanding is an established research area with numerous significant findings on how both expression discrimination and production in children. Since 1980's, this area of study has seen momentum with widespread use of video and facial coding systems.

The assessment of facial expression production in children require facial expression recognition systems appropriate for their age group. The state-of-the-art facial expression recognition systems[1] are all developed using facial images of adult humans to capture full breadth of facial action patterns which may not be appropriate in recognizing less developed facial expressions in children. There remain few studies on children's facial expression recognition possibly due to a scarcity of ground-truth data sets collected from children. The publication of child data sets, such as the Child Affective Facial Expression (CAFE)[2, 3] database, may help to facilitate new research into facial expression production by children, which in turn, may facilitate automatic tracking and quantitative analysis of facial expressions in child developmental studies.

Compared to adults, children are certainly different and less matured in their facial shape and expression patterns. Unlike adult facial expressions, children's expressions can be incomplete, ambiguous, and difficult to reproduce[4]. With the growing application of human-computer interaction (HCI), it has also become important to develop custom user and age specific systems for facial expression recognition. Although there is a great body of literature focused on applying machine learning and deep learning techniques to the classification of facial expressions produced by adults[1], few works apply these

*kiftekha@odu.edu; https://sites.wp.odu.edu/VisionLab/

Applications of Machine Learning, edited by Michael E. Zelinski, Tarek M. Taha, Jonathan Howe,
Abdul A. S. Awwal, Khan M. Iftekharuddin, Proc. of SPIE Vol. 11139, 1113911
© 2019 SPIE · CCC code: 0277-786X/19/$21 · doi: 10.1117/12.2530397

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

methods to the facial expressions produced by children[5-8]. Automated methods for the classification of facial expressions produced by children are an important component for the development of HCI systems that target child users, especially those designed for treatment, intervention, or training of children. Prior studies have observed the trajectory of improvement in the ability to produce facial movements from infancy to adulthood. It is known that children are not adept in producing all of constituent facial muscle actions for some facial expressions, especially the negative expressions of fear, sadness, and anger[9, 10]. Thus, due to age-related developmental differences in production of facial expressions, classifier models trained on adult data may not generalize well to research problems targeting children.

Omitting contempt, which is difficult for children to produce, there are six basic expressions of emotion that were first identified as having strong evidence of universality by Paul Ekman and colleagues: anger, disgust, fear, happiness, sadness, surprise[11]. These universal emotions have become the fundamental classes that are used in facial expression classification and published in facial expression databases. Ekman and colleagues also published the Facial Action Coding System (FACS)[12], a dictionary of the fine movements of the human facial muscles that can be used to annotate all human facial behavior, but also describe the prototypical definitions of the six basic expressions of emotion.

In the design of a model for classifying facial expressions, it is important that the training data are based on the prototypic expression definitions defined by FACS. Producing a FACS-validated database is time-consuming as it requires manual coding by FACS experts. While databases of adult faces with FACS annotations, such as the Extended Cohn-Kanade (CK+)[13, 14] database, have existed for some time, this is a limitation of current child databases. While it does not include FACS annotations, CAFE comes closer to bridging this limitation by having a photographer trained in the Specific Affect Coding System (SPAFF)[15], an observational coding system heavily influenced by FACS, pose the child models.

In this study, we propose to overcome the issue of limited training samples of child facial expressions by leveraging the useful information from well-established domain of adult facial expressions. We train a deep learning model for adult facial expression classification using the CK+ database. Then, we fine tune this model for classification of child expressions using the CAFE database. Using the same model architecture, we demonstrate the advantage of transfer learning over other training paradigms. We evaluate our models using person-independent[1] 10-fold cross validation.

This paper is organized as follows. Section 2 describes relevant background on deep learning and transfer learning applied to facial expression classification. Section 3 provides details on the data, preprocessing, model, and experiments. Section 4 reports and discusses the results. Section 5 concludes and indicates directions for future work.

2. BACKGROUND REVIEW

Deep learning has demonstrated the state-of-the-art superior performance in a variety of pattern recognition applications, including facial expression classification. Unlike traditional machine learning pipelines, which require feature extraction and feature selection steps prior to training, deep learning models perform feature extraction and feature selection as an integral part of the training procedure.

One class of deep architectures are convolutional neural networks (CNNs), which are appropriate for applications involving image data. CNNs are a deep, feed-forward neural network. CNNs are made up of an input, an output, and at least one convolutional layer[16]. Convolutional layers scan their input with a small, trainable kernel and produce one or more feature maps. This format lends the advantages of parameter sharing and local connectivity, which reduce the total number of parameters and computational cost for training the model. CNNs may also have pooling layers that subsample the layer input to produce a layer output of reduced size, also reducing the total number of parameters and computational cost. One popular type of pooling is maximum pooling, which keeps the maximum value for each local subset of the layer input. Fully connected layers have connections with trainable weights between every neuron in one layer and every neuron in the next layer.

In transfer learning, a deep learning model that has been trained for one task is fine-tuned for another task[16]. Notably deep learning algorithms require large number of training samples to ensure optimal accuracy and generalizability of the recognition system. However, many recognition tasks can be limited by the sample size, similar to our study of child facial expressions. Transfer learning methods have shown success in areas where the target domain with limited samples can leverage training patterns learned from large datasets to extend the application of deep neural networks. In transfer learning, a trained network with large number of image samples is then fine-tuned with the limited available samples of the target domain. One way of fine tuning the model is to freeze the early layers such that the weights in these layers do not receive updates. Using the existing weight values as a starting point, the last few layers are trained on data for the new task.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

There are very few studies have developed models for classifying CAFE[5-8]. Rather than focusing on prototypic expressions, many of these studies classify all images in the database[5, 6]. In Ref. 7, a variety of traditional machine learning models are trained for classification of the full CAFE dataset, open mouth subset and closed mouth subset. The best model reported for the closed mouth subset is linear kernel SVM with 59.375% average 10-fold cross validation accuracy[7].

Transfer learning has been applied in child facial expression classification task by training a CNN using a subset of the Affect from the Internet (Affectnet)[17] database, which contains labeled images of adults or children from the Internet, then further training on the NIMH Child Emotional Faces Picture Set (NIMH-ChEFS)[18] child database[8]. Rather than freezing early layers to preserve the feature information learned from Affectnet, all layers are updated in a continued learning paradigm. The NIMH-ChEF data set is limited in that it has only 534 photographs and only five classes, 'angry', 'fearful', 'happy', 'neutral', and 'sad', which is fewer samples and classes compared to our proposed CAFE data set.

3. METHODOLOGY

3.1 Adult Database

The Extended Cohn-Kanade (CK+)[13, 14] data set consists of 593 FACS-coded sequences collected from 123 adults. Of the 593 total sequences, 327 sequences have validated emotion labeling. These labels include 'angry', 'contempt', 'disgust', 'fear', 'happy', 'sad', and 'surprise'. Each sequence begins from a neutral expression and ends with the peak emotional expression.

Following established practice in literature[1], we consider the last three frames of each expression sequence for inclusion in the expression class, which yields a total of 1254 images for the six basic facial expressions plus neutral: 135 'angry' images, 177 'disgust' images, 75 'fear' images, 207 'happy' images, 84 'sad' images, 249 'surprise' images, and 327 'neutral' images. We disregard the 'contempt' expression as it is not present in the target, child database.

3.2 Child Database

The CAFE[2, 3] data set consists of 1192 color photographs of 154 untrained child models (64 males) imitating the six basic facial expressions of 'angry', 'disgust', 'fearful', 'happy', 'sad', and 'surprised', plus 'neutral'. The data set includes racially and ethnically diverse children between 2 and 8 years of age (mean = 5.3 years; range = 2.7 – 8.7 years). The photographs are the work of a professional photographer trained in the Specific Affect Coding System who coached the child models to produce the expressions in a lab setting. All photos show frontal views of the child models on the same off-white background with overhead lighting. Each model is covered from the neck down with an off-white sheet. The photographs have been cropped to a square image with the child's chin 1/6 of the image height in pixels from the bottom and the child's forehead 1/6 of the image height in pixels from the top. The child's face is aligned in each image so that the points on the eye contour are equidistant from the center line of the image. CAFE contains two categories of posed expression data: mouth open and mouth closed. Expressions 'happy', 'sad', 'angry', 'fearful', and 'neutral' have both mouth open and mouth closed poses. For 'disgust' expression, poses with closed mouth and tongue protrusion are available. All 'surprised' expressions are posed with the mouth open except the image labeled as 11069-surprise_F-AA-04, which is posed with the mouth closed.

To eliminate ambiguity in expression recognition due to open and closed mouth appearances, we exclude all instances with mouth open (with exception of the surprised expression) or tongue protrusion from the CAFE data set. We also exclude the image 11069-surprise_F-AA-04 of the surprised expression that is posed with closed mouth. This yields a total of 709 images (278 males): 121 'angry' images, 96 'disgust' images, 79 'fearful' images, 120 'happy' images, 62 'sad' images, 102 'surprised' images, and 129 'neutral' images.

3.3 Preprocessing

OpenCV (https://opencv.org/), imutils (https://github.com/jrosebr1/imutils), and dlib (http://dlib.net/) libraries are used to preprocess the images from both adult and child sets of data. The dlib face detector is used to first detect the face in each image. The detector is based on linear classification of Histogram of Oriented Gradients features combined with image pyramid and sliding window techniques to locate faces at various scales and locations in the input image. The pretrained 68-coordinate facial landmark detector from the dlib library, which has been trained on the iBUG 300-W[19] data set, is then used to extract landmarks on the face. Using these landmarks, all faces are normalized such that the face is centered and rotated such that the eyes are level horizontally. The faces are scaled such that all faces are 256 by 256 pixels and the images are cropped such that the left eye is 30% of the image width in pixels from the left edge, minimizing background around the face.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3.4 Person-independent Cross-validation

We use 10-fold cross validation to evaluate our models. The folds are generated through person-independent sampling[1]. Each database contains keys associated with individual participants. We sort the database files by these participant keys then consider the number in the sort order modulo ten as the fold placement. For example, the 1st, 11th, 21st images are placed in the first fold, the 2nd, 12th, 22nd in the second fold, and so on.

3.5 CNN Model

We design and train a CNN model for classification of the six basic facial expressions plus neutral in adults and children. Our model architecture consists of three convolutional layers with 16, 32, and 64 feature maps, respectively. We select the ReLU activation function and convolutional kernel size of 3x3 for all layers. At the convolutional layers, we also apply batch normalization and dropout with a fraction of 0.25. After each convolutional layer, 2x2 maximum pooling is applied to reduce spatial dimension and aggregate feature information. Following the convolutional layers are a ReLU-activated fully connected layer with 128 hidden units and a softmax classification layer. Batch normalization and dropout with a fraction of 0.50 are applied at the fully connected layer.

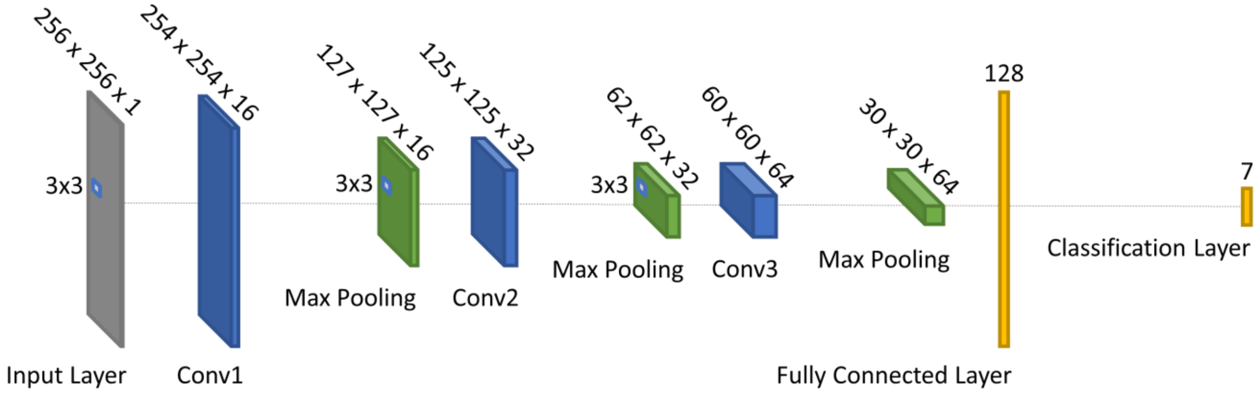

Figure 1. CNN model architecture for 7-class facial expression classification. The architecture consists of an input layer (gray), three convolutional layers (blue) each with a 3x3 kernel size, 2x2 maximum pooling (green), and fully connected layers (yellow). Batch normalization and dropout are applied at each convolutional layer.

To address the class imbalance, we utilize weighted categorical cross entropy[20] as our loss function, defined as follows:

$$J(\theta) = \frac{1}{7}\sum_{i=1}^{7}[-y_i \log(\hat{y}_i)\lambda - (1-y_i)\log(1-\hat{y}_i)\,(1-\lambda), \tag{1}$$

$$\lambda^{(j)} = \left(\frac{N}{M}\right)^{-1}, \tag{2}$$

where N is the number of samples in the positive class and M is the total number of samples.

We train our model using the Adam optimizer. Adam adapts the learning rate for each of the network weights by estimating the first and second moments of the gradient (mean and centered variance) and using these to scale the individual learning rates. We also use early stopping, monitoring the loss and stopping training after the loss stops decreasing.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3.6 Experiments

Using the architecture described in Section 3.5, we train and evaluate models trained for facial expression classification through 10-fold person-independent cross validation under 5 different conditions:

(A) The model is trained on CK+ data and evaluated on CK+ data.

(B) The model is trained on the CAFE data and evaluated on the CAFE data.

(C) The model obtained from condition A is evaluated on the CAFE data.

(D) The pretrained weights obtained from condition A are used to initialize the model, which is then trained and evaluated using the CAFE data.

(E) The pretrained model from condition A is fine-tuned on the CAFE data through transfer learning.

Through 10-fold person-independent cross-validation, we obtain ten models trained on the CK+ database under condition A. We select one of these models for use in conditions C, D, and E based upon the desirable training behavior. Under condition B, we train ten models from scratch on the CAFE database. In condition C, CAFE is considered as the test set for the model trained on the CK+ database. As the choice of initialization for the weights of a deep model has a strong regularizing effect on the performance of the model[16], the pretrained weights from the model trained on CK+ are used to initialize the model for condition D. Then, all layers are trained. For condition E, the pretrained weights from the model trained on CK+ are loaded for all layers, but during training, the weights are updated for the final CNN layer, fully connected layer, and classification layer only. All other layers are frozen.

4. RESULTS AND DISCUSSION

Following experimental condition A, we train ten models on the CK+ database through 10-fold person independent cross validation. The training and validation accuracies for each of the models is reported in Table 1. We select Model 9 for use in conditions B-E for its performance and desirable training behavior, shown in Figure 2.

Table 1. CK+ Training and Validation Accuracies for 10-fold Person Independent Cross Validation Models

Model Number	Number of Epochs	Training Accuracy	Validation Accuracy
0	6	95.323%	90.909%
1	6	97.392%	89.090%
2	6	95.953%	93.636%
3	6	96.223%	89.090%
4	11	98.113%	96.330%
5	12	98.652%	97.247%
6	8	96.765%	89.286%
7	10	97.663%	99.083%
8	7	96.406%	92.661%
9	**18**	**99.551%**	**98.165%**

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

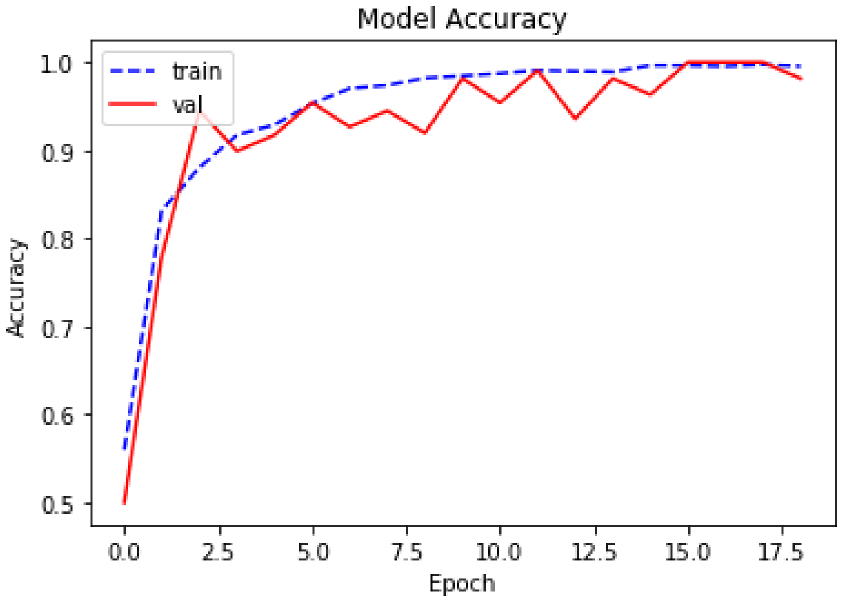

Figure 2. Training and validation accuracy versus epoch for the selected model trained on the CK+ data set, showing desirable training behavior. The model stops training after 19 epochs when the early stopping criterion is met. The training and validation accuracies after 19 epochs are 99.551% and 98.165%, respectively.

Average 10-fold cross validation accuracies and standard deviation for experimental conditions A-E are summarized in Table 2.

Table 2. Average Training and Testing Accuracies from 10-fold Person Independent Cross Validation for Conditions A-E

Condition	Training Accuracy	Testing Accuracy
(A) Train and Test on CK+	97.204% ± 1.248%	93.550% ± 3.729%
(B) Train and Test on CAFE (randomly initialized weights)	94.349% ± 3.478%	63.516% ± 9.317%
(C) Train on CK+, Test on CAFE	99.551% (Model 9)	46.505% ± 5.278%
(D) Train and Test on CAFE (initialize with CK+ Model 9 Weights)	96.431% ± 2.094%	62.442% ± 14.744%
(E) Train on CK+, Fine-Tune and Test on CAFE	**97.369% ± 1.177%**	**76.033% ± 7.058%**

The poor performance of the models trained with condition C demonstrates the poor generalization of a model trained on adult expression data to the classification of child facial expressions. With the exception of condition C, the 10-fold average cross-validation performance for all other conditions outperform the only other classification study using the closed mouth subset of the CAFE database of which we know, which achieves a 59.375% average 10-fold cross validation accuracy using linear kernel SVM [7].

Of the four experimental conditions that are evaluated on CAFE (B-E), the best average 10-fold person independent performance of 76.033% is achieved by condition E, training on CK+ then fine-tuning the last three network layers on CAFE. Condition E outperforms both condition B and condition D, which have similar average 10-fold person independent cross validation accuracies. This suggests that the early layers of the models trained on CK+ for condition E better represent the generic facial expression patterns underlying both child and adult expressions, compared to the features learned from the CAFE data alone (condition B). The poorer comparative performance and large standard deviation of the models

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

trained with experimental condition D suggest that the weight updates to early layers during training on CAFE corrupt the representation of generic facial expression patterns previously learned from CK+.

Figure 3 shows the training and validation accuracies versus epoch for the best of the ten models trained for experimental condition E.

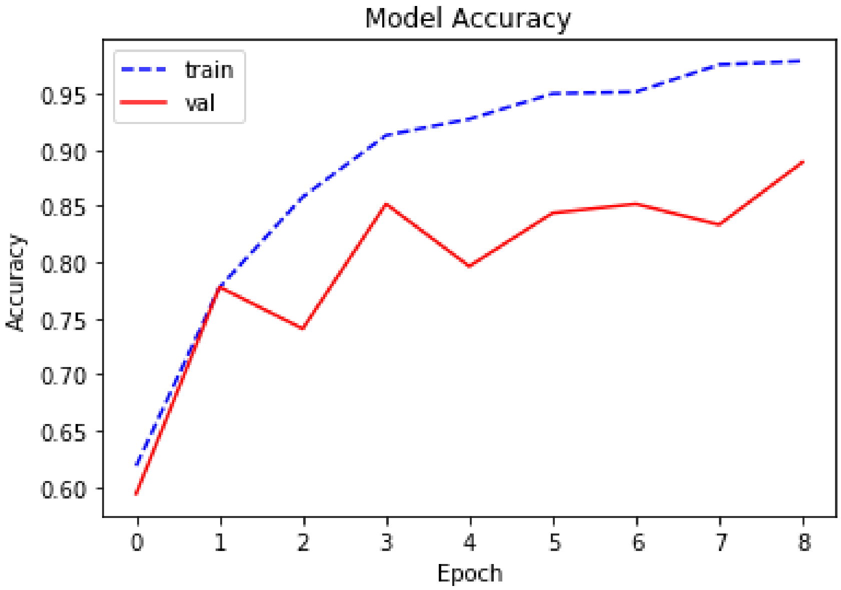

Figure 3. Training and validation accuracy versus epoch for the best model trained on the CAFE database. The model stops training after 8 epochs when the early stopping criterion is met. The training and validation accuracies after 8 epochs are 97.903% and 88.889%, respectively.

5. CONCLUSION AND FUTURE WORK

In this study, we demonstrate the advantage of using transfer learning to learn the underlying generic facial expression patterns from adult facial expression data and report improved performance over existing machine-learning based methods for the mouth closed subset of CAFE. As CK+ is still a relatively small facial expression database, we plan to investigate the effect of using larger facial expression databases for training the base/adult model in the future. We also plan to conduct a more comprehensive study to investigate the effect of different model architectures, parameter selections, and training paradigms on the performance and generalizability of our child facial expression recognition model.

We believe that development of facial expression recognition software and system will be one of the cornerstones of future HCI systems focused on children and child developmental studies. In the future, it is worth investigating how such automated systems can continuously monitor a child's mental and developmental states at home to facilitate means for early screening of emotional and behavioral disorders, and neurodevelopmental disorders like ASD.

REFERENCES

[1] S. Li, and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, (2018).
[2] V. LoBue, and C. Thrasher, [The Child Affective Facial Expression (CAFE) set] Databrary, (2014).
[3] V. LoBue, and C. Thrasher, "The child affective facial expression (CAFE) set," Databrary, 10, B7301K (2014).
[4] P. M. Cole, "Children's spontaneous control of facial expression," Child development, 1309-1321 (1986).
[5] E. Benhamou, D. Wolhandler, A. Zvirin *et al.*, "CHILD FACIAL EXPRESSION DETECTION," (2018).
[6] S. Nagpal, M. Singh, M. Vatsa *et al.*, "Expression Classification in Children Using Mean Supervised Deep Boltzmann Machine." 0-0.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[7] L. Roa Barco, “Analysis of facial expressions in children: Experiments based on the DB Child Affective Facial Expression (CAFE),” (2016).

[8] A. Lopez-Rincon, "Emotion Recognition using Facial Expressions in Children using the NAO Robot." 146-153.

[9] X. Gao, D. Maurer, and M. Nishimura, “Similarities and differences in the perceptual structure of facial expressions of children and adults,” Journal of Experimental Child Psychology, 105(1-2), 98-115 (2010).

[10] O. Houstis, and S. Kiliaridis, “Gender and age differences in facial expressions,” The European Journal of Orthodontics, 31(5), 459-466 (2009).

[11] P. Ekman, and D. Keltner, “Universal facial expressions of emotion,” Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture, 27-46 (1997).

[12] P. Ekman, J. C. Hager, and W. V. Friesen, [Facial action coding system : the manual] Research Nexus, Salt Lake City(2002).

[13] T. Kanade, J. F. Cohn, and T. Yingli, "Comprehensive database for facial expression analysis." 46-53.

[14] P. Lucey, J. F. Cohn, T. Kanade *et al.*, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression." 94-101.

[15] J. A. Coan, and J. M. Gottman, “The specific affect coding system (SPAFF),” Handbook of emotion elicitation and assessment, 267-285 (2007).

[16] I. Goodfellow, Y. Bengio, and A. Courville, [Deep learning] MIT press, (2016).

[17] 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, 10(1), 18-31 (2017).

[18] H. L. Egger, D. S. Pine, E. Nelson *et al.*, “The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli,” Int J Methods Psychiatr Res, 20(3), 145-56 (2011).

[19] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou *et al.*, "300 faces in-the-wild challenge: The first facial landmark localization challenge." 397-403.

[20] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro *et al.*, “Learning from Imbalanced Data Sets with Weighted Cross-Entropy Function,” Neural Processing Letters, (2019).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Nov 2019
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use