An Interactive Facial-Expression Training Platform for Individuals with Autism Spectrum Disorder

Christina Tsangouri, Wei Li[+], Zhigang Zhu**
* Dept. of Comp. Sci.. and [+]Dept of Electrical Eng..
City College of New York, New York, USA
ctsango000@citymail.cuny.edu, wli3@ccny.cuny.edu,
zhu@cs.ccny.cuny.edu

Farnaz Abtahi, Tony Ro[+]*
*Dept. of Comp. Sci. & [+]Dept. of Psychology
CUNY Graduate Center, New York, NY
fabtahi@gradcenter.cuny.edu , tro@gc.cuny.edu

***Abstract*—Emotions are an incredibly important aspect of human communication. Individuals with Autism Spectrum Disorder (ASD) suffer from significant challenges in perceiving, and understanding others' emotions, and responding emotionally in an appropriate manner. In order to help them to improve their facial expression recognition and response skills, EmoTrain was developed as an interactive platform that utilizes deep learning to track a user's facial expressions from a mobile device's camera and assists a user in learning to understand facial expressions and how to perform them. The paper details the design of the EmoTrain platform, the deep learning model and game engine for the platform, and evaluates EmoTrain's effectiveness.**

Keywords—Emotion recognition, assistive technology, autism spectrum disorder, deep learning

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a group of developmental disabilities that are characterized by social-communication impairments that cause significant deficits in the areas of social interaction, communication and language skills, and repetitive behaviors and interests. In particular, these deficits cause difficulties in the perception of faces and the expressions of faces, understanding emotional states, and the perception of gaze direction. Studies [1] of the face processing skills of people with ASD have shown that these impairments are widespread and present from an early age, and affect both the perception and memory of faces. Studies [2-5] have demonstrated that computer-based instruction is more effective in individuals with ASD versus traditional instruction. Due to recent advances in computer vision and deep learning, new avenues are now being explored in computerized assistive and intervention programs for people with disabilities. In this paper we propose a novel face expression-training platform, EmoTrain, which integrates our work in real-time facial expression recognition and assistive technology for people with ASD. As a summary, we have made the following contributions:

1) We implement real-time facial expression recognition on a mobile device, which includes our methods for creating an emotion recognition model, and our framework for achieving real-time performance from the built-in camera of standard mobile technology.

2) We design a novel EmoTrain's interface and game engine, and the tasks that the users are expected to perform for the training platform.

3) We demonstrate the effectiveness of the EmoTrain platform by evaluating the facial expression recognition and facial behavior response skills of the subjects with ASD who received training with EmoTrain.

II. RELATED WORK

There exists a lot of interesting work relating to applications of human facial expression recognition. The MIT Media Lab has created 'Affectiva', an application for analyzing your smile, which began as an effort to help people on the autism spectrum who have difficulty reading emotion, and is now being commercialized to help businesses understand their customers[2]. In [3] Cockburn, et al propose including real-time expression recognition in an existing dynamic game for children with ASD as a means for improving the efficacy of the gamefied intervention platform. In their work they utilize the Computer Expression Recognition Toolbox (CERT), which analyzes facial expressions in real-time and can classify them into 7 basic emotions and 30 facial action units from the Facial Action Coding System. In [4] Tanaka, et al propose *Let's Face It!*, a computer-based intervention program, that is comprised of 7 interactive computer games. These games are created to enhance the face recognition skills, holistic face processing skills, and attention to the eye region in children with autism. The results from their study were promising, indicating that a short-term intervention program produces significant improvements in the face processing skills of children with autism. In [5] Golan, et al evaluate an animated series designed to enhance emotion comprehension in children with autism spectrum conditions, *The Transporters.* The children who participated in this study exhibited significant improvement in emotion recognition in all the task levels 'The Transporters' intervention program trained and tested them on.

Studies have demonstrated that there are several advantages to using computer software for instruction with ASD individuals. In [12] the impact of computer instruction vs traditional behavioral techniques for vocabulary acquisition for children with ASD was examined. The study demonstrated that the children learned more, paid more attention, and were more motivated in the educational software program that was designed based on behavioral learning principles than in an education program with human instructors. In [13] a computerized intervention program for teaching adults with ASD to recognize emotions in faces and voices, *Mind Reading,* was evaluated. The adults who used *Mind Reading* for even a relatively short period of time showed significant improvement in their emotion recognition skills. The use of educational software for individuals with ASD is more effective for a myriad of reasons. Individuals with ASD prefer computerized educational environments since they are predictable and consistent, and devoid of the social factor which causes them stress. With computer-based instruction these individuals have the ability to work at a pace that suits their learning capacities, and can repeat lessons until they are mastered [10].

III. Real-Time Emotion Computing

A. Model Construction

The emotion recognition model is based on a machine-learning approach, which uses convolutional neural networks (CNNs) – one of the most effective deep learning structures. A CNN consists of two major components: a forward-pass estimation and a back-propagation mechanism. While the CNN model is being trained, the forward-pass mechanism obtains a classification prediction based on the image it took as input. This prediction is compared to the image's label and a loss of the estimation is measured. The back-propagation mechanism updates the CNN's parameters using the loss [6].

Previous work has demonstrated that deep learning based models perform better in facial emotion recognition tasks than traditional approaches. But in order to develop a robust emotion recognition model a large and balanced dataset is necessary for training. Unbalanced or small datasets lead to some emotions being hard to detect. Datasets therefore play a crucial role in classification models. Since most of the existing datasets were either not large enough or imbalanced we utilized a game-based data collection approach to collect a new dataset that was sufficiently large and balanced [7]. This approach uses a game-based facial expression interface, similar to EmoTrain, which prompts users to match the images on the screen with their own faces.

The game-based data collection interface was created by first fine-tuning a CNN model that was pre-trained on an existing dataset of face expression images that we collected from the web [8]. . The user's face images from the device camera running the game interface are then analyzed by the CNN model, labeled, and added to our new dataset. Using this approach, we collected approximately a game-based dataset of 15,000 images, which were used to train a fine-turned CNN model, which forms the basis for the EmoTrain platform.

B. Framework

In order to achieve real-time emotion computing performance we designed and implemented a framework consisting of the following elements (Fig.1):

(1). An *application* is implemented on a mobile device that accesses the built-in camera, and performs face detection on the camera feed.

(2). A *web service* on the main server handles all requests from the application, and receives the face images detected.

(3). An *emotion-computing server* runs the model detailed in the previous section. The model computes the probability of 7 basic human emotions of each face image received. These emotions are: happy, neutral, angry, surprise, fear, sad, and disgust. The emotion probabilities are then sent to the mobile application.

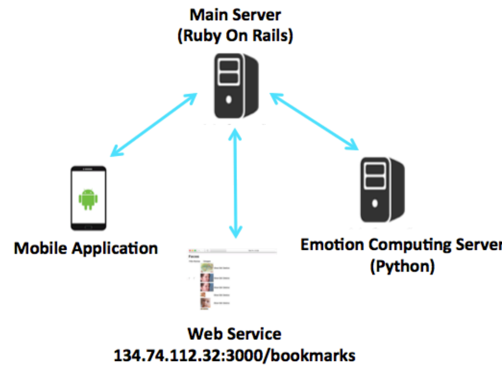

Fig 1. Real-time emotion computing framework

IV. EmoTrain

A. ASD Facial Expression Deficits

A typical individual that does not have ASD processes faces in a holistic manner and based on configural information. Studies [9] have demonstrated that individuals with ASD exhibit deficits in their face processing abilities, and in particular in discrimination and recognition of faces. Although some individuals over time develop some strategies for recognizing basic emotions, there remains a general impairment in identifying more complex facial expressions [10]. Due to these deficits individuals with ASD have difficulties in understanding facial expressions and reciprocating emotionally [3]. In [11] the ability for individuals with ASD for automatic mimicry of facial expressions is explored. Automatic mimicry refers to the automatic response that is elicited when observing another person's emotional behavior and is an important factor in social functioning. When a typically developing individual sees an emotional expression he/she would spontaneously respond with a similar expression. For example, that individual would usually smile to a smile or frown to a frown. The data from this study supported the view that individuals with ASD exhibited a deficit in automatic mimicry.

B. Face Expression Training Platform

EmoTrain is a platform that is designed to target deficits in recognizing and labeling facial expressions, and reciprocating emotionally and to help teach these face processing skills to individuals with ASD. EmoTrain is designed to teach individuals with ASD to recognize 7 basic human emotions and to also perform these 7 facial expressions. When playing EmoTrain, users are asked to attempt to match the images of face expressions they are presented with on the screen. By incorporating our work in real-time emotion recognition from a mobile camera, the platform is able to track the user's expressions in real-time and judge if the user is performing those expressions correctly. In [3] it is demonstrated that training in facial expression mirroring is essential to developing recognition skills. This sort of training in mimicking expressions allows us to improve the skills of individuals with ASD in facial expression recognition, and in automatic emotional reciprocity.

C. Interface

The game logic for EmoTrain is as follows: a user is presented with images of various faces representing the seven different emotions and has to perform that same expression in order to score. The EmoTrain interface is shown in Fig. 2. On the right half of the screen is the game scene, on the top left corner is the live video stream from the front-facing camera of the device, and on the bottom left corner is displayed a visualization of the emotions detected in the user's face in real-time.

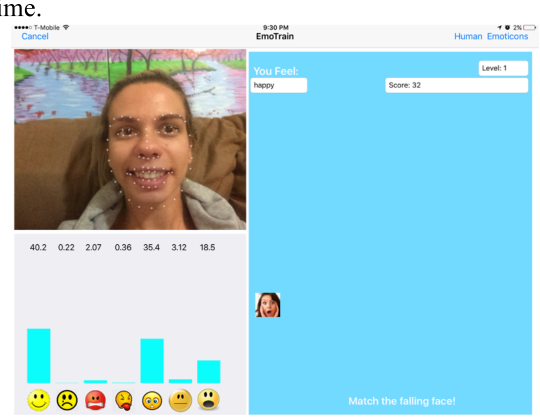

Fig 2. Screenshot of EmoTrain interface

Each face image target enters the screen from the top and the user has to match that emotion before it reaches the bottom in order to score. The face image disappears if the user performs a facial expression that matches the target image, which is judged by the CNN-based emotion detector. The EmoTrain platform accesses the front facing camera on the user's device and sends the camera frames to our server running the CNN model – which analyzes each image and generates a probability vector for the seven emotions and sends this back to the app. The app then compares the probability vector with the image label for the face image target. In order to combat inaccurate emotion probability results and make it easier for a user to score each emotion category is assigned a predefined threshold. This threshold varies for each emotion since some emotions are more difficult to mimic voluntarily such as anger and fear. If the threshold is reached for the emotion category that is the same as the image target label then that is considered a match and the user scores.

In order to make the platform more engaging and interactive for the user we have added certain visualizations and multimedia. On the real-time video of the user's face on the upper left corner OpenCV and Dlib's face landmark detector are utilized to track the user's face on the screen and 68 landmarks on the face. On the lower left corner of the screen a bar graph live visualization is displayed that shows the probability of each emotion the CNN model predicts for the user's face.

V. Platform Evaluation

In order to gauge the effectiveness of the EmoTrain platform, 9 subjects participated in an evaluation study. The participants were all youth (ages 18-25) who were diagnosed with ASD. The participants received training with the EmoTrain platform 2 times a week for 2 weeks, for 20 minutes each session. In total each participant received 80 minutes of training. The participants were administered a survey before they began using the platform and after they had completed all their sessions. The surveys consisted of two sections: the first section was a general survey to gauge interest and the second section was an assessment to determine the subject's facial expression recognition abilities. The general part of the survey administered before a subject started using EmoTrain was about the subject's online gaming preferences, and the types of mobile devices they use along with their proficiency in using them. The assessment part of the pre-use survey asked a user to identify the facial expressions in 14 images, where there were 2 images for each of the 7 basic emotion categories. The assessment questions were simple: "How is he/she feeling?" The assessment was multiple choice with 7 choices for each emotion and an 8th choice of 'I'm not sure'. The general part of the survey administered after a subject completed all his/her sessions with EmoTrain was about the subject's experience with EmoTrain such as difficulties with the platform and recommendations for improvement. The assessment part of the post-use survey asked a user to perform the same expression identification task as the pre-assessment but with a different set of 14 images.

The effectiveness of the platform was analyzed by determining if the facial expression recognition skills and the facial reciprocity skills of the participants improved. The improvement of the facial recognition skills of the subjects was determined by looking at the subjects' responses to the assessments administered before and after the study. The improvement of the facial reciprocity skills of the subjects was determined by looking at the subject's score for each session. If the score increased significantly for each session with EmoTrain then we concluded that the facial reciprocity skills for that participant had improved. In Fig. 3 we compare each subject's session 1 total score with session 4 total score. From this we see that all the participants' total score increased each session of game play with EmoTrain. This demonstrates that even with a relatively short training period of 80 minutes the facial expression recognition and expression reciprocity skills of the participants in this study significantly improved. In Figure 4 we see each subject's pre-assesment in comparison to each subject's post-assessment survey. Every subject was scored out of 14, which is the total images asked to identify. From this figure we see that most of the participants' scores in the post-assessment show a significant improvement from their scores from the pre-assessment in facial expression recognition.

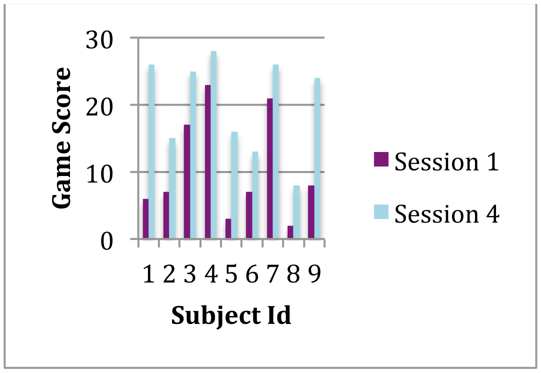

Fig 3. Comparison of score results

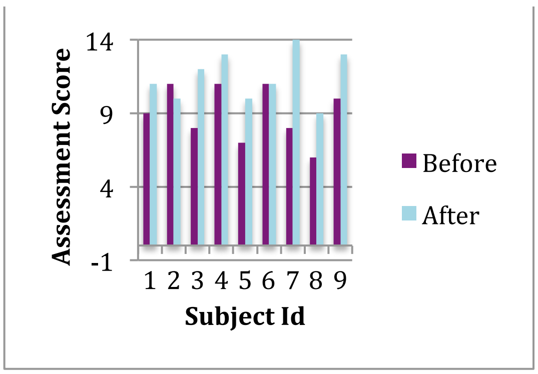

Fig 4. Comparison of pre-assessment and post-assesment

The general survey that was administered to the participants upon completion with the sessions demonstrated that 100% claimed to really enjoy the game, and 90% would recommend EmoTrain to a friend. Most of the participants found the interface fairly easy to understand and use, with only 20% stating that it was difficult to match the face expression targets.

CONCLUSION

In this paper we present EmoTrain, an interactive platform based on deep learning that is designed to help improve face expression recognition and reciprocity skills in individuals with Autism Spectrum Disorder. EmoTrain is a gamified training platform that prompts user's to match the face expression images shown with their own faces in real-time. By utilizing our work in real-time emotion recognition from a mobile camera, EmoTrain's interface tracks the user's face from the device's front facing camera and can determine if the user makes the correct face expression. Using this method of mimicry training, the objective is to help people with ASD overcome their impairment in face processing skills such as perception of faces, understanding emotional states, and emotional reciprocity. The effectiveness of this platform was evaluated by administering training with EmoTrain to a group of participants diagnosed with ASD. By analyzing the data collected from the participants' performance while playing on EmoTrain over time and comparing their face processing skills before and after playing, we demonstrated that the platform is effective.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation under award # EFRI – 1137172, the 2015 NSF EFRI-REM pilot program at the City College of New York, the NSF GARDE Program (award #1160046)), the City College Zahn Innovation Center, VentureWell, and Goodwill Day Services – Bridges to Success Program.

REFERENCES

[1] M. Behrmann, et. al, "Configural processing in autism and its relationship to face processing," in Neuropsychologia, vol. 44, pp. 110-129, 2006.

[2] D. McDuff, et. al, "Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In-the-Wild"," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on , vol., no., pp.881-888, 23-28 June 2013.

[3] J. Cockburn, et. al, "SmileMaze: A Tutoring System in Real-Time Facial Expression Perception and Production in Children with Autism Spectrum Disorder," in Proceedings of Intl. Conference on Automatic Face and Gesture Recognition, 2008.

[4] J. Tanaka, et. al, "Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! Program," in Journal of Child Psychology and Psychiatry, vol. 51, pp. 944-952, 2010.

[5] O. Golan, et. al, "Enhancing Emotion Recognition in Children with Autism Spectrum Conditions: An Intervention Using Animated Vehicles with Real Emotional Faces," in Journal of Autism and Developmental Disorders, vol. 40, pp. 269-279, 2010.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[7] W. Li, F. Abtahi, C. Tsangouri, and Z. Zhu. "Towards An "In-The-Wild" Emotion Dataset Using a Game-Based Framework". CVPRW 2016, Las Vegas.

[8] W. Li, M. Li, Z. Su, and Z. Zhu. "A deep-learning approach to facial expression recognition with candid images." In Machine Vision Applications (MVA), 2015 14th IAPR Internation Conference on, pp. 279-282. IEEE, 2015.

[9] S. Faja, E. Aylward, R. Bernier, and G. Dawson, "Becoming a Face Expert: A Computerized Face-Training Program for High-Functioning Individuals with Autism Spectrum Disoders," in Developmental Neuropsychology, vol. 33, pp. 1-24, 2008.

[10] O. Golan and S. Baron-Cohen, "Systemizing Empathy: Teaching adults with Asperger Syndrome or high-functioning autism to recognize complex emotions using interactive multimedia," in Development and Psychopathology, vol. 8, pp. 591-617, 2006.

[11] D. McIntosh, A. Reihmann-Decker, P. Wienkelman, and J. Wilbarger, "When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism," in Developmental Science, vol. 9, pp. 295-302, 2006.

[12] M. Moore and S. Calvert, "Brief Report: Vocabulary Acquisition for Children with Autism: Teacher or Computer Instruction," in Journal of Autism and Developmental Disorders, vol. 30, pp. 359-362, 2000.

[13] M. Heimann, K. Nelson, T. Tjus, and C. Gillberg, "Increasing Reading and Communication Skills in Children with Autism Through an Interactive Multimedia Program," in Journal of Autism and Developmental Disoders, vol. 25, no. 5, 1995.