RBE 595 Deep Learning for Robot Perception Final Project
Emotion Recognition in Children with Autism Spectrum Disorder

Prepared by: Katie Gandomi and Kritika Iyer

***Abstract*—The ability to assess facial expressions can play a pivotal role in human-robot interactions. PABI, a humanoid social robot, can use this ability to detect affect responses in children with Autism Spectrum Disorder (ASD) during Applied Behavioral Analysis (ABA) therapies. In this paper, two convolutional neural networks are created for the purpose of detecting emotions in children with ASD. The Kaggle FER'13 dataset is used for training and the accuracies of these models are within 15%-5% of the best solution for this data. Validations were performed on videos previously taken by PABI of ABA therapies. PABI can use facial expression recognition to help children with their social skills and customize therapy sessions. This paper solely focuses on the emotions of various children and the accurate classification of these detected emotions.**

***Keywords*—Emotion tracking, Deep learning, ASD**

I. Introduction

A. Emotion Tracking and Deep Learning

Facial expressions are used to convey emotion in everyday conversation. A human can deduce whether another person is happy, sad, or angry simply through visual inputs, and this ability plays a pivotal role in typical social interactions. Some of these emotions in children is shown in figure 1. Human-robot exchanges can be enhanced with this information as it can serve as a source of feedback on the quality of the interaction itself.

Fig. 1: Different emotion of children

Emotion recognition is the task of classifying these social subtleties typically through facial expressions. A number of different methods have been proposed to accomplish this task, and deep learning approaches such as [7], [6] have been particularly successful. This is accredited to features in the different expressions being learned rather than explicitly dictated.

In this paper, facial expressions are identified on autistic children through a socially-assistive robotic platform. With this added information, therapy sessions can be improved and metrics such as the childs emotional response to treatment can be measured. The robot can then dynamically change or customize a therapy session based on this data. In turn, this can improve the overall quality of the treatment and drive child-robot social interactions.

B. Autism Spectrum Disorder

Autism Spectrum Disorder(ASD) is a neurological disorder which can affect motor and social skills. This generally presents itself in children with various symptoms and varied intensities of these symptoms[8]. With early intervention, many children with ASD can be incorporated and function properly in society[9].

PABI(Penguin for Autism Behavioral Interventions) is a socially assistive robot which helps enhance the social aspects of ASD affected children[1]. When people look at other people, there is a lot of information the facial expressions can give. The amount if information is hard for ASD affected children to assimilate and can become overwhelmed. This is why a lot of ASD affected children do not make eye contact. For learning however, eye contact is crucial. For this, robots can be used as there is not a lot of information provided other than the basic expressions. Over the years, robotic intervention has proved to be more effective for ASD affected children[10].

C. Social Implications

The social impairments of children with ASD ranges from speech issues to social interactions on a daily basis[11]. They wish to have more social interactions but are not capable of doing so. They end up feeling lonely and set aside in society[12][13]. Due to the inability to assess situations and peoples feelings this affects learning especially in the crucial first 10 years of their life. Through early intervention during this time, most of these can be resolved and the child can lead a relatively normal life.

II. Literature Review

A. Prior Art

The concept of facial expression recognition (FER) for human emotions is not new. In the paper [2], six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) are considered the most universal to all human beings. A number of approaches have been applied throughout the years to solve the FER problem. Existing detectors have used Histogram of Oriented Gradients (HoG), Local Binary Patterns Histograms (LBPH) and Gabor methods where hyper-parameters are carefully tuned to give the best recognition accuracies. Traditional machine learning approaches such as support vector machines, dictionary learning, and binary classifiers have also been employed successfully to classify facial expressions [5]. However, these methods lack the ability to generalize well beyond their datasets and categorize emotions without significant exaggeration. Convolutional neural networks (CNN) have shown a remarkable ability to assess emotions in the wild. Often overfitting can be an issue if the dataset is not large enough, but this can largely be overcome with data augmentation and dropout regularization[7].

Typically, there are three main steps taken in FER solutions: face registration, feature extraction, and classification. The first step generally requires localizing the face(s) in the image and isolating them from the rest of the picture. Feature extraction can be based on facial geometries such as landmarks or on pixel intensities. Classification is then performed to determine the emotion that is most likely represented in the facial sub-image.

Fig. 2: PABI participating in ABA therapy with Discrete Trial Training via Tablet with a Therapist and Child.

B. Penguin for Autism Behavioral Intervention

PABI is developed in the WPI Automation and Interventional Medicine (AIM) Robotics Lab. This accessible social robot is designed for effective intervention in children with autism, and its small, cartoonish, non-anthropomorphic form makes it ideal for this task [1]. PABI participates in applied behavioral analysis (ABA) via discrete trial training (DTT) on a custom tablet application(Fig 2). Briefly, the therapist prompts the child to select a flashcard from a given set and records the childs response. PABI participates in these sessions by making meaningful expressions and utterances that prompt the child, provide reinforcement, and facilitate response data collection.

As a robotic system, PABI is comprised of a 3-DOF, 2-DOF , and 1-DOF stage for its eyes, neck, and beak respectively. PABIs eyes consist of two ultra-wide view fisheye cameras. The robot also has two highly compliant wings actuated by cables. Embedded within PABI is a computer with an Intel Pentium i7 quad-core processor with a solid state hard drive. The current software running on the robot features face detection and tracking. This allows the robot to move in order to maintain the child in the center of PABIs field of view. Additionally, the robot tracks the users head pose and eye gaze which is current used to rudimentarily assess the childs attention to therapy sessions.

C. Social Robots and Deep Learning

Some socially assistive robots use imitation to aid the children. Robots such as Robota[14] and Zeno[15] , imitate the therapist and make sure the child follows. Romibo[16] shows emotion to help the child slowly improve interactions and get used to other peoples emotions. Keepon[17] is a small robot that uses a non-verbal method to convey emotions. It vibrates on a countertop to convey emotions to the child. Kimset[18] is an anthropomorphic robot with a head which can recognize and show emotions. Kaspar[19] a humanoid robot acts as a companion to the child to help with daily interactions. These are some of the robots used to aid children affected with ASD improve their social skills, PABI is another such robot which will be used to carry out this project.

III. Problem Statement

The problem to be addressed by this paper, is the accurate classification of emotions in children with autism spectrum disorder on a socially-assistive robotic platform. The developed software will be built to identify facial expression from a video feed. This requires the following 2 key tasks to be performed: isolation of face sub-regions in the image and classification of depicted emotion, where the final task will be performed via a convolutional neural network.

IV. Methodology

A. Preparing the Dataset

A suitable database in both size and diversity is important to prevent overfitting during the training phase. The Kaggle Facial Expression Recognition Challenge (KFERC) dataset[3] from 2013 was selected. It consists of approximately 30,000 labeled 48x48 images of up to 7 different emotions. The data was initially plotted for the 7 different emotions and the following distribution was obtained(Fig 3). As can be seen from the graph, the data was not spread evenly across the 7 cases. Another issue was "bad" data within the dataset, which represented either data that did not show a face or that incorrectly classified an emotion(Fig 5).

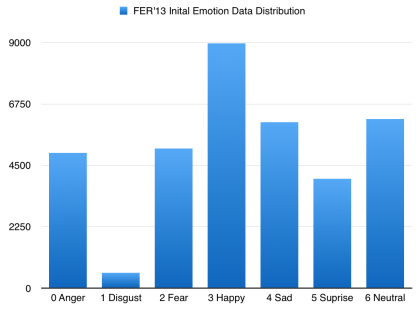

Fig. 3: Initial distribution of the 7 emotions in Kaggle Facial Emotion Recognition Dataset

Fig. 4: Images above where all incorrectly characterized as 'anger'. Top row are images that don't characterize a face properly and bottom row are images that don't characterize the emotion 'anger'.

To improve this, the team sifted through all images in the dataset and removed the incorrectly characterized frames. Since the emotion disgust had so little data when compared with the other 6, it was combined with anger since they both exhibit a similar negative sentiment. To prevent data bias, all categories were augmented until each had 6000 images(Fig **??**).

Fig. 5: Images show typical good characterizations from the FER'13 dataset.

B. Model Experimentation and Architecture

The dataset now had 6000 images for each label. These pictures where then compressed into a pickle file with the corresponding labels. This pickle file was then loaded, shuffled and split into train and test data. Once this was done, a model was devised with many layers of convolution, max-pooling, activation, flattening and dropout layers. The data was trained on this model and the initial accuracy was 47.5%. After a lot of tweaking of parameters, the accuracy improved to 57.22%. Looking at the confusion matrix, fear was not being detected properly and a further look into the dataset of fear showed how varied the emotion of fear can be expressed so the team decided to take fear out and train the model on 5 labels, namely: happy,sad,neutral,anger/disgust and surprise. The data was re-pickled with their corresponding labels. The accuracy then increased to 65% and reached a maximum of 68.50%. The model has a slightly tough time differentiating between sad and neutral but detects a slightly exaggerated sad emotion. This is hard to train as a lot of times humans find it difficult to differentiate between sad and neutral so it can be equally hard for a network to differentiate with small amounts of data. Some experimentation with removing surprise was also done but did not give fruitful results. The final model that was arrived is shown in Fig 6.

V. Validations & Experimentation

A. Validation

The best accuracy for the six emotions (anger/disgust, fear, happy, sad, surprised, and neutral) was 57.22%. The best accuracy using five emotions (same as above, but not including fear) was 68.50%.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 16, 46, 46)	160
conv2d_2 (Conv2D)	(None, 16, 44, 44)	2320
batch_normalization_1 (Batch	(None, 16, 44, 44)	176
activation_1 (Activation)	(None, 16, 44, 44)	0
max_pooling2d_1 (MaxPooling2	(None, 8, 22, 44)	0
conv2d_3 (Conv2D)	(None, 32, 20, 42)	2336
conv2d_4 (Conv2D)	(None, 32, 18, 40)	9248
batch_normalization_2 (Batch	(None, 32, 18, 40)	160
activation_2 (Activation)	(None, 32, 18, 40)	0
max_pooling2d_2 (MaxPooling2	(None, 16, 9, 40)	0
conv2d_5 (Conv2D)	(None, 32, 7, 38)	4640
batch_normalization_3 (Batch	(None, 32, 7, 38)	152
activation_3 (Activation)	(None, 32, 7, 38)	0
max_pooling2d_3 (MaxPooling2	(None, 16, 3, 38)	0
conv2d_6 (Conv2D)	(None, 32, 1, 36)	4640
batch_normalization_4 (Batch	(None, 32, 1, 36)	144
activation_4 (Activation)	(None, 32, 1, 36)	0
dropout_1 (Dropout)	(None, 32, 1, 36)	0
flatten_1 (Flatten)	(None, 1152)	0
dense_1 (Dense)	(None, 128)	147584
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 5)	165

Total params: 182,061
Trainable params: 181,745
Non-trainable params: 316

Fig. 6: Best model for 5 labels accuracy 68.50%

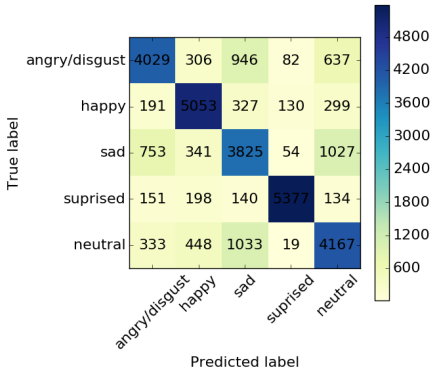

Fig. 7: This figure shows the confusion matrix created for the five emotion model

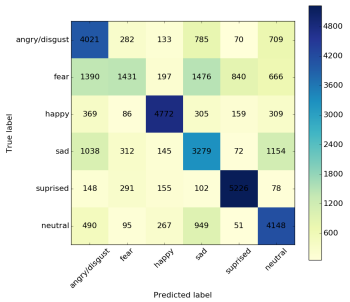

Fig. 8: This figure shows confusion matrix created for the six emotion model

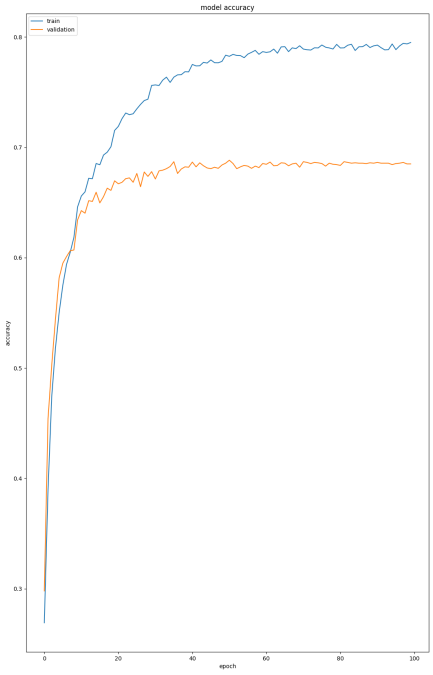

Fig. 9: Accuracy plot for 68.50%

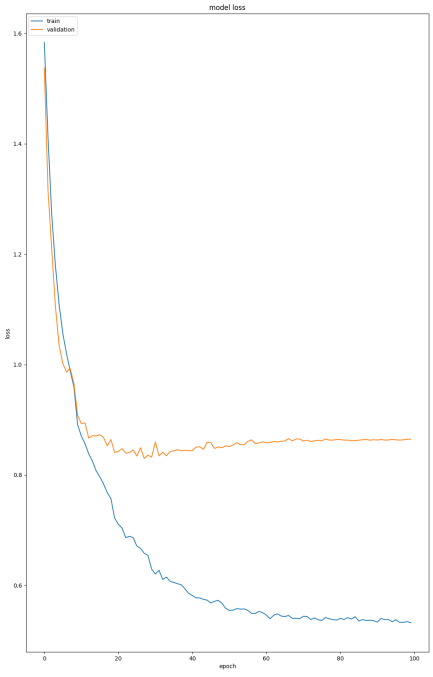

Fig. 10: Loss plot for 68.50%

It is worth noting that for the FER'13 Kaggle Dataset the best accuracy recorded accuracy has been 71.2%. The graphs show the accuracy and loss plots for each model(Fig 9,10).

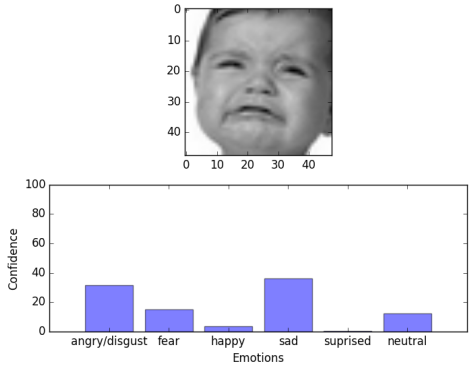

Fig. 11: This image shows the softmax output for the six emotion model. The given image characterizes 'sad' and the graph shows the confidence for the different emotions.

The softmax output was observed for various images and their corresponding emotion confidences were plotted. Human emotions are complex and are rarely exhibited independently. For example, one might feel happy and surprised, or sad and angry. The team noticed that when an inaccurate prediction was made by the model, the true value was usually the second highest rated emotion.

The team also visualized different layers in the model to get a better understanding of how decisions were being made. The figures below illustrate layers two and three of the six emotion network.

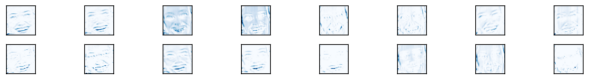

Fig. 12: This figure shows the visualization of layer three in the six emotion model

B. Experimentation with PABI

Before starting experimentation with PABI, the team created a JSON and h5 file for exporting and loading the models independently. A real-time video emotion detector was first created to determine emotions via a web-cam as shown in the figure below. The performance of the model could be successfully, applied to a 30fps live video feed without lag.

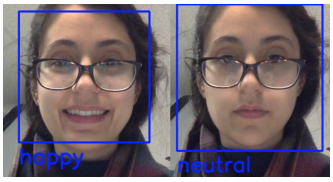

Fig. 13: This figure shows a sample from the created real-time video emotion detector using the created models.

Next, the team considered, videos taken previously by PABI of real applied behavioral analysis therapy sessions of children with autism spectrum disorder. The models each performed within their expected accuracies with the most common misclassification coming from differentiating the subtle differences between sad and neutral.

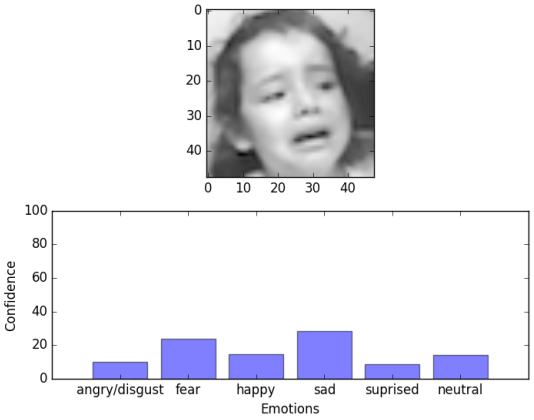

Fig. 14: This figure shows the softmax output for a sample image taken from the PABI videos. It shows correct classification of sad.

Fig. 15: This figure shows the emotion detector CNN working on previously recorded ABA therapy sessions by PABI.

VI. Conclusion

In conclusion, two convolutional neural networks with created in this project, one trained across six emotions and the other across five. The best accuracies for each

model were 57.22% and 68.50% respectively. This results in a 14.35% and a 2.72% difference from the best recorded model accuracy for this data set of 71.2%. The model was assessed for real-time performance on a computer web camera and validated on applied behavioral analysis therapies of children with autism spectrum disorder recorded by PABI. The success of the model in these two areas, shows the effectiveness of this approach, however, in the future more data would likely need to be obtained to give the models greater confidence.

VII. Future Work

Emotion detection is a very useful piece of technology for understanding the reactions of humans towards any given stimulus provided by the robot. In the future, after this project is completed, it is possible to use reinforcement learning to customize the learning experience to each child. PABI has a variety of features that can provide positive reinforcement for a child when a task is properly completed. Each child may respond to a certain cue better than another one. By detecting if the child responds better or worse to certain cues, the therapy session can be customized to each student.

References

[1] Dickstein-Fischer L.A, at al., An Affordable Compact Humanoid Robot for Autism Spectrum Disorder Interventions in Children, in Proceedings of IEEE Engineering in Medicine and Biology Society, 2011, pp. 5319-5322.

[2] Ekman Paul and Friesen Wallace V.. Constants Across Cultures in the Face and Emotions. Journal of Person- ality and Social Pyschology,

[3] Goodfellow Ian, Dumitru Erhan, Pierre-Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in rep- resentation learning: A report on three machine learning contests, 2013.

[4] Mavadati, S.M. , Mahoor M. H., K. Bartlett, P. Trinh, and J. F. Cohn. Disfa: A spontaneous facial action intensity database. Affective Computing, IEEE Transactions on, 4(2):151160,2013. 1, 2, 5

[5] Mollahosseini, Ali, David Chan, and Mohammad H. Mahoor. "Going deeper in facial expression recognition using deep neural networks." Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016.

[6] Romero, P., Cid, F., and Nnez, P. (2013, September). A novel real time facial expression recognition system based on candide-3 reconstruction model. In Proceedings of the XIV Workshop on Physical Agents (WAF 2013), Madrid, Spain (pp. 18-19).

[7] Song, Inchul, Hyun-Jun Kim, and Paul Barom Jeon. "Deep learning for real-time robust facial expression recognition on a smartphone." Consumer Electronics (ICCE), 2014 IEEE International Conference on. IEEE, 2014.

[8] Jessica Wright (2015). Cognition and behaviour: Motor skills affect speech in autism. Spectrum news. http://sfari.org/news-and-opi nion/in- brief/2013/cognition-and-behavior-motor-skills-affect-speech-in-autism

[9] Autism is not a disease. Times of India. http://timesofindia.indiatimes.com/city/ahmedabad/Autism-is-not-a-disease/articleshow/19333810.cms

[10] Cabibihan, John-John, et al. "Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism." International journal of social robotics 5.4 (2013): 593-618.

[11] Feil-Seifer, David, and Maja J. Matari. "Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders." Experimental robotics. Springer, Berlin, Heidelberg, 2009.

[12] White, Susan Williams, Kathleen Keonig, and Lawrence Scahill. "Social skills development in children with autism spectrum disorders: A review of the intervention research." Journal of autism and developmental disorders 37.10 (2007): 1858-1868.

[13] Mller, Eve, Adriana Schuler, and Gregory B. Yates. "Social challenges and supports from the perspective of individuals with Asperger syndrome and other autism spectrum disabilities." Autism 12.2 (2008): 173-190.

[14] Billard, Aude, et al. "Building robota, a mini-humanoid robot for the rehabilitation of children with autism." Assistive Technology 19.1 (2007): 37-49.

[15] Torres, Nahum A., et al. "Implementation of interactive arm playback behaviors of social robot Zeno for autism spectrum disorder therapy." Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments. ACM, 2012.

[16] Shick, Aubrey. "Romibo robot project: an open-source effort to develop a low-cost sensory adaptable robot for special needs therapy and education." ACM SIGGRAPH 2013 Studio Talks. ACM, 2013.

[17] Kozima, Hideki, Marek P. Michalowski, and Cocoro Nakagawa. "Keepon." International Journal of Social Robotics 1.1 (2009): 3-18.

[18] Breazeal, Cynthia, and Brian Scassellati. "A context-dependent attention system for a social robot." rn 255 (1999): 3.

[19] Dautenhahn, Kerstin, et al. "KASPARa minimally expressive humanoid robot for humanrobot interaction research." Applied Bionics and Biomechanics 6.3-4 (2009): 369-397.