Emotion and Attention Level Detection for Children with ASD using Deep Learning

By, Kritika Iyer

Directed Research Report

Professor Gregory Fischer

Introduction

Autism Spectrum Disorder

Autism Spectrum Disorder(ASD) is a condition where the brain does not develop at normal speed. This causes deficients in social and motor skills which are essential for daily interactions and tasks [1]. Social and motor skills are important for a growing child as they are main facets for learning. ASD can be identified before a child turns three years old. Early intervention can play a big role in rehabilitating the child into normal society [2]. It has also been seen that robotic intervention proves more efficient for social skill rehabilitation [3] [4].

The social impairments of children with ASD ranges from speech issues to social interactions on a daily basis[3]. A characteristic quality of a child with ASD is the lack of eye contact during interactions with other people [7]. They wish to have more social interactions but are not capable of doing so. They end up feeling lonely and set aside in society[5][6]. Due to the inability to assess situations and peoples feelings this affects learning especially in the crucial first 10 years of their life. Through intervention during this time, most of these can be resolved and the child can lead a relatively normal life.

PABI and Other Rehabilitative Robots

There are many robots which have been used to rehabilitate children with ASD. Some socially assistive robots have been mentioned in the following paragraph. Robots such as Robota[8] and Zeno[9] , imitate the therapist and make sure the child follows. Romibo[10] shows emotion to help the child slowly improve interactions and get used to other peoples emotions. Keepon[11] is a small robot that uses a non-verbal method to convey emotions. It vibrates on a counter-top to convey emotions to the child. Kimset[12] is an anthropomorphic robot with a head which can recognize and show emotions. Kaspar[13] 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(Penguin for Autistic Behavioral Intervention) is another such robot which will be used to carry out this project.

PABI(Penguin for Autistic Behavioral Intervention) is a social robot designed in the WPI Automation and Interventional Medicine (AIM) Robotics Lab. This is used to help children with ASD improve upon their social skills [14]. PABI uses a tablet as shown in Fig.1 to communicate with the child and ask questions. On each correct answer PABI displays an action such as chirping or flapping it's wings along with words of praise which acts as positive reinforcement. PABI has other visual and audio queues which helps the child perform and learn. PABI is equipped with a face detection algorithm and is capable of detecting each child and is a useful feature for this project.

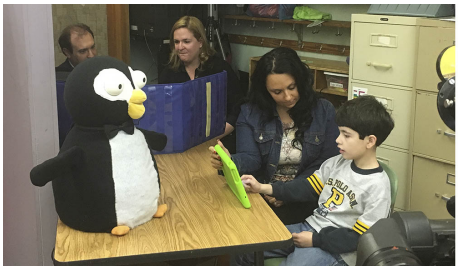


Figure 1: PABI robot with tablet and child

Emotion and Attention Level Detection using Deep learning

PABI can detect faces and it would be nice to further PABI's capabilities by making PABI adept at detecting human emotion. This can be achieved by deep learning. Deep learning has been used in robots to detect the best grasping positions for different objects, 3D object classification, emotion detection and other applications involving vision [17] [18] [19] [20]. Deep learning is a part of Machine learning which uses Neural Networks to learn a certain task. A neural network consists of many neurons/nodes connected to each other by functions which hold a certain value/weight in the determination of the output. There can be many layers of neurons and the layers in between can be visible or hidden. In Fig.2 there is one hidden layer and hence this is a single layer network. The various weights for functions can be seen and all the cumulative relationships between each node gives the output Y. There are different types of Deep learning algorithms, broadly classified as, supervised and unsupervised.

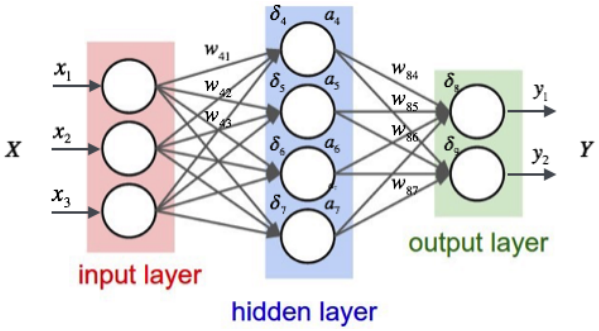


Figure 2: single layer neural network

For supervised learning the input and output are fed to the algorithm and the algorithm comes up with an output Y. This output is then compared to that of the actual/supervised output. The weights are then adjusted accordingly and the next set of input and output is given. In this way large data is trained upon and weights are adjusted to provide the correct answer [15].

In Unsupervised learning there is no output and only and input X. This input goes through the network and the output of this algorithm is basically a recreation of the input. The output Y is matched with the input X and weights changed to get a more accurate reconstruction [16]. For example, for feature extraction images of the number 8 can be given to the algorithm. This figure is then convoluted and reconstructed by the algorithm. The reconstruction is then compared to the original picture and weights are changed in accordance to that. By doing this the algorithm slowly learns what features exists in the number 8.

There are many ways in which emotions have been tracked in robots [31] and as softwares.Some of these emotions are based on facial expression and feature extraction [28] while some forms of emotion recognition occurs by assessing speech [25] [26] [27] and also through EEG signals [29]. Emotion recognition by facial expression has been done by using various methods like using deep learning [30] , semantic based trees [32] and hierarchical Binary decision trees [33]. Emotion tracking in socially assistive robots has been done in a few robots. Some robots like keepon and romibo are capable of exhibiting emotion but not detecting it [11] [10]. Robots like 'Kismet', 'Kaspar' and FACE can exhibit and detect emotions [12] [13] [34]. For PABI, face detection is possible using trees and this project will make emotion detection using deep learning possible.

Attention level detection is an important feature in Robots that deals with human interaction as it is a good measure for the quality of this interaction. This metric however is difficult to measure. For the cases of PABI there are two behavioral analysts who go through videos frame by frame and label the data for the child being on and off task. Using this data it is possible to know how well the sessions are going and where improvements need to be made. To reduce the load of the analysts, we would like to introduce and algorithm that might be able to perform at the same level as the analyst.

Problem Statement

PABI has an Existing Face detection algorithm and this can be used to better the experience the child has with PABI. Last Semester I worked on an emotion detection algorithm which can detect 5 emotions(Happy, sad, neutral, Anger/disgust, surprise) to an accuracy of 68.5%. Using better data the accuracy can be increased. This will allow PABI to detect the child, the child's emotion and the reaction of the child to a certain action. Using a similar model attention level can also be detected. This can help detect if the child is paying attention to the task or not. This will give us information about how well the session is going and how well the child is improving over a period of time.

Methodology

Feature Extraction

For emotion detection first many databases were taken and images as a whole were processed. In the previous semester this resulted in a 65% accuracy. After much manipulation and consideration this dataset was then made into a list of arrays consisting of [x,y] pixel points which refers to 68 feature points. The face is detected and taking that the feature points are detected. This can be seen in Fig.3. The white dots are the feature points and the black box is the face detected.

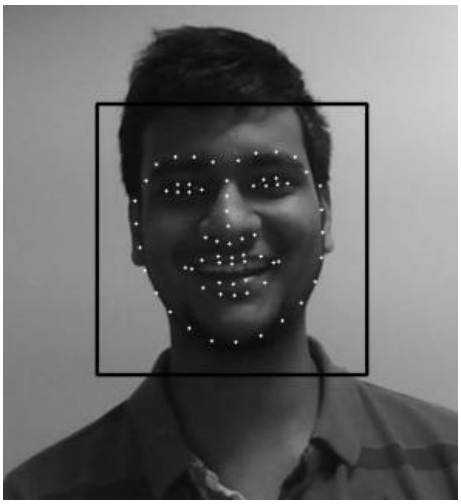


Figure 3: Feature Extraction

After storing all the feature points for all the images in the data the model was trained. This resulted in an increase to 85.19% accuracy. The same was used for attention level as the data was a lot of 300 x 300 RGB images.

Deep learning model

For images the network is convolutional in nature with multiple layers of max pooling, convolutions, dropouts and activations. When feature points were used a fully connected neural network was used with 3 dense layers.

For emotion detection the input is 136 values and output is 3 values (positive, negative, neutral) for the 3 emotions being detected. For attention level the input is 136 points and the output is 2 values for on task or off task determination.

Attention level

The following image(Fig.4) shows the flow of information and the working of the attention level algorithm that ended in a cumulative csv file. In this there were 5 children who had around 11-15

videos of their sessions with PABI. These were assesed by 2 behavioral analysts. Some videos were taken for validation and some for testing and some for training. The data was then used to train 3 different models. One model was trained for each child separately, another removed one whole child and trained on 4 children and another model was trained on all the children removing one video from each child for validation.

Using this information results from behavioral analyst(BCBA) 1 and 2 are compared to the model results to create a csv hence making computation of results easier.

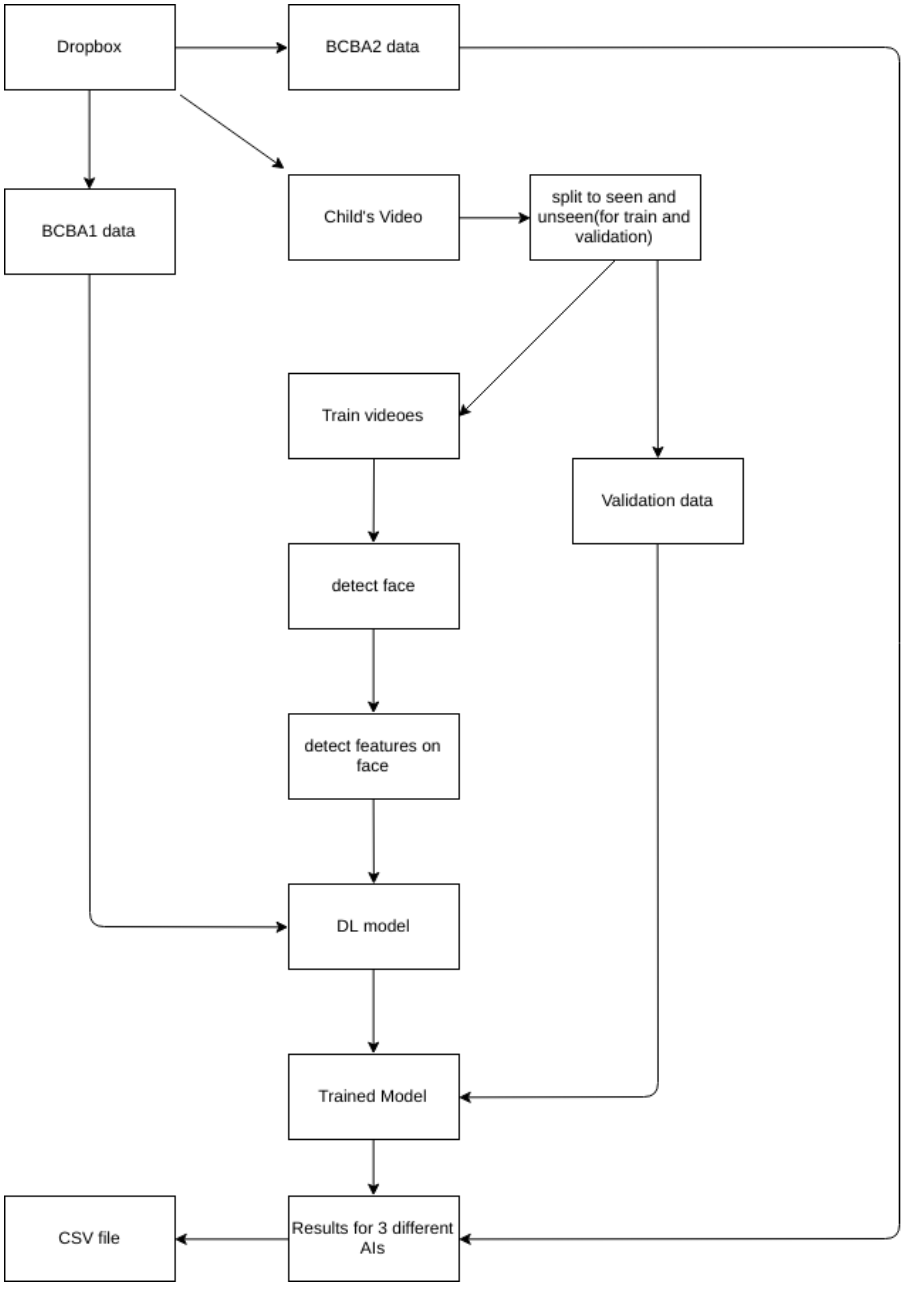


Figure 4: Flowchart

Assumptions

For this project to obtain good results in the time given the following assumptions were made.

- The BCBA1 data is treated as the ground truth
- Optical flow is not taken into consideration at the moment but will be incorporated later

- taken into consideration that only the data in which a face could be detected is data I am using.
- If a face is not detected there is no assumption on the state of the child. There is simply no output

Results

Emotion Detection

For emotion detection the optimally trained model had an accuracy of 85.19% The graph output for training and test accuracy looked like that shown in Fig.5 and the loss is shown in Fig. 6.This model was tested on validation data of another dataset and the results were promising. This shows that the model is robust in nature. The representation of this can be seen in the confusion matrix in Fig.7.The model was also tested on real-time video feeds as well as existing videos. It worked pretty well and the result for this is shown in Fig.8

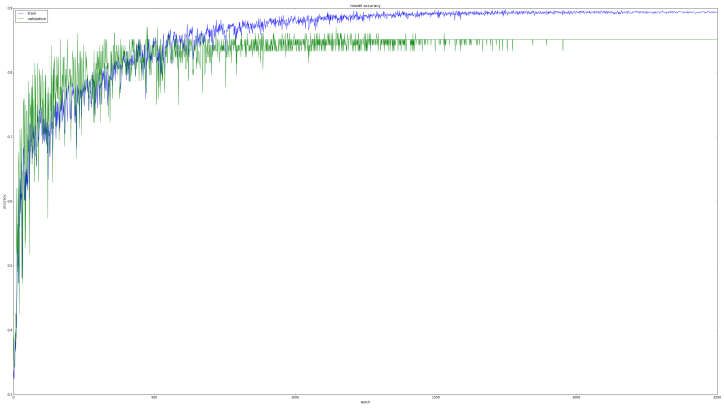


Figure 5: Accuracy plot for emotion detection

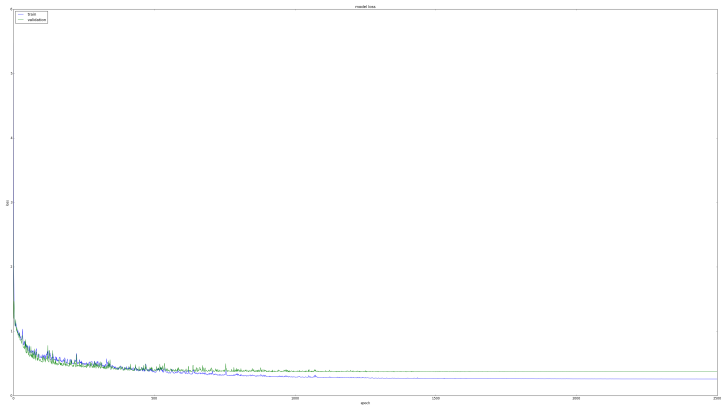


Figure 6: Loss graph for emotion detection

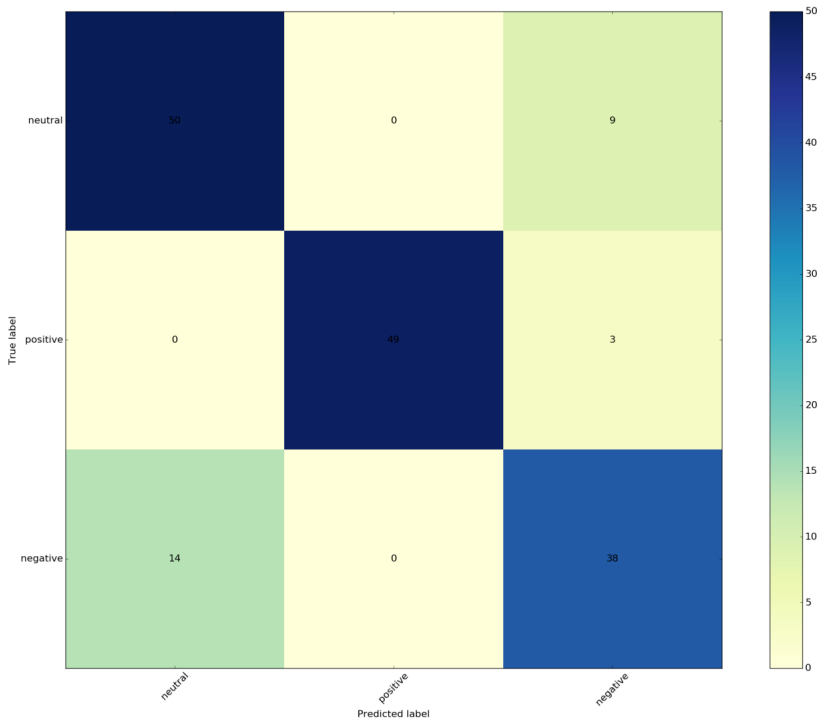
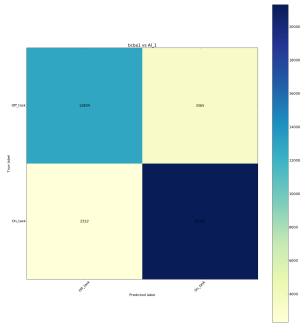


Figure 7: Confusion matrix for emotion detection

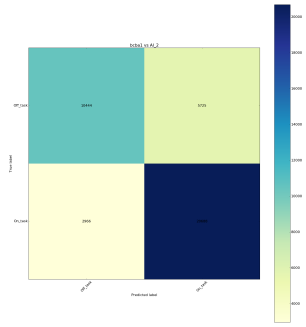
Figure 8: Application of emotion detection on a video

Attention Level Detection

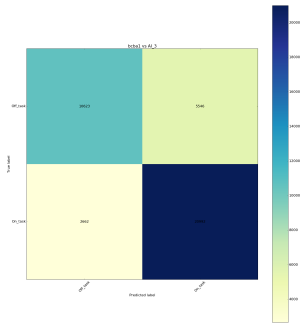
There were 3 algorithms that were trained. AI1,AI2 and AI3. These algorithms are trained with different train and validation data. for AI1 the model is trained separately for each child and validated on each child. The validation data is 2 videos out of all the videos of that child. The test data is the rest of the data for that child. For AI2 the model is trained on all the data together. The validation data is all the videos of one child. AI3 is also all the videos together except one video from each child. All the videos have been manually labeled by behavioral analysts 1 and 2 and the results of all of these have been compared.



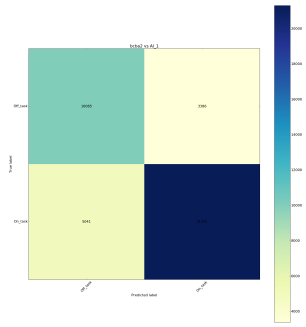
(a) AI 1 with respect to BCBA 1



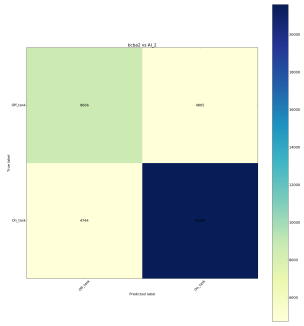
(b) AI 2 with respect to BCBA 1



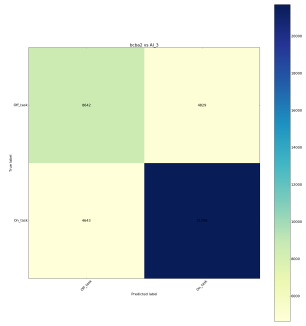
(c) AI 3 with respect to BCBA 1



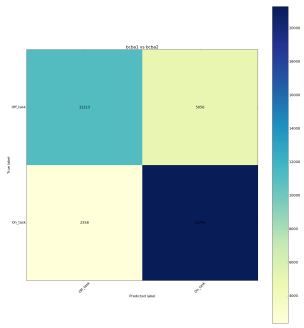
(d) AI 1 with respect to BCBA 2



(e) AI 2 with respect to BCBA 2



(f) AI 3 with respect to BCBA 2



(g) BCBA 2 with respect to BCBA 1

Figure 9: Confusion matrixes

Future work

For the future work these algorithms can be improved upon and the assumptions made could be educed to get a more robust and accurate model for the detections. The range of emotions maybe widened and detected better with other data or other processing of data. The csv files that were made as a result of this project could be properly studied and this would give a better idea on the information we have.

A build on to this could also be a reinforcement algorithm to detect how well the sessions are going. This would give us the ability to customize sessons for each child. The algorithm may look something like Fig 10.

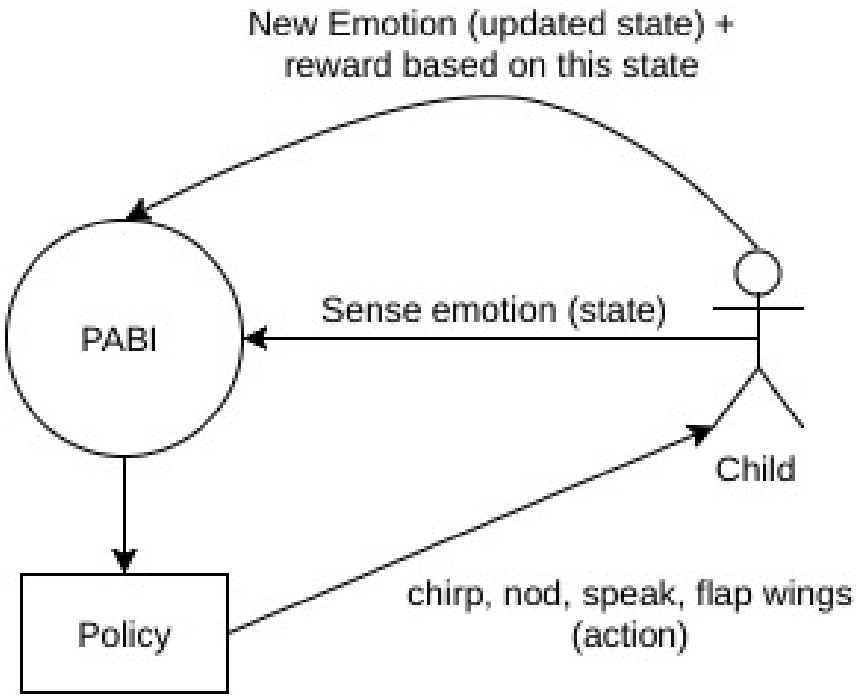


Figure 10: Possible Architecture for reinforcement learning

References

[1] Lord, C., Cook, E. H., Leventhal, B. L., & Amaral, D. G. (2000). Autism spectrum disorders. Neuron, 28(2), 355-363.

[2] Smith, T. (1999), Outcome of Early Intervention for Children With Autism. Clinical Psychology: Science and Practice, 6: 3349. doi:10.1093/clipsy.6.1.33

[3] Feil-Seifer D., Matari M.J. (2009) Toward Socially Assistive Robotics for Augmenting Interventions for Children with Autism Spectrum Disorders. In: Khatib O., Kumar V., Pappas G.J. (eds) Experimental Robotics. Springer Tracts in Advanced Robotics, vol 54. Springer, Berlin, Heidelberg

[4] Cabibihan, JJ., Javed, H., Ang, M. et al. Int J of Soc Robotics (2013) 5: 593. https://doi.org/10.1007/s12369-013-0202-2

[5] 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.

[6] 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.

[7] Handbook of Autism and Pervasive Developmental Disorders, Assessment, Interventions, and Policy. John Wiley & Sons; 2014 [Retrieved 24 December 2014]. ISBN 1-118-28220-5. p. 301.

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

[9] 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.

[10] 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.

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

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

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

[14] 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.

[15] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[16] Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009, June). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th annual international conference on machine learning (pp. 609-616). ACM.

[17] Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., & Quillen, D. (2016). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. The International Journal of Robotics Research, 0278364917710318.

[18] Lenz, I., Lee, H.,& Saxena, A. (2015). Deep learning for detecting robotic grasps. The International Journal of Robotics Research, 34(4-5), 705-724.

[19] Socher, R., Huval, B., Bath, B., Manning, C. D., & Ng, A. Y. (2012). Convolutional-recursive deep learning for 3d object classification. In Advances in Neural Information Processing Systems (pp. 656-664).

[20] Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 443-449). ACM.

[21] Sutton, R. S., Barto, A. G., & Williams, R. J. (1992). Reinforcement learning is direct adaptive optimal control. IEEE Control Systems, 12(2), 19-22.

[22] Matari, M. J. (1997). Reinforcement learning in the multi-robot domain. Autonomous Robots, 4(1), 73-83.

[23] Smart, W. D., & Kaelbling, L. P. (2002). Effective reinforcement learning for mobile robots. In Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on (Vol. 4, pp. 3404-3410). IEEE.

[24] Beom, H. R., & Cho, H. S. (1995). A sensor-based navigation for a mobile robot using fuzzy logic and reinforcement learning. IEEE transactions on Systems, Man, and Cybernetics, 25(3), 464-477.

[25] Han, K., Yu, D., & Tashev, I. (2014). Speech emotion recognition using deep neural network and extreme learning machine. In Fifteenth Annual Conference of the International Speech Communication Association.

[26] Kwon, O. W., Chan, K., Hao, J., & Lee, T. W. (2003). Emotion recognition by speech signals. In Eighth European Conference on Speech Communication and Technology.

[27] Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., ... & Narayanan, S. (2004, October). Analysis of emotion recognition using facial expressions, speech and multimodal information. In Proceedings of the 6th international conference on Multimodal interfaces (pp. 205-211). ACM.

[28] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, 44(3), 572-587.

[29] Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. The Scientific World Journal, 2014.

[30] Kahou, S. E., Pal, C., Bouthillier, X., Froumenty, P., Glehre, ., Memisevic, R., ... & Mirza, M. (2013, December). Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM on International conference on multimodal interaction (pp. 543-550). ACM.

[31] Rani, P., Liu, C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in humanrobot interaction. Pattern Analysis and Applications, 9(1), 58-69.

[32] Zhang, L., Jiang, M., Farid, D., & Hossain, M. A. (2013). Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Systems with Applications, 40(13), 5160-5168.

[33] Lee, C. C., Mower, E., Busso, C., Lee, S., & Narayanan, S. (2011). Emotion recognition using a hierarchical binary decision tree approach. Speech Communication, 53(9), 1162-1171.

[34] Pioggia, G., Sica, M. L., Ferro, M., Igliozzi, R., Muratori, F., Ahluwalia, A., & De Rossi, D. (2007, August). Human-robot interaction in autism: FACE, an android-based social therapy. In Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on (pp. 605-612). IEEE.