Applied Behavior Analysis Therapy for Autistic Children through Audio & Video for Psychological Intervention using Artificial Intelligence

Rubini.P[1], Midhunchakkaravarthy[2], Hemalatha[3]

[1] Lincoln University College, Malaysia; [2] Lincoln University College, Malaysia; [3]Panimalar Engineering College, India

prubini.cs@gmail.com, midhun@lincoln.edu.my, pithemalatha@gmail.com

---

Abstract:

Autism Spectrum Disorder (ASD) is a neurological condition that impacts an individual's cognitive, emotional, physical, and social well-being. Recognizing the gestures of autistic children is crucial for preventing meltdowns and self-harm. In this study, we introduce a novel method to identify gestures by detecting poses through a person pose estimation technique. The features extracted from the pose estimation are then used to develop a gesture classification model using supervised learning algorithms. Our proposed model achieved the highest accuracy with the Random Forest technique, exhibiting evaluation metrics of 83% precision and 71% recall.

Keywords: Autistic Children; Gesture Identification; Person Pose Estimation; Supervised learning; Random Forest Technique

---

Introduction

Autism Spectrum Disorder (ASD) impacts social communication and interaction in both children and adults. Early detection and intervention can enhance the health, development, and mental well-being of affected children. Children with autism often respond differently to sensory inputs; for example, they might close their ears in response to loud noises or shut their eyes when disturbed by flickering fluorescent lights or fluttering curtains and posters. Research indicates that before a meltdown, children exhibit signs of distress, known as the rumble stage [1][2]. Observable involuntary cues of anxiety, such as head banging, self-scratching, wrist biting, kicking, and hand-flapping, typically precede a meltdown. Identifying these behaviors early can help parents and teachers manage the situation more effectively.

This research is organized as follows: Section II discusses related work on pose estimation and gesture recognition. Section III explains the proposed new framework for gesture detection.

Related work

Recently, artificial intelligence has made significant strides in various domains, including image classification, speech recognition, text classification, and more. Numerous studies have been conducted on autism detection utilizing machine learning and deep learning techniques. Authors in [1] propose a deep learning approach to identify autism meltdowns. A classifier was developed to predict gestures from images that can be used to identify meltdowns. The model was trained using a RCNN supported by the pretrained GoogleNet model and achieved a validation accuracy of 93%. The training process included a loss classifier with a marginal loss of 0.4%. The authors in [3] propose a single-shot model using a box-free, downside-up approach for instance segmentation and pose estimation. A Convolutional Neural Network (CNN) is used to group key points for person pose estimation by detecting significant points and their displacements. The model was trained on the COCO dataset [4], achieving an average precision of 0.67 on the COCO test dataset using a single-scale inference model, and 0.69 using a multi-scale inference model, outperforming previous models. Specifically, it achieved an average precision of 0.42 for the person category. The classification of children as ASD based on body movement behaviors is discussed in [5].This study involved twenty-four autistic children and twenty-five children with neurodevelopmental disorders participating in a virtual reality event. Body movement variations were recorded using a depth-sensor camera, revealing that children with ASD exhibited more bodily movements compared to others. The head, body, and foot movements demonstrated the highest classification accuracy, achieving 83%.

PROPOSED METHODOLOGY

The proposed work introduces a novel method for identifying gestures in autistic children. The input image, captured via a webcam, is processed through the framework. The architecture is divided into two stages: the first stage focuses on person pose estimation, and the second stage handles gesture recognition, as illustrated in Fig. 1.

target classes namely “close ears” and “close eyes”.

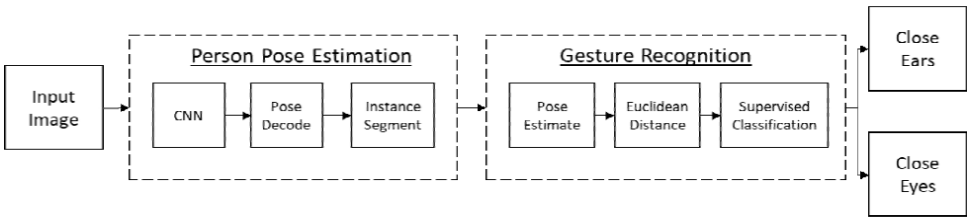

Fig. 1. Proposed Architecture

A. Person Pose Estimation

For this analysis, the COCO dataset [4] was utilized, specifically selecting images that depict actions such as closing ears and eyes. A CNN model with a ResNet-101 backbone was employed, using a batch size of 8 images and trained over 30 epochs. The model predicts key points of a person in the image through a hybrid classification and regression approach, generating heat maps and displacements. Subsequent steps involve predicting short-range offsets and mid-range pairwise offsets for the pose estimation model. Person segmentation maps and long-range offsets are utilized in the instance

segmentation module to produce pose estimates for 17 key points on the body, which are then used for the gesture recognition model.

B. Gesture Recognition

Seventeen pose key points coordinates are utilized to construct a feature vector for the gesture classification model based on Euclidean distance. Examples of pose key points include leftEye, rightEye, leftEar, rightEar, nose, leftWrist, rightWrist, among others. The distances between the wrist and the positions of the ears and eyes are computed using Euclidean Distance as shown in equation (1). These distances serve as features for building a binary classifier to detect gestures such as "close ears" and "close eyes." Various classifiers are tested, including Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Random Forest.

$$d_{1,2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Learning Algorithms

REFERENCES

[1] V. S. P. Patnam, F. T. George, K. George and A. Verma, "Deep Learning Based Recognition of Meltdown in Autistic Kids," 2017 IEEE International Conference on Healthcare Informatics (ICHI), 2017, pp. 391-396, doi: 10.1109/ICHI.2017.35.

[2] Erikk Linstead, and Renee, “An Application of Neural Networks to Predicting Mastery of Learning Outcomes in the Treatment of Autism Spectrum Disorder,” IEEE 14th International Conference on Machine Learning and Applications, California, 2015.

[3] George Papandreoua, Tyler Zhuu, Liang-Chieh Chenn, Spyros Gidaris, Jonathan Tompson, “PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model.” ECCV 2016.

[4] Lin, Tsung-Yii, M. Mairre, Serge J. Belongiee, James Hayes, P. Perona, D. Ramanan, Piotr Dollár and C. L. Zitnick. “Microsoft COCO: Common Objects in Context.” ECCV (2014).

[5] Alcañiz Raya and Mariano, “Machine Learning and Virtual Reality on Body Movements' Behaviors to Classify Children with Autism Spectrum Disorder.” Journal of clinical medicine vol. 9,5 1260. 26 Apr. 2020, doi:10.3390/jcm9051260.