Event detection in therapy sessions for children with Autism

Alexandre Soli Soares
Jônata Tyska Carvalho
Universidade Federal de Santa Catarina

Guilherme Ocker Ribeiro
Mateus Grellert
Universidade Federal de Santa Catarina

ABSTRACT

Autism spectrum disorder (ASD) impacts communication and cognitive development of children and adults, has a worldwide prevalence of 1% on children and affects not only the people with this disorder, but also their family and the surrounding community. In the family circle, individuals on the spectrum require greater support and attention relative to its cognitive capacity, impacting the mental and emotional health and even the financial life of families. The lack of infrastructure, professionals, and public health policies to deal with ASD is a known problem, specially in low income countries. To mitigate this issue, computer-aided ASD diagnosis and treatment represent a powerful ally, reducing the workload of professionals and allowing a better overall therapeutic experience. This paper intends to investigate how machine learning techniques can help specialists by providing an automated analysis of ASD recorded therapy sessions. The proposed solution is capable of handling large amounts of video data, filtering out irrelevant frames and keeping only relevant scenes for posterior analysis. Our results show that the proposed solution is capable of reducing manual checks by up to 51.4%, which represents a significant workload reduction for health experts. This solution will hopefully provide researchers, therapists and specialists with a tool that assists the automated identification of features and events of interest in video-recorded therapy sessions, reducing the amount of time spent on this task.

KEYWORDS

autism spectrum disorder machine learning therapy.

1 INTRODUCTION

Autism spectrum disorder (ASD) affects the cognitive development and communication skills of children and adults, limiting the functional capacity of these individuals. ASD has a worldwide prevalence of 1% on children and it affects not only the people with this disorder, but also their families and the surrounding community[Zeidan et al. 2022]. In the family circle, individuals on the spectrum require greater support and attention relative to their cognitive capacity, impacting the mental and emotional health and even the financial life of families[Kołakowska et al. 2017].

Individuals in the spectrum are usually accompanied by a doctor along with frequent sessions of ASD therapy. These sessions contain qualitative and quantitative data about the patient's progress and evolution that a lot of times go unnoticed. Data that is generated by questions like if the patient responded to a stimulus, how many responses happened, and whether the patient engaged in a person-to-person activity or preferred not to. But the way data is collected is far from standard, thus not maintaining a concise record of the patient[Ramirez-Duque et al. 2018].

One way to tackle this problem is to video record each therapy session, composing a database with diverse features including the ones mentioned above and many others. But again, which standard is going to be used to collect this information, and who is going to collect it? Manually reviewing and taking notes of a therapy session could take more time than the session itself, which ends up being unfeasible for a large amount of data and/or when the responsible therapist treats other patients[Kołakowska et al. 2017].

This study aims to provide an automated event detection mechanism that is capable of filtering out irrelevant scenes of recorded therapy sessions to assist professionals in their analysis. The designed system is capable of processing a large amount of video data, and automatically detect interactions between actors of the scenes. This is done by using state-of-the-art machine learning techniques for object detection combined with an ad hoc heuristic. The detected events are then used to build a user-friendly timeline of interactions for each therapy session footage.

2 PROPOSED SOLUTION

Considering the length and the number of therapy sessions needed for an individual with ASD, and the need to review these recordings for documenting relevant events for the therapy assessment, we propose an end-to-end tool capable of providing an automated analysis of actors interactions in ASD therapy sessions. The proposed system can recognize different pairs of interactions among the actors, including child-toy, child-therapist and therapist-toy.

2.1 Dataset

The data used for training and testing the proposed solution are videos of therapy sessions for children with ASD. It was provided by the european project PlusMe (https://www.plusme-h2020.eu/) as supplementary material published in [Sperati et al. 2020].

2.2 Proposed solution

The framework is composed of 3 main steps, the first one process the video-recorded session through a computer vision tool able to detect and classify the existing actors in each frame. Next, these detections are processed in combination with each other to identify the overlaps between them. Finally, based on these overlaps, an interaction timeline is built showing in which frames of the video there are potential interactions, as shown in figure 1.

2.2.1 Actors detection tool - YOLOv5 -. For this task, we employed the YOLOv5 model available from the Ultralytics repository under the GNU General Public License v3.0 and implemented it with the open-source machine learning framework PyTorch 2. The only modifications made to the default hyper-parameters were changing the batch-size to 1 and image size to 256x256 pixels. Such modifications had to be done due to memory limitations of the training environment (GPU) detailed later. The network was trained using transfer learning with the default YOLO network (trained on COCO dataset), and the training dataset.

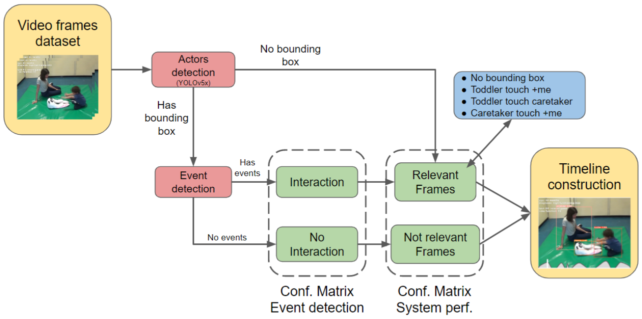

Figure 1: Proposed solution outline. First, the content of a recorded therapy session is loaded into a Computer vision tool, generating a bounding box for each actor. Next, a heuristic-based event detection mechanism outputs relevant information about interactions.

2.2.2 Bounding box event detection -. The proposed event detection tool computes whether two bounding boxes intersect for each frame. Since a simple overlap, does not always mean an interaction, we treated this as a "possible interaction event". Such predictions are then evaluated against the "ground truth" to evaluate the accuracy of the event detection tool. For sake of simplicity, at the current state of our system, the interaction between the actors is defined by any overlap between their bounding boxes.

2.2.3 Event detection timeline construction -. Finally, we use the detected bounding box interactions information from the previous step, group them accordingly and merge this information as an overlay on the original video, highlighting the actors interaction frames and enabling a faster evaluation of the interaction between them in the recorded video therapy.

3 RESULTS AND DISCUSSIONS

3.1 Actors detection tool - YOLOv5

It used the extra-large model YOLOv5x, which produces better results in nearly all cases but has more parameters. This model was trained for 189 epochs, in a batch size of 1 and image size of 256x256 due to GPU limitations. We used a NVIDIA GTX 1650 GPU with 4 Gb VRAM for training the model. The training took around 16 hours. The best results were found at 89th epoch achieving a precision of 0.94, recall of 0.82, and mAP@0.5 of 0.90.

3.2 Events detector tool

Since the tool predicts the existence of interactions of the bounding boxes, we can evaluate these predictions by manually asserting if there is in fact an interaction at the given frame (ground truth). As the objective is to point out which frames are important, it succeeds featuring an almost zero number of False Negatives.

With prior knowledge of this particular dataset and to properly evaluate the events detector algorithm, we applied a filter beforehand and sent all the frames where the actors detection tool failed to provide a bounding box for any of the actors directly to the end of the system. These filtered frames are annotated as relevant for the final analysis and bypass the event detector algorithm entirely.

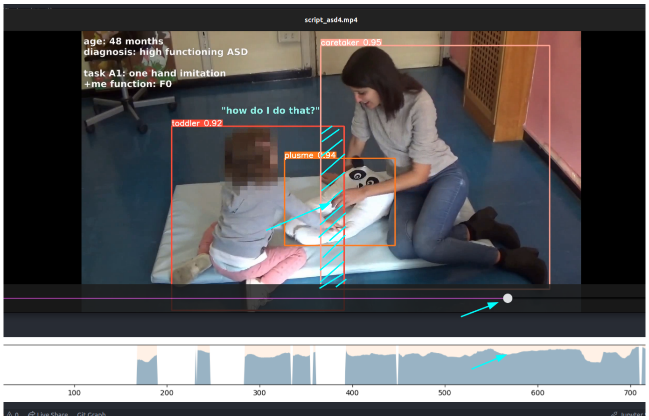

Figure 2: A screenshot from the prototype output of the proposed framework, highlighting the bounding-boxes interaction and the timeline construction tool with the relevant frames in evidence.

Table 1: Reduction of analysed video data considering different types of interaction

Video	Frames	Td&Ct	Ct&+me	Td&+me	All
Video 1	100	34.0%	0.00%	35.0	0.0%
Video 2	100	51.4%	17.8%	18.8	17.8%

3.3 System prediction performance

In the final step of the system, where we join both actor's and event's detection streams, it's possible to assess the prediction of each frame to compose the highlighted snippets that the system proposes. Like a binary prediction, again, we can evaluate the results in a confusion matrix that also features a low number of False Negatives, really filtering only relevant scenes.

3.4 Video length reduction

Random samples of videos (test set) show that it is possible to filter relevant frames for one kind of interaction making it 49% of the whole video, resulting in a reduction of time required for analysis up to 51%, as shown in Table 1. A Sample of the interaction timeline can be seen in figure 2.

REFERENCES

Agata Kołakowska, Agnieszka Landowska, Anna Anzulewicz, and Krzysztof Sobota. Oct. 2017. "Automatic recognition of therapy progress among children with autism." *Scientific Reports*, 7, 1, (Oct. 2017). DOI: 10.1038/s41598-017-14209-y.

Andrés A. Ramirez-Duque, Anselmo Frizera-Neto, and Teodiano Freire Bastos. 2018. "Robot-Assisted Diagnosis for Children with Autism Spectrum Disorder Based on Automated Analysis of Nonverbal Cues." In: *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, 456–461. DOI: 10.1109/BIOROB.2018.8487909.

Valerio Sperati et al.. 2020. "Acceptability of the transitional wearable companion "+me" in children with autism spectrum disorder: a comparative pilot study." *Frontiers in psychology*, 11, 951.

Jinan Zeidan, Eric Fombonne, Julie Scorah, Alaa Ibrahim, Maureen S Durkin, Shekhar Saxena, Afiqah Yusuf, Andy Shih, and Mayada Elsabbagh. 2022. "Global prevalence of autism: a systematic review update." *Autism Research*, 15, 5, 778–790.