Early Screening for Children with Fetal Alcohol Spectrum Disorder via Natural Viewing Behavior

Chen Zhang¹, James Reynolds², Douglas P. Munoz², Laurent Itti¹,³

¹Neuroscience Graduate Program, University of Southern California, ²Department of Biomedical and Molecular Science, Queen’s University, ³Department of Computer Science, University of Southern California, ⁴Center for Neuroscience Studies, Queen’s University

Abstract
Fetal alcohol spectrum disorder (FASD) is one of the most prevalent neurodevelopmental disorders caused by prenatal alcohol consumption [1]. Children with FASD may suffer from growth retardation, structural and functional brain abnormalities, and cognitive and behavioral impairments [2-4]. Costs associated with FASD, in areas such as medical care, special education and facilities, and social services, can run into billions of dollars annually [5,6], placing high psychological pressures and large economic burdens on families and society. An easy, objective and widely-deployable early screening procedure is of significant importance because it can lead to earlier intervention and treatment, which is crucial for neurodevelopmental disorders, to mitigate the disruptive effects of the disorder and/or frequency of secondary comorbidities. Yet, the current diagnostic process or proposed screening tools are challenging and time- and money-consuming, vary across clinics and countries, leading to inconsistent diagnoses and treatments [7]. A large proportion of the at-risk population cannot receive timely diagnosis, particularly individuals with limited access to clinical resources. Thus, an objective, time- and cost-efficient screening and diagnosis system with globally unified standards is desperately needed in order to help target the at-risk population, accelerate the clinical process, and improve diagnostic accuracy and efficiency.

Children with FASD show altered natural viewing behaviors compared to the age-matched typically developing controls (TD) [8,9]. The recording of natural viewing behaviors can be applied to individuals as young as 2 months old, without any requirement of professionals. Previous studies have shown that oculomotor and attentional features extracted during free viewing of videos or static images could help differentiate clinical populations with various high-prevalent neurodevelopmental disorders [10, 11]. However, those studies based on SVM or convolutional neural networks either required a longer experiment time or didn't achieve the state-of-art classification accuracy [8,10]. In this study, we hypothesize that the entire eye movement dynamic and its temporal dependencies, together with the local visual attentional parameters and the global visual scene understanding which might reflect the internal viewing strategies, can further help the detection of children with FASD during less than 5-minute video watching. We used deep neural networks (DNN) containing multiple long-short-term-memory (LSTM) layers followed by a video selection classifier to give out the final prediction of the risk of having FASD.

We recorded the eye movements of participants while they naturally watched short video clips of various contents (Figure 1). Those raw eye traces (x, y coordinates) were treated as the oculomotor features of the inputs of networks. The local bottom-up attentional features were acquired on the scanpath of eyes from the saliency maps derived using the Itti-Koch saliency model [12,13], while the top-down attentional features were derived from the inter-observer map of a group of healthy young adults along the eye traces. We used the multi-resolution gist vectors of each video frames (which were used to accurately classifying the scenes) around the eye positions as another group of global features to reflect the understanding of the video [14]. The inputs of each participant were down-sampled and augmented 10 times to maintain an appropriate length while approximately preserving the natural viewing dynamics. The network structure of each video clips is shown in Figure 2. The masking layer was used to mask out any blinks or ocular drifts and the multiple LSTM layers were used to learn the temporal dependencies along the oculomotor, attention and visual understanding domains. The fully connected layer was used for a prediction and a voting step was followed to get one label and prediction probability for each participant.
per video clip. A final video clip selection and L1-regularized logistic regression were used to make the ultimate prediction and the profile of distinguishing videos.

We achieved a classification accuracy of above 78%, with sensitivity and specificity to be 73% and 83% on a completed separated testing set of almost equal size of the training, which is comparable to or better than the results of other studies strongly rely on demographic, behavioral, physical, psychometric, face morphometric and neuroimaging analysis, and history of maternal alcohol consumption [9,15-18]. We also discovered that eliminating any feature group or reversing the eye sequences resulted in a significantly drop of the testing accuracy, indicating that the disorder has impacts on all the aspects of oculomotor control, attention allocation and visual understanding, and the FASD and TD groups had different viewing strategies. Further analysis of distinguishing videos showed that the FASD group demonstrated different visual processing regarding color, motion, flicker of texture of the video and altered behaviors in top-down modulation.

The proposed screening procedure only takes a few minutes to run with almost no age restriction. Anyone who is trained to use the eye tracker, which is just a laptop connected to a high-speed camera, can administrate this procedure. Compared to the standard diagnostic process, our proposal is much more accessible and easier to implement with early or pre-school age children, and could be widely used at clinics, schools, or health units where young children are seen routinely, across different regions and areas. The data can be analyzed through the same pipelines and thus all the screening estimations are achieved through the same standards. Such a screening tool could be applied as an annual screening tool for children at risk of FASD with high-capacity and low-cost, which is particularly beneficial for populations with limited healthcare resources. It can also have substantial impact of the referral and diagnostic process, and expected maximized long-term benefits to the tested individuals and to society.

Figure 1. Natural viewing behavior recording.

Figure 2. DNN architecture.
References