Exploiting visual behaviour for autism spectrum disorder identification

Giuliano Arru and Pramit Mazumdar and Federica Battisti
Roma Tre University, Rome, Italy
giuliano.arru@uniroma3.it, pramit.mazumdar@uniroma3.it, federica.battisti@uniroma3.it

***Abstract*—In this contribution, a model for revealing the presence of autism spectrum disorder by exploiting visual information is developed. This condition is characterized by a deficit in social behaviour and nonverbal interactions such as specific facial expressions, reduced eye contact, and body gestures. Advancements in multimedia technologies can help in understanding symptoms for early detection of the disorder. In the proposed model, both the image content and the viewing behaviour are used for defining relevant features to be used in a machine learning-based classifier. A training phase is realized by taking multiple images and scanpaths representing the viewing behaviour of persons affected and not by the disorder. The influence of specific objects in the scene is considered. Finally, the number of fixations towards centre of the scene and duration for which a subject looked at the central area is also considered. A decision tree based classifier is used for training the model. The achieved results show that by taking into account the semantic and image features extracted from content, fixation, and center-bias, it is possible to estimate the presence of autism spectrum disorder. The results obtained in the performed experiments are promising even if they show room for improvement.**

***Index Terms*—Autism spectrum disorder, Object detection, Fixations, Centre bias, Classification, Visual saliency**

I. Introduction

Autism Spectrum Disorder (ASD) or Autism Spectrum Condition is a developmental disorder that affects communication and behaviour. Although autism can be diagnosed at any age, it is said to be a *developmental disorder* since symptoms generally appear in the first two years of life [1]. The early detection of this disorder may strongly improve the quality of life of the patient. As shown in literature, i.e., [2], the adoption of early intervention programs, starting from age 3-5, allow half of the children to gain enough skills to be mainstreamed for kindergarten. The methods usually adopted for ASD diagnosis rely on behavioural, historical, and parent-report information [3].

Recently, there has been an increasing interest in applying human bio-signals, particularly eye movements, to recognize the emotional gist of a scene such as its valence. The related studies exploit the advancements in both hardware and software technologies. In fact, eye tracking system are now available thus allowing to record precise information about the eye positions and eye movement and to infer about the subject's focus on a screen. At the same time, machine learning techniques now allow the research to design very accurate classifiers for different purposes. The ASD early diagnosis can benefit from the joint use of eye tracker and machine learning. In fact, the eye tracker is a non-invasive assessment tool and does not require advanced motor responses or language. Therefore, its use is relevant in the study of young children and infants diseases. In [4], the importance of features extracted from eye movements in the classification of images into pleasant, neutral, and unpleasant categories is addressed. A machine learning approach is used for analyzing the performance of features by learning a support vector machine and exploiting various feature fusion schemes. In [5] a tool for supporting the ASD diagnosis is proposed. The authors design a graphical user interface for visualizing the gaze estimation collected on videos shown to the patient. In [6] a study for assessing the predictive reasoning abilities of Typically Developing (TD) infants and 2-year-old children with ASD by exploiting eye-tracking is performed. A goal-based action was shown to the subjects. The results revealed differences in early predictive reasoning abilities. When predicting the action without kinematic cues, younger TD children show goal-based visual predictions, whereas in older TD children the visual prediction was not systematic. ASD children generated location-based predictions, suggesting that their visual predictions may reflect visual motor perseveration. In [7] the atypical visual-attention patterns of ASD patient is studied. In the performed experiment, participants were asked to perform two tasks (browsing and searching) while looking at a web page. The gaze was collected and used for training a machine learning classifier. The achieved results show that the differences in the way ASD subjects process web content might be used for the development of serious games for autism screening. A similar approach is adopted in [8] where, by exploiting data that describe the saccades of the patients sight, a high percentage of ASD classification is obtained. However the number of subjects is very limited (six subjects). In [9] atypical attention towards stimuli and their features is used for detecting ASD. An annotated dataset of 700 complex natural scene images has been used for testing a 3-layered saliency model incorporating pixel-level, object-level, and semantic-level attributes on 5551 annotated objects. The performed subjective test show that ASD subject has a stronger image center bias regardless of object distribution, reduced saliency for faces and for locations indicated by social gaze, yet a general increase in pixel-level saliency at the expense of semantic-level saliency. The importance of the face feature is addressed in [10]. Results show that social orienting is actually not qualitatively impaired in ASD patient and that decreased

978-1-5386-9214-1/19/$31.00 ©2019 IEEE
DOI 10.1109/ICMEW.2019.00123

attention to faces cannot be generalized across contexts. The authors also find poor evidence about the hypothesis that ASD individuals show excess mouth and diminished eye gaze compared to TD individuals. Also in [11] the eye-tracking system is exploited for understanding the visual attention to faces and objects in children with ASD. They obtained lower scores for face recognition and social-emotional functioning tasks but exhibited similar patterns of visual attention to the ones gathered from non-ASD subjects. In [12], the possibility of using visual based techniques for revealing the strategies adopted by ASD subjects when processing social information is addressed.

In this contribution, we exploit the visual behaviour of subjects while exploring images to take out distinct features. Those are extracted from three components: image content, fixations/viewing behaviour and bias towards centre of a scene. Finally, this set of features is used to train a decision-tree based classifier for identifying persons affected by autism spectrum disorder.

II. Proposed method

In Figure 1, the block diagram of the proposed RM3ASD model for classifying a subject as ASD or TD from its scanpath is depicted. The scanpath data consists of the coordinates of fixations and their duration.

Features are extracted from: image content, fixation points, and bias towards the centre of the image. In more details:

1) **Image content**: human attention is strongly dependent on the image content. For this reason, an object detection is performed for identifying the presence of objects in the image. To this aim, the You Only Look Once (YOLO) detector [13] is used. The detected objects are processed to extract content based features. Not all fixations for an image may belong to detected objects. In fact, persons with ASD tend to look at areas with no socially prominent objects [9]. Therefore, area not belonging to detected objects is considered as unique region. All fixation points lying in this region are considered as fixations on non-objects. Based on this assumption, the count, duration, and rate of fixations/duration on both objects and non-objects are computed (features Cnf1 to Cnf8, Table I). The rate of fixation/duration is computed as the fraction of fixations/duration on an object/non-object to the total number of fixations/duration.

   Persons with ASD tend to explore the content of images in a different way with respect to the TD persons. To cope with this evidence, we consider the spatial span of the fixation points as a feature for classifying ASDs from TDs [14]. This measure models the fact that the subject having a large visualization area but very low fixation coverage might have missed important parts of the image [15]. To compute the image coverage (Cnf9), a binary map is generated from the fixations and a threshold of 0.1 is set. Cnf9 is the percentage of non-zero pixels.

   Refixation (Cnf10) is a relevant factor in the study of human fixations that directly depends on the content of the image. They account for the number of times a subject returns to a region considering the total session of image viewing. Our intuition behind using this factor as a feature lies in the fact that refixation is a general tendency of users while viewing a multimedia content. Therefore, this can be used as a tool for separating a TD person from a person affected by ASD.

   In RM3ASD model, we integrate a feature that accounts for the saliency of an image. The saliency detection by combining Simple Priors (SDSP) technique proposed in [16] is exploited due to its low computational complexity and good performance in estimating saliency of 2D images.

   Next, the saliency weight of the fixated pixel and the duration of fixations is used for computing a saliency influenced feature: Cnf11. If F is the fixation data of any subject u on image i, then Cnf11 is computed as,

$$Cnf11_i^u = \frac{\sum_{j \in F} S_{map}^{ij} * duration_j}{\sum_{j \in F} duration_j} \quad (1)$$

   where, S_{map}^{ij} is the saliency at fixated pixel j on image i, and $duration_j$ is the duration of the fixation j.

2) **Fixations**: The set of fixations performed by a subject represents its scanpath. Whereas, saccades are the intervals between two consecutive fixations [15]. Saccadic movements over an image vary from person to person. Thus, we explore a set of features from the saccadic movement and the fixations of eyes, denoted as ‘Fxf’ type in Table I. Fixation per duration (Fxf2) is computed as the ratio of number of fixations to the total duration of fixations.

   Additionally, we perform clustering for extracting Region-Of-Interests (ROIs) based on the fixations and not on the content which influences the fixations. This is due to the fact that, an ASD subject is less focused towards the socially relevant stimuli. For this purpose we employ the mean shift clustering approach [17]. It is performed in two steps: first, the weighted mean of the nearby points based on a multivariate Gaussian kernel function is assigned to each fixation. The mean shift property tends to move all fixations towards area of high density or convergence. Then, a standard distance based clustering approach is used to group the nearest fixations. By tuning the distance threshold/scale factor, it is possible to define the size of clusters: high threshold corresponds to few cluster with higher dimension. In this work, we set this distance/scale factor threshold to 100. After clustering the fixations on the input image to understand the region of interest, the following features are computed:

   - Number of region of interests in the input image;
   - Mean duration of fixations in the region of interests;
   - Maximum duration of fixations in the ROIs.

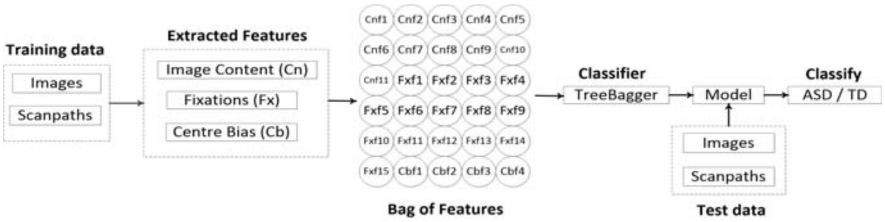

Fig. 1: Block diagram of the proposed RM3ASD model.

Image Content		Fixations		Center Bias	
Type	**Feature name**	**Type**	**Feature name**	**Type**	**Feature name**
Cnf1	No of fixations on objects	Fxf1	Total number of fixations	Cbf1	Image and ROI centre distance
Cnf2	No of fixations on non-objects	Fxf2	Fixation per duration	Cbf2	Fixation distance from image centre
Cnf3	Fixation duration on objects	Fxf3	Total duration of fixations	Cbf3	Number of fixations near image centre
Cnf4	Fixation duration on non-objects	Fxf4	Mean duration of fixations	Cbf4	Fixation duration near image centre
Cnf5	Fixation rate on objects	Fxf5	Longest fixation duration		
Cnf6	Fixation rate on non-objects	Fxf6	Index position of the longest fixation		
Cnf7	Duration rate on objects	Fxf7	Standard deviation of duration		
Cnf8	Duration rate on non-objects	Fxf8	Standard deviation of saccadic amplitudes		
Cnf9	Image coverage	Fxf9	Mean of saccadic amplitudes		
Cnf10	Refixation	Fxf10	Min. of saccadic amplitudes		
Cnf11	Image saliency	Fxf11	Max. of saccadic amplitudes		
		Fxf12	Total saccadic amplitudes		
		Fxf13	No of region of interests (ROIs)		
		Fxf14	Mean duration of fixations in ROIs		
		Fxf15	Max. duration of fixations in ROIs		

TABLE I. List of features extracted from image content, eye fixations and bias towards centre in the model RM3ASD.

3) **Centre bias**: as shown in [9], an ASD person has a stronger centre bias irrespective of the distribution of objects in the image. To incorporate this aspect into the RM3ASD model, multiple features are extracted from the input image and corresponding fixations that reflect the bias of the subject towards the centre of a scene. In more details, the following features are extracted:
   - Mean distance between image centre and ROI centre;
   - Mean distance of fixations from image centre;
   - Number of fixations close to the centre of the image. A distance radius of 100 pixels is set and any pixel within this threshold from the centre is considered as near;
   - Time spent by a subject near the centre of the image. This is computed from the duration of fixations near the centre;

III. Experimental results

The performance of the RM3ASD model is evaluated using a dataset of eye movements for children with ASD and TD [18] for the “Saliency4ASD Visual attention modeling for Autism Spectrum Disorder” Grand Challenge at IEEE ICME 2019. The dataset consists of normal 2D images and corresponding scanpath of subjects with ASD and TD condition. First, the training dataset was provided with 300 images and the corresponding scanpaths of ASD and TD subjects. Finally, the test dataset was provided with 100 images and scanpaths of subjects who were not labelled as ASD or TD.

The training dataset is restructured as <SubjectId, ImageId, X-coordinate fixation, Y-coordinate fixation, duration of fixation, ClassId>, with random SubjectId. It is useful to underline that in the dataset there is no strict correspondence between the fixation file index and the subject: the subject generating the fixation data_1 for image 1 may not be the one who generated the fixation data_1 for image 2.

The model is trained with the TreeBagger supervised learning function [19]. It is based on the random forest algorithm [20] and uses bagging method (bootstrap aggregation) for training. This algorithm is selected due to its accuracy in classifying data similar to the one used in this work (all numeric features and a final numeric class label for classification) [21]. The parameter for the number of trees in TreeBagger is set to 128. The model is evaluated using standard classification metrics such as sensitivity or recall, specificity or true negative rate and accuracy. Sensitivity or recall or detection rate depicts the performance of a model in identifying persons who test positive for ASD. Specificity or true negative rate is the fraction of TDs that are correctly identified on the total number of TDs in the dataset. Accuracy is the fraction of the number of correct predictions over the number of predictions by the model. Additionally, other popular metrics such as precision, F1, and AUC-ROC score are also computed for evaluating the performances of the proposed model. Table II shows the average performance (%) of our model for the test dataset. The average results are obtained by computing the mean of the individual scanpath classification results for each image. The 60% accuracy of the model on the test dataset

Accuracy	Recall	Specificity	F1	Precision	AUC-ROC
59.30	68.43	50.56	61.64	56.96	59.50

TABLE II. Performance of RM3ASD model on test dataset.

can be considered promising for future improvements. First, studies on subjects with ASD demonstrated that the behaviour of persons with this disorder change with age [10]. The fixation behaviours of a toddler (around 21 months old), young children (below 10 years) and grown up children (around 20 years) are different. Objects, such as face, are explored for a longer time by toddlers, but not by the grown up children. Toddlers can even reach the same fixation duration on human faces like the typically developed toddlers [22]. There is no visible difference in fixation duration between objects and non-objects for adults with ASD. Therefore, the ASD subject age is very important for detecting the condition. Thus, selecting appropriate features for modeling fixation behaviour of subjects require adaptation on basis of age. However, the proposed RM3ASD model does not include age information since this feature is not available in the given dataset. Second, there exists a difference in fixation behaviour among ASD grown children and TD ones. However, if a scene contains both human faces and social objects of circumscribed interest (i.e., plants or clothing) then fixation behaviour of a young child does not differ with that of a TD one [23]. Interestingly, the grown-up ASD children demonstrate a bias towards non-objects in a scene with faces. Therefore, ASD condition cannot be generalized over age, fixation behaviour, and image content. Whereas, in RM3ASD we generalized all features over the subjects.

IV. Conclusions

In this contribution, a model for ASD identification has been presented. It is based on the analysis of the visual behaviour of subjects while exploring images. The performed experiments highlighted the presence of aspects that could be taken into account for improving the performances of the system. First of all, object semantic (e.g., the type of object being looked at) is not considered in the current study. However, it is worth exploring the type of objects present in a scene. In fact, studies show that persons with ASD tend not to look at eye region of a face, instead look at the mouth region [24]. Moreover, future work could address an in-depth study on the areas of fixations coverage missed by the person with ASD. This analysis might provide an insight on what does not attract the attention of a person with ASD.

References

[1] "Autism spectrum disorder," https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml.

[2] J.H. Elder, C.M. Kreider, S.N. Brasher, and M. Ansell, "Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships," *Psychology Research and Behavior Management*, vol. 10, pp. 283–292, 2017.

[3] S.M. Brenda, J.B. Stacey, and L.S. Richard, "Asperger syndrome diagnostic scale (ASDS)," Pro-Ed, USA, 2001.

[4] H.R. Tavakoli, A. Atyabi, A. Rantanen, S.J. Laukka, S. Nefti-Meziani, and J. Heikkil, "Predicting the valence of a scene from observers eye movements," *PLoS ONE*, vol. 10, no. 9, pp. 1–19, 2015.

[5] K. Higuchi, S. Matsuda, R. Kamikubo, T. Enomoto, Y. Sugano, J. Yamamoto, and Y. Sato, "Visualizing gaze direction to support video coding of social attention for children with autism spectrum disorder," in *International Conference on Intelligent User Interfaces*. ACM, 2018, pp. 571–582.

[6] S. Krogh-Jespersen, Z. Kaldy, A.G. Valadez, A.S. Carter, and A.L. Woodward, "Goal prediction in 2-year-old children with and without autism spectrum disorder: An eye-tracking study," *Autism Research*, vol. 11, no. 6, pp. 870–882, 2018.

[7] V. Yaneva, L.A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting autism based on eye-tracking data from web searching tasks," in *Internet of Accessible Things*. 2018, pp. 16:1–16:10, ACM.

[8] R. Carette, F. Cilia, G. Dequen, J. Bosche, J-L. Guerin, and L. Vandromme, "Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach," in *Internet of Things Technologies for HealthCare*. 2018, pp. 75–81, Springer.

[9] S. Wang, M. Jiang, X.M. Duchesne, E.A. Laugeson, D.P. Kennedy, R. Adolphs, and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking," *Neuron*, vol. 88, no. 3, pp. 604–616, 2015.

[10] G. Quentin, H. Nouchine, B. Sophie, and Bernadette R., "Visual social attention in autism spectrum disorder: Insights from eye tracking studies," *Neuroscience & Biobehavioral Reviews*, vol. 42, pp. 279 – 297, 2014.

[11] J.C. McPartland, S.J. Webb, B. Keehn, and G. Dawson, "Patterns of visual attention to faces and objects in autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 41, no. 2, pp. 148–157, 2011.

[12] Blakemore S.J. Boraston Z., "The application of eye-tracking technology in the study of autism," *Journal of Physiology*, vol. 581, no. 3, pp. 893–898, 2008.

[13] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 7263–7271.

[14] D.S. Wooding, "Eye movements of large populations: II. deriving regions of interest, coverage, and similarity using fixation maps," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 518–528, 2002.

[15] Z. Bylinskii, M.A. Borkin, N.W. Kim, H. Pfister, and A. Oliva, "Eye fixation metrics for large scale evaluation and comparison of information visualizations," in *Workshop on Eye Tracking and Visualization*. Springer, 2015, pp. 235–255.

[16] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," in *International Conference on Image Processing*. IEEE, 2013, pp. 171–175.

[17] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[18] H. Duan, G. Zhai, X. Min, Z. Che, Y. Fang, X. Yang, J. Gutirrez, and P. Le Callet, "A dataset of eye movements for the children with autism spectrum disorder," in *Multimedia Systems Conference*. ACM, 2019.

[19] "Treebagger class," https://uk.mathworks.com/help/stats/treebagger-class.html, 2019.

[20] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] C. Heylman, R. Datta, A. Sobrino, S. George, and E. Gratton, "Supervised machine learning for classification of the electrophysiological effects of chronotropic drugs on human induced pluripotent stem cell-derived cardiomyocytes," *PloS one*, vol. 10, no. 12, pp. 1–15, 2015.

[22] K. Chawarska, S. Macari, and F. Shic, "Context modulates attention to social scenes in toddlers with autism," *Journal of Child Psychology and Psychiatry*, vol. 53, no. 8, pp. 903–913, 2012.

[23] N. J. Sasson and E. W. Touchstone, "Visual attention to competing social and object images by preschool children with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 44, no. 3, pp. 584–592, 2014.

[24] S. Wang, J. Xu, M. Jiang, Q. Zhao, R. Hurlemann, and R. Adolphs, "Autism spectrum disorder, but not amygdala lesions, impairs social attention in visual search," *Neuropsychologia*, vol. 63, pp. 259–274, 2014.