ASD Saliency Map Prediction and Discrimination Based on SalGAN

DiXi Yao, ChuMeng Liang, and ZhanDa Zhu

Computer Science, {Jimmyyao18,horstein,dazz993}@sjtu.edu.cn

***Abstract*—This paper proposed a method predicting the saliency map of children with Autism Spectrum Disorder(ASD). Instead of building up a brand new model for prediction, we refer to the saliency prediction method for Typical Developing(TD) children. By fine-tuning Sal-CFS-GAN model with our ASD saliency map dataset, this method gets generalized effectively on predicting ASD saliency map. Based on this method, we use a voting mechanism supported by several correlation metrics to discriminate saliency maps of ASD and TD. By comparing the correlation similarity of predicted saliency of two groups and the given saliency, we can predict whether the ground truth comes from children with ASD.**

***Index Terms*—ASD, Saliency Prediction, Sal-CFS-GAN, Voting Mechanism**

I. Introduction

People with autism spectrum disorder (ASD) perform atypically when it refers to the job of viewing the world. Patients with ASD will pay more attention to idiosyncratic objects or meaningless regions and less focus on social objects than healthy humans do [1]. [2] shows the provident difference between the saliency map of children with ASD and those with TD, introducing the possibility to discriminate ASD/TD with saliency recognition.

However, when we use generalized methods to predict the ASD/TD with saliency maps, the performance is not as good as we expect. Generally speaking, difficulties focus on two aspects. First, fixed points on saliency maps of ASD children are with strongly central bias which seems to be randomly distributed. This impeded the present methods from accurately generating saliency map with correct location of the bias. Second, traditional CNN methods may take the way of over fitting when trying to discriminate saliency maps from ASD and TD, since some of them are really similar. See in 1.

To resolve those two problems, we combine the method proposed by [3] and [4]. We assume that the basic structure of generating saliency maps is generalized enough to apply on both TD and ASD. We focus our jobs in the following two aspects:

1) We build up a ASD saliency prediction model based on a mature saliency prediction method for TD with fine-tuning on ASDSal dataset. Introducing a new loss function, this model is more capable when fitting the central bias of ASD saliency maps.
2) We use a voting mechanism to discriminate the saliency maps from ASD and TD. We firstly predict the TD saliency map with the model proposed above. Statistical correlation metrics are computed to measure the similarity of the predicted sample and two given samples.

Three authors are with the Department of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China

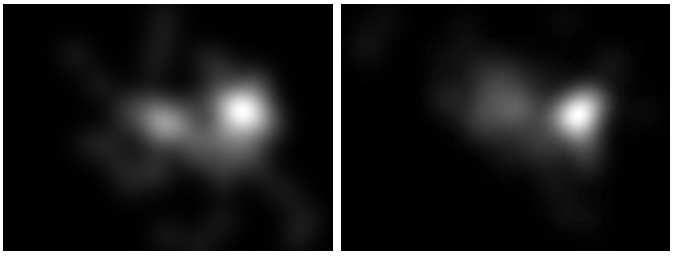

Fig. 1. The left and right show the ground truth saliency maps of ASD and TD. It is hard even for people to tell if either of them comes from ASD or TD.

II. Related Works

A. Saliency Prediction Model

To solve the problem of generating the saliency map of ASD people and classifying ASD people, a commonly adopted method is constructing saliency prediction model. The saliency prediction model are mainly based on visual attention knowledge. There are many ways to construct saliency prediction models.Haung etal. [5] designed SALICON based on adaptive DNN. Cornia etal. [4] proposed ML-Net which using multi-level features to solve the problem. Marcella etal. [6] adopting an RNN based model. SalGAN [7] adopting an adversarial learning path to deal with the problem. The generated model is the base model in traditional salicency prediction method, and with that, we can generate our target, which is the salicency prediction map. Then we use a discriminator to judge between the prediction map and ground truth. With the learning of generative adversarial networks (GAN) method, we can train a perfect generator.

We also adopting a SalGAN based method to solve our problem. Zhai etal. [3] proposed GazeGAN model. They solved the problem of lacking data when generating saliency map. The method of data augmentation transformation (DAT) is proposed. We used the novel GAN model and adpoted the similar DATmethod. Then we transfer the problem of generating generative saliency map into our particular problem, generating the saliency map of ASD people.

B. Dataset

Social difficulties are the hallmark features of Autism Spectrum Disorder (ASD) and can lead to atypical visual attention

Fig. 2. The demonstration of fixation points on a sampled image. The images from left to right are original image, fixation points of ASD people and fixation points od TD people respectively

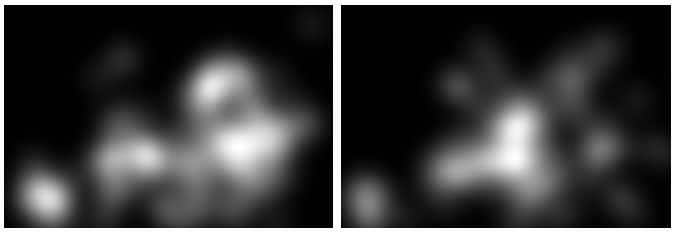

Fig. 3. The demonstration of saliency maps on a sampled image. The images from left to right are fixation points of ASD people and fixation points of TD people respectively

towards stimuli. We use the dataset in the ICME challenge 19 [8], [9]. It consists of 300 natural scene images and the corresponding eye movement data collected from 14 children with ASD and 14 healthy controls. In particular, fixation maps and scan-paths are available in the dataset. Hence, our model is trained with the fixation maps and ground truth is the saliency map generated by scan-paths. During the challenge [10], many methods are proposed to solve the problem of generating ASD children saliency maps based on the dataset. We use the similar evaluating methods and evaluate our method on the dataset.

Before constructing our model, we take some policies to prepossess the dataset. The raw data includes the original images and scan-paths captured by eye-moving equipment. We first generate the fixation points map, which is point out the fixation points in the image. Then we use them to generate the ground truth saliency maps. The dataset also provides the information about the control group, which is the TD group. We use the TD group in our method to further verify the effectiveness and efficiency of our method. The example of fixation points we used are shown in Fig 2. The saliency maps generated of the same image are shown in Fig. 3. We also design and generate the heat maps of visual attention to better understand the patterns of saliency model and visual attention, which can help to better design models, shown in Fig 4. We can observe that the fixation of TD people are more centralized and can lay attention on the main context of the image while ASD people's attention is distracted and fail to focus on particular information on the image. The discrepancy of attention patterns of different kinds of people ensure the feasibility of classifying ASD people with saliency prediction method.

III. Methods

A. Task1

1) Base Model: We use Sal-CSC-GAN model proposed by [3] as our base model for fine-tuning. Sal-CSC-GAN is a complex network architecture specifically designed for

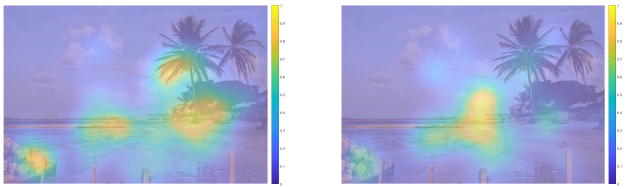

Fig. 4. The demonstration of heat maps on a sampled image. The images from left to right are heat maps of ASD people and fixation points od TD people respectively

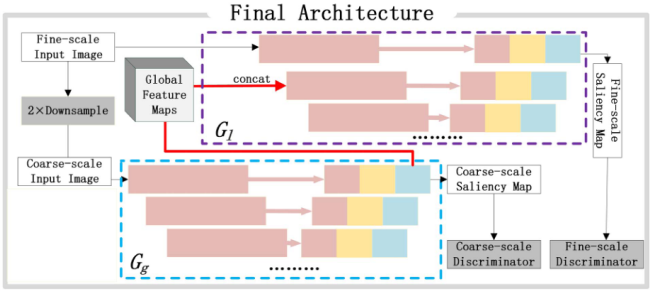

Fig. 5. Model structure

the porpose of saliency map generating based on U-Net and SalGAN. A generator and a discriminator are built up to finish the task of generating saliency maps based on image inputs and discriminating the generated saliency and the ground truth saliency. The generator aims to output saliency maps that are geometrically similar to the ground truth saliency maps and that can confused the discriminator, while the discriminator tries to tell if the given saliency map is a real one or just one generated by the generator. By playing the game mentioned above, the generator can gain a considerable improvement on the performance of saliency generating.

Compared with traditional SalGAN used to generate saliency maps, Sal-CSC-GAN refines mainly in introducing the "Central-Surround" mechanism of human vision to the CNN architecture. To extract high level information, the model not only generate the saliency map in the original scale images, but also generate that in downsample ones. By doing that, the network can sketch the location of hot points approximately. With U-Net applied as a data pre-processing module, this information is extracted as global feature maps and referred when the generator generates the final saliency map.

2) Optimizing Loss function: In the training of generator, we firstly trivial adopt L1Loss between ground truth and prediction saliency map as loss function. While, we found that loss function mat effect the performance of training generator. We choose the loss function similar to [4]:

$$\mathcal{L}(S_i, \hat{S}_i) = \frac{1}{M \times N} \Sigma_{i=1}^{M \times N} \left(\frac{1}{\alpha - S_i} (S_i - \hat{S}_i) + \beta \times R_i \right) \quad (1)$$

The S and $\hat{S}$ represent prediction map and ground truth respectively. The loss function is composed of two parts. The first part is similar to the Mean Squared Error (MSE) with a weight function. The reason is that the loss should give the same importance to high and low ground truth values, even

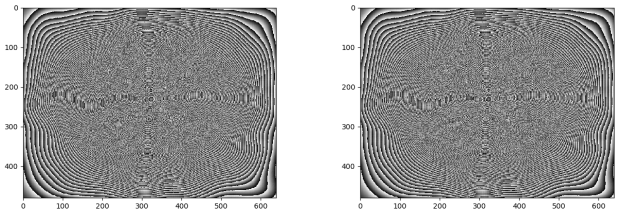

Fig. 6. The average saliency map of ASD people and TD people over all training data

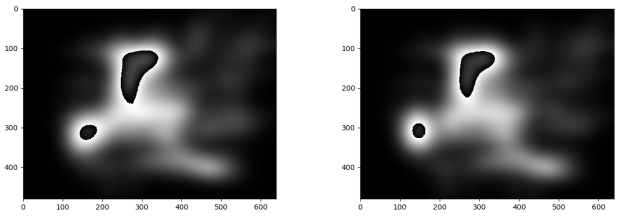

Fig. 7. The average saliency map of ASD people and TD people over all a sampled mini-batch

though the majority of ground truth pixels are close to zero. The second part is more like giving a L2 regulation. For each mini-batch, we calculate the R_i:

$$R_i = (\hat{S}_i - B_i)^2 \tag{2}$$

The bias map B is the average saliency map computed over the ground-truth in a mini-batch. If we use the average map over all training data, there will have few effects. As shown in Fig 6, it will be very hard to tell the difference between ASD people and TD people. However, if we use mini-batch shown in Fig 7, we can lead the model to learn the differences. Hence, we can lead the generator have more intention to form the patterns of ASD people. We simply set α to 1.1 and β to 0.1.

B. Task2

On the basis of task1, we have a well-trained saliency map prediction model. In the task2, two unknown scan-paths information are provided with a given image. We need to judge which scan-path is from the ASD person and which is from the TD person. First, we use the given information to generate the fixation points and saliency maps of two people. The example of saliency maps of two undefined people are shown in Fig 8. We can see some differ a lot while some are similar. Hence, we take a ranking method to solve the problem. Since, we have a well-trained model, these time we use the output of generator in ASD GazeGAN model as the ground truth, and define the prediction map in task1 as the grounf truth. Then we calculate the differences of provided saliency map and ASD prediction saliency map respectively. We use multi-metrics [11] to fairly compare the two maps. We first calculate the metrics for each image, then we compare the metrics one by one. For example, we first compare the AUC score of participant one and participant 2. The one with higher AUC scores 1. Then, we do the same operation on other 3

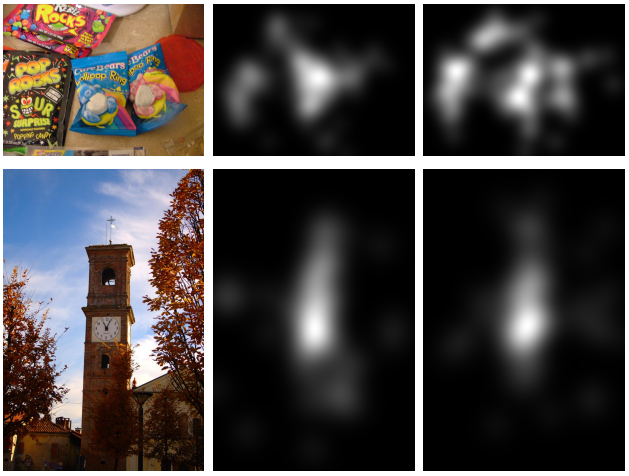

Fig. 8. The demonstration of saliency maps of undefined people. The images in each group from left to right are original images and two saliency maps respectively

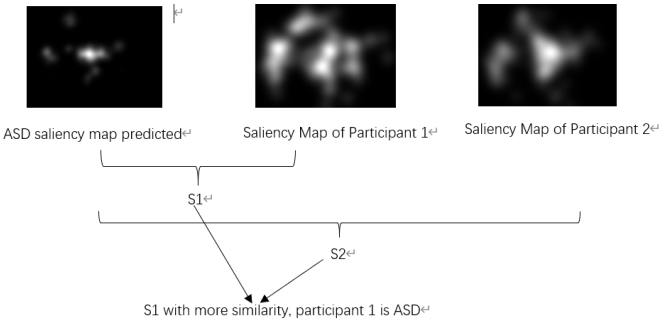

Fig. 9. The Overall Flow Chart of Dealing with Task2

metrics. Adding up all the scores two sample get, the sample with higher score is labeled as TD, since it is more similar to generated saliency map, while the other is labeled as ASD. The overall procedure is shown as Fig 8

IV. Experiments

A. Implementation Details and Experiments Setup

We use python and `Pytorch` framework to implement our GAN model.

B. Task1

1) Training Phase: We use the conventional training method and hyper-parameters in [3] to train our GAN model. We fine-tune the ASD saliency model on the basis of pre-trained model, which has been well trained on TD people. We finetune for 15 epochs. The change of Loss is shown in Fig 10. The loss is composed of three parts, the first part is the loss used to train generator. We use MCELoss as GAN loss and the proposed loss as the VGG loss. The second part is the loss to train the discriminator. Loss of real and fake image are respectively designed. The third part includes some metrics related losses. They are used to better insctruct the training phase. The samples of training result is shown in Fig 11. We can see there is little difference between predicted map and the ground truth, at least it is hard for human eyes.

TABLE I
PERFORMANCE OF SALIENCY GENERATING

	AUC_Borji	AUC_Judd	sAUC	NSS	CC	SIM
Our method						
FSalCSCGAN	0.6458	0.8035	0.5	1.4347	**0.7444**	**0.6765**
Other participants						
SU&UR [12]	**0.786**	**0.818**	-	**1.510**	0.681	0.623
UR&SU [13]	0.785	0.811	-	1.419	0.682	0.631
JUFE [14][15]	0.769	0.79	-	1.245	0.600	0.587
IITJ	0.667	0.683	-	0.656	0.316	0.468

TABLE II
PERFORMANCE OF ASD/TD CLASSIFICATION

	acc	recall	precision	F1_score	Cohen's k	AUC	specificity
Our method							
vanilla voting	**0.733**	0.733	**0.733**	**0.733**	**0.467**	**0.733**	**0.733**
Other participants							
TUM [16]	0.598	0.717	0.574	0.632	0.201	0.644	0.484
R3U [17]	0.593	0.684	0.570	0.616	0.189	0.595	0.506
UM [18][19]	0.557	**0.877**	0.532	0.658	0.127	0.564	0.251
	0.579	0.592	0.563	0.570	0.158	0.579	0.566
	0.574	0.594	0.568	0.568	0.149	0.575	0.556
UCD&UK [20][21]	0.551	0.635	0.527	0.546	0.106	0.613	0.471
	0.542	0.741	0.522	0.610	0.091	0.575	0.351
	0.539	0.807	0.519	0.629	0.089	0.544	0.282
ECNU [22]	0.516	0.705	0.504	0.585	0.041	0.521	0.337
	0.446	0.397	0.429	0.412	-0.110	0.445	0.493
	0.420	0.442	0.413	0.427	-0.159	0.421	0.399

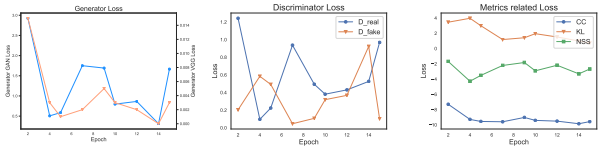

Fig. 10. The change of loss during the fine-tuning phase. The graphs from left to right show the loss of the generator, the loss of the discriminator and the metrics related loss respectively.

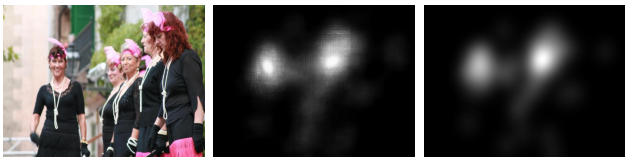

Fig. 11. The sample of our trainig result. The images from left to right is the original image, predicited saliency map and ground truth respectively

2) Evaluation Metrics: For task1, we adopt some metrics provided in [11] and we use `MATLAB` to implement these metrics. We use different kinds of metrics to fairly evaluate our methods. The metrics we use including the followings. Before calculating, we firstly turn our prediction map into binary images.

1) We use the generated saliency prediction map and fixation maps to calculate the score.
   - Cross-correlation metric (CC)
   - Similarity metric (SIM)
2) We use the generated saliency prediction map and fixation points to calculate the score.
   - AUC_Borji_score (AUC_Borji)
   - AUC_Judd_score (AUC_Judd)
   - shuffled AUC metric (sAUC)
   - Normalized Scan-path Saliency metric (NSS)

3) Dataset: We use ASDSal300 dataset to train and evaluate our model. The first 270 samples are used for fine-tuning the base model, making it capable on generating ASD saliency maps. The rest 30 samples are took as testset to evaluate the performance of our model.

4) Evaluation on Dataset: By applying our model on the testset of ASDSal300 dataset, we get 30 saliency maps. Using those generated saliency and ground truth provided by the dataset, we computed the average values of metrics mentioned above. We compare the performance of our model and the performance of submitted work in ICME2019 game. See in Table I. Our method performs better providently in 5 of 6 metrics compared to almost all submitted works, leading to the conclusion its superiority.

C. Task2

1) Evaluation Metrics: Four metrics mentioned in Task 1 are picked up as voting benchmark of one voter separately in Task 2:

- AUC_Borji_score (AUC_Borji)
- shuffled AUC metric (sAUC)
- Cross-correlation metric (CC)
- Normalized Scan-path Saliency metric (NSS)

For the voting model, each voter would score one for one sample in a sample pair if it gains a higher score in the related metric than the other. The final score is computed by adding up all four scores one sample gains.

We use various metrics to do fair comparison. The metrics we use include: accuracy, recall score, precision scroe, specificity, Area Under Cureve (AUC), F1_score, Cohen's Kappa coefficience (Cohen's k). Since our method is based on voting mechanism, the output of classifying is either 0 or 1.

2) Evaluation on Dataset: For vanilla voting model, we directly compute the correlation metrics mentioned above between two input samples and the generated TD saliency map with the generator in Task 1. We determine the input sample with higher metric score in all as TD and the other as ASD. We compute the classification accuracy among 30 samples. We compare the performance of our model and the performance of submitted work in ICME2019 game. See in Table II.Our method does better than all of those submitted works.

V. Discussion

Our discussion focuses on evaluating the performance of our classification model:

Compared with other proposed work, we give up the method of DNN when it refers to the classification task. Since we have already got a saliency generating model in saliency generating task, we can transform the classification problem as one of comparing the similarity of saliency maps. We can directly determine TD as one sample's label if it comes as the more similar one in the sample pair to our generating TD saliency. This method is more straight-forward than DNN, leading to a better effect consequently.

-

VI. Conclusion

In this paper, we introduce a TD saliency generating model to resolve the problem of ASD saliency generating. We fine-tune the base model with a new term of loss on ASDSal300 dataset. Our model's performance is quite near to the first one of ICME2019 game. Based on this model, we then propose a brand new method to discriminate the saliency of ASD and TD, which leads to performance superior than all participants of ICME2019 game in almost all benchmark.

Acknowledgement

We give our sincerest appreciation to Prof.Zhai and Dr.Sun for giving us this chance to contribute our efforts in the work of ASD saliency prediction and discrimination. We also thank them for their generous help when we met difficulties.

References

[1] S. J. W. G. Dawson and J. M. Partland, "Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies," 2005.

[2] X. M. D. E. A. L. D. P. K. R. A. S. Wang, M. Jiang and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking," 2015.

[3] Z. Che, A. Borji, G. Zhai, X. Min, G. Guo, and P. L. Callet, "How is gaze influenced by image transformations? dataset and model," 2019.

[4] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "A deep multi-level network for saliency prediction," 2016.

[5] X. Huang, C. Shen, X. Boix, and Q. Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[6] C. Marcella, B. Lorenzo, S. Giuseppe, and C. Rita, "Predicting human eye fixations via an lstm-based saliency attentive model," *IEEE Transactions on Image Processing*, vol. 27, pp. 5142–5154, 2016.

[7] J. Pan, C. Canton, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, and X. a. Giro-i Nieto, "Salgan: Visual saliency prediction with generative adversarial networks," in *arXiv*, January 2017.

[8] H. Duan, G. Zhai, X. Min, Z. Che, and P. L. Callet, "A dataset of eye movements for the children with autism spectrum disorder," in *the 10th ACM Multimedia Systems Conference*, 2019.

[9] H. Duan, G. Zhai, X. Min, Y. Fang, Z. Che, X. Yang, C. Zhi, H. Yang, and N. Liu, "Learning to predict where the children with asd look," 2018, pp. 704–708.

[10] J. G. a, Z. C. B, G. Z. B, and P. L. C. C, "Saliency4asd: Challenge, dataset and tools for visual attention modeling for autism spectrum disorder," *Signal Processing: Image Communication*, 2020.

[11] M. Kümmerer, Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit/tübingen saliency benchmark," https://saliency.tuebingen.ai/.

[12] L. H. A. N. W. Wei, Z. Liu and O. L. Meur, "Saliency prediction via multi-level features and deep supervision for children with autism spectrum disorder," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[13] Z. L. L. H. W A. Nebout, W. Wei and O. L. Meur, "Predicting saliency maps for asd people," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[14] B. W. Y. Fang, H. Huang and Y. Zuo, "Visual attention modeling for autism spectrum disorder," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[15] Y. Z. W. J. H. H. Y. Fang, H. Zhang and J. Yan, "Visual attention prediction for autism spectrum disorder with hierarchical semantic fusion," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[16] M. Startsev and M. Dorr, "Classifying autism spectrum disorder based on 695 scanpaths and saliency," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[17] F. B. G. Arru, P. Mazumdar, "Exploiting visual behaviour for autism spectrum disorder identification," *International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[18] M.-L. S. Y. Tao, "Sp-asdnet: Cnn-lstm based asd classification model using observer scanpaths," *IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[19] F. B. G. Arru, P. Mazumdar, "Early detection of children with autism spectrum disorder based on visual exploration of images," *Signal Processing: Image Communication*, 2019.

[20] S.-C. S. C. C.-N. C. S. O. C. Wu, S. Liaqat, "Predicting autism diagnosis using image with fixations and synthetic saccade patterns," *IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China*, 2019.

[21] P. R. D. S. S. C. C. N. C. S. Liaqat, C. Wu and S. Ozonoff, "Predicting asd diagnosis in children with synthetic and image-based eye gaze data," *Signal Processing: Image Communication*, 2019.

[22] J. Y. S. Xu and M. Hu, "A new bio-inspired metric based on eye movement data for classifying asd and typically developing children," *Signal Processing: Image Communication*, 2019.