Uncertainty Inspired Autism Spectrum Disorder Screening

Ying Zhang, Yaping Huang(✉), Jiansong Qi, Sihui Zhang, Mei Tian, and Yi Tian

Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, China
yphuang@bjtu.edu.cn

Abstract. People with autism spectrum disorder (ASD) show distinguishing preferences for specific visual stimuli compared to typically developed (TD) individuals, opening the door for objective and quantitative screening by eye-tracking data analysis. However, existing eye-tracking-based ASD screening approaches often assume that there are no individual differences and that all stimuli contribute equally to the prediction of an ASD. Consequently, a fixed number of images are usually selected by a pre-defined strategy for further training and testing, ignoring the distinct characteristics of various subjects viewing the same image. To address the aforementioned difficulties, we propose a novel Uncertainty-inspired ASD Screening Network (UASN) that dynamically modifies the contribution of each stimulus viewed by different subjects. Specifically, we estimate the uncertainty of each stimulus by considering the variation between the subject's fixation map and the ones of the two clinical groups (i.e., ASD and TD) and further utilize it for weighting the training loss. Besides, to reduce the diagnosis time, instead of the shuffle-appeared mode of image viewing, we propose an uncertainty-based personalized diagnosis method to dynamically rank the viewing images according to the preferences of different subjects, which can achieve high prediction accuracy with only a small set of images. Experiments demonstrate the superior performance of our proposed UASN.

Keywords: ASD Screening · Eye-tracking · Data Uncertainty

1 Introduction

Autism Spectrum Disorder (ASD) has been a prevalent neurodevelopmental disorder worldwide that one in 44 kids aged 8 years in the United States suffers from it as reported in 2021 [15] and there is still a steady and substantial growth in the population.

However, diagnosing ASD relies on subjective evaluations that are expensive and clinically demanding.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_39.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 399–408, 2023.
https://doi.org/10.1007/978-3-031-43904-9_39

Seminal works [3,6,12,16,17,19] have pointed out that eye movement patterns of people with ASD play an irreplaceable vital role in identifying ASD.

Early efforts [7,13,18,20,21] usually focus on low-level behavior features combined with machine learning algorithms to identify autism, while in recent years, the eye-tracking data driven method [2,10,14] boosts the performance of ASD screening by utilizing deep neural networks (DNNs) which extract high-level semantic information of eye movement.

However, existing deep-learning-based methods usually define an image-ranking strategy as a pre-processing step to select a certain number of images. During training, each visual stimulus is treated equally, ignoring the distinct contributions of different stimuli. Besides, during the diagnosis procedure, a fixed number of images are shown to a subject which takes a relatively long time, thereby leading to poor cooperation of subjects, especially little kids.

To tackle the above issues, in this paper, we propose a novel Uncertainty-inspired ASD Screening Network, named UASN, to distinguish the importance of each visual stimulus for different individuals. Despite the success of uncertainty in computer vision [1,5,22,23], to our best knowledge, this is the first attempt to introduce uncertainty estimation into ASD screening. Our uncertainty-inspired UASN can enforce the model learning from more distinctive gaze patterns during training. Meanwhile, when the model is deployed in the real clinical scenario, we further design an efficient personalized diagnosis strategy, which can dramatically reduce the diagnosis time without a performance drop.

Specifically, the uncertainty in UASN works in two ways to ensure both higher accuracy and lower time consumption. On the one hand, given an input gaze pattern, we estimate the uncertainty by comparing the difference between the fixation map and the ones of ASD and TD groups. The uncertainty will be assigned a lower value for a larger disparity, suggesting the importance of the given gaze pattern for identifying a certain individual. Subsequently, guided by the estimated uncertainty, we design a truncated weighting loss to select the most distinctive gaze patterns and further dynamically adjust the contributions made by different stimuli, resulting in a more efficient classification. On the other hand, how to reduce the diagnosis time is also a key factor in real clinical applications, especially for preschool children. To achieve this goal, we propose a personalized diagnosis method that ranks the stimuli according to the estimated uncertainty. Instead of the random shuffle mode for image viewing, we recommend a top similar or dissimilar stimulus for the next viewing according to the decision of the previous gaze patterns. Following the proposed protocol, our method achieves state-of-the-art performance while spending much less diagnosis time.

In general, our contributions can be summarized as follows: 1) we propose the first usage of the Uncertainty-inspired ASD Screening Network, named UASN, for identifying ASD people; 2) we estimate the uncertainty of each gaze pattern and further design a truncated weighting loss, which can enforce the model to dynamically adjust the contributions of different gaze patterns during training; 3) we design a personalized online diagnosis protocol that can dramatically reduce the diagnosis time without losing accuracy; 4) we conduct comprehensive

experiments on the Saliency4ASD benchmark and achieve state-of-the-art performance only using 1/5 visual stimuli compared with other leading approaches.

2 Uncertainty Inspired ASD Screening

Our UASN is built upon traditional DNNs and consists of two novel stages: 1) uncertainty-guided training, and 2) uncertainty-guided personalized diagnosis. During training, we estimate an uncertainty value for each gaze pattern and further apply it for weighting the training loss. Besides, for a more simplified diagnosis procedure, we design an uncertainty-based strategy that adaptively selects the most discriminative images based on the subject's gaze behaviors.

2.1 Uncertainty Guided Training

Figure 1 illustrates the detailed training process of UASN. Firstly, we extract the features of gaze patterns by taking the temporal eye tracking information as input and resulting in the classification prediction. Then, we design an uncertainty estimation module to compute the uncertainty values of each subject on all the visual stimuli. Moreover, we further apply the estimated uncertainty to weight the training samples by a truncated loss in a reasonable manner.

Gaze Pattern Feature Extraction. Formally, by collecting a group of ASD and TD subjects $S = \{s_i\}_{i=1}^{M}$'s eye movement data on a set of images $X = \{x_j\}_{j=1}^{N}$, we get the corresponding scanpaths which comprise each fixation point's position and duration in the temporal order. The labels of the two clinical groups are denoted as $Y = \{y_i\}_{i=1}^{M} \in \{0, 1\}$.

To generate the discriminative features of the given gaze pattern (i-th subject watching j-th image), we first feed the image x_j into a ResNet-50 [9] with the last max pooling layer removed to learn a 2048-dimension visual feature. Then the visual feature sequence taken from the scanpath of i-th subject is fed into a Long Short Term Memory (LSTM) network [8]. A final hidden state is obtained at the end of the sequence which is then fed into a fully connected (FC) layer followed by a sigmoid function to get the prediction result for the i-th subject viewing j-th image, which is denoted as $\hat{y}_i^j$. The network can be optimized by the binary cross-entropy (BCE) loss:

$$\mathcal{L}_{\text{bce}} = -\frac{1}{NM} \sum_i \sum_j (y_i^j \log \hat{y}_i^j + (1 - y_i^j) \log(1 - \hat{y}_i^j)), \tag{1}$$

where $\hat{y}_i^j$ and y_i^j denote the predicted and ground truth labels of i-th subject respectively. If belonging to ASD, $y_i^j = 1$ for all images $\{x_j\}_{j=1}^{N}$, otherwise 0.

Uncertainty Estimation. We believe that different images contribute unequally to a subject's final classification due to the subject's unique preferences for viewing images. As a result, we estimate an uncertainty value for each gaze pattern based on the variance between its fixation map and the ones

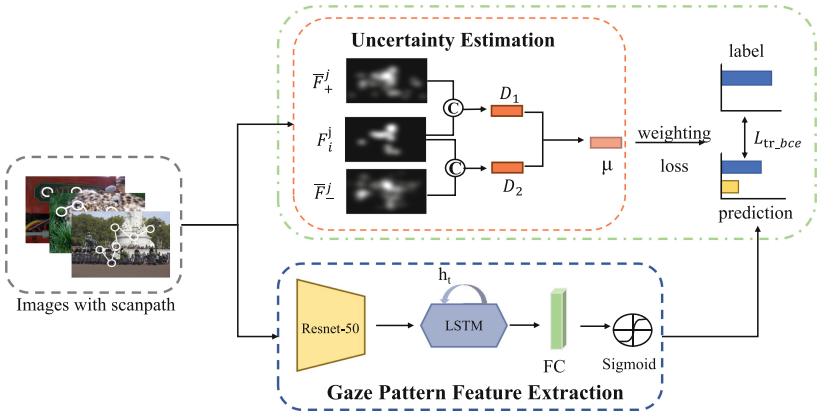

Fig. 1. Illustration of the uncertainty-guided training module of UASN. Images with corresponding gaze patterns are fed into the Gaze Pattern Feature Extraction module to get the prediction result. By applying the Uncertainty Estimation module, we compute an uncertainty for each gaze pattern and further use it to weight the training BCE loss to get a more reasonable as well as efficient result.

of the two clinical groups (*i.e.*, ASD and TD). For instance, an ASD's fixation map on a discriminative image should appear more similar to the ASD group's averaged one than the TD group's so the variance is supposed to be large.

We first generate the fixation map for each gaze pattern according to the fixation data. Then, given two groups' fixation maps on each image in the dataset, for each subject, we apply cosine similarity to compute an uncertainty measurement on each image. Let F_i^j denote the fixation map of the i-th subject's fixation map on the j-th image, and $\bar{F}_+^j$, $\bar{F}_-^j$ denote the fixation maps of ASD and TD group for j-th image respectively. The uncertainty can be written as:

$$D_i^j = \left| C(F_i^j, \bar{F}_+^j) - C(F_i^j, \bar{F}_-^j) \right|, \tag{2}$$

$$\mu_i^j = 1 - D_i^j, \tag{3}$$

where C is the cosine similarity function, D_i^j is the distinguishability of the j-th image when viewed by the i-th subject and μ_i^j denotes the uncertainty. Specifically, for images that do not contain certain subjects' eye-tracking data, we reasonably set the μ_i^j to be large because we assume that the absence of eye-tracking data is due to a lack of interest, and further signals ineffectiveness.

Truncated Weighting Loss. Upon obtaining the uncertainty value μ_i^j, we can utilize the uncertainty to re-weight the training loss by teaching the model which images to trust and which to discredit. We hope that the larger the μ is, the less the image contributes to the final classification, so the corresponding loss needs to be reduced correspondingly.

However, for some gaze patterns that are confusing and much more difficult to distinguish between ASD and healthy people, it is more suitable to discard those

unreliable patterns. Considering this, we finally propose a truncated weighting BCE loss for training. Specifically, when the estimated uncertainty value is larger than a pre-defined threshold t, we set the corresponding loss to zero. In summary, The final loss function is denoted as:

$$\mathcal{L}_{\mathrm{tr\_bce}} = \sum_i \sum_j \mathbb{I}_{[\mu \leq t]}(1 - \mu_i^j)\mathcal{L}_{\mathrm{bce}}, \tag{4}$$

where $\mathbb{I}$ is the indicator function and t is the threshold. Only when the condition of $[\mu \leq t]$ is met, is the value of the indicator function set to 1, and the $\mathcal{L}_{\mathrm{tr\_bce}}$ remains. Specifically, for a more reasonable computation, we do not simply set the t to be a fixed value. Instead, we determine t with an adaptive technique. For each subject, we sort the uncertainty values from low to high and choose the 1/3 of the images with the lowest uncertainty and retain the contributions they make to the prediction while discarding the remainder.

2.2 Uncertainty Guided Personalized Diagnosis

On the basis of training our model in an uncertainty-guided way, we are encouraged to go deeper to simplify the diagnostic procedure. To this end, we incorporate personalized diagnosis into our work, taking into account the gaze behavior features of each subject. Figure 2 presents the workflow of our personalized diagnosis protocol after completing the training process in Fig. 1. Specifically, by extracting features of images and forming a feature bank, we selectively choose the most suitable images to update the viewing list according to the subject's viewing pattern.

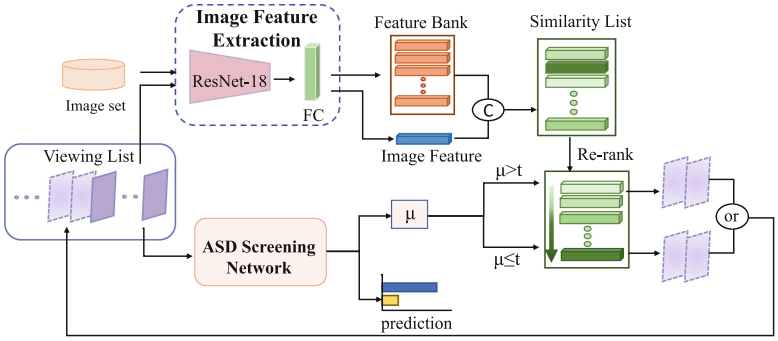

Fig. 2. Illustration of our proposed uncertainty-guided personalized diagnosis module. We feed the gaze patterns in the current viewing list into the ASD screening network to get a prediction as well as the uncertainty value μ. The μ is further used to refresh the viewing list with images based on the similarity-based ranking result.

Image Feature Extraction. First, we assume that the visual similarity between images brings the potential of personalized diagnosis that similar images

may contribute similarly in distinguishing an individual. From this point, we extract the feature of all the 300 images in the dataset (the Image set in Fig. 2) using a ResNet-18 [9] followed by an FC to get a 128-dimension feature, thereby forming a feature bank getting prepared for the further dynamical ranking procedure.

Similarity-Based Image Ranking. Then, we build a viewing list simulating the diagnosis procedure where images are shown to a subject one by one and the list is updated in real-time. In the beginning, we select an image from the image set randomly to initialize the viewing list. When a trial is completed, the viewing list is subsequently updated. In each trial, we generate an average feature of all images in the viewing list. We then compute the cosine similarity between the feature of the current image and images in the feature bank to obtain a similarity list. We sort the list in similarity-ascending order.

Uncertainty-Based Viewing List Updating. To determine which images should be included in the viewing list, a method based on uncertainty is developed. First, we feed the image in the list and the corresponding eye-tracking data into the ASD screening network which is composed of the Uncertainty Estimation module and the Gaze Pattern Feature Extraction module in Fig. 1 to get the uncertainty and the prediction result. By pre-defining a threshold p, we separate the scenario into a positive case and a negative one. When the average uncertainty value is larger than p, we consider it negative so we select the top K dissimilar images to join the viewing list and vice versa. After T trials, we achieve a relatively confident prediction.

3 Experiments

3.1 Dataset and Experimental Settings

Dataset. So far, only Saliency4ASD [4] is publicly released for the evaluation of ASD screening. It consists of 300 images from a public dataset collected by Judd *et al.* [11] and the eye movement data collected from 14 kids with ASD and 14 with TD. For each image, the scanpath of each subject is provided, allowing us to derive the single fixation map for additional uncertainty computation.

Evaluation Protocol. We employ the leave-one-subject-out cross-validation method for evaluating the model's performance. Specifically, for the Saliency4ASD dataset, we perform a 28-round validation with each round selecting only one subject for testing while the remaining 27 subjects work as the training data.

Metrics. We follow the previous works [2,10] to adopt the accuracy, sensitivity, specificity, and AUC to evaluate the performance of the prediction for each subject. Besides, we assess the performance of the prediction for each scanpath to generate a more strict measurement. We still adopt accuracy, sensitivity, and specificity, called Acc_I, Sen_I, and Spe_I. Accordingly, for the previously used three metrics, we denote them as Acc_S, Sen_S, and Spe_S.

Implementation Details. The experimental setting mostly follows the [2] during training. Besides, we manually set the uncertainty values of those images without some certain subjects' eye movement data to be $1–10^{-5}$ which is a relatively large margin that rarely contributes to the classification. For personalized diagnosis, we initialize the viewing list by randomly selecting an image. By considering the gaze pattern with an uncertainty level lower than $p = 0.9$ a positive case, we choose the top K most similar images to update the viewing list. We set K to be 1 and update the viewing list $T = 20$ trials in total.

3.2 Comparison with State-of-the-Art

We conduct extensive experiments to explore whether our proposed model outperforms the state-of-the-art. We compare our UASN with [2] which selects a fixed 100 images subset out of a total of 300 images according to a Fisher-Score-based image selection strategy for training and testing. We design the following three UASN variants: 1) **UASN-noUnLoss** only uses uncertainty to choose the top most discriminative images and removes the weighting loss procedure for training; 2) **UASN-Fixed** selects a 100-image subset with the lowest uncertainty for each subject during training and testing without dynamically adjusting the diagnosis process; 3) **UASN-Dynamic** uses 100 images with low uncertainty for training. In the diagnosis process, we adopt the proposed uncertainty-guided personalized diagnosis to recommend a small number of images, which significantly reduces the diagnosis time while maintaining high accuracy.

Table 1. Comparison on both the subject and scanpath level with state-of-the-art [2].

Method	Subject level metrics				Scanpath level metrics		
	Acc_S	Sen_S	Spe_S	AUC	Acc_I	Sen_I	Spe_I
[2]	0.93	0.93	0.93	0.98	0.59	0.58	0.59
UASN-noUnLoss	**1**	**1**	**1**	**1**	0.76	0.75	0.72
UASN-Fixed	**1**	**1**	**1**	**1**	0.74	0.71	0.72
UASN-Dynamic	**1**	**1**	**1**	**1**	**0.79**	**0.78**	**0.78**

Table 1 shows the results. We can see that our model's three variants all outperform the baseline model by all evaluation metrics. The result demonstrates that introducing uncertainty during training can reach 100% accuracy. When removing the uncertainty weighting loss part, the UASN-noUnLoss model's performance drops slightly in scanpath level metrics than the best UASN-Dynamic, *i.e.*, 3% (Acc_I), 3% (Sen_I) and 6% (Spe_I). Besides, the personalized diagnosis strategy achieves the same accuracy but largely reduces the number of images to 1/5 (20 images), which decreases the diagnosis time dramatically. In terms of scanpath level metrics, our UASN-Dynamic outperforms the previous leading model [2] by 20% in Acc_I and Sen_I, and 19% in Spe_I.

Table 2. Comparison of using images of different uncertainty levels. The "top 100", "middle 100" and "bottom 100" denote the three subsets of 100-image with the lowest, the middle, and the highest uncertainty level.

Method	Selected ImageSet	Acc_S	Sen_S	Spe_S	AUC	Acc_I	Sen_I	Spe_I
UASN-Fixed	top-100	**1**	**1**	**1**	**1**	0.74	0.71	0.72
	middle-100	0.78	0.54	0.93	0.86	0.56	0.50	0.56
	bottom-100	0.54	0.25	0.83	0.59	0.50	0.46	0.51
UASN-Dynamic	**top-100**	**1**	**1**	**1**	**1**	**0.79**	**0.78**	**0.78**
	middle-100	0.64	0.64	0.64	0.68	0.52	0.50	0.58
	bottom-100	0.38	0.33	0.742	0.39	0.47	0.45	0.50

The results suggest introducing uncertainty both in the training and testing stage achieves the best performance with a quite small image subset at classifying the two clinical groups.

3.3 Ablation Study

Effect of the Uncertainty Estimation. To verify the effect of uncertainty estimation, we divide the 300 images into three non-overlapping subsets based on the ascending order of uncertainty level, denoted as top-100, middle-100, and bottom-100 subsets. Table 2 shows the results. It is not surprising that UASN-Dynamic with top-100 achieves the best performance and the large performance drop on both UASN-Fixed and UASN-Dynamic approves our model's strong capability of selecting the most discriminative images for classifying ASD and TD. We further visualize the relation between Acc_I and uncertainty, as well as give some samples of ASD and TD's fixation maps with different uncertainty values to support the effectiveness of our method. Details are given in Fig. I and Fig. II of the supplementary material.

Effect of Different Inference Strategies. During the diagnosis process, we need to define the number of trials (T) and the number of images (K) selected to append to the viewing list. The results are given in Table 3. We tried various permutations of T and K and finally found that appending one image each trial

Table 3. Comparison of how different inference strategies influence our model's performance. T denotes the number of inference trials during the diagnosis and K denotes the number of images inferred in each trial.

T	K	Acc_S	Sen_S	Spe_S	AUC	Acc_I	Sen_I	Spe_I
20	1	**1**	**1**	**1**	**1**	**0.79**	**0.78**	**0.78**
10	2	0.96	0.93	**1**	0.99	0.75	0.74	0.77
10	1	0.96	0.93	1	1	0.73	0.71	0.74

and performing 20 trials for one subject obtained the best performance while cost the least diagnosing time.

Effect of Different Similarity Measurements and Backbones. We conduct experiments on replacing the similarity measurement method and the model's backbone. Results can be referred to in the Table I and Table II of supplementary.

Computational Cost. Another crucial metric for assessing our model's efficiency is the time interval between every two viewing list update rounds during the personalized diagnosis process. Upon conducting experiments, we get an average of 0.038 s for each interval, demonstrating that our dynamic update strategy causes no delay in the clinical diagnosis.

4 Conclusion

In this paper, we present UASN, a novel ASD screening approach, inspired by uncertainty. The uncertainty benefits the ASD diagnosis in two ways: a weighted truncated training loss that enables the model to learn the most discriminative and effective features of gaze patterns and a personalized procedure that dynamically ranks the stimuli according to the subject's gaze behaviors. Comprehensive results show superior performance in classifying ASD people.

Acknowledgements. This work was supported by the Beijing Natural Science Foundation (M22022, L211015) and the National Natural Science Foundation of China (62271042, 61906013).

References

1. Chang, J., Lan, Z., Cheng, C., Wei, Y.: Data uncertainty learning in face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020
2. Chen, S., Zhao, Q.: Attention-based autism spectrum disorder screening with privileged modality. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1181–1190 (2019)
3. Corbetta, M., Shulman, G.L.: Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. **3**(3), 201–215 (2002)
4. Duan, H., et al.: A dataset of eye movements for the children with autism spectrum disorder. In: Proceedings of the 10th ACM Multimedia Systems Conference, pp. 255–260 (2019)
5. El Ghaoui, L., Lanckriet, G.R.G., Natsoulis, G., et al.: Robust classification with interval data (2003)
6. Frazier, T.W., et al.: A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. J. Am. Acad. Child Adolesc. Psychiatry **56**(7), 546–555 (2017)
7. Freeth, M., Chapman, P., Ropar, D., Mitchell, P.: Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with ASD? Evidence from eye-tracking. J. Autism Dev. Disord. **40**(5), 534–547 (2010)

8. Graves, A.: Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850 (2013)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Jiang, M., Zhao, Q.: Learning visual attention to identify people with autism spectrum disorder. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3267–3276 (2017)
11. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2106–2113 (2009). https://doi.org/10.1109/ICCV.2009.5459462
12. Klin, A., Lin, D.J., Gorrindo, P., Ramsay, G., Jones, W.: Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature **459**(7244), 257–261 (2009)
13. Li, B., Sharma, A., Meng, J., Purushwalkam, S., Gowen, E.: Applying machine learning to identify autistic adults using imitation: an exploratory study. PLoS ONE **12**(8), e0182652 (2017)
14. Liu, W., Li, M., Yi, L.: Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. Autism Res. **9**(8), 888–898 (2016)
15. Maenner, M.J., et al.: Prevalence and characteristics of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2018. MMWR Surveill. Summ. **70**(11), 1 (2021)
16. McPartland, J.C., Webb, S.J., Keehn, B., Dawson, G.: Patterns of visual attention to faces and objects in autism spectrum disorder. J. Autism Dev. Disord. **41**(2), 148–157 (2011)
17. Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., Piven, J.: Visual scanning of faces in autism. J. Autism Dev. Disord. **32**(4), 249–261 (2002)
18. Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C.C., Malige, A.: Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. Biol. Psychiat. **79**(8), 657–666 (2016)
19. Sasson, N.J., Turner-Brown, L.M., Holtzclaw, T.N., Lam, K.S., Bodfish, J.W.: Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. Autism Res. **1**(1), 31–42 (2008)
20. Thabtah, F.: Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. In: Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp. 1–6 (2017)
21. Wang, S., et al.: Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron **88**(3), 604–616 (2015)
22. Wang, Z., Li, Y., Guo, Y., Fang, L., Wang, S.: Data-uncertainty guided multi-phase learning for semi-supervised object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4568–4577 (2021)
23. Xu, Y., et al.: Data uncertainty in face recognition. IEEE Trans. Cybern. **44**(10), 1950–1961 (2014)