Prediction of gaze direction using Convolutional Neural Networks for Autism diagnosis

Dennis Núñez-Fernández[1], **Franklin Porras-Barrientos**[1]
Macarena Vittet-Mondoñedo[1], **Robert H. Gilman**[2], **Mirko Zimic**[1]
[1]Laboratorio de Bioinformática y Biología Molecular, Universidad Peruana Cayetano Heredia, Peru
[2]Department of International Health, Johns Hopkins University, USA
`{dennis.nunez, franklin.barrientos.p, macarena.vittet.m}@upch.pe`
`rgilman1@jhmi.edu, mirko.zimic@upch.pe`

Abstract

Autism is a developmental disorder that affects social interaction and communication of children. The gold standard diagnostic tools are very difficult to use and time consuming. However, diagnostic could be deduced from child gaze preferences by looking a video with social and abstract scenes. In this work, we propose an algorithm based on convolutional neural networks to predict gaze direction for a fast and effective autism diagnosis. Early results show that our algorithm achieves real-time response and robust high accuracy for prediction of gaze direction.

1 Introduction

Around 1 in 160 children worldwide is affected by Autism spectrum disorder (ASD). It generates a deficit social interaction [2] and result in delay in cognitive development [6]. Recent studies have shown that early intervention for children with ASD is effective in improving quality of life, every dollar spent on early intervention helps to save eight dollars in special education [4, 5]. The reasons to the low utilization of the gold standard diagnostic tools are the duration of the tests and the extensive training for the technician [1], and developing countries have very few of them. Recent studies have shown strong evidence for utilizing gaze direction as an early biomarker of ASD [8, 9, 12, 13]. Indeed, children with ASD show a preference for geometric scenes rather than social scenes [9, 12].

In recent years, several approaches to gaze direction recognition were proposed and some open source eye-tracking algorithms are currently available; however, these algorithms demand extensive calibration, several settings and training processes that are not appropriate for young children. For instance, most of current gaze direction systems for the ASD diagnosis need to be evaluated under controlled environments by using expensive devices that require holding the head to avoid undesirable movements [9, 12]. None of such systems are appropriate in children due its restless behavior.

In a more recent work [11], eye movements are used to asses ASD. Based on the gaze patterns and using K-means clustering and a support vector machine classifier, they are able to identify children with ASD with an accuracy of 88.51%. Nonetheless, since this process involves a high-acuity eye tracker, it is not a scalable screening process. In [3], a CNN-based approach is used to predict gaze in a natural social interaction and assess ASD in children. They developed two CNNs, one for face detection and another for gaze prediction. Despite gaze direction is precisely predicted by the CNNs, wearing glasses involves equipment, which disturb child's attention.In addition, to date there are several popular open source tools for eye tracking. One of the most popular is `https://github.com/pupil-labs/pupil`, which provides an accurate tool for eye tracking, however, this system makes use of glasses. As explained above, external hardware is not suitable for children. Another popular tool for eye tracking is `http://www.pygaze.org`, however, calibration is difficult and eye region should be in a fixed position, which is difficult to set in children.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

2 Methodology

Our proposed system recognizes gaze direction based on images obtained from a video sequence. Face and eye detection are performed by cascade classifiers using LBP and Haar features that were found with the Viola-Jones mechanism [14]. Face detection uses LBP cascades over the whole image due its better performance, and eye detection employes Haar cascades into the facial region since they are much smaller than the full image. Later, we generate a single square image with both eye regions. Finally, the CNN classifies it into right, left or vague direction, see Fig. 1.

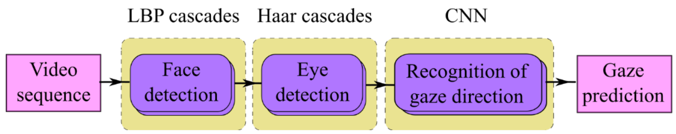

Figure 1: Diagram for the proposed system

The dataset was collected in our research facilities, the Laboratory of Bioinformatics and Molecular Biology, Universidad Peruana Cayetano Heredia, Peru. We enroled 30 adults between 22-35 years old, working in our laboratory. The videos were recorded under a controlled environment and using a standard web camera. The eye glace directons were three: right, left and vague. After frame extraction, we obtained a total of 420 images, which increased to 66,750 after data augmentation.

The proposed CNN is a variation of the LeNet model [10], with 60K learnable parameters. The training of the proposed CNN has been carried out on the 80% of the collected dataset (53,400 images) and testing on the remaining 20% (13,350 images). The CNN input are 72x72 pixel binary images, following the architecture: C(5x5)-S(2x2)-C(5x5)-S(2x2)-FC(120)-FC(3), where C: Conv. layer, S: Sub sampling, FC: Full connection. We employed Caffe framework [7].

3 Early Results

For all shuffle testing on adult dataset, we obtain 96.01% of accuracy. However, for a rigorous testing, we evaluated our model using a 5-fold cross-validation and using different groups of people who do not appear in the training dataset. For testing on the dataset using 3 classes and employing 5-fold cross-validation, we obtained an average accuracy of 89.54%. Tests were conducted with people who do not appear in the training dataset.

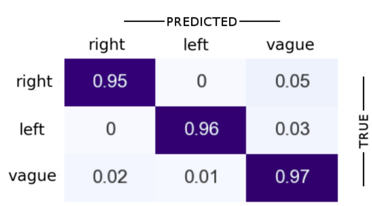

Figure 2: Confusion matrix for three classes

4 Conclusions

In this work we have presented the first results of a CNN-based methodology for eye glance prediction using a web camera with the aim to help in autism diagnosis. The system recognizes three gaze directions and works on a desktop PC. We show that our proposed method achieves a high classification accuracy of 96.01% for testing. Furthermore, the system shows a real-time response of about 90 ms. The previous results demonstrate that the proposed system is a useful, fast, effective and accessible autism diagnosis tool.

References

[1] Natacha Akshoomoff, Christina Corsello, and Heather Schmidt. The role of the autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. *The California school psychologist : CASP*, 11:7–19, 2006. 17502922[pmid].

[2] American Psychiatric Association. Autism spectrum disorder. *In: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, American Psychiatric Association, Arlington, VA 2013. P*, 50, 2013.

[3] Eunji Chong, Katha Chanda, Zhefan Ye, Audrey Southerland, Nataniel Ruiz, Rebecca M. Jones, Agata Rozga, and James M. Rehg. Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):43:1–43:20, September 2017.

[4] Flavio Cunha and James Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, May 2007.

[5] Orla Doyle, Colm P. Harmon, James J. Heckman, and Richard E. Tremblay. Investing in early human development: Timing and economic efficiency. *Economics & Human Biology*, 7(1):1 – 6, 2009.

[6] Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders-autism and developmental disabilities monitoring network, united states, 2006. *MMWR Surveillance Summaries*, 58:1–20, 2009.

[7] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 675–678, New York, NY, USA, 2014. ACM.

[8] Warren Jones, Katelin Carr, and Ami Klin. Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers With Autism Spectrum Disorder. *JAMA Psychiatry*, 65(8):946–954, 08 2008.

[9] Ami Klin, David J. Lin, Phillip Gorrindo, Gordon Ramsay, and Warren Jones. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, 459(7244):257–261, 2009.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[11] Wenbo Liu, Ming Li, and Li Yi. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism research : official journal of the International Society for Autism Research*, 9, 04 2016.

[12] Karen Pierce, David Conant, Roxana Hazin, Richard Stoner, and Jamie Desmond. Preference for Geometric Patterns Early in Life as a Risk Factor for Autism. *JAMA Psychiatry*, 68(1):101–109, 01 2011.

[13] Karen Pierce, Steven Marinero, Roxana Hazin, Benjamin McKenna, Cynthia Carter Barnes, and Ajith Malige. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biological Psychiatry*, 79(8):657–666, Apr 2016.

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, Dec 2001.