Autism Spectrum Disorder Recognition with Deep Learning

Jongmin Shin, Jinwoo Choi
Kyung Hee University
shinpaul14@khu.ac.kr

Abstract

Since it is common to have touch-screen devices, it is less challenging to draw sketches anywhere and save them in vector form. Current research on sketches considers coordinate sequence data and adopts sequential models for learning sketch representation in sketch understanding. In the sketch dataset, it has become customary that the dataset is in vector coordinate format. Moreover, the popular dataset does not consider real-life sketches, sketches from pencil, pen, and paper. Art psychology uses real-life sketches to analyze patients. ETRI presents a unique sketch dataset for sketch recognition of autism spectrum disorder in pixel format. We present a method to formulate the dataset for better generalization of sketch data. Through experiments, we show that pixel-based models can produce a good performance.

1. Introduction

A sketch is a universal tool for communication and visualization. It is a tool for analyzing a person's mental condition or status in art psychology. In this paper, our study aims to implement deep learning in screening autism spectrum disorder scenarios applicable using art psychology. The task is to classify the given sketches into two classes, autism spectrum disorder (ASD) or typically developing (TD). Autism spectrum disorder, known as ASD, is a neurological and developmental disorder that affects how people interact with others, communicate, learn, and behave. People diagnosed with ASD has a different approach to normal conceptual understanding of learning, moving, and focusing[11]. The symptoms of ASD appear at an early age of youth. The characteristic of ASD children in art psychology in screening ASD shows non-verbal and spatial expression.

ETRI has developed a sketch medical dataset for ASD screening. The data is from children in the range of 5 to 12 years old, 34 ASD diagnosed children, and 36 non-disabled children. The gathered data was filtered to balance the data between ASD and TD. The dataset contains drawing a person tasks and free drawing tasks to examine the drawing characteristics of ASD children and non-disabled children. We have split the data into two different datasets in the provided dataset since the domain gap between the drawing a person task and the free drawing task is too big.

In the drawing person task, the subject is asked to draw male and female. Furthermore, in the free drawing task, a subject can draw any drawing, which usually ends up as a person, building, objects, and signs. The free drawing dataset contains 58 sketch drawings, and it is split into two classes, ASD and TD. Each sketch is drawn from a single subject, which means every drawing is separately drawn. The person drawing dataset contains 100 sketch images split into two classes. In each class, 50 sketches are drawn from 25 subjects, where each subject draws two sketches of male and female.

In splitting the data to train and test set, the typical norm randomly splits the dataset. However, we implement ID split in our datasets, splitting the dataset by subject ID or number. Due to the characteristics of the person drawing dataset, each subject drew two sketches, male and female. Splitting the data set by subjects allows the model to generalize sketches in each class better rather than randomly selecting train and test data.

Our sketch dataset is different from popular sketch datasets such as QuickDraw[4] and Tu-Berlin[3]. These two large sketch datasets are drawn from tablets which can be taken as 2D images or sequential data. A sketch is formed with multiple strokes, and it consists of a sequence of data points. The models that deal with sequential data are commonly used in the two datasets. Unlike the popular sketch datasets, our datasets are not from tablets nor consist of sequential data points. With the difference in the formulation process of the datasets compared to sequential data-based datasets, we could not test the dataset in recent sketch recognition models [2,7,8,9] that take sequential coordinates as input.

2. Method

Our dataset is evaluated with Resnet[5] and Vision Transformer, ViT [6] models pre-trained on ImageNet. Models known for image classification are selected since the dataset does not have sequential coordinates and is a pixel-based dataset. We did not evaluate with Sketch-a-Net[1], a CNN model designed explicitly for sketches, due to result in [2]. Where Resnet18 and 50 have higher top-1 and top-5 accuracy than Sketch-a-Net. Moreover, the gap between the sequential and pixel-based models is not unrecoverable to only use sequential models for sketch recognition tasks. Due to the constraints of medical datasets, only data augmentation was resizing the input data into 224 by 224. Adding additional data augmentation could transform the characteristics of a specific class. The last layer of the model changed the output number to 2 since we have only two classes. To optimize the model, we adopted an SGD optimizer with a learning rate of 0.001. For loss, a cross-entropy loss is used.

The dataset was split by 80% train and 20% test by ID split method. In the dataset, sketch images are separated by each class, and the naming convention of each ASD sketch is "A 인 15-001- 001". If the image starts with "A," it means it is an ASD subject's sketch, and the number represents the ID number of the subjects. ID split randomly selects a number in the range of the total subject number in each class. For example, only 25 subject IDs exist in each class in the Person dataset, where ID split would randomly select 5 IDs for a specific class test set. Therefore, the person drawing dataset will be split into 80 sketch images in the training set and 20 sketches in the test set. The free drawing dataset training set contains 46 sketches, and the test set has 12 sketch images. Considering the number of sketches in the dataset, we implement a 5-fold cross-validation of the ID split train and test set.

Model Name	Dataset	Split method	Averaged Top-1 Accuracy
Resnet18	Free drawing	ID Split	96.35 (±04.82)
Resnet34	Free drawing	ID Split	93.24 (±05.17)
Resnet50	Free drawing	ID Split	95.01 (±04.54)
Resnet101	Free drawing	ID Split	95.19 (±04.64
Resnet152	Free drawing	ID Split	95.1 (±06.00)

Table 1. Mean averaged accuracy of performance on Free drawing dataset with Resnet

Model Name	Dataset	Split method	Averaged Top-1 Accuracy
vit_b_16	Free drawing	ID Split	97.15 (±03.53)
vit_b_32	Free drawing	ID Split	97.60 (±04.79)
vit_l_16	Free drawing	ID Split	95.9 (±03.40)
vit_l_32	Free drawing	ID Split	93.49 (±08.30)

Table 2. Mean averaged accuracy of performance on Free drawing dataset with ViT

3. Experiment Results

From the result in table. 1, Resnet18 outperformed other Resnet models with deeper networks, such as Resnet34, 50, 101, and 152. Resnet18 has the highest averaged top1 accuracy. An averaged top-1 accuracy of 5 different variations of the train test set. Compared with Resnet152, about 1% higher, and the standard deviation is 1% lower. In table. 2, we see a similar pattern in the result, with larger models underperforming compared to smaller models where vit_l_32 has the lowest averaged accuracy and highest standard deviation. Compared with the best result in vision transformer models, vit_l_32 is 4% lower, and the standard deviation of 4% higher.

Model Name	Dataset	Split method	Averaged Top-1 Accuracy
Resnet18	Person drawing	ID Split	78.94 (±08.92)
Resnet34	person drawing	ID Split	79.80 (±06.00)
Resnet50	person drawing	ID Split	69.26 (±09.35)
Resnet101	person drawing	ID Split	78.50 (±12.50)
Resnet152	person drawing	ID Split	76.46 (±08.59)

Table 3. Performance of ASD sketch recognition tasks on ASD person drawing dataset with Resnet

Model Name	Dataset	Split method	Averaged Top-1 Accuracy
vit_b_16	Person drawing	ID Split	74.99 (±11.28)
vit_b_32	Person drawing	ID Split	85.75 (±07.41)
vit_l_16	Person drawing	ID Split	88.84 (±04.54)
vit_l_32	Person drawing	ID Split	73.42 (±05.79)

Table 4. Performance of ASD sketch recognition tasks on ASD person drawing dataset with Vit

The result of the person drawing dataset is summarized in table 3, and table 4. We further compare various types of models in person drawing dataset. One similar pattern between the free drawing dataset result and person drawing dataset is model with the largest parameters underperforms and usually has the lowest averaged accuracy. If free drawing dataset, all the results have a higher average accuracy of over 90%. However, person drawing dataset no result is higher than 90%, and the standard deviation is higher. Through table 3 Resnet18 and 34 have highest accuracy of 78.94% and 79.8, compared to other larger Resnet models. On the other hand, in table 4, the result of ViT models shows a different pattern than the Resnet models. The smallest model and largest models have a low accuracy of under 75%. In contrast, vit_l_16 achieved an average accuracy of 88.84% and the lowest standard deviation of 4.54%.

From all the results, we can see vision transformer model performed better than the Resnet models. According to the [10], vit models are better than CNN(convolutional neural network) models in shape recognition. Especially in the person drawing dataset where the sketch is only drawn with a pencil, large ViT models with less texture bias and high shape emphasis show why ViT large 16 has the state-of-art result in this dataset.

4. Conclusion

In this work, we tested a unique sketch dataset with off-the-shelf image classification models to learn sketch representation for the autism spectrum disorder sketch recognition task. Points out the difference between popular sketch datasets and our ASD sketch datasets. Our dataset is much harder to learn due to being a pixel-based dataset and a real-life sketch on paper. The results on the ASD sketch dataset show that models learning sketch representation from pixel-based sketch datasets can compete with sequential models. Furthermore, using data augmentation methods on person drawing dataset can potentially increase the averaged accuracy, which can be studied in future.

5. Acknowledgements

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2019-0-00330, Development of AI Technology for Early Screening of Infant/Child Autism Spectrum Disorders based on Cognition of the Psychological Behavior and Response)
This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1070997).

References

[1] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy M Hospedales, and Chen-Change Loy. Sketch me that shoe. In Proceedings of the IEEE Conference on 9 Computer Vision and Pattern Recognition, 2016. 2, 4.1, 1, 4.1

[2] H. Lin, Y. Fu, X. Xue, and Y.-G. Jiang, "Sketch-BERT: Learning sketch bidirectional encoder representation from transformers by selfsupervised learning of sketch gestalt," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 6758-6767.

[3] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? SIGGRAPH, 2012. 4.1

[4] David Ha and Douglas Eck. A neural representation of sketch drawings. In ICLR, 2018. (document), 1, 2, 3.1, 4.1, 4.1, 4.4

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016. 4.1, 4.1

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In ICLR, 2021.

[7] David Ha and Douglas Eck. A neural representation of sketch drawings. In ICLR, 2018. (document), 1, 2, 3.1, 4.1, 4.1, 4.4

[8] Sepp Hochreiter and Jurgen Schmidhuber. Long short- ¨ term memory. Neural computation, 9(8):1735-1780, 1997. 2, 4.1, 4.1

[9] P. Xu, C. K. Joshi, and X. Bresson, "Multi-graph transformer for freehand sketch recognition," 2019, arXiv:1912.11258.

[10] Muzammal Naseer, Kanchana Ranasinghe, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2021. Intriguing Properties of Vision Transformers. arXiv preprint arXiv:2105.10497 (2021).