Mobile-based Clock Drawing Test for Detecting Early Signs of Dementia

Hongchao Jiang[1], **Yanci Zhang**[2], **Zhiwei Zeng**[3], **Jun Ji**[3], **Yu Wang**[4], **Ying Chi**[4], **Chunyan Miao**[1,2,3]
[1]Alibaba-NTU Singapore Joint Research Institute (JRI), Nanyang Technological University (NTU), Singapore
[2]School of Computer Science and Engineering (SCSE), NTU, Singapore
[3]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY), NTU, Singapore
[4]Alibaba Group, Hangzhou, China
{hongchao001, yanci001, i160001}@e.ntu.edu.sg, {tonggou.wangyu, xinyi.cy}@alibaba-inc.com

Abstract

Dementia is one of the major causes of disability and dependency among older people. Early detection is the key for preserving the quality of life of the patients and reducing caring costs. The Clock Drawing Test (CDT) is commonly used by clinicians to screen for early signs of dementia. We build an automated CDT that runs on mobile platforms, enabling convenient and frequent self-monitoring and testing at minimal costs. Our system combines both a spatial-temporal approach and a purely image-based deep learning approach to analyze and evaluate the hand-drawn clocks based on established clinical criteria. Our system produces scores that are highly correlated with expert human raters.

(a) Log in (b) Register (c) Draw (d) Score

Figure 1: Screens of our automated Clock Drawing Test.

Introduction

Dementia is a chronic neurodegenerative disease that leads to cognitive deficits like memory loss, poor concentration, and mood changes. There is a new case of dementia diagnosed every 3 seconds, with an estimated global societal economic cost of USD 1 trillion (Prince 2015). There is currently no cure for dementia, hence early detection is crucial for managing the symptoms and containing the associated caring costs. The Clock Drawing Test (CDT) is commonly used for screening dementia as it is quick to conduct and easily understood across different cultures. Test subjects are asked to draw a clock depicting a given time, after which a clinician evaluates the drawing based on established criteria, e.g. the correct placement of digits and clock hands. The effectiveness of CDT can be seen from various studies which report an average recall and sensitivity of 85% for detecting cognitive deficits (Shulman 2000). However, CDT, like most clinical screening tests, requires expert administration and scoring, making them unsuitable for frequent and long-term monitoring. Automating CDT enables a more proactive approach to the early detection of dementia by supporting more frequent, objective, and cost-efficient screening.

In prior works on automating CDT, tablet PC and digitizer stylus are often used to capture spatial-temporal information of the drawing process. (Müller et al. 2019) used only the kinematic information collected from the stylus (e.g., velocity, pressure) to differentiate between healthy individuals and those with dementia. Other approaches try to detect the clock components first, before extracting semantic features based on the clinical scoring criteria. (Souillard-Mandar et al. 2016) used k-means to cluster the drawing strokes to meaningful clock components. Then, features based on the positions of the components are extracted and used to train classifiers like SVM and random forests. (Harbi 2017) built an ontology to describe what a well-drawn clock looks like and created an inference engine with a set of fuzzy rules.

In our work, we take a hybrid approach to automatic scoring of the clock drawings. A spatio-temporal model can make use of temporal order to disentangle overlapped components or components in close proximity. A purely image-based deep learning model learns semantic patterns that allows it to be more resistant to noise. We heuristically combine these two models to achieve better scoring performance. To the best of our knowledge, deep learning methods have not been explored for this task as large-scale patient data is hard to obtain. To tackle this problem, we introduce a method to synthetically generate clock drawings which can be used to train such deep models.

To validate our system[1], we collected a total of 100 clock drawings from human subjects aged from 21 to 55 years old. Our system achieved a high Pearson correlation value of 0.763 with the scores provided by expert human raters when using the same scoring criteria.

[1]Video demonstration: https://youtu.be/E5fNkuffIEw

Automated CDT

The tester first logins to his/her account (Figure 1a). If the test subject is being assessed for the first time, relevant demographic data are recorded (Figure 1b). Clinical assessment scores (MMSE) and diagnosis are also recorded and will be used for model training in the future. The system then proceeds to generate a time for the test subject to draw (Figure 1c). Once the subject is done drawing, our system analyzes the image and outputs a final score based on the quality of the drawing (Figure 1d).

AI Engine

The AI engine (Figure 2) consists of two main modules: (1) segmentation and recognition of clock components, and (2) scoring based on the positions of the detected components.

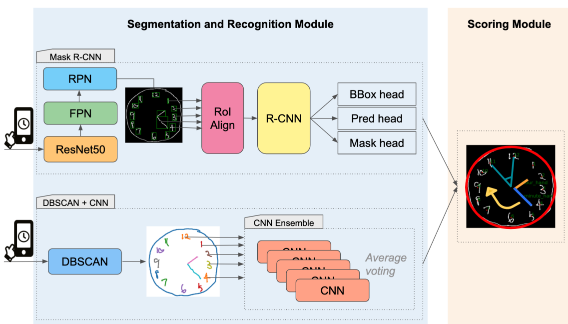

Figure 2: AI engine which uses both spatial and spatio-temporal information for analyzing the clock drawings.

Segmentation and Recognition Module

DBSCAN relies on cues like pauses and transitions during the drawing process to segment the clock components, while Mask R-CNN is data-driven and relies on learning image patterns. We heuristically combine the two models based on the quality of the drawings. The drawing is first processed by DBSCAN due to its ability to effectively segment components in close proximity. However, its performance degrades when there are stray strokes or noise that it cannot meaningfully cluster. In the case when DBSCAN forms an excessive number of clusters (i.e. > 15), we switch to Mask R-CNN instead which is more noise resistant.

DBSCAN + CNN Spatial-temporal data collected from the application is of the form $\mathbf{x}_t = \begin{bmatrix} r_t^x & r_t^y & t \end{bmatrix}^T$, where (r_t^x, r_t^y) is the position of the user's finger on the phone screen at a given time t. We use DBSCAN to cluster all the $\mathbf{x}_t$ from a given drawing episode. Each cluster is seen as a unique clock component. To recognize the digits, we trained a simple CNN with 2 convolutional layers and spatial dropout using the MNIST dataset. Numbers that are not in MNIST (i.e., 10, 11, 12) are generated by combining the single digits (e.g., "1" + "0" = "10"). Several CNNs are trained on this augmented MNIST dataset to form a CNN ensemble.

Mask R-CNN A Mask R-CNN with ResNet50 (He et al. 2016) backbone and Feature Pyramid Network (Lin et al. 2017) is used. To train the model, we created a dataset of synthetic clock drawings. We first extract the clock face outline from the Quick Draw dataset (Jongejan et al. 2017). Then, we calculate the centroid and radius of the outline and use them to determine the placement of clock numbers. The synthetic process gives us multiple degrees of freedom, such as leaving out digits, altering digit sequence, or varying digit positions, to generate drawings over a wide spectrum, from poorly-drawn to well-drawn clocks. We approximate the clock hands using straight lines and placed them at an angle based on the time to be depicted. To increase the variety of our synthetic drawings, ticks and different arrow heads are added to the clock face randomly.

Scoring Module

Our scoring module is largely based on (Mendez, Ala, and Underwood 1992) and (Royall, Cordes, and Polk 1998). It is a rule-based system which evaluates whether certain criteria are satisfied. The criteria are chosen based on whether they are self-contained and easy to translate into programmable rules. For qualitative criteria, such as "numbers have consistent spacing", we find suitable thresholds experimentally that matches human perception of consistency. Other criteria like "time drawn is correct" is more straightforward. The normalized cross product is calculated between the clock hands and each digit, with the smallest value being the digit that the hand is pointing to. We also implement correction mechanisms to account for mistakes made by the detection module. When evaluating the criteria "all digits are in the correct clockwise sequence", we use Levenshtein distance (Levenshtein 1966) to compare our sorted digit sequence and the ground truth. Given a digit that is not in sequence with neighboring digits, we replace the offending digit (e.g., 5-8-7 is corrected to 5-6-7). Some examples of drawings scored by the system can be found in Figure 3.

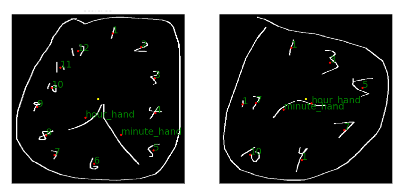

Figure 3: Examples of drawings evaluated by our system: average clock (left, score 33); bad clock (right, score 15).

Conclusion

No cure currently exists for dementia hence early intervention is paramount. Once clinically validated and deployed, our automated CDT could introduce a more proactive approach to dementia screening as seniors can self-monitor their cognitive health with frequent and accessible testing which can help to identify dementia signs at a very early stage. Clinical studies are being planned in our collaborating hospitals in China to further evaluate our automated CDT with healthy and demented seniors.

Acknowledgements

This research is supported, in part, by Alibaba Group through Alibaba Innovative Research (AIR) Program and Alibaba-NTU Singapore Joint Research Institute (JRI) (Alibaba-NTU-AIR2019B1), Nanyang Technological University, Singapore, and by the Singapore Ministry of Health under its National Innovation Challenge on Active and Confident Ageing (NIC Project No. MOH/NIC/COG04/2017).

References

Harbi, Z. 2017. *Automatic interpretation of clock drawings for computerised assessment of dementia*. Ph.D. thesis, Cardiff University.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jongejan, J.; Rowley, H.; Kawashima, T.; Kim, J.; and Fox-Gieg, N. 2017. The Quick, Draw! - AI Experiment. *Google AI Experiments* https://github.com/googlecreativelab/quickdraw-dataset.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, 707–710.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Mendez, M. F.; Ala, T.; and Underwood, K. L. 1992. Development of scoring criteria for the clock drawing task in Alzheimer's disease. *Journal of the American Geriatrics Society* 40(11): 1095–1099.

Müller, S.; Herde, L.; Preische, O.; Zeller, A.; Heymann, P.; Robens, S.; Elbing, U.; and Laske, C. 2019. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *Scientific Reports* 9(1): 1–10.

Prince, M. J. 2015. *World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International.

Royall, D. R.; Cordes, J. A.; and Polk, M. 1998. CLOX: an executive clock drawing task. *Journal of Neurology, Neurosurgery & Psychiatry* 64(5): 588–594.

Shulman, K. I. 2000. Clock-drawing: is it the ideal cognitive screening test? *International journal of geriatric psychiatry* 15(6): 548–561.

Souillard-Mandar, W.; Davis, R.; Rudin, C.; Au, R.; Libon, D. J.; Swenson, R.; Price, C. C.; Lamar, M.; and Penney, D. L. 2016. Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. *Machine learning* 102(3): 393–441.